



**Uma proposta de coextensividade entre termo técnico,  
grupo nominal e item lexical no português brasileiro:  
um estudo com base em ferramentas da linguística de corpus  
sob o arcabouço de teoria sistêmico-funcional**

*A proposal of coextensiveness between technical term,  
nominal group, and lexical item in Brazilian Portuguese:  
a study based on corpus linguistics' software within the  
framework of systemic-functional theory*

Júlia Santos Nunes Rodrigues

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil  
juliasnrodrigues@ufmg.br

<http://orcid.org/0000-0002-7673-1833>

Kícila Ferregueti

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil  
Kfo2008@ufmg.br

<http://orcid.org/0000-0002-1919-0073>

Adriana S. Pagano

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil  
apagano@ufmg.br

<http://orcid.org/0000-0002-3150-3503>

**Resumo:** Objetivo: Sob a perspectiva do trabalho de Pearson (1998) e utilizando as ferramentas da linguística de corpus disponíveis para o português brasileiro, a pesquisa apresentada neste artigo busca verificar em que medida a coextensividade entre termo técnico, grupo nominal e item lexical pode ser considerada válida para o português brasileiro. Método: Para a verificação desta coextensividade, um corpus de artigos acadêmicos sobre o domínio experiencial do autocuidado em Diabetes Mellitus foi compilado eletronicamente. Esse corpus foi inserido no software concordanciador AntConc (ANTHONY, 2019) e três palavras-chave foram extraídas com base no

corpus de referência CALIBRA (FIGUEREDO; PAGANO; FERREGUETTI, 2014). O contexto de cada uma dessas palavras foi analisado por meio da ferramenta clusters/n-grams do AntConc, considerando os clusters/n-grams com o número mínimo de dez ocorrências. Resultados: A investigação desses clusters/n-grams formados à direita e à esquerda de cada uma das palavras-chave selecionadas mostrou que a coextensividade entre termo técnico, grupo nominal e item lexical nem sempre pode ser identificada, embora a utilização da ferramenta de cluster/n-grams pode ser considerada eficaz para buscar por itens lexicais que estão em coextensividade à ordem do grupo nominal, em razão da existência de pelo menos um grupo nominal em cada cluster/n-gram analisado. Conclusão: Ainda que os programas utilizados em Pearson (1998) não estejam totalmente difundidos para o português brasileiro, a abordagem sistêmico-funcional para o grupo nominal e para o item lexical em conjunto com as ferramentas do software concordanciador utilizado se mostraram eficientes para a análise proposta neste artigo.

**Palavras-chave:** item lexical; grupo nominal; termo técnico; linguística de corpus; sistêmico-funcional; co-extensividade.

**Abstract:** Objective: Drawing on Pearson (1998) and using corpus linguistics tools available for Brazilian Portuguese, we report on a study aimed at exploring to what extent the concepts of technical term, nominal group and lexical item are coextensive in Brazilian Portuguese. Method: A corpus of academic articles on the experiential domain of Diabetes Mellitus self-care was compiled and queried in AntConc, a concordancing software (ANTHONY, 2019). Using as a reference corpus CALIBRA (FIGUEREDO; PAGANO; FERREGUETTI, 2014), three keywords were extracted analysed with AntConc tool clusters/n-grams, considering clusters/n-grams with a minimum number of ten occurrences. Results: Analysis of clusters/n-grams to the right and left of each of the selected keywords showed that technical term, nominal group and lexical item cannot always coextensive. The use of cluster/n-grams tool can be considered effective to search for lexical items that are coextensive to the order of the nominal group, due to the existence of at least one nominal group in each cluster/n-gram analyzed. Conclusion: Although the programs used by Pearson (1998) are not fully available to Brazilian Portuguese, a systemic-functional approach to nominal group and lexical item together with the tools of the concordancing software used in this paper proved to be efficient for the analysis herein proposed.

**Keywords:** lexical item; nominal group; technical term; corpus linguistics; systemic functional theory; coextensiveness.

Recebido em 10 de outubro de 2020

Aceito em 07 de janeiro de 2021

## **1 Introdução**

O registro de uso de corpus é antigo. Na Grécia Antiga, foi criado o Corpus Helenístico de Alexandre, o Grande. Na Antiguidade e Idade Média, os corpora com citações da Bíblia foram desenvolvidos. No século XX, muitos pesquisadores utilizavam os corpora para trabalhos de descrição da linguagem. No entanto, apesar desses registros, o uso de corpora nessas fases era restrito ao aprendizado de línguas, sendo que todo o processo de elaboração dos corpora era feito de forma manual em razão da inexistência de recursos tecnológicos (SARDINHA, 2004, p. 3).

Os corpora como conhecemos hoje, com milhões de palavras, com textos compilados, majoritariamente, de forma automática, construídos para suprir diversas demandas linguísticas e sendo viabilizados por softwares de diferentes origens tem início com o lançamento do corpus Brown (Brown Corpus of Standard American English) no início da década de 1960 (SARDINHA, 2004). Desde essa época até os dias atuais, a Linguística de Corpus tem evoluído muito, principalmente a partir do uso de computadores pessoais nos anos de 1980.

Hoje em dia já existem corpora compilados para uma grande variedade de línguas, corpora utilizados para diferentes finalidades, como, tradução, criação de dicionários e gramáticas, processamento de linguagem natural, terminologia, etc. Há também softwares livres, como o AntConc (ANTHONY, 2019), por exemplo, que auxiliam pesquisadores no desenvolvimento de estudos que utilizam corpora. Contudo, a evolução tecnológica no âmbito da Linguística de Corpus não pode ser vista de forma homogênea para todas as línguas. No âmbito da língua inglesa, por exemplo, sobretudo no contexto britânico, a Linguística de Corpus tem uma disponibilidade maior de recursos tecnológicos em razão de investimentos financeiros em pesquisas dessa área, bem como pelo fato dos estudos de corpora, como conhecemos hoje, terem iniciado no contexto dessa língua. Esse investimento em tecnologia pode ser observado com clareza no trabalho de Pearson (1998), que serviu de base para o estudo do presente artigo, haja vista o desenvolvimento de um anotador morfossintático (CLG tagger) e de um programa de padrão de correspondência para aquele trabalho.

Apesar dessas limitações, este artigo tem como principal objetivo apresentar soluções que podem ser utilizadas como forma de diminuir o abismo existente entre os recursos tecnológicos disponíveis para o

contexto da língua inglesa e ainda incipientes para o português brasileiro. Essas soluções baseiam-se ora em ferramentas da própria Linguística de Corpus ora no suporte de outras teorias linguísticas, no caso, a teoria sistêmico-funcional (FERREGUETTI, 2018; FIGUEREDO, 2007; HALLIDAY; MATTHIESSEN, 2014).

Nesse sentido, a pesquisa detalhada no presente artigo faz uso de um corpus de artigos acadêmicos sobre autocuidado em Diabetes Mellitus, escritos em português brasileiro, para extração de termos técnicos dessa área do conhecimento, por meio das ferramentas de lista de palavras-chave (Keyword List), clusters e n-grams (Clusters/N-grams) e lista de concordância (Concordance) do software concordanciador AntConc (ANTHONY, 2019). As análises do trabalho em questão são pautadas na teoria sistêmico-funcional principalmente na propriedade estabelecida por Halliday (2002, p. 59-60) denominada co-extensividade (coextensiveness), a qual pode ocorrer entre o item lexical e a escala de ordens da gramática – morfema, palavra, grupo/frase preposicional e oração – sobretudo no que se refere às ordens da palavra e do grupo/frase preposicional (FERREGUETTI, 2018; FIGUEREDO, 2007).

Partindo dessa propriedade, espera-se verificar em que medida a correspondência entre termo técnico, grupo nominal e item lexical pode ser considerada válida para o português brasileiro. Em paralelo à verificação dessa hipótese, soluções de como os recursos tecnológicos desenvolvidos para/pela a linguística de corpus alinhados ao arcabouço linguístico da teoria sistêmico-funcional são testadas a fim de suprimir recursos da linguística de corpus difundidos para a língua inglesa (cf. PEARSON, 1998) e pouco (ou nada) disseminados para o português brasileiro.

## **2 Fundamentação teórica**

### **2.1 A Linguística de Corpus e o estudo de termos técnicos**

Dentre as diferentes vertentes linguísticas, a Linguística de Corpus e a Terminologia são conhecidas como subáreas da linguística cujos objetivos estão relacionados ao estudo de termos técnicos sob diferentes abordagens. Na obra “Linguística de Corpus”, Sardinha (2004) afirma que:

A Linguística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (SARDINHA, 2004, p. 3).

Sob essa perspectiva, a Linguística de Corpus lida com termos técnicos em seu contexto e cotexto de uso, ocupando-se com o desenvolvimento de ferramentas que consigam examinar uma grande quantidade de textos de maneira automática (ALMEIDA; CORREIA, 2008).

Por outro lado, a Terminologia, a partir do século XVII, pode ser entendida sob dois vieses diferentes: i. “Como conjunto de termos de uma área técnica ou científica” ou ii. “Como disciplina de natureza linguística que estuda esse conjunto de termos” (ALMEIDA, 2004, p. 31). Cada vez mais as teorias de Terminologia entendem o texto especializado de uma determinada área do conhecimento como seu principal ponto de estudo, sendo o termo técnico compreendido como:

uma condição especial da palavra, um signo linguístico dotado de significado e significante, e atrelado a uma determinada unidade e corpo de conhecimentos historicamente estabelecidos. Desse modo, terminologias deixam de ser unidades “estranhas” ou “artificiais”, índices de uma língua à parte da língua-sistema, e passam a ser vistas como palavras que têm ou adquirem um estatuto peculiar em uma dada situação de comunicação (FINATTO, 2007, p. 224).

Apesar da Linguística de Corpus e da Terminologia serem subáreas da linguística, cada uma apresenta uma série de particularidades que inviabiliza a sobreposição de uma subárea sobre a outra. Uma das implicações diz respeito ao termo técnico que para a Linguística de Corpus precisa apresentar uma determinada frequência para ser considerado um possível termo técnico, bem como necessita ter seu contexto analisado para certificar-se que se trata de um termo técnico, ao passo que para a Terminologia a frequência de um dado termo não compreende um critério de seleção para que esse termo seja entendido como técnico (FINATTO, 2007). Outra implicação compreende a elaboração dos corpora compilados para os estudos terminológicos, visto

que a extração de termos considerados candidatos a técnicos é diretamente influenciada pela forma como os textos dos corpora são manipulados pelos pesquisadores.

Do ponto de vista da complementariedade, no entanto, os recentes avanços tecnológicos relacionados à Linguística de Corpus se mostram relevantes também para o progresso da Terminologia enquanto disciplina. O desenvolvimento e o aprimoramento de ferramentas computacionais podem viabilizar a gestão e manipulação de grandes bases textuais, beneficiando as pesquisas terminológicas baseadas em corpora (ALMEIDA; CORREIA, 2008).

Ainda no tange a complementariedade entre essas duas subáreas, está a distinção conceitual entre palavra e termo, uma vez que os manuais de Terminologia designam termo como aquilo que geralmente é caracterizado como palavra. Contudo, por meio da análise de corpora, o pesquisador consegue ter acesso ao contexto e cotexto em que determinado termo aparece, o que pode favorecer a seleção desse termo como candidato a termo técnico (ALMEIDA; CORREIA, 2008).

A subseção seguinte apresenta o trabalho que norteou a pesquisa abordada no presente artigo. Esse estudo deixa claro como os avanços tecnológicos da Linguística de Corpus podem beneficiar diretamente as pesquisas na área de Terminologia.

## **2.2 O trabalho de Pearson (1998)**

A pesquisa desenvolvida por Jennifer Pearson publicada no livro *Terms in Context* de 1998 teve como um dos principais objetivos estabelecer parâmetros metodológicos que fossem capazes de auxiliarem terminologistas, lexicógrafos e linguistas de corpus a lidar com termos técnicos de uma forma mais automatizada. Para além disso, a autora explica conceitos técnicos inerentes à Linguística de Corpus, estabelece as diferenças entre palavra e termo técnico e apresenta um panorama de metodologias utilizadas por outras pesquisas que também enfocaram na definição e extração de termos.

Três corpora com domínios distintos foram utilizados nesse trabalho. O primeiro chamado de “Nature corpus”, com 230.000 palavras, lida com o uso especializado da linguagem, visto que envolve artigos acadêmicos publicados no periódico *Nature* ao longo do ano de 1989. O segundo denominado “International Telecommunications Union (ITU)

corpus”, com 4.7 milhões de palavras, engloba os textos do manual da União Internacional de Telecomunicações da Europa (International Telecommunications Union CCITT Handbook) conhecido como “The Blue Book”, escrito por membros do ITU para aprendizes da área. O terceiro nomeado como “GCSE corpus”, com um milhão de palavras, compreende uma série de livros didáticos sobre história, geografia, biologia, química, sociologia e política, cujo intuito é promover o ensino de disciplinas do currículo escolar.

Encerrado o processo de compilação desses corpora, um anotador morfossintático, CLG tagger, desenvolvido pelo grupo de Linguística de Corpus da Universidade de Birmingham, foi utilizado para gerar uma anotação automática das classes de palavra de cada um dos três corpora.

Em seguida, uma busca manual por sinais linguísticos, como, “i.e.” ou “e.g.” foi realizada em cada um dos corpora. Isso se deu pela hipótese levantada sobre a possibilidade de coocorrência de termos técnicos com alguns sinais linguísticos específicos.

Essa busca manual por padrões de coocorrência de termos técnicos e sinais linguísticos resultou em conjuntos de padrões que indicavam como os termos técnicos eram formados para cada um dos três corpora. Esses padrões eram formados por meio de uma sequência de etiquetas, como, por exemplo, adj + noun + noun. Esses conjuntos de padrões foram inseridos em um programa de padrão de correspondência que fora treinado para selecionar nos textos de cada corpus palavras que estavam em concordância com aqueles padrões obtidos manualmente, ou seja, o programa de padrão de correspondência rastreava os textos de cada corpus e extraía todos os conjuntos de palavras que se encaixavam nos *inputs* dados pelos pesquisadores a partir dos padrões obtidos manualmente. Então, esses padrões gerados automaticamente pelo programa de padrão de correspondência eram concebidos pelos pesquisadores como possíveis termos técnicos.

O próximo passo dessa pesquisa foi a investigação desses possíveis termos técnicos com o propósito de refinar quais deles poderiam ser considerados termos técnicos de fato. Para isso, os pesquisadores responsáveis lançaram mão das seguintes estratégias:

**Referência genérica:** a presença ou ausência de referência endofórica no cotexto em que determinado termo aparece pode levar esse termo a ser considerado genérico ou individual, neste último caso, o termo poderia ser realizado pelo próprio nome, não por um termo qualquer.

**Termo sinalizado:** o possível termo técnico não pode ser precedido por uma série de determinantes.

**Termo não-sinalizado:** o possível termo técnico pode não ser precedido por um determinante nem ser precedido apenas por artigos indefinidos. No entanto, somente essa estratégia não é suficiente para definir se um dado termo é técnico de fato. Tal estratégia só é considerada suficiente se o termo em questão está em coocorrência com algum dos sinais linguísticos a seguir:

- i. “por exemplo”, “ou seja”;
- ii. “chamado de”, “conhecido como”, “denominado por”, “também chamado de”, “geralmente conhecido como” e
- iii. Obedecer à construção: “possível termo técnico” seguido de “o termo” ou “esse processo” ou “esse método” ou “esse instrumento”, etc.

Por fim, Pearson detalha os resultados encontrados para cada um dos três corpora, bem como aponta dificuldades e limitações relacionadas à pesquisa de extração de termos técnicos em corpora especializados.

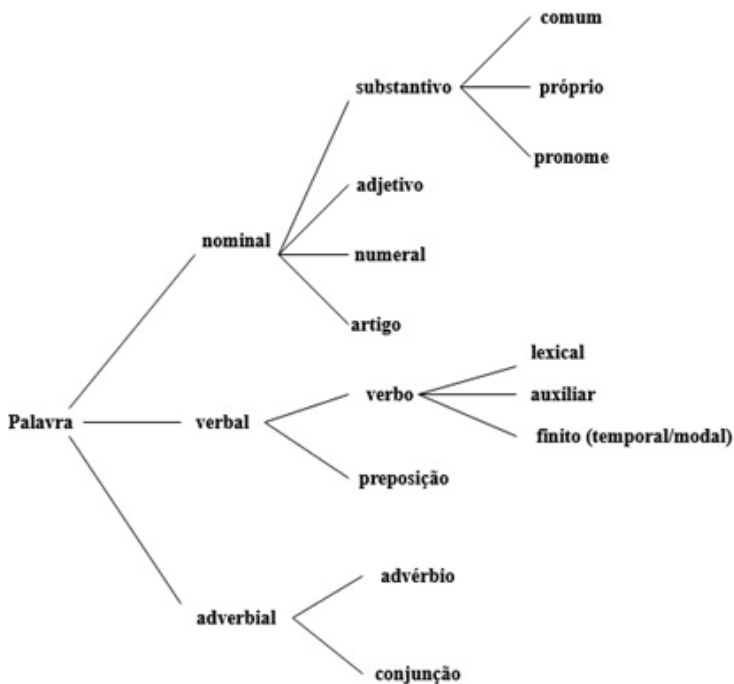
### 2.3 O termo técnico realizado por grupo nominal sob a perspectiva da TSF

No âmbito da teoria sistêmico-funcional (TSF), a gramática é organizada em uma escala de ordens hierárquicas em que cada ordem é formada pelas ordens que estão imediatamente abaixo. O português brasileiro conta com quatro ordens: oração, grupo, palavra e morfema. Nesse sentido, a oração é constituída por grupos, que por sua vez, são constituídos por palavras, que são constituídas por morfemas (cf. FERREGUETTI, 2018; FIGUEREDO, 2007).

Ainda com relação ao sistema linguístico do português, em um primeiro nível de delicadeza (*delicacy*), há grupos de cinco tipos: nominal, verbal, adverbial, conjuntiva e preposicional (FIGUEREDO, 2007). Isso significa dizer que para cada grupo as palavras que os constituem são, geralmente, da classe correspondente, ou seja, as palavras que constituem o grupo nominal, por exemplo, são, na sua maioria, da classe de palavras nominal. A Figura 1 apresenta a distribuição de classes de palavra de acordo com as concepções da TSF.



FIGURA 1 – Distribuição das classes de palavra segundo a TSF



Fonte: Traduzido e Adaptado de Halliday e Matthiessen (2014, p. 75).

A Figura 1 identifica que as três principais classes de palavra são nominal, verbal e adverbial. As palavras nominais compreendem os substantivos (comum, próprio e/ou pronome), adjetivos, numerais e/ou artigos. As palavras verbais englobam os verbos (lexical, auxiliar e/ou finito) e as preposições. E as palavras adverbiais envolvem os advérbios e as conjunções.

No que tange o grupo nominal, objeto de discussão desta seção, Halliday e Matthiessen (1999) afirmam que o grupo nominal apresenta duas funções primárias denominadas como Qualidade e Ente. Essas funções são responsáveis pela ideia de permanência que é intrínseca ao grupo nominal. Em outras palavras, os elementos permanentes são capazes de se repetirem ao longo do texto, por isso são mais duradouros e podem participar de eventos distintos. Em contrapartida, os elementos transitórios frequentemente realizados pelo grupo verbal representam os eventos do texto (FIGUEREDO, 2007).

No que diz respeito à estrutura do grupo nominal em português brasileiro, as funções de Ente e de Qualidade são as duas principais. O Ente representa, linguisticamente, os seres do mundo, é o núcleo semântico do grupo nominal e pode ser realizado pelos substantivos (comum, próprio ou pronome); e a Qualidade é responsável por definir qual o subconjunto de seres o escritor ou falante daquele texto se refere. Para além dessas duas funções, o grupo nominal pode contar ainda com as funções do Dêitico, Numerativo, Epíteto, Classificador e Qualificador. O Dêitico determina um subconjunto do Ente, sendo que essa determinação pode ser em relação à definição, especificidade e/ou localização. O Numerativo aponta alguma característica numérica ao subconjunto do Ente. O Epíteto indica alguma qualidade do Ente, a qual pode englobar traços objetivos do subconjunto do Ente, bem como alguma avaliação do falante. O Classificador envolve a relação de hiponímia e é responsável por delimitar o Ente em relação a uma subclasse. O Qualificador tem a função de caracterizar o Ente do grupo nominal por meio de frase preposicional,<sup>1</sup> oração encaixada ou oração não-finita (FIGUEREDO, 2007; HALLIDAY; MATTHIESSEN, 2014). O Quadro 1 mostra as classes de palavra mais prováveis para cada uma das funções do grupo nominal.

QUADRO 1 – Funções do grupo nominal em PB e as respectivas classes de palavra mais frequentes

Função	Classes de palavra mais provável		Outras possibilidades de ocorrência	
<b>Ente</b>	substantivo	pronome pessoal	verbo (fenômeno)	
<b>Dêitico</b>	artigo	pronome		
<b>Numerativo</b>	numeral		pronome	
<b>Epíteto</b>	adjetivo		verbo	
<b>Classificador</b>	adjetivo	substantivo	pronome	numeral

Fonte: Figueredo (2007, p. 226).

<sup>1</sup> De acordo com a Teoria Sistêmico-funcional, a frase preposicional corresponde à ordem localizada entre a palavra e oração na escala de ordem da gramática. Ela é constituída por uma preposição + um grupo nominal (cf. FERREGUETTI, 2018; HALLIDAY; MATTHIESSEN, 2014).

Os dados do Quadro 1 revelam que para todas as funções do grupo nominal as palavras relacionadas à classe nominal são mais prováveis de aparecer nos grupos nominais em português brasileiro. É importante ressaltar que os verbos são as palavras localizadas fora da classe nominal que apresentam ocorrências no grupo nominal. Isso pode ser explicado pelas metáforas gramaticais, frequentes em determinados tipos de texto (cf. *grammatical metaphor* – HALLIDAY; MATTHIESSEN, 2014).

Tendo detalhado os principais pontos teóricos relacionados ao grupo nominal em português brasileiro, a correspondência entre grupo nominal e termo técnico pode ser entendida por meio da perspectiva trinocular (de baixo, de cima e ao redor) inerente à teoria sistêmico-funcional. A análise do termo técnico pela perspectiva de baixo revela que termos técnicos são constituídos por palavras da classe nominal e são frequentemente realizados por grupos nominais, que operam na ordem da oração como Participantes. A análise pela perspectiva de cima indica que os termos técnicos funcionam semanticamente como participantes. E a análise pela perspectiva ao redor aponta que os termos técnicos podem ser co-extensivos (coextensiveness) ao grupo nominal, ou seja, os termos técnicos podem compartilhar de funções inerentes ao grupo nominal, como, Ente e Qualidade, por exemplo (cf. HALLIDAY, 1961).

### **3 Metodologia**

A metodologia deste artigo é dividida em duas partes. A primeira parte aborda a compilação do corpus utilizado, bem como descreve o corpus que serviu como corpus de referência para a presente pesquisa. A segunda explica como se deu a adaptação do trabalho de Pearson (1998) para o português brasileiro, uma vez que o presente estudo foi baseado na obra *Terms in Context* de Jennifer Pearson.

#### **3.1 O corpus**

Os textos selecionados para constituírem o corpus utilizado nesta pesquisa compreendem artigos acadêmicos que retratam pesquisas desenvolvidas no âmbito do Diabetes Mellitus tipo II. O domínio experiencial (HALLIDAY; MATTHIESSEN, 1999; HAO, 2015) desses

textos é, portanto, essa condição crônica, especificamente, o autocuidado em Diabetes Mellitus.<sup>2</sup>

Esses artigos acadêmicos foram extraídos de publicações da área das Ciências da Saúde a partir da palavra-chave “autocuidado em diabetes mellitus” digitada na aba de busca do Google Acadêmico.<sup>3</sup> Foram coletados 40 artigos originalmente escritos em português brasileiro, publicados entre 2010 e 2019, gerando, ao todo, 133.232 *tokens*, sendo que a seleção por artigos desse período de tempo é justificada pela necessidade de textos atuais, capazes de retratar padrões produzidos na última década.

Cada artigo foi salvo separadamente em um arquivo no bloco de notas e nominado com o Título da Revista\_ano de publicação, sendo que, para dois ou mais artigos de um determinado periódico publicados no mesmo ano, a nomeação se deu da seguinte forma Título da Revista\_ano de publicaçãoob.

É importante destacar que os *abstracts*, quadros, gráficos, anexos, figuras, tabelas e referências bibliográficas de todos os artigos selecionados para o corpus foram excluídos quando cada artigo foi transferido para o respectivo arquivo do bloco de notas, dado que a utilização desse material poderia gerar um corpus poluído, sem contribuições significativas para os resultados do presente estudo.

### 3.2 O corpus de referência

Para que a lista de palavras-chave pudesse ser gerada, recorreu-se à parte escrita monológica do corpus CALIBRA (Catálogo da Língua Brasileira) como corpus de referência. A seleção por essa parte do CALIBRA é justificada pelo fato do corpus utilizado para a análise da

---

<sup>2</sup> É importante mencionar que este artigo faz parte de uma pesquisa de doutorado em desenvolvimento pela primeira autora deste artigo, a partir do apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Tal pesquisa está localizada na área de Estudos Linguísticos no escopo do Programa de Pós-graduação em Estudos Linguísticos da Faculdade de Letras da UFMG e também no âmbito do Projeto Empoder@ – Protótipo conceitual e metodológico para avaliação de intervenções orientadas ao autocuidado em diabetes, uma parceria entre o Laboratório Experimental de Tradução (LETRA) da FALE/UFMG, a Escola de Enfermagem da UFMG e o Departamento de Estatística do ICEx/UFMG.

<sup>3</sup> Disponível em: <https://scholar.google.com.br/?hl=pt>.

presente pesquisa ser formado por artigos acadêmicos, um tipo de texto escrito e monológico, tal como os textos do corpus de referência. Essa similaridade entre os textos do corpus de análise e os textos do corpus de referência, mesmo que mínima, pode garantir a coerência dos resultados obtidos para o presente artigo.

O CALIBRA compreende um corpus com cerca de um milhão de palavras (*tokens*), compilado de acordo com a tipologia do contexto de cultura (HALLIDAY, 1978). Essa tipologia é determinada por meio de cinco variáveis: i. Especialização (especializado/não especializado); ii. Papel da língua na situação (constitutivo/auxiliar); iii. Modo de produção (escrito/falado); iv. Modo de interação (monólogo/falado); v. Processo sociossemiótico (explorar/compartilhar/explicar/relatar/recriar/fazer/recomendar/habilitar). O Quadro 2 resume todos os tipos de texto presentes no CALIBRA e destaca aqueles que foram utilizados como corpus de referência nesta pesquisa.

QUADRO 2 – CALIBRA – Distribuição dos tipos de texto segundo o contexto de cultura

	PRODUÇÃO		escrito		falado	
	INTERAÇÃO		diálogo	monólogo		diálogo
ESPECIALIZAÇÃO	PAPEL	PROCESSO				
especializada	constitutivo	EXPLICAR	"yuhoo respostas"	livro texto	palestra	debate
não-especializada		RELATAR	questionário	reportagem	depoimento	entrevista
		RECRIAR	quadrinhos	conto	causo	teatro de improviso
		COMPARTILHAR	e-chat	blog (diário)	vlog (diário)	bate-papo
	auxiliar	FAZER	carta comercial	receita	instruções	co-operação
especializada	constitutivo	RECOMENDAR	auto-ajuda	anúncios	orações	consulta médica
		HABILITAR	perguntas mais frequentes	panfletos	orientações	perguntas e respostas
		EXPLORAR	carta ao editor	artigo acadêmico	discurso	discussão

Fonte: Adaptado de Figueredo, Pagano e Ferregueti (2014).

O Quadro 2 mostra que a parte do corpus de referência utilizada para a geração da lista de palavras-chave conta com textos dos tipos: livro texto, reportagem, conto, blog (diário), receita, anúncios, panfletos e artigo acadêmico. Todos originalmente escritos em português e monológicos.

Em relação ao número de textos e de *tokens* da parte do CALIBRA utilizada como corpus de referência para este artigo, a Tabela 1 informa todas as quantidades.

TABELA 1 – O corpus de pesquisa: número de textos e de tokens

<b>Processo sociossemiótico</b>	<b>Número de textos</b>	<b>Número de <i>tokens</i></b>
Compartilhar	19	9.608
Explicar	24	28.841
Explorar	18	18.552
Fazer	56	32.646
Habilitar	53	58.875
Recomendar	25	36.912
Recriar	39	33.708
Relatar	30	33.142
<b>TOTAL</b>	<b>264</b>	<b>252.284</b>

Fonte: Elaborada para fins deste artigo.

### 3.3 A adaptação do trabalho de Pearson (1998) para o português brasileiro

Como detalhado anteriormente (cf. Referencial Teórico), o estudo publicado em Pearson (1998) contou com alguns recursos que ainda são pouco difundidos para o português brasileiro, como, por exemplo, o anotador morfossintático automático e o programa de padrão de correspondência. Em função disso, os parâmetros metodológicos estabelecidos para a presente pesquisa se baseiam em alguns dos passos metodológicos definidos em Pearson (1998), excluindo as fases dependentes dos recursos tecnológicos desenvolvidos para aquele trabalho. Portanto, não se pode afirmar que a metodologia da presente pesquisa replica, *ipsis litteris*, aquela descrita no estudo de Pearson (1998), uma vez que várias modificações metodológicas tiveram que ser implementadas.

Após a compilação do corpus de pesquisa, detalhado na primeira subseção desta metodologia, todos os textos desse corpus foram inseridos no *software* concordanciador AntConc. Em seguida houve a exclusão de todos os itens gramaticais, bem como os metadados que serviam para documentação do corpus. Isso se deu pelo ajuste das configurações de “Tool Preferences > WordList” do AntConc, a partir da inserção de uma lista de itens gramaticais importada para o concordanciador, permitindo que uma lista de palavras-chave fosse gerada sem o ruído causado pela presença desses itens gramaticais.

Com a lista de palavras-chave gerada, as três palavras dessa lista que tiveram maior número de ocorrências foram selecionadas para terem seus cotextos investigados. Como o uso do programa de padrão de correspondência em português brasileiro pode ser considerado restrito a algumas áreas do conhecimento, a ferramenta de geração de clusters/N-grams do AntConc foi determinante para que se pudesse verificar a existência de grupos nominais formados à esquerda e à direita da palavra-chave em questão. A busca pelos clusters/N-grams se deu sempre primeiro à esquerda e, depois, à direita da palavra-chave, sendo a extensão de cinco palavras para cada direção, contando com a palavra-chave em questão. Todos os clusters/N-grams formados à esquerda e à direita da palavra-chave que tivessem o mínimo de dez ocorrências deveriam ter seus respectivos cotextos examinados, por meio da ferramenta de concordância do AntConc, que permite verificar quais palavras estão em coocorrência com determinado cluster/N-gram e/ou palavra de busca. Quando a busca pela extensão de cinco palavras à esquerda e/ou à direita da palavra-chave não resultasse em clusters/N-grams com o mínimo de dez ocorrências, a extensão deveria ser reduzida de uma em uma palavra, isto é, quatro palavras, três palavras, até que clusters/N-grams com o número mínimo de dez ocorrências fossem obtidos.

Após a extração de cada cluster/N-gram à esquerda e à direita da palavra-chave que tivesse o número mínimo de dez ocorrências, deu-se início ao processo de análise desses clusters/N-grams. Essa análise compreende a busca por possíveis grupos nominais presentes nesses clusters/N-grams, bem como a anotação das funções exercidas por cada palavra que compõe o grupo nominal em análise.

É importante mencionar que esse processo de busca e análise de grupos nominais presentes em clusters/N-grams pode ser visto como uma forma semiautomática de pesquisa por termos técnicos, visto que em português brasileiro termos técnicos são comumente coextensivos ao grupo

nominal (cf. FIGUEREDO, 2007). Além disso, os artigos acadêmicos, que constituem o corpus de pesquisa utilizado no presente estudo, compreendem um tipo de texto em que as chances de aparecer termos técnicos são altas dado o grau de especialidade da linguagem utilizada nesse tipo de texto (HALLIDAY, 1978). Esse uso especializado da linguagem pode favorecer também o aparecimento de termos técnicos realizados por mais de uma palavra, isto é, grupos nominais que contam com mais de uma função, por exemplo, Ente e Qualificador (cf. FIGUEREDO, 2007).

O último passo metodológico do presente estudo foi verificar dentre os grupos nominais investigados anteriormente quais poderiam ser caracterizados como termos técnicos. Essa verificação foi feita pela análise da coocorrência desses grupos nominais com algum (alguns) dos sinais linguísticos apontados no trabalho de Pearson (1998) (cf. Referencial Teórico), bem como pela análise de alguma particularidade do corpus de pesquisa utilizado e/ou algum sinal linguístico que não foi mencionado no estudo de Pearson (1998), mas que se mostrou relevante para a presente pesquisa.

## 4 Resultados

Como detalhado nas seções de metodologia e de referencial teórico, o presente trabalho se baseia na pesquisa apresentada em Pearson (1998), a qual propõe uma metodologia para auxiliar terminologistas, lexicógrafos e aqueles que se dedicam à linguística de corpus a identificar termos técnicos por meio de ferramentas da linguística de corpus utilizando três corpora diferentes (cf. seção 3 Metodologia).

Na obra, a autora relata que a pesquisa se inicia quando os textos dos corpora são importados para um anotador morfossintático (*P.O.S. tagger*), desenvolvido especialmente para aquele trabalho, a fim de isolar os itens da classe de palavras nominal<sup>4</sup> e, posteriormente, verificar se eles formavam padrões, isto é, se aqueles itens nominais extraídos do anotador morfossintático eram frequentemente encontrados juntos nos corpora. A hipótese apresentada nesse estudo sugere que esses padrões

---

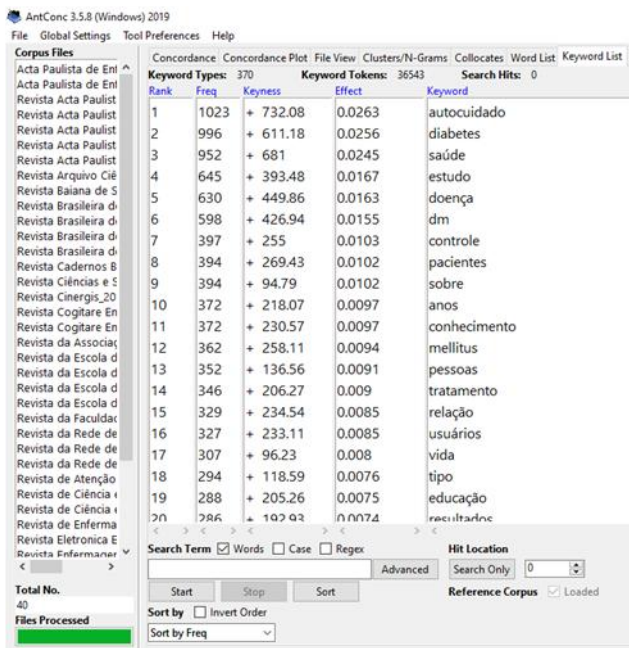
<sup>4</sup> Para a teoria Sistêmico-funcional (HALLIDAY; MATTHIESSEN, 2014, p. 75), há três classes de palavras – nominal (nominal), verbal (verbal) e adverbial (adverbial) – a classe nominal, sobre a qual o presente artigo enfoca, engloba os adjetivos (adjective); numerais (numeral), determinantes (determiner) e os substantivos (noun) que, por sua vez, podem ser comum (common), próprio (proper) ou pronomes (pronoun).



de coocorrência entre itens da classe de palavras nominal poderiam identificar termos técnicos comuns para as áreas de conhecimento dos corpora utilizados naquela pesquisa.

Entretanto, como ainda não há um *software* livre que faça esse tipo de processo de anotação morfosintática para o português brasileiro, sem que haja o desenvolvimento de um *script* em linguagem de programação e a utilização de um corpus de treinamento que auxilie na exatidão dos resultados, o presente estudo optou por fazer uma busca semiautomática dos itens da classe de palavras nominal no corpus utilizado neste trabalho (cf. Metodologia). Para isso, o corpus de artigos acadêmicos sobre Diabetes Mellitus foi importado para o *software* concordanciador AntConc. Após a extração dos itens gramaticais do corpus (cf. Metodologia), uma lista de palavras-chave foi gerada a partir do *upload* da parte escrita monológica do CALIBRA, a qual foi utilizada nesta pesquisa como corpus de referência. A seguir, a parte inicial da lista de palavras-chave é apresentada.

FIGURA 2 – Captura de tela com o início da lista de palavras-chave proveniente do corpus de artigos acadêmicos



Fonte: Elaborada para fins deste artigo.

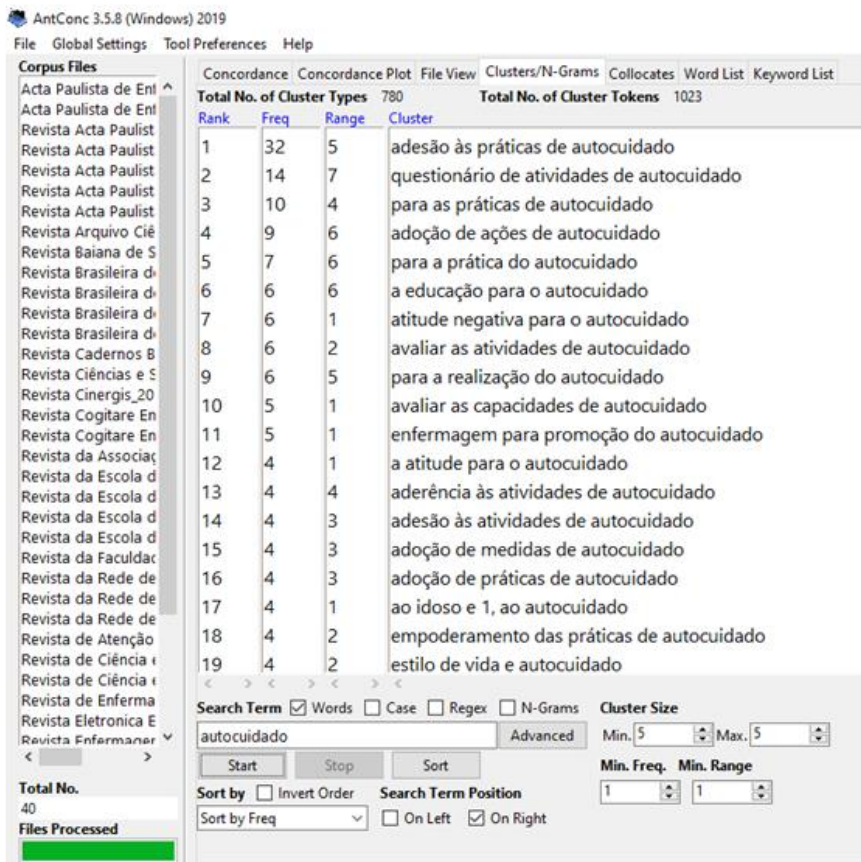
A Figura 2 mostra uma lista com as palavras-chave mais frequentes a partir da comparação que o *software* concordanciador faz com o corpus de referência utilizado. Os índices da coluna de frequência (Freq) detalham em ordem decrescente o número de vezes que determinada palavra-chave apareceu no corpus, comprovando que as palavras autocuidado (1023 ocorrências), diabetes (996 ocorrências) e saúde (952 ocorrências) são as três palavras-chave mais frequentes e por isso foram selecionadas para serem investigadas neste artigo.

Seguindo os passos metodológicos descritos em Pearson (1998), o próximo passo compreende a análise das coocorrências, ou seja, quais são as palavras da classe nominal que frequentemente coocorrem com outras palavras dessa mesma classe, sugerindo a possibilidade de termos técnicos daquela área de conhecimento. Como no presente artigo não foi utilizado um anotador morfossintático para selecionar somente as palavras da classe nominal, a ferramenta de clusters/N-grams do *software* concordanciador AntConc serviu para examinar qual(is) palavra(s) da classe nominal coocorre(iam) com cada uma das três palavras-chave selecionadas anteriormente.

Antes de apresentar os clusters/N-grams mais frequentes para cada uma das palavras-chave, é importante lembrar que a busca por clusters/N-grams objetiva encontrar termos técnicos da área de conhecimento dos textos do corpus de maneira semiautomática, adaptando a metodologia proposta por Pearson (1998) para o contexto do português brasileiro. Nesse sentido, grupos nominais devem estar instanciados nos clusters/N-grams gerados para as palavras-chave – autocuidado, diabetes e saúde (cf. Referencial Teórico).

Finalizado os esclarecimentos acerca da correspondência entre termo técnico, grupo nominal e clusters/N-grams, inicia-se o detalhamento dos clusters/N-grams formados à esquerda da palavra-chave com maior número de ocorrências – “autocuidado”.

FIGURA 3 – Captura de tela dos clusters/N-grams formados à esquerda da palavra-chave “autocuidado”



Fonte: Elaborada para fins deste artigo.

Na Figura 3, tem-se os clusters/N-grams formados com a palavra-chave autocuidado, sendo que esses clusters/N-grams abrangem cinco palavras à esquerda do item de busca, uma vez que o tamanho do cluster, marcado na parte inferior à direita da Figura 3 em *Cluster Size*, é cinco, tanto para o tamanho mínimo quanto para o tamanho máximo. Seguindo a metodologia do presente trabalho, os clusters/N-grams com o mínimo de dez ocorrências são selecionados para serem analisados de maneira mais profunda (cf. Metodologia). Portanto o cotexto em que os clusters/N-grams “adesão às práticas de autocuidado”, “questionário de

atividades de autocuidado” e “para as práticas de autocuidado” aparecem são explorados a seguir.

FIGURA 4 – Captura de tela do primeiro cluster/N-gram formado à esquerda da palavra-chave “autocuidado”: “adesão às práticas de autocuidado”

Hit	KWIC
1	do protocolo Compasso para promover a adesão às práticas de autocuidado em diabetes
2	e adequação culturalmente para promover a adesão às práticas de autocuidado em diabetes
3	estratégia inovadora capaz de incentivar a adesão às práticas de autocuidado, uma vez
4	desenvolver um protocolo para promover a adesão às práticas de autocuidado, cuja finalidade
5	cultural protocolo Compasso para promover a adesão às práticas de autocuidado em diabetes
6	abordar os principais temas sobre a adesão às práticas de autocuidado em diabetes
7	- alvo e este foi nomeado Compasso: adesão às práticas de autocuidado em diabetes
8	seria um protocolo para promover a adesão às práticas de autocuidado dentro do
9	que diz respeito à promoção da adesão às práticas de autocuidado em DM2,
10	telefônicas contextualizadas e de incentivo à adesão às práticas de autocuidado. Além disso,
11	significativo do escore mediano referente à adesão às práticas de autocuidado em diabetes (
12	0,001). Conclusão: A visita domiciliar promoveu à adesão às práticas de autocuidado com diabetes
13	estudo, as variáveis dependentes foram: adesão às práticas de autocuidado relacionadas à
14	autocuidado, respectivamente. O ESM mede a adesão às práticas de autocuidado do usuário
15	7 dias. Para indicar melhora quanto à adesão às práticas de autocuidado, deve-se
16	na linha de base quanto à adesão às práticas de autocuidado (p=0,894), mas
17	de empoderamento (p<0,001; Tabela 2). Quanto à adesão às práticas de autocuidado, a comparação
18	a visita domiciliar foi efetiva para adesão às práticas de autocuidado com diabetes,
19	visita domiciliar como estratégia educativa para adesão às práticas de autocuidado com diabetes
20	eramento também encontram resultados positivos na adesão às práticas de autocuidado ao abordarem
21	. CONCLUSÃO A visita domiciliar promoveu à adesão às práticas de autocuidado com diabetes
22	demonstrada a eficácia das orientações na adesão às práticas de autocuidado com os
23	2015. Foram coletados dados com relação à adesão às práticas de autocuidado, ao empoderamento
24	uma prática educativa direcionada para a adesão às práticas de autocuidado em diabetes,
25	, foram utilizados os instrumentos validados de adesão às práticas de autocuidado para o
26	glicada. O instrumento ESM mede a adesão às práticas de autocuidado do usuário
27	dias. Para indicar melhora quanto à adesão às práticas de autocuidado, deve-se
28	menor que 0,001, Tabela 3). Em relação à adesão às práticas de autocuidado (ESM), obtiveram-
29	glicêmico e os comportamentos para a adesão às práticas de autocuidado podem ser
30	controle do diabetes. Com relação à adesão às práticas de autocuidado, as medianas
31	condição, aumento do empoderamento e da adesão às práticas de autocuidado, principalmente para
32	na intervenção individual; e quanto à adesão às práticas de autocuidado foi observada

Fonte: Elaborada para fins deste artigo.

O primeiro cluster/N-gram formado com cinco palavras à esquerda de “autocuidado” é “adesão às práticas de autocuidado”, como mostra a Figura 4. Nesse cluster/N-gram, há dois grupos nominais, a saber:

QUADRO 3 – Os dois grupos nominais presentes em “adesão às práticas de autocuidado” e as respectivas funções exercidas na ordem do grupo nominal

<b>1º grupo nominal</b>	adesão	às práticas de autocuidado	
	<b>Ente</b>	<b>Qualificador</b>	
<b>2º grupo nominal</b>	às	práticas	de autocuidado
	<b>Dêitico</b>	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Contudo, o primeiro grupo nominal surge de uma metaforização<sup>5</sup> em que “aderir” sofreu um processo de metaforização gramatical para “adesão”, havendo a nominalização de “aderir” para “adesão”, comum em textos em que o uso da linguagem especializada é predominante. Por essa perspectiva prevista pela teoria sistêmico-funcional, esse primeiro grupo nominal não existiria, já que esse cluster/N-gram seria uma oração: “aderir às práticas de autocuidado”. Em razão disso, “as práticas de autocuidado” passa a ser entendido como o único grupo nominal desse cluster/N-gram.

O segundo cluster/N-gram formado a partir da palavra-chave “autocuidado” é “questionário de atividades de autocuidado”. A Figura 5 traz as linhas de concordância para esse cluster/N-gram.

<sup>5</sup> Segundo os pressupostos da teoria sistêmico-funcional, a metáfora gramatical é frequentemente encontrada em textos científicos, como os artigos acadêmicos, uma vez que é nesse tipo de texto em que a linguagem especializada pode ser identificada. Uma das formas possíveis de se analisar e compreender a metáfora gramatical é por meio das mudanças de ordem, isto é, uma oração é metaforizada e torna-se um grupo nominal (cf. HALLIDAY; MATTHIESSEN, 2014).

FIGURA 5 – Captura de tela do segundo cluster/N-gram formado à esquerda da palavra-chave “autocuidado”: “questionário de atividades de autocuidado”



Fonte: Elaborada para fins deste artigo.

Assim como o primeiro cluster/N-gram formado à esquerda da palavra-chave “autocuidado”, este segundo cluster/N-gram também apresenta dois grupos nominais.

QUADRO 4 – Os grupos nominais presentes em “questionário de atividades de autocuidado” e as respectivas funções exercidas na ordem do grupo nominal

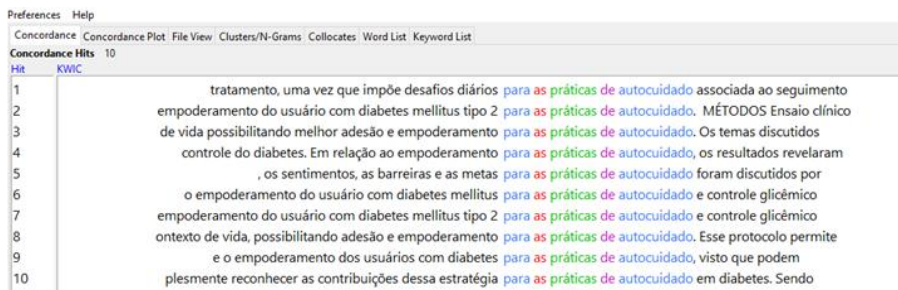
<b>1º grupo nominal</b>	Questionário	de atividades de autocuidado
	<b>Ente</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	Atividades	de autocuidado
	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Embora haja dois grupos nominais no cluster/N-gram “questionário de atividades de autocuidado”, pela análise do contexto das linhas de concordância da Figura 5, é possível afirmar que o primeiro grupo nominal, que engloba todo cluster/N-gram, é o único que pode ser considerado, uma vez que ele corresponde ao nome do questionário utilizado nas pesquisas reportadas nos artigos selecionados para o corpus. Em outras palavras, o segundo grupo nominal “atividades de autocuidado” apresenta um significado diferente daquele sinalizado pela análise das linhas de concordância do cluster/N-gram em questão, cujo objetivo deve ser revelar o nome do instrumento usado.

O terceiro e último cluster/N-gram formado à esquerda da palavra-chave “autocuidado” é “para as práticas de autocuidado”. A Figura 6, a seguir, mostra as linhas de concordância desse cluster/N-gram.

FIGURA 6 – Captura de tela do terceiro cluster/N-gram formado à esquerda da palavra-chave “autocuidado”: “para as práticas de autocuidado”



Fonte: Elaborada para fins deste artigo.

Esse cluster/N-gram é formado pela preposição “para”, a qual não exerce função no grupo nominal presente no cluster/N-gram, por isso tal palavra não aparece no Quadro 5, que aborda as funções presentes no grupo nominal em questão.

QUADRO 5 – O grupo nominal presente em “para as práticas de autocuidado” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	as	práticas	de autocuidado
	Dêitico	Ente	Qualificador

Fonte: Elaborado para fins deste artigo.

Encerrado o detalhamento dos cluster/N-grams formados à esquerda da palavra-chave “autocuidado” que tiveram o número mínimo de dez ocorrências no corpus, torna-se necessário apresentar os clusters/N-grams formados à direita dessa palavra-chave. Nesta segunda configuração, os clusters/N-grams também seguem o padrão de busca de cinco palavras à direita de “autocuidado” como mostra a Figura 7 a seguir.



FIGURA 7 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “autocuidado”

AntConc 3.5.8 (Windows) 2019  
File Global Settings Tool Preferences Help

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Cluster Types 925 Total No. of Cluster Tokens 1018

Rank	Freq	Range	Cluster
1	10	6	autocuidado de pacientes com diabetes
2	9	2	autocuidado com os pés e
3	7	6	autocuidado com o diabetes (qad
4	4	1	autocuidado aos portadores de diabetes
5	4	4	autocuidado de pessoas com diabetes
6	4	2	autocuidado dos pacientes com dm
7	4	3	autocuidado para o controle do
8	4	1	autocuidado às pessoas com dm
9	3	1	autocuidado com diabetes mellitus tipo
10	3	1	autocuidado da pessoa portadora de
11	3	3	autocuidado do usuário com diabetes
12	3	2	autocuidado e os fatores que
13	3	1	autocuidado em diabetes mellitus via
14	2	2	autocuidado com o diabetes e
15	2	2	autocuidado com o diabetes mellitus
16	2	1	autocuidado com os pés orientadas
17	2	2	autocuidado com os pés. o
18	2	2	autocuidado das pessoas com dm
19	2	1	autocuidado das pessoas portadoras de

Search Term  Words  Case  Regex  N-Grams Cluster Size  
 Advanced Min. 5 Max. 5  
 Start Stop Sort  
 Sort by  Invert Order Search Term Position  
 On Left  On Right  
 Min. Freq. 1 Min. Range 1

Total No. 40  
Files Processed

Fonte: Elaborada para fins deste artigo.

A lista de clusters/N-grams formados à direita da palavra-chave “autocuidado”, apresentada na Figura 7, revela que somente um cluster/N-gram teve o mínimo de dez ocorrências – “autocuidado de pacientes com diabetes”. Adiante, as linhas de concordância, que mostram os cotextos em que esse cluster/N-gram aparece, estão retratadas na Figura 8.



FIGURA 8 – Captura de tela do primeiro cluster/N-gram formado à direita da palavra-chave “autocuidado”: “autocuidado de pacientes com diabetes”



Fonte: Elaborada para fins deste artigo.

Esse cluster/N-gram possui dois grupos nominais, como mostra as classificações destacadas no Quadro 6 a seguir.

QUADRO 6 – Os grupos nominais presentes em “autocuidado de pacientes com diabetes” e as respectivas funções exercidas na ordem do grupo nominal

<b>1º grupo nominal</b>	autocuidado	de pacientes com diabetes
	<b>Ente</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	pacientes	com diabetes
	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Os dois grupos nominais encontrados no cluster “autocuidado de pacientes com diabetes” apresentam significados diferentes. No primeiro grupo nominal o que está qualificado é o Ente “autocuidado” pela frase preposicional “de pacientes com diabetes”. No segundo grupo nominal o que está qualificado é o Ente “pacientes” pela frase preposicional “com diabetes”. Como a palavra-chave que norteou a busca pelos clusters/N-grams foi “autocuidado” não seria coerente considerar o segundo grupo nominal de forma desmembrada do primeiro, em razão das diferentes instanciações para a função do Ente. Nesse caso, portanto, há a correspondência exata entre o cluster/N-gram e o grupo nominal.

A segunda palavra-chave mais frequente no corpus é “diabetes”, como mostra a Figura 1. Assim, os resultados apresentados a seguir dizem respeito aos dados encontrados para tal palavra. A Figura 9 traz

os clusters/N-grams formados à esquerda de “diabetes”, considerando o intervalo de cinco palavras.

FIGURA 9 – Captura de tela dos clusters/N-grams formados à esquerda da palavra-chave “diabetes”

Concordance		Concordance Plot		File View		Clusters/N-Grams		Collocates		Word List		Keyword List			
Total No. of Cluster Types						794						Total No. of Cluster Tokens		996	
Rank	Freq	Range	Cluster												
1	19	8	para o autocuidado em diabetes												
2	18	8	de autocuidado com o diabetes												
3	15	6	práticas de autocuidado em diabetes												
4	11	6	para o controle do diabetes												
5	10	6	autocuidado de pacientes com diabetes												
6	10	4	com a condição do diabetes												
7	6	4	o conhecimento sobre o diabetes												
8	5	2	de atitudes psicológicas do diabetes												
9	5	2	programa de educação em diabetes												
10	5	1	práticas de autocuidado com diabetes												
11	4	1	a condição crônica do diabetes												
12	4	4	a sociedade brasileira de diabetes												
13	4	1	autocuidado aos portadores de diabetes												
14	4	4	autocuidado de pessoas com diabetes												
15	4	1	conhecimento e autocuidado em diabetes												
16	4	2	mudança de comportamento em diabetes												
17	4	3	programas de educação em diabetes												
18	4	4	tempo de diagnóstico do diabetes												
19	3	3	ao conhecimento aeral do diabetes												

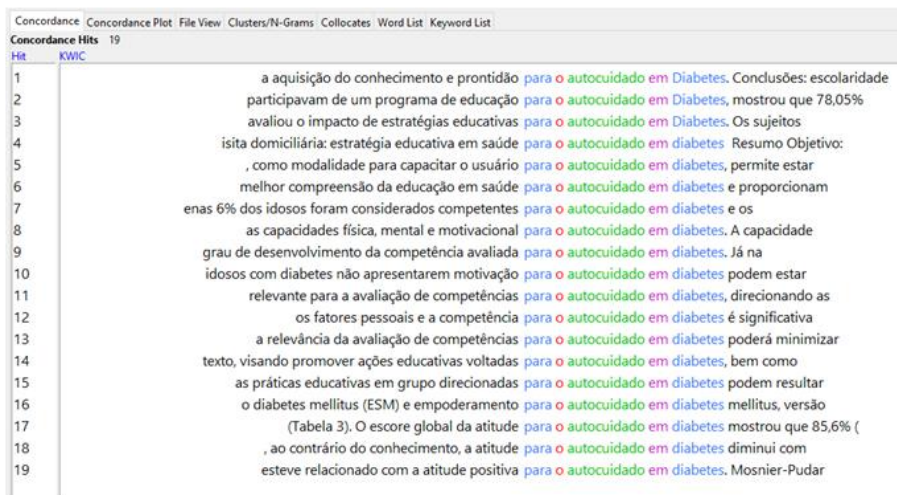
Search Term	<input checked="" type="checkbox"/> Words	<input type="checkbox"/> Case	<input type="checkbox"/> Regex	<input type="checkbox"/> N-Grams	Cluster Size
diabetes	Advanced			Min. 5	Max. 5
Start	Stop	Sort		Min. Freq. 1	Min. Range 1
Sort by	<input type="checkbox"/> Invert Order	Search Term Position			
Sort by Freq	<input type="checkbox"/> On Left	<input checked="" type="checkbox"/> On Right			

Fonte: Elaborada para fins deste artigo.

Seguindo a metodologia previamente desenvolvida para este artigo, os clusters/N-grams que têm no mínimo dez ocorrências são examinados, são eles: “para o autocuidado em diabetes”, “de autocuidado com o diabetes”, “práticas de autocuidado em diabetes”, “para o controle do diabetes”, “autocuidado de pacientes com diabetes” e “com a condição do diabetes”.

As linhas de concordância do primeiro cluster/N-gram formado à esquerda de “diabetes” estão na Figura 10 adiante.

FIGURA 10 – Captura de tela do primeiro cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “para o autocuidado em diabetes”



Fonte: Elaborada para fins deste artigo.

A primeira palavra do cluster/N-gram em questão é a preposição “para”, a qual não apresenta função no grupo nominal “o autocuidado em diabetes” que está no cluster/N-gram, destacado na Figura 10. O Quadro 7 destaca as funções deste grupo nominal.

QUADRO 7 – O grupo nominal presente em “para o autocuidado em diabetes” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	o	autocuidado	em diabetes
	Dêitico	Ente	Qualificador

Fonte: Elaborado para fins deste artigo.

O segundo cluster/N-gram formado à esquerda da palavra-chave “diabetes” compreende a expressão “de autocuidado com o diabetes”. Para viabilizar o entendimento de tal cluster/N-gram, as linhas de concordância em que ele aparece no corpus estão destacadas na Figura 11.

FIGURA 11 – Captura de tela do segundo cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “de autocuidado com o diabetes”

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits		18					
Hit	KWIC						
1		autocuidado foi o Questionário de Atividades de Autocuidado com o Diabetes, previamente validado					
2		objetivo deste estudofoi avaliar as atividades de autocuidado com o diabetes em pessoas					
3		pesquisa foi o Questionário de Atividades de Autocuidado com o Diabetes, versão traduzida,					
4		Grupo controle. Foram utilizados os questionários de Autocuidado com o diabetes e Diabetes					
5		controle (Tabela 3), o efeito no escore de autocuidado com o diabetes (ΔESM) no					
6		2012, com uso de questionário de atividades de autocuidado com o diabetes e instrumento					
7		discutem a baixa adesão às atividades de autocuidado com o diabetes, descrevendo possíveis					
8		utilizou-se o Questionário de Atividades de Autocuidado com o Diabetes (QAD) versão					
9		últimos 12 meses. O Questionário de Atividades de Autocuidado com o Diabetes (QAD) permitiu					
10		-PAID) e o Questionário de atividades de autocuidado com o Diabetes (QAD). O					
11		foi avaliado pelo Questionário de Atividades de Autocuidado com o Diabetes, em que					
12		foi denominado de Questionário de Atividades de Autocuidado com o Diabetes (QAD). O					
13		. Aplicou-se o Questionário de Atividades de Autocuidado com o diabetes que abordou					
14		2013. Utilizou-se o Questionário de Atividades de Autocuidado com o Diabetes (QAD). Este					
15		do estudo, o Questionário de Atividades de Autocuidado com o Diabetes mostrou que,					
16		por meio do Questionário de Atividades de Autocuidado com o Diabetes (QAD), versão					
17		strável, específico para avaliação das atividades de autocuidado com o diabetes, e possui 15					
18		por meio do Questionário de Atividades de Autocuidado com o Diabetes (QAD), versão					

Fonte: Elaborada para fins deste artigo.

O cotexto onde o cluster/N-gram em questão aparece evidencia que o grupo nominal localizado nesse cluster/N-gram faz parte de um grupo nominal maior, “Questionário de atividades de autocuidado com o diabetes”, que, por sua vez, teve a outra porção analisada no segundo cluster/N-gram mais frequente para a palavra-chave “autocuidado”. Por esse motivo, a preposição “de” presente no cluster “de autocuidado com o diabetes” funciona como preposição na frase preposicional “de autocuidado”, que qualifica “atividades”. O Quadro 8 mostra, portanto, as funções presentes no grupo nominal “autocuidado com o diabetes”.

QUADRO 8 – O grupo nominal presente em “de autocuidado com o diabetes” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	autocuidado	com o diabetes
	Ente	Qualificador

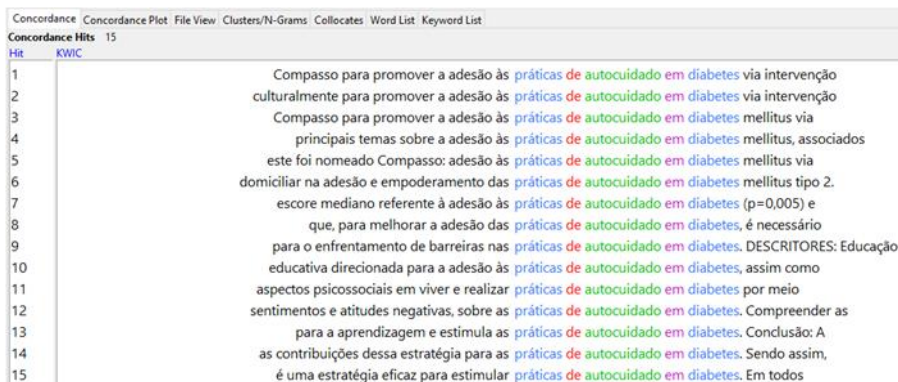
Fonte: Elaborado para fins deste artigo.

As funções, destacadas no Quadro 8, que constituem o grupo nominal presente no cluster/N-gram em questão, demonstram que a frase preposicional “com o diabetes” exerce a função de qualificar o ente

“autocuidado”, ressaltando o tipo de autocuidado, no caso, aquele que se refere à condição crônica diabetes.

O terceiro cluster/N-gram mais frequente formado à esquerda da palavra-chave “diabetes” diz respeito à “práticas de autocuidado em diabetes”. As linhas de concordância em que esse cluster/N-gram aparece estão detalhadas na Figura 12.

FIGURA 12 – Captura de tela do terceiro cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “práticas de autocuidado em diabetes”



Fonte: Elaborada para fins deste artigo.

Esse terceiro cluster/N-gram conta com dois grupos nominais cujas funções estão no Quadro 9 a seguir.

QUADRO 9 – Os grupos nominais presentes em “práticas de autocuidado em diabetes” e as respectivas funções exercidas na ordem do grupo

<b>1º grupo nominal</b>	práticas	de autocuidado em diabetes
	<b>Ente</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	autocuidado	em diabetes
	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

O primeiro grupo nominal presente no Quadro 9 engloba o cluster/N-gram em questão por completo. Neste grupo nominal, a frase preposicional “de autocuidado em diabetes” qualifica o Ente “práticas”,

delimitando quais os tipos de prática são frequentemente abordados no corpus utilizado. O segundo grupo nominal “autocuidado em diabetes” tem como Ente “autocuidado” e como qualificador “em diabetes”, este responsável por indicar qual o tipo de autocuidado os textos do corpus tratam. Fica evidente na análise das funções dos dois grupos nominais que seus significados são diferentes, enquanto no primeiro grupo nominal o que está qualificado são as “práticas”, no segundo o que está qualificado é o “autocuidado”. Entretanto, como um dos principais objetivos deste artigo é investigar a correspondência entre cluster/N-grams, grupo nominal e termo técnico, o primeiro grupo – práticas de autocuidado em diabetes – tende a se aproximar com maior exatidão dessa correspondência, por isso, a partir deste momento, apenas esse grupo é considerado como grupo nominal presente no terceiro cluster/N-gram formado à esquerda da palavra-chave “diabetes”. Para além disso, o pressuposto de que os clusters/N-grams devem ter no mínimo dez ocorrências reforça também a escolha pelo primeiro grupo nominal, já que não se sabe qual a frequência do segundo grupo nominal no corpus.

O quarto cluster/N-gram formado à esquerda da palavra-chave “diabetes” trata-se da expressão “para o controle do diabetes”. A Figura 13 mostra as linhas de concordância onde este cluster/N-gram pode ser encontrado no corpus.

FIGURA 13 – Captura de tela do quarto cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “para o controle do diabetes”

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits		11					
Hit	KWIC						
1		que interferem no desenvolvimento destes cuidados para o controle do diabetes. Em relação					
2		autocuidado, constitui- -se a peça principal para o controle do diabetes mellitus (DM),					
3		as variáveis de conhecimento e autocuidado para o controle do diabetes mellitus. Nesse					
4		e longo prazo apresenta resultados favoráveis para o controle do diabetes mellitus. Outros					
5		ao controle glicêmico e às habilidades para o controle do diabetes mellitus. O					
6		o usuário na aquisição de habilidades para o controle do diabetes mellitus e					
7		apresentando baixo conhecimento e pouco cuidado para o controle do diabetes mellitus. Porém,					
8		população. Uma das formas de colaborar para o controle do diabetes é por					
9		sua condição e dos comportamentos necessários para o controle do diabetes. Com relação					
10		ou não as medidas de autocuidado para o controle do diabetes. Comumente, a					
11		o processo de ensino e aprendizagem para o controle do diabetes, torna-se					

Fonte: Elaborada para fins deste artigo.

A análise dos cotextos em que este cluster/N-gram aparece possibilita afirmar que o “para” funciona como uma conjunção na ordem



da oração. Segundo o trabalho de Figueredo (2007), essa função não está diretamente relacionada ao grupo nominal em português brasileiro, ou seja, palavras da classe adverbial exercem prototipicamente funções no grupo adverbial, ao passo que palavras da classe nominal tem funções prototípicas no grupo nominal. Por essa razão, o grupo nominal presente neste cluster/N-gram envolve apenas a porção: “o controle do diabetes”. O Quadro 10 salienta as funções desse grupo nominal.

QUADRO 10 – O grupo nominal presente em “para o controle do diabetes” e as respectivas funções exercidas na ordem do grupo nominal

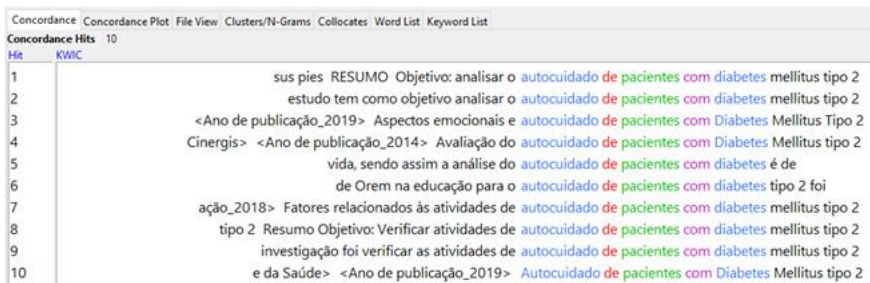
<b>Grupo nominal</b>	o	controle	do diabetes
	<b>Dêitico</b>	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Em conjunto com a preposição “de” a palavra-chave “diabetes” opera como um grupo nominal formado por uma frase preposicional, cuja função é qualificar o Ente “controle”, caracterizando o tipo de controle, no caso, aquele relacionado ao Diabetes Mellitus.

O quinto cluster/N-gram formado à esquerda de “diabetes” é “autocuidado de pacientes com diabetes”. A seguir, a Figura 14 revela as linhas de concordância com os cotextos em que este cluster/N-gram aparece.

FIGURA 14 – Captura de tela do quinto cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “autocuidado de pacientes com diabetes”



Fonte: Elaborada para fins deste artigo.

Tal como os exemplos anteriores, o cluster/N-gram em questão “autocuidado de pacientes com diabetes” conta com dois grupos nominais, como mostra o Quadro 11.

QUADRO 11 – Os grupos nominais presentes em “autocuidado de pacientes com diabetes” e as respectivas funções exercidas na ordem do grupo nominal

<b>1º grupo nominal</b>	autocuidado	de pacientes com diabetes
	<b>Ente</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	pacientes	com diabetes
	<b>Ente</b>	<b>Qualificador</b>

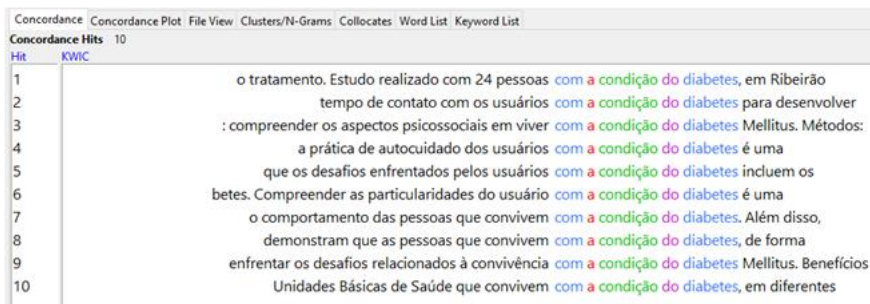
Fonte: Elaborado para fins deste artigo.

O primeiro grupo nominal “autocuidado de pacientes com diabetes” envolve todo o cluster/N-gram que está sendo examinado, enquanto que o segundo grupo nominal está em relação de dependência (hipotaxe) com o primeiro nominal. Isso salienta que os dois grupos nominais apresentam significados diferentes. No primeiro grupo nominal, o Ente é “autocuidado”, e a qualificação se dá por “de pacientes com diabetes”, que classifica o tipo de autocuidado abordado nos artigos acadêmicos do corpus. No segundo grupo nominal, o Ente é “pacientes”, e a qualificação é “com diabetes”, identificando que os tipos de pacientes mencionados no corpus são aqueles que tem diabetes. Tendo em mente o fato de que o primeiro grupo nominal “autocuidado de pacientes com diabetes” envolve todo o cluster/N-gram em questão, e, portanto, tem o mínimo de dez ocorrências requisitadas na metodologia deste artigo, bem como apresenta significado mais abrangente e coerente ao contexto em que aparece nos textos do corpus, este grupo nominal passa a ser considerado o único válido para este cluster/N-gram.

O sexto e último cluster/N-gram formado à esquerda da palavra-chave diabetes, considerando a extensão de cinco palavras e com o mínimo de dez ocorrências, é “com a condição do diabetes”. A Figura 15 traz as linhas de concordância que mostram os cotextos em que este cluster/N-gram aparece no corpus.



FIGURA 15 – Captura de tela do sexto cluster/N-gram formado à esquerda da palavra-chave “diabetes”: “com a condição do diabetes”



Fonte: Elaborada para fins deste artigo.

Investigando as linhas de concordância da Figura 15, observa-se que o cluster/N-gram em questão começa com a preposição “com”, a qual em conjunto com o restante do cluster/N-gram instanciado pelo grupo nominal “a condição do diabetes”, tem a função de qualificar os Entes que estão fora do escopo do cluster/N-gram, como, por exemplo, “24 pessoas”, “os usuários” e “as pessoas”. Nesse sentido, se a preposição “com” fosse considerada como parte do grupo nominal presente no cluster/N-gram, a extensão deste cluster/N-gram precisaria ser modificada, o que causaria incoerência com a metodologia estabelecida para o presente estudo, bem como poderia causar alterações nos resultados encontrados até então. Por esses motivos, o grupo nominal presente no cluster/N-gram “com a condição do diabetes” diz respeito à “a condição do diabetes”. As funções deste grupo nominal estão destacadas no Quadro 12 a seguir.

QUADRO 12 – O grupo nominal presente em “com a condição do diabetes” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	a	condição	do diabetes
	Dêitico	Ente	Qualificador

Fonte: Elaborado para fins deste artigo.

Neste grupo nominal, o qualificador “do diabetes” tem a função de identificar o tipo de “condição” que foi mencionada com maior frequência nos textos do corpus utilizado na presente pesquisa. “Condição”, por sua vez, funciona como o Ente do grupo nominal. Ainda neste grupo nominal, aparece o determinante “a” que opera como dêitico.

Encerrada a análise dos clusters/N-grams formados à esquerda da palavra-chave “diabetes” que obtiveram o número mínimo de dez ocorrências, a Figura 16 apresenta os clusters/N-grams formados à direita desta palavra-chave.

FIGURA 16 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “diabetes” com extensão de cinco palavras

Concordance				Clusters/N-Grams			
Total No. of Cluster Types			Total No. of Cluster Tokens				
Rank	Freq	Range	Cluster				
1	6	6	diabetes mellitus (dm) é uma				
2	5	1	diabetes mellitus tipo 2 em uso				
3	4	4	diabetes mellitus tipo 2 (dm2) é				
4	3	1	diabetes mellitus via intervenção telefônica				
5	3	2	diabetes por mais de 5 anos				
6	2	2	diabetes (qad), versão traduzida, adaptada				
7	2	2	diabetes cadastrada nas unidades de				
8	2	2	diabetes consistem em melhorar o				
9	2	1	diabetes e fazer uso de				
10	2	1	diabetes empowerment scale-short form				
11	2	2	diabetes há cinco anos ou				
12	2	2	diabetes knowledge questionnaire (dkn-a				
13	2	1	diabetes knowledge scale (dkn-a				
14	2	2	diabetes mellitus destaca-se por				
15	2	2	diabetes mellitus e hipertensão arterial				
16	2	1	diabetes mellitus em todas as				
17	2	1	diabetes mellitus que possuem úlceras				
18	2	2	diabetes mellitus tipo 2 resumo objetivo				
19	2	2	diabetes mellitus tipo 2 (dm2). de				

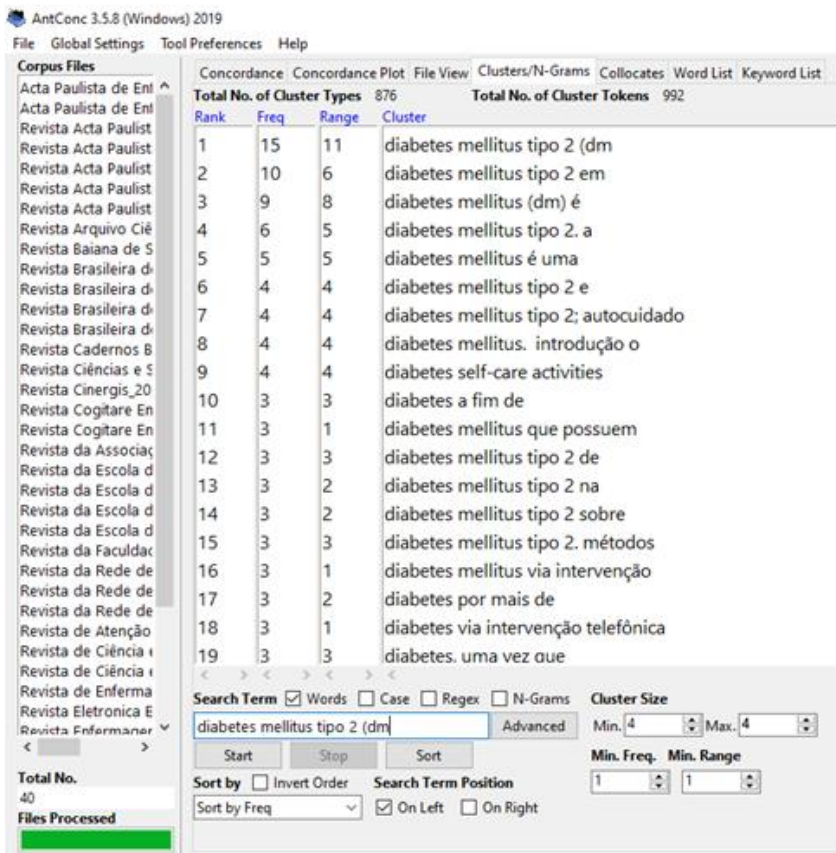
  

Search Term	<input checked="" type="checkbox"/> Words	<input type="checkbox"/> Case	<input type="checkbox"/> Regex	<input type="checkbox"/> N-Grams	Cluster Size
diabetes	Advanced			Min. 5	Max. 5
Start	Stop	Sort		Min. Freq. Min. Range	
Sort by	<input type="checkbox"/> Invert Order	Search Term Position		1	1
Sort by Freq	<input checked="" type="checkbox"/> On Left	<input type="checkbox"/> On Right			

Fonte: Elaborada para fins deste artigo.

Observando a coluna de frequências (Freq) da Figura 16, é possível afirmar que não há nenhum cluster/N-gram com o mínimo de dez ocorrências, condição estabelecida na metodologia do presente artigo. Por isso, o tamanho mínimo e máximo do cluster foi reduzido para quatro, como mostra a Figura 17 a seguir.

FIGURA 17 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “diabetes” com extensão de quatro palavras



Fonte: Elaborada para fins deste artigo.

Reduzindo a extensão do cluster/N-gram para quatro palavras à direita da palavra-chave “diabetes”, dois clusters/N-grams aparecem com o mínimo de dez ocorrências. O primeiro diz respeito à “diabetes mellitus tipo 2 (dm)”. Adiante, as linhas de concordância, que demonstram onde este cluster/N-gram ocorre nos textos do corpus, estão representadas na Figura 18.

FIGURA 18 – Captura de tela do primeiro cluster/N-gram formado à direita da palavra-chave “diabetes”: “diabetes mellitus tipo 2 (DM)”



Fonte: Elaborada para fins deste artigo.

O cotexto deste cluster/N-gram envolve o nome da condição crônica que todos os textos do corpus utilizado abordam, seguido da sigla que identifica a condição aos pares, uma vez que o público-alvo principal dos artigos acadêmico são pesquisadores, assim como os autores desse tipo de texto que, na maior parte das vezes, também ocupam esse papel de pesquisador. O Quadro 13 expõe as funções deste grupo nominal.

QUADRO 13 – O grupo nominal presente em “diabetes mellitus tipo 2 (DM)” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	Diabetes	Mellitus	Tipo 2
	<b>Ente</b>	<b>Classificador</b>	<b>Classificador</b>

Fonte: Elaborado para fins deste artigo.

Seguindo a iniciativa salientada no trabalho de Bowker e Pearson (2002) sobre o fato de se considerar variações de um item como um único termo, em que Diabetes Mellitus tipo 2, Diabetes Mellitus e DM2 corresponderiam a um mesmo termo, ou seja, o significado de todas essas variações é o mesmo, todas dizem respeito à condição crônica Diabetes Mellitus, embora cada ocorrência possa ser contabilizada. Considerando esta perspectiva, o presente artigo computa as ocorrências de DM e Diabetes Mellitus tipo 2, presentes no cluster/N-gram em questão, de forma separada, mas considera que ambos os termos tem significados análogos. Nesse sentido, “Diabetes” funciona como Ente do grupo nominal.

“Mellitus” e “Tipo 2” como Classificadores de “Diabetes”, inserindo essa condição na classe dos mellitus e do tipo 2, não do tipo 1, por exemplo.

O segundo cluster/N-gram formado à direita da palavra-chave “Diabetes” é “diabetes mellitus tipo 2 em”. A Figura 19 exhibe as linhas de concordância em que esse cluster/N-gram aparece.

FIGURA 19 – Captura de tela do segundo cluster/N-gram formado à direita da palavra-chave “diabetes”: “diabetes mellitus tipo 2 em”

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 12							
Hit	KWIC						
1		para o autocuidado de portadores de	diabetes mellitus tipo 2 em	uso de insulina			
2		em um grupo de brasileiros com	diabetes mellitus tipo 2 em	uso de insulina.			
3		etapa pré-teste, participaram 50 usuários com	diabetes mellitus tipo 2 em	uso de insulina			
4		aduzida e adaptada, participaram 150 usuários com	diabetes mellitus tipo 2 em	uso de insulina			
5		em um grupo de 150 pessoas com	diabetes mellitus tipo 2 em	uso de insulina.			
6		em um grupo de usuários com	diabetes mellitus tipo 2 em	uso de insulina,			
7		elevação das taxas da ocorrência de	diabetes mellitus tipo 2 em	todo o mundo.			
8		emocionais e autocuidado de pacientes com	Diabetes Mellitus Tipo 2 em	Terapia Renal Substitutiva			
9		.2014> Avaliação do autocuidado de pacientes com	Diabetes Mellitus tipo 2 em	uma unidade de			
10		intervenções de enfermagem aos usuários com	diabetes mellitus tipo 2 em	nível de atenção			
11		um estudo realizado com 150 usuários com	diabetes mellitus tipo 2 em	Londrina/PR. Esse			
12		atividades de autocuidado de pacientes com	diabetes mellitus tipo 2 em	seguimento ambulatorial, e			

Fonte: Elaborada para fins deste artigo.

A análise das linhas de concordância aponta que o grupo nominal “diabetes mellitus tipo 2” presente no cluster/N-gram em questão faz parte de grupos nominais maiores, como: “50 usuários com diabetes mellitus tipo 2 em uso de insulina” e “avaliação do autocuidado de pacientes com diabetes mellitus tipo 2 em uma unidade”. Em função disso, a preposição “em” que aparece ao final deste segundo cluster/N-gram opera como a preposição presente nas frases preposicionais (preposição + grupo nominal) que sucedem este cluster/N-gram. A depender da linha de concordância examinada, essas frases preposicionais podem indicar o lugar onde os pacientes/usuários com Diabetes Mellitus abordados nos artigos acadêmicos do corpus realizam seus tratamentos ou podem informar sobre a maneira pela qual esses pacientes/usuários recebem tratamento, no caso, pelo uso de insulina.

Ainda com relação ao grupo nominal “diabetes mellitus tipo 2” presente no cluster/N-gram em questão, a análise das linhas de concordância da Figura 19 indica que esse grupo nominal opera junto com as preposições “com/de”, constituindo uma frase preposicional, a qual qualifica o Ente que a antecede. Esse Ente é realizado por “usuários”, “pessoas”, “portadores”, “pacientes”, “grupo” ou “taxas”, dependendo de qual linha de concordância está em pauta.



Tendo em mente essas particularidades do cluster/N-gram “diabetes mellitus tipo 2 em” sobre o fato do grupo nominal que aparece nesse cluster/N-gram apresentar função tanto com as palavras que o antecede quanto com aquelas que o sucede inviabiliza a análise desse grupo nominal de forma isolada. Em outras palavras, para compreender o grupo nominal “diabetes mellitus tipo 2” nos casos destacados pelas linhas de concordância da Figura 19, é primordial entender as funções desse grupo em relação ao que vem antes e depois dele. No entanto, como a metodologia do presente artigo estabelece que a extensão do cluster/N-gram não pode ultrapassar as cinco palavras, os dados obtidos para este grupo nominal, e, conseqüentemente, para o cluster/N-gram em que ele está presente não estão incluídos nos resultados deste trabalho.

A terceira e última palavra-chave selecionada para o presente estudo diz respeito ao item “saúde”. A Figura 20 destaca os clusters/N-grams formados à esquerda desta palavra.

FIGURA 20 – Captura de tela dos clusters/N-grams formados à esquerda da palavra-chave “saúde”

The screenshot shows the AntConc 3.5.8 (Windows) 2019 interface. The main window displays a table of clusters/N-grams for the search term "saúde". The table has four columns: Rank, Freq, Range, and Cluster. The search term is "saúde" and the cluster size is set to 5. The search is sorted by frequency.

Rank	Freq	Range	Cluster
1	14	8	profissionais da área da saúde
2	11	4	profissional da área da saúde
3	10	6	uma unidade básica de saúde
4	9	7	que os profissionais de saúde
5	7	7	acesso aos serviços de saúde
6	7	2	de educação para a saúde
7	7	5	na unidade básica de saúde
8	6	5	a equipe multiprofissional de saúde
9	6	3	ações de promoção da saúde
10	5	2	a educação para a saúde
11	5	4	ações de educação em saúde
12	5	3	da unidade básica de saúde
13	5	2	na atenção primária de saúde
14	4	1	de letramento funcional em saúde
15	4	4	do conselho nacional de saúde
16	4	3	na atenção primária à saúde
17	4	2	o letramento funcional em saúde
18	4	3	para cuidar da sua saúde
19	4	4	que a educação em saúde

The interface also shows a list of corpus files on the left, with "Revista Brasileira de" selected. The search term "saúde" is entered in the search field, and the cluster size is set to 5. The search is sorted by frequency.

Fonte: Elaborada para fins deste artigo.

Levando em consideração os passos metodológicos definidos para a presente pesquisa, os cluster/N-grams que contam com o mínimo de dez ocorrências tem seus resultados discutidos. Para a palavra-chave “saúde”, os clusters/N-grams que estão em concordância com esse padrão de ocorrências são: “profissionais da área da saúde”, “profissional da área da saúde” e “uma unidade básica de saúde”.

Considerando a possibilidade reportada no trabalho de Bowker e Pearson (2002) sobre as variações de um mesmo termo serem associadas a um significado em comum, o primeiro e segundo clusters/N-grams – “profissionais da área da saúde” e “profissional da área da saúde” – podem ser contabilizados de maneira separada, mas compreendidos sob um mesmo significado. A variação entre estes clusters/N-grams corresponde apenas à flexão de número entre “profissionais”, que aparece no plural no primeiro cluster, e “profissional”, que aparece no singular no segundo cluster. As Figuras 21 e 22 abordam as linhas de concordâncias, onde estão localizados estes clusters/N-grams.

FIGURA 21 – Captura de tela do primeiro cluster/N-gram formado à esquerda da palavra-chave “saúde”: “profissionais da área da Saúde”

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 14							
Hit	KWIC						
1		deste cenário, um dos desafios para os <b>profissionais da área da Saúde</b> é buscar alternativas					
2		um Comitê de Juízes composto por nove <b>profissionais da área da Saúde</b> . A escolha desses					
3		. De acordo com a validação realizada pelos <b>profissionais da área da Saúde</b> , o Compasso foi					
4		adequado culturalmente, podendo ser utilizado por <b>profissionais da área da Saúde</b> atuantes em práticas					
5		usuário e um acompanhamento contínuo junto aos <b>profissionais da área da saúde</b> pode vir a					
6		a esta intervenção. Educação realizada por diferentes <b>profissionais da área da saúde</b> Diferentes profissionais de					
7		intervenção educativas contaram com a participação de <b>profissionais da área da Saúde</b> (Enfermeiro, Nutricionista, Fisioterapeuta,					
8		o controle glicêmico. Nesse sentido, cabe aos <b>profissionais da área da Saúde</b> utilizar ferramentas que					
9		e que algumas vezes são negligenciados pelos <b>profissionais da área da Saúde</b> . A efetividade cuidado					
10		não estou querendo muito não (U7). Os <b>profissionais da área da Saúde</b> foram citados como					
11		barreira importante de ser trabalhada, sobretudo pelos <b>profissionais da área da Saúde</b> , os quais devem					
12		diabetes é uma necessidade premente para os <b>profissionais da área da Saúde</b> , os quais podem					
13		do vínculo entre os participantes e os <b>profissionais da área da Saúde</b> . Seguindo a perspectiva					
14		contato e acesso aos saberes de diferentes <b>profissionais da área da saúde</b> . A equipe multiprofissional,					

Fonte: Elaborada para fins deste artigo.

FIGURA 22 – Captura de tela do primeiro cluster/N-gram formado à esquerda da palavra-chave “saúde”: “profissional da área da Saúde”

Hit	KWIC
1	que viabiliza uma comunicação efetiva entre o profissional da área da Saúde e o usuário
2	é uma maneira eficaz de aproximar o profissional da área da Saúde aos principais dificultadores
3	do Compasso poderá oferecer subsídios para o profissional da área da Saúde planejar e implementar
4	o usuário pela prática do autocuidado, o profissional da área da saúde deve atuar como
5	so resultante de corresponsabilização juntamente com o profissional da área da saúde e construído por
6	estratégias educativas. Este processo requer que o profissional da área da saúde e o usuário
7	elaboração de um plano de metas entre profissional da área da saúde e usuário, é
8	quatro e no máximo 12 contatos com um profissional da área da Saúde, totalizando 14 horas de
9	em um curto espaço de tempo o profissional da área da Saúde tenha acesso ao
10	de apoio podem auxiliar. A presença do profissional da área da Saúde nas falas do
11	acompanhamento contínuo, a fim de que o profissional da área da Saúde consiga nortear as

Fonte: Elaborada para fins deste artigo.

A análise das linhas de concordância das Figuras 21 e 22, respectivamente, indica que o pressuposto revelado por Bowker e Pearson (2002) sobre a variação e o significado de palavras análogas encontradas no mesmo corpus pode ser considerado válido para os clusters um e dois formados à esquerda da palavra-chave “saúde”, uma vez que a variação de número em profissionais e profissional não gerou alterações expressivas de significado entre os dois clusters/N-grams. Por esse motivo, o Quadro 14 destaca o grupo nominal que aparece nestes clusters/N-grams, bem como as funções presentes neste grupo.

QUADRO 14 – Os grupos nominais presentes em “profissionais/profissional da área da saúde” e as respectivas funções exercidas na ordem do grupo nominal

<b>1º grupo nominal</b>	profissionais/profissional	da área da saúde
	<b>Ente</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	área	da saúde
	<b>Ente</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Antes de discutir as funções presentes neste grupo nominal, faz-se importante ressaltar que há um segundo grupo nominal dentro de “profissionais/profissional da área da saúde”. Este grupo nominal corresponde a “área da saúde”, em que “área” funciona como Ente e “da saúde” como Qualificador, este último responsável por caracterizar a área

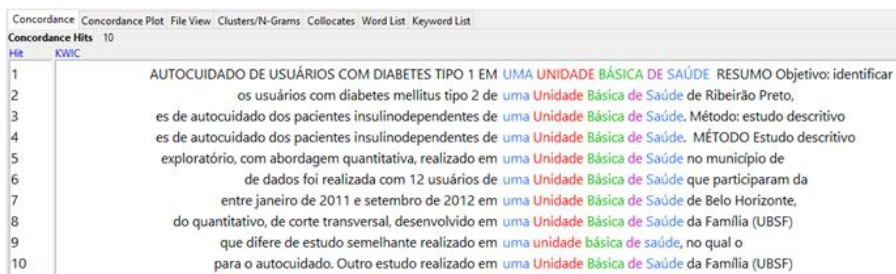


mencionada nos textos do corpus. Entretanto, como esse grupo nominal “área da saúde” está em relação de hipotaxe (dependência) com o grupo nominal “profissionais/profissional da área da saúde” e não há como prever se a frequência mínima desse grupo nominal dependente é dez, como acontece com o grupo nominal “profissionais/profissional da área da saúde”, esse segundo grupo nominal não será explorado neste artigo.

Considerando apenas o grupo nominal “profissionais/profissional da área da saúde”, o Quadro 14 mostra que a frase preposicional “da área da saúde” funciona como Qualificador dos Entes profissionais e profissional, caracterizando-os dentro da área da saúde, não da educação ou advocacia, por exemplo.

O terceiro cluster/N-gram formado à esquerda da palavra-chave “saúde” refere-se a “uma unidade básica de saúde”. Por meio das linhas de concordância, a Figura 23 demonstra o contexto em que este cluster/N-gram aparece no corpus.

FIGURA 23 – Captura de tela do terceiro cluster/N-gram formado à esquerda da palavra-chave “saúde”: “uma unidade básica de saúde”



Fonte: Elaborada para fins deste artigo.

A análise das linhas de concordância da Figura 23 revela que todo o cluster/N-gram “uma unidade básica de saúde” engloba um grupo nominal, cujas funções estão sinalizadas no Quadro 15.

QUADRO 15 – O grupo nominal presente em “uma unidade básica de saúde” e as respectivas funções exercidas na ordem do grupo nominal

Grupo nominal	uma	unidade	básica	de saúde
	Dêítico	Ente	Qualificador	Qualificador

Fonte: Elaborado para fins deste artigo.

O Quadro 15 identifica que o item “unidade” funciona como Ente do grupo nominal em questão. “Básica” e “de saúde” funcionam como Qualificadores e são responsáveis por inserir o Ente “unidade” em uma classe de unidades específica, no caso, aquelas que são básicas e destinadas a lidar com a saúde. Por fim, o determinante não-seletivo e não-específico “um” opera como Dêitico, demonstrando que nos artigos acadêmicos selecionados para o corpus da presente pesquisa as unidades básicas de saúde foram mencionadas de forma ampla, sem o enfoque no nome ou nos detalhes que poderiam identificar quais unidades básicas de saúde os estudos se referiam.

No que diz respeito às buscas pelos clusters/N-grams formados à direita da palavra-chave “saúde”, as Figuras 24 e 25 deixam evidente que as buscas com a extensão de cinco palavras e com a extensão de quatro palavras, respectivamente, não apresentaram os resultados esperados, uma vez que não houve clusters/N-grams com o mínimo de dez ocorrências, quantidade determinada na metodologia deste artigo.

FIGURA 24 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “saúde” com extensão de cinco palavras

Concordance		Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List	
Total No. of Cluster Types				907	Total No. of Cluster Tokens				948
Rank	Freq	Range	Cluster						
1	3	3	saúde coletiva > < ano de publicação						
2	3	1	saúde da família do município						
3	3	1	saúde de ribeirão preto, sp						
4	3	3	saúde de uma cidade do						
5	3	3	saúde e a qualidade de						
6	3	1	saúde no sudeste do brasil						
7	2	1	saúde (lilacs), medline (via ebSCO						
8	2	1	saúde a para orientar as						
9	2	1	saúde adequados, sendo, essas habilidades						
10	2	2	saúde com elevadas taxas de						
11	2	2	saúde como uma ferramenta para						
12	2	1	saúde da família da zona						
13	2	2	saúde da família de um						
14	2	1	saúde da família, em teresina						
15	2	2	saúde deve incluir atividades de						
16	2	1	saúde devem envolver a pessoa						
17	2	2	saúde e aceitação social. as						
18	2	1	saúde e contexto de vida						
19	2	2	saúde e o bem-estar						

Search Term  Words  Case  Regex  N-Grams Cluster Size  
 saúde Advanced Min. 5 Max. 5  
 Start Stop Sort Min. Freq. Min. Range  
 Sort by  Invert Order Search Term Position  
 Sort by Freq  On Left  On Right

Fonte: Elaborada para fins deste artigo.

FIGURA 25 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “saúde” com extensão de quatro palavras

Rank	Freq	Range	Cluster
1	9	7	saúde de belo horizonte
2	4	4	saúde da família (esf
3	4	3	saúde e aceitação social
4	3	3	saúde da família (ubsf
5	3	1	saúde da família do
6	3	1	saúde de ribeirão preto
7	3	3	saúde de uma cidade
8	3	3	saúde e a qualidade
9	3	3	saúde e recursos disponíveis
10	3	1	saúde no sudeste do
11	3	3	saúde para o autocuidado
12	3	1	saúde. as oficinas foram
13	2	1	saúde (lilacs), medline (via
14	2	1	saúde a para orientar
15	2	1	saúde adequados, sendo, essas
16	2	2	saúde com elevadas taxas
17	2	2	saúde como uma ferramenta
18	2	2	saúde da família (csf
19	2	2	saúde da família com

**Search Term**  Words  Case  Regex  N-Grams **Cluster Size** Min. 4 Max. 4  
 saúde Advanced  
 Start Stop Sort  
**Sort by**  Invert Order **Search Term Position** Min. Freq. 1 Min. Range 1  
 Sort by Freq  On Left  On Right

Fonte: Elaborada para fins deste artigo.

Para evitar incoerência com os pressupostos metodológicos estabelecidos para o presente estudo, a busca por clusters/N-grams formados à direita da palavra-chave “saúde” se deu com a extensão de três palavras que resultou em dois clusters/N-grams, com o mínimo de dez ocorrências cada, como mostra a Figura 26 a seguir.

FIGURA 26 – Captura de tela dos clusters/N-grams formados à direita da palavra-chave “saúde”

Concordance		Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List	
Total No. of Cluster Types				755		Total No. of Cluster Tokens			949
Rank	Freq	Range	Cluster						
1	42	15	saúde da família						
2	10	9	saúde e a						
3	9	7	saúde de belo						
4	8	8	saúde e o						
5	5	5	saúde. no entanto						
6	4	2	saúde do idoso						
7	4	3	saúde e aceitação						
8	4	3	saúde para a						
9	4	3	saúde por parte						
10	4	3	saúde, além de						
11	4	4	saúde, bem como						
12	4	4	saúde, o que						
13	4	4	saúde; diabetes mellitus						
14	4	4	saúde> <ano de						
15	3	3	saúde coletiva> <ano						
16	3	2	saúde da população						
17	3	1	saúde de ribeirão						
18	3	3	saúde de uma						
19	3	2	saúde dos idosos						

Search Term	<input checked="" type="checkbox"/> Words	<input type="checkbox"/> Case	<input type="checkbox"/> Regex	<input type="checkbox"/> N-Grams	Cluster Size
saúde	Advanced				Min. 3 Max. 3
Start	Stop	Sort		Min. Freq. Min. Range	
Sort by	<input type="checkbox"/> Invert Order	Search Term Position		1	1
Sort by Freq	<input checked="" type="checkbox"/> On Left	<input type="checkbox"/> On Right			

Fonte: Elaborada para fins deste artigo.

O primeiro cluster/N-gram formado à direita da palavra-chave “saúde”, que apresentou o mínimo de dez ocorrências, corresponde a “saúde da família”. Adiante, a Figura 27 traz as linhas de concordância em que tal cluster/N-gram aparece.

FIGURA 27 – Captura de tela do primeiro cluster/N-gram formado à direita da palavra-chave “saúde”: “saúde da família”



Fonte: Elaborada para fins deste artigo.

A análise das linhas de concordância detalhadas na Figura 27 aponta que o cluster/N-gram “saúde da família” compreende um grupo nominal que está em relação de dependência com outro(s) grupo(s) nominal(is) que varia(m) conforme a linha de concordância examinada, a saber: “os profissionais das equipes de estratégia da saúde da família”, “unidades de estratégia de saúde da família”, “centro de saúde da família”, etc. Entretanto, o cluster/N-gram em questão não abarca essa dependência entre os grupos nominais, porque a busca se deu pela extensão de três

palavras à direita da palavra-chave “saúde”, e a dependência é encontrada à esquerda dessa palavra-chave.

Observando as linhas de concordância da Figura 27 de maneira aprofundada, é possível afirmar que das 42 ocorrências mais da metade (23) aborda a “saúde da família” como uma estratégia de assistência aos usuários do sistema único de saúde do Brasil, que ora pode ser viabilizada por uma equipe/profissionais, ora por uma unidade/centro. Em razão disso, o Quadro 16 detalha as funções que podem ser encontradas nos grupos nominais que consideram a “saúde da família” como equipe ou como unidade, respectivamente.

QUADRO 16 – Os grupos nominais possíveis segundo a análise do cluster/N-gram “saúde da família” e as respectivas funções exercidas na ordem do grupo nominal

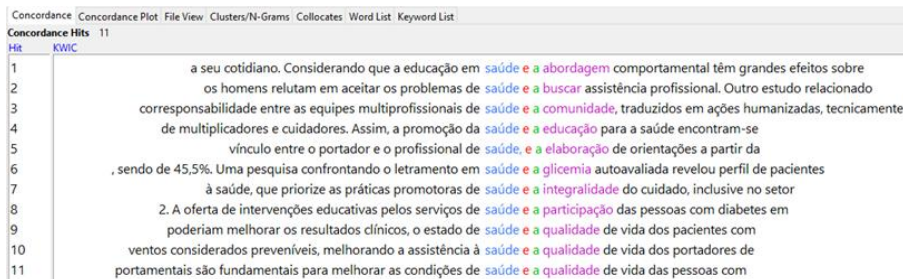
<b>1º grupo nominal</b>	equipes	de saúde	da família
	<b>Ente</b>	<b>Qualificador</b>	<b>Qualificador</b>
<b>2º grupo nominal</b>	unidades	de saúde	da família
	<b>Ente</b>	<b>Qualificador</b>	<b>Qualificador</b>

Fonte: Elaborado para fins deste artigo.

Os dados do Quadro 16 indicam que a principal diferença entre os grupos nominais “equipes de saúde da família” e “unidades de saúde da família” diz respeito à realização do Ente, que no primeiro refere-se às “equipes”; e no segundo às “unidades”. Em comum, os dois grupos nominais têm as duas qualificações realizadas por frases preposicionais “de saúde” e “da família”, respectivamente, ambas responsáveis por identificar o tipo das equipes ou das unidades mencionadas nos textos do corpus utilizado nesta pesquisa.

O segundo e último cluster/N-gram formado à direita da palavra-chave “saúde” com o mínimo de dez ocorrências é “saúde e a”. A princípio, tal cluster/N-gram não apresenta sentido completo, pois termina com “a”, uma palavra da classe nominal > determinante que precisa coocorrer com outra(s) para fazer sentido numa oração e/ou grupo nominal, por exemplo. A seguir, a Figura 28 destaca as linhas de concordância em que este cluster/N-gram aparece.

FIGURA 28 – Captura de tela do segundo cluster/N-gram formado à direita da palavra-chave “saúde”: “saúde e a”



Fonte: Elaborada para fins deste artigo.

O exame das linhas de concordância da Figura 28 indica que não há um padrão de palavras da classe nominal que se repete após o determinante “a”. Além disso, o grupo nominal “saúde”, presente no cluster/N-gram em questão, parece funcionar em conjunto com as preposições de/em, formando uma frase preposicional, responsável por qualificar diferentes Entes, a depender da linha de concordância analisada. Como não há, portanto, um padrão de coocorrências em relação ao cluster/N-gram “saúde e a” para que algum grupo nominal possa ser extraído e investigado, os resultados encontrados para tal cluster/N-gram não foram contemplados no presente artigo.

A seguir, o Quadro 17 resume todos os grupos nominais investigados neste artigo, derivados das buscas por clusters/N-grams à esquerda e à direita das palavras-chave selecionadas.

QUADRO 17 – Resumo dos clusters/N-grams e grupos nominais analisados

Palavra-chave	Clusters/N-grams	Grupo nominal
autocuidado	adesão às práticas de autocuidado	as práticas de autocuidado
	Questionário de atividades de autocuidado	Questionário de atividades de autocuidado
	para as práticas de autocuidado	as práticas de autocuidado
	autocuidado de pacientes com diabetes	autocuidado de pacientes com diabetes



<b>diabetes</b>	para o autocuidado em diabetes	o autocuidado em diabetes	
	de autocuidado com o diabetes	autocuidado com o diabetes	
	práticas de autocuidado em diabetes	práticas de autocuidado em diabetes	
	para o controle do diabetes	o controle do diabetes	
	autocuidado de pacientes com diabetes	autocuidado de pacientes com diabetes	
	com a condição do diabetes	a condição do diabetes	
	diabetes mellitus tipo 2 (DM	diabetes mellitus tipo 2	
	diabetes mellitus tipo 2 em	diabetes mellitus tipo 2	
<b>saúde</b>	profissionais da área da saúde	profissionais da área da saúde	
	profissional da área da saúde	profissional da área da saúde	
	uma unidade básica de saúde	uma unidade básica de saúde	
	saúde da família	equipes de saúde da família	unidades de saúde da família
	saúde e a	-	

Fonte: Elaborado para fins deste artigo.

De acordo com os passos metodológicos adaptados do trabalho de Pearson (1998) para a presente pesquisa, a análise dos dados do Quadro 17 aponta que nem todos os grupos nominais examinados podem ser entendidos como termos técnicos, visto que: i. Alguns compreendem nomes de instrumentos que foram utilizados nas pesquisas reportadas nos artigos acadêmicos, como, “Questionário de atividades de autocuidado” e “de autocuidado com o diabetes”, por exemplo, os quais correspondem ao título de um questionário (Questionário de atividades de autocuidado com o diabetes); ii. Outros indicam o estabelecimento onde ocorreu parte da pesquisa, como, por exemplo, “uma unidade básica de saúde”; iii. Outros ainda dizem respeito ao tipo de profissional e/ou equipe que contribuiu para a realização da pesquisa, como, “profissionais da área da saúde”, “profissional da área da saúde” e “equipes de saúde da família” ou caracterizam o tipo de sujeito que participou da pesquisa, como, por exemplo, usuários/pessoas “com a condição do diabetes”.

Por outro lado, há grupos nominais que podem ser compreendidos como termos técnicos sem que haja a necessidade de uma consulta aprofundada ao contexto onde tal grupo nominal aparece, como, por exemplo, “diabetes mellitus tipo 2”. Isso pode ser explicado pelo fato deste grupo nominal se referir ao nome de uma condição crônica, sendo, conseqüentemente, um termo técnico da área das Ciências da Saúde.



Por último, tem-se os grupos nominais que precisam de uma investigação mais criteriosa para serem validados como termos técnicos. Essa investigação compreende uma sequência de passos estabelecidos a partir da adaptação para o português brasileiro da metodologia proposta no estudo de Pearson (1998). Dentre os grupos nominais detalhados no Quadro 17, os que podem ser caracterizados como termos técnicos, após a análise dos cotextos onde aparecem, são: “adesão às práticas de autocuidado/em diabetes”, “o autocuidado em diabetes” e “autocuidado de pacientes com diabetes”. A análise mostrou que esses grupos nominais podem ser caracterizados como termos técnicos, porque referem-se a um conjunto de estratégias e/ou competências utilizadas pelos profissionais da Saúde para verificar o empoderamento e os índices de autocuidado dos usuários com a condição crônica do Diabetes Mellitus. O Quadro 18 mostra alguns exemplos que apontam para essa espécie de relação de hiponímia prevista em Pearson (1998) como forma de identificar termos técnicos em corpus (cf. item iii da Subseção 2.2 O trabalho de Pearson (1998) na seção 2 Fundamentação Teórica).

QUADRO 18 – Termo técnico coextensivo a grupo nominal: extração e exemplificação das relações de Pearson (1998)

Termo técnico coextensivo a um grupo nominal	Exemplos extraídos do corpus
adesão às práticas de autocuidado/práticas de autocuidado em diabetes	Exemplos de relação extraídas pelo contexto
	<p>O instrumento ESM mede a adesão às práticas de autocuidado do usuário com diabetes. Tem o escore total de oito pontos e abrange questões referentes às atividades de autocuidado, relacionadas à alimentação e à atividade física dos últimos sete dias. Para indicar melhora quanto à adesão às práticas de autocuidado deve-se obter um escore mínimo de cinco pontos.</p>
	<p>instrumento (ESM) para medir – adesão às práticas de autocuidado</p>
	<p>Com relação à adesão às práticas de autocuidado, as medianas de pontuação no GI aumentaram após o processo educativo, e a comparação entre os grupos intervenção e controle em relação a essa variável evidenciou diferença estatisticamente significativa (p=0,026), indicando uma melhora nas práticas de autocuidado.</p>
	<p>adesão às práticas de autocuidado – essa variável</p>
<p>intervenção e 111 do Grupo controle. Foram utilizados os questionários de Autocuidado com o diabetes e Diabetes Empowerment Scale-Short Form para comparação entre grupos na linha de base, assim como entre o antes e depois intragrupo. O nível de significância foi 0,05. Resultados: O grupo intervenção apresentou aumento estatisticamente significativo do escore mediano referente à adesão às práticas de autocuidado em diabetes: (p=0,005) e à escala de empoderamento (p&lt;0,001). Conclusão: A visita domiciliar promoveu à adesão às práticas de autocuidado com diabetes mellitus tipo 2.</p>	
<p>adesão às práticas de autocuidado em diabetes – escore mediano</p>	

<b>autocuidado de pacientes com diabetes</b>	Objetivo: analisar o autocuidado de pacientes com diabetes mellitus tipo 2 na Estratégia Saúde da Família, em Teresina-PI.
	autocuidado de pacientes com diabetes – como único objetivo da pesquisa
	Com exceção da idade e do histórico familiar de DM2, quaisquer outras variáveis podem ser controladas através da mudança no estilo de vida, sendo assim a análise do autocuidado de pacientes com diabetes é de grande importância para regular problemas relacionados com a doença. A educação do paciente na contribuição
	autocuidado de pacientes com diabetes – como outras variáveis

Fonte: Elaborado para fins deste artigo.

O primeiro termo técnico coextensivo ao grupo nominal destacado no Quadro 18 compreende “adesão às práticas de autocuidado/em diabetes”. A segunda coluna desse quadro indica as relações estabelecidas entre esse grupo nominal e o contexto em que ele aparece no corpus, as quais tornam-se relevantes para a caracterização desse grupo nominal como um termo técnico de fato. A primeira relação refere-se ao fato de que a “adesão às práticas de autocuidado/em diabetes” compreende algo passível de ser contabilizado de acordo com as atitudes tomadas pelos usuários com essa condição crônica. Isso pode ser feito por meio de um instrumento, bem como pela análise do escore mediano da pesquisa. Além disso, a “adesão às práticas de autocuidado” é referenciada como “essa variável”, aspecto apontado por Pearson (1998) como indicativo de termo técnico (cf. item iii da Subseção 2.2 O trabalho de Pearson (1998) na seção 2 Fundamentação Teórica).

Ainda com relação aos dados apresentados no Quadro 18, o segundo termo técnico coextensivo ao grupo nominal diz respeito ao “autocuidado de pacientes com diabetes”. Com base nas premissas adotadas no estudo de Pearson (1998) e adaptadas para o presente artigo, esse grupo nominal pode ser caracterizado como termo técnico por duas razões principais: i. Por compreender o único objeto de análise de uma das pesquisas reportadas nos artigos acadêmicos que compõem o corpus da presente pesquisa e ii. Por estabelecer uma relação de hiponímia com “outras variáveis”, demonstrando que o “autocuidado de pacientes com diabetes” pode ser definido como uma variável.

Por fim, a análise do contexto do grupo nominal “o autocuidado em diabetes”, presente no Quadro 17 como candidato a termo técnico,

não apontou resultados significativos que indicassem que se tratava de um termo técnico de fato. Contudo, é importante salientar que tal análise, assim como as demais, foi feita com base nos resultados descritos no estudo de Pearson (1998) e adaptados para o português brasileiro neste artigo, o que pode influenciar nos resultados obtidos no presente trabalho.

## 5 Conclusão

Este artigo buscou mostrar como a linguística de corpus, no que concerne o uso das ferramentas disponíveis no *software* concordanciador AntConc, pode funcionar como um recurso acessível para extração de candidatos a termos técnicos em textos especializados, mesmo quando não se tem para o português brasileiro todos os artifícios mencionados no estudo de Pearson (1998), como, por exemplo, o programa de padrão de correspondência e o anotador morfossintático (CLG tagger) desenvolvidos para aquela pesquisa. Contudo, é relevante destacar que o conhecimento e a aplicação dos pressupostos teóricos da linguística sistêmico-funcional acerca do grupo nominal em português brasileiro foram imprescindíveis para suprimir a ausência desses artifícios detalhados em Pearson (1998).

Considerando a ferramenta de geração de clusters/N-grams do AntConc por meio da seleção de palavras-chave de uma lista gerada pela ferramenta de Keywords do mesmo *software* concordanciador, os resultados obtidos indicam que a coextensividade existente entre cluster/N-gram, termo técnico, grupo nominal e item lexical nem sempre funciona de maneira exata, apesar da presença de pelo menos um grupo nominal dentro de todos os clusters/N-gram gerados. Mas, por se tratar de uma forma semiautomática de extração de termos técnicos coextensivos ao grupo nominal, a utilização dessa ferramenta pode ser avaliada como um recurso útil para a busca por termos técnicos em textos especializados, como artigos acadêmicos, por exemplo.

A análise dos grupos nominais que apareceram nos clusters/N-grams revelou que o tipo de texto, artigo acadêmico, bem como o domínio dos textos, diabetes mellitus, selecionados para constituírem o corpus utilizado no presente estudo podem ter influenciado nos tipos de grupos nominais encontrados. Muitos dos grupos nominais examinados faziam parte de outro grupo nominal maior, como, por exemplo, o grupo nominal “Questionário de atividades de autocuidado com o diabetes” formado

pelos clusters/N-gram “Questionário de atividades de autocuidado” gerado para a palavra-chave “autocuidado” e “de autocuidado com o diabetes” gerado para a palavra-chave “diabetes”. A maior parte dos grupos nominais analisados apresentou também pelo menos uma frase preposicional com função de Qualificador, como “autocuidado de pacientes com diabetes”. Houve também grupos nominais que foram resultado de uma metáfora gramatical (nominalização), por exemplo, “adesão às práticas de autocuidado em diabetes”. Essas características – grupos nominais com muitas funções, constituídos por mais de uma palavra e/ou metafóricos – são frequentemente encontradas nos textos em que a linguagem empregada é especializada, cujo autor e leitor são pares e ambos possuem conhecimento experto na área de domínio do texto.

Para além das adaptações do trabalho Pearson (1998) que se fizeram necessárias no escopo da linguística de corpus, resultados diferentes foram observados no que compreende as distinções entre os sistemas linguísticos do inglês, abordados na pesquisa de Pearson (1998), e do português brasileiro, abordado no presente artigo. A principal diferença refere-se aos determinantes com função de dêitico. Em Pearson (1998), a autora menciona que os candidatos a termos técnicos que estavam sinalizados, ou seja, apresentavam algum artigo definido exercendo a função de dêitico não poderiam ser classificados como candidatos a termo técnico. No presente artigo, entretanto, esse aspecto não pode ser considerado relevante, visto que a presença de dêiticos realizados por artigos definidos não se mostrou um impasse para que determinados grupos nominais pudessem estar em coextensividade aos termos técnicos.

Por outro lado, uma espécie de relação de hiponímia mencionada no estudo de Pearson (1998) como uma das maneiras de caracterizar determinado candidato a termo técnico como técnico de fato (cf. Metodologia – item iii) também foi observada neste artigo, sobretudo, nos contextos dos grupos nominais “adesão às práticas de autocuidado/em diabetes” e “autocuidado de pacientes com diabetes” que apresentaram essa relação com “essa variável” e com “outras variáveis”, respectivamente, e puderam, posteriormente, serem caracterizados como termos técnicos. É válido mencionar que essa espécie de relação de hiponímia sugerida por Pearson (1998) como uma maneira de caracterizar e encontrar um termo técnico em um texto especializado também já foi

reportada no trabalho de Figueredo *et al.* (2019) como uma das maneiras possíveis de caracterizar um item lexical em português brasileiro.

Ainda no campo dos fatores concordantes entre o estudo de Pearson (1998) e a pesquisa detalhada neste artigo encontram-se dois aspectos que podem ser considerados limitantes, mas passíveis de serem investigados em pesquisas futuras. O primeiro deles está relacionado à ideia de consultas a especialistas na área dos textos que constituem o corpus de pesquisa, no caso um profissional da área da Saúde. Esses especialistas funcionariam como uma fonte extralinguística capaz de corroborar se determinado termo poderia ser considerado um termo técnico de fato. O segundo aspecto corresponde ao tamanho do corpus de pesquisa utilizado que pode interferir nos resultados encontrados, ou seja, com um corpus de maiores proporções, a quantidade de termos candidatos a termos técnicos também poderia ser maior, embora a análise manual proposta no presente trabalho poderia tornar-se inviável pela demanda extra de tempo a ser despendido.

### **Declaração de autoria**

Júlia Santos Nunes Rodrigues: compilação do corpus utilizado na pesquisa, anotação da pesquisa, escrita das seções metodologia e resultados, escrita resumo e abstract e revisão do artigo.

Kícila Ferregueti: auxílio na compilação do corpus de referência, escrita da seção de fundamentação teórica, formatação do artigo nos moldes da revista, incluindo a formatação das referências e revisão do artigo.

Adriana S. Pagano: supervisão da pesquisa, escrita das seções introdução e conclusão e revisão do artigo.

### **Referências**

ALMEIDA, G. M. D. B.; CORREIA, M. Terminologia e corpus: relações, métodos e recursos. In: TAGNIN, S. E. O.; VALE, O. A. (org.). *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008. p. 67-90.

ALMEIDA, L. B. Identidade científica da Terminologia. In: \_\_\_\_\_. *Curso básico de Terminologia*. São Paulo: Edusp, 2004. p. 25-96.

ANTHONY, L. AntConc Homepage, *Laurence Anthony Website*, Tokyo, Version 3.5.8, 2019. Disponível em: <https://www.laurenceanthony.net/software/antconc/>. Acesso em: 22, Fevereiro, 2019.

BOWKER, L.; PEARSON, J. *Working with Specialized Language. A Practical Guide to Using Corpora*. 1. ed. London; New York: Routledge, 2002. DOI: <https://doi.org/10.4324/9780203469255>

FERREGUETTI, K. *A Frase preposicional com função de qualificador no grupo nominal: um estudo de equivalentes textuais no par linguístico inglês e português brasileiro*. 2018. 154f. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, 2018.

FIGUEREDO, G. *Uma descrição sistêmico-funcional do grupo nominal em português brasileiro*. 2007. 292f. Dissertação (Mestrado em Estudos Linguísticos) - Faculdade de Letras, Universidade Federal de Minas Gerais, 2007.

FIGUEREDO, G. P. *et al.* O léxico como um recurso linguístico para a produção de significado no texto: um estudo de caso com protocolos de investigação. *Estudos da Língua(gem)*, Vitória da Conquista, v. 17, n. 3, p. 37-59, 2019. DOI: <https://doi.org/10.22481/el.v17i3.5928>

FIGUEREDO, G. P.; PAGANO, A. S.; FERREGUETTI, K. Os sistemas textuais de focalização na organização funcional da gramática do Português Brasileiro. *D.E.L.T.A.*, São Paulo, v. 30, n. 2, p. 309-352, 2014. DOI: <https://doi.org/10.1590/0102-445080334301692532>

FINATTO, M. J. B. Exploração terminológica com apoio informatizado: diálogos entre terminologia e linguística de corpus. *In: LORENTE, M. et al. (org.). Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Barcelona: Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra, 2007. p. 221-230.

HALLIDAY, M. A. K. Categories of the Theory of Grammar. *In: \_\_\_\_\_ . Collected Works of M.A.K. Halliday*. London; New York: Continuum, 1961. p. 37-88.

HALLIDAY, M. A. K. *Language as Social Semiotic. The Social Interpretation of Language and Meaning*. 1. ed. London: Edward Arnold, 1978.

HALLIDAY, M. A. K. *On Grammar*. 1. ed. London; New York: Continuum, 2002.

HALLIDAY, M. A. K.; MATTHIESSEN, C. *Construing Experience as Meaning: A Language Based Approach to Cognition*. London: Cassell, 1999.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. *An Introduction to Functional Grammar*. 3. ed. London: Routledge, 2014. DOI: <https://doi.org/10.4324/9780203431269>

HAO, J. *Construing biology: An Ideational Perspective*. 2015. These (PhD of Linguistics) – Department of Linguistics University of Sydney, Sydney, Sydney, 2015.

PEARSON, J. *Terms in Context*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 1998.

SARDINHA, B. Visão geral da Linguística de Corpus. In: SARDINHA, T. B. *Linguística de Corpus*. São Paulo: Editora Manole, 2004. p. 1-42.

