

## SUBSTANTIVOS EM ARTIGOS CIENTÍFICOS: O VOCABULÁRIO TÉCNICO E ACADÊMICO NA PERSPECTIVA DA LINGUÍSTICA DE *CORPUS*

---

MONIQUE VIEIRA MIRANDA \*  
APARECIDA DE ARAÚJO OLIVEIRA\*\*

---

### RESUMO

Foram analisados os 30 substantivos mais recorrentes em um *corpus* misto de artigos acadêmicos de diferentes áreas de conhecimentos, utilizando-se ferramentas da Linguística de *Corpus*. O objetivo foi classificar esses itens lexicais como integrantes das categorias: vocabulário acadêmico, vocabulário técnico e vocabulário de uso geral. Foi realizada uma análise semântica não automática de 35.320 linhas de concordância – número correspondente à soma dos totais de ocorrências em contexto para cada um desses substantivos –, a fim de se obter seus significados de uso mais comum. Concluiu-se que o número de lexemas de uso acadêmico é maior no *corpus* composto por diferentes áreas que o obtido na análise de *subcorpora* separados por área de conhecimento, que apresentam termos técnicos entre os substantivos mais frequentes.

PALAVRAS-CHAVE: vocabulário acadêmico, vocabulário técnico, substantivos, *corpus*.

---

### 1. CONSIDERAÇÕES INICIAIS

O domínio dos gêneros acadêmicos é importante em todas as etapas do ensino superior, tanto para os falantes nativos de uma língua

---

\* Mestre em Linguística Teórica e Descritiva pela Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil. E-mail: nk.miranda@gmail.com

\*\* Doutora em Linguística pela Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil. Professora Adjunta na Universidade Federal de Viçosa (UFV), Minas Gerais, Brasil. E-mail: aparecida.oliveira@ufv.br

quanto para os estrangeiros. A habilidade de ler e compreender um desses gêneros depende, em grande parte, do tamanho do vocabulário que cada indivíduo domina, especialmente daqueles que desejam aprendê-lo em uma segunda língua (DELL'ISOLLA, 2005). Além de um conhecimento básico do idioma, o estudante precisará ter um domínio do vocabulário acadêmico e do vocabulário específico de sua área de conhecimento.

Em termos de vocabulário de uso geral, Berber Sardinha (2004) e Biderman (1998 apud DELL'ISOLA, 2005, p. 25-26) propõem listas das três mil palavras mais frequentes do Português do Brasil (PB), retiradas de amostras da linguagem oral e escrita. Para o ensino da língua inglesa, a *General Service List* (GSL) (WEST, 1953), com as duas mil palavras mais frequentes, é uma das mais conhecidas e úteis. Essa lista ainda continua atual e cobre entre 80% e 90% das palavras em um texto, segundo Chung e Nation (2003).

Do total geral de palavras (*tokens*) em textos da esfera acadêmica, o volume dos vocabulários acadêmico (presente na “Academic Word List”, de Coxhead, 2000) e técnico fica em torno de 8,5% e 5% respectivamente (CHUNG; NATION, 2003; NATION, 2001). Apesar de esses números parecerem pouco significativos, esses itens lexicais são fundamentais para o sucesso no entendimento do texto e devem ser aprendidos justamente porque não costumam aparecer em outros gêneros.

No presente estudo, buscamos identificar os diferentes usos dos 30 substantivos mais frequentes em um *corpus* formado por artigos científicos de seis áreas de conhecimento distintas, escritos em português do Brasil, desambiguando formas pertencentes a mais de uma classe gramatical e classificando esses substantivos como parte do vocabulário *geral*, *acadêmico* ou *técnico*, cujas definições serão apresentadas mais adiante. Através de ferramentas da Linguística de *Corpus*, foram obtidas linhas de concordância, colocações, coligações e funções textuais das palavras de uso prototipicamente acadêmico. Entendemos que esta pesquisa será útil para o ensino de Português para fins acadêmicos, especialmente para falantes nativos de outras línguas.

## 1. FUNDAMENTOS TEÓRICOS

### 1.1 A TERMINOLOGIA EMPREGADA

Neste artigo, o termo *palavra* é sinônimo de *forma* e essas expressões podem se alternar ao longo do texto, com o sentido atribuído por Trask (2004, p. 218) à “palavra ortográfica”, e por Sautchuk (2010, p. 5), correspondendo a cada unidade escrita linearmente, com espaços em branco em ambos os lados. Difere, portanto, de *lexema* ou *item lexical*, que se trata de uma unidade do léxico constituída de uma ou mais palavras e, portanto, tem sua correspondência mais próxima na forma lexicográfica, proposta em dicionários (TRASK, 2004; CRUSE, 2006). Cruse (2006, p. 92) trata um lexema em termos de “uma associação entre a forma e o sentido”, desconsiderando certas variações de forma e de significado<sup>1</sup>.

Portanto, as diferenças de natureza flexional não criam lexemas distintos nos pares “aquisição”/”aquisições” e “apresenta”/”apresentou”. Já as derivações, como “casar”/”casamento”, e os afixos, por exemplo, “considerar”/”desconsiderar” produzem lexemas distintos.

Com relação ao sentido, adotamos definições consagradas na literatura (DUBOIS et al., 2009). Assim, *polissemia* é o fenômeno pelo qual um lexema pode assumir vários sentidos, relacionados entre si. Esse item é chamado “polissêmico” e esses sentidos geralmente surgem na língua por meio de extensão semântica a partir de um sentido primário. Um exemplo de *lexema polissêmico* encontrado no *corpus* é “consumo”, com os sentidos de ‘utilização, pela população, das riquezas, materiais, artigos produzidos’ (ANT, GEO e GEN)<sup>2</sup> e ‘ingestão, utilização’ (NUT e ZOO)<sup>3</sup>. Por sua vez, a *homonímia* ocorre quando dois ou mais lexemas apresentam a mesma forma gráfica e/ou fonológica, associada a sentidos distintos, não relacionados entre si. Como exemplo desse fenômeno, a forma “dado” (GER) aparece nos *subcorpora* com os sentidos de ‘informação’, ‘objeto cúbico, cujas faces são marcadas por números, naipes, figuras etc. e geralmente numeradas por pontos de um a seis’, e como o particípio do verbo “dar”, constituindo, assim, três *lexemas homônimos*.

Ainda sobre o termo lexema, nós o equiparamos, neste artigo, à denominação *lema*, cuja definição, proposta por Francis e Kučera

(1982, p. 1), é “conjunto de formas lexicais tendo o mesmo radical e pertencente à mesma classe de palavras, distinguindo-se apenas em termos de desinências (flexões de gênero, número, modo, pessoa) e/ou grafia”<sup>4</sup>.

*Concordâncias* são amostras dos contextos em que uma palavra ocorre num *corpus*, dispostas em uma sequência de linhas, cuja extensão é definida em termos do número de palavras à direita e à esquerda do termo de busca. Como avalia Berber Sardinha (2004, p. 106-107), as concordâncias são “indispensáveis no estudo da colocação e da padronização lexical e, por isso, fundamentais, na investigação de *corpora*”. No exemplo abaixo, o termo de busca é “medida”.

**FIGURA 1** – LINHAS DE CONCORDÂNCIA DA PALAVRA “MEDIDA” NO SOFTWARE KITCONC 3.0.

ran para a participação efeitos agregados em maior ou menor	medida	intencionais, abrindo terreno para a ...
ds acabariam se desempenhando como fuzzwords. Porém, em boa	medida	, o fim da Guerra Fria subjaz à conflu...
iliar conteúdo, substância ou ... Na sua trajetória, em boa	medida	acompanhando o espírito crítico do se...
io, o teste que se buscou efetuar foi o de verificar em que	medida	preponderavam os vínculos dos conselh...
munidades quilombolas (Soyer, 2009, 2010). Como se sabe, na	medida	em que não faz sentido tomar a noção ...
ficultaram a constituição de comunidades de ex-escravos, na	medida	em que colocaram na condição de liber...
eensão da situação a que se viram inicialmente expostos, na	medida	em que está na base de suas formas de...
mentos importantes na fundação da comunidade de negros, na	medida	em que eles reagiram à tal proibição ...
arcação das fronteiras externas com o mundo dos brancos, na	medida	em que resultou na própria redefiniçã...
versão ao luteranismo dos negros que viviam em solidez, na	medida	em que possibilita considerá-la como ...
sobre a situação dos negros na região fossem acionadas, na	medida	em que tal categoria constitui uma no...

Fonte: Elaborada pelas autoras.

*Colocados* são palavras que ocorrem com frequência à direita ou à esquerda de um item lexical (SARDINHA, 2004, p. 188) e *colocações* são, segundo Firth (1968, p. 181 apud XIAO; McENERY, 2006, p. 105), “declarações dos lugares habituais daquela palavra”<sup>5</sup>. Para Cruse (2006, p. 27), uma colocação é qualquer sentença bem formada gramaticalmente, que não cause estranheza. Em “*an excellent performance*” [um desempenho excelente], pode ser dito que “*excellent*” é um colocado regular de “*performance*” porque talvez ocorra mais comumente nesse contexto que outros adjetivos com sentido semelhante. Ou ainda, trata-se de uma sequência de palavras que geralmente ocorrem juntas, que mantêm unidade semântica. Nesse caso, dependendo do colocado, uma mesma palavra pode expressar

diferentes significados, por exemplo, “*high*” em “*a high wind*” [vento muito forte], “*high seas*” [águas internacionais], “*high office*” [posto de comando] e “*have a high opinion of*” [ter uma opinião muito positiva sobre]. Quantitativamente, destaca Hoey (1991), o termo “colocação” somente se aplica aos casos em que itens lexicais coocorram com frequência mais que aleatória. No presente trabalho, empregou-se a definição de colocação fornecida por Hoey.

*Coligações* são combinações de itens lexicais em que o colocado é uma palavra gramatical (TAGNIN, 2005, p. 30), por exemplo, uma combinação entre verbo e preposição (“confiar em”), substantivo e preposição (“aptidão para”), adjetivo e preposição (“severo com”), advérbios com preposições (“diferentemente de”) ou a combinação de uma ou mais preposições com um sintagma nominal (“em mãos”; “de acordo com”).

Uma *lista de frequência* ou *lista de palavras*, como o próprio nome indica, mostra a frequência de ocorrências de cada forma distinta num *corpus* e a porcentagem que essas ocorrências representam no *corpus*. É comum o emprego do termo inglês *token* para se referir a cada ocorrência de uma forma, e do termo *type* em referência a cada forma distinta (SARDINHA, 2004, p. 94), que pode ser uma palavra ou uma coligação, ou uma colocação, dependendo do contexto em que esse termo aparece. Neste estudo, um *type* corresponde a cada lexema analisado, isto é, formas de um mesmo substantivo, no singular e no plural. A forma “de”, por exemplo, é o *type* mais frequente e aparece 54.652 vezes (*tokens*), representando 6,21% das 879.556 ocorrências (*tokens*) de 40.560 formas distintas (*types*) no *corpus* GER.

Por fim, a expressão *anotação* (ou *etiquetagem*) *morfossintática* corresponde à adição de um código indicativo da classe gramatical a que pertence cada item no texto, além de outras informações mais específicas, como pessoa, número, tempo verbal, etc. *Esse processo é o passo essencial para diversos tipos de análise, incluindo a análise sintática* (ou *parsing*), mas pode ser útil diretamente para a desambiguação entre homônimos (McENERY; WILSON, 2001, p. 36), como empregada no estudo. O trecho sublinhado acima assume a seguinte forma após a anotação morfossintática pelo anotador categorial da página Lex-Center, da Universidade de Lisboa.

<p><s> Esse/DEM#ms processo/PROCESSO/CN#ms é/SER/V#pi-3s o/DA#ms passo/PASSO/CN#ms essencial/ESSENCIAL/ADJ#ms para/PREP diversos/DIVERSO/ADJ#mp tipos/TIPO/CN#mp de/PREP análise/ANÁLISE/CN#fs ,\*//PNT incluindo/INCLUIR/V#ger a/DA#fs análise/ANÁLISE/CN#fs sintática/SINTÁTICO/ADJ#fs .//PNT </s></p>

## 1.2 TIPOS DE VOCABULÁRIO

Neste estudo, utilizaremos a classificação do vocabulário proposta por Chung e Nation (2003), que o divide em: palavras de alta frequência, vocabulário acadêmico, vocabulário técnico e palavras de baixa frequência. Para os autores, uma lista de palavras de alta frequência deve incluir as duas mil palavras mais frequentes de um idioma. No PB, Berber Sardinha (2004) apresenta uma lista das três mil palavras mais frequentes, baseada no *corpus* Banco do Português. Já as palavras de baixa frequência, que correspondem às restantes da língua e não se enquadram nas outras divisões, ocorrem, com maior frequência, em gêneros literários.

Os dois tipos de vocabulário estudados neste trabalho – acadêmico e técnico – são descritos e exemplificados a seguir.

Segundo Chung e Nation (2003, p. 91-92), o *vocabulário acadêmico* é constituído por palavras que não são de uso geral, mas são frequentes em diferentes textos de disciplinas variadas, e geralmente expressam as ações e avaliações do autor. Elas evocam noções gerais que, na língua portuguesa, podem ser codificadas por “método”, “função” e “fator”, termos que podem eventualmente estar contidos na lista de palavras mais frequentes da língua.

O papel do vocabulário acadêmico nos três principais processos acadêmicos – pesquisa, análise e avaliação – é enfatizado por Martin (1976). Para esse autor, a *pesquisa* envolve etapas de formulação, investigação, análise, conclusões e relato de resultados. O vocabulário de *análise* inclui formas verbais simples e compostas, que devem ser ensinadas com suas respectivas regências. Por fim, a *avaliação*, a despeito da objetividade almejada em textos acadêmicos, pressupõe o conhecimento de adjetivos e advérbios.

O lexema “teste” (e testes)<sup>6</sup>, que aparece nos exemplos (1) e (2), é um representante típico do vocabulário acadêmico, uma vez que expressa uma ação envolvida na fase da pesquisa. Nos exemplos (1) e (2) abaixo, seu significado corresponde ao sentido de ‘experimento’ ou ‘método ou conjunto de procedimentos que levam a um diagnóstico’, e se assemelha à acepção 5 (ver nota 2) atribuída à área de medicina, pelo Dicionário Houaiss. Embora não esteja entre os trinta substantivos mais frequentes no *corpus* geral, “teste(s)” ocorre com esse sentido nos *subcorpora* de todas as áreas investigadas, evidentemente, com frequência maior em áreas que incluem experimentação em suas metodologias: ANT (7 ocorrências desse sentido em 8 ocorrências totais neste subcorpus), GEO (4 em 5 ocorrências totais), LING (45 em 45), GEN (96 em 96), NUT (112 em 112) e ZOO (157 em 157). Essas amostras contrastam com (3) e (4), que representam a acepção 1 do verbete mencionado (ver nota 6), de uso geral na língua portuguesa.

(1)

Os genótipos utilizados no *teste* [acadêmico] foram: Thap Maeo, Cai-pira, Pacovan Ken, Ambrosia, PV 42-53, PA 42-44, FHIA 01, FHIA 18 e a cultivar Prata-Anã utilizada como testemunha, que é suscetível à antracnose. (GEN)

(2)

Não se pode saber com certeza qual sem mais *testes* [acadêmico] com foco nas diferenças entre o sistema pro-drop default e o sistema pro-drop de espanhol baseado em traços de concordância. (LING)

(3)

O que há de interessante nesses momentos de transição é que sua liminaridade de vida curta está muitas vezes encravada na liminaridade dilatada, em que a transição final é, como foi, ensaiada em uma série de minitransições: as provas, *testes* [uso geral] e provações das cerimônias de puberdade, por exemplo. (ANT)

(4)

A instantaneidade das transformações possibilita que cada uma delas possa ser livremente imaginada, sem temer o *teste* [uso geral] da prática, como abrigo de segurança e confiança desejada com ardor. (GEO)

Para Chung e Nation (2003), o *vocabulário técnico* é constituído por dois níveis. No primeiro, palavras mais próximas à área observada. Na área de Anatomia, investigada por esses autores, são exemplos desse nível, na língua inglesa, termos como “*ribs*” [costelas], “*neck*” [pescoço] e “*organs*” [órgãos], que podem ocorrer em outros contextos com o mesmo significado, pertencer à lista de palavras mais frequentes de uso geral ou ao vocabulário acadêmico, e, ainda, ocorrer em textos de outras áreas, nas quais possam ser considerados termos técnicos. No segundo nível, encontram-se itens lexicais que dificilmente ocorrem fora da área específica, tais como “*sternum*” [esterno], “*pectoralis*” [peitoral]<sup>7</sup> e “*trachea*” [traqueia].

Para melhor observar esta distinção, seguem abaixo outros trechos de artigos acadêmicos retirados do *corpus* deste trabalho.

(5)

Nas regiões tropicais, a utilização do *capim-elefante* (*Pennisetum purpureum* Schum.) e de seus *clones* (*Pennisetum* spp.) em forma de *capineira*, sob irrigação, constitui prática promissora e bastante difundida, por seu potencial de acúmulo de forragem, que varia de acordo com o *genótipo* e o ambiente no qual está inserido. (GEN)

(6)

As *cultivares* de melancia disponíveis no mercado brasileiro são, na sua grande maioria, de origem americana e suscetíveis a muitas doenças e pragas. (GEN)

(7)

...o CMI apresenta *correlação* positiva menos importante e o CNM apresenta correlação negativa. (GEO)



(8)

Este autor atribuiu a inexistência de *correlação* [técnico da área de estatística] entre as duas medidas de *dissimilaridade* [técnico da área de estatística] à diferença do controle genético nos diferentes tipos de *caracteres* [técnico na área de genética] analisados. (GEN)

Os itens lexicais em itálico no exemplo (5) são avaliados como pertencentes ao primeiro nível, porque são considerados vocabulário técnico em Melhoramento de Plantas, mas também podem ocorrer fora de textos específicos desta área, inclusive como termos técnicos. Já o substantivo “cultivar”<sup>8</sup>, que aparece em sua forma plural em (6) pouco provavelmente ocorrerá fora de disciplinas das áreas agrárias e, inclui-se, portanto, no segundo nível da classificação. Os exemplos (7) e (8) contêm o substantivo “correlação”<sup>9</sup> com o sentido 2 (ver nota 9), o qual, mesmo ocorrendo em textos de áreas variadas do conhecimento, é um termo técnico de estatística, do mesmo modo que o substantivo “dissimilaridade” em (8).

Dessas categorias de vocabulário, o acadêmico é normalmente escolhido para o ensino, pelo fato de, ao contrário do vocabulário técnico, não exigir conhecimento específico de diferentes disciplinas. Além disso, acredita-se que o vocabulário técnico normalmente seja aprendido pelo aluno no contato com a literatura de cada área específica, na qual ele tende a ter alta frequência.

## 2. MATERIAL E MÉTODOS

### 2.1 O *CORPUS*

Seis *subcorpora* foram elaborados a partir de artigos científicos de seis áreas de conhecimento distintas, especificados na TABELA 1. Os artigos foram escolhidos semialeatoriamente em páginas de diferentes periódicos eletrônicos de suas respectivas áreas. Foi usado como critério, dentro de cada área, o número médio de palavras por artigo, de modo a se ter um equilíbrio no *corpus*. De cada artigo foram retirados o título, o resumo, as citações em língua estrangeira, os cabeçalhos, tabelas, figuras, subtítulos, agradecimentos e referências, de modo a diminuir a repetição massiva do conteúdo (como nos resumos), ou da linguagem menos acadêmica (agradecimentos).

Cada *subcorpus* é constituído por aproximadamente 150 mil palavras, e a diferença entre o número de artigos em cada um é devida aos padrões típicos de cada área; por exemplo, os artigos publicados nas áreas de ciências humanas são normalmente mais extensos, como os de Antropologia. Os artigos de ciências biológicas e exatas distinguem-se daqueles de Antropologia porque geralmente apresentam muitas tabelas, que são excluídas na criação do *corpus*, e frequentemente também não têm uma seção dedicada à revisão da literatura, características que os tornam menos extensos que os demais.

Buscamos representar diferentes áreas de conhecimento, sendo elas ciências biológicas (Zoologia, Melhoramento de Plantas), ciências da saúde (Nutrição), ciências humanas (Antropologia, Geografia) e linguística, letras e artes (Aquisição da Linguagem). E, ao fazer isso, nossa preocupação foi equilibrar minimamente o tamanho dos artigos no *corpus*. Temos consciência de que o nível de especialização de alguns artigos, em particular, aqueles sobre Aquisição da Linguagem, criou um padrão de textos que, muito provavelmente, não seja representativo da área de Linguística, em função da presença massiva de pesquisas experimentais. Esse é um aspecto a ser aprimorado no *corpus*.

**TABELA 1** – COMPOSIÇÃO DO *CORPUS*

	ANTROPOLOGIA	AQUISIÇÃO DA LINGUAGEM	GEOGRAFIA	MELHORAMENTO DE PLANTAS	NUTRIÇÃO	ZOOLOGIA
ARTIGOS	21	30	31	60	48	50
TOTAL DE <i>TOKENS</i>	147.569	142.852	149.079	150.789	147.831	141.436

Fonte: Elaborada pelas autoras.

## 2.2 O SOFTWARE DE ANÁLISE

Os textos coletados foram formatados como arquivos “.txt” para serem lidos pelo software livre Kitconc© 3.0 (MOREIRA FILHO, 2008), que tem, entre seus muitos méritos, sua interface em

língua portuguesa. Esse aplicativo fornece listas com as frequências das palavras, concordâncias e colocados, e permite a visualização do contexto de cada *token*, além de calcular sua dispersão no *corpus*. Após a etiquetagem morfossintática (MOREIRA FILHO, 2013) dos textos para a desambiguação de homônimos, foram obtidas listas de frequência para o *corpus* de cada área separadamente e uma com todos os *subcorpora* reunidos, GER. Isso permitiu uma comparação a respeito do tipo de vocabulário frequente nos sete conjuntos.

Como as listas de frequências geradas pelo Kitconc© apresentam uma contagem por cada forma ou palavra ortográfica, foram somados os números de *tokens* de cada *type* que constituem um lexema. Assim, por exemplo, os números de *tokens* de “trabalho” (825) e “trabalhos” (223) foram somados de modo que os 1048 *tokens* totais permitiram a inclusão das duas formas entre os trinta substantivos mais frequentes do *corpus* misto.

A análise levou em consideração o sentido com que cada lexema foi empregado, suas colocações e concordâncias. Os vocabulários acadêmico e técnico foram identificados, segundo a divisão do vocabulário proposta do Chung e Nation (2003). Além dos colocados, também buscamos identificar coligações com a palavra.

A seguir, foram criadas listas de concordâncias de cada forma dos 30 substantivos mais frequentes no *corpus* misto, o que permitiu a identificação de significados e de categorias de vocabulário acadêmico e técnico, e o vocabulário de uso geral. As categorias de uso tiveram como ponto de partida sentidos dispostos no dicionário Houaiss (2009).

### 3. RESULTADOS E DISCUSSÃO

Quando um *corpus* de determinada área é analisado numa lista de frequência, à parte das palavras que não pertencem ao sistema fechado da língua, as que se sobressaem são as que pertencem ao vocabulário técnico (CHUNG; NATION, 2003; SURTASYAH et al., 1994). Esta afirmação é corroborada pelos dados reunidos no ANEXO, que lista os 30 primeiros substantivos dos sete *corpora* desse trabalho, com alta frequência de itens lexicais técnicos de suas respectivas áreas, como “fruto” e “genótipo”, em GEN, e “língua”, em LING.

A lista do *corpus* GER foi comparada com a lista das três mil palavras mais frequentes elaborada por Berber Sardinha (2004). Dentre os 30 substantivos, 22 (“estudo”, “criança”, “valor”, “grupo”, “área”, “peso”, “dado”, “processo”, “forma”, “resultado”, “produção”, “ano”, “população”, “consumo”, “região”, “número”, “trabalho”, “caso”, “vez”, “dia”, “sistema” e “meio”) estão presentes nas primeiras mil palavras da lista mencionada, quatro (“análise”, “animal”, “atividade” e “língua”) estão entre as primeiras mil e duas mil palavras e uma (“tabela”) se encontra entre as primeiras duas mil e três mil palavras.

Dos substantivos mencionados, houve ocorrências técnicas com os substantivos “língua”, “população”, “idade”, “peso”, “valor”, “sistema”, “região”, “produção”, “processo”, “número”, “meio”, “forma”, “caso”, “atividade”, “análise” e “área”. Porém, conforme explicam Chung e Nation (2003), nem todas as ocorrências ao longo dos diferentes *corpora* pertencem ao mesmo tipo de vocabulário. Assim os substantivos listados anteriormente podem pertencer também a outros tipos, como “análise”, que tem duas ocorrências em vocabulário de uso geral na locução “em última análise”, 646 ocorrências técnicas em função de termos colocados, como em “análise combinatória”, “análise morfométrica” e “análise de covariância” e 814 ocorrências acadêmicas com sentido “estudo pormenorizado”.

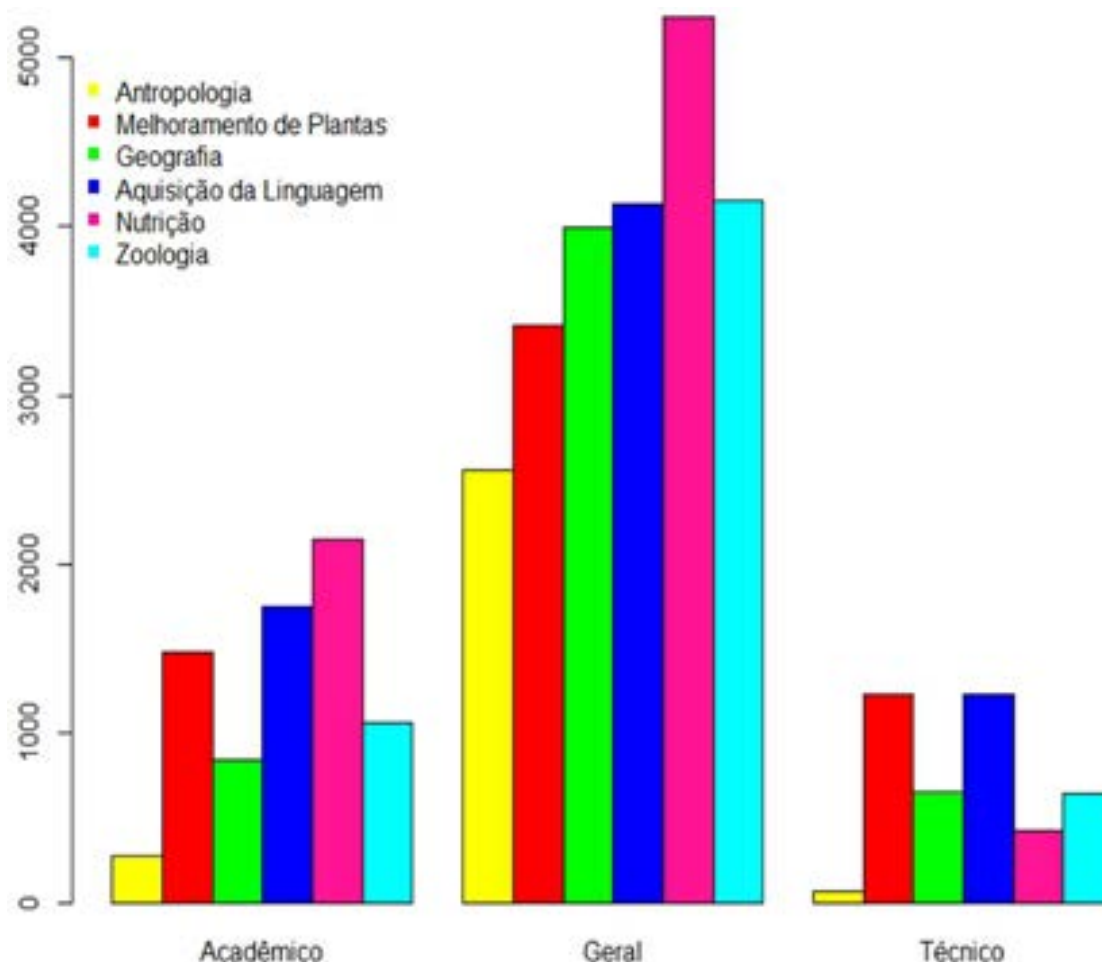
Os três itens da lista deste artigo que não estão na lista de Berber Sardinha (2004) são “fruto”, “característica” e “idade”. “Fruto” ocorreu em vocabulário majoritariamente técnico, com sentido de ‘parte produtiva do vegetal, que sai da flor’. Já “característica” e “idade” pertencem, em grande parte, ao vocabulário geral, com sentido respectivamente de ‘traço, propriedade’ e ‘tempo de vida decorrido’. O item “idade” não está entre as palavras mais frequentes da lista de Berber Sardinha, mesmo fazendo parte do vocabulário de uso geral. Uma hipótese seria que isso ocorre porque o *corpus* montado pelo autor citado é composto por textos jornalísticos, os quais geralmente omitem a palavra “idade” quando se referem à idade de alguém, colocando apenas o número entre vírgulas.

A análise aqui proposta foi a do *corpus* misto, pois esperávamos que o vocabulário acadêmico ganhasse mais destaque na lista desse *corpus*, conforme o estudo desenvolvido por Sutarsyah et al. (1994).

Esse efeito pode ser percebido na lista em ANEXO, pois palavras que ocorrem nos *subcorpora* por área, mesmo não muito frequentes, tendem a ganhar relevância quando os *subcorpora* são unificados. É o caso de ‘forma’ e ‘atividade’. Elas só entraram na lista de GER em função do número total de ocorrências quando reunidos os *subcorpora* e as suas formas singular e plural.

A fim de identificar cada categoria de vocabulário (acadêmico, técnico e geral) em cada um dos trinta itens lexicais mencionados, todas as 35.320 ocorrências foram analisadas introspectivamente de acordo com o sentido em que foram utilizadas. A Figura 2 corresponde ao gráfico que sintetiza as análises.

**FIGURA 2** - GRÁFICO QUE REPRESENTA A FREQUÊNCIA DE CADA VOCABULÁRIO NOS SEIS *SUBCORPORA*.



Fonte: Elaborado pelas autoras.

É interessante observar no gráfico a diferença em relação aos conjuntos de barras referentes à área de Antropologia (1ª à esquerda), em que há volume menor de itens nos vocabulários acadêmico e técnico. Isso ocorre devido à longa extensão da maioria dos artigos nessa área e às metodologias de pesquisa adotadas nesse campo, que fazem com que boa parte dos lexemas recorrentes nestes textos seja de itens lexicais pertencentes ao vocabulário geral, como “caso”, “população” e “região” e a grande incidência dessas palavras se sobressaia em relação ao número de substantivos do vocabulário técnico.

Isso também ocorre em Nutrição, por pertencerem ao vocabulário de uso geral as palavras mais frequentes do *corpus*, como “peso”, “criança” e “anos”. Porém, em contraste com os demais *subcorpora* em NUT, o volume de itens lexicais acadêmicos também é grande, motivado pela maior incidência do lema “estudo”, com 939 ocorrências. Como se verá abaixo, observamos que este substantivo é empregado com maior frequência em referência a pesquisas de outros autores. Também há grande quantidade de palavras pertencentes ao vocabulário geral em Aquisição de Linguagem, isso porque alguns de seus principais lexemas pertencem a este vocabulário, como “criança” e “língua”.

Das três áreas com maior volume de vocabulário acadêmico, Melhoria de Plantas e Aquisição de Linguagem são também as que apresentam a maior quantidade de substantivos no vocabulário técnico. Acreditamos que isso se deva aos modelos de pesquisa adotados, que incluem experimentação e quantificação, além do emprego de termos técnicos de cada área. Esse resultado vai ao encontro da afirmação de Dubois et al. (2009, p. 471-472) de que, em vocabulários especializados, os termos tendem a ser “monossêmicos”, e, portanto, muito específicos, o que aumentaria a quantidade de lexemas diferentes nessas áreas.

Desta análise do *corpus* GER, observou-se que apenas seis (“análise”, “dado”, “estudo”, “resultado”, “tabela” e “trabalho”) dos 30 substantivos possuíam sentido prototipicamente acadêmico. Entretanto, esse número corresponde a um volume significativo de ocorrências (24,2% do total de linhas de concordância analisadas). Na seção abaixo, essas palavras são analisadas, considerando-se suas ocorrências, sentidos, colocados e possíveis coligações.

### 3.1 PALAVRAS PROTÍPICAMENTE ACADÊMICAS DE GER

#### - **Análise(s):**

Este item lexical apresentou três sentidos abrangentes nos *subcorpora* analisados, cuja proporção em cada *corpus* está representada na TABELA 2.

**TABELA 2** – ANÁLISE(S): OCORRÊNCIAS POR *CORPUS* X SENTIDO

ANÁLISE(S)	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) ‘ESTUDO PORMENORIZADO’	65	142	169	163	217	58
(VOC. TÉCN.) LOCUÇÃO TERMO TÉCNICO;	7	276	41	130	109	83
(VOC. GERAL) LOCUÇÃO “EM ÚLTIMA ANÁLISE”	-	2	-	-	-	-

Fonte: Elaborada pelas autoras.

As ocorrências agrupadas como ‘termo técnico’ referem-se àquelas que compõem o vocabulário técnico e, neste caso, só pertencem a ele quando em conjunto com outra(s) palavra(s). A frequência da locução “em última análise” não chega a ser representativa nesse conjunto de 1.462 *tokens*. Os itens agrupados no sentido acadêmico de ‘estudo pormenorizado’ o foram por funcionarem como parte do processo acadêmico de pesquisa (BAKER, 1988; MARTIN, 1976).

#### - **Dado(s):**

Para o vocabulário acadêmico, este é um item lexical importante, que é usado com frequência muito maior no plural (1.445 *tokens*) que no singular (52 *tokens*), majoritariamente com o sentido de ‘material coletado para ou resultado de investigação, cálculo ou pesquisa’.

Sua alta frequência no *corpus* misto justifica-se pelo fato de se referir ao “objeto” investigado em qualquer pesquisa e estar presente em praticamente todas as partes de um artigo científico, na classificação de Martin (1976): formulação, investigação, análise, conclusões e relato de resultados. Suas ocorrências e classificação de vocabulário podem ser observadas na TABELA 3.

Apesar de majoritariamente acadêmico, houve sete ocorrências de locuções técnicas e um uso sem modificador em LING, quando foi retirado o foco do significado, em favor do significante, e a forma, portanto, não se configura como um lexema.

**TABELA 3** – DADO(S): OCORRÊNCIAS POR *CORPUS* X SENTIDO

<b>DADO(S)</b>						
TIPO DE VOCABULÁRIO/SENTIDO	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) RESULTADO DE INVESTIGAÇÃO, CÁLCULO OU PESQUISA	35	241	125	583	375	130
(VOC. TÉCN.) LOCUÇÃO TERMO TÉCNICO;	-	7	-	-	-	-
SEQUÊNCIA DE FONEMAS [DADO]	-	-	-	1	-	-

Fonte: Elaborada pelas autoras.

### - Estudo(s)

Um sinônimo próximo de “trabalho” como obra acadêmica, porém, com sentido mais específico, esse lexema é encontrado 2.404 vezes no *corpus* misto (TABELA 4). Entretanto, curiosamente, foi usado menos vezes em referência ao próprio texto (15,8%) e preferido (74,2%) em relação à obra de outros autores, identificados ou não. Houve dois sentidos acadêmicos: um relacionado à pesquisa científica e outro às áreas acadêmicas de conhecimento.

**TABELA 4** – ESTUDO(S): OCORRÊNCIAS POR *CORPUS* X SENTIDO

<b>ESTUDO(S)</b>						
TIPO DE VOCABULÁRIO/SENTIDO	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) OBSERVAÇÃO, EXAME; ATO DE ESTUDAR	6	69	54	61	9	3
(VOC. ACAD.) INVESTIGAÇÃO CIENTÍFICA SOBRE DETERMINADO ASSUNTO	88	152	257	449	930	326

Fonte: Elaborada pelas autoras.



### - Resultado(s)

Neste item lexical, houve 1.123 ocorrências com sentido acadêmico de ‘resolução de alguma questão’ e 121 ocorrências com sentido de ‘consequência’, em vocabulário geral. O último uso geralmente associa-se com o colocado “como”, em posição anterior ao lexema, com 23 ocorrências. A relação de ocorrências e a classificação dos vocabulários se encontram na TABELA 5.

**TABELA 5 – RESULTADO(S): OCORRÊNCIAS POR CORPUS X SENTIDO**

<b>RESULTADO(S)</b>						
TIPO DE VOCABULÁRIO/SENTIDO	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) RESOLUÇÃO SOBRE ALGUM ASSUNTO	17	208	71	291	280	256
(VOC. GERAL) O QUE RESULTA, É CONSEQUÊNCIA	30	3	45	22	7	14

Fonte: Elaborada pelas autoras.

O colocado mais evidente à direita deste item lexical é “obtido(s)”, na posição N+1, com 92 *tokens*, opondo-se à possibilidade de “encontrado(s)”, que possui apenas 39 *tokens*, na mesma posição. Assim, podemos afirmar que *resultados são obtidos* a partir de certo esforço e, nem tanto, *encontrados*, termo que pode sugerir a ideia de ‘ao acaso’.

### - Tabela(s)

O único sentido deste item lexical foi ‘agrupamento coerente de dados’, com 921 ocorrências em GER. Este é um termo característico da esfera de gêneros acadêmicos, servindo para ilustrar de forma não verbal os dados obtidos ou analisados no estudo. Vale observar, na TABELA 6, que a quantidade de ocorrências em textos da área de ciências humanas (ANT, GEO e LING) é inferior ao das ciências exatas e saúde (GEN, NUT e ZOO), em função de, nestas últimas, as análises serem geralmente quantitativas e, assim, seus resultados serem visualizados com maior clareza em tabelas.

Uma construção desse lexema é “apresentado(s) na(s) tabela(s)” com 48 ocorrências, em lugar de outras como “demonstrado(s) na(s) tabela(s)” ou “observado(s) na(s) tabela(s)”.

**TABELA 6** – TABELA(S): OCORRÊNCIAS POR *CORPUS* X SENTIDO

<b>TABELA(S)</b>						
TIPO DE VOCABULÁRIO/SENTIDO	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) AGRUPAMENTO						
COERENTE DE CÁLCULOS ELABORADOS PREVIAMENTE	9	415	60	64	206	167

Fonte: Elaborada pelas autoras.

### - Trabalho(s)

Esse lexema tem suas ocorrências demonstradas na TABELA 7. No vocabulário acadêmico, com o sentido de ‘obra acadêmica realizada’, ocorre 671 vezes no *corpus* misto, majoritariamente em referência ao próprio texto (58,1%), fazendo uso de demonstrativos como “deste”, “desse” e “presente”. Com menor frequência, portanto, refere-se a textos de outros pesquisadores (122 ocorrências, 18,1% do total). Por essa razão, também aparece em diferentes seções do texto: na descrição/apresentação da pesquisa; na avaliação de pesquisas, como parte do referencial teórico; no relato de resultados.

**TABELA 7** – TRABALHO(S): OCORRÊNCIAS POR *CORPUS* X SENTIDO

<b>TRABALHO(S)</b>						
TIPO DE VOCABULÁRIO/SENTIDO	ANT	GEN	GEO	LING	NUT	ZOO
(VOC. ACAD.) OBRA ACADÊMICA REALIZADA	56	231	91	117	63	113
(VOC. GERAL) EXERCÍCIO DE ATIVIDADE	80	23	54	6	14	6
(VOC. GERAL) ATIVIDADE PROFISSIONAL REGULAR	97	1	45	-	51	-

Fonte: Elaborada pelas autoras.

A colocação “presente *trabalho*” é preferida, com 18,6% dentre suas ocorrências, em comparação com outras quase sinônimas, como “presente *estudo*” (241 *tokens*, correspondendo a 10% das ocorrências totais desse lema) e “presente *pesquisa*” (15 *tokens*, 2,9% do lema destacado).

Assim, em relação às coligações, um reflexo do que foi dito acima é a baixa ocorrência da construção “trabalho(s) de” em que o complemento da preposição é o sobrenome do autor (14 *tokens*), que não chega a configurar uma coligação no melhor uso do termo, por ser pouco frequente.

#### 4. CONSIDERAÇÕES FINAIS

Com esta pesquisa, pudemos observar alguns padrões textuais e de vocabulário em artigos acadêmicos. Nosso estudo corrobora a afirmação de Sutarsyah et al. (1994), de que o vocabulário acadêmico se destaca quando textos de diferentes áreas do conhecimento são reunidos; em *corpora* de textos de áreas específicas, o vocabulário técnico é o mais frequente. Com os seis itens prototipicamente acadêmicos identificados, concluímos que uma próxima pesquisa deverá considerar a inclusão de outras áreas de conhecimento, atentando para as particularidades de cada área e subárea, tais como os fatos de “tabela” ser um item mais frequente nas áreas exatas do que na área de ciências humanas e a aquisição de linguagem não representar outras subáreas da linguística. Daí a necessidade de buscar um equilíbrio na elaboração do *corpus*, pois os diferentes graus de especialização dos textos têm influência direta nos resultados. Num *corpus* misto maior e mais variado, inclusive dentro de cada área de conhecimento, deve ser fortalecida a tendência a um maior volume de vocabulário acadêmico em contrapartida a um decréscimo do volume relativo de vocabulário específico de cada área.

Apesar de a maioria dos substantivos investigados estar entre as três mil palavras mais frequentes do PB, isso não significa que eles não possam fazer parte dos vocabulários acadêmico e técnico. Conforme a literatura consultada, o que os caracteriza como vocabulário acadêmico é sua distribuição e função semelhante nas diferentes áreas. Em relação

ao vocabulário técnico, boa parte das palavras assim classificadas formam termos compostos, ou seja, expressões técnicas, com sentido variado, de acordo com o seu colocado. Observamos também que certos termos técnicos perpassam diferentes áreas, como é o caso de termos de estatística, que podem ser encontrados em qualquer área de pesquisa que trabalhe com dados quantitativos.

Finalmente, consideramos importante expandir essa lista de substantivos para aumentar suas possibilidades de aplicação para fins didáticos.

#### NOUNS IN RESEARCH ARTICLES: TECHNICAL AND ACADEMIC VOCABULARY FROM A CORPUS LINGUISTICS PERSPECTIVE

##### ABSTRACT

We analyzed the 30 most frequent nouns in a mixed *corpus* of research articles from different areas of knowledge, using tools of *Corpus Linguistics*. Our goal was to classify these items as part of the categories: academic vocabulary, technical vocabulary or general use vocabulary. We carried out a non-automatic semantic analysis of 35,320 concordance lines – the number corresponding to the sum of total occurrences in context for each noun – in order to obtain their most frequent meanings. We concluded that the number of academic lexemes was higher in the mixed *corpus* than in the individual *subcorpora* of specific areas, which presented mainly technical terms among their most frequent nouns.

KEYWORDS: academic vocabulary, technical vocabulary, nouns, *corpus*.

---

#### LOS SUSTANTIVOS EN ARTÍCULOS CIENTÍFICOS: EL VOCABULARIO TÉCNICO Y ACADÉMICO EN EL CONTEXTO DE LINGÜÍSTICA DE CORPUS

##### RESUMEN

Fueron analizados los 30 sustantivos más frecuentes en un *corpus* mixto de artículos académicos de diferentes áreas del saber, empleando herramientas de la Lingüística de *Corpus*. El objetivo fue clasificar esos elementos léxicos como elementos integrantes de las categorías de vocabulario académico, vocabulario técnico y vocabulario de uso común. Se realizó un análisis semántico no automático de 35.320 líneas de concordancia – número correspondiente a la suma total de las ocurrencias en contexto para cada uno de estos sustantivos

–, con la finalidad de obtener sus significados más utilizados. Se concluyó que el número de lexemas de uso académico es más alto en el *corpus* compuesto de diferentes áreas que los obtenidos en los análisis de *corpus* separados por áreas del saber, los cuales presentaron términos técnicos entre los sustantivos más frecuentes.

PALABRAS-CLAVE: vocabulario académico, vocabulario técnico, sustantivos, *corpus*.

---

## 5. NOTAS

- 1 Tradução nossa do original: “Basically a lexeme is an association between form and meaning which ignores certain types of variation both on the form side and on the meaning side.” (CRUSE, 2006, p. 92)
- 2 Os sete conjuntos usados na pesquisa recebem os seguintes nomes: os *subcorpora* ANT – antropologia –, GEN – melhoramento genético –, GEO – geografia –, LING – aquisição de linguagem –, NUT – nutrição –, ZOO – zoologia – e o *corpus* GER – união dos seis anteriores.
- 3 Esses e outros sentidos, salvo quando indicado, são acepções encontradas no Dicionário Houaiss da Língua Portuguesa (2003) ou em sua versão eletrônica (2009).
- 4 Tradução nossa do original: “[...] a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and / or spelling” (FRANCIS; KUČERA, 1982, p. 1).
- 5 Tradução nossa do original: “collocations of a given word are statements of the habitual or customary places of that word”.
- 6 “<sup>2</sup>teste *Datação*: sXX (s. m.): 1 exame crítico ou prova das qualidades de uma pessoa ou coisa; [...] 5 Rubrica: medicina. exame ou ensaio destinado a estabelecer um diagnóstico. Ex.: *t. de alergia*.” (DICIONÁRIO Eletrônico Houaiss da Língua Portuguesa 3.0)
- 7 Na língua portuguesa, é bastante provável que “peitoral” pertença ao primeiro nível de vocabulário técnico.
- 8 “Uma cultivar é resultado de melhoramento em uma variedade de planta que a torne diferente das demais em sua coloração, porte, resistência a doenças.

A nova característica deve ser igual em todas as plantas da mesma cultivar, mantida ao longo das gerações.” (CULTIVARES protegidas)

- 9 “**correlação** *Datação*: 1675 (s. f.): **1** correspondência, similitude, analogia entre pessoas, coisas, ideias relacionadas entre si; **2** Rubrica: estatística. interdependência de duas ou mais variáveis; **3** Rubrica: geometria. transformação que associa, no plano, pontos a linhas e linhas a pontos e, no espaço, pontos a planos e planos a pontos” (DICIONÁRIO Eletrônico Houaiss da Língua Portuguesa 3.0)

## 6. REFERÊNCIAS

BAKER, M. Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, v. 4, n. 2. 1988.

BIDERMAN, M. T. C. O vocabulário fundamental no ensino do Português como segunda língua. In: SILVEIRA, R. C. P. da (Org.). *Português língua estrangeira: perspectivas*. São Paulo: Cortez Editora, 1998. p. 73-91.

CHUNG, T. M.; NATION, P. Technical vocabulary in specialized texts. *Reading in a Foreign Language*, v. 15, n. 2. 2003.

CORRELAÇÃO. In: *Dicionário Eletrônico Houaiss da Língua Portuguesa 3.0*. Rio de Janeiro: Objetiva, 2009.

COXHEAD, A. A new academic word list. *TESOL Quarterly*, v. 2, n. 34, p. 213-238, 2000.

CRUSE, A. *A glossary of semantic and pragmatics*. Edinburgh: Edinburgh University Press, 2006.

CULTIVARES protegidas. In: BRASIL. *Ministério da Agricultura*. Disponível em: <<http://www.agricultura.gov.br/vegetal/registros-autorizacoes/protecao-cultivares/cultivares-protegidas>>. Acesso em: 10 maio 2015.

DELL'ISOLA, R. L. P. *O sentido das palavras na interação leitor-texto*. Belo Horizonte: Faculdade de Letras da UFMG, 2005.

DUBOIS, J. et al. *Dicionário de linguística*. São Paulo: Cultrix, 2009.

FIRTH, J. A synopsis of linguistic theory 1930-1955. In: PALMER, F. R. *Selected papers of J. R. Firth 1952 – 1958*. Bloomington: Indiana University Press, 1968.

- FRANCIS, N.; KUČERA, H. *Frequency Analysis of English Usage: lexicon and grammar*. Boston: Houghton Mifflin, 1982.
- HOEY, M. *Pattern of Lexis in Text*. Oxford: Oxford University Press, 1991.
- HOUAISS, A. et al. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva, 2009.
- MARTIN, A. V. Teaching Academic Vocabulary to Foreign Graduate Students. *TESOL Quarterly*, v. 10, n. 1. p. 91-98. 1976.
- McENERY, T.; WILSON, A. *Corpus linguistics*. 2. ed. Edinburgh: Edinburgh University Press, 2001.
- MOREIRA FILHO, J. L. *Kitconc 3.0*. 2008. Disponível em: <<http://www.fflch.usp.br/dl/li/x/?p=394>> Acesso em: 8 ago. 2013.
- MOREIRA FILHO, J. *Etiquetador Morfossintático*. 2013. Disponível em: <<http://www.fflch.usp.br/dl/li/x/?p=727>>. Acesso em: 24 jun. 2014.
- NATION, I. S. P. *Learning Vocabulary in Another Language*. Cambridge Applied Linguistics. Cambridge: Cambridge University Press, 2001.
- NLX-natural Language and Speech Group. *LX Suite*. Universidade de Lisboa, Departamento de Informática. Disponível em: <<http://lxcenter.di.fc.ul.pt/services/pt/LXServicesSuitePT.html>>. Acesso em: 12 maio 2015.
- SARDINHA, T. B. *Linguística de Corpus*. Barueri: Manole, 2004.
- SAUTCHUCK, I. *Prática de morfossintaxe: como e por que aprender análise (morfo)sintática*. 2. ed. Barueri: Manole, 2010.
- SUTARSYAH, C. et al. How useful is EAP vocabulary for ESP? A *corpus* based study. *RELC Journal*, v. 25, n. 2, p. 34-50. 1994.
- TAGNIN, S. E. O. *O jeito que a gente diz: expressões convencionais e idiomáticas*. São Paulo: Disal, 2005.
- TESTE. In: *Dicionário Eletrônico Houaiss da Língua Portuguesa 3.0*. Rio de Janeiro: Objetiva, 2009.
- XIAO, R.; MCENERY, T. Collocation, semantic prosody, and near synonymy: a cross-linguistic perspective. *Applied Linguistics*, v. 27, n. 1. 2006.

---

Submetido em 15 de maio de 2015.

Aceito em 27 de outubro de 2015.

Publicado em 23 de novembro de 2016.

---

## APÊNDICE

**TABELA 8** – LISTA DE LEMAS DOS 30 PRIMEIROS SUBSTANTIVOS DOS SETE *CORPORA*

GERAL	ANTROPOLOGIA	AQUISIÇÃO LING.	GEOGRAFIA	MEL. PLANTAS	NUTRIÇÃO	ZOOLOGIA
ESTUDO	RELAÇÃO	CRIANÇA	ÁREA	FRUTO	ESTUDO	ANIMAL
CRIANÇA	FORMA	LÍNGUA	PROCESSO	GENÓTIPO	CONSUMO	DIA
GRUPO	MODO	AQUISIÇÃO	CIDADE	PLANTA	ALIMENTAR	PESO
DADO	GRUPO	DADO	POPULAÇÃO	MÉDIA	PESO	VALOR
ÁREA	TRABALHO	PALAVRA	RELAÇÃO	ACESSO	ALIMENTO	ESTUDO
ANÁLISE	FAMÍLIA	ESTUDO	REGIÃO	VALOR	CRIANÇA	DIETA
VALOR	REPRESENTAÇÃO	SUJEITO	FORMA	NÚMERO	SAÚDE	PERÍODO
FORMA	PESSOA	PRODUÇÃO	ESPAÇO	SELEÇÃO	GRUPO	RESULTADO
PESO	POLÍTICA	PROCESSO	ESTUDO	TABELA	ANOS	CASO
RESULTADO	CASO	FORMA	USO	ANÁLISE	IDADE	AMOSTRA
PROCESSO	SOCIEDADE	SISTEMA	MUNICÍPIO	GRUPO	DADO	TRATAMENTO
PRODUÇÃO	VEZ	RELAÇÃO	ATIVIDADE	MÉTODO	PREVALÊNCIA	PRODUÇÃO
ANO	PROCESSO	VOGAL	SOLO	CARACTERÍSTICA	ADOLESCENTES	IDADE
CONSUMO	PRODUÇÃO	RESULTADO	VALOR	CARACTERE	ANÁLISE	AUMENTO
POPULAÇÃO	PRÁTICA	CASO	LOCAL	POPULAÇÃO	GORDURA	RELAÇÃO
REGIÃO	SENTIDO	ANÁLISE	ANÁLISE	CORRELAÇÃO	ALIMENTAÇÃO	TEMPERATURA
TRABALHO	ANO	FALA	AMBIENTE	VARIAÇÃO	MULHER	REGIÃO
CASO	PARTICIPAÇÃO	REGRA	MEIO	EFEITO	RESULTADO	FORMA
NÚMERO	REGIÃO	LINGUAGEM	SISTEMA	ESTIMATIVA	VALOR	RAÇÃO
VEZ	COMUNIDADE	POSIÇÃO	CAMPO	RESULTADO	MÊS	ALIMENTOS
ANIMAL	HOMEM	EXEMPLO	TRABALHO	LOCAL	INGESTÃO	AVE
TABELA	TEMPO	CONTEXTO	RISCO	DADO	FATOR	LEITE
DIA	GERAÇÃO	TIPO	PARTE	PRODUÇÃO	POPULAÇÃO	ENZIMA
SISTEMA	ESTADO	ERRO	OCUPAÇÃO	TRABALHO	INDIVÍDUO	CONSUMO
LÍNGUA	ATOR	ESTRUTURA	ESTADO	MARCADOR	VIDA	CARCAÇA
MEIO	MEIO	ESCRITA	CASO	INTERAÇÃO	SEXO	VACA
CARACTERÍSTICA	PODER	PB	RIO	AVALIAÇÃO	RISCO	TABELA
ATIVIDADE	FATO	PORTUGUÊS	ANO	RELAÇÃO	INTERVENÇÃO	GANHO
IDADE	LUGAR	SEGMENTO	DESENVOLVIMENTO	MELHORAMENTO	RELAÇÃO	MÉDIA
FRUTO	VIDA	USO	FLUXO	ÍNDICE	EXCESSO	PROTEÍNA