

O papel das colocações acadêmicas no processo de aprimoramento da escrita

The role of academic collocations in the process of writing improvement

Giseli Aparecida CECÍLIO*

Universidade Estadual Paulista (UNESP)

Adriane ORENHA-OTTAIANO**

Universidade Estadual Paulista (UNESP)

Tanara Zingano KUHN***

Universidade de Coimbra (CELGA-ILTEC)

RESUMO: Este artigo apresenta um estudo sobre as colocações acadêmicas identificadas em dois *corpora* de pesquisa formados por artigos científicos em Português do Brasil, das áreas de Letras/Linguística (*The Scientific Electronic Library Online - SciELO Brasil*) e Engenharias (CoPEP, Kuhn, 2018). Para isso, temos como fundamentação teórica a Linguística de Corpus (Biber, 2006; Kuhn, 2017; Kuhn; Ferreira, 2020) e a Fraseologia (Tagnin, 1989; Orenha-Ottaiano, 2004, 2015, 2016, 2020). O objetivo foi investigar a variação de sentido dessas colocações acadêmicas em função das especificidades temáticas de cada área. A partir dessa análise, propôs-se a inclusão dos verbetes examinados na plataforma *on-line* Dicionário de Colocações Acadêmicas (DICA), com o propósito de auxiliar estudantes no desenvolvimento de habilidades e competências de leitura e produção textual, exigidas para o contexto acadêmico.

PALAVRAS-CHAVE: Colocações Acadêmicas. Dicionário de Colocações Acadêmicas Fraseologia. Linguística de Corpus.

ABSTRACT: This paper presents a study on academic collocations identified in two research corpora composed of scientific papers in Brazilian Portuguese, from the fields of Literature/Linguistics (*The Scientific Electronic Library Online – SciELO Brasil*) and Engineering (CoPEP, Kuhn, 2018). Therefore, we have a theoretical framework: Corpus Linguistics (Biber, 2006; Kuhn, 2017; Kuhn; Ferreira, 2020) and Phraseology (Tagnin, 1989; Orenha-Ottaiano, 2004, 2015, 2016, 2020). The aim was to investigate the variation in meaning of these academic collocations in light of the thematic specificities of each field. Based on this analysis, the inclusion of the examined entries in the online platform *Dicionário de Colocações*

* Doutoranda em Estudos Linguísticos pela Universidade Estadual Paulista, São José do Rio Preto, SP, email: giselicecilio@unesp.br.

** Professora Livre-Docente em Fraseologia e Fraseografia baseadas em Corpus Programa de PósGraduação em Estudos Linguísticos na Universidade Estadual Paulista, São José do Rio Preto, SP, email: adriane.ottaiano@unesp.br.

*** Professora e Membro da Direção Executiva (CELGA-ILTEC, UC) Universidade de Coimbra Centro de Estudos de Linguística Geral e Aplicada, na Faculdade de Letras, Largo da Porta Férrea, Coimbra, Portugal, email: tanarazingano@uc.pt.

Acadêmicas (DICA) was proposed, with the purpose of assisting students in developing the reading and writing skills and competences required in the academic context.

KEYWORDS: Academic Collocations, Academic Collocations Dictionary. Phraseology. Corpus Linguistic.

Introdução

É notório o crescente número de estudantes que todos os anos ingressam em universidades públicas e particulares no Brasil. Almeida et al. (2012 *apud* Kuhn, p. 41, 2017), destaca que esse aumento deve-se a vários fatores, mas, principalmente, ao fenômeno de “massificação” do ensino superior. Os programas destinados a fundos de investimentos na educação, nacionais e transnacionais ligados a empresas do setor educacional privado e público contribuem para esse fenômeno. Entre esses programas, estão o SiSU (Sistema de Seleção Unificado), desenvolvido pelo MEC, cujo objetivo é selecionar candidatos para as vagas disponibilizadas pelas universidades públicas através do ENEM (Exame Nacional do Ensino Médio), entre outros que também corroboram para o processo de massificação, como o FIES (Fundo de Financiamento Estudantil), criado também pelo MEC, com o objetivo de financiar a educação acadêmica de graduandos que e o PROUNI (Programa Universidade para Todos) para estudantes de baixa renda ingressar em cursos de graduação em instituições particulares, que seriam beneficiadas com isenções fiscais, cujo processo de concessão se dá a partir da nota obtida no ENEM.

Um dos alicerces desse processo de “massificação” foi, e ainda é, a instituição privada, que flexibiliza os meios de acesso para os alunos ingressantes, seja, por valores acessíveis, seja, pela modalidade EAD, favorecendo esse crescimento. Parte dos estudantes saem da realidade do ensino médio entusiasmados com essa nova etapa, entretanto, o processo de transitoriedade da escola para a universidade, por vezes, é impactante, gerando dificuldades.

Esses aspectos ficam evidentes quando os alunos se deparam com a linguagem dos textos científicos que, de acordo com Fischer, (2007, p. 51) “está presente ou compõe os gêneros discursivos do meio acadêmico” e, tão logo iniciam suas atividades acadêmicas, os alunos de graduação, por exemplo, precisam produzir textos como: resumos, resenhas, entre outros tipos de trabalhos que demonstrem seu conhecimento

acadêmico, enquanto que alunos de pós-graduação precisam desenvolver textos para apresentações em congressos, artigos, bem como suas dissertações ou teses, com o objetivo de divulgar seus trabalhos em revistas, congressos, simpósios, seminários, haja vista o crescente número de publicações de textos acadêmico-científicos pelas mais diversas áreas de conhecimento.

É importante salientar que, ao abordarmos as dificuldades na escrita dos alunos, estamos nos referindo aos gêneros utilizados no ambiente acadêmico até então desconhecidos pelos estudantes, e que as universidades buscam cumprir seu papel na formação de seus alunos, em relação ao conhecimento desses gêneros. Entretanto, Biber (2006) constata que o processo para adquirir uma escrita acadêmica adequada é desafiador. De acordo com o pesquisador, além de os estudantes precisam se ajustar à “realidade universitária”, o que se faz necessário para alcançar seus objetivos, muitos obstáculos devem ser superados para se adequarem aos padrões acadêmicos esperados.

Dessa maneira, tendo em vista a situação descrita acima, propomos um estudo a partir das colocações acadêmicas encontradas em artigos científicos em português brasileiro, que não possuam um valor terminológico dentro dos textos das áreas selecionadas, com a finalidade de contribuir para o aprendizado dos gêneros discursivos.

Dessa maneira, selecionamos duas grandes áreas, seguindo a classificação da “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES), que a define como conjunto de “diversas áreas do conhecimento, em virtude da afinidade de seus objetos, métodos cognitivos e recursos instrumentais refletindo contextos sociopolíticos específicos”².

Para conduzir esta pesquisa desenvolvida no mestrado, foram compilados dois *corpora*, quais sejam, um *subcorpus* da grande área de Engenharias (SCEs) com 441.591 palavras, criado a partir do CoPEP (*Corpus* de Português Escrito em Periódicos), desenvolvido por Kuhn e Ferreira (2020), um *corpus* de artigos do *Scielo* (*Scientific Library On-line*), da grande área de Letras/Linguística (CLL), contendo 476.191 palavras.

A decisão de fazer uma pesquisa voltada para os gêneros acadêmicos se deu por vários motivos, entre eles: 1) as dificuldades que permeiam estudantes, no que diz respeito

² Definição retirada do site Capes - <https://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>

à escrita acadêmica, 2) a escassez de materiais comprovados por alguns pesquisadores e, principalmente, 3) a necessidade visível de alternativas pedagógicas para amenizar essas dificuldades na escrita.

Pensando nessas “alternativas pedagógicas”, e observando a relevância das colocações no contexto acadêmico, propomos o levantamento e análise de colocações acadêmicas que serão inseridas em uma plataforma de Dicionários de Colocações Acadêmicas (DICA)³, em português do Brasil e em inglês, como um dos trabalhos desenvolvidos no âmbito do Grupo de Pesquisa do CNPq "FRASCORP - Fraseologia e Colocações Baseado em Corpora", bem como do projeto de pesquisa “Desenvolvimento de uma metodologia e aprimoramentos de recursos lexicográficos para uma Plataforma On-Line de Dicionários de Colocações Acadêmicas em Português e Inglês” (Processo CNPq CNPq/MCTI/FNDCT N° 18/2021).

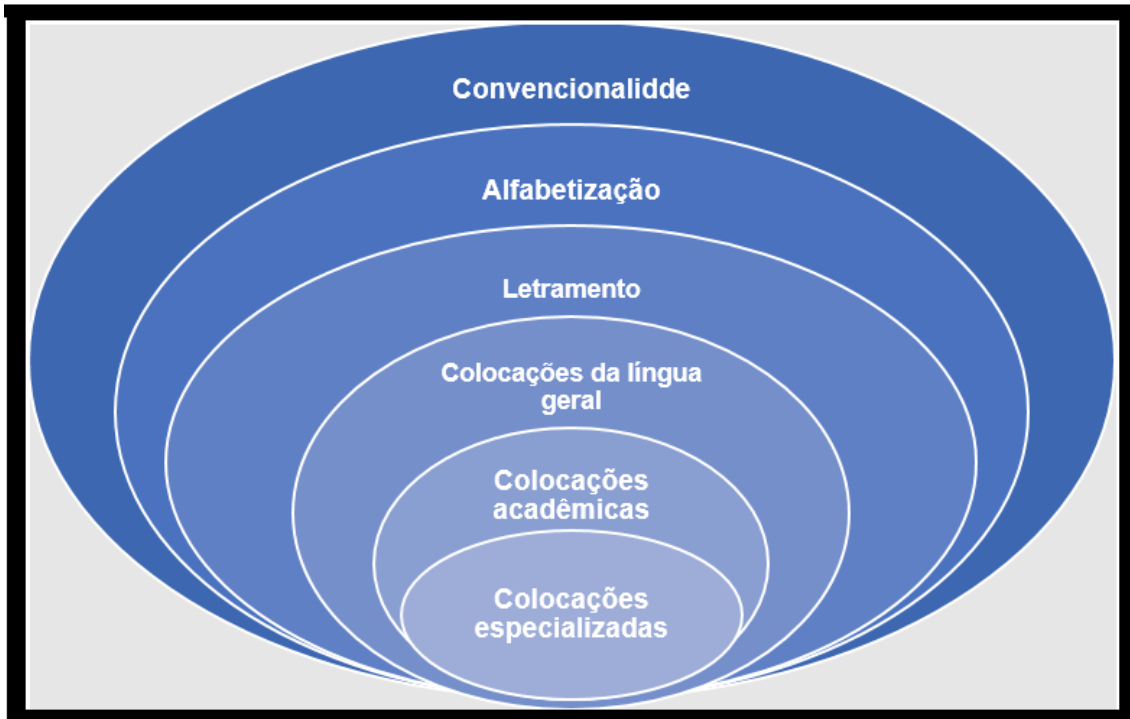
1 O Percurso para o conhecimento das colocações

Como mencionamos na introdução deste artigo, são vários os desafios dos estudantes quando inseridos na realidade acadêmica, entre eles: a) os gêneros discursivos do meio acadêmico (Fisher, 2007, p. 51) e, b) as lacunas preexistentes do período anterior à universidade (Alves, 2016, p. 1). Dessa maneira, podemos, a partir desses apontamentos, adentrar nas questões da convencionalidade e, assim, das colocações da língua geral.

Se pudéssemos descrever todo o percurso do desenvolvimento da linguagem de um indivíduo até a sua inserção no meio acadêmico, seria necessário que abarcássemos várias teorias, no entanto, para descrever as etapas, de modo mais didático e breve, iniciando pelas de etapas de aprendizagem primárias, passando pelas colocações da língua geral, colocações acadêmicas, chegando às colocações especializadas, elaboramos um organograma expandido a partir do organograma desenvolvido por Orenha-Ottaiano (2016):

³ O DICA pode ser acessado pelo site: <https://www.godica.com.br/>. Os dicionários de colocações acadêmicas ainda estão em processo de desenvolvimento e serão disponibilizados em breve futuro.

Organograma 1: Etapas da aprendizagem – da convencionalidade às colocações especializadas



Fonte: Elaboração própria.

Através deste organograma, podemos verificar o processo que envolve o indivíduo em contato com a linguagem, partindo dos princípios da convencionalidade da língua que, segundo Tagnin (1989), pode acontecer em diferentes níveis da língua; da alfabetização que compreende a aquisição da leitura e escrita (Soares, 2002); seguido do letramento que foca “os aspectos sócio-históricos da aquisição de um sistema escrito por uma sociedade” (Tfouni, 1988 *apud* Soares, 2002, p.3).

Retomando a convencionalidade que se faz presente em todas as etapas da linguagem, temos o nível sintático, o qual se refere à relação arbitrária que existe entre uma palavra e seu significado, bem como nos níveis fonológico, morfológico, sintático, semântico e pragmático. Todavia, por se tratar de uma pesquisa que tem como foco as colocações, priorizamos o nível sintático e semântico.

Sob a ótica de Fillmore convencionalidade (1979), pode ser compreendida como “o conjunto dos elementos linguísticos, cuja coocorrência não é explicada sintática ou semanticamente, mas, sim, pelo uso”. Orenha-Ottaiano (2015, p. 844), salienta que a

convencionalidade está intimamente ligada à fluência na língua. Ou seja, independente de nacionalidade, conhecer as convenções depende do nível de conhecimento que se tem de língua e, por conseguinte, dos fraseologismos, onde se insere as colocações. Isso implica em um discurso, em uma leitura ou escrita fluente, ou seja, sem impedimentos para a comunicação, em qualquer esfera da sociedade, seja em inglês, português, espanhol, francês, etc.

Desse modo, da convencionalidade, que abarca todas as etapas da aquisição da linguagem, adentramos às colocações que, de acordo com Orenha-Ottaiano (2004, p. 22), trata-se de “uma das categorias mais relevantes” no ensino-aprendizagem de uma língua estrangeira. Segundo Orenha-Ottaiano (2020), as colocações “são combinações recorrentes, arbitrárias, convencionalizadas na língua, abrangentes, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito. São, ainda, combinações específicas de um idioma e de uma cultura e, como tal, a colocabilidade de seus elementos pode variar significativamente de um idioma para outro, de uma área para outra, sendo constituído, portanto, por sua própria rede de colocações”.

Pensando nesse processo de transitoriedade, da linguagem geral para a linguagem científica, é preciso pensar em novas maneiras de auxiliar estudantes das mais variadas áreas a usar a linguagem de modo coerente no seu contexto acadêmico. Assim, novas possibilidades pedagógicas surgem através do estudo das colocações, de modo particular, das colocações acadêmicas. Diante disso, podemos nos perguntar: qual o papel das colocações acadêmicas para auxiliar estudantes no aprimoramento de sua escrita? É importante mencionar que na língua geral, usamos diariamente, de acordo com Orenha-Ottaiano (2020), “blocos pré-fabricados” de palavras, e o mesmo ocorre dentro da escrita científica. Dessa maneira, destaca a pesquisadora que, ao utilizarmos estas “unidades fixas” no dia a dia, mantemos um discurso natural, o que não é diferente quando precisamos produzir nossos textos científicos. Dessa forma, as colocações assumem um enorme papel somatório no auxílio daqueles que precisam melhorar sua escrita acadêmica.

No entanto, acreditamos que esses materiais pedagógicos precisam vir acompanhados de meios facilitadores, ou seja, de meios que simplifiquem o acesso a eles:

1) que sejam didáticos e de fácil compreensão para seu manuseio; 2) que sejam de fácil acesso; e 3) que sejam gratuitos.

É exatamente essas características que a plataforma DICA possui. Sua página inicial nos mostra todas as informações de modo simples e objetivo, apresentando uma interface, embora ainda em desenvolvimento, prática e funcional, dando acesso rápido às ferramentas de busca e aos recursos, em construção, da plataforma.

Figura 1: Plataforma on-line de Dicionários de Colocações Acadêmicas



Fonte: DICA (<https://www.godica.com.br/pt>)

A Plataforma on-line de Dicionários de Colocações Acadêmicas, embora como mencionado acima, esteja em desenvolvimento, nela será possível consultar as colocações acadêmicas com exemplos contextualizados. A busca poderá ser feita, inicialmente, em português, inglês, espanhol, francês ou chinês, com filtros adicionais para selecionar o tipo de colocação desejada como adjetiva, adverbial, nominal, verbal ou complexa, nas mais variadas áreas do conhecimento.

A plataforma de estudos colocacionais foi desenvolvida pela Profa. Dra. Adriane Orenha-Ottaiano, na Universidade Paulista “Júlio de Mesquita Filho” (UNESP/IBILCE), e é o resultado de mais de 10 anos de pesquisa.

O DICA faz parte da plataforma digital PLATCOL, a qual representa um avanço significativo na consolidação de práticas linguísticas com a finalidade auxiliar a produção no contexto acadêmico. Desenvolvido com base em *corpora* especializados e

metodologia pautada em dados empíricos, o Dicionário oferecerá aos usuários um grande número de colocações amplamente utilizadas em gêneros acadêmicos, especialmente na redação de artigos, dissertações e teses.

A relevância do dicionário reside em sua proposta pedagógica e funcional, ao proporcionar aos usuários subsídios para a escolha lexical mais naturalizada, evitando traduções literais ou combinações inadequadas que comprometem a clareza textual do texto produzido. Além disso, o dicionário atende a múltiplos perfis, desde estudantes de pós-graduação até tradutores e docentes, possibilitando consultas em diversos idiomas, os quais serão expandidos para outros futuramente, o que o torna um recurso relevante tanto para a produção quanto para a tradução acadêmica.

Do ponto de vista prático, o DICA se destaca por sua interface intuitiva, pois permitirá buscas personalizadas por área do conhecimento e exemplos contextualizados, o que reforça seu valor como instrumento de formação linguística contínua. A ferramenta também se mostra eficaz para o ensino e aprimoramento da escrita acadêmica, contribuindo para que os usuários internalizem padrões colocacionais convencionais da escrita científica em diversas línguas.

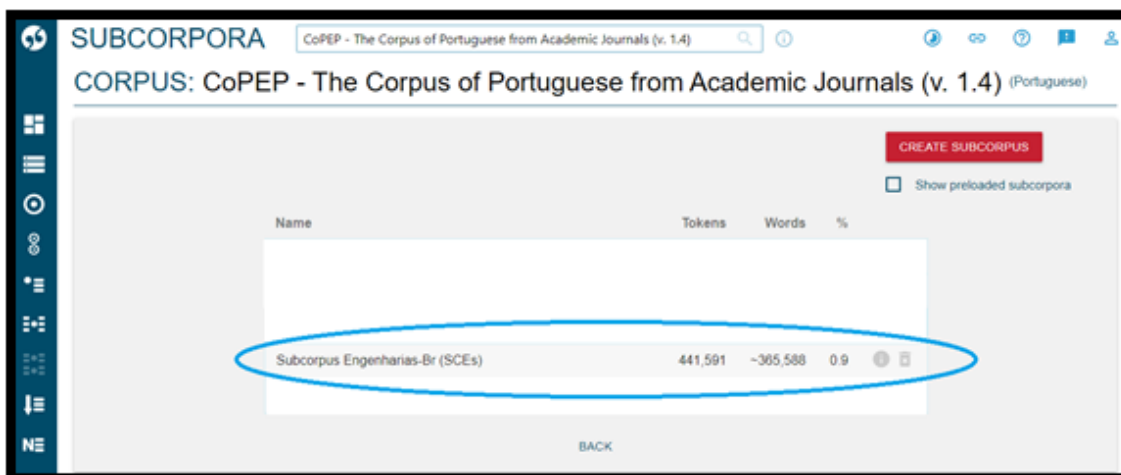
Dessa forma, o DICA não se limita aos propósitos de um dicionário tradicional, mas se configura como uma plataforma de apoio didático e linguístico orientada por princípios da Linguística de Corpus, da Fraseologia e da Lexicografia. Sua utilização por pesquisadores, estudantes e tradutores auxiliará na qualidade da escrita científica, especialmente em contextos de internacionalização e publicação em língua estrangeira.

2 Procedimentos metodológicos para a compilação dos *Corpora*

Com o propósito de desenvolver um estudo que contemplasse as questões levantadas acima selecionamos, no Sketch Engine, o sub*corpus* de português brasileiro de Engenharias pertence ao CoPEP (Kuhn; Ferreira, 2018), um *corpus* com 40 milhões de palavras, balanceado entre as variedades Português Brasileiro e Português Europeu e que cobre seis áreas de conhecimento. De acordo com os procedimentos adotados por Kuhn (2017), os metadados do *corpus* foram criteriosamente anotados, permitindo fazer pesquisas avançadas. É o primeiro *corpus* internacional de português acadêmico de que temos conhecimento. Seu desenvolvimento se deu a partir do SciELO, como mencionado

na seção anterior. Contudo, fizemos um recorte somente do português do Brasil, restringindo nossa pesquisa à grande área de Engenharias. Desse modo, o corpus de Engenharia que usamos neste trabalho contém 441,591 *tokens*.

Figura 2: *Layout da página do Subcorpus Engenharias -Br (SCEs) – CoPEP Engenharias*



Fonte: Printscreen da página do site Sketch Engine (<https://www.sketchengine.eu/#ca-subcorpora/>)

A seguir, trataremos da compilação do corpus da área Letras/Linguística.

2.1 Compilação do corpus de artigos da grande área Letras/Linguística

Dentre os 16 periódicos apontados na tabela 2, de ARS (São Paulo) até o de *Trabalhos em Linguística Aplicada*, descartamos, logo de início, o periódico *Per Musi*, por tratar sobre a reflexão científica na área de música. Depois, excluímos outros oito periódicos, alguns por conterem assuntos muito específicos em seus artigos, entre estes, assuntos relacionados às artes cênicas, outros por conter muitos textos em outras línguas.

Sendo assim, optamos pelas seguintes revistas para compor o corpus de artigos da grande área letras/linguística:

Quadro 1: Dados sobre as revistas selecionadas para a compilação do *Corpus* de Artigos

Número de Artigos Selecionados para a Pesquisa	
Revista Alfa	12 artigos
Revista Brasileira de Linguística Aplicada	12 artigos

Trabalhos em Linguística	11 artigos
Revista Alea	14 artigos
Revista Letras de Hoje	12 artigos
Estudos de Literatura Brasileira Contemporânea	15 artigos

Fonte: SciELO Brasil.

Selecionamos, portanto, seis revistas, optando por iniciar a constituição a partir dos anos de 2015 a 2019, pois pretendemos trabalhar com o maior número possível de textos que fossem recentes e com uma variedade de revistas, a fim de ganharmos na diversidade de estilo dos periódicos, porém, buscando um número semelhante entre as revistas, a fim de manter um balanceamento entre as mesmas.

A partir das revistas selecionadas, retiramos artigos escritos em outras línguas como inglês, espanhol, entre outras, utilizando o processo de escolha que seguia os seguintes critérios: 1) líamos o título, 2) verificamos se o artigo se encontrava na língua portuguesa e, 3) para finalizar, líamos o resumo do artigo.

Relembrando, nosso objetivo era manter o equilíbrio entre o número de palavras/tokens dos *corpora* de Engenharia e de Letras e Linguística. Assim, ao final desse processo, o CLL contabilizou 476.191 tokens.

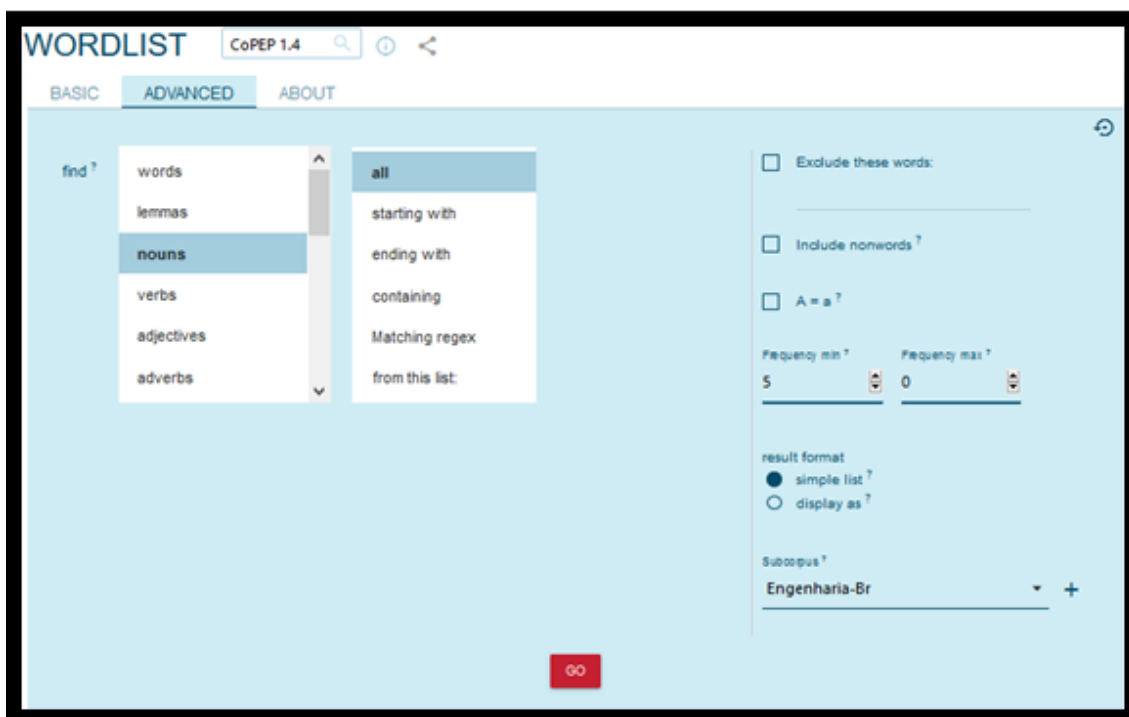
2.2 Procedimentos adotados para seleção dos dados

Conforme informamos na Introdução, delimitamos o escopo das nossas análises aos 27 substantivos mais frequentes que ocorrem em ambos os corpora. Essa delimitação se deu por três razões. Primeiramente, entendemos que o fato de haver equivalência de formas não garante, *a priori*, que os sentidos serão os mesmos nas duas áreas. Parte-se da hipótese de que algumas colocações possam apresentar sentidos comuns às duas áreas, enquanto outras tendem a assumir significados distintos, em função de seus contextos específicos. Em segundo lugar, em função de restrições de espaço e tempo, a determinação desse critério restringe os dados a uma quantidade com a qual podemos investigar. Finalmente, como nosso trabalho vai além da extração de colocações, pois

envolve também suas análises e a proposta de verbetes-piloto, o trabalho com um número menor de dados serve para testar a metodologia.

Para tanto, inicialmente criamos uma lista com os 100 substantivos de maior ocorrência nos *corpora* das áreas de Engenharias e Letras/Linguística. Essas listas foram geradas a partir da função *WordList* do *Sketch Engine*.

Figura 3: *WordList* do *Subcorpus* de Engenharias-Br (SCEs)



Fonte: Dados obtidos a partir da ferramenta *WordList* do programa *Sketch Engine* (<https://www.sketchengine.co.uk/>)

Seguindo o mesmo processo da figura 3, também para Letras/Linguística, identificamos os 100 substantivos mais frequentes nas áreas selecionadas para a presente pesquisa. Em seguida, partimos para a identificação dos que eram comuns às duas áreas, processo este que foi manual e que, no final, nos levou à seleção de 27 substantivos em ambas as áreas acadêmicas, conforme abordaremos no subtópico a seguir.

2.3 Identificação dos Substantivos em comum no CLL e no SCEs

Após gerar a lista dos 100 substantivos mais frequentes em cada uma das grandes áreas, Engenharia e Letras/Linguística, buscamos identificar manualmente quais

eram comuns às duas áreas. Feito isto, encontramos 27 substantivos em comum nas respectivas áreas.

Um fato que torna ainda mais premente a necessidade de se considerar a possível variação de sentido das palavras acima elencadas, em função das colocações que integram e das áreas de conhecimento em que são empregadas, é que muitas delas também são as mais frequentes da língua geral. Observemos a tabela abaixo, que apresenta os 100 substantivos mais frequentes do SCEs, CLL e *subcorpus* Brasileiro do *Portuguese Web* 2011(pt Ten Ten 11), que é um corpus de textos da Web disponível no *Sketch Engine*⁴. A parte brasileira contém mais de 3,5 bilhões de *tokens*.

Tabela 1: Dados obtidos a partir da ferramenta *WordList* do programa *Sketch Engine*

Subcorpus de Engenharias			Corpus de Letras/Linguística			Corpus de Portuguese Web 2011 (Brazilian variety)		
Ordem	Substantivo	Freq.	Ordem	Substantivo	Freq.	Ordem	Substantivo	Freq.
6	Tempo	888	3	Forma	1135	1	Ano	8.041.492
7	Caso	817	4	Relação	937	4	Trabalho	3.684.040
10	Sistema	753	7	Veza	685	5	Veza	3.539.759
12	Processo	734	8	Trabalho	661	8	Projeto	3.109.806
13	Resultado	729	9	Exemplo		9	Tempo	3.079.055
14	Trabalho	645	10	Processo	646	12	Forma	2.866.177
15	Função	629	12	Tempo	615	23	Caso	2.203.301
17	Forma	622	15	Autor	588	26	Processo	2.079.300
23	Exemplo	476	18	Meio	560	29	Grupo	2.002.963
25	Estudo	456	23	Estudo	514	31	Sistema	1.963.022
26	Dado	451	25	Caso	510	39	Relação	1.759.417
27	Análise	442	26	Uso	508	41	Ponto	1.737.398
29	Tipo	432	27	Pesquisa	498	42	Meio	1.706.697
31	Relação	416	35	Produção	447	45	Escola	1.649.578
35	Ponto	373	38	Análise	422	12	Exemplo	1.516.172

⁴ <https://www.sketchengine.eu/pttnten-portuguese-corpus/>

38	Pesquisa	362	55	Tipo	329	63	Resultado	1.435.537
43	Escola	342	59	Grupo	294	68	Pesquisa	1.404,816
45	Vez	334	61	Função	288	69	Tipo	1.400.619
63	Ano	287	65	Ponto	285	74	Produção	1.331.864
69	Grupo	271	67	Condição	281	92	Uso	1.159.781
70	Produção	268	69	Sistema	279	99	Estudo	1.121.056
72	Ideia	265	81	Ideia	251			
73	Autor	263	82	Ano	249			
74	Uso	260	84	Dado	247			
80	Projeto	246	86	Escola	241			
83	Meio	236	88	Resultado	237			
96	Condição	209	90	Projeto	232			

Fonte: Elaboração Própria.

Podemos observar que, dos 27 substantivos de maior frequência nos *corpora*, 21 deles são os mais frequentes no *corpus* de língua geral, fazendo parte do vocabulário de todos nós diariamente. Todavia, o grande desafio é utilizar essas palavras academicamente.

De fato, pesquisadores mostraram que a maior dificuldade dos alunos é usar palavras da língua geral de forma apropriada em gêneros acadêmicos (Kosem, 2010; Todd, 2017; *apud* Kuhn, 2017, 66-7), principalmente quando essas palavras precisam ser combinadas dentro do texto científico. Diante desses dados apresentados, faz jus o estudo não somente das palavras que se especializam em cada contexto, mas também das colocações que encontramos em comum nas duas áreas, e que assumem significados distintos, especializando-se, em função da área de conhecimento ao qual pertencem.

Para seguir, portanto, com a proposta de identificar e analisar as colocações acadêmicas, utilizamos a ferramenta *Word Sketch*, a fim de encontrar as colocações e, a *Concordance*, para verificar o contexto das ocorrências. Isso nos permite fazer uma análise comparativa e contrastiva, pois podemos observar como cada substantivo se comporta dentro de ambos os *corpora*.

Na tabela abaixo, mostramos os procedimentos metodológicos que foram seguidos, com descrição das ações desenvolvidas em cada uma delas.

Quadro 2: Dados sobre as revistas selecionadas para a compilação do *Corpus* de Artigos

1	Identificação dos 100 substantivos mais frequentes
	Elaboração das listas dos 100 primeiros substantivos de cada <i>corpus</i> , organizados por frequências através da ferramenta <i>WordList</i> ;
2	Identificação dos substantivos comuns às duas áreas
	Identificação dos substantivos comuns às duas áreas, totalizando 27, seguida da elaboração de uma tabela com estes substantivos mais frequentes do SCEs, CLL e <i>subcorpus</i> Brasileiro do <i>Portuguese Web 2011</i> (pt Ten Ten 11);
3	Identificação dos candidatos a colocações
	Identificação dos candidatos a colocações acadêmicas por meio das ferramentas <i>Word Sketch</i> e <i>Concordance</i> para identificar o contexto das ocorrências;
4	Identificação manual do número de colocações
	Identificação do número de colocações acadêmicas comuns nas duas áreas, para cada um dos 27 substantivos comuns;
5	Elaboração da lista de candidatas a colocações
	Após identificação do número de colocações encontradas e comuns nos dois <i>corpora</i> , elaboração de uma lista dessas possíveis colocações acadêmicas;
6	Análise das colocações selecionadas por frequência nos <i>Corpora</i>
	Análise do sentido das colocações a partir da análise do seu contexto de uso (nos exemplos), através da função <i>Concordance</i> ;
7	Proposta dos verbetes-piloto
	Desenvolvimento de um modelo para inserção na plataforma PLATCOL, com exemplos contextualizados e notas para a melhor compreensão quanto ao uso das colocações acadêmicas.

Fonte: Elaboração própria

Tendo apresentado os passos metodológicos desta pesquisa, trataremos, no item seguinte, da análise dos dados extraídos.

3 Análise dos Dados

A partir dos 27 substantivos encontrados em comum nas duas áreas da pesquisa, foi possível identificar 44 colocações adjetivas (doravante CAs) e 26 colocações verbais (doravante CVs) comuns aos dois *corpora*, como podemos observar no quadro abaixo.

Tabela 2: Número de colocações encontradas a partir dos 26 substantivos

SUBSTANTIVO	CA (BASE+COLOCADO)	CV (COLOCADO+BASE)
tempo	0	1
caso	3	0
sistema	2	0
processo	0	1
resultado	9	8
trabalho	2	0
função	0	1
forma	5	2
exemplo	3	1
estudo	2	1
dado	2	1
análise	5	2
tipo	0	0
relação	0	2
ponto	2	0
pesquisa	2	1
escola	3	0
vez	0	1
ano	1	1
grupo	2	0
produção	1	0
ideia	0	1
autor	0	0
uso	0	1
projeto	0	0
meio	0	0
condição	0	2
total de colocações	44	26

Fonte: Elaboração própria.

Em alguns casos, tivemos um número significativo de colocações, como aquelas geradas a partir dos substantivos: *resultado*, com nove CAs e 8 CVs, *forma*, com 5 CAs e, *análise*, com 5 CAs. Por outro lado, tivemos os itens lexicais: *autor*, *tipo*, *projeto* e *meio* sem nenhuma colocação em comum, em ambos *os corpora*, totalizando 70 colocações em comum quanto à forma, nas duas áreas acadêmicas, o que, para essa pesquisa, seria inviável. Precisávamos limitar esses dados para testar os procedimentos metodológicos deste estudo e, desse modo, antes de realizar esse recorte dos dados, elaboramos a lista de possíveis colocações, segundo mostra quadro abaixo:

Quadro 3: Parte da lista de candidatas a colocações acadêmicas comuns (CLL e SCEs)

	A	B	C	D
1	SUBSTANTIVO	CA (BASE+COLOCADO)	CV (COLOCADO+BASE)	
2	tempo	xxx	ter tempo	
3	caso	caso particular	xxx	
4		caso específico	xxx	
5		caso analisado	xxx	
6	sistema	sistema computacional	xxx	
7		sistema operacional	xxx	
8	processo	xxx	envolver processo	
9	resultado	resultado obtido	comparar resultado	
10		resultado alcançado	obter resultado	
11		resultado encontrado	apresentar resultado	
12		resultado final	produzir resultado	
13		resultado preciso	alcançar resultado	
14		resultado observado	encontrar resultados	
15		resultado apresentado	ter resultados	
16		resultado positivo	xxx	
17		resultado importante	xxx	
18	trabalho	trabalho realizado	xxx	
19		trabalho anterior	xxx	
20	função	xxx	assumir função	
21	forma	forma geral	dar forma	
22		forma possível	ter forma	
23		forma explícita	xxx	
24		forma sistemática	xxx	
25		forma simples	xxx	
26	exemplo	exemplo analisado	apresentar exemplo	
27		exemplo apresentado	xxx	
28		exemplo anterior	xxx	

Fonte: Elaboração própria.

Para cada substantivo comum, fizemos um *Word Sketch* em cada um dos *corpora*, comparamos os resultados e identificamos as colocações que ocorreram em ambos. Nos casos em que as combinações não pareciam ser colocações, verificamos o seu uso em contexto, usando a ferramenta *Concordance*. Por fim, inserimos as colocações comuns em um documento do *Excel*.

Após a elaboração da lista, realizamos o processo de recorte dos dados da análise, como mencionado na metodologia e, por questões de restrições de espaço e tempo, embora relevantes, em relação à determinação desse critério, foi necessário reduzir os dados analisados a uma quantidade com a qual podíamos trabalhar neste artigo, a fim de testar a metodologia adotada que vai além do fato de extrair as colocações acadêmicas, adicionando a extração, as análises e a proposta de verbetes-piloto.

Desse modo, selecionamos, para este artigo, por questões delimitativas, dois substantivos de maior frequência, a partir do SCEs, *tempo* e *caso* e, optamos por selecionar as colocações que eram mais frequentes nos *corpora*. Sendo assim, selecionamos as colocações *ter tempo* e *caso particular* que farão parte do DICA.

De forma a demonstrar nosso entendimento acerca da variação de sentidos das palavras em função das colocações em que ocorrem e das áreas de conhecimento, trazemos aqui *caso*. Por meio da ferramenta *Word Sketch*, buscamos verificar as colocações adjetivas (CA), substantivo+adjetivo (base+colocado), e colocações verbais (CV), verbo+substantivo (colocado+base).

Adotamos o critério de frequência dos substantivos para iniciar nossas análises. Assim, começamos com *tempo*, com 888 ocorrências, no SCEs, e 615, no CLL, e terminamos com *forma*, com 622 ocorrências, no SCEs, e 1135, no CLL, no entanto, para este artigo, com a finalidade de testar a metodologia adotada, realizaremos 2 análises.

Primeiramente, apresentamos uma análise envolvendo o substantivo *tempo*, que constitui o foco em uma CV, seguida de uma CA com o substantivo *caso*. Na figura a seguir, são exibidas todas as ocorrências do substantivo *tempo* encontradas nos dois *corpora* de pesquisa. Antes, contudo, apresentamos algumas acepções atribuídas a esse substantivo, a fim de evidenciar que seus sentidos variam independentemente da área temática. Para fins ilustrativos, seguem, de forma resumida, algumas definições extraídas do Dicionário Houaiss (2020).

- 1) duração relativa das coisas que cria no ser *humano* a ideia de presente, passado e futuro;
- 2) determinado período considerado em relação aos acontecimentos nele ocorridos;
- 3) certo período da vida que se distingue de outros;
- 4) *período específico*, segundo quem fala, de quem se fala ou sobre quem se fala;
- 5) época na qual se vive;
- 6) oportunidade

para a realização de alguma coisa; 7) período indefinido; 8) conjunto de condições meteorológicas; 9) época propícia para certos fenômenos ou atividades; 10) hora em local específico; 11) cada um dos períodos em que se dividem as partidas de determinados jogos; 12) duração cronometrada; 13) dimensão que permite identificar dois eventos que, caso contrário, seriam idênticos e que ocorrem no mesmo ponto do espaço; 14) categoria verbal que indica o momento em que se dá o fato expresso pelo verbo ou o tempo em que transcorrem; 15) subdivisão da categoria tempo; 16) unidade abstrata de medida do tempo musical, a partir da qual se estabelecem as relações rítmicas; 17) andamento no sentido de velocidade.

Figura 5: Dados obtidos a partir da ferramenta *Word Sketch* com o substantivo *tempo*.

substantivo+adjetivo		verbo+substantivo	
Letras/Linguística	Engenharias	Letras/Linguística	Engenharias
tempo + adjetivo	tempo + adjetivo	verbo + tempo	verbo + tempo
real 17 ...	computacional 56 ...	passar 8 ...	economizar 3 ...
presente 13 ...	médio 106 ...	ganhar 5 ...	modelar 3 ...
verbal 10 ...	hábil 3 ...	haver 21 ...	demandar 5 ...
lmemorial 2 ...	discreto 5 ...	correr 3 ...	minimizar 6 ...
circular 2 ...	máximo 14 ...	permanecer 3 ...	requerer 4 ...
linear 2 ...	contínuo 14 ...	significar 3 ...	reduzir 5 ...
remoto 2 ...	total 37 ...	percorrer 2 ...	representar 6 ...
mítico 2 ...	razoável 3 ...	demandar 2 ...	mostrar 4 ...
sensível 2 ...	aceitável 3 ...	levar 4 ...	Incluir 4 ...
inteiro 2 ...	necessário 14 ...	pôr 2 ...	considerar 4 ...
livre 2 ...	suficiente 3 ...	perder 2 ...	apresentar 4 ...
atual 2 ...	constante 3 ...	remeter 2 ...	ter 6 ...
distinto 2 ...	disponível 4 ...	viver 2 ...	
histórico 2 ...	mínimo 3 ...	surgir 2 ...	
contemporâneo 2 ...	inicial 4 ...	utilizar 2 ...	
específico 2 ...	diferente 3 ...	ocorrer 2 ...	
tempo + adjetivo participial	tempo + adjetivo participial	fazer 4 ...	
	gastar 16 ...	ter 5 ...	

Autor: *Sketch Engine*

No que se refere à figura 5, identificamos *ter tempo* como uma possível colocação verbal presente tanto no CLL como no SCEs. A fim de verificarmos se o sentido é o mesmo nos dois *corpora*, buscamos, através das linhas de concordâncias, o contexto em que ela ocorre:

Figura 6: *Concordance* do CLL em destaque CV *ter tempo*.



Fonte: Dados obtidos a partir da ferramenta *Concordance* do programa *Sketch Engine* (<https://www.sketchengine.co.uk/>)

No CLL, encontramos 4 linhas de concordância com o substantivo *tempo*. Já, pela visualização dos contextos um pouco mais ampliados, *terei tempo de comentá-la*, *quando tem bom tempo*, *enquanto tenho tempo* e a *construção tem percorrido o tempo e o espaço*, é possível observar que os sentidos de tempo variam. Ao analisarmos as linhas de concordância, constatamos que, na primeira e na terceira linhas de concordância, *tempo* se refere ao tempo cronológico. Já, na linha 2, *tempo* se refere à condição climática. Por fim, na linha 4, *tempo* tem o sentido de período específico, de duração. Trazemos a concordância contextualizada desse uso em um trecho do artigo do CLL:

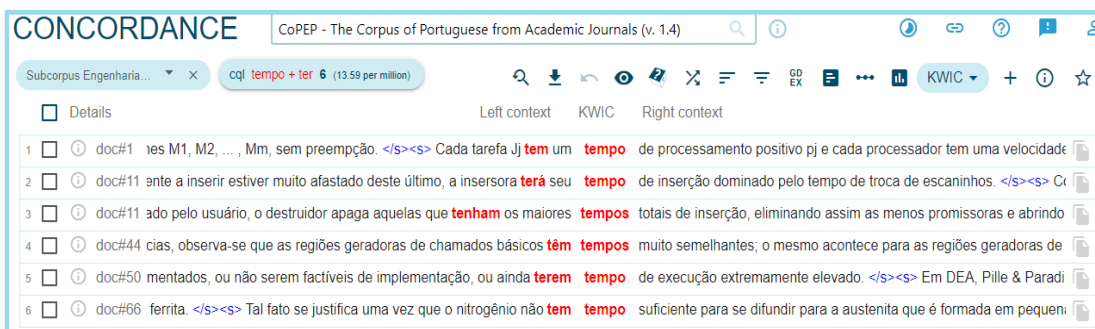
Figura 7: Concordance do CLL em destaque CV *ter tempo*

e mascarado, como indicam referência e predicação (esses preconceitos "mascarados"). </s><s> As referências (vergonha, receio) e a predicação (surpresos) indexalizam que o termo negro ainda é construído como ruim e ofensivo. </s><s> Podemos dizer que tal construção tem percorrido o tempo e diversos espaços, visto que como mencionado por [Munanga (1986) [HYPERLINK: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-18132015000100053&lng=pt&nrm=iso&tlng=pt]), desde os tempos da chamada ciência da raça, negras/os são qualificados como ruins, inferiores etc. A participante Eunice traz à tona a cristalização do problema racial no Brasil

Fonte: Dados obtidos a partir da ferramenta *Concordance* do programa *Sketch Engine*.

Em relação ao SCEs, temos, nas 6 linhas de concordâncias, as seguintes combinações: *tem um tempo*, *terá seu tempo*, *tenham os maiores tempos*, *têm tempos*, *terem tempo* e *tem tempo*.

Figura 8: Concordance do SCEs em destaque CV *ter tempo*



Fonte: Dados obtidos a partir da ferramenta *Concordance* do programa *Sketch Engine*

Podemos observar algumas palavras intermediárias e, o que elas fazem, é contribuir para o sentido da palavra *tempo*, ou mesmo da colocação como um todo. Assim, *ter tempo*, usado sem artigo indefinido, como na linha 6, indica que *tempo* é a duração ou velocidade, enquanto na linha 1, *tem um tempo*, a colocação possui um significado diferente, temos um período, um momento dentro do espaço, ou seja, presente, passado ou futuro. Isso tudo é fundamental para mostrar que, ainda que a forma seja a mesma, os sentidos são diferentes mesmo dentro de uma mesma área e, principalmente, em outras áreas.

No que diz respeito às CAs, podemos observar as seguintes ocorrências com o substantivo *caso*:

Figura 9: Dados obtidos a partir da ferramenta *Word Sketch* com o substantivo *caso*

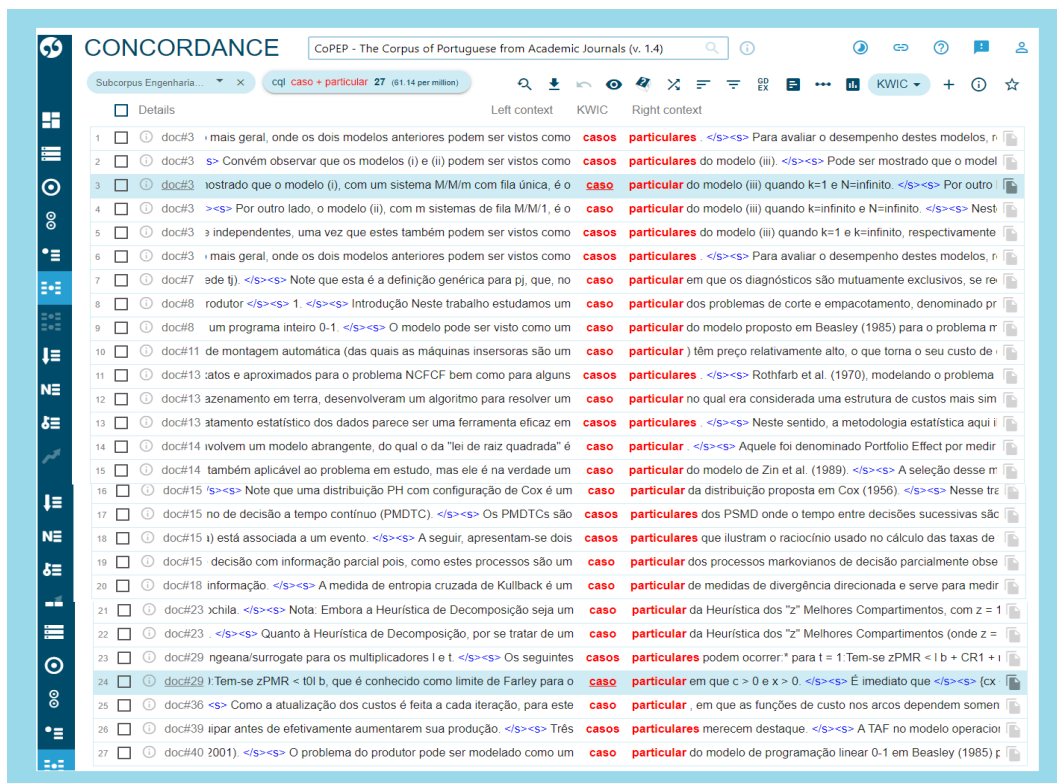
Colocação Adjetiva		Colocação Verbal	
substantivo+adjetivo		verbo+substantivo	
Letras/Linguística	Engenharias	Letras/Linguística	Engenharias
<p>caso + adjetivo</p> <p>vivo 5 ...</p> <p>particular 5 ...</p> <p>inexplicável 2 ...</p> <p>excepcional 2 ...</p> <p>específico 9 ...</p> <p>típico 2 ...</p> <p>brasileiro 2 ...</p> <p>caso + adjetivo participial</p> <p>discutir 4 ...</p> <p>narrar 2 ...</p> <p>analisar 2 ...</p>	<p>caso + adjetivo</p> <p>contrário 49 ...</p> <p>bidimensional 8 ...</p> <p>particular 27 ...</p> <p>multidimensional 4 ...</p> <p>prático 7 ...</p> <p>uniforme 2 ...</p> <p>crítico 2 ...</p> <p>específico 3 ...</p> <p>comum 2 ...</p> <p>caso + adjetivo participial</p> <p>univariado 4 ...</p> <p>bivariado 2 ...</p> <p>analisar 4 ...</p> <p>descrever 2 ...</p> <p>encontrar 2 ...</p> <p>realizar 3 ...</p>	<p>verbo + caso</p> <p>fossar 2 ...</p> <p>contar 3 ...</p> <p>ir 2 ...</p> <p>partir 3 ...</p> <p>haver 2 ...</p>	<p>verbo + caso</p> <p>exemplificar 2 ...</p> <p>resolver 4 ...</p> <p>estudar 2 ...</p> <p>considerar 5 ...</p> <p>analisar 2 ...</p> <p>haver 2 ...</p>

Fonte: Sketch Engine.

Como podemos observar na figura 9, há três colocações que ocorrem nas duas áreas com o substantivo *caso* com a estrutura substantivo **base** + adjetivo **colocado**, quais sejam, *caso particular*, *caso específico* e *caso analisado*. Todavia, analisaremos a de maior frequência: *caso particular*, com cinco ocorrências no CLL e 27 ocorrências no SCEs.

Desse modo, damos continuidade à nossa análise, buscando identificar os aspectos semânticos dos candidatos a colocações, a partir das linhas de concordâncias, conforme pode ser observado no SCEs.

Figura 10: Dados obtidos a partir da ferramenta *Concordance do Sketch Engine*



Fonte: Sketch Engine.

Optamos por analisar a *CA caso particular*, em ambas as áreas, examinando o conteúdo de cada linha de concordância e, quando necessário, em um contexto maior.

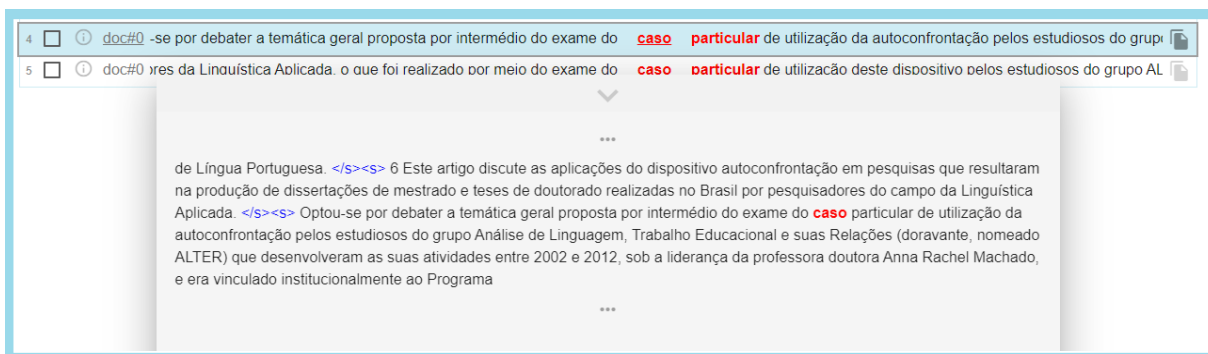
Podemos perceber na construção *vistos como casos particulares* que temos um complemento: *vistos como*. Nas linhas de concordâncias um, dois, três, quatro, cinco, seis e nove, *caso particular* assume o sentido de diferença, ou seja, fora dos aspectos

padronizados, o que nos apresenta uma informação importante que devemos destacar: se observarmos as linhas de concordâncias, veremos que grande parte das colocações vêm acompanhadas de um substantivo, mostrando as particularidades desses casos específicos.

Podemos perceber também algumas características de unicidade, de descoberta. Nas linhas um e dois, há a distinção de vários *casos particulares* e, na linha três, temos *desenvolver o caso particular*, especificidade. Na linha quatro, ampliamos a linha de concordância do texto e verificamos que, nessa situação, os sentidos de *caso* também são mantidos em ambos os contextos.

Nas linhas 7 até a 27, catalogamos *no caso particular*, *um caso particular*, *como casos particulares*, *alguns casos particulares* e, com isso, podemos identificar características de especificidades. Concluímos que, no SCEs, a colocação *caso particular* tem dois sentidos que caminham muito próximos de “diferente” e de “especificidade”.

Figura 11 – Linhas de Concordância da Colocação *Caso Particular* no *Corpus CLL*



Fonte: dados obtidos a partir da ferramenta *Concordance* do *Sketch Engine*.

Na figura 11, temos duas ocorrências da colocação *caso particular* no CLL, o que parece se tratar de uma colocação especializada, pois em ambos os casos temos o substantivo *exame* que vem somar à colocação *exame do caso particular*, cuja pesquisa diretamente no Google nos trouxe 15.900 ocorrências, em sua maioria nas áreas jurídica, da Linguística e da Filosofia, mostrando se tratar de um processo, uma análise minuciosa sobre determinado assunto.

No que diz respeito aos exemplos, quando houver muitas ocorrências com sentido semântico diferentes, iremos enumerá-los e, se houver necessidade de informação adicional, incluiremos uma nota de entrada, seguindo as orientações dos lexicógrafos.

É possível perceber que a plataforma é desenvolvida a partir de elementos pedagógicos, não dando apenas a colocação pronta ao usuário da plataforma, pelo contrário, insere de modo simples e, de certa forma, didático, ao tripé teórico que foi a base para o seu desenvolvimento: a Fraseologia, a Linguística de *Corpus* e a Lexicografia.

Conclusão

A proposta deste estudo buscou extrair e analisar, do ponto de vista léxico-semântico, as colocações acadêmicas presentes nos dois *corpora* de pesquisa, sendo um da área de Letras/Linguística e, outro, de Engenharias, a partir de artigos do SciELO.

Para desenvolver a pesquisa, buscamos, por meio da testagem dos procedimentos metodológicos, encontrar os substantivos comuns nas duas áreas e, em seguida, verificar a existência de colocações com os mesmos substantivos. Encontradas e extraídas essas colocações, verificamos, no âmbito deste trabalho, como cada uma se comportava dentro de cada área acadêmica, com o auxílio do programa *Sketch Engine* (Kilgarrif, et al, 2004).

A pesquisa mostrou que a metodologia adotada para as análises já apresentou dados relevantes, a fim de que possamos dar continuidade na busca por padrões colocacionais nas áreas selecionadas deste artigo, bem como para os demais tipos de colocações existentes. A análise das colocações acadêmicas mostrou-se importante para sua inserção em um dicionário de colocações acadêmicas que já se encontra em desenvolvimento, o DICA.

Pretende-se, assim, dar continuidade ao processo de extração de dados, com o objetivo de ampliar o conjunto de colocações a serem investigadas e, conseqüentemente, contribuir para o aprimoramento e enriquecimento da plataforma DICA. Almeja-se aplicar a metodologia testada nas colocações das áreas de Letras/Linguística e

Engenharias a outros campos do conhecimento, bem como estendê-la aos demais tipos de colocações identificados, de acordo com a taxonomia adotada neste estudo.

REFERÊNCIAS

ALVES, M. F.; MOURA, L. O. B. M. A escrita de artigo acadêmico na universidade: autoria x plágio. **Ilha Desterro**, Florianópolis, v. 69, n. 3, p. 77-93, dez. 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-80262016000300077&lng=en&nrm=iso. Acesso em: 23 jun. 2025.

BIBER, D. **University Language: a corpus-based study of spoken and written registers**. Philadelphia: John Benjamins, 2006. DOI: <https://doi.org/10.1075/scl.23>

DICA – **Plataforma de Dicionários de Colocações Acadêmicas**. Disponível em: <https://www.godica.com.br/>. Acesso em: 04 fev. 2025.

Dicionário Eletrônico Houaiss da Língua Portuguesa (2020). CD-ROM. Editora Objetiva Limitada.

FILLMORE, C. Innocence: a second idealization for linguistics. In: BERKELEY LINGUISTICS SOCIETY, 5., 1979, Berkeley. **Proceedings of the 5th Annual Meeting of the Berkeley Linguistics Society**. Berkeley: [s. n.], p. 63-76, 1979.

FISCHER, A. **A construção de letramentos na esfera acadêmica**. 2007. 341 f. Tese (Doutorado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2007.

KILGARRIFF, A. et al. The Sketch Engine. In: EURALEX INTERNATIONAL CONGRESS, 11., 2004, Lorient. **Proceedings**. Lorient: Université de Bretagne-Sud, pp. 105-116, 2004.

KUHN, T. Z. **A Design Proposal of an On-line Corpus-Driven Dictionary of Portuguese for University Students**. 2017. 421 f. Tese (Doutorado em Estudos Linguísticos) – Universidade de Lisboa – Faculdade de Letras, Lisboa, 2017.

KUHN, T. Z.; FERREIRA, J. P. O Corpus de Português Escrito em Periódicos - CoPEP. **DELTA**, São Paulo, v. 36, n. 2, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502020000200408&lng=pt&nrm=iso. Acesso em: 23 set. 2020.

ORENHA-OTTAIANO, A. **A compilação de um glossário bilíngue de colocações, na área de jornalismo de negócios, baseado em corpus comparável**. 2004. 246 f. Dissertação (Mestrado em Estudos Linguísticos e Literários) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

ORENHA-OTTAIANO, A. Collocations workbook: um material de apoio pedagógico on-line e baseado em corpus para o ensino de colocações em inglês. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 23, n. 3, pp. 833-881, 2015.

ORENHA-OTTAIANO, A. The compilation of an online corpus-based bilingual collocations dictionary. In: CORPAS PASTOR, G. (org.). **Computerised and corpus-based approaches to phraseology: monolingual and multilingual perspectives**. Genebra: Editions Tradulex, v. 1, pp. 486-493, 2016.

ORENHA-OTTAIANO, A. **A phraseographical methodology and model for an online corpus-based multilingual collocations dictionary platform**. 2020. Projeto pela Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo FAPESP nº 2020/01783-2).

ORENHA-OTTAIANO, A. The creation of an On-line English Collocations Platform to help develop collocational competence. **Phrasis. Rivista di studi fraseologici e paremiologici**, 2020.

ORENHA-OTTAIANO, A. **Tela de Abertura da Plataforma On-line de Dicionários de Colocações**. Disponível em: <http://www.institucional.grupogbd.com/dicionario/sobre>. Acesso em: 29 maio 2025.

SOARES, M. Novas práticas de leitura e escrita: letramento na cibercultura. *Educação & Sociedade*, Campinas, v. 23, n. 81, p. 143-160, dez. 2002. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-73302002008100008&lng=pt&nrm=iso. Acesso em: 23 jun. 2025.

TAGNIN, S. E. O. **Expressões idiomáticas e convencionais**. São Paulo: Ática, 1989.