

RESUMO

A teoria do aparelhamento ou primação lexical formulada pelo professor e pesquisador britânico Michael Hoey (*Lexical priming: A new theory of words and language*; Abingdon, UK: Routledge, 2005) propõe que a configuração das línguas resulta de processos mentais individuais alimentados por encontros entre indivíduos e palavras. Conforme a teoria, experiências prévias preparam, aparelham ou primam os usuários das línguas a empregar certas palavras e outros recursos linguísticos em determinados contextos, domínios e gêneros textuais. Este trabalho teve como objetivo explorar a relação entre aparelhamento psicológico (*psychological priming*) – elemento central da teoria – e variação disciplinar, com foco na escrita acadêmica em português. Para tanto, foram compilados dois *corpora* especializados. O primeiro é composto por 24 artigos publicados na *Revista Paulista de Pediatria*; o segundo, por 24 artigos da *Revista de Administração de Empresas*. Com o auxílio de ferramentas de programas de análise linguística, foram estudados aspectos colocacionais, semânticos e distributivos de duas palavras-chave comuns a ambos os *corpora*: *estudo* e *pesquisa*. Como resultado, observaram-se diferenças entre os dados de pediatria e administração. Todavia, ao passo que as diferenças relativas a *pesquisa* são bem marcadas, as diferenças relativas a *estudo* são um pouco mais sutis. Os resultados mostram a relevância do fator campo de conhecimento ou disciplina para o aparelhamento psicológico de autores de artigos acadêmicos em língua portuguesa. Ainda, indicam a existência de palavras mais e menos propensas a assumir comportamentos distintos em disciplinas diversas.

Palavras-chave: Teoria da primação lexical. Linguística de *corpus*. Escrita acadêmica.

ABSTRACT

Michael Hoey's lexical priming theory (*Lexical priming: A new theory of words and language*; Abingdon, UK: Routledge, 2005) proposes that the shape of languages results from individual mental processes fed by encounters between individuals and words. According to the theory, previous experiences prepare, load, or prime language users to employ certain words and other linguistic resources in particular contexts, domains, and genres. The present study aimed to explore the relationship between psychological priming—central for the theory—and disciplinary variation, with a focus on academic writing in Portuguese. To this end, two specialized corpora were compiled. The first is composed of 24 articles published in *Revista Paulista de Pediatria* (pediatrics); the second consists of 24 articles from *Revista de Administração de Empresas* (management). With the aid of corpus linguistics software tools, collocational, semantic, and distributive aspects of *estudo* and *pesquisa*—two keywords shared by the corpora—were studied. As a result, differences between pediatrics and management data were observed. However, while distinguishing features of *pesquisa* are very clear, differences concerning *estudo* are a little more subtle. The results show the relevance of field of knowledge or discipline to the psychological priming of authors of academic articles in Portuguese. Also, they indicate the existence of words with differing degrees of discipline-specific primings.

Keywords: Lexical priming theory. Corpus linguistics. Academic writing.

¹ Professor de Comunicação no Centro Paula Souza (CPS). Mestre em Letras e doutorando em Estudos Linguísticos e Literários em Inglês pela Universidade de São Paulo (USP). São Paulo/SP, Brasil. ORCID: <https://orcid.org/0000-0003-0284-3223>. Endereço de e-mail: rodrigo.aragao@fatec.sp.gov.br.



1 INTRODUÇÃO

A teoria do aparelhamento ou primação lexical (*lexical priming*) formulada pelo pesquisador e professor britânico Michael Hoey propõe que as línguas têm como força motriz o aparelhamento psicológico (*psychological priming*), processo individual resultante de experiências comunicativas que predispõe, prepara ou prima (*prime*) os usuários das línguas a optarem por um ou outro recurso linguístico em determinado contexto, domínio, estilo e gênero textual (HOEY, 2004, 2005, 2013). A configuração de uma língua seria, então, o produto de incontáveis aparelhamentos individuais que se moldam, confirmam, modificam continuamente (HOEY, 2005). A mente humana é vista como um programa de computador de análise linguística que associa palavras e combinações de palavras preferencialmente a certas outras palavras (*colocações*), a certos conjuntos semânticos (*associações semânticas*), a certas funções discursivas (*associações pragmáticas*), a certos padrões gramaticais (*coligações*), a certas classes gramaticais (*categorias gramaticais*), a certas relações coesivas (*colocações textuais*), a certas relações semânticas (*associações semânticas textuais*) e a posições específicas no texto (*coligações textuais*); tais associações preferenciais ou associações-primas (*primings*) manifestam-se coletivamente em conjuntos de dados linguísticos ou *corpora*, sendo estes o alvo dos estudos de primação lexical (HOEY, 2005, 2013).

A teoria do aparelhamento lexical tem sido usada para o estudo da conformidade e não conformidade a padrões coletivos de usos linguísticos. Uma linha consolidada de pesquisa refere-se a usos criativos da língua. Hoey (2007), por exemplo, mostra como a adequação e a fuga às normas manifestam-se em textos literários, podendo explicar a criatividade literária. Goatly (2017) detalha como enunciados humorísticos fogem de padrões associativos, concluindo que a ideia de imposição autoral de aparelhamento (*overriding of priming*) é capaz de elucidar mecanismos do humor. Outra linha estabelecida de pesquisa refere-se à aprendizagem de línguas estrangeiras, sobretudo o inglês. Contrastam-se dados de aprendizes com dados autênticos da língua de interesse a fim de verificar em que medida os usuários não nativos se aproximam das formas-padrão da língua que estudam (cf. HADIKIN, 2013; OOI, 2013; JANTUNEN, 2017).

Embora a relação entre aparelhamento psicológico e domínio esteja pressuposta nos estudos de primação lexical, a variação de campos de conhecimento ou disciplinas acadêmicas tem escapado do escopo desses estudos. Há trabalhos sobre textos religiosos (SALIM, 2012), textos publicitários (CUNHA, 2017), textos de política (DUGUID; PARTINGTON, 2017), textos de divulgação



científica (HOEY, 2017); todavia, não se localizam estudos dedicados a possíveis diferenças entre textos de disciplinas acadêmicas diversas quanto ao aparelhamento psicológico.

Este trabalho dirige-se a essa lacuna. Por meio da comparação de indícios de aparelhamento psicológico encontrados em artigos acadêmicos em língua portuguesa de pediatria e administração, pretende-se avaliar em que medida o campo de conhecimento ou disciplina é relevante para os processos associativos mentais inerentes à escrita acadêmica. São enfocados três tipos de associações-primas: colocacionais, por meio do estudo de combinações ou conglomerados de palavras; semânticos, por meio do estudo de conjuntos semânticos; e posicionais, por meio do estudo de distribuições de palavras. Além do aspecto teórico, justificam este estudo possíveis aplicações dos resultados para o ensino e a aprendizagem de escrita acadêmica em português, já que semelhanças em associações-primas de dados linguísticos de disciplinas distintas autorizam um ensino generalista e diferenças fundamentam um ensino especializado.

2 *CORPORA* E PROCEDIMENTOS DE ANÁLISE

Dois coleções de artigos acadêmicos formam a base dos *corpora* deste estudo. A primeira é composta por 24 textos publicados entre 2016 e 2018 na *Revista Paulista de Pediatria*; a segunda é constituída de 24 textos da *Revista de Administração de Empresas* datados, em quase sua totalidade, de 2018. A *Revista Paulista de Pediatria* e a *Revista de Administração de Empresas* foram escolhidas entre os periódicos da Scientific Electronic Library Online do Brasil (<https://www.scielo.br>), biblioteca eletrônica nacional de excelência, com base nas disciplinas ou campos que representam, respectivamente pediatria e administração. Pediatria e administração, por sua vez, foram escolhidas por integrarem duas grandes áreas de conhecimento, respectivamente ciências da saúde e ciências sociais aplicadas. Partiu-se do pressuposto de que textos provenientes de grandes áreas distintas são mais capazes de revelar diferenças disciplinares do que aqueles provenientes de uma mesma área. O tamanho dos *corpora* foi definido com base em pesquisas passadas acerca do gênero artigo acadêmico. Swales (1981), por exemplo, examinou 48 textos de três grandes áreas: 16 de ciências duras (*hard sciences*), 16 de biologia/medicina e 16 de ciências sociais. Anthony (1999) estudou 12 textos de uma única disciplina: engenharia de *software*. Samraj (2002) examinou 24 textos de dois ramos da biologia: 12 de comportamento da vida selvagem (*wildlife behavior*) e 12 de biologia da conservação (*conservation biology*). Embora a abordagem deste estudo difira daquela empregada nas pesquisas citadas (análise de gênero), entende-se que pequenas coleções de 12, 16 ou 24 artigos são suficientes para evidenciar





características linguísticas próprias de um campo ou disciplina. A diferença entre os períodos de publicação dos artigos de pediatria e administração deve-se à preferência por textos que, em seus arquivos eletrônicos do tipo .pdf, contêm indicação de licenças de uso da Creative Commons (cf. CREATIVE COMMONS BRASIL, ca. 2022). Deu-se preferência a tais textos, pois a reprodução de fragmentos extensos dos originais poderia ser necessária para a apresentação de resultados. Ao passo que a *Revista Paulista de Pediatria* inclui menção a licenças da Creative Commons nos arquivos de 2016, 2017 e 2018, a *Revista de Administração de Empresas* faz isso apenas a partir de 2018. Os artigos de pediatria foram selecionados ao acaso entre aqueles publicados entre 2016 e 2018. Já no caso dos textos de administração, todos os artigos publicados em 2018 em português foram incluídos, exceto aqueles de uma edição especial dedicada a um evento (cf. TONELLI; ZAMBALDI, 2018). Um texto de 2017 foi ainda selecionado ao acaso para totalizar 24 artigos de administração. Foram priorizados artigos recentes com o intuito de lançar luz sobre aspectos atuais da escrita acadêmica.

Os textos dos artigos foram convertidos para o formato .txt com codificação UTF-8, que é usada em análises linguísticas auxiliadas por programas de computador. Na conversão, foram excluídos dados dos autores e das revistas, palavras-chave, *abstract*, *keywords*, figuras, tabelas, referências e agradecimentos; foram mantidos títulos de seções, resumos e notas de rodapé ou fim de texto.

A etapa inicial da análise foi realizada na plataforma eletrônica Sketch Engine (<https://sketchengine.eu/>) com o uso do *software* de mesmo nome, ambos desenvolvidos pela empresa Lexical Computing CZ (ca. 2022). O *software* Sketch Engine analisa dados em diferentes línguas, inclusive o português, com funções convencionais de programas linguísticos, como geração de listas de palavras frequentes (*wordlist*), identificação de palavras típicas de um *corpus* em relação a outro (*keywords*) e geração de linhas de concordância (*concordance*). Os arquivos no formato .txt foram compilados na plataforma para a constituição de dois *corpora*: AA-Ped (artigos acadêmicos de pediatria) e AA-Adm (artigos acadêmicos de administração). A Tabela 1 apresenta dados básicos dos *corpora*.

Tabela 1: Visão geral dos *corpora*

<i>Corpus</i>	Artigos	<i>Tokens</i>	Palavras	Frases
AA-Ped	24	82,519	69,257	2,493
AA-Adm	24	175,596	142,906	5,554
Total	48	258,115	212,163	8,047

Fonte: Elaborada pelo autor com o auxílio do Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022)





Inicialmente, procedeu-se à análise de palavras-chave (*keywords*) não com o intuito de identificar palavras de uso exclusivo de um ou outro *corpus*, mas sim com o propósito de identificar palavras de uso comum a ambos. O Sketch Engine usa uma técnica própria para o cálculo da tipicidade (*keyness*) de palavras denominada *matemática simples* (KILGARRIFF, 2009; LEXICAL COMPUTING LTD., 2015). Mais especificamente, a seguinte fórmula é empregada com esse objetivo:

$$\frac{fpm_{estudo} + N}{fpm_{referência} + N}$$

Na fórmula, fpm_{estudo} refere-se à frequência da palavra por milhão de *tokens* (palavras e outros elementos cercados por espaços) no *corpus* de estudo; $fpm_{referência}$ corresponde à frequência da palavra por milhão de *tokens* no *corpus* de referência; N é um parâmetro definido pelo analista que permite direcionar a identificação de palavras-chave para itens raros (N próximo ou igual a 0,001) ou itens comuns (N próximo ou igual a 1.000.000). A análise foi realizada com os critérios dispostos na Tabela 2.

Tabela 2: Critérios usados para identificar palavras-chave de uso comum com o Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022)

Critério	Descrição	Opção
Atributo de busca	É possível buscar palavras-chave com os atributos palavra, lexema (forma dicionarizada da palavra), etiqueta (<i>tag</i>) e parte do discurso, entre outras opções	Palavra
Foco: 0,001–1.000.000	O foco recai sobre palavras raras com números próximos a 0,001 e sobre palavras comuns com números próximos a 1.000.000	1.000
Frequência mínima	Estabelece a frequência mínima que a palavra-chave deve ter no <i>corpus</i> de estudo	10
Frequência máxima	Estabelece a frequência máxima que a palavra-chave pode ter no <i>corpus</i> de estudo	Zero (sem efeito)
Número máximo de itens	Estabelece o número máximo de palavras-chave	100
A = a	Estabelece se haverá distinção entre letras maiúsculas e minúsculas (desativado) ou não (ativado)	Ativado
Ao menos um caractere alfanumérico	Restringe a análise a palavras compostas por ao menos uma letra ou número	Ativado
Apenas caracteres alfanuméricos	Restringe a análise a palavras compostas por letras, números e hífens	Ativado
Inclusão de não palavras	Inclui na análise números, sinais de pontuação e outros <i>tokens</i> que não começam com letras	Desativado

Fonte: Elaborada pelo autor





Primeiro, adotou-se como *corpus* de estudo AA-Ped e como *corpus* de referência AA-Adm. Em seguida, os papéis inverteram-se: AA-Adm foi adotado como *corpus* de estudo e AA-Ped como *corpus* de referência. Como resultado, foram geradas duas relações de 100 palavras cada. A fim de privilegiar itens representativos dos conjuntos de artigos em detrimento daqueles limitados a poucos textos, foram excluídas das relações todas as palavras-chave com presença em menos de 50% dos textos que compõem cada um dos *corpora*. Isto é, foram mantidas apenas as palavras que aparecem em pelo menos 12 artigos de pediatria e 12 artigos de administração. Ainda, foram excluídas das relações as palavras cuja incidência em um dos *corpora* fosse inferior a 50, uma vez que a escassez numérica dificulta a observação de padrões linguísticos. A Tabela 3 apresenta a relação das palavras-chave comuns dos artigos acadêmicos de pediatria (foco em AA-Ped); a Tabela 4, a relação das palavras-chave comuns dos textos de administração (foco em AA-Adm).

Tabela 3: Palavras-chave comuns de artigos acadêmicos de pediatria (*corpus* de estudo: AA-Ped; *corpus* de referência: AA-Adm)

Posição	Palavra	Frequência Geral		Frequência por Milhão de <i>Tokens</i>		Prop. de Artigos com a Palavra (%)		Score
		AA-Ped	AA-Adm	AA-Ped	AA-Adm	AA-Ped	AA-Adm	
1	<i>anos</i>	179	85	2.169,2	484,1	91,7	95,8	2,14
2	<i>estudo</i>	364	292	4.411,1	1.662,9	100,0	100,0	2,03
3	<i>foram</i>	424	373	5.138,2	2.124,2	100,0	100,0	1,97
4	<i>foi</i>	578	540	7.004,4	3.075,2	100,0	100,0	1,96
5	<i>uso</i>	130	73	1.575,4	415,7	70,8	75,0	1,82
6	<i>média</i>	89	57	1.078,5	324,6	75,0	66,7	1,57
7	<i>presente</i>	84	59	1.017,9	336,0	87,5	70,8	1,51
8	<i>número</i>	84	62	1.017,9	353,1	91,7	70,8	1,49
9	<i>ou</i>	363	465	4.399,0	2.648,1	100,0	100,0	1,48
10	<i>qualidade</i>	94	88	1.139,1	501,2	62,5	70,8	1,43
11	<i>com</i>	914	1.364	11.076,2	7.767,8	100,0	100,0	1,38
12	<i>não</i>	477	692	5.780,5	3.940,9	100,0	100,0	1,37
13	<i>período</i>	73	66	884,6	375,9	79,2	62,5	1,37

Fonte: Compilada pelo autor com o apoio do Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022)

**Tabela 4:** Palavras-chave comuns de artigos acadêmicos de administração (*corpus* de estudo: AA-Adm; *corpus* de referência: AA-Ped)

Posição	Palavra	Frequência Geral		Frequência por Milhão de <i>Tokens</i>		Prop. de Artigos com a Palavra (%)		Score
		AA-Adm	AA-Ped	AA-Adm	AA-Ped	AA-Adm	AA-Ped	
1	<i>se</i>	631	153	3.593,5	1.854,1	100,0	95,8	1,61
2	<i>pesquisa</i>	347	82	1.976,1	993,7	95,8	87,5	1,49
3	<i>assim</i>	237	51	1.349,7	618,0	95,8	83,3	1,45
4	<i>relação</i>	459	130	2.614,0	1.575,4	100,0	95,8	1,40
5	<i>esse</i>	272	73	1.549,0	884,6	95,8	91,7	1,35

Fonte: Compilada pelo autor com o apoio do Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022)

Das palavras listadas nas Tabelas 3 e 4, escolheram-se duas que, em artigos acadêmicos em português, assumem sentido semelhante: *estudo* e *pesquisa*. Procedeu-se, então, ao exame de aspectos colocacionais, semânticos e distributivos das palavras em ambos os *corpora*. A análise de aspectos colocacionais foi feita com a função de identificação de combinações ou conglomerados de palavras (*n-grams*) do Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022). Buscaram-se combinações contendo as palavras *estudo* e *pesquisa* em ambos os *corpora* com as seguintes especificações: palavra como atributo de busca; frequência mínima no *corpus* equivalente a duas ocorrências; frequência máxima desativada; equivalência entre maiúsculas e minúsculas ativada; extensão das combinações entre duas e seis palavras. Em seguida, verificaram-se os números de combinações exclusivas e compartilhadas dos *corpora* e compararam-se as frequências normalizadas por milhão de *tokens* das combinações compartilhadas. Números altos de combinações exclusivas e diferenças expressivas nas frequências por milhão de *tokens* das combinações compartilhadas foram assumidos como evidência da relevância do fator campo ou disciplina para o aparelhamento psicológico.

A análise de aspectos semânticos deu-se por intermédio de conjuntos semânticos. A identificação dos conjuntos semânticos foi realizada manualmente a partir de listas de colocados geradas pelo Sketch Engine (LEXICAL COMPUTING CZ, ca. 2022). Buscaram-se colocados para *estudo* e *pesquisa* em ambos os *corpora* com o uso da ferramenta de linhas de concordância (*concordance*). Para a geração de linhas, foram adotadas as seguintes especificações: palavra como atributo de busca; substantivo como categoria gramatical visada; equivalência entre maiúsculas e minúsculas ativada. Em seguida, para a extração dos colocados das linhas de concordância geradas, as seguintes especificações foram usadas: palavra sem distinção de maiúsculas como atributo de busca; alcance de cinco palavras antes e cinco depois do termo de busca ou nóculo; frequência mínima no *corpus* equivalente a duas ocorrências; frequência mínima no alcance adotado igual a duas ocorrências; grau de associação medido por logDice, em que valores negativos indicam ausência de significância





estatística para a coocorrência de palavras e valores positivos alcançam um máximo teórico igual a 14 (RYCHLÝ, 2008). Os conjuntos semânticos foram identificados mediante a análise de linhas de concordância dos 100 primeiros colocados de cada uma das listas resultantes, excluídos sinais de pontuação, números e símbolos matemáticos. O sentido e o conjunto semântico dos colocados foram determinados pela observação de seu entorno nas linhas de concordância. No caso da palavra *pesquisa* no *corpus* AA-Ped, a lista continha apenas 75 colocados; o estudo semântico deteve-se sobre as linhas de todos eles. Diferenças nos conjuntos semânticos de uma mesma palavra entre os *corpora* foram assumidas como evidência da importância do fator campo de conhecimento ou disciplina para o aparelhamento psicológico.

A análise de aspectos distributivos foi adotada como abordagem indireta para o estudo de coligações textuais, isto é, associações entre palavras e posições no texto (HOEY, 2005). Nesta etapa, usou-se a função de geração de gráficos *plot* do *software* linguístico AntConc versão 4.0.3 (ANTHONY, 2022). Primeiro, os arquivos-base .txt dos *corpora* foram abertos com o *software*. Em seguida, empregou-se a função para cada uma das palavras nos dois conjuntos de dados. Isso foi feito com a seleção da medida de dispersão D de Juilland, que varia de zero (ocorrências concentradas em um único ponto do texto) a 1 (ocorrências distribuídas ao longo de todo o texto). Os valores de D foram registrados separadamente e, depois, convertidos em gráficos de linhas que mostram os padrões de distribuição das palavras nos artigos que compõem os *corpora*. Diferenças entre os padrões de distribuição foram assumidas como evidência de relação relevante entre aparelhamento psicológico e campo de saber ou disciplina.

3 INDÍCIOS DE APARELHAMENTO PSICOLÓGICO

No *corpus* AA-Ped, foram encontradas 251 combinações de palavras contendo *estudo* com duas ou mais ocorrências. Desse total, 43 (17,1%) são comuns ao outro *corpus*. Já em AA-Adm, observaram-se 181 combinações com *estudo* de duas ou mais ocorrências. As 43 combinações compartilhadas correspondem, assim, a 23,8% do total identificado nos artigos de administração. Sob outro ângulo, 82,9% das combinações com *estudo* de AA-Ped e 76,2% das combinações presentes em AA-Adm são exclusivas. Exemplos de combinações exclusivas do *corpus* de pediatria são *estudo transversal* (13 ocorrências), *tempo de estudo* (13 ocorrências), *objetivo deste estudo* (6 ocorrências), *estudo multicêntrico* (4 ocorrências), *estudo de revisão* (4 ocorrências), *estudo foi aprovado pelo comitê* (4 ocorrências), *estudo retrospectivo* (2 ocorrências) e *estudo descritivo de corte transversal* (2 ocorrências). Exemplos de

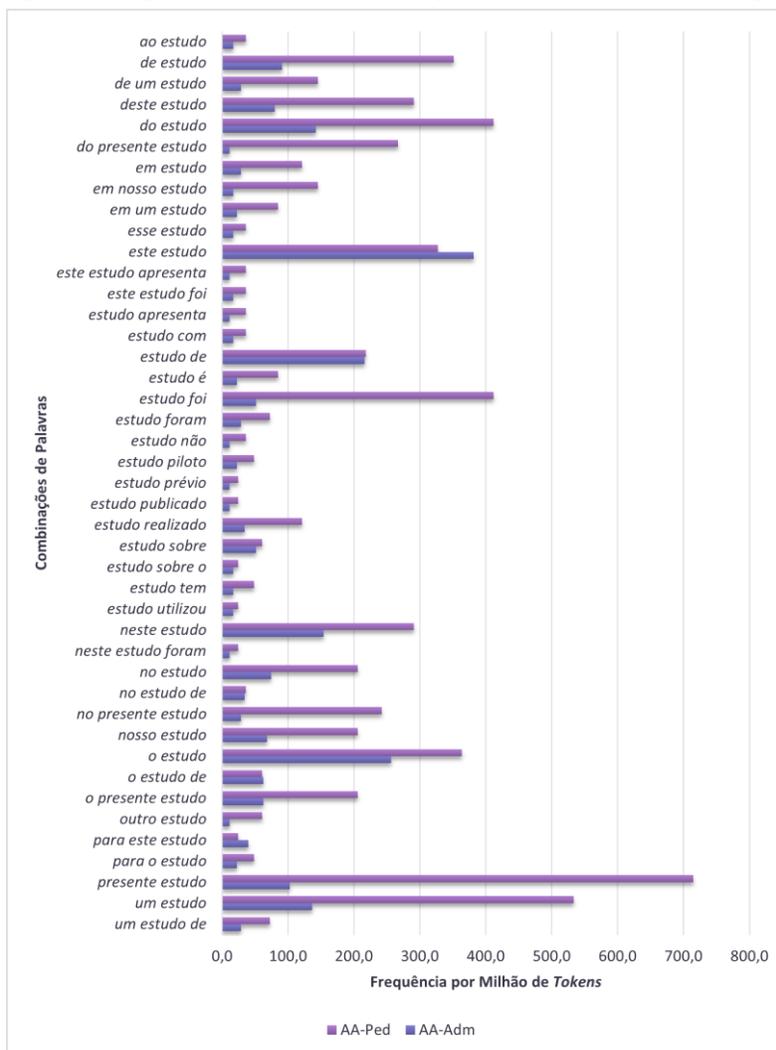




combinações exclusivas dos artigos de administração são *estudo contribui* (7 ocorrências), *campo de estudo* (7 ocorrências), *estudo propõe* (6 ocorrências), *estudo revela* (4 ocorrências), *o estudo realizado* (4 ocorrências), *estudo sugere* (3 ocorrências), *estudo historiográfico* (2 ocorrências) e *os resultados deste estudo* (2 ocorrências).

As combinações comuns a ambos os *corpora* apresentam, quase em sua totalidade, frequências normalizadas por milhão de *tokens* bastante distintas, como mostra a Figura 1. Exceções são *estudo de* (AA-Ped: 218,1; AA-Adm: 216,4), *estudo sobre* (AA-Ped: 60,6; AA-Adm: 51,3), *estudo sobre o* (AA-Ped: 24,2; AA-Adm: 17,1), *estudo utilizou* (AA-Ped: 24,2; AA-Adm: 17,1), *no estudo de* (AA-Ped: 36,4; AA-Adm: 34,2) e *o estudo de* (AA-Ped: 60,6; AA-Adm: 62,6).

Figura 1: Frequências de combinações de palavras contendo *estudo* nos *corpora*



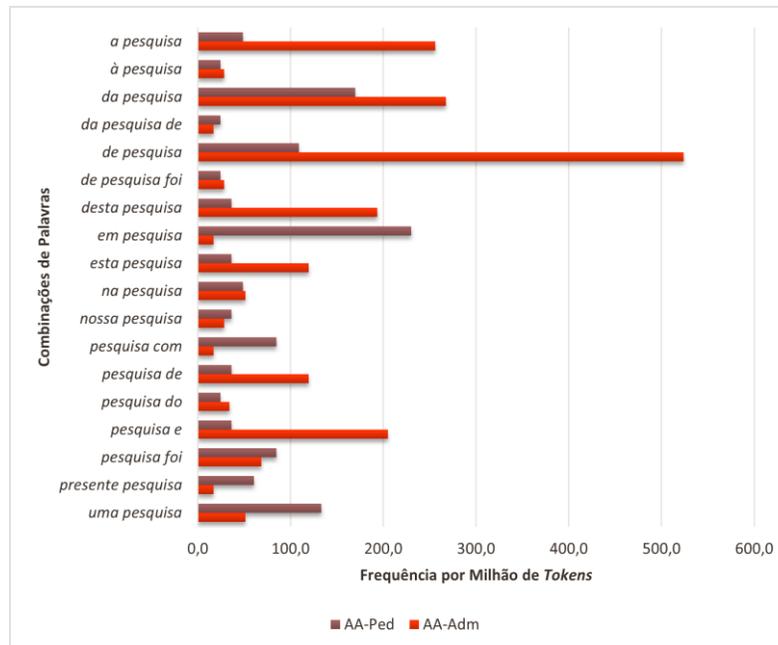
Fonte: Elaborada pelo autor



Quanto às combinações contendo a palavra *pesquisa*, foram encontradas 110 combinações em AA-Ped e 237 em AA-Adm. Apenas 18 combinações são comuns a ambos os *corpora*, o que corresponde a 16,4% das combinações de AA-Ped e 7,6% das combinações de AA-Adm. Sob outro ângulo, 83,6% das combinações com *pesquisa* encontradas em AA-Ped e 92,4% das combinações de AA-Adm são exclusivas. São exemplos de combinações exclusivas de AA-Ped *ética em pesquisa* (18 ocorrências), *pesquisa foi aprovada* (4 ocorrências), *pesquisa com seres humanos* (4 ocorrências), *pesquisa recente* (3 ocorrências), *da presente pesquisa* (3 ocorrências), *trata-se de uma pesquisa* (3 ocorrências), *pesquisa do hospital* (2 ocorrências) e *projeto de pesquisa* (2 ocorrências). São exemplos de combinações exclusivas de AA-Adm *pesquisa e desenvolvimento* (28 ocorrências), *nesta pesquisa* (21 ocorrências), *resultados desta pesquisa* (8 ocorrências), *pesquisa sobre* (5 ocorrências), *modelo de pesquisa* (5 ocorrências), *questão de pesquisa* (4 ocorrências), *hipóteses de pesquisa* (4 ocorrências) e *pesquisa bibliográfica* (3 ocorrências).

As combinações com *pesquisa* comuns a ambos os *corpora* apresentam, em sua maioria, frequências por milhão de *tokens* distintas (Figura 2). Exceções são *à pesquisa* (AA-Ped: 24,2; AA-Adm: 28,5), *da pesquisa de* (AA-Ped: 24,2; AA-Adm: 17,1), *de pesquisa foi* (AA-Ped: 24,2; AA-Adm: 28,5), *na pesquisa* (AA-Ped: 48,5; AA-Adm: 51,3) e *nossa pesquisa* (AA-Ped: 36,4; AA-Adm: 28,5).

Figura 2: Frequências de combinações de palavras contendo *pesquisa* nos *corpora*



Fonte: Elaborada pelo autor



Em conjunto, o alto número de combinações exclusivas de um e outro *corpus* e as diferenças marcadas nas frequências normalizadas das combinações compartilhadas indicam que, em textos oriundos de disciplinas distintas pertencentes a um único gênero acadêmico em português, palavras-chave comuns com alta distribuição e razoável incidência tendem a aparecer em conglomerados diversos. Sob uma perspectiva mais ampla, os resultados sugerem que palavras presentes em artigos de periódicos de campos distintos em português são predispostas ou primadas ao uso com companhias diferentes conforme o campo. As alegações referentes à relevância do domínio para o aparelhamento psicológico (HOEY, 2005, 2013) parecem, portanto, válidas para disciplinas acadêmicas, ao menos no âmbito do gênero artigo acadêmico em língua portuguesa. Os resultados, além disso, convergem com aqueles de Hyland (2008), que estudou *corpora* compostos por artigos acadêmicos, teses de doutorado e dissertações de mestrado em inglês de engenharia elétrica, biologia, administração e linguística aplicada (cerca de 3 milhões e 400 mil palavras). Hyland (2008) identificou apenas cinco conglomerados comuns aos quatro campos; ainda, mais da metade de todas as 50 combinações mais frequentes de cada um dos campos mostrou-se exclusiva. Assim como autores de textos acadêmicos em inglês, autores de artigos acadêmicos em português aparentemente são primados a usar uma ou outra combinação de palavras dependendo do campo de conhecimento.

O exame dos colocados de *estudo* produziu sete conjuntos semânticos tanto para o *corpus* AA-Ped como para AA-Adm (Tabela 5). Como pode ser observado, há cinco conjuntos comuns a ambos os *corpora*: AÇÃO COMO AGENTE, CONTEÚDO DO ARTIGO, METODOLOGIA DE PESQUISA E ESTATÍSTICA, O PRESENTE ESTUDO e TIPO DE ESTUDO. Destes, quatro apresentam colocados em comum, sendo que um tem exatamente os mesmos colocados nos dois *corpora* (O PRESENTE ESTUDO).

O exame dos colocados de *pesquisa*, por sua vez, gerou sete conjuntos semânticos para AA-Ped e cinco para AA-Adm (Tabela 6). Diferentemente dos conjuntos semânticos de *estudo*, porém, há apenas dois grupos compartilhados pelos *corpora*: A PRESENTE PESQUISA e METODOLOGIA DE PESQUISA E ESTATÍSTICA. Destes, apenas A PRESENTE PESQUISA apresenta os mesmos colocados para ambos os *corpora*, exceto *nesta*, que é exclusivo de AA-Ped.





Tabela 5: Conjuntos semânticos de *estudo* (substantivo) nos dois *corpora* (semelhanças em destaque)

<i>Corpus</i>	CONJUNTO (Colocados)
AA-Ped	AÇÃO COMO AGENTE (<i>analisou, avaliou, mostrou</i>)
	AÇÃO COMO ALVO (<i>aprovado, feito, realizado</i>)
	CONTEÚDO DO ARTIGO (<i>discussão, limitações, método, métodos, objetivo, resultados</i>)
	METODOLOGIA DE PESQUISA E ESTATÍSTICA (<i>abordagem, achados, amostra, associação, comitê, dados, desenho, fator, participantes, participar, participaram, período, prevalência, variáveis</i>)
	PEDIATRIA (<i>adolescentes, anos, caps, crianças, idade, lactentes, mães, tea</i>)
	O PRESENTE ESTUDO (<i>deste, este, neste, nosso, presente</i>)
	TIPO DE ESTUDO (<i>descritivo, multicêntrico, observacional, piloto, revisão, transversal</i>)
AA-Adm	AÇÃO COMO AGENTE (<i>considera, contribui, identificou, mostra, propõe, revela, sugere, utilizou</i>)
	ADMINISTRAÇÃO (<i>comportamento, consumidores, indústria</i>)
	CONTEÚDO DO ARTIGO (<i>conclusão, conclusões, contribuições, discussão, implicações, limitações, método, metodologia, objetivo, resultado, resultados, resumo</i>)
	METODOLOGIA DE PESQUISA E ESTATÍSTICA (<i>abordagem, amostras, campo, dados, desenho, efeito, evidências, fatores, hipóteses, modelo, natureza, objeto, participantes, pesquisa, pesquisadores, respondentes, teoria, variáveis</i>)
	OBJETIVO DE PESQUISA (<i>compreender, entender, identificar</i>)
	O PRESENTE ESTUDO (<i>deste, este, neste, nosso, presente</i>)
	TIPO DE ESTUDO (<i>caso, eventos, piloto</i>)

Fonte: Elaborada pelo autor

CAPS = Centro de Atenção Psicossocial; TEA = transtorno do espectro autista

Tabela 6: Conjuntos semânticos de *pesquisa* (substantivo) nos dois *corpora* (semelhanças em destaque).

<i>Corpus</i>	CONJUNTO (Colocados)
AA-Ped	AÇÃO COMO ALVO (<i>feita, realizada</i>)
	A PRESENTE PESQUISA (<i>desta, esta, nossa, presente</i>)
	ESCOPO (<i>brasil, nacional</i>)
	ÉTICA EM PESQUISA (<i>aprovada, aprovado, aprovou, assinatura, cep, comitê, envolvendo, ética, humanos, seres</i>)
	INSTITUIÇÕES DE PESQUISA (<i>centro, faculdade, instituição, universidade</i>)
	METODOLOGIA DE PESQUISA E ESTATÍSTICA (<i>aceitaram, dados, estudo, informações, participar, prevalência, projeto, qualitativa, resultados, transversal</i>)
	PEDIATRIA (<i>aleitamento, crianças, enfermeiras, hospital, mães, materno, medicina, saúde</i>)
AA-Adm	ADMINISTRAÇÃO (<i>consumidor, consumidores, gestores, serviços</i>)
	A PRESENTE PESQUISA (<i>desta, esta, nesta, nossa, presente</i>)
	CONTEÚDO DO ARTIGO (<i>contribuição, hipótese, hipóteses, implicações, metodologia, objetivo, questão, resultado, resultados</i>)
	METODOLOGIA DE PESQUISA E ESTATÍSTICA (<i>abordagem, amostra, categorias, construto, contexto, dados, estudo, método, modelo, responder, sujeitos, teórica, teórico, variáveis</i>)
	TIPO DE PESQUISA (<i>bibliográfica, campo, curso, pré-compra, qualitativa</i>)

Fonte: Elaborada pelo autor

CEP = comitê de ética em pesquisa



As diferenças nos conjuntos semânticos e seus componentes convergem com os resultados referentes aos conglomerados de palavras. Todavia, sob a ótica semântica, encontram-se mais pontos de encontro envolvendo a palavra *estudo* entre os dados de pediatria e administração. Parece haver um denominador comum entre as disciplinas que associa *estudo* a aspectos metodológicos e estatísticos, assim como ao texto que trata do próprio texto (metatexto). Em particular, a palavra *estudo* ocorre em ambos os *corpora* próxima de menções a componentes dos artigos, como *resultados*, *objetivo* e *limitações*, e após determinantes que mostram que o estudo em questão é aquele reportado no próprio artigo (*este estudo*, *nosso estudo*, etc.). De um lado, os resultados da análise semântica dão sustentação às alegações feitas por Hoey (2005, 2013) acerca da relevância do domínio para associações-primas. Autores de artigos de disciplinas distintas aparentemente são primados a usar palavras com grupos semânticos diferentes. Por outro lado, as semelhanças observadas entre os conjuntos semânticos de *estudo* sugerem que há palavras com menor propensão a diferenças disciplinares. *Pesquisa* parece mais predisposta ou primada que *estudo* a associações semânticas com especificidade disciplinar.

Por último, a análise da distribuição das ocorrências de *estudo* e *pesquisa* nos *corpora* produziu os gráficos das Figuras 3 e 4. Conforme já explicado, a medida adotada (D de Juilland) equivale a zero quando as ocorrências da palavra se concentram em um único ponto do texto e a 1 quando as ocorrências estão espalhadas por todo o texto. Como é possível constatar, *estudo* tem distribuição razoavelmente similar nos dois conjuntos de dados, mas exibe maior concentração espacial (menor dispersão) em AA-Adm. Já *pesquisa* tem distribuição bastante diversa nos *corpora*. Ao passo que nos dados de administração a palavra exibe em sua maioria (13 artigos) valores de dispersão próximos ou maiores que 0,6, nos dados de pediatria ela exibe valores inferiores a esse patamar na maior parte dos casos – só há dois artigos de pediatria com valores de dispersão acima de 0,6. Destaca-se, ainda, que há seis artigos de pediatria com valores de dispersão iguais a zero para *pesquisa* e somente um texto de administração com essa característica.

As distribuições das ocorrências de *estudo* e *pesquisa* nos *corpora* convergem com os resultados das análises colocacionais e semânticas. As duas palavras exibem comportamentos distributivos distintos em um e outro *corpus* e, portanto, em uma e outra disciplina, o que está de acordo com as proposições de Hoey (2005, 2013). Ainda assim, porém, constata-se que *pesquisa* é mais propensa que *estudo* a diferenças de uso entre artigos de uma e outra disciplina. *Estudo* parece ser uma palavra disciplinarmente menos marcada que *pesquisa*.

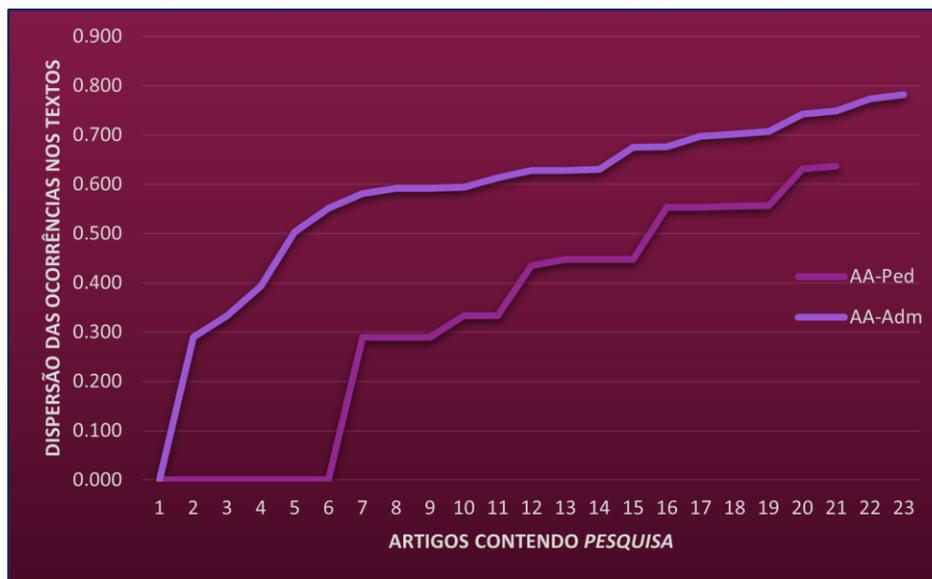


Figura 3: Dispersão das ocorrências da palavra *estudo* nos corpora



Fonte: Elaborada pelo autor

Figura 4: Dispersão das ocorrências da palavra *pesquisa* nos corpora



Fonte: Elaborada pelo autor

4 CONCLUSÕES, IMPLICAÇÕES E SUGESTÕES DE PESQUISA

O objetivo deste trabalho foi avaliar em que extensão o campo de conhecimento ou disciplina é relevante para o aparelhamento psicológico de autores de textos acadêmicos. Para alcançar esse objetivo, procedeu-se à análise de aspectos colocacionais, semânticos e distributivos das palavras *estudo* e *pesquisa* em coleções de artigos acadêmicos em português de pediatria e





administração. Como resultado, observaram-se diferenças quanto às combinações ou conglomerados que contêm as duas palavras enfocadas, aos conjuntos semânticos de uma e outra palavra e à distribuição das palavras nos artigos que compõem os *corpora*. Coletivamente, os resultados indicam a relevância do fator disciplina para os processos associativos mentais próprios da escrita acadêmica, corroborando as alegações de Hoey (2004, 2005, 2013) quanto à relação entre aparelhamento psicológico e domínio.

Apesar das diferenças nas associações-primas de ambas as palavras nos dados, constatou-se que *estudo* apresenta mais semelhanças de uso nos artigos de pediatria e administração que *pesquisa*. Não só há cinco conjuntos semânticos comuns entre os *corpora*, como a distribuição das ocorrências da palavra é relativamente similar nos artigos das duas disciplinas. Os resultados sugerem, assim, a existência de palavras mais e menos predispostas a exibir diferenças associativas relacionadas ao campo de conhecimento ou disciplina. *Estudo* parece ter um número menor de atributos disciplinarmente marcados em relação a *pesquisa*.

As implicações dos resultados para o ensino e a aprendizagem de escrita acadêmica em português podem ser consideradas sob duas óticas: a ótica do professor e a do aprendiz. Para o professor, a principal mensagem deste estudo é a de que autores de disciplinas diferentes possivelmente são primados a usar uma mesma palavra típica de artigos acadêmicos em português de formas distintas. Não há sinais de fórmulas lexicais aplicáveis igualmente a todas as áreas. Então, o ensino de escrita acadêmica requer um escopo disciplinar bem definido para que possa tratar adequadamente de aspectos léxico-gramaticais, semânticos e discursivos. Um ensino de escopo abrangente, generalista, provavelmente é incapaz de detalhar especificidades associativas disciplinares, ficando, pois, restrito a uma abordagem superficial.

Na mesma linha defendida por Hyland (2008), considera-se fundamental que o ensino de escrita acadêmica se baseie em dados, e não em suposições. Iniciativas como a de Kuhn (2017), cujo *corpus* de português acadêmico é organizado por áreas do conhecimento, são imprescindíveis para embasar esse ensino. São necessários *corpora* com componentes separados por campos ou disciplinas para fundamentar atividades pedagógicas especializadas que deem conta de particularidades associativas. De posse de dados organizados por disciplina, o professor pode buscar associações-primas úteis para seus alunos. Nessa busca, que pressupõe inúmeros encontros com palavras, o próprio professor será primado ao emprego de certos recursos linguísticos segundo as disciplinas de interesse, o que é especialmente vantajoso quando a formação de quem ensina difere daquela de quem aprende (por exemplo, um docente da área de letras frente a estudantes de engenharia).





Para o aprendiz, este estudo revela a necessidade de uma consciência disciplinar – assim como também sugere Hyland (2008) quanto à aprendizagem de escrita acadêmica em inglês. É comum que estudantes de graduação e pós-graduação – público natural dos cursos de escrita acadêmica no Brasil – tenham contato com textos acadêmicos de campos diversos durante seu período de formação. Por exemplo, um estudante de letras normalmente lê textos não só de linguística, mas também de literatura, sociologia, psicologia e filosofia. Um estudante de engenharia de produção se depara com textos não só de engenharia, mas também de administração, economia, física e estatística. Situações parecidas (aulas, etc.) em um ambiente comum (sala de aula, biblioteca, etc.) podem criar a ilusão de unidade no discurso acadêmico. Todavia, é possível que haja diferenças na escrita de uma e outra disciplina, assim como mostram este e outros estudos (HYLAND, 2008; REN, 2021; etc.). Naturalmente, dentro de uma mesma grande área do conhecimento, as diferenças possivelmente são menos marcadas do que as mostradas neste artigo. Ainda assim, pressupõe-se que existam. Dessa forma, é importante que o aprendiz reconheça as disciplinas dos textos que lê a fim de situar as combinações que encontra, tornando-se capaz de, em algum momento, filtrar aquelas que podem ser replicadas nos textos que produz; é essencial que esteja engajado em um processo de aprendizagem direcionada por dados (cf. SCOTT, 2011). Assim como o falante proficiente de português adequa o nível de formalidade de sua fala ao contexto da situação comunicativa em que se encontra, o aprendiz de escrita acadêmica deve adequar os recursos de sua escrita à disciplina e ao gênero textual em foco.

Para pesquisas futuras acerca da relação entre aparelhamento psicológico e disciplina acadêmica, algumas sugestões podem ser feitas. A primeira é um estudo semelhante a este acerca de palavras-chave comuns a artigos de outras disciplinas, como engenharia e biologia. Os resultados poderiam confirmar as conclusões deste trabalho ou não. A segunda sugestão é um estudo das mesmas palavras (*estudo e pesquisa*) em artigos de dois ramos da administração (por exemplo, *marketing* e gestão de pessoas) ou de dois ramos da medicina (por exemplo, pneumologia e ortopedia). Os resultados poderiam mostrar se em disciplinas mais próximas as diferenças associativas são menos marcadas ou não (pressupõe-se que sejam). A terceira e última sugestão é uma análise de um conjunto maior de palavras-chave comuns a *corpora* de artigos de disciplinas diversas para elucidar o que caracteriza palavras mais e menos propensas a exibir associações distintivas entre disciplinas.



REFERÊNCIAS

ANTHONY, L. Writing research article introductions in software engineering: How accurate is a standard model? **IEEE Transactions on Professional Communication**, v. 42, n. 1, p. 38-46, 1999.

ANTHONY, L. **AntConc**. Versão 4.0.3. Tokyo: Waseda University, 2022. Disponível em: <https://www.laurenceanthony.net>. Acesso em: 24 fev. 2022.

CREATIVE COMMONS BRASIL. **Sobre as licenças**. [ca. 2022]. Disponível em: <https://br.creativecommons.net/licencas/>. Acesso em: 28 abr. 2022.

CUNHA, A. L. da. **Análise do priming lexical das composições sintagmáticas em textos publicitários**: Por uma metodologia em que o léxico seja o ponto central do ensino de língua portuguesa. 2017. 211 f. Tese (Doutorado em Linguística Aplicada) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017. Disponível em: <http://hdl.handle.net/1843/LETR-AN6KNY>. Acesso em: 22 fev. 2022.

DUGUID, A.; PARTINGTON, A. Forced lexical primings in transdiscursive political messaging: How they are produced and how they are received. In: PACE-SIGGE, M.; PATTERSON, K. J. (ed.). **Lexical priming: Applications and advances**. Amsterdam: John Benjamins, 2017. p. 67-92.

GOATLY, A. Lexical priming in humorous discourse. **European Journal of Humour Research**, v. 5, n. 1, p. 52-68, 2017. Disponível em: <http://dx.doi.org/10.7592/EJHR2017.5.1.goatly>. Acesso em: 22 fev. 2022.

HADIKIN, G. S. You know Korean English? Lexical priming in short strings of Korean spoken English. **International Journal of Language Studies**, v. 7, n. 1, p. 59-78, 2013. Disponível em: <http://www.ijls.net/pages/volume/vol7no1.html>. Acesso em: 22 fev. 2022.

HOEY, M. The textual priming of lexis. In: ASTON, G.; BERNARDINI, S.; STEWART, D. (ed.). **Corpora and language learners**. Amsterdam: John Benjamins, 2004. p. 21-41.

HOEY, M. **Lexical priming: A new theory of words and language**. Abingdon, UK: Routledge, 2005.

HOEY, M. Lexical priming and literary creativity. In: HOEY, M.; MAHLBERG, M.; STUBBS, M.; TEUBERT, W. **Text, discourse, and corpora: Theory and analysis**. London: Continuum, 2007. p. 7-30.

HOEY, M. Lexical priming. In: CHAPELLE, C.A. (ed.). **The encyclopedia of applied linguistics**. London: Wiley-Blackwell, 2013. p. 3342-3347.

HOEY, M. Cohesion and coherence in a content-specific corpus. In: PACE-SIGGE, M.; PATTERSON, K. J. (ed.). **Lexical priming: Applications and advances**. Amsterdam: John Benjamins, 2017. p. 3-40.





HYLAND, K. As can be seen: Lexical bundles and disciplinary variation. **English for Specific Purposes**, v. 27, p. 4-21, 2008.

JANTUNEN, J. H. Lexical and morphological priming: A holistic phraseological analysis of the Finnish time expression *kello*. In: PACE-SIGGE, M.; PATTERSON, K. J. (ed.). **Lexical priming: Applications and advances**. Amsterdam: John Benjamins, 2017. p. 253–271.

KILGARRIFF, A. Simple maths for keywords. In: CORPUS LINGUISTICS CONFERENCE, 2009, Liverpool. **Proceedings of [...]**. Liverpool: University of Liverpool, 2009. Páginas não numeradas.

KUHN, T. Z. **A design proposal of an online corpus-driven dictionary of Portuguese for university students**. 2017. 421 f. Tese (Doutorado em Linguística Aplicada) – Universidade de Lisboa, Lisboa, 2017. Disponível em: <http://hdl.handle.net/10451/32013>. Acesso em: 26 fev. 2022.

LEXICAL COMPUTING CZ. **Sketch Engine**. Brno, Czech Republic: Lexical Computing CZ, [ca. 2022]. Disponível em: <https://www.sketchengine.eu/>. Acesso em: 5 mar. 2022.

LEXICAL COMPUTING LTD. **Statistics used in the Sketch Engine**. July 8, 2015. Disponível em: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf>. Acesso em: 24 fev. 2022.

OOI, V. Lexical priming and Asian learners of English. **Learner Corpus Studies in Asia and the World**, v. 1, p. 31-41, 2013. Disponível em: http://www.lib.kobe-u.ac.jp/infolib/meta/pub/G0000003kernel_81006673. Acesso em: 22 fev. 2022.

REN, J. Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. **Journal of English for Academic Purposes**, v. 50, artigo n. 100968, 2021.

RYCHLÝ, P. A lexicographer-friendly association score. In: RECENT ADVANCES IN SLAVONIC NATURAL LANGUAGE PROCESSING, 2., 2008, Brno. **Proceedings of [...]**. Brno: Masaryk University, 2008, p. 6-9.

SALIM, H. Y. A. **Collocation and other lexical relationships in translations of the Quran: A corpus based application of lexical priming theory to a unique theological text**. 2012. 320 f. Tese (Doctor of Philosophy) – University of Liverpool, Liverpool, 2012.

SCOTT, M. Aprendizagem direcionada por dados: Uma homenagem a Tim Johns (1936–2009). In: VIANA, V.; TAGNIN, S. E. O. (org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2011. p. 7-11.

SAMRAJ, B. Introductions in research articles: Variations across disciplines. **English for Specific Purposes**, v. 21, n. 1, p. 1-17, 2002.

SWALES, J. **Aspects of article introductions**. Birmingham: The University of Aston, 1981.

TONELLI, M. J.; ZAMBALDI, F. Ciência responsável e impacto social da pesquisa em administração. **Revista de Administração de Empresas**, São Paulo, v. 58, n. 3, p. 215-216, 2018. Disponível em: <https://doi.org/10.1590/S0034-759020180301>. Acesso em: 28 abr. 2022.





Artigo recebido em: 22/03/2022

Artigo aprovado em: 29/04/2022

Artigo publicado em: 03/05/2022

COMO CITAR

ARAGÃO, R. M. L. de. Indícios de aparelhamento psicológico em artigos acadêmicos de pediatria e administração. **Diálogo das Letras**, Pau dos Ferros, v. 11, p. 1-19, e02202, 2022.

