

**DESENVOLVIMENTO DA PROFICIÊNCIA ESCRITA EM INGLÊS
ESPECIALIZADO:
UM ESTUDO DE *CORPUS* DE *ABSTRACTS* EM CIÊNCIAS DA SAÚDE
PROMOTING PROFICIENCY IN ABSTRACT WRITING: A CORPUS-
DRIVEN STUDY IN HEALTH SCIENCES**

Ana Luiza Pires de Freitas¹
alf@via-rs.net

Maria José BocornyFinatto²
mariafinatto@gmail.com

Resumo: Este artigo apresenta um projeto de pesquisa de doutorado em curso e o respectivo estudo piloto sobre o desenvolvimento da proficiência escrita em língua inglesa a partir da produção de *abstracts* de textos de autoria de estudantes universitários e pesquisadores do campo de Ciências da Saúde com base no uso de itens lexicais. O estudo piloto manipula 112 textos de periódicos internacionais especializados, totalizando 30.371 palavras e 159 textos, incluindo publicações em periódicos internacionais e manuscritos inéditos, produzidos em universidades brasileiras, perfazendo 30394 palavras. O *corpus* foi analisado através do *software* Antconc (ANTHONY, 2011) pela função *clusters/n-grams*, buscando-se ocorrências de pacotes de 4 quatro palavras gráficas. Os resultados da amostra sugerem que os autores brasileiros tendem a se apoiar mais em estruturas pré-fabricadas para construir o discurso acadêmico, divergem do padrão estrutural ao empregarem a mesma frequência de pacotes lexicais e frasais, bem como tendem a adotar expressões que funcionalmente se relacionam à pesquisa e ao texto, mas não ao leitor. Tais constatações preliminares carecem de maior averiguação através do estudo completo, que cotejará um acervo de dados três vezes mais extenso.

Palavras-chave: Inglês para Propósitos Acadêmicos. Proficiência escrita. Pacotes lexicais. Abstract. Linguística de Corpus. Ciências da Saúde.

Abstract: This article introduces a pilot study and the central arguments of an ongoing PhD dissertation on the topic of written proficiency enhancement in English as an additional language based on the use of repeated lexical forms in the production of abstracts in academic texts by students and researchers in the field of health sciences. The empirical study explores 112 texts from international journals adding up to 30371 tokens and 159 texts produced in Brazilian universities amounting 30394 tokens. The texts were run through the Antconc Software (ANTHONY, 2011) in its n-grams/clusters feature, so as to track down 4 graphic word-occurrences of lexical bundles. The outcomes suggest that when producing their texts Brazilian authors tend to find support in a more abundant use of lexical bundles than the international researchers do. National researchers also diverge from the standard structural pattern by adopting the same rate of lexical and structural bundles. Furthermore as they seem to favor the use of extended collocations that functionally relate to research and textual aspects, rather than to the reader. Such preliminary findings should be

¹Docente da Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA) e Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (PPG-Letras-UFRGS).

²Docente e Coordenadora do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (PPG-Letras-UFRGS).

viewed with caution, as they need further investigation in the complete study which will deal with three times more data than this trial did.

Keywords: English for Specific Purposes. Academic writing. Lexical bundles. Abstract. Corpus Linguistics. Health Sciences.

1 Introdução

Este artigo apresenta os pontos principais e um estudo piloto organizado para introduzir a estrutura empírica de uma pesquisa de doutorado, cujo tema é o ensino da produção de *abstracts*³ de gêneros acadêmicos do campo de conhecimentos da saúde em contexto de ensino superior. O objetivo é, além de descrever e analisar a tessitura discursiva a partir da área especializada em que os textos se realizam, produzir um ambiente virtual de aprendizagem (AVA)⁴, na forma de um espaço interativo para construção de conhecimento, no qual o usuário tenha a oportunidade de ser coautor, ao alimentar o conjunto de ferramentas tecnológicas com seus próprios textos, e aprendiz, ao ser levado a qualificar a escrita de modo significativo para suas necessidades acadêmicas. A adoção de um conjunto de recursos tecnológicos em ambiente digital de acesso público visa a promover a sistematização didática das conclusões da pesquisa, bem como a contribuir com os avanços do campo de ensino e aprendizagem de EAP⁵

Nossa pesquisa se orienta pela compilação de um *corpus*, um conjunto de dados criteriosamente selecionados como matéria-prima para o estudo e que direcionam a investigação. É o *corpus*, desse modo, que nos fornecerá insumos sobre o que é mais importante ensinar em se tratando de escrita acadêmica com base nas fontes selecionadas. Partimos, portanto, de fontes que ocupam um lugar no mundo da produção científica a partir das quais construiremos nossas conclusões sobre a configuração da linguagem no contexto sob exame. Para entender a tessitura ou a constituição lexical do acervo, adotamos a noção de *lexical bundles*, em língua portuguesa ‘pacotes lexicais’⁶, que se refere a agrupamentos de

³Do mesmo modo como Cilveti e Pérez (2006), optamos por utilizar o termo *abstract* no lugar de *resumo*, que seria o correspondente em língua portuguesa, uma vez que o último é polissêmico, podendo referir-se ao texto inicial tratado em um trabalho ou, ainda, ao ato de coletar e redigir as ideias principais de um texto.

⁴Os ambientes virtuais de aprendizagem (AVA) integram as diversas ferramentas para auxiliar a criação e a oferta de cursos mediados por tecnologias, surgidas a partir do desenvolvimento da internet e da popularização do uso do computador (BELMONTE; GROSSI, 2010). Ao longo do artigo, apresentaremos o AVA que nossa pesquisa pretende construir.

⁵English for Academic Purposes.

⁶Por questões de objetividade, optamos por abordar o tema de modo sucinto neste texto, embora voltemos a ele nas próximas seções. A tese, no entanto, o abordará de modo detalhado. Decidimos, outrossim, pela adoção do termo pacote lexical, a exemplo de Berber Sardinha (2000) como correspondente a *lexical bundle*, por ser essa a expressão consagrada pelos autores com os quais nos identificamos (BIBER, 2006; BIBER; CONRAD, 1999;

palavras formados por três ou mais elementos gráficos, que costumam se repetir ao longo de um *corpus* com uma frequência determinada, na forma de blocos fixos a integrar o repertório dos usuários de uma língua.

Para atingir os objetivos descritos, a tese propõe dar conta das seguintes questões, cujo escopo pretendemos detalhar ao longo das próximas páginas: 1-Que tipo de variabilidade lexical ocorre entre os gêneros⁷ integrantes do *corpus* pesquisado? Como são os elementos lexicais fixos e os variáveis? 2-Quais são os pacotes lexicais significativamente mais frequentes e como se comportam estrutural e funcionalmente?

É sobre isso que debruça esta pesquisa, voltada para o campo de divulgação de conhecimentos da saúde. A meta é a qualificação de aprendizes de inglês especializado e de pesquisadores brasileiros, já que, mesmo para publicações em língua portuguesa, é uma convenção que os *abstracts* sejam produzidos em uma língua adicional⁸. No campo da saúde, essa língua é o inglês, por se tratar da *língua franca* para as publicações da área. Por extensão, pretendemos que o estudo também possa qualificar o ensino de língua inglesa em contexto acadêmico, especialmente em uma época em que o nosso país está em franca expansão de acordos de cooperação internacional no ensino superior e o lema dentre os pesquisadores é *publish or perish*⁹.

Em sequência, desenvolvemos o texto com a seguinte configuração: a seção 2, a seguir, introduz a problemática que gerou o estudo, a seção 3, traz o foco da pesquisa, a seção 4, descreve como o *corpus* se desenha, apresenta a metodologia da investigação e oferece ao leitor o modelo que idealizamos para nosso AVA, a seção 5, estabelece os aportes teóricos norteadores, a seção 6, disponibiliza o pequeno estudo piloto e, por fim, a seção final, tece algumas conclusões com base no trabalho desenvolvido até aqui.

BIBER et al., 1999; BIBER et al., 2004; CONRAD; BIBER, 2005; CORTES, 2004; HYLAND, 2008a, dentre os principais). Não obstante, existe grande divergência de denominações e definições na literatura sobre combinações de unidades lexicais.

⁷A noção de gêneros textuais, de acordo com Ciapuscio (2003), origina-se da conceitualização de gênero discursivo na obra de Bakhtin (1997), que propôs a existência dos gêneros do discurso e os definiu como “tipos relativamente estáveis de enunciados”. No campo da Análise do Discurso e ESP, a noção foi apresentada por John Swales (1981), fortemente relacionada aos conceitos de propósito comunicativo e comunidade discursiva. Esse autor define gênero como “uma classe de eventos comunicativos compartilhada pelos membros de uma comunidade” (SWALES, 1993, p. 58).

⁸O termo *língua adicional* refere-se à aprendizagem e ao uso de qualquer língua, outra que não a materna, que promove o acréscimo de práticas sociais. A adoção desse conceito reflete a necessidade de se entender a língua como um artefato cognitivo, cultural e social (HELLERMAN, 2008), em oposição à visão cognitivista tradicional que gera a dicotomia *língua estrangeira* versus *segunda língua*. Embora reconheçamos que o termo *língua estrangeira* ainda seja amplamente adotado como paradigma nas pesquisas em Linguística Aplicada, ao longo desta investigação, serão respeitadas as referências às fontes citadas, mas adotada a denominação *língua adicional* para as considerações de nossa autoria.

⁹Publique ou pereça.

2 Situando o problema da pesquisa

Para aqueles de nós com familiaridade no âmbito das práticas sociais em leitura ou mesmo em escrita acadêmicas, geralmente pesquisadores, professores ou estudantes, que partilhemos saberes próprios de nossas áreas de conhecimento, o contato com um texto acadêmico comporta expectativas quanto ao formato ou quanto ao modo como o conteúdo costuma ser desenvolvido. Ao longo de nossa interação com o mundo escrito especializado, vamos construindo competência para o reconhecimento das convenções, regras e normas que regem os *contratos* da ciência. Estudiosos do campo das práticas textuais denominam tal experiência de letramento (ROJO, 2012) ou de leitura de mundo (FREIRE, 1989). Quanto mais ampla seja a condição de letramento de alguém, tanto mais vasta é a possibilidade de exercer sua cidadania e atuar com plenitude no mundo em que vive. No contexto acadêmico, a capacidade de redigir e compreender pesquisas em língua inglesa de modo fluente e adequado para os princípios das comunidades a que estamos atrelados são habilidades importantes, para que nos tornemos membros plenos dessas comunidades e para que nossas formas de registro se tornem instrumento de nossa participação no mundo do conhecimento especializado. Esse mundo, de acordo com Hoffmann (2015), culmina por sofrer uma rígida formalização nas últimas décadas, por conta, especialmente, das exigências do ensino superior e das revistas especializadas, como base nas demandas dos gêneros textuais relacionados a esses contextos. Assim, “a produção exemplar de importantes especialistas, também conduziu à elaboração de planos de construção textual relativamente fixos”(HOFFMANN, op. cit, p. 28). É nessa conjuntura que surge nosso texto alvo.

O *abstract*, síntese do texto científico, seja de um artigo, de uma tese ou dissertação, é um gênero importante no domínio acadêmico, funcionando quase como um *cartão de visita* dos textos que precede. Atualmente, mesmo em periódicos especializados de acesso restrito, o *abstract* é de domínio público¹⁰, e sua leitura é importante por permitir que o leitor obtenha uma visão geral do trabalho e identifique seus tópicos de modo objetivo. A partir dessa identificação, o leitor decidirá, inclusive, se o texto do artigo, na íntegra ou em alguma parte em especial, merece ser acessado e lido.

Em função do propósito de “sumarizar, indicar e predizer o conteúdo e a estrutura do texto integral que segue” (MOTTA-ROTH e HENGES, 2013, p. 152), mesmo quando o artigo completo está disponibilizado para consulta, é comum ler-se primeiramente o *abstract*,

¹⁰Destaque-se que grande parte dos periódicos especializados da área da saúde são acessíveis somente sob assinatura.

que guia o leitor através do texto, além de influir e facilitar suas escolhas de pesquisa mais aprofundadas. Outra característica importante desse gênero de texto destacada por Swales e Feak (2009) é que, quando trabalhos científicos são submetidos a avaliações em eventos acadêmicos, os *abstracts* são balizadores de aceite ou rejeição das propostas.

Apesar de serem textos curtos, os *abstracts* são textos de alta densidade em função de agregarem um grande número de informações. Geralmente compostos em um só parágrafo, cabe-lhes sintetizar as pesquisas que introduzem de modo claro e adequado, a fim de serem tanto compreendidos quanto acolhidos pelas comunidades científicas especializadas. Além disso, produzir um *abstract* eficiente em um língua adicional inclui muitas questões que transcendem a escolha das palavras ou das estruturas gramaticais. Como apontam Dayrell (2010) e Hyland (2009), a redação de *abstracts* proficientes em outra língua está relacionada “à adesão às convenções e características do discurso¹¹ acadêmico”.

A globalização e conseqüente processo de expansão das fronteiras tecnológicas e científicas também vêm motivando a expansão dos estudos em EAP voltados para o campo das Ciências da Saúde, assim como a especialização das metodologias de ensino da língua e as pesquisas em Linguística Aplicada. Acrescente-se que o interesse por orientar e auxiliar alunos na escritura de registros de pesquisa fomenta a possibilidade de divulgação dessas pesquisas internacionalmente (SWALE e FEAK, 2009), quer através de publicações de artigos, quer através de acesso a congressos. John Swales (2004), um dos principais estudiosos do campo de EAP, preconiza a importância dos gêneros de textos acadêmicos serem amplamente descritos, a fim de que se possam desenvolver materiais de aprendizagem e metodologias adequadas ao ensino.

Gatehouse (2001), pesquisador de gêneros acadêmicos em Ciências da Saúde, apontou três focos importantes para o sucesso comunicacional quando se desenvolve um currículo nesse segmento: o trabalho com o léxico especializado a partir do contexto de uso, o contraste desse léxico em relação à língua geral e a prática de uma gama de habilidades acadêmicas de modo integrado. Nos contextos de ensino superior de Ciências da Saúde de nosso conhecimento, todavia, as abordagens mais frequentes para a aproximação dos gêneros especializados ainda são desenvolvidas por conta da leitura de textos com vistas à compreensão geral e/ou abordagens impressionísticas para estimativas de padrões de uso adequados do discurso científico. O trabalho a partir da conjugação de aspectos empíricos

¹¹Esta pesquisa segue a linha de Cabré (2002), Ciapuscio e Kuguel (2002) e Guimarães (2006) no que tange a não estabelecer diferenças entre texto e discurso. Por questão de espaço, não discutiremos aqui esta opção.

com base em *corpora*, estudando como o léxico se comporta, tomando por base as necessidades de aprendizagem específicas dos participantes ainda está aquém de suas possibilidades nos Cursos do campo da saúde no Brasil. Desse modo, há aspectos lacunares a serem explorados no que tange ao cruzamento entre EAP e Ciências da Saúde, especialmente em termos de estudos que enfoquem questões probabilísticas de adoção de padrões lexicais e contexto de uso, com vistas a qualificar tanto os programas brasileiros de ensino de EAP, quanto a formação de educadores linguísticos. Afinal, escritores competentes no registro acadêmico em inglês, participantes de forma ativa e crítica nos eventos internacionais, terão melhores condições de tomar parte na rede de comunicação e ocupar espaço na comunidade científica internacional.

3 O foco da pesquisa

A estrutura que opera para inteirar o leitor sobre a pesquisa descrita e mesmo para convencê-lo a ler o texto completo, a partir da qual o *abstract* se organiza, já foi descrita por Swales (1990) como estabelecida a partir de um formato padrão, não no sentido limitante, mas em um sentido organizacional, até mesmo porque os *abstracts* em Ciências da Saúde apresentam adaptações em relação aos de outras áreas de conhecimento acadêmico. De acordo com Motta-Roth e Henges (2013), esses textos-síntese dos trabalhos científicos podem ser construídos como resposta às seguintes questões: “Por que o estudo foi realizado? Que conceito(s) é(são) central(is) no trabalho? Como o estudo foi realizado? Que resultados foram obtidos? Qual a significação desses resultados para a área?” (p. 24). Perguntas essas que o modelo proposto por Swales (1990; 1994) divide sucintamente nas seguintes partes: (1) introdução, (2) propósito¹², (3) metodologia, (4) resultados e (5) conclusões. É importante que os autores desses escritos sejam capazes de identificar tal caracterização e reconhecer os itens lexicais que os materializam. Assim, uma vez que assumimos que as línguas humanas seguem um princípio previsível (SINCLAIR, 1991)¹³, é de nosso interesse conhecer aspectos característicos dessa previsibilidade no contexto de nosso estudo. Afinal, os termos¹⁴ e grupos de palavras adotados para expressar o conhecimento especializado, assim como a organização dos textos que os contêm, não são obra da livre escolha de quem produz esses registros. Por

¹²Destaque-se que esta ordem não é necessariamente fixa e que, nos textos de nosso *corpus*, introdução e propósito, na maioria das vezes, são o mesmo item.

¹³Noção a ser desenvolvida na seção 5 deste texto.

¹⁴Toma-se *termo* aqui como unidade formada por uma ou várias palavras que designa um conceito de forma unívoca no interior de uma domínio de conhecimento (KRIEGER e FINATTO, 2004, p. 87).

isso, propõe-se, em especial, identificar elementos fixos e variáveis, assim como as sequências de palavras mais frequentes nos contextos estudados. Esses, por sua vez, são elementos-chave para auxiliar aprendizes da escrita e leitura em língua inglesa a desvendar a produção de *abstracts* proficientes.

De modo mais específico, a tese de doutoramento que estamos desenvolvendo se encaixa em uma verificação empírica do impacto de padrões recorrentes na estrutura textual e por isso busca identificar “as formas lexicais mais frequentes encontradas em um registro”¹⁵ (BIBER et al., 2004, p. 376)¹⁶, denominadas pacotes lexicais ou *lexical bundles*, como mencionamos na introdução. Tais colocações estendidas, na maioria das vezes, apresentam significado não idiomático, ou seja, podem ser compreendidas pela coordenação das partes (a exemplo de *in order to*, *in the case of* e *on the other hand*). Tratam-se, portanto, de grupamentos de palavras que costumam acontecer no discurso mais constantemente do que o acaso determinaria e operam para conferir coerência aos textos sem, no entanto, constituírem estruturas gramaticais completas. Pacotes lexicais são reconhecidos na forma de sequências de três, quatro, cinco ou mais palavras contíguas que ocorrem com uma frequência determinada (BIBER et al., 1999), segundo a tradição dos estudos de Gêneros Acadêmicos (BIBER et al., 1999; SWALES, 1990) e da Linguística de *Corpus* (LC) (MCENERY et al., 2006) pelos quais nos orientamos. Conrad e Biber (2005) argumentam, inclusive, que seria interessante ensiná-las no formato de um só bloco, como se funcionassem do mesmo modo que uma palavra.

Por fim, como desdobramento às respostas às questões postuladas para o estudo, pretendemos organizar um AVA, propondo uma abordagem de ensino para escritura de *abstracts*, partindo da descrição dos aspectos léxico-gramaticais e dos padrões organizacionais levantados. Dito de outro modo, queremos averiguar quais são os modos de dizer dos *abstracts* do campo da saúde encontrados em nosso acervo de textos e sistematizar as descobertas para ensiná-las melhor. Para organizar esse recorte e emprestar à pesquisa os contornos desejados, partimos de entendimentos basilares de que trataremos a seguir: são eles

¹⁵Observe-se que Biber adota aqui a perspectiva de *registro* para analisar o texto, um conceito peculiar as suas reflexões linguísticas. Na verdade, o conceito de registro para variedade de língua situacionalmente definida (BIBER et al., 1999, p. 15) é adotado em muitos estudos em Linguística de Corpus e se distingue do de gênero, uma vez que o primeiro se volta para o nível linguístico do discurso, ou seja, aqui se está falando em restrições linguísticas do léxico, enquanto o segundo, o conceito de gênero, se volta para o objetivo do discurso e para as suas restrições discursivas em oposição às linguísticas (BIBER e CONRAD, 2009).

¹⁶“The most frequent recurring lexical frequencies in a register” (BIBER et al., 2004, p. 276). Todas as traduções apresentadas ao longo deste texto são de nossa inteira responsabilidade.

a visão de língua e a concepção de ensino. É com base nesses dois entendimentos que estabelecemos o pressuposto conceitual ao qual nos afiliamos para sustentar a investigação.

Primeiramente, adotamos a concepção de língua na condição de atividade social, histórica e cognitiva (BAKHTIN, 1997)¹⁷, que significa entender a relação com a escrita acadêmica como uma atividade interativa verbal, voltada para os pares na comunicação (KOCH, 1995). Também compreendemos que nosso objeto de estudo vem sendo moldado ao longo do tempo por um grupo de participantes que compartilham de uma comunidade que estabelece parâmetros para a sua atividade e seu discurso (SWALES, 1990). Além disso, a condição de atividade cognitiva implica que o texto acadêmico representa um todo de sentido por meio do qual o conhecimento especializado se manifesta (HOFFMANN, 2004).

O segundo entendimento é nossa visão de ensino e, por conseguinte, de aprendizagem, cujo ponto de chegada é o desenvolvimento da competência comunicativa (HALLIDAY, 1973), para que o aprendiz possa agir no mundo que, em nosso contexto, é o mundo do conhecimento especializado. Embora a língua seja a mesma, ela assume um contorno distinto na comunicação especializada. A fim de promover proficiência linguística, um construto variável baseado nas necessidades do interlocutor (MACNAMARA, 2011), é preciso que o ensino se volte para desenvolver competência estratégica (CANALE & SWAIN, 1980), no intuito de levar o usuário a dar conta das diferenças que separam o texto da língua geral e o da língua de especialidade. Não se trata exclusivamente de ensinar o léxico especializado, porque esse léxico só faz sentido quando analisado em relação ao seu pano de fundo, ao seu contexto de uso (HUNSTON, 2002), justamente como estamos propondo analisá-lo aqui, para, por fim, ensiná-lo posteriormente com base nos achados da investigação.

Por fim, a hipótese que assumimos para construir a investigação parte da relação com a estrutura retórica do texto em questão. Ao nos aproximarmos do termo *retórica*, somos remetidos à arte do convencimento aristotélico, que é, naturalmente, parte do propósito do texto científico. Podemos mesmo dizer, no caso desse tipo de texto, que “alguns dos objetivos de uma boa retórica seriam dar credibilidade às proposições e convencer o leitor do que está sendo apresentado, além de organizar o texto de *forma coerente e coesa*”¹⁸ (POSSAMAI, 2004, p. 173). A esse respeito, Hoffmann (2015) postula inclusive que, com a especialização das sociedades contemporâneas, as pesquisas em linguagens especializadas têm presenciado um retorno à retórica em conjugação a novos procedimentos de estatística linguística.

¹⁷ Entendimento compartilhado por Douglas Biber (1988) e John Swales (1990).

¹⁸ Grifo nosso.

Pela perspectiva da Teoria da Estrutura Retórica (MANN & THOMPSON, 1988), um texto, em nosso caso um *abstract*, estrutura-se de modo *coerente* em função de existir um objetivo para suas partes, alguma razão plausível para sua inclusão, que é evidente aos olhos de um leitor experiente. Em se tratando de uma língua adicional, diríamos que há um motivo para a inclusão de cada parte aos olhos de um leitor *fluyente*. Há, desse modo, uma função desempenhada por cada segmento que compõe o texto. Pela lente de ESP ou dos estudos de gêneros acadêmicos desenvolvidos por Swales (1990), os textos apresentam modelos que ordenam sua estrutura e que servem para orientar os membros especialistas da comunidade discursiva que os adotam, a fim de colocar em prática seus objetivos compartilhados. As motivações dos membros, por sua vez, formatam a estrutura esquemática do texto e influenciam, assim como delimitam, as escolhas de conteúdo e de estilo. No caso do *abstract*, Swales (op. cit.) chamou atenção para sua arquitetura, a que nos referimos acima como *estrutura retórica*, a qual serve para sintetizar as partes principais do estudo a que o *abstract* se refere.

Isso posto, a crença a partir da qual a investigação em curso se fundamenta é de que, para realizar os propósitos de proficiência discursiva, o aprendiz brasileiro da língua inglesa precisa ser capaz de construir *abstracts* que traduzam uma estrutura retórica fluente, em que cada parte ou item lexical represente um valor para formar um todo de sentido. Em outras palavras, parece razoável argumentar que aprender a combinar adequadamente as palavras para construir textos em um determinado gênero escrito pode ajudar o acadêmico a fazer escolhas linguísticas apropriadas e, com isso, desenvolver referências acerca do tipo de linguagem que se espera que produza. Porque, só desse modo, ele poderá ser membro pleno da comunidade referida por Swales.

É em função dessa premissa, que correlaciona proficiência escrita acadêmica em língua inglesa com a escritura de *abstracts* adequados para um campo específico de conhecimento em questão, que defendemos os aspectos eleitos para análise como candidatos a alavancar um ensino e uma aprendizagem de ESP mais qualificados.

4 Os materiais e métodos

Primeiramente, organizamos nosso *corpus* de estudo, denominado *Corpus CISA* (nomeado pela aglutinação das palavras *ciência e saúde*), compreendendo *abstracts*¹⁹ dos campos de Farmácia, Medicina e Nutrição a partir de:

- a) artigos científicos publicados nos periódicos *Annals of Internal Medicine*, *American Journal of Clinical Nutrition* e *British Journal of Pharmacology*, selecionados a partir do critério de avaliação *Qualis A* pelos parâmetros da CAPES²⁰, bem como em função de estarem disponíveis eletronicamente de forma gratuita;
- b) publicações dos pesquisadores docentes da Universidade Federal da Ciências da Saúde de Porto Alegre (UFCSPA), da Universidade Federal do Rio Grande do Sul (UFRGS), da Universidade Federal de São Paulo (UNIFESP) e da Universidade Federal Paulista (UNESP) com base no banco de dados dessas universidades;
- c) pesquisas dos alunos dos Programas de Pós-Graduação das referidas universidades;
- d) trabalhos de conclusão dos Cursos de Graduação de discentes das referidas universidades.

O critério para a organização do *Corpus CISA* é da ordem de representatividade e abrangência, uma vez que compreende produções textuais de diferentes gêneros, quais sejam artigos, teses, dissertações e monografias de conclusão de Curso. Além disso, a adoção de quatro universidades brasileiras com estudos no campo da saúde enriquece e diversifica o acervo de textos. Defendemos que, uma vez que o gênero *abstract* está presente em diferentes tipos de texto, é relevante investigar suas semelhanças e diferenças. De mais a mais, os textos científicos escolhidos são igualmente representativos das comunidades acadêmicas em que atuamos e que, portanto, serão prioritariamente beneficiárias das conclusões deste trabalho. Acreditamos, desse modo, que os critérios de organização do *corpus* diferenciam nosso estudo de outros já realizados acerca do tema de organização lexical em contexto acadêmico. Por fim, cumpre que o produto final para o desenho de um *corpus* deve ser balanceado em tipos/variedades/estilos para dar conta do objetivo: o ensino.

A análise do *corpus* de estudo visa a promover um cotejamento entre os quatro *subcorpora*, especialmente em relação às publicações dos periódicos internacionais, uma vez

¹⁹A Associação Brasileira de Normas Técnicas (ABNT) classifica os *abstracts* em indicativos, informativos, críticos ou recensão. *Abstract* indicativo ou descritivo se caracteriza como um sumário narrativo; nesse tipo de texto, descrevem-se os principais tópicos do texto original e indicam-se sucintamente seus conteúdos. *Abstract* informativo ou analítico é o tipo que reduz o texto em tamanho, mantendo-se, porém, as ideias principais. Os tipos informativo e crítico ou recensão devem dispensar a leitura do texto original para o conhecimento do assunto. Para este trabalho, adotamos *abstracts* do tipo informativo, pois, do mesmo modo como Feltrim et al. (2003), defendemos a noção de que um *abstract* representativo de um texto acadêmico deve ser preferencialmente do tipo informativo e descrever os propósitos, a metodologia, os resultados e as conclusões do trabalho.

²⁰Avaliação de 2012.

que a criteriosa revisão demandada com vistas a qualificar os textos para publicação cria um parâmetro de qualidade. Ainda, a sequência pretende espelhar a ordem de análise dos *subcorpora*, uma vez que o estudo visa a explorar os dados com vistas a analisar as diferenças de frequência dos padrões lexicais entre ‘a’ e ‘d’ e, especialmente, contribuir para a qualificação da produção acadêmica de *abstracts* produzidos pelos segmentos ‘c’ e ‘d’.

O *corpus* está sendo organizado na proporção de 330 *abstracts* para a coleção nacional e 330 *abstracts* para a coleção internacional, o que, como base na média dos textos já compilados, gerará aproximadamente 180 mil palavras. Isso posto, tendo em mente que “o tamanho de um *corpus* está longe de ser uma questão tranquila para as pesquisas em LC, e são as perguntas de pesquisa que vão determinar se a extensão do acervo é adequada para dar conta de analisar um fenômeno linguístico” (MCENERY, 2015²¹), defendemos que, não obstante estarmos diante de um *corpus* de pequeno porte, seja possível dar conta das perguntas de pesquisa estabelecidas. Ademais, para chegar a esse montante, analisamos outros trabalhos significativos na exploração do gênero *abstract* com coleções de proporção semelhante a nossa (a exemplo de BERBER SARDINHA, 2003; DAYRELL, 2009b; DAYRELL; ALUÍSIO, 2008; FENG, 2006).

O acervo está sendo preservado a partir de arquivos eletrônicos em formato *.txt*, sendo que cada *abstract* corresponde a um arquivo e cada *subcorpus* a uma pasta. Gera-se a lista de frequência de pacotes lexicais pelo ponto de corte de três ocorrências em pelo menos três textos do acervo em função da limitação do tamanho do *corpus*²². Para as ocorrências, adotamos o critério de quatro palavras gráficas, a exemplo de outras pesquisas que tratam desse mesmo item lexical (BIBER et al., 2004, CHEN; BAKER, 2010, CORTES, 2004; HYLAND, 2008a). Seguiu-se o princípio norteador de que os pacotes lexicais de três palavras são muito frequentes e geralmente se abrigam em sequências de quatro palavras²³ (BIBER; BARIBIERE, 2007; BIBER et al., 1999; CORTES, 2004), pacotes menores do que os de três palavras podem ser fragmentos de texto apenas, enquanto os mais longos tendem a ser mais escassos. O *software* adotado para a análise lexical é o AntConc (ANTHONY, 2011), que se encontra disponível de modo gratuito para o usuário pesquisador.

O trabalho de análise dos pacotes lexicais se orienta por dois eixos: estruturalmente, enfocando a composição lexical (BIBER et al., 1999, pp. 997–1025, estabelecida a partir dos

²¹Comunicação pessoal em Lancaster University, 01/06/2015.

²²Biber et al. (1999) adotam o critério de um mínimo de 10 ocorrências por milhão de palavras em, pelo menos, cinco textos. Seguimos aqui a mesma linha de Dayrell (2009b), que também trabalhou com essa linha de corte para um *corpus* de pequena proporção.

²³A exemplo de *as a consequence of* que abriga *as a consequence*.

padrões mais usuais encontrados pelo autor para pacotes lexicais de quatro palavras gráficas a partir de textos acadêmicos),²⁴ e funcionalmente, em relação à estrutura retórica dos textos (BIBER, 2006 e BIBER et al. 2004, adaptado por HYLAND, 2008a)²⁵. Essas categorias se organizam da seguinte forma:

A. Distribuição Estrutural:

➤ Categoria	Exemplo
1. <i>Noun phrase + of</i>	<i>the end of the, the nature of the</i>
2. <i>Other noun phrases</i>	<i>the fact that the, one of the most</i>
3. <i>Prepositional phrase + of</i>	<i>at the end of, as a result of</i>
4. <i>Other prepositional phrases</i>	<i>on the other hand, at the same time</i>
5. <i>Passive + prep phrase fragment</i>	<i>is shown in figure, is based on the</i>
6. <i>Anticipatory it + verb/adj</i>	<i>it is important to, it is possible that</i>
7. <i>Be + noun/adjectival phrase</i>	<i>is the same as, is a matter of</i>
8. <i>Others</i>	<i>as shown in figure, is likely to be</i>

B. Distribuição Funcional:

➤ Categoria	Exemplo
1. Orientação para a Pesquisa	
a) Local	in the present study
b) Procedimento	the role of the
c) Quantificação	a wide range of
d) Descrição	the structure of the
e) Tópico	in hospitals in Brazil
2. Orientação para o Texto	
a) Marcas de Transição	on the other hand
b) Marcas de Resultado	the results suggest that
c) Marcas de Estrutura	in the present study
d) Marcas de Limites	in the presence of
3. Orientação para o Participante	
a) Características de Posicionamento	are likely to be
b) Características de Engajamento	as can be seen

²⁴Para verificação das categorias lexicais, adotamos o analisador CLAWS (*Constituen Likelihood Automatic Word-Tagging System*) (Lancaster University Center for Computer Corpus Research in Language, 2000).

²⁵A opção pelos modelos adaptados deu-se em função da simplificação, uma vez que Biber (op. cit.) trabalhou com coleções de dados muito mais amplas do que a nossa e, portanto, com muito mais variedades de registros e gêneros.

Outrossim, as combinações de unidades lexicais especializadas de maior incidência, a partir das linhas de concordância para os respectivos nódulos, organizarão as prioridades pedagógicas da ferramenta para escrita acadêmica contida no AVA voltado para ESP²⁶.

A ferramenta de suporte à escrita contida no ambiente projetado tem por meta apoiar a construção de diferentes tipos de *abstracts*, voltados para artigos, monografias de conclusão de cursos de graduação, dissertações, teses e trabalhos para apresentações em eventos do campo da saúde. Caberá ao usuário, assim, eleger o tipo e a subárea para seu *abstract*. Feitas as escolhas, o autor acessará o *template* e, ao concluir seu texto, receberá um *feedback* da própria ferramenta, sem, no entanto, alterações automáticas no texto original, para que o possa verificar e comparar as sugestões feitas. O texto ficará armazenado no sistema e alimentará a base de *feedbacks*, formando um *corpus*. O SCIPO²⁷ (FELTRIM et al., 2003), por exemplo, também é um portal de escrita acadêmica para língua inglesa disponível aos usuários brasileiros, mas nossa ferramenta pretende ocupar um campo de especialização em Ciências da Saúde, além de organizar os recursos pedagógicos de forma generificada, destacando a função social dos textos no mundo dos usuários. Ademais, pretendemos abrir espaço interativo para participação dos consulentes.

Por ora, relacionamos os aspectos que idealizamos como principais componentes do ambiente:

- área de apresentação da ferramenta, seus objetivos e público-alvo;
- tutorial para usuários;
- exemplificação de *abstracts* adequados para a área da saúde a partir de aspectos mais genéricos (modelos de Swales, 1990 e Swales e Feak, 2009) e das três revistas especializadas escolhidas para desenvolvimento do estudo;
- apresentação sucinta do que constituem pacotes lexicais, além de exemplos a partir das principais *lexical bundles* aferidos no estudo, bem como a descrição de aspectos relevantes das dez mais frequentes²⁸;
- inclusão de espaço para um *corpus* embutido;
- área de cadastro de arquivos de *abstracts*;
- preservação do texto original do usuário e, a partir dele, produção da revisão, apontando, em cores, o que pode ser alterado para melhorá-lo;

²⁶Exemplo de AVA desenvolvido por pesquisadores vinculados à UFRGS para estudo de língua alemã: <<http://www.ufrgs.br/virtualern/VL/index2.php>>.

²⁷Disponível em: <<http://www.nilc.icmc.usp.br/~scipo/>>.

- espaço para avaliação da ferramenta e do AVA pelos usuários, bem como um sistema de registro de suas contribuições;
- sistema de correio eletrônico conjugado, informando aos administradores sobre a ocorrência de mensagens e comentários;
- *links* a *sites* com tutoriais para escrita acadêmica, a fim de encorajar o desenvolvimento da aprendizagem e a autoria dos usuários nos processos de escrita;
- espera para vídeos de aulas e para roteiros de apresentações em congressos.

5 Os subsídios teóricos

Alguns dos principais entendimentos que servem como pressupostos da pesquisa de doutoramento em curso são oriundos da Linguística de *Corpus* (LC), a Linguística das Linguagens Especializadas (LLE) os estudos em *English for Academic Purposes* (EAP) e o campo do Processo de Línguas Naturais (PLN). Contextualizemos agora de que forma esses aportes se relacionam para dar forma ao arcabouço teórico que organiza nossa investigação.

Adotamos a metodologia LC especialmente por se tratar de uma investigação empírica, que se ocupa da coleta e análise de um conjunto de dados linguísticos organizados a partir de *corpora*. Também nos valemos dos fundamentos da LC devido ao entendimento de língua como um sistema probabilístico (HALLIDAY, 1991; 1992), verificável através de coleções de dados organizados criteriosamente com o intuito de servirem para a pesquisa de uma língua ou variedade linguística. De acordo com essa noção, é possível apontar e quantificar padrões de regularidade, destacando uma correlação entre tais traços e os contextos situacionais de uso da linguagem. Cotejando esses aspectos, pode-se reconhecer que uma língua não se resume ao preenchimento de espaços vazios de forma incerta, uma vez que o ambiente linguístico atua sobre a co-seleção de itens lexicais. Por essa perspectiva, prevalece que a língua se orienta por um princípio idiomático, postulado por John Sinclair (1991), para quem a seleção de itens seria guiada por padrões maiores do que simplesmente palavras, em oposição ao princípio da livre escolha, segundo o qual seria possível escolher, apenas com restrições gramaticais, os itens que compõem um enunciado. Nos termos desse pesquisador britânico, “um usuário da língua tem a sua disposição um grande número de sintagmas²⁹ semi-

²⁹ Combinação de duas formas ou unidades linguísticas elementares em que uma, funcionando como determinante, cria um elo de subordinação com outra, dita determinada: sintagma lexical, locucional, nominal, verbal, oracional, etc. (Dicionário Online de Língua Portuguesa).

pré-construídos, que constituem escolhas únicas, mesmo que eles pareçam ser analisáveis em segmentos” (SINCLAIR, 1991, p. 120).

Esse princípio mostrou-se inovador e envolveria a seleção simultânea de estruturas maiores, nas quais as palavras não estariam, no caso, sujeitas à discriminação do escritor fluente no domínio de um gênero textual. Ao residir no princípio idiomático, a proposta de Sinclair possibilita a expansão dos padrões fraseológicos de uso da linguagem. Observe-se que é por nos identificarmos com esse princípio que nossa unidade de análise são os pacotes lexicais, “sequências de palavras que costumam ocorrer conjuntamente no discurso” (BIBER, 1990, p. 262).

Acrescente-se aqui um aspecto interessante em relação às investigações em LC: o fato de nosso campo metodológico voltar-se para o uso da língua a partir do critério frequência não explica em si os padrões linguísticos. A frequência, na verdade, aponta aspectos a serem explicados pelo pesquisador.

Partilhamos também da perspectiva investigativa da LLE, pois nos ocupamos da verificação do *modus dicendi* característico de textos de uma área científica. A partir de nossa bagagem de leitura acadêmica, por exemplo, podemos relacionar certos pacotes lexicais, como *need for further studies* como expressão típica da prosa acadêmica, diferentemente de *have a good one*, que caracteriza a conversa cotidiana. Uma particularidade da comunicação entre especialistas em um tema, afinal, é a precisão da informação compartilhada, que se materializa por uma linguagem de especialidade³⁰. Os modos de nomear da ciência remetem à linguagem de especialidade que Hoffmann (1985) define como “o conjunto de todos os recursos linguísticos que são utilizados em um âmbito comunicativo, delimitado por uma especialidade, a fim de garantir a compreensão entre as pessoas que trabalham nesse âmbito” (p. 53). Em outras palavras, há uma preocupação com o todo de sentido de um texto, assim como com as escolhas de termos típicos através dos quais as comunidades discursivas são identificadas. É em função do recorte voltado para o texto, que, por sua vez, extrapola os limites do termo, sem, no entanto, descartá-lo, que a investigação aqui apresentada se identifica com a Terminologia Textual³¹ (CIAPUSCIO,

³⁰Outrossim, concordamos com Cabré (1999), quando essa pesquisadora observa que não é apenas o tema que determina o caráter especializado, mas também a maneira segundo a qual esse tema é abordado, além dos interlocutores e da situação comunicativa, a exemplo dos pesquisadores e da comunidade científica.

³¹Grafada com inicial maiúscula, em oposição à ‘terminologia’, que alude a um conjunto de termos, a designação se refere ao campo de estudos interdisciplinar filiado à Linguística Aplicada, que se ocupa dos fenômenos da comunicação técnica e científica, a partir de uma abordagem prescritiva ou descritiva, conforme a linha adotada. Estudo do vocabulário científico, que toma como paradigma a unidade terminológica, e também, mais contemporaneamente, estudo das práticas textuais e suas convencionalidades, incluindo a comunicação entre leigos e especialistas (FINATTO, EVERS, OLIVEIRA e SILVA, 2010; FINATTO, 2011). Para maior

2003). Ou seja, o texto é visto como “um todo de sentido e de comunicação, constituído pelo conteúdo expresso e também pelo quadro geral das escolhas ou convenções do seu modo de dizer” (FINATTO, 2012, p. 156). O aspecto frequência de classes de palavras que designam os objetos da atividade especializada também é de especial interesse para o estudo das linguagens especializadas, já que é em função da frequência de manifestações linguísticas específicas que as linguagens especializadas podem ser definidas, o que, por sua vez, estabelece um cruzamento com a LC. O foco de atenção em frequência para a LLE abrange o estudo do uso e dos padrões de uso a partir de, entre outros aspectos, combinações de palavras mais ou menos fixas, elementos retóricos e argumentativos, colocações especializadas, condições pragmáticas da comunicação e condições do gênero textual, aspectos que se relacionam à unidade de análise de nosso estudo.

Agregamos, outrossim, o ponto de vista dos estudos em EAP, com base nos postulados de Halliday, MacIntosh e Stevens (1964), dada nossa preocupação particularmente pedagógica com a promoção de usuários fluentes de inglês para finalidades acadêmicas, a partir de um ensino que promova autonomia e empoderamento. Por ser orientada pelos discursos de comunicação especializada, essa área de estudos aplicados se volta para os níveis de especialização dos textos, sua relação com os usuários, os níveis de *expertise* que possuem e os propósitos que perseguem para aprendizagem da língua inglesa nos contextos profissionais ou acadêmicos. Essa abordagem, a propósito, materializa pedagogicamente os estudos de LLE. São também várias as razões pelas quais os estudos de EAP têm demonstrado aproximação com as investigações em LC, a começar pela disponibilidade de exemplos de uso autêntico da linguagem, originárias de situações reais de comunicação e não de sentenças inventadas e, com isso, pouco frequentes. Além disso, a aplicação de uma abordagem de cunho empírico na área de línguas adicionais altera o foco de atenção tradicional do *falante nativo* para um conjunto de dados que representa o uso efetivo da língua que se deseja descrever. Com isso, transpõe-se a barreira de que o *falante não nativo* de uma língua estaria em desvantagem ao realizar um estudo dessa língua. Estando a investigação da linguagem calcada em dados, a necessidade de recorrer a falantes que tenham o inglês como língua materna perde o sentido. De mais a mais, no caso do Brasil, muitos educadores linguísticos se sentem desvalorizados em seus conhecimentos a partir dessa dicotomia clássica entre falantes

aprofundamento das noções, vide Zilio (2010), em ‘Terminologia Textual e Linguística de *Corpus*: Estudo em Parceria’.

nativos e não nativos³². Portanto, a LC também é saudada pelos que se interessam por *EAP* por estar em condições de romper com o paradigma tradicional e fornecer aporte ao trabalho de formação de professores-pesquisadores.

Por fim, em função da produção de nosso objeto de aprendizagem, voltamo-nos para o paradigma aplicado aos problemas de geração e compreensão automática de línguas humanas naturais, que se instaura a partir da área do Processamento de Línguas Naturais (PLN) (NUNES, 2008; NUNES et al. 2010). Sabe-se que as tecnologias da informação e da língua são mediadoras centrais do ingresso nos centros de conhecimento contemporâneos. “Essa temática perpassa a interatividade e contempla aspectos do multilinguismo e do acesso à informação, que promovem a cidadania, o desenvolvimento tecnológico e a redução do abismo digital” (LIMA et al., 2007, p.1).

O gênero *abstract* tem sido alvo de atenção dos pesquisadores do campo de PLN por meio da criação de ferramentas eficazes que podem auxiliar o processo de escrita acadêmica. Dentre as implementações baseadas em *corpus*, são significativas as pesquisas brasileiras de Aluísio et al. (2001; 2005) que desenvolveu a ferramenta AMADEUS e de Feltrim et al. (2003), que apresentam o Sistema SciPo³³ e o Scipo Farmácia (ALUÍSIO et al., 2001; 2005), para suporte à produção textual, as quais têm auxiliado pesquisadores na redação de seus escritos. Igualmente importantes são as pesquisas de Genoves et al. (2007), que desenvolveram ferramenta AZEA³⁴ para detecção automática da estrutura esquemática de *abstracts*, e de Dayrell et al. (2012). Esta última estudiosa, também com a participação do autor da AZEA, aperfeiçoou a ferramenta anterior na forma de um classificador com possibilidade de etiquetagem múltipla (para o caso de determinada sequência de palavras corresponder a mais de um movimento retórico) que foi denominado MAZEA³⁵.

Nossa ferramenta pretende fornecer uma contribuição ao agregar uma combinação das atividades de produção e revisão textual. A existência desses recursos representa o sonho de muitos pesquisadores e sua ausência, um pesadelo, no dizer de Feltrim et al. (2002). Muitos avanços foram feitos no Brasil, graças às possibilidades abertas pelo PLN. Não obstante, Dayrell et al. (2012, p. 1608) observam que há ainda um caminho significativo a percorrer. “Seria particularmente útil desenvolver mais investigações acerca da regularidade nos

³²A trajetória clássica dos estudos aquisicionistas em aprendizagem de línguas estrangeiras postula o falante nativo como parâmetro para o aprendiz, sem considerar aspectos tais como quem é esse falante em termos de formação, que campos de conhecimento domina e de que registro ou dialeto é falante, dentre outros aspectos (RAJAGOPALAN, 2004; WILLIAMS, 2006).

³³Disponível em: <<http://www.nilc.icmc.usp.br/~scipo/>>. Acesso em 08 jun. 2013.

³⁴Argumentative Zoning for English Abstracts.

³⁵Multi-able Argumentative Zoning for English Abstracts.

padrões lexicais e gramaticais dos movimentos retóricos dos *abstracts* em inglês”. Nossa pesquisa de doutoramento pretende agregar esforços para melhor entender tais regularidades e variações, contrastando-as a características de *abstracts* de textos acadêmicos de gêneros variados através da observação de pacotes lexicais recorrentes ou típicos.

6 O estudo piloto

Para desenvolver o estudo piloto que dá suporte à pesquisa em maior escala, trabalhamos com 271 textos do *Corpus CISA*³⁶, 112 dos quais originários dos três periódicos internacionais selecionados, quais sejam *Annals of Internal Medicine*, *American Journal of Clinical Nutrition* e *British Journal of Pharmacology*, e 159 a partir de publicações, teses, dissertações e trabalhos de conclusão de curso da Universidade Federal de Ciências da Saúde de Porto Alegre (UFSCPA) e da Universidade Federal do Rio Grande do Sul (UFCSPA). Para tanto, foi adotada a seguinte distribuição:

1- Acervo internacional:

a) *American Journal of Clinical Nutrition*: 33 textos

b) *Annals of Internal Medicine*: 36 textos

c) *British Journal of Pharmacology*: 43 textos

Total de Palavras: 30371 / Total de Tipos: 4351

2- Acervo Nacional:

a) Publicações (UFRGS: 47 e UFCSPA: 25): 72

b) Teses e Dissertações (UFRGS: 4 e UFCSPA: 24): 28

c) TCCs (UFRGS: 34 e UFCSPA: 25): 59

Total de Palavras: 30394 / Total de Tipos: 4749

Primeiramente, o acervo foi convertido para *.txt* e dividido em dois blocos para o tratamento analítico pelo uso do *software* AntConc (ANTHONY, 2011) por intermédio da função *clusters/n-grams*, que permite a identificação de colocações estendidas. Dadas as dimensões do *corpus* manipulado, que contabilizou apenas 60.765 palavras, foi necessário fazer algumas adaptações no roteiro metodológico, para dar conta do objetivo de levantar indícios do funcionamento do acervo quanto à variabilidade lexical. Uma vez que o quesito frequência é nosso indicador, lidando com um terço do *corpus* da pesquisa completa, não seria possível gerar evidências realistas cotejando *subcorpora* ainda menores de um acervo já de pequenas proporções. Por esse motivo, decidimos por estudar o *corpus* nacional em relação ao *corpus* internacional na rodada analítica aqui apresentada e reservar o cotejamento

³⁶Por questão de limitação de espaço, não foi possível incluir a coleção de *abstracts* integrantes do estudo neste texto.

dos subcorpora brasileiros para a etapa de análise do acervo completo. Os dois blocos de texto foram analisados em termos de distribuição estrutural (BIBER et al., 1999) e retórica (BIBER, 2006 e BIBER et al. 2004, adaptado por HYLAND, 2008a), tendo sido necessário estabelecer algumas categorias estruturais não descritas dentre as mais frequentes, para lidar com as ocorrências encontradas.

Em segmento, apresentamos as ocorrências geradas a partir dos dois blocos de textos, no *layout* original do *software*:

Figura 1: *American Journal of Clinical Nutrition, Annals of Internal Medicine e British Journal of Pharmacology/ Table 1. American Journal of Clinical Nutrition, Annals of Internal Medicine e British Journal of Pharmacology*

Rank	Freq	Range	N-gram
1	4	3	but little is known
2	4	3	little is known about

Fonte: AntConc.

Figura 2: Abstracts do Acervo de Produções UFCSPA e UFRGS/ Table 2. Abstracts from the UFCSPA and UFRGS Electronic Archives

Total No. of N-Gram Types			24	Total No. of N-Gram Tokens	
Rank	Freq	Range	N-gram		
1	15	3	rio grande do sul		
2	12	3	of this study was		
3	11	3	this study was to		
4	9	3	of rio grande do		
5	8	3	the aim of this		
6	7	3	aim of this study		
7	6	3	in relation to the		
8	5	3	as well as the		
9	5	3	grande do sul brazil		
10	5	3	in the field of		
11	5	3	state of rio grande		
12	5	3	study was to evaluate		
13	5	3	the state of rio		
14	4	3	evaluate the correlation between		
15	4	3	in children and adolescents		
16	4	3	mean age was xb		
17	4	3	this study aimed to		
18	4	3	to evaluate the correlation		
19	4	3	was to evaluate the		
20	3	3	conclusions the results of		
21	3	3	it was used a		
22	3	3	the mean age was		
23	3	3	the results of the		
24	3	3	was evaluated using the		

Fonte: AntConc.

Inicialmente, chama atenção o pequeno número de pacotes encontrados, mas é preciso relacionar alguns aspectos atribuídos a essa peculiaridade. Logicamente, a dimensão do *corpus* é fator desencadeador, sendo importante lembrar que os pacotes lexicais presentes em determinado gênero se configuram como tal pela frequência estatística de ocorrência³⁷. Nossa busca aqui se orientou pelo ponto de corte mínimo de três ocorrências a cada três textos de cada acervo trabalhado, para que pudéssemos nos certificar de estarmos realmente lidando

³⁷Hyland (2008c), por exemplo, encontrou 130 pacotes lexicais de quatro palavras gráficas em um *corpus* de 3.5 milhão de palavras de textos acadêmicos. No entanto, por trabalhar com um *corpus* de tamanho grande, esse autor optou pelo ponto de corte de 20 ocorrências a cada milhão de palavras e que incidissem em pelo menos 10% dos textos do acervo, frequência essa que não seria adequada para o montante de dados atual.

com colocações estendidas, que caracterizam e auxiliam na construção de significado na prosa acadêmica e não apenas com idiosincrasias do estilo de cada autor. Além disso, como também seguimos o critério, segundo o qual pacotes de três palavras se abrigam em pacotes de quatro palavras gráficas (BIBER; BARIBIERE, 2007; BIBER et al., 1999; CORTES, 2004), mantivemos a busca por pacotes dessa última extensão em detrimento do tamanho do *corpus*. Isso posto, um primeiro item se destaca no que tange às diferenças nos *corpora* contrastados, considerando que trabalhamos com dois acervos de dimensões muito próximas: as publicações internacionais tendem a se apoiar menos em pacotes lexicais do que as produções das universidades brasileiras, o que poderia confirmar a tradição de que usuários menos experientes buscam mais suporte em expressões pré-fabricadas (DE COCK, 2000; HYLAND, 2008c), porque o texto que produzem tende a ser mais formulaico. Note-se que o acervo brasileiro não apenas apresenta um número maior de variedades de ocorrências (23 x 2)³⁸, mas também revela uma frequência de uso mais elevada em 13 das 23 ocorrências mais frequentes

Em relação às ocorrências do acervo internacional, a análise estrutural demonstra categorias não descritas dentre as mais usuais para artigos científicos:

Quadro 1: Categorização Estrutural do Acervo Internacional/ Table 3. *Structural Categorization of the International Corpus*

➤ Categorização	➤ Ocorrências
Coordinating conjunction + General determiner + Past Participle	✓ <i>But little is known</i>
General Determiner + Past Participle + Preposition	✓ <i>Little is known about</i>

Fonte: elaborada pelas autoras.

Embora Biber et al. (1999:1020) descrevam colocações estendidas de caráter frasal, tais quais as relacionadas na Figura 3, como pouco frequentes nos textos acadêmicos, uma vez que as de caráter nominal costumam ser mais abundantes, Hyland (2008c, p. 52-53) observa que tais colocações contendo Voz Passiva, na verdade, representam uma forma importante de destacar relações lógicas nas pesquisas científicas. Uma vez, no entanto, que os pacotes lexicais se apresentam ao olho do analista como estruturas incompletas, para averiguar tais

³⁸Embora, se observadas as linhas de concordância do AnticConc na Figura 2, percebam-se 24 ocorrências, foi preciso limpar uma delas, pois o *corpus*, por questões de preservação de *layout* de apresentação dos textos, manteve os subtítulos dos passos dos *abstracts* originais. Neste exemplo: *Conclusions Of the results of the*.

aspectos precisamos nos valer do contexto de onde se originam. Abrindo manualmente as linhas de concordância que dão acesso aos textos, foi possível constatar que todas as ocorrências contendo Voz Passiva são empregadas no momento em que os autores descrevem por que os estudos foram realizados, ou seja, na introdução/propósito, quando explicitam a motivação e destacam a existência de uma lacuna que dá sentido à pesquisa realizada. Tal constatação parece corroborar o aspecto mencionado acerca da motivação para o uso da estrutura e reforça a questão da forte relação existente entre forma e função do padrão fraseológico textual.

Quadro 2: Categorização Funcional do Acervo Internacional/*Table 4: Functional Categorization of the International Corpus*

➤ Categorização	➤ Ocorrências
Orientação para a Pesquisa	
Quantificação	✓ But little is known ✓ Little is known about

Fonte: elaborado pelas autoras.

Avançando na busca de entendimento qualitativo para os pacotes lexicais mais frequentes, a caracterização da estrutura retórica parece reforçar a dimensão mencionada anteriormente acerca do foco no estudo científico em si, uma vez que, funcionalmente, as ocorrências se orientam para a própria pesquisa, para descrever a existência de um nicho que valida o estudo, característica descrita na literatura (SWALES, 1990; HYLAND, 2008c) como típica do artigo acadêmico, cujo foco é especialmente voltado para validar a construção de conhecimento.

Para as ocorrências do acervo brasileiro, a distribuição estrutural³⁹ respeitou a seguinte classificação:

Quadro 3: Categorização Estrutural do Acervo Brasileiro/*Table 5. Structural Categorization of the Brazilian Corpus*

³⁹A classificação se atém às ocorrências em língua inglesa e às ocorrências lexicalmente aceitas no idioma. Não estão incluídos, portanto, Rio Grande do Sul, Grande do Sul Brazil e 'it was used a'.

➤ Categorização	➤ Ocorrências
<i>Noun phrase + of</i>	✓ <i>the aim of this</i> ✓ <i>the state of rio</i> ✓ <i>the results of the</i>
<i>Noun + of + noun phrase</i> (<i>Other noun phrases</i>)	✓ <i>state of rio grande</i> ✓ <i>aim of this study</i>
<i>Noun + be + infinitive</i>	✓ <i>study was to evaluate</i>
<i>Noun phrase + be</i>	✓ <i>the mean age was</i>
<i>Noun phrase + verb + prep</i>	✓ <i>this study aimed to</i> ✓ <i>this study was to</i>
<i>Of + noun phrase</i> ⁴⁰	✓ <i>of Rio Grande do</i>
<i>Of + noun phrase + be</i>	✓ <i>of this study was</i>
<i>Verb phrase</i>	✓ <i>to evaluate the correlation</i>
<i>Verb phrase+ prep</i>	✓ <i>evaluate the correlation between</i>
<i>Adjectival phrase + be + complement</i>	✓ <i>mean age was xb</i>
<i>Be + infinitive form + article</i>	✓ <i>was to evaluatet he</i>
<i>Passive + verb + article</i>	✓ <i>was evaluated usingt he</i>
<i>Prepositional phrase + of</i>	✓ <i>in the field of</i>
<i>Other prep phrases</i>	✓ <i>in relation to the</i> ✓ <i>as well as the</i> ✓ <i>in children and adolescents</i>

Fonte: elaborado pelas autoras.

Várias categorias estruturais precisaram ser nomeadas para dar conta do *corpus* brasileiro, dentre elas ‘*noun + be + infinitive*’, ‘*noun phrase + be*’, ‘*noun phrase + verb + prep*’, ‘*of + noun phrase*’, ‘*of + noun phrase + be*’, ‘*verb phrase*’, ‘*verb phrase+ prep*’, ‘*adjectival phrase + be + complement*’, ‘*be + infinitive form + article*’ and ‘*Passive + verb + article*’, porque não se encontram dentre as mais frequentes descritas por Biber et al. (1999).

Constata-se uma distribuição equilibrada entre colocações estendidas de caráter nominal e frasal, característica que novamente vai de encontro àquilo que a literatura sobre o tema tem descrito. Por outro lado, constata-se uma frequência maior do padrão ‘*noun phrase*’

⁴⁰Optamos pela classificação *of + noun phrase* por reconhecimento do composto Rio Grande do Sul.

em associação a outras estruturas (*Noun phrase + of, Noun + of + noun phrase, Noun phrase + be, Noun phrase + verb + prep, Of + noun phrase, Of + noun phrase + be*). Verifica-se também a metade da frequência de uso de Voz de Passiva do *corpus* internacional. Logicamente, estamos tratando com uma coleção muito pequena de dados, mas poderia haver aqui, se confirmado na pesquisa maior, um indício de que muitos dos blocos de língua adotados pelos pesquisadores incluídos em nosso *corpus* são menos adotados por estudiosos publicados internacionalmente. Além disso, cotejando as ocorrências a um estudo de Hyland (2008c) que levantou 50 pacotes lexicais mais frequentes em 120 artigos acadêmicos, em uma coleção de 730 mil palavras, apenas duas das colocações estendidas de nosso acervo nacional encontram-se entre as listadas por esse pesquisador: ‘*the results of the*’ e ‘*as well as the*’. Por certo, deve-se ter cautela com generalizações, especialmente porque o estudo mencionado não incluiu apenas colocações estendidas do campo da saúde. No entanto, nossos dados apresentam apenas uma ocorrência em comum (*as well as the*) também com um estudo de Byrd e Coxhead (2010) com 875 mil palavras em um *corpus* acadêmico de ciências biológicas, que apontou 35 pacotes lexicais dentre os mais frequentes. Tendo em mente, portanto, que o uso eficiente de um padrão de língua implica “sensibilidade quanto às preferências de usuários especializados para certas sequências de palavras em relação a outras”⁴¹ (HYLAND, op. cit.), podemos estar diante de um aspecto significativo do acervo.

Quadro 4: Categorização Funcional do Acervo Brasileiro/ Table 6. *Functional Categorization of the Brazilian Corpus*

➤ Categorização	➤ Ocorrências
Orientação para a Pesquisa	
Local	<ul style="list-style-type: none"> ✓ the state of rio ✓ state of Rio Grande ✓ of Rio Grande do
Procedimento	<ul style="list-style-type: none"> ✓ evaluate the correlation between ✓ to evaluate the correlation ✓ was evaluated using the
Descrição	<ul style="list-style-type: none"> ✓ the mean age was

⁴¹“[...] a sensitivity to the preferences of expert users for certain sequences of words over others.” (HYLAND, 2008c, p. 44).

	✓ mean age was xb
Tópico	✓ in the field of ✓ in children and adolescents
Orientação para o Texto	
Marcas de Transição	✓ as well as
Marcas de Resultado	✓ the resultof the
Marcas de Estrutura	✓ the aim of this ✓ aim of this study ✓ this study aimed to ✓ this study was to ✓ study was to evaluate ✓ of this study was ✓ was to evaluate the
Marcas de Limites	✓ in relation to the

Fonte: elaborado pelas autoras.

Conforme relaciona a Figura 6, duas dentre as três categorias de distribuição funcional estão presentes no acervo: orientação para a pesquisa e para o texto. Poderíamos relacionar a frequência de ocorrências de orientação para a pesquisa com um forte senso de importância de disseminação dos estudos por parte dos acadêmicos brasileiros. De acordo com a literatura, no campo as ciências experimentais, para pesquisas empíricas, muitos desses pacotes lexicais contribuem para a descrição dos contextos (*the state of Rio, state of Rio Grande, of Rio Grande do*) e tópicos de pesquisa (*in the Field of, in children and adolescents*) e para tanto se valem frequentemente da estrutura *noun phrases+ of* (HYLAND, 2008a). A orientação para o texto, por sua vez, aponta um engajamento com a conexão de ideias, com o direcionamento dos leitores para o texto e com a especificação de limitações. Tais pacotes tendem a incluir preposições, como é o caso aqui⁴².

Por outro lado, a ausência da terceira categoria descrita, qual seja de orientação para o participante⁴³, que se materializa através de expressões características de posicionamento e

⁴²Dado, no entanto, a presença da palavra ‘*evaluate*’ no campo retórico procedimento e no campo marcas de estrutura, cumpre explicitar que, pela abertura manual das linhas de concordância foi possível desambiguar as ocorrências e constatar que as do campo pesquisa dizem respeito a um procedimento experimental, enquanto ‘*this study was to evaluate*’ e ‘*was to evaluate the*’ se referem ao objetivo do estudo.

⁴³No entanto, é possível, conforme descrito por Byrd e Coxhead (2010), por exemplo, que a alguns pacotes lexicais seja atribuída mais de uma função retórica por vezes. De todo modo, um cotejamento criterioso das

engajamento do leitor, podem evidenciar um apagamento da voz do autor no texto. Dentre possíveis expressões de orientação para o leitor, encontram-se os ‘*hedges*’, formas de modular o discurso para expressar uma asserção, que são tão caras ao texto acadêmico contemporâneo, tais como ‘*it seems*’, ‘*it might be suggested that*’, ‘*it is useful that further studies*’. Alguns estudos destacam que aprendizes com menos proficiência ou maturidade em termos de redação acadêmica tendem a se expressar de modo mais categórico, e sua escrita apresenta menos marcas de modalização (KENNEDY e THORP, 2007; CHEN e BAER, 2010). Cabe ressaltar que, conforme verificado, o acervo internacional também não inclui essa categoria, o que creditamos à extensão, uma vez que, por meio dessas expressões, os autores experientes no domínio da prosa acadêmica transmitem aos leitores seu posicionamento epistêmico.

7 Perspectivas para a pesquisa em maior escala

A oportunidade de tecer este primeiro estudo empírico e elicitar alguns caminhos de entendimento do padrão lexical dos textos apontou perspectivas significativas para a pesquisa maior. Inicialmente, as questões de uso precisam ser entendidas em relação a um contexto específico de aplicabilidade, ou seja, deve-se ter cautela com as generalizações. Trabalhamos com um montante bastante pequeno de dados e percebemos a relevância de ampliar a base para o próximo cotejamento, a fim de estender os argumentos – ou descobrir novos desdobramentos, já que é o *corpus* que nos vai dar as respostas - para o padrão fraseológico do contexto eleito. Outrossim, mesmo com um acervo de dimensões tão pequenas, foi possível perceber que autores diferentes costumam adotar modos de dizer semelhantes, pois existe um forte componente formulaico nos textos acadêmicos, que o fator de frequência determina. Além disso, a forma como as análises estruturais e funcionais se completam parece indicar não apenas o *status* estritamente linguístico dos usuários, mas também um componente cultural ou ideológico de parte dos autores adotarem um determinado padrão lexical. Essas asserções destacam o significado de olharmos com atenção para os pacotes lexicais a serem organizados a partir do acervo completo.

Por fim, a pesquisa em maior escala certamente tecerá asserções mais ricas quanto a padrões de variabilidade lexical do *corpus*, porque explorará também os diferentes gêneros do acervo e fornecerá indícios mais específicos para a organização do AVA que a sistematizará. Para desenvolver tal percurso, reiteramos nossas escolhas metodológicas.

ocorrências listadas à categorização de Hyland (2008a,b,c) indicou que, de fato, não estamos diante de expressões atreladas à categoria de orientação para o participante.

Adotamos a LC, em função da crença na importância de trabalhar com coleções de dados amplos, a fim de perceber os padrões de previsibilidade da língua. Do mesmo modo, defendemos os pressupostos da LLE, em função de que “a especificidade das linguagens especializadas se expressa principalmente pela frequência de uso de determinados recursos linguísticos estatisticamente comprováveis” (FINATTO e KILLIAN, 2015, p. 38), pontos que a aproximam da LC. Alinhamo-nos também ao campo de EAP, porque esse segmento trata do ensino da língua adicional em consonância com as preocupações de desenvolver a língua em especialidade. Outrossim, voltamo-nos ainda para o suporte do campo de PLN para construir nosso AVA, em função da possibilidade de interface com os esforços no campo do ensino. Não obstante, entendemos que caminhamos aqui em direção a desenvolver um trabalho de ensino da língua adicional com real significado para os aprendizes, uma vez que pautado por padrões e uso reais a partir da área das Ciências da Saúde.

Referências

ABNT. Associação Brasileira de Normas Técnicas, 2013. Disponível em: <<http://www.abnt.org.br/>>. Acesso em: 01 jun. de 2013.

ALUÍSIO, Sandra Maria et al. How to Learn the Many Unwritten ‘Rules of the Game’ of the Academic Discourse: A Hybrid Approach Based on Critiques and Cases. In: **Proceedings of the IEEE International Conference on Advanced Learning Technologies**. Madison, Wisconsin. Los Alamitos, CA: IEEE Computer Society, 1, 2001. p. 257-260.

ALUÍSIO, Sandra Maria et al. Evaluating Scientific Abstracts with a Genre-specific Rubric. In **The 12th International Conference on Artificial Intelligence in Education - AIED**, 18-22 July, Amsterdam. 2005. p. 738-740.

ANTHONY, Lawrence. AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University, 2011. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/>>. Acesso em: 25 jan. 2015.

BAKHTIN, Mikhail. (Em Russo, 1979). **Estética da criação verbal**. 2. ed. São Paulo: Martins Fontes, 1997.

BELMONTI, Vanessa; GROSSI, Márcia Gorett Ribeiro. **Ambientes virtuais de aprendizagem: um panorama da produção nacional**. Disponível em: <www.abed.org.br/congresso2010>, Belo Horizonte, 2010. Acesso em: 01 jan. de 2014.

BERBER SARDINHA, Antônio Paulo. Linguística de corpus: histórico e problemática. **D.E.L.T.A.**, v. 16, n.2, p. 323-367, 2000.

BERBER SARDINHA, Antônio Paulo. Análise de gênero e linguística de corpus: identificação das unidades internas do gênero por meio da padronização lexical. **DIRECT Papers**, São Paulo/Liverpool, n. 51, 2003.

BIBER, Douglas. **Variations across speech and writing**. Cambridge: Cambridge University Press, 1988.

BIBER, Douglas. Methodological issues regarding corpus-based analyses of linguistic variation. **Literary and linguistic computing**, v. 5, p. 257-269, 1990.

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

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

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

BIBER, Douglas et al. **Grammar of spoken and written English**. London: Longman, 1999.

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

BIBER, Douglas; BARBIERE, Federica. Lexical bundles in university spoken and written registers. **English for specific purposes**, v. 26, p. 263-286, 2007.

BYRD, Pat. COXHEAD, Averil. *On the other hand: lexical bundles in academic writing and in the teaching of EAP*. **University of Sydney Papers in TESOL**, v.5, p. 31-64, 2010.

CABRÉ, Maria Teresa. El discurso especializado o la variación funcional determinada per la temática: nuevas perspectivas. In: CABRÉ, Maria Teresa. (Org.) **La terminología: representación y comunicación; elementos para una teoría de base comunicativa y otros artículos**. Barcelona: IULA, 1999, p. 151-173.

CABRÉ, Maria Teresa. Textos especializados y unidades de conocimiento: metodología y tipologización. In: GARCIA PALACIOS, Joaquin.; FUENTES, Maria Teresa (Eds.). **Texto, terminología y traducción**. Salamanca: Almar, 2002. p. 15-36.

CANALE, Merrill; SWAIN, Michael. Theoretical bases of communicative approaches to second language teaching and testing. **Applied Linguistics**, v. 1, p. 1-47, 1980.

CANDLIN, Christopher; HYLAND, Ken. Introduction: Integrating approaches to the study of writing. In: CANDLIN, Christopher; HYLAND, Ken. (Eds.) **Writing: texts, processes and practices**. New York: Longman, 2009. p. 1-18.

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

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

CIAPUSCIO, Guiomar; KUGUEL, Inés. Hacia una tipología del discurso especializado: aspectos teóricos y aplicados. In: GARCÍA PALACIOS, Joaquín; FUENTES, María Teresa. **Entre la terminología, el texto y la traducción**. Salamanca: Almar, p. 37-73, 2002.

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

CONRAD, Susan; Biber, Douglas. The frequency and use of lexical bundles in conversation and academic prose. In: TEUBERT, Wolfgang; MAHLBERG, Michaela (Eds.). **The corpus approach to lexicography**. Thematischer Teil von Lexicographica. Internationales Jahrbuch für Lexikographie, 2004. p. 56–71.

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

DAYRELL, Carmen. Lexical bundles in English abstracts: A corpus-based study of published and non-native graduate writing. In: **Proceedings of the Corpus Linguistics Conference, 2009**. Liverpool, 2009. Disponível em: <<http://www.liv.ac.uk/english/CL2009/index.htm>> Acesso em: 01 jun. 2015.

DAYRELL, Carmen. *Corpora* no Ensino do Inglês Acadêmico: Padrões Léxico-Gramaticais em *abstracts* de pós-graduandos brasileiros. In: VIANA, Vander; TAGNIN, Stella. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010. p. 137-171.

DAYRELL, Carmen; ALUÍSIO, Sandra. Using a comparable corpus to investigate lexical patterning in English abstracts written by non-native speakers. **LREC Workshop on Comparable Corpora, 2008**. p. 61-66, 2008. Disponível em: URL: http://repository.dlsi.ua.es/251/1/workshops/w12_P .Acesso em: 01 jan.2014.

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

DAYRELL, Carmen; CANDIDO JR., Arnaldo ; LIMA, Gabriel; MACHADO JR., Danilo ; COPESTAKE, Ann ; FELTRIM, Valéria Delisandra; TAGNIN, Stella Ester; ALUÍSIO, S. Rhetorical Move Detection in English Abstracts: Multi-label Sentence Classifiers and their Annotated Corpora, 2012. In: LREC 2012, Istambul. **Proceedings of the 8th International Conference on Language Resources and Evaluation**. 2012.

FELTRIM, Valéria Delisandra; NUNES, Maria das Graças Volpe; ALUÍSIO, Sandra Maria. Um corpus de textos científicos para apoiar a construção de um sistema de auxílio à escrita técnica. 2002. In: **12o Intercâmbio de Pesquisas em Linguística Aplicada**, 2002, São Paulo. Caderno de Resumos.

FELTRIM, Valéria Delisandra; ANTIQUEIRA, L.; NUNES, Maria das Graças Volpe. A construção de uma ferramenta de auxílio à escrita de resumos acadêmicos em português. 2003. In: **IV Encontro Nacional de Inteligência Artificial (ENIA)**. Campinas. Anais do IV ENIA.

FENG, Haiying. Perspectives: Working Papers in English and Communication. **Spring**, v. 17, n. 1, p. 16-51, 2006.

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

FINATTO, Maria José Bocorny. Dos dicionários aos modos de perceber a língua: contribuições de um trabalho pioneiro. In: TEIXEIRA, Marlene; FLORES, Valdir Nascimento (Orgs.). **O sentido na linguagem: uma homenagem à professora Leci Borges Barbisan**. 1. ed. Porto Alegre: EDIPUCRS, 2012. p. 17-32.

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

FINATTO, Maria José Bocorny; KILLIAN, Cristiane. Texto-Comentário 7. In: FINATTO, Maria José Bocorny; ZILIO, Leonardo. **Textos e termos por Lothar Hoffmann**. Porto Alegre: Palotti, 2015, p.123-26. Disponível em: <<http://www.ufrgs.br/ppglettras/>>. Acesso em: 01 jun. 2015.

FREIRE, Paulo. **A importância do ato de ler: em três artigos que se completam**. São Paulo: Autores Associados, 1989.

GATEHOUSE, Kristen. Key Issues in English for Specific Purposes (ESP) Curriculum Development. **The internet TESL journal**, v. VII, n. 10, 2001. Disponível em: <http://iteslj.org/> Acesso em maio de 2014.

GENOVES JR., Luiz; FELTRIM, Valéria Delisandra; DAYRELL, Carmen ; ALUÍSIO, Sandra. Automatically detecting schematic structure components of English abstracts. In: **Proceedings of the RANLP'2007, workshop on natural language processing for educational resources**. Borovets, Bulgaria, 2007. p. 23-29.

GUIMARÃES, Elisa. **A articulação do texto**. 10. ed. São Paulo: Ática, 2006.

HALLIDAY, Michael Alexander Kirkwood. **Explorations in the functions of language**. London: Edward Arnold, 1973.

HALLIDAY, Michael Alexander Kirkwood. *Corpus* studies and probabilistic grammar. In: AIJMER, Karin; ALTENBERG, Bengt (Org.). **English corpus linguistics: Studies in honour of Jan Svartvik**. London: Longman, 1991. p. 30-43.

HALLIDAY, Michael Alexander Kirkwood. Language as system and language as instance: The *corpus* as a theoretical construct. In: SVARTVIK, J. (Org.). *Directions in Corpus Linguistics*. **Proceedings of nobel symposium 82**, Stockholm, 4-8. Berlin, New York: De Gruyter, 1991. p. 4-8.

HALLIDAY, Michael Alexander Kirkwood; MCINTOSH, Angus; STREVEN, Peter. **The linguistic sciences and language**. Teaching. Longmans' Linguistic Library. London: Longmans, 1964.

HELLERMAN, John. **Social actions for classroom language learning**. Clevedon, U.K.: Multilingual Matters, 2008.

HOFFMANN, Lothar. **Kommunikationsmittel Fachsprache: eine Einführung**. 2. Aufl. Tübingen: Narr, 1985.

HOFFMANN, Lothar. O papel das linguagens especializadas desde meados do século XX. Tradução de ZILIO, Leonardo; PICKBRENNER, Minka. In: Maria José Bocorny; ZILIO, Leonardo. **Textos e termos por Lothar Hoffmann**. Porto Alegre: Palotti, 2015, p.21-34. Disponível em: <http://www.ufrgs.br/ppglettras/> Acesso em: jun. de 2015.

HUNSTON, Susan. **Corpora in applied linguistics**. London: Cambridge University Press, 2002.

HYLAND, Ken. "As can be seen: lexical bundles and disciplinary variation". **English for Specific Purposes**, v. 27, p. 4–21, 2008a.

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

HYLAND, Ken. "Academic clusters: text patterning in published and postgraduate writing". **International Journal of Applied Linguistics**, v. 18, n.1, p. 41–61, 2008c.

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

KENNEDY, Christopher; THORP, Dilys. A corpus investigation of linguistic responses to an IELTS Academic Writing task. In: TAYLOR, Lynda ; FALVEY Peter (Eds.). **IELTS collected paper: research in speaking and writing assessment**. Cambridge: Cambridge University Press, p. 316-78, 2007.

KOCH, Ingedore Villaça. O texto: construção de sentidos. **Organon**, v. 9, n. 23, p. 21-27, 1995.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à terminologia – teoria e prática**. São Paulo: Contexto, 2004.

LIMA, Vera Lúcia; STRUBE, NUNES, Maria das Graças Volpe; VIEIRA, Renata. Desafios do processamento de línguas naturais. **SEMISH-Seminário Integrado de Software e Hardware**. v.34, n.1, 2007. **Anais do XXVII Congresso da SBC**, 2007. p. 2202-2216.

MANN, William; THOMPSON, Sandra. Rhetorical Structure Theory: Toward a functional theory of text organization. **Text**, v. 8, n.3, p. 243-281, 1988.

MCENERY, Tony; XIAO, R.; TONO, Y. **Corpus-based language studies**. USA/Canada: Routledge, 2006.

MCNAMARA, Tim. **Language testing**. Oxford/New York: Oxford University Press, 2000.

MOTTA-ROTH, Désirée.; HENGES, Graciela Rabuske. **Produção textual na universidade**. São Paulo: Parábola, 2013.

NUNES, Maria das Graças. **O processamento de línguas naturais: para quê e para quem?** Notas Didáticas do ICMC, N. 73, Instituto de Ciências Matemáticas e de Computação. São Carlos-SP, maio, 12p, 2008.

NUNES, Maria das Graças; ALUÍSIO, Sandra Maria; PARDO, Thiago. Um panorama do Núcleo Interinstitucional de Linguística Computacional às vésperas de sua maioridade. **LinguaMÁTICA**, v. 2, n. 2, p. 13-27, 2010.

PICKBRENNER, Minka; ZILIO, Leonardo. Texto-comentário 2. In: FINATTO, Maria José Bocorny; ZILIO, Leonardo. **Textos e termos por Lothar Hoffmann**. Porto Alegre: Palotti, 2015, p. 15-20. Disponível em: <<http://www.ufrgs.br/ppgletras/> Acesso em: jun. de 2015>.

POSSAMAI, Viviane. **Marcadores textuais do artigo científico em comparação português-inglês: um estudo sob a perspectiva da tradução**. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2004.

RAJAGOPALAN, Kanavillil. The Concept of “World English” and its implications for ELT. **ELT Journal**, v. 58, n. 2, p. 111-17, 2004. Disponível em: <<http://eltj.oxfordjournals.org/cgi/content/abstract/58/2/111>> Acesso em: maio de 2014.

ROJO, Roxane. **Multi-letramento na escola**. São Paulo: Parábola, 2012.

SINCLAIR, John. **Corpus, concordance, collocation**. Oxford: OUP, 1991.

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

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

SWALES, John Malcolm. **Genre and engagement**. La Revue Belge de la Philologie et L’Histoire. London: Routledge, 1993.

SWALES, John Malcolm. **Research genres: exploration and applications**. Cambridge: Cambridge University Press, 2004.

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

ZILIO, Leonardo. Terminologia textual e linguística de *corpus*: estudo em parceria. In: LOPES PERNA, Cristina Becker; DELGADO KOCH, Heloísa Orsi; FINATTO, Maria José Bocorny (Orgs.). **Linguagens especializadas em corpora**. Porto Alegre: EDIPUCRS, 2010, p.128-151.

WILLIAMS, Geoffrey. Challenging the native-speaker norm: A corpus-driven analysis of scientific usage. In: BARNBROOK, Geoff; DANIELSSON, Pernilla; MAHLBERG, Michaela (Eds.). **Meaning texts: the extraction of semantic information from monolingual and multilingual corpora**. London/New York: Continuum International Publishing Group, 2006. p. 115–27.