

UM OLHAR PARA A LINGUAGEM NA PRODUÇÃO DE ARTIGOS CIENTÍFICOS

Cintia Ferreira de Almeida¹

Jane Marian²

RESUMO

O objetivo desta pesquisa é apresentar uma metodologia baseada na linguística de corpus com propósito de auxiliar acadêmicos de diferentes áreas na escrita de artigos científicos. Para ilustrar os procedimentos foi compilado um corpus da área de Terminologia, esse é constituído de 143 artigos científicos de três revistas conceituadas da área de estudo. Para análise e extração dos dados foi utilizado o programa computacional *AntConc* e suas ferramentas *Keyword*, *Wordlist* e *Concordance*. Os resultados demonstraram que a partir da compilação e processamento dos dados para análises, o corpus pode apresentar resultados surpreendentes que dificilmente seriam observados sem o auxílio das ferramentas computacionais. Portanto, conclui-se que a Linguística de Corpus pode contribuir significativamente para o desenvolvimento das habilidades nas produções textuais.

Palavras-chave: Linguística de Corpus. Terminologia. Escrita Acadêmica. Estudo da Linguagem. Léxico Especializado.

¹ Aluna do 3º ano do curso de Letras Português – Inglês da FAE Centro Universitário. Bolsista do Programa de Apoio à Iniciação Científica (PAIC 2016/2017). *E-mail*: cintia_almeida24@hotmail.com

² Doutora em Estudos da Tradução pela UFSC. Professora da FAE Centro Universitário. *E-mail*: jane.marian@fae.edu

INTRODUÇÃO

Um dos maiores desafios do estudante ao ingressar na universidade, certamente, são os obstáculos linguísticos, ou seja, novos conhecimentos científicos e vocabulários específicos introduzidos pelo professor. Nessa fase inicial de imersão no curso e ambiente acadêmico muitos estudantes acabam desistindo, pois consideram impossível a missão de compreender o que o professor está falando.

Galvão (2004, p. 242) observa que após o vestibular o aluno é inserido em um contexto de bibliografias especializadas indicadas pelo curso. Segundo a autora, esse contato técnico-científico é marcante, pois até então o indivíduo estava habituado a fazer leituras e produções com termos e estruturas bem diferentes, utilizando vocabulários do cotidiano para expressar seu conhecimento geral. Essa mudança entre a produção de textos do cotidiano para os textos técnico-científicos geram insatisfações para ambos, professores e acadêmicos. Para Galvão (2004, p. 242) os docentes acreditam que a limitação nas produções textuais estão relacionadas à imaturidade ou a uma provável deficiência na formação escolar e, do outro lado, os estudantes consideram o texto acadêmico, muitas vezes, chatos e difíceis de compreender.

De fato, a problemática faz refletir sobre o processo de aquisição do conhecimento. Como adquirir novas informações sem compreender a terminologia empregada, uma vez que, o ser humano aprende a partir de conexões que faz entre o conhecimento já adquirido e o novo?

Neste sentido, a aprendizagem ocorre de forma organizada e sistematizada, considerando as novas expressões e termos específicos da área, pois são fundamentais para o desenvolvimento cognitivo do aprendiz no que se refere ao conteúdo técnico-científico que será adquirido ao longo dos anos.

Krieger (2006, p. 190) ressalta a importância que o domínio da terminologia representa para a produção da escrita. Para a autora, a linguagem especializada é necessária para que a produção textual seja confiável e, no ambiente acadêmico, é comum perceber dificuldades, por parte dos acadêmicos, em reconhecer a terminologia especializada. Isso sucede independente da área de estudo, sobretudo, nos primeiros anos das graduações.

Segundo Almeida (2002, p. 157) “para aprender, o aluno precisa entender, organizar, armazenar e evocar a informação” isso decorre porque a aprendizagem é um processo de construção do conhecimento e não acontece de forma aleatória, mas sistematizada. Por isso, o professor é o principal responsável no processo de amadurecimento cognitivo do acadêmico, ensinando não somente o conteúdo, mas também, estratégias de desenvolvimento do aprendizado.

Normalmente, a terminologia é internalizada pelo estudante de forma natural com o passar dos semestres e após muitas leituras de textos científicos. No entanto, esse processo é lento e diferente para cada estudante, pois depende de vários fatores que irão estimulá-los ou não durante o decorrer dos anos. Nesta perspectiva, questiona-se como o professor poderia acelerar esse processo de reconhecimento terminológico para que o aprendiz sofra menos ao longo do percurso.

A LC (Linguística de Corpus) é uma área que estuda a linguagem por meio de textos autênticos, ou seja, produzidos naturalmente em um contexto real da fala ou da escrita. Por meio de ferramentas eletrônicas é possível extrair palavras-chave da área de estudo e analisá-las em seu contexto de uso.

Esta abordagem vem sendo utilizada por diferentes estudiosos para o ensino de idiomas (DUTRA e SILERO, 2010; MARIAN e LIMA, 2016), para os estudos da tradução (SARDINHA, 2002; BAKER, 1995; COULTHARD, 2005; MARIAN, 2010; TEIXEIRA, 2008), estudos linguísticos em geral e elaboração de material didático (BIBER, 1998; SINCLAIR, 2004; TOGNINI-BONELLI, 200, SARDINHA, 2004), estudos terminológicos e extrações de terminologias para a elaboração de glossários e dicionários (MARIAN, 2016; MAIA, 2003; CARVALHO, 2007; FROMM, 2005; LEPRE, 2007; MATSIRA, 2008; PORTOLAN, 2011) e, recentemente, vem sendo utilizado para o ensino materno, principalmente, para áreas específicas de diferentes cursos com o objetivo de desenvolvimento nas habilidades de produção textual (GALVÃO, 2004; GAVIOLI, 2005).

O objetivo desta pesquisa é desenvolver uma metodologia de pesquisa baseada na LC com o uso do programa AntConc desenvolvido por Laurence Antony. *Este software está disponível gratuitamente no site³ do autor e serve para análises linguísticas e extração de termos específicos de uma determinada área. Como amostra serão utilizados artigos científicos da área da Terminologia, com o intuito de ilustrar ao acadêmico como ele pode compilar seu próprio banco de dados e a partir desse método desenvolver uma melhor escrita acadêmica, baseando-se no processo de análise lexical por meio de um corpus especializado.*

1 LINGUÍSTICA DE CORPUS

Para Sardinha (2004, p. 37), a LC faz a parte do estudo da linguística aplicada que se ocupa em fazer a investigações da linguagem. Essa ciência se expandiu com a ascensão das novas tecnologias, sobretudo, com o surgimento dos computadores na década de 1960, o que contribuiu muito para o desenvolvimento da pesquisa.

³ <http://www.laurenceanthony.net/software.html>

Nos anos de 1980, os microcomputadores e a elaboração de softwares específicos para a área da pesquisa de linguagem facilitaram para o avanço dos estudos, fazendo a LC ganhar mais destaque. No entanto, o estudo baseado em corpus não é uma prática recente, pois surgiu antes da Grécia Antiga com Alexandre, O Grande, com o corpus helenístico e durante a Idade Média os corpora eram baseados na bíblia.

No século XX, os linguistas Boas e Fries e o educador Thorndike ganharam destaque nas pesquisas baseadas em corpus, pois se empenharam no estudo e descrição da linguagem por meio de análises linguísticas (SARDINHA, 2004, p.3). Coletados manualmente, as pesquisas de corpora se tornaram por muitas vezes exaustivas. O processo de coleta e armazenamento ainda é utilizado nos dias atuais, porém, as ferramentas computacionais proporcionaram ao estudo velocidade e confiabilidade.

Distintamente da perspectiva da gramática gerativista de Chomsky (1975), que defendia a ideia que a escolha do léxico é determinada por fatores como a intuição do indivíduo e a introspecção, Sardinha (2004, p. 31), partindo da concepção da Linguística de Corpus, destaca que as variações das palavras não são feitas de maneira aleatória, mas existem possibilidades prováveis entre escolhas lexicais e padrões linguísticos que devem ser considerados. Um dos fatores que comprovam a relevância dessa afirmação são os centros de pesquisa criados com o intuito de estudar a linguagem de forma empírica.

Sardinha (2002, p.18) destaca a Grã-Bretanha como um dos maiores centros de pesquisa da linguagem em universidades como Birmingham, Brighton e Lancaster. Países como Noruega, Suécia e Dinamarca também contém locais dedicados especificamente para a pesquisa em linguística de corpus.

Por meio da combinação entre o estudo da Terminologia e da Linguística de Corpus, a presente pesquisa tem como propósito apresentar uma metodologia de pesquisa em que os termos relevantes na escrita acadêmica são extraídos e analisados. Quando o acadêmico descobre que precisa prestar atenção tanto no conteúdo da área de estudo quanto nos elementos linguísticos textuais, ele aprimora significativamente suas habilidades e produz melhores textos de valor científico. Para exemplificar o processo de coleta, organização e análise dos termos técnicos foram utilizados artigos da área de Terminologia que servem para ilustrar como funciona a pesquisa baseada na linguística de corpus.

1.1 A LINGUÍSTICA DE CORPUS E SUA CONTRIBUIÇÃO PARA O APRENDIZADO DA LINGUAGEM ACADÊMICA ESPECIALIZADA

A LC trata da coleta e da exploração de textos e desempenha um papel importante para a pesquisa de uma língua ou variedade linguística (SARDINHA, 2004, p. 3). A partir do estudo da LC a linguagem é explorada em um tempo muito mais rápido e os dados são mais confiáveis, pois conta-se com a colaboração de ferramentas computacionais. No presente artigo foi utilizado o *software* AntConc, sendo possível analisar diferentes funções que determinados termos de uma área podem desempenhar.

Os corpora especializados são compilados com o propósito de servirem como amostra da linguagem específica, por exemplo: medicina, negócios, filosofia, direito etc. Gavioli (2006, p. 9) afirma que os corpora de textos especializados são instrumentos úteis em fornecer informações isoladas de palavras lexicais, gramaticais ou elementos textuais para um trabalho com turmas de diferentes áreas do conhecimento, principalmente, por ser extremamente fácil de manusear e sem custos.

Segundo Gavioli (2006, p. 55) em 1962 Barber desenvolveu o estudo em *ESP* (*English for Specific Purpose*). Barber escreveu um artigo baseado em corpora com 23 mil palavras, formado por pesquisas de três artigos científicos de áreas diferentes (engenharia, química e astronomia) no qual o autor calculava manualmente os dados obtidos. Em sua pesquisa, o linguista analisou o tamanho das frases, o número de orações por frases, os modais mais frequentes e aspectos relacionados aos tempos verbais utilizados nos textos. Nesse artigo, o autor utilizou os dados para dar suporte a sua hipótese, segundo Barber o gerúndio não precisaria ser ensinado nos cursos de produção científica, pois a frequência dessa forma verbal era extremamente baixa. Atualmente, os programas de computadores desenvolvidos para análises linguísticas permitem gerar automaticamente essas operações com um número muito maior de dados e de forma mais confiável.

Hunston (2002, p. 110) menciona que os alunos precisam ser orientados para a escrita acadêmica e devem ter consciência sobre aspectos linguísticos que caracterizam esse gênero textual. O autor observa que o trabalho com corpus deve complementar o estudo de gêneros textuais. Existem alguns estudos que analisam as fraseologias empregadas, outros observam como os escritores de diferentes disciplinas criam relacionamentos com os seus leitores, entre outras investigações realizadas.

Gledhill (1995, 2000 apud Gavioli, 2005, p. 57) utilizou um corpus de artigos científicos sobre câncer para observar os movimentos retóricos (introdução, método, resultados, discussão e conclusão) e utilizou a ferramenta computacional de análises linguísticas *WordSmith Tools* para buscar palavras-chave. Neste trabalho, o autor verificou que algumas palavras são significantes em determinadas seções do artigo científico. Dessa forma, afirmou que alguns padrões lexicais são frequentes para diferentes seções e essas informações são importantes no ensino da escrita.

Weber (2001, p. 14) apresentou um experimento em que os alunos extraíram elementos frequentes de textos jurídicos e usaram o concordanciador para encontrar os padrões lexicais mais frequentes. Assim, pode-se dizer que o estudo da LC abriu caminhos para novas propostas de estudo do léxico e forneceu suporte para a descrição da linguagem. Segundo Gavioli (2002, p. 58) essa abordagem auxilia os professores de ensino da escrita especializada, pois fornece rapidamente uma ideia de quais itens lexicais são mais recorrentes em textos de determinadas áreas de estudo (medicina, direito, administração, etc.), além de fornecer aos acadêmicos alguns *frame words* de palavras que causam dificuldades na compreensão e produção de textos.

Gavioli (2006, p. 90) aponta que uma dificuldade tradicional encontrada por professores de idiomas é lidar com o vocabulário técnico e o conceito deles. Nesse sentido, a autora apresenta uma problemática entre escolhas estilísticas e retóricas versus o significado dos termos e conceitos. De um lado, os elementos retóricos e estilísticos que deveriam ser ensinados pelo professor de línguas ou linguistas e, do outro, o significado dos termos que deveriam ser apresentados por professores especialistas da área. Apesar de haver uma tentativa dos dois lados interagirem, segundo a autora, isso não ocorre. De acordo com Gavioli (2006, p. 57), com o estudo baseado na LC, o linguista pode explorar a terminologia especializada e construir uma ponte entre essas duas vertentes.

Para Olson (2008, p. 1), a escrita científica exige mais do que apenas saber utilizar uma estrutura adequada. O domínio de termos específicos da área também faz parte do conjunto de fatores que são relevantes para o acadêmico produzir uma escrita com um valor científico considerável.

1.2 ANÁLISES LINGUÍSTICAS DE EXTRAÇÃO TERMINOLÓGICA E A FERRAMENTA COMPUTACIONAL ANTCONC

O objeto desta pesquisa é a Terminologia, ou seja, o estudo e análise de termos técnicos dentro de um contexto (ALBERTS-FRANCO, 2015, p. 183). A contribuição da LC proporciona ao presente estudo uma melhor análise mediante a utilização de ferramentas computacionais, as quais auxiliam na pesquisa sobre a frequência que os termos são utilizados comparando diferentes corpora da área.

Por meio da pesquisa em diferentes artigos que exploraram diferentes terminologias, pode-se perceber que os programas mais utilizados quando se trata de LC são: *WordSmith Tools* e *AntConc*. Há inúmeros trabalhos desenvolvidos na área de estudo dos termos especializados que utilizam dessas mesmas ferramentas. Para a elaboração desta pesquisa, foi utilizado o software *AntConc* (3.4.4). A ferramenta computacional foi escolhida a partir de alguns critérios que incluem: a fácil acessibilidade do *software*, por se tratar de uma ferramenta gratuita; e a facilidade para estudantes de qualquer área compreender e saber utilizá-la para organizar o corpus do seu campo de estudo.

AntConc é uma ferramenta computacional desenvolvida por Laurence Anthony, PhD em Linguística Aplicada pela Universidade de Birmingham, que serve para auxiliar o estudo da linguística de corpus. As funções que compõem esse software são: *Concordance*; *Concordance Plot*; *File View*; *Clusters*; *Collocates*; *Word List* e *Keyword List*.

O *Concordance* apresenta o termo pesquisado e as diferentes linhas de concordância presentes nos corpora. Dessa forma, essa ferramenta pode ser utilizada para contribuir na produção da escrita, pois assim é possível identificar de que maneira os termos são utilizados em situações concretas.

O *Concordance Plot* mostra em forma de código de barras como o termo pesquisado aparece nos arquivos selecionados, fazendo com que seja possível identificar em qual parte do texto ele aparece. Logo, a pesquisa dos termos é contextualizada na situação real de uso, o que auxilia o pesquisador a compreender como utilizar. Ao clicar em uma das barras é possível abrir a função *File View*.

O *File View* identifica os diferentes contextos que um termo aparece em um mesmo corpus de forma individual. Essa função torna mais clara a observação de diferentes papéis que um mesmo termo pode desempenhar dentro de um mesmo campo de estudo.

A ferramenta *Clusters* exhibe as combinações frequentes de uma palavra pesquisada com dois ou mais termos, e permite ao pesquisador perceber como diferentes derivações de uma mesma palavra são utilizadas dentro de um mesmo texto.

Word List apresenta uma lista com todas as palavras dos corpora por ordem de frequência. Com essa função é possível pesquisar na lista se determinada palavra está presente nela e qual é a frequência que ela aparece.

Keyword List mostra uma lista de palavras-chave a partir da comparação entre um corpus maior (corpus de referência) e o corpus do campo específico (corpus de estudo). Tornando possível identificar quais são os termos mais utilizadas na área estudada. Essa ferramenta proporciona ao pesquisador uma clara identificação dos termos mais recorrentes que dão propriedade na área dos corpora analisados. Para a presente pesquisa foram utilizadas essas ferramentas descritas acima.

1.3 A TERMINOLOGIA E O ESTUDO DA LINGUAGEM ESPECIALIZADA PARA A ESCRITA CIENTÍFICA

Segundo Pavel e Nolet (2002, p.17) “a terminologia faz parte da Linguística Aplicada, que compreende o trabalho em lexicografia especializada, tradução, redação e ensino de línguas”. Esses estudos são divididos pelo objetivo da pesquisa que pode ser a análise do léxico comum ou do léxico especializado. Na investigação do léxico comum está inserida a lexicografia que tem por objetivo organizar dicionários de uma língua partindo do léxico total da mesma. Por outro lado, a análise do léxico especializado,

com a Terminologia e Terminografia, se ocupa em estudar o corpus de uma área e formular um glossário de termos especializados (FILHO, 2013, p. 6).

Com a finalidade de organizar linguagens especializadas, a Terminologia surge em 1931 com o engenheiro austríaco Eugen Wüster, que propôs por meio da obra “Introdução Geral da Terminologia” (TGT) fazer da terminologia uma disciplina e pretendia retirar possíveis ambiguidades nas linguagens (CABRÉ, 2003, p.166). No início dos anos 90 esse estudo foi criticado por Cabré (1999, p. 3) e pesquisadores da Linguística Aplicada com a publicação da Teoria Comunicativa da Terminologia (TCT).

A TCT defende a ideia de que uma unidade lexical pode desempenhar papel de termo dependendo do contexto, partindo desse princípio um termo pode assumir diferentes significados em função da situação. Além disso, esse estudo analisa os aspectos terminológicos da linguagem em seus aspectos social, cognitivo e linguístico. Social porque tem por objetivo estudar termos levando em conta a adequação do trabalho à área que está sendo estudada. A perspectiva cognitiva leva em conta o domínio da linguagem especializada, tendo em vista que é o conhecimento da área-objeto que permite a identificação de sua terminologia. Por último a perspectiva linguística que diz respeito à elaboração de um modelo que leve em consideração a competência e o desempenho do produtor do texto.

Outra ideia defendida na TCT é a de que as variações ocorrem de maneira natural, ou seja, a variação não deve ser restringida, mas levada em conta e descrita na terminologia. Assim, a organização da terminologia auxilia a escrita especializada de diferentes estudos, pois por meio do conhecimento do léxico a escrita científica tem maior prestígio.

Profissionais de diferentes áreas precisam dominar conceitos e termos de suas áreas, para que assim possam atuar de forma eficaz. Tendo isso em vista, a presente pesquisa tem como objetivo auxiliar tanto professores quanto acadêmicos que estão ingressando em diferentes cursos por meio de uma metodologia de coleta análises de dados linguísticos que permitam perceber e compreender a escrita científica de forma mais rápida e eficaz.

2 METODOLOGIA DE PESQUISA

A metodologia empregada visa analisar e extrair dados empíricos de textos relacionados às situações reais de uso. A análise dos dados é realizada de forma qualitativa e quantitativa, pois além da coleta de termos específicos da área de estudo, também serão analisados diferentes contextos em que os mesmos aparecem.

Esse método visa a observação de informações autênticas da produção escrita e, conseqüentemente, extração e compreensão linguística de termos da área de estudo. O objetivo desse procedimento é desenvolver e melhorar, no que se refere a questões linguísticas, a produção da escrita acadêmica especializada.

O procedimento utilizado para chegar aos resultados desta pesquisa consistem em três passos (MARIAN, 2014, p. 201), são eles: (i) projeto do corpus; (ii) construção do corpus; e (iii) processamento do corpus.

2.1 PROJETO DO CORPUS

O Projeto do corpus consiste na seleção dos textos, ou seja, o desenho ou descrição dos corpora a serem analisados na pesquisa. É importante analisar qual é a origem dos corpora, gênero textual, área de estudo, representatividade, modo (escrito ou falado), idioma e temporariedade.

2.1.1 Descrição do Corpus de Estudo

O corpus de estudo desta pesquisa é formado a partir de artigos sobre estudos da terminologia, compilados de revistas científicas especializadas nessa área, são elas: *Gtlex*, *Tradterm* e *Riterm*. Ao todo são 143 artigos sobre terminologia escritos em língua portuguesa, compreendidos no período de 2010 a 2016. O grau de representatividade que esse corpus tem é considerado de pequeno porte e, mesmo assim, eficaz para chegar ao objetivo desta pesquisa.

2.1.2 Descrição do Corpus de Referência

O corpus de referência utilizado foi o NILC (Núcleo Institucional de Linguística Computacional), disponível no site⁴ para download. Composto por textos escritos, contendo diferentes gêneros textuais de linguagem geral, apresenta 40 milhões de palavras em língua portuguesa. Segundo classificação de Sardinha (2004, p. 26) é um corpus de tamanho grande, portanto, válido para ser comparado ao corpus de estudo.

2.2 CONSTRUÇÃO DO CORPUS

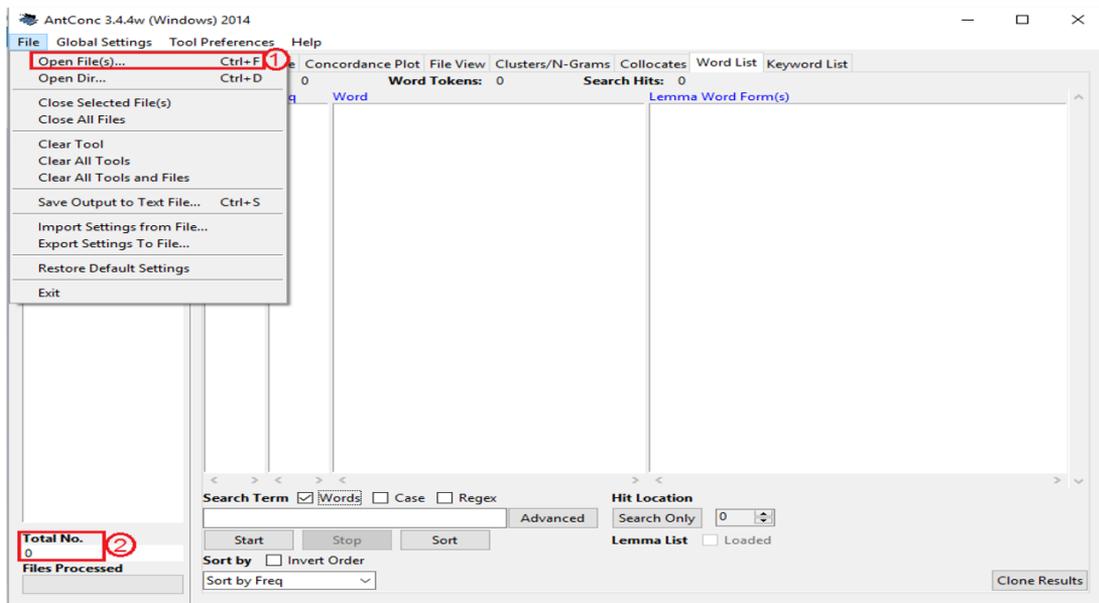
O processo de construção do corpus equivale a coleta de dados, formatação e “limpeza” dos textos. A coleta dos corpora de estudo foram realizadas por meio de downloads de artigos presentes em revistas científicas eletrônicas. Depois da coleta os textos foram convertidos do formato pdf para txt, pois é o único aceito na leitura do software *AntConc*. Após a conversão foi realizada a limpeza dos textos, ou seja, foram removidos todos os símbolos, espaços, marcas e números que não são relevantes para a pesquisa.

⁴ Disponível em: <<http://www.nilc.icmc.usp.br/nilc/index.php/tools-and-resources>> acesso em: 20 de jun. 2017

2.3 PROCESSAMENTO DO CORPUS

O processamento dos corpora foi gerado por meio da ferramenta computacional *AntConc*, e nela utilizadas as funções *Concordance*, *Word List* e *Keyword List*. A primeira ferramenta utilizada foi o *Word List*, que organiza uma lista de palavras do corpus por ordem de frequência. Para isso, realizou-se a inserção de arquivos no programa, seguindo os passos abaixo:

Figura 1: AntConc 3.4.4 (inserindo corpus de estudo)



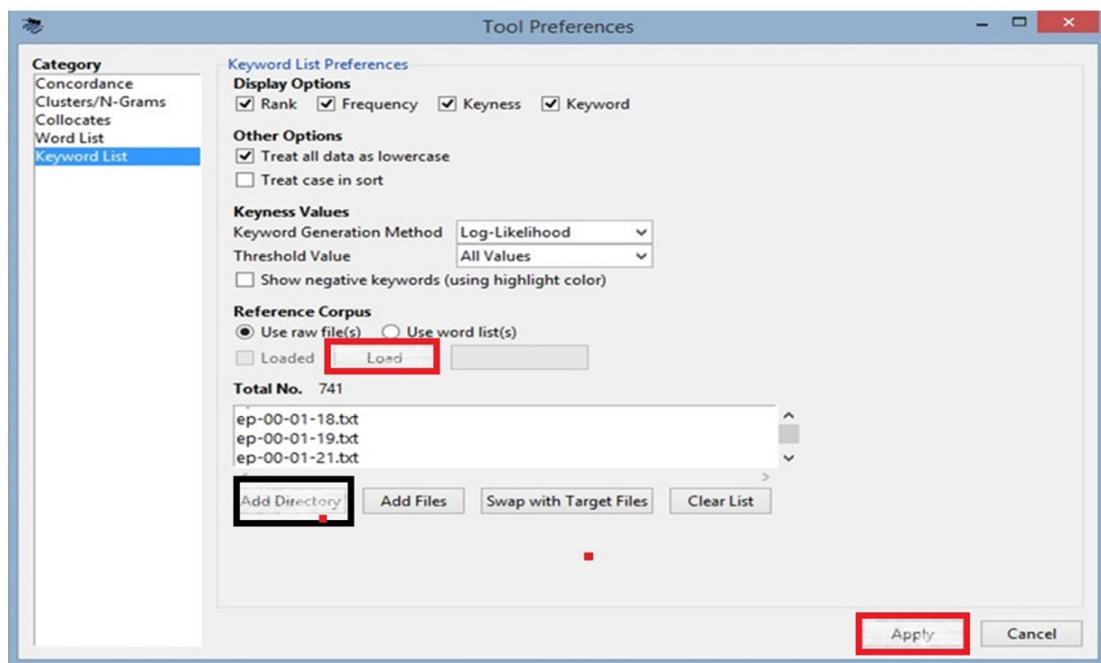
Fonte: Programa computacional de análises linguísticas AntConc⁵ de Laurence Antony

Em (1) “Open Files” foram selecionados os arquivos em txt de uma pasta do computador (em que foram salvas durante a seleção dos corpora) para o processamento dos dados. Em (2) a ferramenta apresenta o número de arquivos que foram selecionados. Para gerar a lista de palavras, é necessário clicar em “start”.

A segunda função utilizada foi a *Keyword List* que extrai as palavras-chave do corpus por meio da comparação entre o corpus de estudo com o corpus de referência. Para a utilização dessa ferramenta foi preciso clicar na opção “Keyword list” e depois em “Tools preferences”.

⁵ Disponível em: <<http://www.laurenceanthony.net/software.html>> Acesso em: 20 jun. 2017.

Figura 2: AntConc 3.4.4 (inserindo corpus de referência)



Fonte: Programa computacional de análises linguísticas AntConc de Laurence Antony

Em “Tools preferences” foi clicada na opção “Keyword list”, localizada na parte esquerda do programa. Na função “add directory” realizou-se a seleção da pasta de corpus e dos corpora de referência. Em seguida foi pressionado o botão “Load” para o programa carregar os dados.

Logo que os dados foram baixados, o próximo passo foi clicar em «Apply», no canto inferior do programa. Depois disso, o programa direcionou para a página inicial, onde foi clicado em “start”, localizado na parte inferior para que as palavras-chave fossem apresentadas.

Figura 3: AntConc 3.4.4 (extraído as palavras-chave)

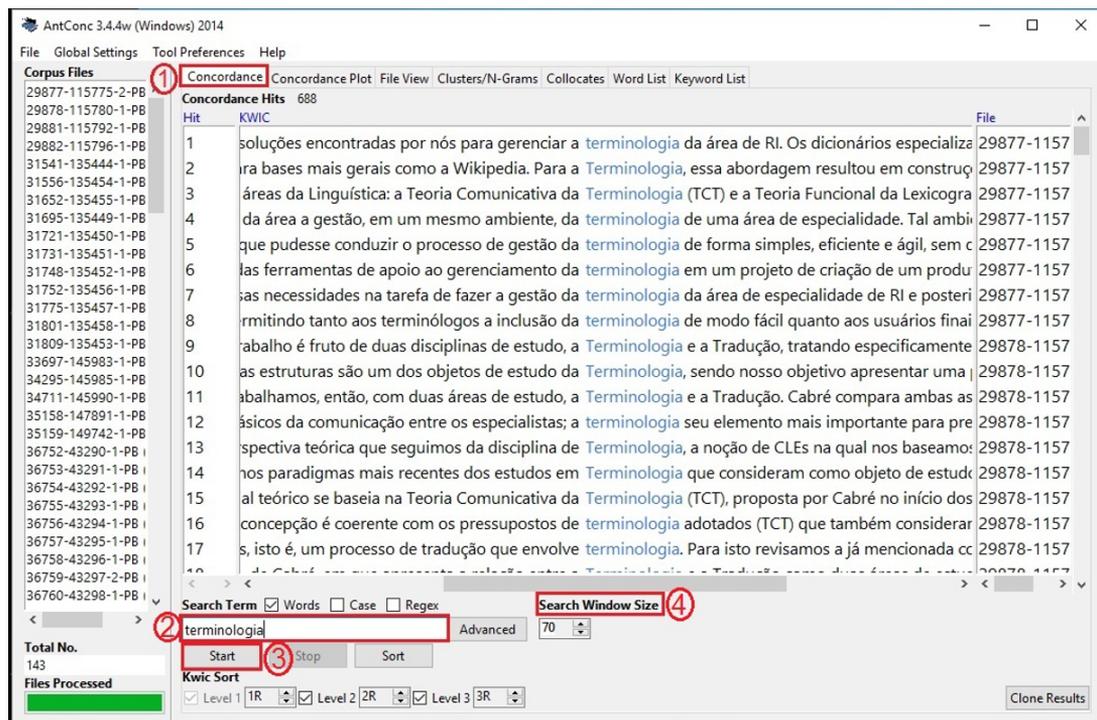
The screenshot shows the AntConc 3.4.4w (Windows) 2014 interface. The main window displays a concordance table with the following columns: Rank, Freq, Keyness, and Keyword. The search term is 'tradutor'. The table lists 59 results, with the keyword 'palavras' appearing at rank 59. The interface also includes a 'Corpus Files' list on the left, a 'Search Term' field containing 'tradutor', and various search options like 'Words', 'Case', and 'Regex'. The 'Hit Location' is set to 'Search Only' with a value of 0. The 'Reference Corpus' is checked as 'Loaded'. The 'Start' button is highlighted with a red box.

Rank	Freq	Keyness	Keyword
29	337	2539.031	dicionário
30	349	2517.422	brasileira
31	399	2501.573	literatura
32	323	2091.267	equivalência
33	275	2088.775	poema
34	727	2081.297	estudo
35	264	2076.609	dicionários
36	636	2003.584	línguas
37	434	1989.206	inglês
38	278	1963.463	brasileiro
39	252	1957.660	caráter
40	252	1937.203	ato
41	583	1881.957	estudos
42	504	1880.197	unidades
43	272	1860.650	variação
44	236	1837.403	legendas
45	224	1774.307	lexicais
46	325	1744.440	intérpretes
47	5718	1724.230	como
48	241	1708.887	atividades
49	325	1697.011	alunos
50	343	1691.168	teoria
51	221	1691.065	terminológica
52	380	1661.800	traduzir
53	275	1659.913	contextos
54	206	1652.713	tradutório
55	392	1611.495	obras
56	219	1592.179	poesia
57	235	1562.258	textual
58	1372	1521.845	termos
59	871	1501.477	palavras

Fonte: Programa computacional de análises linguísticas AntConc de Laurence Antony

Após selecionar os termos que interessam para o estudo, o acadêmico poderá observar como a terminologia é utilizada em diferentes contextos e quais as palavras que, normalmente, coocorrem com ela.

Figura 4: AntConc 3.4.4 (concordância da palavra terminologia)



Fonte: Programa computacional de análises linguísticas AntConc de Laurence Antony

Para analisar os termos em contexto o programa computacional *AntConc* apresenta a ferramenta *Concordance*, que apresenta o contexto que os termos estão inseridos em diferentes artigos científicos. Com a aba *concordance* aberta (1), é possível pesquisar (2) um termo específico para visualizar diferentes contextos em que o termo aparece dentro do corpus. Ao digitar e clicar em “start” (3) o termo aparece centralizado e destacado em cor azul. Para escolher o tamanho da linha de concordância é necessário aumentar o número de caracteres em *search window size* (4).

Os procedimentos acima podem servir de auxílio, sobretudo, na escrita acadêmica. A visualização de termos técnicos dentro de um contexto específico facilita a compreensão de um vocabulário novo.

Muitos projetos são desenvolvidos seguindo essa metodologia na área de estudo da Linguística de Corpus, porém, é possível notar que a maioria das pesquisas, ferramentas computacionais e corpora compilados são voltados para a língua inglesa e tradução.

3 RESULTADOS E DISCUSSÃO

Com base na comparação entre o corpus de referência e o corpus especializado, a função *Keyword* do software *AntConc* (3.4.4) forneceu uma lista de palavras-chave do corpus de estudo.

Quadro 1: Termos mais frequentes do corpus de estudo

continua

Frequência	Palavras-chave
3950	Tradução
2287	Língua
1285	Tradutor
1802	Texto
690	Terminologia
625	Corpus
1095	Termo
430	Léxico
695	Linguagem
516	Linguística
337	Dicionário
323	Equivalência
272	Variação
173	Lexicografia
737	Análise
140	Corpora
194	Especializada
161	Ferramenta
328	Unidade
337	Ensino
99	Significação
122	Vocabulário
114	Semântica
381	Conceito
151	Metodologia
909	Relação
232	Fonte
492	Palavra
54	Terminografia
352	Definição
164	Aprendizagem
202	Significado
154	Frase
49	Lexicologia
290	Expressão
166	Identificação

Quadro 1: Termos mais frequentes do corpus de estudo

continua

Frequência	Palavras-chave
58	Colocações
521	Contexto
444	Dados
54	Enunciação
115	Ocorrência
162	Frequência
63	Ficha
84	Correspondência
35	Coletânea
34	Extração
80	Variedade
35	Referencial
101	Classificação
116	Comparação
70	Frequentes
345	Formação
77	Ideologia
73	Descrever
149	Diferença
229	Conteúdo
45	Nomenclatura
104	Formal
92	Norma
71	Especificidades
119	Estratégias
36	Recorrentes
155	Técnicas
33	Assimilação
18	Concordanciador
358	Produção
54	Interpretar
118	Identidade
114	Artigos
146	Específico
196	Procedimentos
685	Exemplo
32	Compilação
107	Representação
137	Técnica
31	Manuais
65	Relevância
35	Recorrente
49	Pragmática

Frequência	Palavras-chave
208	Construção
369	Comunicação
118	Reflexão
136	Analisar
720	Sentido
158	Competência
41	Paralelo
203	Referência
437	Sistema
217	Resultados
117	Tratamento

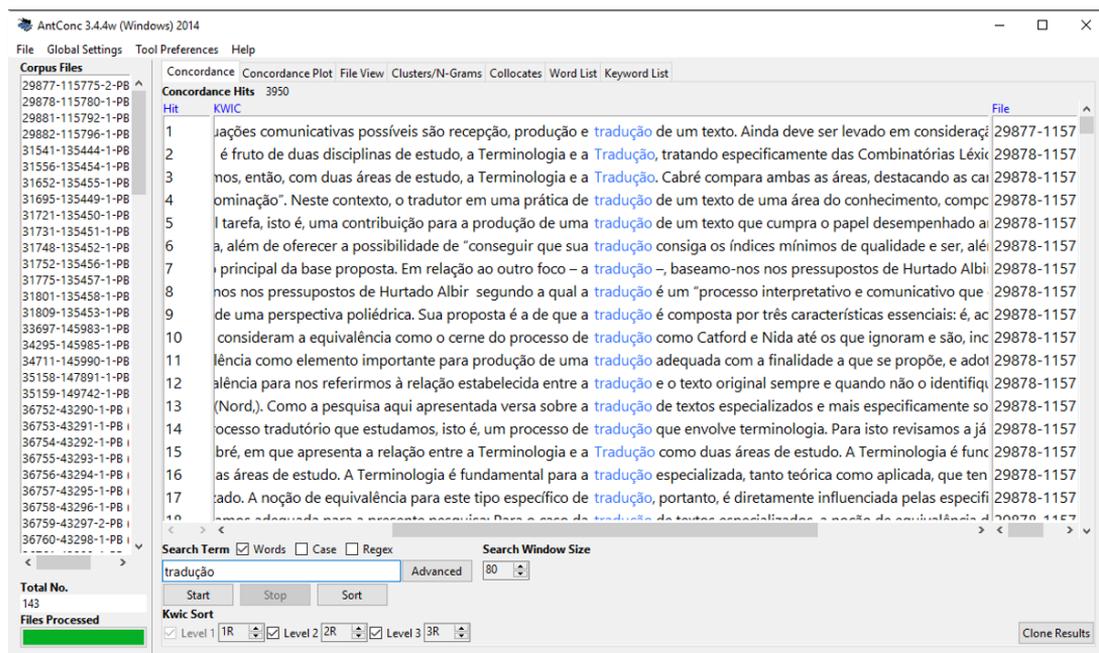
FONTE: Terminologia extraída do corpus de estudo da área de terminologia

Os dados coletados representam os termos mais frequentes apresentados no corpus de estudo. A análise dos termos mais recorrentes no corpus proporcionam melhor compreensão de uso no contexto real da escrita acadêmica. Por exemplo: (i) quando o acadêmico possui dúvidas quanto ao seu uso em contexto; (ii) quando quer saber quais palavras acompanham determinado item lexical; (iii) para obter informações sobre o conteúdo da área; (iv) compreender o significado da palavra dentro do seu contexto; (v) identificar elementos gramaticais; (vi) observar generalidades estilísticas do gênero textual; e observar especificidades dos movimentos retóricos (introdução, metodologia, resultados, discussão e conclusão).

3.1 TRADUÇÃO

O primeiro termo da lista de palavras-chave é “tradução”, que no singular obteve 3950 ocorrências e no plural “traduções” atingiu 739 ocorrências. Percebe-se que é um termo relevante para a área da Terminologia. Esse termo aparentemente não possui nenhum mistério para qualquer acadêmico. No entanto, muitas vezes o aluno precisa perceber como ela é empregada no seu contexto e quais são as possibilidades lexicais que essa palavra proporciona ao texto. Segundo Sinclair (1991, p. 101) um grande número de frases construídas constituem escolhas únicas, ou seja, existem modelos prontos que são padronizados ao uso em determinadas situações. Para isso a ferramenta *Concordance* possibilita uma melhor análise de como o termo foi utilizado no corpus.

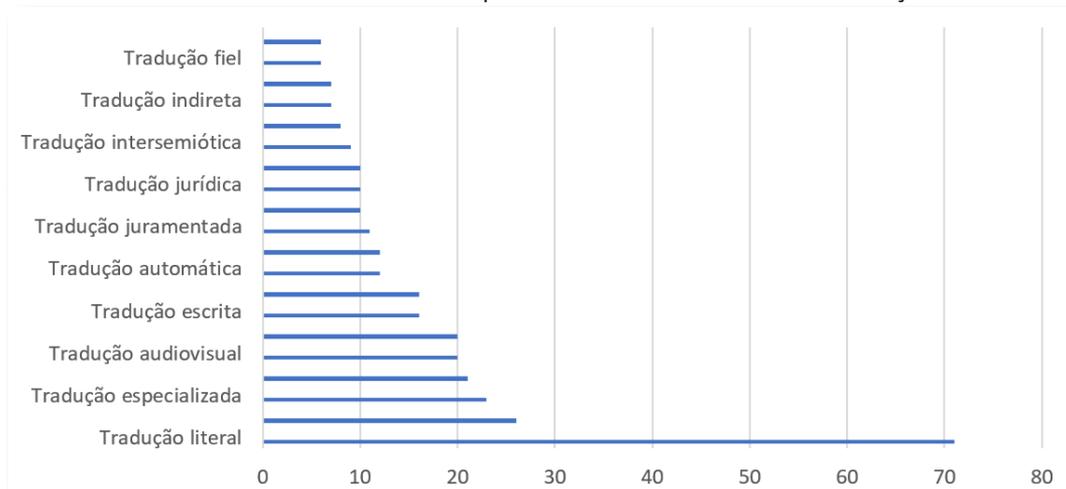
Figura 5: Linha de concordância do termo Tradução



Fonte: Corpora comparáveis da área da Terminologia

A linha de concordância mostra a palavra no centro e seu contexto ao redor. Pode-se perceber com isso que “tradução” está, normalmente, associada aos termos: terminologia, processo de tradução, prática de tradução, estudos da tradução, etc. Com essa ferramenta o acadêmico consegue tirar qualquer dúvida referente ao uso da palavra em contexto. Ao clicar na palavra em azul, também pode verificar o artigo completo, caso tenha interesse, pois o programa direciona para a opção “view file”.

Outra ferramenta importante para o acadêmico é o “Cluster”, esta escolha apresenta as palavras que, geralmente, coocorrem com o termo de estudo. Seguem alguns exemplos: Gráfico 1 – Ocorrências de clusters no corpus de estudo do item lexical “tradução”



Fonte: Artigos científicos da área de terminologia.

Os clusters acima mostram quais *são* os termos mais frequentes que ocorrem com a palavra tradução nos artigos científicos das revistas TradTerm, GTLex e Debate Terminológico entre o período de 2010 a 2016. O termo Tradução Literal é o mais frequente no corpus de estudo, apresentando 71 ocorrências. Ao clicar no cluster o programa abre uma lista com os 71 exemplos em uma linha de concordância. Conforme figura apresentada abaixo:

FIGURA 6: Clusters do item lexical Tradução Literal

clusive, foi preciso acrescentar a palavra "nem", e yet foi deslocado para o final da frase. Assim, a *tradução literal* que deveria ser "Ainda não tenho 16", passa a ser "Não tenho nem dezesseis anos: os alunos e, principalmente, porque tal linguagem evidencia a impossibilidade de se adotar uma *tradução literal*. Embora muito conhecido, julga-se importante uma breve contextualização sobre itivo As traduções realizadas pelos alunos ilustram a prevalência de que eles não buscaram uma *tradução literal* para as expressões "Deu uns amassos", "Era só o que faltava" e "Climão, hein?", e ia introdução na língua alvo, ou com a manutenção da forma original da expressão emprestada. *tradução literal*: mais conhecida como tradução "palavra por palavra", é mais comum, segundo os es, a autora buscou fugir da proposta dicotômica até então utilizada, baseada na oposição entre *tradução literal* e livre. Procura, com seu modelo, oferecer ao tradutor "a possibilidade de mover- mesma família. Sua crítica ao modelo de Vinay e Darbelnet reside, principalmente, na dicotomia *tradução literal* x tradução livre (que na terminologia dos autores é tradução direta x tradução obl dutórios. O fato de caracterizarem o empréstimo, por exemplo, como primeiro procedimento da *tradução literal*, destacado como o mais fácil, é criticado por Barbosa, que afirma que em sua con e estrangeirismo por uma explicação, utilizada de acordo com a finalidade da tradução; decalque: *tradução literal* de sintagmas ou frases da língua original para a língua de tradução; e adaptação: ilidade extralinguística das duas línguas em questão na tradução: tradução palavra por palavra e *tradução literal*; - divergência do sistema linguístico: transposição, modulação e equivalência; - di -se para afastar-se de alguns de seus postulados, principalmente o da dicotomia tradução livre x *tradução literal* tão presente na proposta dos franceses, sem impor limites muito rígidos em sua i na língua meta, sofre alterações morfológicas ou gráficas para adequar-se às regras linguísticas; *tradução literal*: considera esta modalidade sinônima de tradução palavra por palavra; transposi n em diferentes tipos de textos. Entre as conclusões de Aubert, podemos apontar o fato de que a *tradução literal*, a transposição e a modulação são as modalidades mais frequentes identificadas i tificadas em suas análises. Destaca-se, como já mencionado anteriormente, o resultado de que a *tradução literal* é a modalidade mais frequente na tradução do inglês para o português, resultado

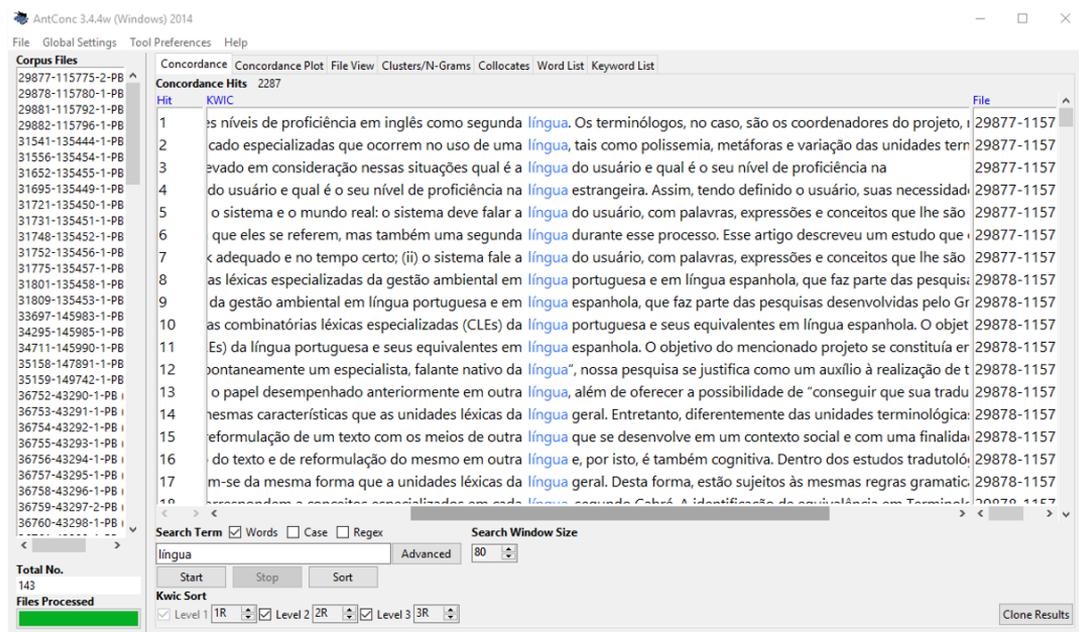
Fonte: Corpus da área de terminologia

A partir da lista dos clusters, apresentada no gráfico 1, pode-se perceber que muitos desses termos evidenciam tipos de traduções (tradução direta, indireta livre, jurídica, etc.). Normalmente, o acadêmico iniciante reconhece a terminologia "tradução", mas desconhece os tipos de traduções. Nesse sentido, o corpus pode fornecer informações e dar *insights* que o estudante jamais teria a olho nu. Reconhecer os colocados de uma palavra auxilia o aprendiz a produzir um texto mais fluente, ou seja, em que o fluxo das informações *ocorram* de forma natural, conforme padrões aceitáveis para a comunidade acadêmica.

3.2 LÍNGUA

O segundo termo analisado é "Língua", que teve 2287 ocorrências no singular e no plural "línguas" 636 vezes.

Figura 7: Linha de concordância do item lexical língua

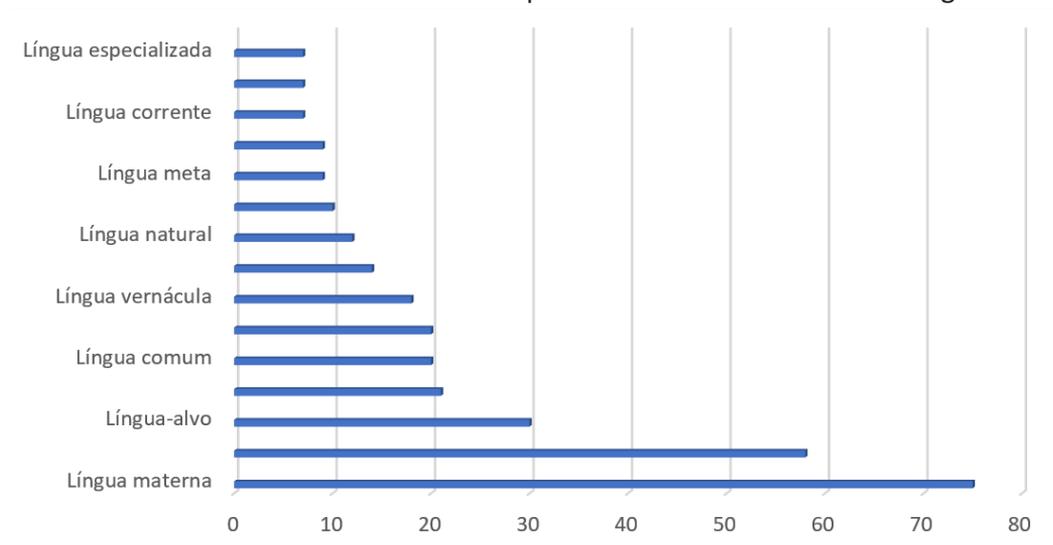


Fonte: Corpora comparáveis da área da Terminologia

A frequência do uso do termo “língua” está diretamente relacionada ao objeto de estudo da terminologia e tradução. É a partir do uso da língua que as análises e elaborações de glossários monolíngues e bilíngues são realizadas. Logo, na linha de concordância 1, pode-se compreender que essa ciência trata da pesquisa e comparação de uma língua estrangeira com a língua materna.

Com isso, pode-se perceber que ao falar de “língua” na Terminologia o produtor do discurso se refere ao processo cognitivo adquirido pelas pessoas para se comunicar, exemplo: “Os usuários finais do glossário são alunos com diferentes níveis de proficiência em inglês como segunda língua” (WILKENS et al. 2012, p.2). Diferente da inferência que ocorre ao se tratar de “língua” na escrita de um profissional da saúde, exemplo: “A cavidade bucal é acometida por diversas enfermidades, muitas das quais ocorrem na língua” (CONCEIÇÃO et al. 2010, p.11). Neste sentido, a questão semântica é influenciada pelo contexto em que a palavra é inserida.

Gráfico 2 – Ocorrências de clusters no corpus de estudo do item lexical “língua”



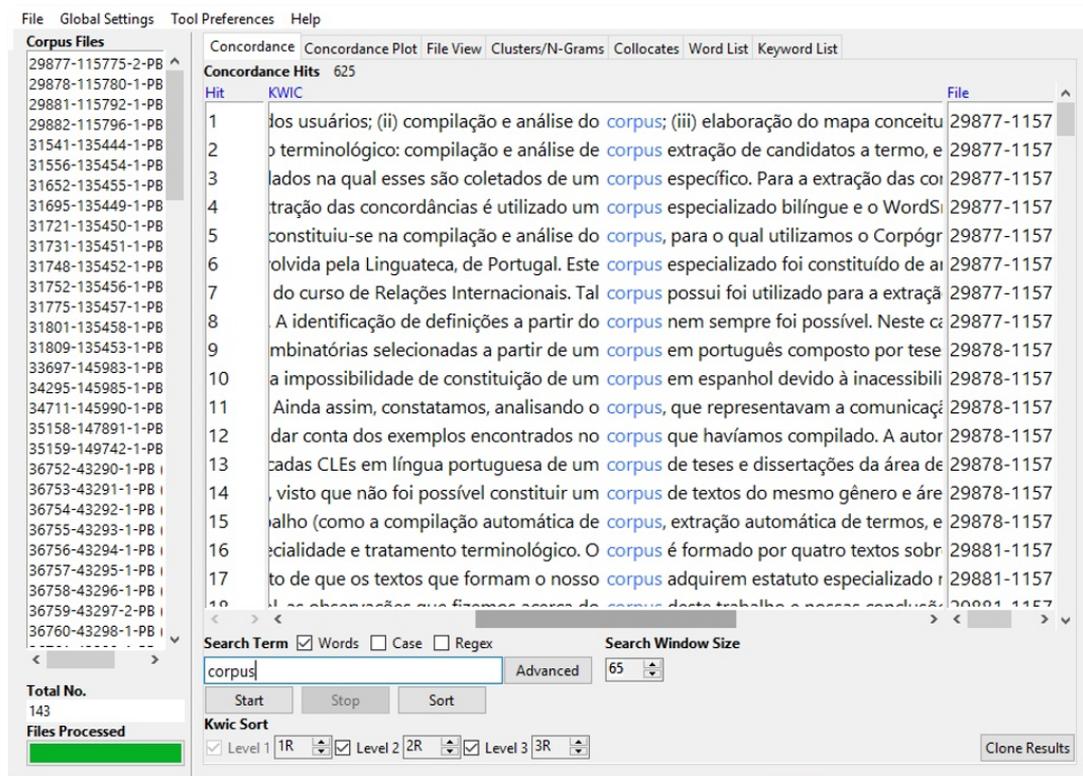
Fonte: Artigos científicos da área de terminologia.

O gráfico acima demonstra com qual frequência os colocados ocorrem dentro dos artigos científicos da área de terminologia. Ao observar esses elementos o acadêmico consegue perceber quais léxicos ou combinações de palavras ele pode selecionar para a escrita acadêmica e produções textuais da sua área de estudo. Desta forma, o aprendiz garante a naturalidade na escrita do gênero textual e compreensão por parte da comunidade acadêmica que já está habituada com os termos da área.

3.5 CORPUS

O termo “corpus” apresentou 635 ocorrências no singular, e no plural (corpora) 140. Esse é um termo com alta frequência na área de terminologia, pois, atualmente, é praticamente impossível falar em extração de termos, compilação de dicionários e glossários sem falar em corpus. Isso ocorre porque a tecnologia e o avanço nos processamentos de dados por meio de softwares, desenvolvidos especialmente para o estudo da linguagem, facilitaram muito a vida dos profissionais de linguística.

Figura 8: Linha de concordância do item lexical Corpus

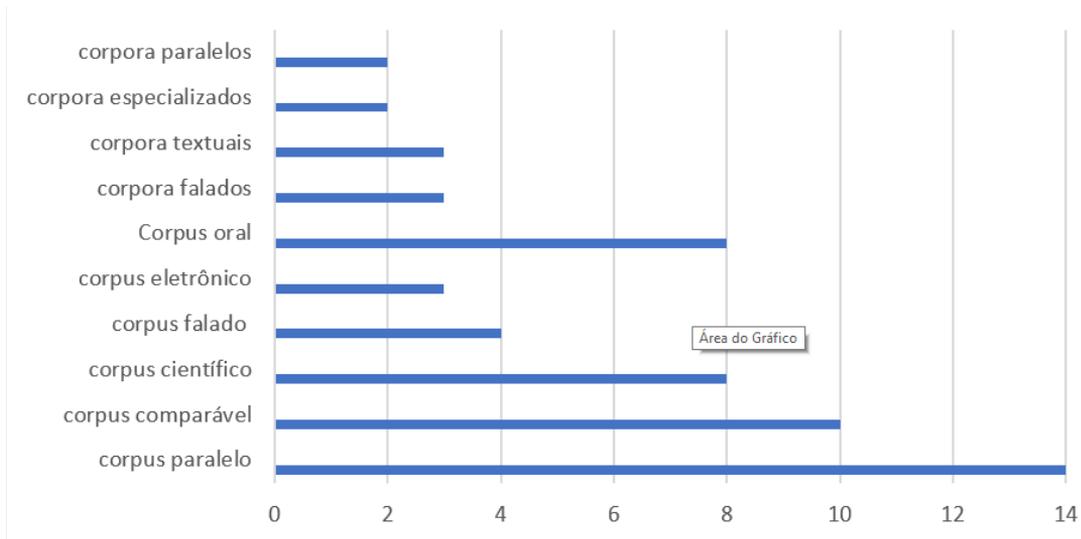


Fonte: Corpora comparáveis da área da Terminologia

Ao encontrar esse termo específico da área, o acadêmico iniciante pode ter dificuldades para fazer o uso dele na produção escrita, mas tal dificuldade pode ser superada com a observação do termo dentro do contexto.

A palavra corpus (plural: corpora) significa texto, portanto, o estudo da linguística de corpus é o estudo da linguagem por meio de textos produzidos naturalmente no seu ambiente real de construção. Exemplos, (i) “Apesar dos obstáculos colocados, os **corpora** permitem analisar maior volume de dados (...)” (GINEZI, 2014, p. 174); (ii) “Este **corpus** especializado foi constituído de artigos acadêmicos com alto fator de impacto, selecionados a partir de temáticas tratadas nos primeiros semestres do curso de Relações Internacionais!” (WILKENS et al. 2012, p.5).

Gráfico 3 – Clusters no corpus de estudo do item lexical “corpus e corpora”



Fonte: Artigos científicos da área de terminologia.

A análise dos clusters também auxilia o pesquisador na identificação dos termos mais frequentes dentro da área de estudo. Segundo Sardinha (2004, p.31) a linguagem nesta perspectiva é vista como um sistema probabilístico, portanto, as escolhas lexicais não ocorrem de forma aleatória e, sim, existem algumas possibilidades de opções que devem ser consideradas. Neste sentido, as ferramentas de análises lexicais facilitam e agilizam a observação destes fenômenos linguísticos.

CONSIDERAÇÕES FINAIS

Nesse artigo foi apresentada uma metodologia de extração e análise de dados baseada na linguística de corpus para o estudo da linguagem acadêmica científica. O processamento dos dados foi realizado com o auxílio da ferramenta computacional *AntConc* e as aplicabilidades que ela oferece. Esta ferramenta foi escolhida por ser gratuita e de fácil compreensão para os estudantes e pesquisadores iniciantes.

O objetivo deste projeto foi demonstrar aos acadêmicos, pesquisadores e demais interessados na produção da escrita científica outro viés de desenvolvimento e aquisição do processo cognitivo para a redação de artigos científicos. Acredita-se que o indivíduo precise refletir e aprender sobre a linguagem e estilo da produção de textos científicos para desenvolver uma composição de maior qualidade que transmita as informações e o real propósito de sua pesquisa de forma fluente e clara para o leitor.

No ambiente de produção acadêmica ainda há pouco espaço para a análise de elementos linguísticos, ou seja, poucas pesquisas sobre léxicos e terminologias vêm sendo desenvolvidas com o propósito de auxiliar na elaboração de uma melhor produção escrita. Nesta perspectiva é que essa metodologia de pesquisa surge, no sentido de auxiliar o pesquisador na organização do seu processo de aprendizagem para a escrita do gênero textual acadêmico, agilizando o processo e dando confiabilidade no que é produzido.

Espera-se que com esse estudo os especialistas de diferentes áreas possam refletir sobre a trajetória de suas próprias redações e produções de textos científicos, valorizando a linguagem e reconhecendo que utilizar as palavras adequadas para cada gênero textual pode fazer com que seus artigos ganhem mais prestígio na comunidade científica.

Portanto, conclui-se que por mais que seja um processo inicialmente árduo, a coleta e análise desses dados faz com que o pesquisador amplie significativamente seu ângulo de conhecimento dentro do processo da produção da escrita científica, pois percebe fatores importantes a partir de exemplos reais produzidos por outros autores. Isso ocorre devido à facilidade de manipulação dos dados que o programa oferece. Com os avanços tecnológicos pode-se observar uma quantidade imensa de dados que jamais seria possível sem os recursos computacionais.

REFERÊNCIAS

ALBERTS-FRANCO, C. Linguística de corpus e terminologia bilíngue: o programa Antconc e a extração de termos em alemão. **The ESpecialist**, São Paulo, v. 36, n. 2, p. 182, 2015. Disponível em: <<https://revistas.pucsp.br/index.php/esp/article/view/23977>>. Acesso em: 10 fev. 2017.

ALMEIDA, L. S. Facilitar a aprendizagem: ajudar aos alunos a aprender e a pensar. **Psicologia Escolar e Educacional**, Campinas, v. 6, n. 2, p. 155-165, dez. 2002. Disponível em: <http://www.scielo.br/scielo.php?pid=S1413-85572002000200006&script=sci_abstract&tlng=pt>. Acesso em: 20 nov. 2016.

BAKER, M. Corpora in translation studies: an overview and some suggestions for future research. *Jonh Benjamins*, Amsterdam, v. 7, n. 2, p. 223-243, 1995. Disponível em: <<https://www.researchgate.net/publication/233656559>>. Acesso em: 20 jan. 2017.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University, 1998.

CABRÉ, M. T. *La terminología: representación y comunicación – elementos para una teoría de base comunicativa e otros artículos*. Barcelona: IULA, 1999.

_____. Theories of Terminology, their description, prescription and explanation. *Terminology*, Amsterdam, v. 9, n. 2, p. 163-199, 2003. Disponível em: <<http://www.jbe-platform.com/content/journals/10.1075/term.9.2.03cab>>. Acesso em: 20 jan. 2017.

CARVALHO, E. M. F. Metodologia de construção de um glossário bilíngue com base em um corpus de domínio técnico. 2007. 81 f. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em: <<http://resitorio.ufsc.br/bitstream/handle/123456789/90029/241755.pdf?sequence>>. Acesso em: 20 jan. 2017.

CHOMSKY, N. **Reflections on Language**. New York: Pantheon Books, 1975.

CONCEIÇÃO, L. D. et al. Estudo retrospectivo de biópsias em língua: aspectos epidemiológicos. **Revista da Faculdade de Odontologia**, Passo Fundo, v. 15, n. 1, p. 11-19, jan./abr. 2010. Disponível em: <revodonto.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-40122010000100003&lng=es&nrm=iso>. Acesso em: 25 nov. 2016.

COSTA FILHO, J. E. Uma proposta de glossário dos termos da teoria da metáfora conceitual. **Debate Terminológico**, Porto Alegre, n. 10, 2013. Disponível em: <seer.ufrgs.br/index.php/riterm/article/view/43964/27609>. Acesso em: 2 fev. 2017.

COULTHARD, R. J. The application of corpus methodology to translation: the JPED parallel corpus and the pediatrics comparable corpus. 2005. 159 f. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2005.

Disponível em: <<http://repositorio.ufsc.br/bitstream/handle/123456789/102257/222327.pdf?sequence=1&isAllowed=y>>. Acesso em: 28 nov. 2017.

DUTRA, D. P.; SILERO, R. P. Descobertas linguísticas para pesquisadores e aprendizes: a linguística de Corpus e o ensino de gramática. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 10, n. 4, p. 909-930, 2010. Disponível em: <www.scielo.br/pdf/rbla/v10n4/a05v10n4.pdf>. Acesso em: 2 fev. 2017.

- FROMM, G. VoTec: A construção de vocabulários eletrônicos para aprendizes de tradução. São Paulo: Universidade de São Paulo. 2007. 215 f. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Universidade de São Paulo, São Paulo, 2007. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855/publico/TESE_GUILHERME_FROMM.pdf>. Acesso em: 3 mar. 2017.
- GALVÃO, M. C. B. A linguagem de especialidade e o texto técnico-científico: notas conceituais. Revista: **Transinformação**, Campinas, v. 16, n. 3, p. 241-251, 2014. Disponível em: <periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/711>. Acesso em: 3 abr. 2017.
- GAVIOLI, L. **Exploring corpora for ESP learning**. Amsterdam: John Benjamins Publishing Company, 2005.
- GINEZI, L. L. Desafios para a construção de um corpus de aprendizes de Interpretação Simultânea. Tradterm, São Paulo, v. 23, p. 165-191, 2014. Disponível em: <<https://www.revistas.usp.br/tradterm/article/view/85576>>. Acesso em: 3 abr. 2017.
- HUNSTON, S. Corpora in applied linguistics. Cambridge: Cambridge University, 2002.
- KADER, C. C. C.; RICHTER, M. G. Linguística de corpus: possibilidades e avanços. **Instrumento**: Revista de Estudo e Pesquisa em Educação, Juiz de Fora, v. 15, n. 1, 2013. Disponível em: <<https://instrumento.ufjf.emnuvens.com.br/revistainstrumento/article/view/2641>>. Acesso em: 3 mar. 2017.
- KRIEGER, M. da G. Do ensino da terminologia para tradutores: diretrizes básicas. **Cadernos de Tradução**, Florianópolis, v. 1, n. 17, p. 190-206, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/traducao/article/view/6862>>. Acesso em: 25 fev. 2017.
- LEPRE, L. A elaboração de glossários bilíngues para a interpretação de textos em inglês com base em um corpus paralelo. 2007. 81 f. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/90658/257757.pdf?sequence=1>>. Acesso em: 10 dez. 2016.
- MAIA, B. Using corpora for terminology extraction: pedagogical and computational approaches. Lódz: Studies in Language, 2003.
- MARIAN, J. O estudo da linguística de corpus para a tradução especializada: elaboração de um glossário da área da informática – manutenção de Computadores. **Cultura e Tradução**, João Pessoa, v. 3, n. 1, p. 197-209, 2014. Disponível em: <periodicos.ufpb.br/index.php/ct/article/viewFile/21567/12170>. Acesso em: 20 nov. 2016.
- _____. O uso de corpora como ferramenta de apoio para tradução: uma análise das ocorrências do item lexical “hearing”. 2010. 86 f. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2010. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/93453/282627.pdf?sequence=1&isAllowed=y>>. Acesso em: 20 nov. 2016.
- _____. Terminologia especializada: um estudo baseado na linguística de corpus. **Caderno PAIC**, Curitiba, v. 16, n. 1, p. 475-487, 2016. Disponível em: <<https://cadernopaic.fae.edu/cadernopaic/article/view/109/108>>. Acesso em: 15 nov. 2016.

MARIAN, J.; LIMA, R. A padronização da linguagem e os colocados: uma análise de erros comuns apresentados por aprendizes de língua inglesa. *Revista da FAE, Curitiba*, v. 19, n. 1, p. 118-129, jan./jun. 2016. Disponível em: <<https://revistafae.fae.edu/revistafae/article/view/78/59>>. Acesso em: 15 nov. 2016.

MATSIRA, M. Extracting terms from an english-greek popular science parallel corpus for translation teaching purpose. 120 f. Dissertação (Mestrado em Linguística) – University of Birmingham, Birmingham, 2008. Disponível em: <[etheses.bham.ac.uk/131](https://theses.bham.ac.uk/131)>. Acesso em: 5 dez. 2016.

OLSON, K. A brief guide to writing the psychology paper. **Harvard College**: Writing Center, p. 1-4, 2008. Disponível em: <https://writingproject.fas.harvard.edu/files/hwp/files/bg_psychology.pdf>. Acesso em: 5 dez. 2016.

PAVEL, S.; NOLET, D. **Manual de terminologia**. Tradução: Enilde Faulstich. Quebec: Departamento de Tradução do Governo Canadense, 2002.

PORTOLAN, A. C. Uma contribuição para a elaboração de um glossário bilíngue na área de pediatria com base em linguística de corpus. 2011, 154 f. Dissertação (Mestrado em Estudos da Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2011. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/106697>>. Acesso em: 3 dez. 2016.

SARDINHA, T. B. Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução, Florianópolis*, v. 1, n. 9, p. 15-59, 2002. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/5980>>. Acesso em: 20 nov. 2016.

_____. **Linguística de corpus**. São Paulo: Manole, 2004.

SINCLAIR, J. *Corpus, concordance, collocation*. Oxford: Oxford University, 1991.

_____. *How to use corpora in language teaching*. Oxford: Oxford University, 2004.

TEIXEIRA, E. D. A linguística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual. São Paulo: Universidade de São Paulo, 2008, 439 f. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Universidade de São Paulo, São Paulo, 2008. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8147/tde-16022009-141747/publico/TESE_ELISA_DUARTE_TEIXEIRA.pdf>. Acesso em: 4 dez. 2016.

TOGNINI-BONELLI, E. *Corpus linguistics at work*. Amsterdam: J. Benjamins, 2001.

WEBER, J.-J. A concordance-and genre-informed approach to ESP essay writing. *ELT Journal*, v. 55, n. 1, p. 14-20, jan. 2001. Disponível em: <citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.4126&rep=rep1&type=pdf>. Acesso em: 20 mar. 2017.

WILKENS, R. et al. Ambientes web de gestão terminológica para a criação de productos terminológicos on-line. *Debate Terminológico, Porto Alegre*, n. 8, p. 16-22, 2012. Disponível em: <seer.ufrgs.br/riterm/article/view/29877>. Acesso em: 25 jan. 2017.