

VIVIAN MEURER LAMEIRA

**A ANÁLISE FUNCIONAL DE ESTRUTURAS LEXICAIS (*LEXICAL FRAMES*):
DADOS LINGÜÍSTICOS EXTRAÍDOS DE *CORPORA*
PARA SUBSIDIAR O ENSINO DE IFA**

Porto Alegre

2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA DE CONCENTRAÇÃO: ESTUDOS DA LINGUAGEM
LINHA DE PESQUISA: LEXICOGRAFIA, TERMINOLOGIA E TRADUÇÃO:
RELAÇÕES TEXTUAIS

A ANÁLISE FUNCIONAL DE ESTRUTURAS LEXICAIS (*LEXICAL FRAMES*):
DADOS LINGÜÍSTICOS EXTRAÍDOS DE *CORPORA*
PARA SUBSIDIAR O ENSINO DE IFA

VIVIAN MEURER LAMEIRA

ORIENTADORA: PROFA. DRA. ANA ELIZA PEREIRA BOCORNY

Dissertação de Mestrado em Estudos da Linguagem,
apresentada como requisito parcial para a obtenção do
título de Mestra em Estudos da Linguagem pelo
Programa de Pós-Graduação em Letras da
Universidade Federal do Rio Grande do Sul.

Porto Alegre

2022

VIVIAN MEURER LAMEIRA

**A ANÁLISE FUNCIONAL DE ESTRUTURAS LEXICAIS (*LEXICAL FRAMES*):
DADOS LINGÜÍSTICOS EXTRAÍDOS DE *CORPORA*
PARA SUBSIDIAR O ENSINO DE IFA**

Dissertação de Mestrado em Estudos da Linguagem,
apresentada como requisito parcial para a obtenção do
título de Mestra em Estudos da Linguagem pelo
Programa de Pós-Graduação em Letras da
Universidade Federal do Rio Grande do Sul.

BANCA EXAMINADORA

Orientadora: Profa. Dra. Ana Eliza Pereira Bocorny

Profa. Dra. Deise Prina Dutra

Prof. Dr. Gustavo Leal Teixeira

Profa. Dra. Rozane Rodrigues Rebechi

CIP - Catalogação na Publicação

Lameira, Vivian Meurer

A análise funcional de estruturas lexicais (lexical frames): dados linguísticos extraídos de corpora para subsidiar o ensino de IFA / Vivian Meurer Lameira. -- 2022.

232 f.

Orientadora: Ana Eliza Pereira Bocorny.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras, Porto Alegre, BR-RS, 2022.

1. Linguística de Corpus. 2. Estudos sobre Gêneros do Discurso. 3. análise de movimentos retóricos. 4. estruturas lexicais. 5. escrita acadêmica. I. Bocorny, Ana Eliza Pereira, orient. II. Título.

AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul (UFRGS), onde realizei toda a minha formação acadêmica.

À professora Ana Eliza Pereira Bocorny, minha orientadora, por acreditar em mim.

Às professoras Anamaria Kurtz de Souza Welp, Cleci Regina Bevilacqua, Sandra Dias Loguercio, Simone Sarmento e Rozane Rodrigues Rebechi, por contribuírem para a minha formação como pesquisadora e professora.

À minha família, por sempre me apoiar em meus estudos.

Aos meus amigos, alunos e aos meus colegas de trabalho, por tornarem meus dias mais alegres.

À Karol Traslatti Soares e ao Gabriel Teixeira Luvizetto, por me auxiliarem com os dados.

RESUMO

Estruturas Lexicais (*Lexical Frames*) são sequências descontínuas de palavras que formam uma estrutura (*frame*) em torno de lacunas variáveis (*slots*) – por exemplo, *the (aim, purpose, objective) of the* (GRAY; BIBER, 2013). Tais unidades representam blocos formulaicos muito importantes na construção do discurso acadêmico (GRAY; BIBER, 2013). Diversas pesquisas já foram realizadas sobre linguagem formulaica em contextos acadêmicos (BIBER et al., 1999; HYLAND, 2008; CORTES, 2004; CORTES, 2013). Poucos estudos, no entanto, tiveram como foco a análise funcional das ELs em resumos de diferentes áreas de especialidade, a partir de um modelo que combine princípios de duas grandes áreas: os Estudos sobre Gêneros do Discurso e a Linguística de *Corpus*. Nessa direção, este estudo investiga o uso e a distribuição das ELs utilizadas na realização linguística das funções retóricas expressas nas seções de resumos de três áreas do conhecimento: (1) Ciências da Computação e da Informação, (2) Física e (3) Medicina e Ciências da Saúde. Em especial, busca-se a identificação e a análise funcional dos referidos blocos formulaicos extraídos dos *corpora*, em uma abordagem dirigida por *corpus* (*data-driven approach*). Para tanto, foram compilados três *corpora* de resumos escritos em inglês, das áreas-alvo, publicados em periódicos revisados por pares. Cada *corpus*, com 1 milhão de palavras, foi compilado com a ferramenta *AntCorGen* (ANTHONY, 2019) e analisado por meio das ferramentas *Sketch Engine* (KILGARRIFF et al., 2004) e *AntConc 4.0.10* (ANTHONY, 2022). Um total de 717 ELs foram extraídas dos três *corpora* estudados. Destas, 159 são da área das Ciências da Computação e da Informação; 154, da área da Física; e 404, da área da Medicina e das Ciências da Saúde. Quanto ao padrão retórico, foi possível constatar que as seções que são convencionais nos resumos estruturados das áreas-alvo são as mesmas elencadas por Swales e Feak (2009). A observação da amostra de 150 ELs, no que tange às funções retóricas que realizam nos resumos acadêmicos, indicou a existência de duas grandes categorias de unidades multipalavras descontínuas: (i) as ELs transparentes e (ii) as ELs opacas. As ELs transparentes representam as unidades que têm a sua função retórica mais facilmente identificável a partir da observação: (i) dos elementos fixos que constituem a sua estrutura; (ii) das palavras que preenchem as lacunas variáveis; e (iii) dos contextos de ocorrência. Quanto a sua tipologia, as ELs transparentes podem ser divididas em 2 tipos, as ELs transparentes retóricas (ELTR) e as ELs transparentes terminológicas (ELTT). As ELTRs realizam linguisticamente as funções retóricas expressas nos gêneros, particularmente as funções relativas à apresentação dos objetivos do trabalho, estando, por essa razão, mais vinculadas a um movimento ou a uma seção retórica. Tais estruturas apresentam, em sua composição, palavras lexicais que indicam função retórica (por exemplo, *aim, purpose, results*). As ELTTs mais vinculadas às áreas de especialidade realizam a função de referir termos, procedimentos e conceitos consagrados nas áreas especializadas. Quanto à frequência de ocorrência nos *corpora*, as ELTTs são menos frequentes do que as ELTRs, sendo necessários pontos de corte mais baixos para que possamos identificá-las. Essas estruturas apresentam, em sua composição, palavras lexicais que indicam vinculação a uma área de especialidade (por exemplo, *risk, hazard patients*). Sugere-se que os dados obtidos neste estudo sejam usados para subsidiar o ensino de Inglês para Fins Acadêmicos (IFA).

Palavras-chave: Linguística de *Corpus*; Estudos sobre Gêneros do Discurso; análise de movimentos retóricos; estruturas lexicais; escrita acadêmica; recursos pedagógicos direcionados por *corpus*.

ABSTRACT

Lexical Frames are discontinuous sequences of words that form a structure (frame) around variable gaps (slots) – for example, *the (aim, purpose, objective) of the* (GRAY; BIBER, 2013). Lexical Frames (LFs) have great pedagogical value for the production of written academic genres in different areas of expertise. Several studies have already been carried out on formulaic language in academic contexts (BIBER et al., 1999; HYLAND, 2008; CORTES, 2004; CORTES, 2013). Few studies, however, have focused on the functional analysis of LFs in abstracts from different specialized areas, based on a model that combines principles from two major areas: Studies on Discourse Genres and Corpus Linguistics. In this direction, this study investigates the use and distribution of the LFs used in the linguistic realization of the rhetorical functions expressed in the sections of abstracts of three areas of knowledge: (1) Computer and Information Sciences, (2) Physics and (3) Medicine and Health Sciences. In particular, the study seeks to identify and functionally analyze the aforementioned formulaic blocks extracted from the corpora, in a data-driven approach. To this end, three corpora of abstracts, written in English, from the target areas, published in peer-reviewed journals, were compiled. Each corpus, with 1 million words, was compiled with AntCorGen (ANTHONY, 2019) and analyzed using Sketch Engine (Kilgarriff et al., 2004) and AntConc 4.0.10 (ANTHONY, 2022). A total of 717 LFs were extracted from the three corpora studied. Of these, 159 are in the area of Computer and Information Sciences; 154, in the area of Physics; and 404, in the area of Medicine and Health Sciences. As for the rhetorical pattern, it was possible to verify that the sections that are conventional in the structured abstracts of the target areas are the same listed by Swales and Feak (2009). The observation of the sample of 150 LFs, regarding the rhetorical functions they perform in the abstracts, indicated the existence of two major categories of discontinuous multiword units: (i) the transparent LFs and (ii) the opaque LFs. The transparent LFs represent the units that have their rhetorical function more easily identifiable from the observation: (i) of the fixed elements that constitute their structure; (ii) the words that fill in the variable slots; and (iii) the contexts of occurrence. Regarding their typology, the transparent LFs can be divided into 2 types, the rhetorically transparent LFs (RTLTF) and the terminologically transparent LFs (TTLTF). The RTLTFs linguistically perform the rhetorical functions expressed in the genres, especially the functions related to the presentation of the objectives of the work and, for this reason, they are more linked to a rhetorical move or section. They have, in their composition, lexical words that indicate rhetorical function (for example, aim, purpose, results). The TTLTFs, more linked to the specialized areas, perform the function of referring to terms, procedures and concepts established in the specialized domains. As for the frequency of occurrence in the corpora, TTLTFs are less frequent than RTLTFs, requiring lower cut-off points so that they can be extracted from the corpus. They have, in their composition, lexical words that indicate a link to a specialized area (for example, risk, hazard patients). It is suggested that the data obtained in this study be used to support the teaching of English for Academic Purposes (EAP).

Keywords: Corpus Linguistics; Studies on Discourse Genres; analysis of rhetorical movements; lexical structures; academic writing; corpus-driven pedagogical resources.

LISTA DE ILUSTRAÇÕES

Quadro 1 – Estrutura retórica de resumos acadêmicos proposta por Swales e Feak (2009, p. 05).	17
Quadro 2 – Taxonomia funcional de pacotes lexicais do discurso acadêmico elaborada por Biber et al. (2004).	25
Quadro 3 – Taxonomia funcional de pacotes lexicais do discurso acadêmico elaborada por Hyland (2008).	26
Quadro 4 – Comparação de Geluso (2019) dos resultados obtidos por Cortes (2004) e por Hyland (2008), a partir de categorias baseadas em Biber et al. (2004)	27
Quadro 5 – Estrutura retórica de resumos acadêmicos proposta por Swales e Feak (2009, p. 05).	28
Quadro 6 – Critérios de compilação dos corpora de estudo.	53
Quadro 7 – Códigos alfa-numéricos utilizados para nomear os corpora do CODISSAE.	54
Quadro 8 – Perguntas de pesquisa e etapas metodológicas realizadas.	57
Quadro 9 – Resultado das lacunas da estrutura lexical “to + the + of the”.	64
Quadro 10 – Identificação das etiquetas de seção dos resumos estruturados das três áreas de estudo que ocorrem mais de 10 vezes.	81
Quadro 11 – Comparação entre a estrutura retórica proposta por Swales e Feak (2009) e a identificada neste estudo para cada uma das áreas-alvo, por meio das etiquetas de seção.	82
Quadro 12 – pFunções retóricas realizadas pelas ELs transparentes.	92
Quadro 13 – Tipologia e caracterização das ELs funcionais.	96
Tabela 1 – Corpora de estudo em números.	55
Tabela 2 – Critérios para a extração das ELs.	61
Tabela 3 – 50 ELs do corpus das Ciências da Computação e da Informação.	62
Tabela 4 – Percentual de resumos estruturados nos corpora.	68
Tabela 5 – Identificação das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.	68
Tabela 6 – Identificação das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.	72
Tabela 7 – Identificação das etiquetas de seção dos resumos estruturados da área de Física.	73
Tabela 8 – Identificação das etiquetas de seção dos resumos estruturados da área de Física.	75
Tabela 9 – Identificação das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.	76
Tabela 10 – Identificação das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.	78
Tabela 11 – Seções que constituem os resumos estruturados das áreas-alvo.	80
Tabela 12 – Número de N-gram types (sem repetição), N-gram tokens (com repetição).	83
Tabela 13 – ELs comuns às áreas do conhecimento estudadas.	85
Tabela 14 – Número de ELs identificadas como “transparentes” nas áreas estudadas.	92
Gráfico 1 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.	72
Gráfico 2 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Física.	76
Gráfico 3 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.	78
Gráfico 4 – Porcentagem das seções que constituem os resumos estruturados das áreas-alvo.	80
Gráfico 5 – ELs recorrentes nos resumos das áreas-alvo.	84

Figura 1 – Layout inicial do AntGram.	32
Figura 2 – Como iniciar a utilização da AntGram (Version 1.3.0)	33
Figura 3 – Primeiros recursos para delimitar a análise do corpus selecionado no AntGram (Version 1.3.0).	34
Figura 4 – Outros recursos disponíveis para a delimitação de resultados no AntGram (Version 1.3.0).	34
Figura 5 – Layout da ferramenta AntGram (Version 1.3.0) para extração de estruturas lexicais.	35
Figura 6 – Resultados da experiência-teste realizada na ferramenta AntGram (Version 1.3.0).....	36
Figura 7 – Palavras que completam a lacuna 1.	37
Figura 8 – Palavras que completam a lacuna 2.	38
Figura 9 – Ferramenta N-Gram Tool do programa Antconc 4.1.0.....	40
Figura 10 – Etapas analíticas da abordagem top-down.	44
Figura 11 – Etapas analíticas da abordagem bottom-up.....	45
Figura 12 – Linhas de concordância da palavra English.	48
Figura 13 – Linhas de concordância aproximadas da palavra English.....	48
Figura 14 – Word Sketch da palavra English.	49
Figura 15 – Página inicial do programa Antcorgen.....	52
Figura 16 – Composição do corpus CODISSAE.....	54
Figura 17 – Localização da funcionalidade CQL Builder na ferramenta Sketch Engine.....	58
Figura 18 – Sintaxe de busca usada para extrair as etiquetas de seção.	59
Figura 19 – Resultado da busca feita com a sintaxe: <s> [{1,3} [word=":"]].....	60
Figura 20 – Exemplo de onde redigir a EL no SE.....	65
Figura 21 – Linhas de concordância do SE.	66

SUMÁRIO

INTRODUÇÃO	12
1 REVISÃO DA LITERATURA	15
1.1 Estudos sobre Gêneros do Discurso	15
1.2 Linguagem formulaica	18
1.2.1 O objeto deste estudo.....	20
1.2.2 A caracterização do objeto deste estudo.....	22
1.2.3 Análise funcional das ELs	23
1.2.4 Metodologia para a extração de Estruturas Lexicais	29
1.2.5 Ferramentas para a extração automática de ELs.....	31
1.3 Linguística de <i>Corpus</i>	40
1.3.1 O histórico da Linguística de <i>Corpus</i>	41
1.3.2 Definição de <i>corpus</i>	42
1.3.3 Abordagens e metodologias analíticas utilizadas na Linguística de <i>Corpus</i>	43
1.3.4 Linguística de <i>Corpus</i> e o Ensino de Línguas Estrangeiras.....	46
2 PROCEDIMENTOS METODOLÓGICOS	51
2.1 Caracterização da pesquisa	51
2.2 Descrição dos <i>corpora</i> e das ferramentas usadas no estudo	51
2.3 Objetivos e questões de pesquisa	56
2.4 Procedimentos analíticos	56
2.4.1 Pergunta de pesquisa 1: Quais seções retóricas constituem os resumos estruturados das áreas-alvo?	58
2.4.2 Pergunta de pesquisa 2: Quais ELs são recorrentes nos resumos das áreas-alvo?	60
2.4.3 Pergunta de pesquisa 3: Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?	61
2.4.4 Análise dos dados para a identificação da função retórica das ELs	65
3 RESULTADOS	67
3.1 Pergunta de pesquisa 1: Quais seções constituem os resumos estruturados das áreas-alvo?..	67
3.2 Pergunta de pesquisa 2: Quais ELs são recorrentes nos resumos das áreas-alvo?	83
3.3 Pergunta de pesquisa 3: Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?	91
CONSIDERAÇÕES FINAIS	98
REFERÊNCIAS	102
APÊNDICES	109
APÊNDICE 1	109
APÊNDICE 2	110

APÊNDICE 3	111
APÊNDICE 4	112
APÊNDICE 5	125
APÊNDICE 6	137
APÊNDICE 7	146
APÊNDICE 8	151
APÊNDICE 9	156
APÊNDICE 10	161
APÊNDICE 11	175
APÊNDICE 12	188
APÊNDICE 13	222
APÊNDICE 14	232

INTRODUÇÃO

O inglês é a língua franca do mundo acadêmico (MAURANEN; HYNNINEN; RANTA, 2010; MAURANEN, 2011; MAURANEN; HYNNINEN; RANTA, 2016), desempenhando um importante papel na disseminação do conhecimento (JENKINS, 2013). Isso porque pesquisadores que almejam ter seus trabalhos publicados em âmbito internacional deparam com a necessidade de redigi-los em língua estrangeira, a qual é, na maioria dos casos, o inglês. Como afirma Dayrell (2014), no entanto, a linguagem produzida por alunos e jovens pesquisadores em contexto acadêmico é diferente daquela produzida por especialistas. Segundo a autora, pesquisadores iniciantes geralmente não estão cientes das convenções do discurso acadêmico em contexto internacional, em termos tanto de estrutura retórica quanto de padrões léxico-gramaticais.

A motivação para a realização deste estudo, portanto, é pedagógica, assim como a de Swales (1981, 1990) ao desenvolver seu esquema analítico de textos (*move analysis*). No desenvolvimento desse esquema, o autor almejava “ajudar alunos avançados para os quais o inglês não é sua primeira língua a melhorarem sua leitura e escrita de AP em inglês” (MORENO, SWALES, 2018, p. 40)¹. Na esteira de Swales, pretendemos, com os resultados de nossa pesquisa, auxiliar pesquisadores brasileiros a produzirem textos acadêmicos em inglês que tenham potencial para serem publicados em periódicos internacionais.

Assim, nesse contexto, assume grande importância o desenvolvimento de competências relacionadas à compreensão e, em especial, à produção de diferentes gêneros acadêmicos em inglês. Dentre os diversos gêneros utilizados na academia, o resumo acadêmico (RA) é um dos mais relevantes. Com o crescimento das publicações especializadas na segunda metade do século XX, o gênero *resumo acadêmico*, conhecido como *abstract* na língua inglesa, surgiu como uma forma de possibilitar o acesso, de maneira sintetizada, às diversas produções acadêmicas (FREITAS, 2016). Nessa conjuntura, é relevante destacar que a produção científica é principalmente difundida por meio de artigos publicados em periódicos acadêmicos, nos quais os resumos são obrigatórios, integrando, desse modo, esse gênero maior que é o artigo acadêmico.

Desse modo, os resumos servem como portas de entrada (VAN LEEUWEN, 1993) para vários gêneros acadêmicos (como artigos, monografias, dissertações e teses), influenciando a decisão de ler os textos completos que os sucedem e impactando o número de citações. Assim, por exporem, de forma concisa, informações importantes presentes no artigo, os resumos acadêmicos permitem que o

¹ No original: “[...] to help advanced students for whom English is not their first language to improve their reading and writing of RAs in English” (MORENO; SWALES, 2018, p. 40).

leitor tenha acesso rápido ao conteúdo do estudo. Por essa razão, os resumos acadêmicos são “o meio mais eficaz de compartilhar os resultados de pesquisas” (TANKÓ, 2017, p. 42)².

Apesar disso, no âmbito do ensino-aprendizagem de Línguas para Fins Acadêmicos (LFA), ainda é pequeno o número de estudos que tratam dos aspectos pedagógicos relacionados à produção de resumos acadêmicos em inglês a partir de um modelo pedagógico que busca relacionar, à estrutura retórica dos gêneros, elementos linguísticos extraídos de *corpora* (MIZUMOTO et al., 2017; CASAL; KESSLER, 2020). Estudos como os de Charles (2007), Chen e Flowerdew (2018), Cotos, Link e Huffman (2017), ao mesmo tempo que revelam a centralidade desse modelo pedagógico, indicam que há, ainda, muito a ser estudado.

Da mesma forma, um grande número de pesquisas já foi realizado sobre linguagem formulaica em contextos acadêmicos (BIBER et al., 1999; HYLAND, 2008; CORTES, 2004, 2013). Poucos estudos, no entanto, tiveram como foco a análise das estruturas lexicais (doravante, ELs)³ na realização linguística das funções retóricas expressas em seções, movimentos e passos de gêneros acadêmicos em diferentes áreas de especialidade, a partir da combinação da perspectiva e dos princípios da área dos estudos sobre gêneros do discurso e da Linguística de *Corpus*, como sugerem Moreno e Swales (2018) e Gray et al. (2020).

A partir do exposto, este estudo tem por **objetivo geral** realizar a análise funcional retórica das ELs em resumos acadêmicos de três áreas do conhecimento: (i) Ciências da Computação e da Informação, (ii) Física e (iii) Medicina e Ciências da Saúde. Quanto aos **objetivos específicos**, trata-se dos seguintes:

1. Identificar o padrão retórico de resumos estruturados das áreas-alvo.
2. Identificar as ELs recorrentes nos resumos das áreas-alvo.
3. Relacionar as ELs identificadas às funções retóricas realizadas nas seções retóricas dos resumos das áreas-alvo.

Os **objetivos específicos** do presente trabalho estão relacionados às seguintes **questões de pesquisa**:

1. Quais seções constituem os resumos estruturados das áreas-alvo?
2. Quais ELs são recorrentes nos resumos das áreas-alvo?

² No original: “Research article abstracts are the most effective means of sharing research results” (TANKÓ, 2017, p. 42).

³ Neste trabalho, utilizaremos o termo *estruturas lexicais* para nos referirmos aos *lexical frames*, ou seja, “sequências descontínuas em que as palavras formam um ‘quadro’ em torno de uma lacuna variável” (GRAY; BIBER, 2013, p. 109).

3. Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?

De forma a responder às questões de pesquisa propostas, os dados aqui utilizados foram obtidos a partir de um estudo empírico e indutivo de três *corpora* especializados. A análise funcional de estruturas lexicais foi realizada com base nos princípios oriundos de duas grandes áreas do conhecimento: Gêneros do Discurso (GD) e Linguística de *Corpus* (LC). Para tanto, buscou-se a convergência desses dois paradigmas analíticos no sentido de estabelecer a relação função-forma, como sugerem Moreno e Swales (2018) e Gray et al. (2020).

Quanto à organização textual, a presente dissertação é composta por quatro capítulos, além desta introdução.

No primeiro capítulo, dividido em quatro seções, apresenta-se o aporte teórico e situa-se este estudo frente a outros trabalhos encontrados na literatura. Na primeira seção, aborda-se a área dos Gêneros do Discurso (GD), mais especificamente aqueles utilizados em contexto acadêmico. Na segunda seção, à luz da revisão de teorias e conceitos relativos à linguagem formulaica, descrevem-se as ELs, objeto deste estudo. Na terceira seção, revisita-se a Linguística de *Corpus*.

No segundo capítulo, dividido em três seções, caracteriza-se a metodologia utilizada neste estudo. Na primeira seção, descrevem-se os *corpora* aqui utilizados. Na segunda seção, apresentam-se as ferramentas mobilizadas para a compilação e para a extração dos dados linguísticos dos *corpora*. Na terceira seção, delineiam-se as etapas metodológicas desta investigação.

No terceiro e último capítulo, descrevem-se e analisam-se os resultados. Chega-se, assim, a um panorama geral das informações obtidas em relação aos objetivos inicialmente propostos.

Finalmente, apresentam-se as considerações finais, nas quais, a partir da retomada dos objetivos e da metodologia propostos, são expostos, de forma resumida, os dados obtidos, as limitações do estudo e as sugestões para pesquisas futuras.

Ao fim do percurso, sugere-se que os dados obtidos neste estudo, a partir da análise funcional proposta, sejam usados para subsidiar o ensino de Inglês Para Fins Acadêmicos (IFA), sobretudo a construção de aplicações pedagógicas para a produção de resumos acadêmicos das áreas de especialidade focalizadas nesta dissertação.

1 REVISÃO DA LITERATURA

Neste capítulo, dividido em três seções, apresenta-se o aporte teórico e situa-se este estudo frente a outros trabalhos encontrados na literatura. Na primeira seção, aborda-se a área dos Gêneros do Discurso (GD), mais especificamente aqueles utilizados em contexto acadêmico. Na segunda seção, à luz da revisão de teorias e conceitos relativos à linguagem formulaica, descrevem-se as ELs, objeto deste estudo. Na terceira seção, revisita-se a Linguística de *Corpus*.

1.1 Estudos sobre Gêneros do Discurso

De acordo com Bakhtin (1997), gêneros são entidades sócio-discursivas e formas de ação social adotadas para agir sobre o mundo e dizer o mundo simultaneamente, elaborados em variadas esferas da atividade humana e capazes de refletir as condições específicas e as finalidades de cada uma dessas esferas. Bakhtin representa um marco inovador no pensamento linguístico ocidental, cruzando fronteiras na década de 1970. Até então, gêneros haviam sido estudados somente pelas perspectivas da retórica e da literatura.

Tal como a língua materna, aprendemos os gêneros na medida em que vamos ouvindo-os e entrando em contato com eles, que se complexificam com o tempo e com a necessidade. O gênero do discurso calcado em uma interpretação da noção seminal de Bakhtin (1999), portanto, diz respeito a inúmeros textos passíveis de serem reconhecidos por se repetirem de modo razoavelmente fixo nos contextos sociais de identificação, como, a título de exemplo, *abstracts*, teses, dissertações e resenhas acadêmicas. Assim, Bakhtin (1997, p. 279) considera os gêneros do discurso como “tipos relativamente estáveis de enunciados”, produzidos de acordo com as condições específicas de cada campo da comunicação verbal.

Apesar de não haver se baseado em concepções bakhtinianas para elaborar seu pensamento, nota-se que, para Swales, o “outro” também desempenha considerável papel na esfera comunicativa. Segundo o autor, os *gêneros* são classes de eventos comunicativos que têm seus propósitos reconhecidos e partilhados pelos membros mais experientes de uma comunidade discursiva. Os estudos de Swales contribuíram para o desenvolvimento de trabalhos relativos ao ensino-aprendizagem de gêneros acadêmicos. A partir desses estudos, compreende-se que a fluência em práticas de escrita e compreensão passa pela reflexão acerca das características comuns a textos de determinado gênero, de modo que identificar traços formais e funcionais desse gênero é fundamental para a realização eficaz de propósitos comunicativos a ele ligados.

Dentre as contribuições de Swales, uma das mais relevantes para esta pesquisa é o seu modelo de análise de movimentos retóricos (*move analysis*), cujo objetivo é, segundo Moreno e Swales (2018,

p. 40), “investigar a estrutura genérica subjacente de artigos de pesquisa (AP) em termos de movimentos e passos retóricos, com fins pedagógicos”⁴. De acordo com os autores, uma das principais aspirações de pesquisadores da área é a identificação de elementos linguísticos (*linguistic features*) característicos de movimentos retóricos. Swales (1990) investigou a macroestrutura da introdução do gênero *artigo acadêmico* através da análise dos padrões organizacionais e retóricos, chamados por ele de movimentos retóricos (*rhetorical moves* ou simplesmente *moves*). Ao fazer isso, o autor desenvolveu um modelo analítico denominado *Create a Research Space* (CARS). Segundo Swales (2004), o movimento retórico é uma “unidade discursiva ou retórica que desempenha uma função comunicativa coerente no discurso escrito ou falado”⁵ (p. 228-229). Kanoksilapatham (2007, p. 23), por sua vez, refere-se ao movimento retórico como “uma seção de um texto que desempenha uma função comunicativa específica”⁶. Outra unidade discursiva que faz parte do modelo de Swales são os passos (*steps*). Tais unidades, conforme o autor, estão inseridas nos movimentos retóricos, sendo partes menores que cumprem a função comunicativa do movimento ao qual pertencem (SWALES, 1990).

Como pontuam Moreno e Swales (2018), poucos estudos, no âmbito da análise dos movimentos retóricos, são baseados em grandes volumes de textos. A ausência de *corpora* anotados retoricamente e o fato de a identificação das funções comunicativas geralmente ser feita de maneira manual tornam o processo de identificação da estrutura retórica difícil e demorado. Estudos recentes (MIZUMOTO, 2017; GRAY et al., 2020) buscaram formas de apresentar as estruturas retóricas em uma ferramenta de auxílio à produção de artigos acadêmicos. Neste estudo, valemo-nos das etiquetas de seção dos resumos estruturados presentes nos *corpora* analisados, como forma de identificar uma tendência macro-estrutural dos resumos da área em estudo (cf. seção 3.1). Também usamos, como referência para nossas análises, o padrão retórico proposto por Swales e Feak (2009) para resumos acadêmicos (cf. Quadro 1), sem que tais categorias restringissem a observação dos dados.

⁴ No original: “[...] investigate the underlying generic structure of research articles (RAs) in terms of moves-and-steps for pedagogical purposes.” (MORENO; SWALES, 2018, p. 41).

⁵ No original: “[...] discursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse” (SWALES, 2004, p. 228-229).

⁶ No original: “[...] a section of a text that performs a specific communicative function” (KANOKSILAPATHAM, 2007, p. 23).

Quadro 1 – Estrutura retórica de resumos acadêmicos proposta por Swales e Feak (2009, p. 05).

Movimentos retóricos	Etiquetas de seção	Perguntas implícitas
Movimento retórico 1	Contextualização/introdução/situação	- O que sabemos sobre o tópico? - Por que o tópico é importante?
Movimento retórico 2	Apresentação da pesquisa/objetivo	- Sobre o que é o estudo?
Movimento retórico 3	Metodologia/materiais/sujeitos/ procedimentos	- Como o estudo foi feito?
Movimento retórico 4	Resultados/ descobertas	- O que foi descoberto com o estudo?
Movimento retórico 5	Discussão/conclusões/implicações/ recomendações	- Qual o significado dos resultados?

Fonte: Traduzido pela autora a partir de Swales e Feak (2009, p. 05).

Como afirmam Gray *et al.* (2020), estudos sobre gêneros acadêmicos escritos produzidos por especialistas são realizados, fundamentalmente, a partir de duas perspectivas: (i) as investigações que tratam da análise de aspectos linguísticos dos textos (por exemplo, padrões de uso lexicais, fraseológicos, léxico-gramaticais) e (ii) os estudos que têm por foco a identificação da estrutura dos textos, especialmente a partir do modelo de análise da estrutura retórica (*move analysis*) de Swales (1981, 2004). Poucos estudos, como afirmam os autores, tratam da combinação dessas duas perspectivas em uma tentativa de fazer um mapeamento da relação entre elementos linguísticos que realizam funções retóricas expressas nas diferentes partes dos textos. A esse respeito, Moreno e Swales (2018) chamam a atenção para a importância de preencher a lacuna entre forma e função (*function-form gap*) (MORENO, SWALES, p. 41) por meio de estudos que estabeleçam “os tipos mais salientes de itens e padrões textuais que ocorrem em um contexto retórico específico em um artigo de pesquisa” (MORENO, SWALES, 2018, p. 41)⁷.

Alguns estudos recentes – como os de Gray *et al.* (2020), Yoon e Casal (2020) e Lu *et al.* (2021) – buscam estabelecer a referida relação entre função e forma. Gray *et al.* (2020), por exemplo, realizam o mapeamento de forma-função nas seções de artigos de pesquisa (Introdução-Métodos-Resultados-Discussão/Conclusão) em 30 disciplinas. Tal mapeamento, conforme os autores, é possível a partir da convergência de dois paradigmas analíticos: a análise de movimentos retóricos (*move analysis*), que revela a estrutura retórica do gênero estudado, e a análise multidimensional, que permite a modelagem de padrões de variação linguística.

Os estudos de Yoon e Casal (2020), Casal e Kessler (2020) e Lu *et al.* (2021), ao tratarem da relação entre forma e função, observam o uso das ELs em gêneros acadêmicos. Yoon e Casal (2020) por exemplo, examinam o uso de ELs, chamadas pelos autores de *p-frames*, na “construção retórica

⁷ No original: “[...] the most salient types of text items, or patterns, occurring in a specific rhetorical context in an RA” (MORENO; SWALES, 2018, p. 41).

de resumos acadêmicos submetidos a conferências na área da Linguística Aplicada” (YOON, CASAL, 2020, p. 536)⁸. Para fazer a análise da distribuição e da associação das ELs com os movimentos retóricos do gênero estudado, os autores, inicialmente, realizaram a etiquetagem retórica do *corpus*, ou seja, manualmente etiquetaram o *corpus* quanto a seus movimentos e passos retóricos. Em seguida, identificaram as ELs de 5 e 6 palavras, sendo mantidas as unidades que apareciam pelo menos 29 vezes a cada milhão de palavras do *corpus*.

O modelo proposto neste estudo, assim como o de Gray *et al.* (2020), também parte da convergência de dois paradigmas analíticos: a análise de movimentos retóricos (*move analysis*) e a extração de ELs, de forma a mapear a relação entre os referidos elementos linguísticos que realizam funções retóricas expressas nas diferentes seções dos resumos das três áreas estudadas. Busca-se, assim, estabelecer uma relação entre função e forma como proposto por Moreno e Swales (2018).

1.2 Linguagem formulaica

Wray (2002) afirma que uma parte considerável da língua que usamos, em nosso dia a dia, apresenta-se em blocos que podem ser fixos ou semifixos e de que a língua não se constitui, opera e evolui somente através da liberdade do sistema, mas também conta com disposições pré-produzidas, em uso, pelos falantes (CORPAS PASTOR, 1997).

Tais considerações alinham-se à teoria de Sinclair (1991) de que a língua é produzida e compreendida de acordo com a interação de dois princípios: o princípio da idiomatidade e o princípio da livre escolha. Conforme o primeiro princípio, o falante “tem à sua disposição um grande número de frases semi pré-construídas que constituem escolhas únicas, embora possam parecer analisáveis em segmentos” (SINCLAIR, 1991, p. 110)⁹. O segundo princípio estabelece que a “cada lacuna em que uma unidade é completada [...] uma grande variedade de escolhas se abre e a única restrição é a gramaticalidade” (SINCLAIR, 1991, p. 109)¹⁰.

Sinclair (1991) considera o princípio da livre escolha um modelo secundário de olhar para a língua. Entretanto, afirma o autor que esse princípio não pode ser descartado, na medida que “um

⁸ No original: “[...] rhetorical construction of Applied Linguistics Conference Abstracts (CAs)” (YOON; CASAL, 2020, p. 536).

⁹ No original: “[...] has available to him or her a large number of semi- preconstructed phrases that constitute single choices, even though they might appear to be analysed into segments” (Sinclair 1991, p. 110).

¹⁰ No original: “At each point where a unit is completed [...], a large range of choice opens up and the only restraint is grammaticalness” (SINCLAIR, 1991, p. 109).

texto ainda tem muitos pontos de troca onde o modelo da livre escolha entra em jogo” (SINCLAIR, 1991, p. 114)¹¹.

Em contrapartida, Sinclair (1991) observa que o princípio da idiomaticidade ocupa um lugar de mais destaque, dado que um grande número de textos apresenta ocorrência de palavras em padrões comuns e que “a maioria das palavras do dia a dia não têm um sentido ou significados independentes, mas são componentes de um rico repertório de padrões de várias palavras que compõem um texto” (SINCLAIR, 1991, p. 108)¹².

Tendo em vista o exposto, neste trabalho, partimos do pressuposto de que a linguagem é padronizada e de que “a padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente” (SARDINHA, 2004, p. 31). Essa característica da linguagem torna possível analisar as sequências fixas e recorrentes de palavras que ocorrem em diferentes gêneros.

O estudo da padronização encontra importante amparo no campo da Fraseologia. A esse respeito, as primeiras ideias sobre Fraseologia são oriundas das contribuições de Bally (1951); no entanto, foi somente a partir dos anos 1990 que esse campo começou a se constituir como uma área independente (GRANGER; PAQUOT, 2008). O termo *fraseologia* abarca diferentes definições e classificações (CORPAS PASTOR, 1997). Aqui, consideramos a Fraseologia como um subcampo da Linguística Aplicada que se dedica à investigação da padronização e da combinação de palavras (GRANGER; PAQUOT, 2008). As principais análises do campo da Fraseologia geralmente distinguem duas tradições de pesquisa: a abordagem fraseológica (*phraseological approach*) e a abordagem baseada na frequência (*frequency-based approach*) (NESSELHAUF'S, 2004; GRANGER, PAQUOT, 2008).

A abordagem fraseológica, também conhecida como “tradição russa” (COWIE, 1998), teve, como precursores, estudiosos da União Soviética e de alguns países da Europa Oriental (COWIE, 1998). Alguns dos teóricos que se destacaram nessa perspectiva, como Vinogradov (1947) e Amosova (1963), veem a Fraseologia como “um *continuum* ao longo do qual as combinações de palavras estão situadas, com as mais opacas e fixas em uma extremidade e as mais transparentes e variáveis na outra” (GRANGER; PAQUOT, 2008, p. 2)¹³. Nesta abordagem, parte-se de critérios linguísticos para identificar, diferenciar e caracterizar esses diferentes tipos de relações lexicais (GRANGER;

¹¹ No original: “[...] a text still has many switch points where the open-choice model comes into play” (SINCLAIR, 1991, p. 114).

¹² No original: “The model of a highly generalized formal syntax with slots into which fall neat lists of words, is suitable only in rare uses and specialized texts. By far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns. Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up a text” (SINCLAIR, 1991, p. 108).

¹³ No original: “[...] y as a continuum along which word combinations are situated, with the most opaque and fixed ones at one end and the most transparent and variable ones at the other.” (CORPAS PASTOR, 1997, p. 2).

PAQUOT, 2008; COWIE, 1998). Como principais legados, essa tradição estabeleceu a fraseologia enquanto disciplina, bem como apresentou critérios possíveis de serem utilizados como suporte para a análise e a categorização de combinações de palavras (GRANGER; PAQUOT, 2008).

Já a abordagem baseada na frequência (NESSELHAUF'S, 2004), igualmente conhecida como “abordagem distributiva” (EVERT, 2004), está associada ao princípio da idiomaticidade de Sinclair (1991) e à Linguística de *Corpus* (cf. adiante). Tal perspectiva, que teve suas origens no trabalho lexicográfico de Firth (1957), parte dos dados linguísticos em uso para depois identificar as regularidades na língua. Firth (1957, p. 11) introduziu a ideia de *colocações*, relacionando-as com a sua famosa frase “*You shall know a word by the company it keeps*”. Dito de outro modo, para se compreender o significado de uma palavra, é importante analisar quais são as outras que frequentemente estão junto a ela. Nessa abordagem, percebe-se a Fraseologia como “a tendência de as palavras não ocorrerem de forma aleatória, ou conforme regras gramaticais, mas em sequências preferidas” (HUNSTON, 2002, p. 137)¹⁴. Desse modo, essa visão gera um grande número de combinações de palavras sem que essas sejam associadas a categorias linguísticas *a priori* (GRANGER; PAQUOT, 2008).

A partir do exposto e da revisão das particularidades gerais de ambas as perspectivas, decidiu-se que a abordagem baseada na frequência é a mais relevante para se alcançarem os objetivos deste trabalho. Por meio da metodologia da Linguística de *Corpus*, foram geradas combinações de palavras com base na frequência (cf. seção 1.3.4) e não em categorias *a priori*. Dessa forma, parte-se do pressuposto de que todas as combinações de palavras extraídas do *corpus* a partir dos critérios estabelecidos, independentemente da categorização, podem ser consideradas objetos deste estudo.

1.2.1 O objeto deste estudo

A identificação de padrões lexicais, entendidos aqui como combinações de palavras que apresentam certa estabilidade e frequência (HYLAND, 2012), típicos dos gêneros acadêmicos e das áreas de especialidade, tem grande importância pedagógica no âmbito do Inglês para Fins Acadêmicos e do ensino-aprendizagem das linguagens especializadas. Para ensinar escritores menos proficientes a produzir textos conforme os padrões considerados convencionais pela comunidade discursiva da área na qual se inserem, é importante ter conhecimento dos referidos padrões.

Textos especializados comunicam informações específicas por meio de elementos linguísticos (termos, unidades terminológicas, unidades fraseológicas, pacotes lexicais) e não linguísticos

¹⁴ No original: “[...] the tendency of words to occur, not randomly, or even in accordance with grammatical rules only, but in preferred sequences” (HUNSTON, 2002, p. 137).

(gráficos, tabelas, imagens, números). A comunicação via meios linguísticos, segundo o princípio da idiomaticidade de Sinclair (1991), acontece através de unidades formulaicas e não com palavras individuais. As relações que essas sequências formulaicas estabelecem entre si e com outros elementos do texto resultam em uma linguagem que pode ser considerada mais fluente e convencional (HERBST, 2011; HYLAND, 2012; NESSELHAUF, 2005; PAQUOT; GRANGER, 2012; PAWLEY; SYDER, 1983).

De acordo com Wray (2002), uma sequência formulaica é “uma sequência, contínua ou descontínua, de palavras ou outros elementos, que é, ou parece ser pré-fabricada” (WRAY, 2002, p. 9)¹⁵. O estudo acerca das sequências formulaicas de palavras que são regularmente usadas pelos falantes tornou-se uma área de grande interesse dos estudos linguísticos, alavancando diferentes terminologias e metodologias de extração para as sequências tanto contínuas quanto descontínuas.

Sequências contínuas recorrentes de palavras são nomeadas na literatura por Cheng, Greaves e Warren (2006, p. 412) como *n-grams*, ou seja, “palavras contíguas que constituem uma frase, ou um padrão de uso, e que se repetem em um *corpus*”¹⁶. Biber et al. (1999, p. 990) utilizam o termo *lexical bundle* para fazer referência a “Expressões recorrentes, independentemente de sua idiomaticidade e independentemente do seu estado estrutural”¹⁷. Conforme Cortes (2015), os *n-grams* e os pacotes lexicais são identificados através de frequências predeterminadas. Ou seja, conforme critérios preestabelecidos relativos à frequência (número de vezes que a unidade ocorre em um *corpus*), tamanho da unidade (número de palavras que constituem a unidade) e dispersão (número de textos do *corpus* em que determinada unidade ocorre).

Sequências descontínuas recorrentes de palavras são nomeadas por Renouf e Sinclair (1991, p. 128) como estruturas colocacionais (*collocational frameworks*) e consistem em “uma sequência descontínua de duas palavras, posicionadas a uma palavra da outra”¹⁸. O termo *conccgram* é utilizado por Cheng et al. (2006, p. 414) para descreverem “todas as permutações de variação constituinte e variação posicional geradas pela associação de duas ou mais palavras”¹⁹. Já Stubbs (2007) e Römer utilizam *phrase frames*²⁰ ou *p-frames* para denominar unidades multipalavras descontínuas. Biber (2009, p. 294), por sua vez, utiliza o termo *formulaic frames* e o define como “Sequências de múltiplas

¹⁵ No original: “[...] a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated” (WRAY, 2002, p. 9).

¹⁶ No original: “[...] contiguous words that constitute a phrase, or a pattern of use, and that recur in a corpus” (CHENG; GREAVES; WARREN, 2006, p. 412).

¹⁷ No original: “[...] recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (BIBER et al., 1999, p. 990).

¹⁸ No original: “[...] discontinuous sequence of two words, positioned at one word remove from each other” (RENOUF; SINCLAIR, 1991, p. 128)

¹⁹ No original: “all permutations of constituent variation and positional variation generated by the association of two or more words” (CHENG et al., 2006, p. 414).

²⁰ Neste trabalho, o termo *phrase-frames* equivale a *estruturas lexicais*.

palavras com uma lacuna variável interna”²¹. Mais recentemente, Gray e Biber (2013, p. 109) adotaram o termo *lexical frame*, que já havia sido utilizado na literatura por Eeg-Olofsson e Altenberg (1994), e o definem como “sequências descontínuas nas quais as palavras formam uma ‘moldura’ em torno de uma lacuna variável”.

Também é importante destacar que, segundo Renouf e Sinclair (1991), *collocational frameworks* são mais frequentes em textos escritos do que falados. Reforçando tal constatação, Biber (2009) afirma que “a maioria dos padrões na redação acadêmica são estruturas formulaicas que consistem em palavras gramaticais²² invariáveis com uma lacuna variável intermediária que é preenchida por palavras lexicais” (BIBER, 2009, p. 275). A partir das afirmações apresentadas acerca do maior volume de sequências descontínuas de palavras em gêneros acadêmicos escritos e considerando que, neste estudo, analisamos o gênero *resumo acadêmico*, escolhemos como nosso objeto de estudo as estruturas lexicais (ELs), definidas como “sequências descontínuas em que as palavras formam um ‘quadro’ em torno de uma lacuna variável” (GRAY; BIBER, 2013, p. 109).

1.2.2 A caracterização do objeto deste estudo

Conforme mencionado por Gray e Biber (2013, p. 133), Eeg-Olofsson e Altenberg foram, possivelmente, os primeiros a investigar as estruturas lexicais (*lexical frames*). Em um estudo usando o London-Lund *Corpus*, Eeg-Olofsson e Altenberg (1994) observaram que as estruturas (*frames*) mais comuns consistiam de palavras gramaticais (por exemplo, pronomes, preposições, conjunções) que serviam para “ancorar uma palavra (de classe aberta) lexicalmente mais rica²³” (EEG-OLOFSSON, ALTERNNERG, 1994, p. 64). Gray e Biber (2013) também mencionam que, ao comparar o discurso escrito e oral, em especial quanto às diferentes formas como as estruturas lexicais estendidas (*extended lexical phrases*) se apresentam, Biber (2009) explica que padrões fraseológicos podem ser sequências relativamente fixas (*bundles*) ou estruturas descontínuas com lacunas variáveis (*frames*). O autor afirma, ainda, que o discurso acadêmico escrito, como já mencionado (cf. seção 1.2.1) depende principalmente de estruturas lexicais com lacunas variáveis (por exemplo, 1*34, 12*4, 1*3*).

Além disso, os autores pontuam que “as estruturas lexicais mais frequentes têm significados não idiomáticos”²⁴, ou seja, é possível compreender seu significado pelo entendimento das palavras

²¹ No original: “Multiword sequences with an internal variable slot” (BIBER, 2009, p. 294).

²² Neste estudo, palavras gramaticais (*function words*) são entendidas como “*the mortar that binds the text together*”. Palavras gramaticais podem ser, por exemplo, pronomes, artigos, preposições, numerais (BIBER et al., 1999, p. 55).

²³ No original: “[...] anchor a lexically richer (open-class) word.” (EEG-OLOFSSON; ALTERNNERG, 1994, p. 64)

²⁴ No original: “[...] most frequent lexical phrases have non-idiomatic meanings” (GRAY; BIBER, 2013, p. 135).

que as compõem, e que geralmente os registros orais “preferem sequências lexicais compostas por verbos e orações, muitas vezes servindo a funções relacionadas à opinião (*stance-related functions*); registros escritos (especialmente registros escritos acadêmicos) geralmente preferem sequências lexicais compostas por substantivos e frases, muitas vezes com funções referenciais (*referential functions*)”²⁵ (GRAY; BIBER, 2013, p. 135).

Assim, neste estudo, entendemos *Estruturas Lexicais* (ELs) como sequências descontínuas de palavras que se constituem de elementos fixos e de lacunas variáveis (por exemplo, 1*34, 12*4, 1*3*) que apresentam certa estabilidade e recorrência.

1.2.3 Análise funcional das ELs

Pesquisas que tratam da linguagem formulaica, especialmente estudos sobre pacotes lexicais, em geral categorizam essas unidades com relação à sua estrutura e à sua forma (BIBER et al., 2004; CORTES, 2004; HYLAND, 2008). A categorização estrutural de pacotes lexicais proposta por Biber et al. (2004) dá conta de 17 tipos distintos divididos em 3 grandes categorias. Cortes (2004) e Hyland (2008) condensam a proposta de Biber et al. (2004) em um número menor de categorias estruturais.

Ainda que seja possível considerar que ELs podem ser analisadas conforme categorias propostas para pacotes lexicais, em função de semelhanças entre os dois tipos de unidades, Gray e Biber (2013) ressaltam que ELs têm características estruturais próprias que não se encaixam nas categorias de análise propostas para pacotes lexicais. Por essa razão, uma classificação estrutural mais adequada às peculiaridades das ELs é proposta por Gray e Biber (2013). Nessa taxonomia, os autores sugerem três grandes categorias: (i) ELs baseadas em verbos; (ii) ELs “com outras palavras lexicais”²⁶; (iii) ELs com palavras gramaticais. As ELs baseadas em verbos contêm, em sua estrutura, um ou mais verbos que podem ser verbos modais, verbos auxiliares ou verbos principais (por exemplo, *must be * to, was * in the, I * going to, what did * do*). Já as ELs com outras palavras lexicais contêm um ou mais substantivos, adjetivos, advérbios, mas não contêm, em sua estrutura, verbos (por exemplo, *on the * hand, it * necessary to, I * no idea*). Por fim, as ELs com palavras gramaticais contêm, em sua estrutura, apenas palavras gramaticais, como pronomes, artigos,

²⁵ No original: “[...] prefer lexical sequences composed of verbs and clauses, often serving stance-related functions; written registers (especially academic written registers) generally prefer lexical phrases composed of nouns and phrases, often with referential functions” (GRAY; BIBER, 2013, p. 135)

²⁶ No original: “Verb based frames: frame contains one or more modal, auxiliary or main verb (e.g. *must be * to, was * in the, I * going to, what did * do*). 2. Frames with other content words: frame contains one or more nouns, adjectives, or adverbs but no verbs (e.g. *on the * hand, it * necessary to, I * no idea*). 3. Function word frames: frame consists of only function words such as prepositions, determiners, conjunctions, pronouns, complementisers, etc. (e.g. *the * of this, in the * that, as * as you, a * in the*)” (GRAY; BIBER, 2013, p. 122).

preposições, numerais (por exemplo, *the * of this, in the * that, as * as you*) (GRAY; BIBER, 2013, p. 122).

Com relação à análise funcional de pacotes lexicais, Biber et al. (2004) sugerem uma taxonomia com três grandes categorias: *Expressões referenciais (Referential expressions)*, *Expressões organizadoras de discurso (Discourse organizing functions)* e *Expressões de opinião (Stance expressions)*²⁷. Como explica Grondona (2015, p. 29), as três categorias “abarcam as principais funções dos pacotes lexicais”. A categoria de *Expressões de opinião* “abrange os pacotes lexicais que expressam opiniões ou avaliações de certeza que constroem outra proposição”. Já a categoria *Organizadores discursivos* “refere-se a pacotes que refletem relações entre discurso anterior e posterior”. Por fim, a categoria *Expressões referenciais* “engloba pacotes lexicais que fazem referência direta a entidades físicas ou abstratas, ou ao contexto, tanto para identificar a entidade ou para sinalizar um atributo particular da entidade como algo importante”. No Quadro 2, na página seguinte, Grondona (2015, p. 28) apresenta a taxonomia de Biber et al. (2004) juntamente com suas subcategorias e exemplos.

²⁷ Neste trabalho, usamos as traduções de Grondona (2015) para os termos *referential expressions*, *discourse organizing functions* e *Stance expressions*.

Quadro 2 – Taxonomia funcional de pacotes lexicais do discurso acadêmico elaborada por Biber et al. (2004).

I. Expressões de opinião	II. Organizadores discursivos	III. Expressões referenciais
A. Postura epistêmica Pessoais – (<i>I don't know if, I think it was</i>) Impessoais – (<i>are more likely to, the fact that the</i>)	A. Introdução ao tópico/foco (<i>what do you think, let's have a look at</i>)	A. Identificação/foco (<i>that's one of the, one of the most</i>)
B. Atitudinais/postura modal B1. Desejo Pessoais – (<i>if you want to, do you want to</i>) B2. Obrigação/diretivos Pessoais – (<i>I want you to, and you have to</i>) Impessoais – (<i>it is important to, it is necessary to</i>) B3. Intenção/predição Pessoais – (<i>I'm not going to, what we're going to</i>) Impessoais – (<i>It's going to be, going to be the</i>) B4. Habilidade Pessoais – (<i>to be able to, to come up with</i>) Impessoais – (<i>can be used to, it is possible to</i>)	B. Elaboração do tópico/clarificação (<i>has to do with, on the other hand</i>)	B. Imprecisão (<i>or something like that, and things like that</i>)
		C. Especificação de atributos C1. Especificação de quantidade (<i>there's a lot of, greater than or equal</i>) C2. Atributos de enquadramento tangíveis (<i>the size of the, in the form of</i>) C3. Atributos de enquadramento intangíveis (<i>the nature of the, in terms of the</i>)
		D. Referência temporal/de lugar/textual D1. Referência de lugar (<i>the United States and, in the United States</i>) D2. Referência temporal (<i>at the same time, at the time of</i>) D3. Dêitico textual (<i>shown in figure N, as shown in figure</i>) D4. Referência multifuncional (<i>the end of the, the top of the</i>)

Fonte: Grondona (2015, p. 28).

Nessa mesma linha, Gray e Biber (2013, p. 135) sugerem que as ELs podem ter as mesmas funções propostas para os pacotes lexicais na construção do discurso: Expressões referenciais (*Local referential expressions*), Expressões organizadoras de discurso (*Discourse organizing functions*) e Expressões de opinião (*Stance expressions*).

Outra taxonomia proposta para a análise funcional de pacotes lexicais é a apresentada por Hyland (2008). Em um estudo que tratou da análise da forma, da estrutura e das funções de pacotes lexicais de 4 palavras a partir de um *corpus* de 3.5 milhões de palavras, composto de artigos de pesquisa, teses de doutorado e dissertações de mestrado de 4 áreas diferentes, Hyland (2008) propõe três categorias de pacotes lexicais: aqueles direcionados à pesquisa, ao texto e aos participantes²⁸. No Quadro 3, na página seguinte, Grondona (2015, p. 27) apresenta a taxonomia de Hyland (2008) juntamente com suas subcategorias e com seus exemplos.

²⁸ Traduções de Grondona (2015).

Quadro 3 – Taxonomia funcional de pacotes lexicais do discurso acadêmico elaborada por Hyland (2008).

Direcionados à pesquisa	Direcionados ao texto	Direcionados ao participante
Contexto – indicam tempo e lugar (<i>at the beginning of, at the same time, in the present study</i>)	Sinalizadores de transição – estabelecimento de links aditivos ou contrastivos entre elementos (<i>on the other hand, in addition to the, in contrast to the</i>)	Recursos atitudinais – transmitem a opinião e avaliação do autor (<i>are likely to be, may be due to, it is possible that</i>)
Procedimento – (<i>the use of the, the role of the, the purpose of the, the operation of the</i>)	Sinalizadores resultativos – marcam relações inferenciais ou causativas entre elementos (<i>as a result of, it was found that, these results suggest that</i>)	Recursos de envolvimento – referem-se diretamente ao leitor (<i>it should be noted that, as can be seen</i>)
Quantificação – (<i>the magnitude of the, a wide range of, one of the most</i>)	Sinalizadores estruturais – marcadores textuais que organizam porções do discurso ou direcionam o leitor para outra parte do texto (<i>in the present study, in the next section, as shown in fig.</i>)	
Descrição – (<i>the structure of the, the size of the</i>)	Sinalizadores de construção – situam argumentos a partir da especificação da limitação de condições (<i>in the case of, with respect to the, on the basis of, in the presence of, with the exception of</i>)	
Tópico – relacionados ao campo de pesquisa (<i>in the Hong Kong, the currency board system</i>)		

Fonte: Grondona (2015, p. 27).

Grondona (2015, p. 26) explica as categorias propostas por Hyland (2008) para pacotes lexicais como integrando três categorias: (i) direcionados à pesquisa: categoria que “abrange os pacotes lexicais que auxiliam os autores dos textos a estruturar suas atividades e experiências do mundo real”; (ii) direcionados ao texto: categoria que “inclui pacotes que relacionam-se à organização do texto e ao significado de seus elementos como uma mensagem ou argumento”; (iii) direcionados ao participante: categoria que “inclui pacotes focados no autor ou leitor do texto. Cada uma dessas categorias abarca subcategorias, esquematizadas no quadro abaixo”. De acordo com Grondona (2015, p. 26), Hyland (2008) “argumenta que a possibilidade de um pacote lexical ser classificado em mais de uma categoria é pequena e que linhas de concordância foram examinadas para que os contextos dos pacotes pudessem auxiliar na determinação de sua função”.

O processo de categorização das unidades multpalavra, sejam elas pacotes lexicais ou ELs, é bastante complexo e trabalhoso. Os procedimentos metodológicos adotados por Biber et al. (2004) e descritos a seguir integram a abordagem indutiva utilizada pelos autores para conduzirem a análise das funções discursivas de pacotes lexicais. Assim, a partir da análise de linhas de concordância que

contêm unidades a serem classificadas em seu contexto discursivo, os pacotes lexicais foram atribuídos a grupos. A partir desse momento, iniciou-se o processo de determinar as funções discursivas associadas a cada um dos grupos. Como afirmam os autores, essa última etapa metodológica “foi influenciada por estudos anteriores sobre as funções discursivas dos traços linguísticos” (BIBER et al., 2004, p. 383). Embora as etapas metodológicas sejam conduzidas com critério, dúvidas surgem e inconsistências ocorrem. O Quadro 4, a seguir, criado por Geluso (2019), compara os resultados obtidos por Cortes (2004) e por Hyland (2008), a partir de categorias baseadas em Biber et al. (2004). Geluso (2019) destaca em negrito algumas inconsistências nesses resultados. Nota-se, por exemplo, que alguns pacotes lexicais como *in the present study* e *on the basis of* aparecem classificados em mais de uma categoria.

Quadro 4 – Comparação de Geluso (2019) dos resultados obtidos por Cortes (2004) e por Hyland (2008), a partir de categorias baseadas em Biber et al. (2004)

Cortes (2004)			Hyland (2008a)		
Category	Sub-category	Example	Category	Sub-category	Example
Referential	Time	<i>at the beginning of at the end of in the course of</i>	Research-oriented	Location/time	<i>at the beginning of in the present study*</i>
	Place	<i>in the present study at the university of the center of the</i>		Procedure	<i>the role of the the purpose of the the use of the</i>
	Descriptive	<i>the depth of the the length of the the size of the</i>		Description	<i>the structure of the the surface of the the size of the</i>
	Quantifying	<i>a large number of an increase in the the mean of the</i>		Quantification	<i>the magnitude of the a wide range of one of the most</i>
			Topic	<i>the currency board system In the Hong Kong</i>	
Text organizers	Contrast/ Comparison	<i>on the other hand in contrast to the</i>	Text-oriented	Transition signals	<i>on the other hand in contrast to the</i>
	Inferential	<i>as a result of as a consequence of on the basis of the results of the</i>		Resultative signal	<i>as a result of it was found that these results suggest that</i>
	Focus	<i>it is important to it is difficult to* there is no evidence</i>		Structuring signals	<i>in the present study in the next section as shown in the figure</i>
	Framing	<i>in addition to the in the case of in the context of in the presence of</i>		Framing signals	<i>in the case of with respect to the on the basis of in the presence of</i>
Stance	Epistemic-impersonal/ probable-possible	<i>may be due to it is possible that it is likely that</i>	Participant-oriented	Stance features	<i>may be due to it is possible that are likely to be</i>
	Other stance	<i>has been shown to was found to be not appear to be</i>		Engagement features	<i>it should be noted that as can be seen</i>

Fonte: Geluso (2019).

Ao buscar entender as funções que as ELs exercem na construção do discurso, acreditamos que a análise funcional deve, também, dar maior destaque às funções retóricas que as ELs realizam linguisticamente nos textos. Se observarmos a taxonomia proposta por Hyland (2008) para pacotes lexicais, em especial os exemplos apresentados nos quadros de Grondona (2015) e Geluso (2019), nota-se que poucas categorias parecem dar conta de funções retóricas. A categoria *Direcionados à pesquisa – procedimentos*, com o pacote lexical *the purpose of the*, e a categoria *Direcionados ao texto – sinalizadores resultativos*, com o pacote lexical *as a result of*, parecem ser os exemplos mais próximos de categorias que tratam das funções retóricas realizadas.

Nessa direção, sugerimos que existem dois subtipos de análise funcional: uma que trata da análise das funções discursivas que unidades multipalavras exercem nos textos; outra que trata da análise das funções retóricas. Chamamos a primeira de *Análise Funcional Discursiva* e a segunda, de *Análise Funcional Retórica*.

Neste estudo, propomos uma *Análise Funcional Retórica*, ou seja, fazemos a análise das funções retóricas que as ELs realizam linguisticamente nos resumos acadêmicos. O Quadro 5, a seguir, mostra a estrutura retórica de resumos acadêmicos proposta por Swales e Feak (2009), tomada como referência para as nossas análises, juntamente com a estrutura retórica identificada a partir das etiquetas de seção dos resumos estruturados de cada uma das áreas estudadas (cf. seção 2.4.1).

Quadro 5 – Estrutura retórica de resumos acadêmicos proposta por Swales e Feak (2009, p. 05).

Movimento retórico	Etiquetas de seção típicas	Perguntas implícitas
Movimento retórico 1	Contextualização/introdução/situação	- O que sabemos sobre o tópico? - Por que o tópico é importante?
Movimento retórico 2	Apresentar a pesquisa/objetivo	- Sobre o quê é o estudo?
Movimento retórico 3	Metodologia/materiais/sujeitos/procedimentos	- Como o estudo foi feito?
Movimento retórico 4	Resultados/ descobertas	- O que foi descoberto com o estudo?
Movimento retórico 5	Discussão/conclusões/implicações/recomendações	- Qual o significado dos resultados?

Fonte: Traduzido pela autora a partir de Swales e Feak (2009, p. 05).

Para a realização da análise funcional retórica proposta no presente trabalho, recorremos a uma abordagem indutiva semelhante à adotada por Biber et al. (2004). Dessa maneira, a análise das funções retóricas das ELs identificadas em nossos três *corpora* de estudo é realizada a partir dos dados extraídos conforme os critérios estabelecidos para tamanho das unidades, número de lacunas,

frequência mínima e dispersão (cf. seção 2.4.2.1). Uma vez extraídas, as ELs são organizadas em três tabelas, uma para cada *corpus* analisado.

Nas tabelas, são incluídas as ELs, as palavras mais recorrentes que preenchem as lacunas 1 (L1) (até três palavras por lacuna) e lacunas 2 (L2) (até nove palavras por lacuna) e os exemplos nos quais as ELs ocorrem. Inicia-se, nesse momento, o processo de determinar as funções retóricas das ELs extraídas dos *corpora* de estudo (cf. seção 2.4.3). Tal processo é realizado em três etapas: (i) observação dos elementos lexicais que não variam (*frame*) nas ELs; (ii) observação das palavras que preenchem as lacunas; (iii) observação das ELs em contexto.

A partir das observações realizadas, as ELs são atribuídas a grupos conforme a função retórica que realizam nas diferentes seções dos resumos de cada uma das áreas-alvo. Como esclarecido, tal categorização toma como referência o padrão retórico proposto por Swales e Feak (2009) para resumos acadêmicos (cf. Quadro 5) e a estrutura retórica identificada a partir das etiquetas de seção dos resumos estruturados de cada área.

Por fim, quanto à distribuição das unidades linguísticas analisadas ao longo do texto, estudos anteriores – como os de Bocorny et al. (2021) e Freitas e Bocorny (2021) – atestam uma grande concentração de ELs em certos passos e movimentos retóricos específicos, o que pode indicar que algumas dessas unidades retóricas são mais formulaicas que outras. Da mesma forma, Gray e Biber (2013, p. 135) ressaltam que “as estruturas lexicais têm distribuições sistemáticas nos textos e servem como sinais importantes para a estrutura interna do discurso dos registros falados e escritos”²⁹.

1.2.4 Metodologia para a extração de Estruturas Lexicais

Grandes avanços ocorreram nas últimas duas décadas com relação aos tipos e às funções da linguagem formulaica. Há ainda, no entanto, muitos aspectos que precisam ser investigados, o que torna a área, segundo Gray e Biber (2013, p. 145), “promissora para estudos futuros”³⁰. As questões metodológicas que dizem respeito à extração das ELs dos *corpora* certamente estão dentre os principais desafios e focos de dificuldade da área.

A esse propósito, Biber (2009) e Gray e Biber (2013) propõem duas metodologias para a extração de ELs. Inicialmente, a abordagem denominada *bundles-to-frames method* foi sugerida por Biber (2009). Nela, parte-se da identificação de pacotes lexicais (sequências contínuas recorrentes) para a posterior identificação das ELs (sequências não contínuas) associadas. Em Gray e Biber (2013),

²⁹ No original: “Previous research has also shown that lexical phrases have systematic distributions within texts, and serve as important signals to the internal discourse structure of spoken and written registers” (GRAY; BIBER, 2013, p. 135).

³⁰ No original: “[...] a highly promising domain of inquiry for future research” (GRAY; BIBER, 2013, p. 145).

portanto, parte-se do pressuposto de que ELs frequentes estão associadas a pacotes lexicais. Como explicam os autores:

Essa abordagem é baseada na suposição de que todas as estruturas lexicais descontínuas de alta frequência serão associadas a pelo menos uma sequência contínua moderadamente frequente. Por exemplo, a estrutura lexical de frequência extremamente alta *on the * of* está associada ao *lexical bundle on the basis of*. Assim, a abordagem dispensa a análise direta do *corpus* para identificar as milhares de possíveis estruturas lexicais descontínuas, partindo do pressuposto de que o conjunto de sequências lexicais contínuas frequentes representará os padrões descontínuos de alta frequência (GRAY; BIBER, 2013, p. 111)³¹.

Embora a abordagem *bundles-to-frames* (BIBER, 2009) tenha levado a conclusões bastante significativas no que concerne às estruturas formulaicas dos gêneros analisados, percebeu-se que a extração de pacotes lexicais para a posterior identificação das ELs associadas excluía a possibilidade de identificação de ELs não associadas a pacotes lexicais.

Em um estudo mais recente, Gray e Biber (2013, p. 109) declaram que, até o momento, poucos estudos buscaram “identificar o conjunto completo de sequências descontínuas em um *corpus*” (GRAY; BIBER, 2013, p. 109)³². Os autores, então, sugerem uma nova metodologia para a extração das ELs denominada abordagem direta (*direct approach*). Tal abordagem, diferentemente do *bundle-to-frames method* (BIBER, 2009), busca “identificar todas as possíveis sequências descontínuas em um *corpus* e determinar quais são recorrentes” (GRAY E BIBER, 2013, p. 121)³³. Sobre a abordagem direta para a extração de ELs, os autores defendem que

identificar todas as sequências descontínuas possíveis em um *corpus* e, em seguida, determinar quais delas são recorrentes - resulta em uma análise consideravelmente mais completa do que a abordagem usada anteriormente, na qual sequências descontínuas são identificadas a partir de sequências contínuas frequentes (GRAY; BIBER, 2013, p. 121)³⁴.

A abordagem direta conforme sugerido por Gray e Biber (2013), portanto, não utiliza os pacotes lexicais frequentes como ponto de partida para a identificação das ELs. Tal abordagem é um processo que requer uma logística complexa, sendo necessária uma quantidade considerável de dados. Ferramentas computacionais (por exemplo, *AntGram*, *AntConc 4.0*, *ConcGram*) podem facilitar bastante esse processo de identificação de ELs, na medida em que tornam possível a extração direta,

³¹ No original: “This approach is based on the assumption that all high-frequency discontinuous frames will be associated with at least one moderately frequent continuous sequence. For example, the extremely high-frequency frame on the * of is associated with the frequent lexical bundle on the basis of. Thus, the approach foregoes direct analysis of the corpus to identify the thousands of possible discontinuous lexical frames, based on the assumption that the set of frequent continuous lexical sequences will represent the high frequency discontinuous patterns” (GRAY; BIBER, 2013, p. 111).

³² No original: “[...] identify the full set of discontinuous sequences in a corpus” (GRAY; BIBER, 2013, p. 109).

³³ No original: “[...] identifying all possible discontinuous sequences in a corpus, and then determining which of those are recurrent” (GRAY; BIBER, 2013, p. 121).

³⁴ No original: “[...] identifying all possible discontinuous sequences in a corpus, and then determining which of those are recurrent – results in a considerably more complete analysis than the previously used approach in which discontinuous sequences are identified from frequent continuous sequences” (GRAY; BIBER, 2013, p. 121).

de forma automática, a partir de um conjunto de critérios de extração previamente estabelecidos. A próxima seção descreve algumas das ferramentas para a extração automática de ELs.

1.2.5 Ferramentas para a extração automática de ELs

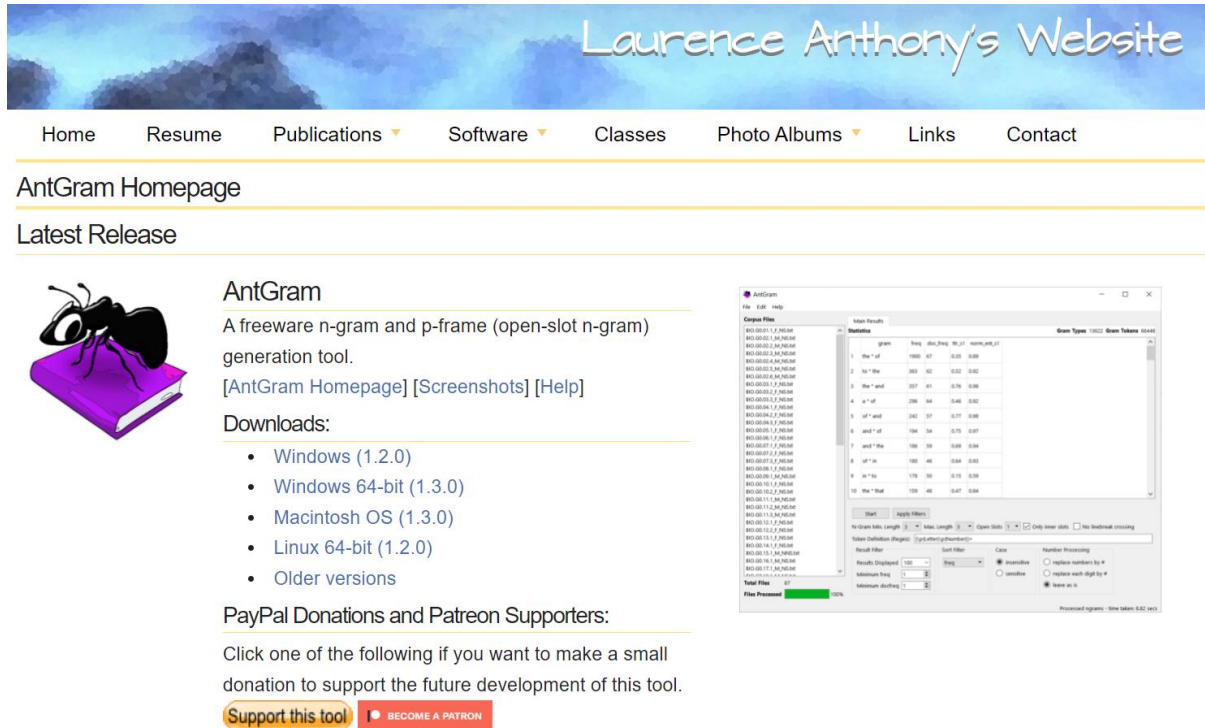
Segundo Anthony (2019), é possível identificar uma predisposição habitual em analisar o vocabulário segundo unidades isoladas (lemas e flemmas). Em estudos recentes, porém, pode-se observar avanços na análise de unidades multipalavras descontínuas a partir de ferramentas desenvolvidas especialmente para esse fim. As ELs, por exemplo, podem ser extraídas por meio de ferramentas especialmente desenvolvidas para esse propósito. Dois exemplos de tais recursos, *AntGram* (ANTHONY, 2019) e *AntConc 4.0.10* (ANTHONY, 2022), são descritos a seguir.

1.2.5.1 *AntGram*

O *AntGram*, desenvolvido por Laurence Anthony em 2019, é uma ferramenta gratuita que extrai *n-grams* e *phrase frames* a partir de um *corpus* de textos (ANTHONY, 2019). Ao utilizar a ferramenta, portanto, é possível inserir um *corpus* que está armazenado no computador e gerar sequências de palavras contínuas (*n-grams*) e descontínuas (*phrase frames*).

Conforme mencionado anteriormente, o *AntGram* é um *software*, ou seja, um programa. Nesse sentido, ao acessar o site do autor, o usuário encontra uma página inicial na qual se apresentam opções gratuitas de *download* do *software* segundo o sistema operacional utilizado pelo usuário, como mostra a Figura 1, na página seguinte.

Figura 1 – Layout inicial do AntGram.




Laurence Anthony's Website

Home Resume Publications Software Classes Photo Albums Links Contact

AntGram Homepage

Latest Release

 **AntGram**
 A freeware n-gram and p-frame (open-slot n-gram) generation tool.
[\[AntGram Homepage\]](#) [\[Screenshots\]](#) [\[Help\]](#)

Downloads:

- Windows (1.2.0)
- Windows 64-bit (1.3.0)
- Macintosh OS (1.3.0)
- Linux 64-bit (1.2.0)
- Older versions

PayPal Donations and Patreon Supporters:

Click one of the following if you want to make a small donation to support the future development of this tool.

[Support this tool](#) [BECOME A PATRON](#)

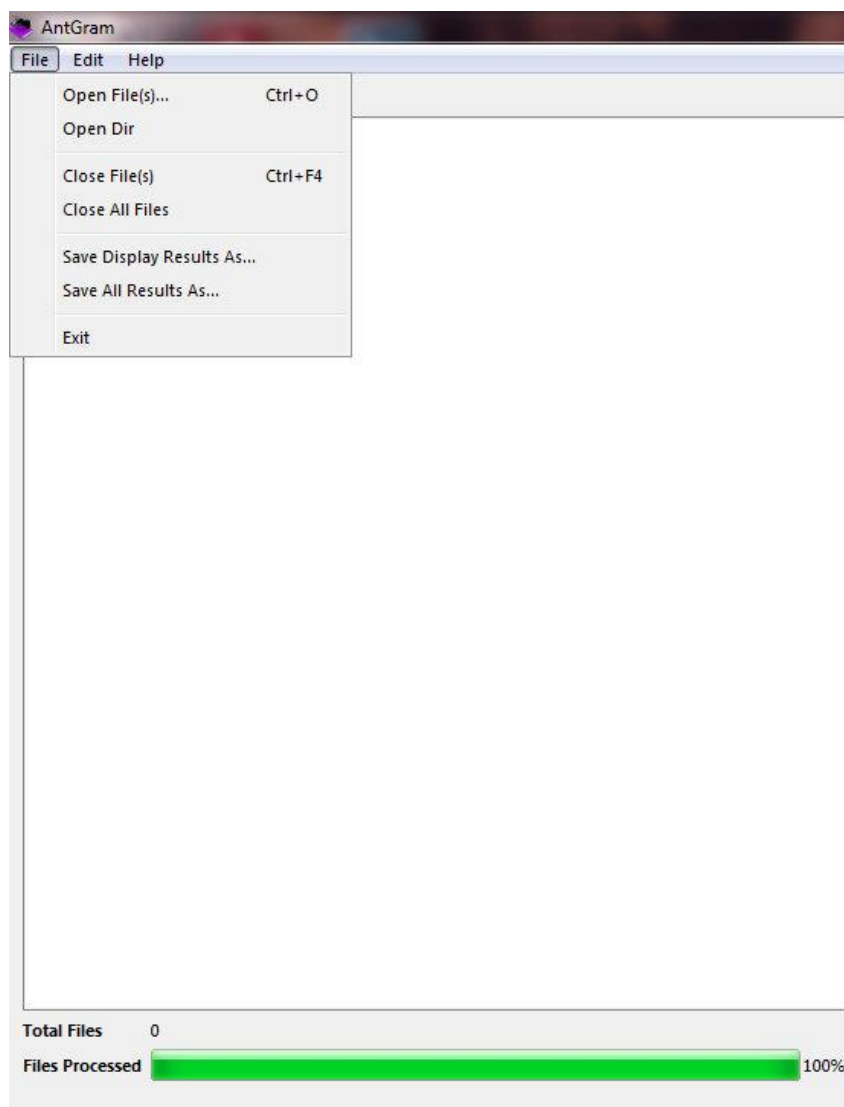
AntGram Main Results

Rank	gram	Freq	Rel_Freq	Rel_Pct	norm_rel_p1
1	the*of	1980	0.07	0.00	0.00
2	the*the	200	0.02	0.00	0.00
3	the*and	207	0.01	0.00	0.00
4	a*of	206	0.04	0.00	0.00
5	of*and	242	0.01	0.00	0.00
6	and*of	104	0.04	0.00	0.00
7	and*the	106	0.04	0.00	0.00
8	of*in	100	0.00	0.00	0.00
9	in*to	176	0.00	0.00	0.00
10	the*but	109	0.00	0.00	0.00

Fonte: Laurence Anthony's Website (2022).

É importante lembrar que, nessa descrição, considera-se a versão *Windows 64-bit* (1.3.0). Sendo assim, inicialmente acessando a opção “*File*”, o usuário pode selecionar o *corpus* armazenado em seu próprio computador (*Open File[s]*) que deseja utilizar na ferramenta (ANTHONY, 2021), (cf. Figura 2), na página seguinte.

Figura 2 – Como iniciar a utilização da *AntGram (Version 1.3.0)*



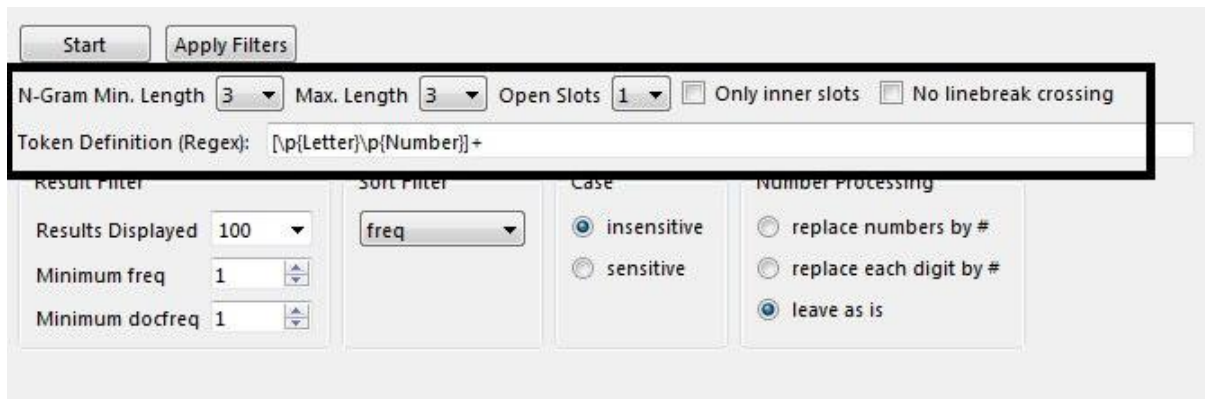
Fonte: *AntGram (Version 1.3.0)* (2021).

Selecionado o *corpus*, o usuário terá de estabelecer os parâmetros que deseja utilizar para extração de *n-grams* (cf. Figura 3), ou seja, o número de palavras da unidade (*n-grams*), o número de lacunas (*slots*), a frequência mínima (*frequency*) e a dispersão (*minimum docfreq*). Quanto ao comprimento, é possível estabelecer o número mínimo (*N.Gram Min. Length*) e máximo (*N.Gram Max. Length*) de *n-grams* que o usuário deseja em uma escala de 1 a 10^{35} (ANTHONY, 2021). Logo ao lado das opções de parâmetro de comprimento, apresentam-se as opções relacionadas às lacunas onde é possível definir o número de lacunas abertas (*Open slots*) em cada *n-gram* de 0 a 10. Também nessa configuração, há a opção de determinar se deseja apenas lacunas internas (*Only inner slots*), isto é, uma opção para que apareçam apenas lacunas dentro dos *n-grams* (por exemplo, *the * of this * is to*); lacunas externas (por exemplo, *the article aims to **); e, ainda, uma outra opção é delimitar

³⁵ Os números de comprimento e de lacunas podem ser menores ou maiores nas diferentes versões do *AntGram*.

se deseja que os *n-grams* apareçam na mesma linha de um arquivo (*no linebreak crossing*) (ANTHONY, 2021).

Figura 3 – Primeiros recursos para delimitar a análise do *corpus* selecionado no *AntGram* (Version 1.3.0).

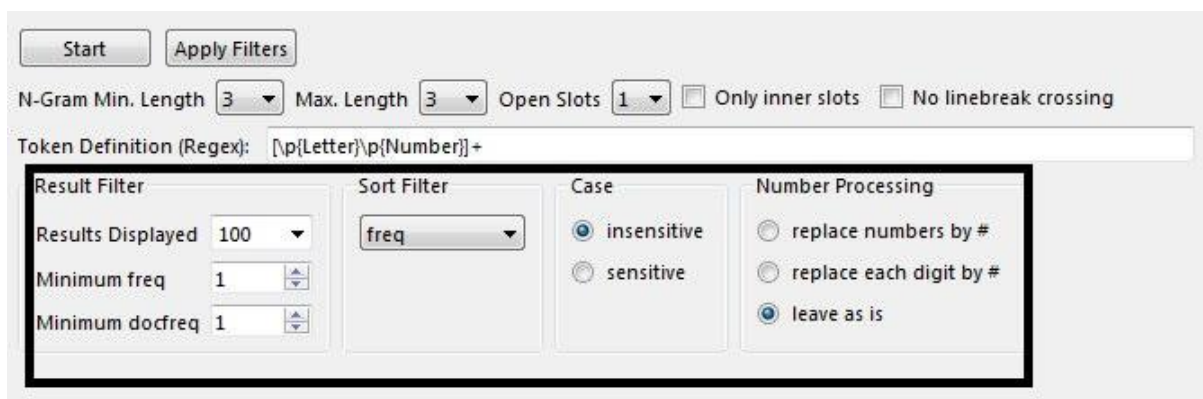


Fonte: *AntGram* (Version 1.3.0) (2021).

Além das opções mencionadas, é igualmente possível ver na figura 3 a opção *Token Definition* (*Regex*), que possibilita, se necessário, alterar a configuração padrão (letras e números Unicode) para uma configuração particular específica que seja necessária ao estudo (ANTHONY, 2021).

Abaixo dessa opção, como ilustra a Figura 4, há o espaço que registra quantos resultados foram obtidos (*Results Displayed*). No entanto, além das delimitações já mencionadas, é preciso estabelecer ou uma frequência mínima (*Min. Freq*), quer dizer, é preciso definir quantas vezes o *n-gram* terá de aparecer no *corpus* para ser validado nos resultados e na frequência do documento (*Min. DocFreq*), ou uma dispersão, isto é, a quantidade de textos em que o *n-gram* terá de aparecer (ANTHONY, 2021).

Figura 4 – Outros recursos disponíveis para a delimitação de resultados no *AntGram* (Version 1.3.0).



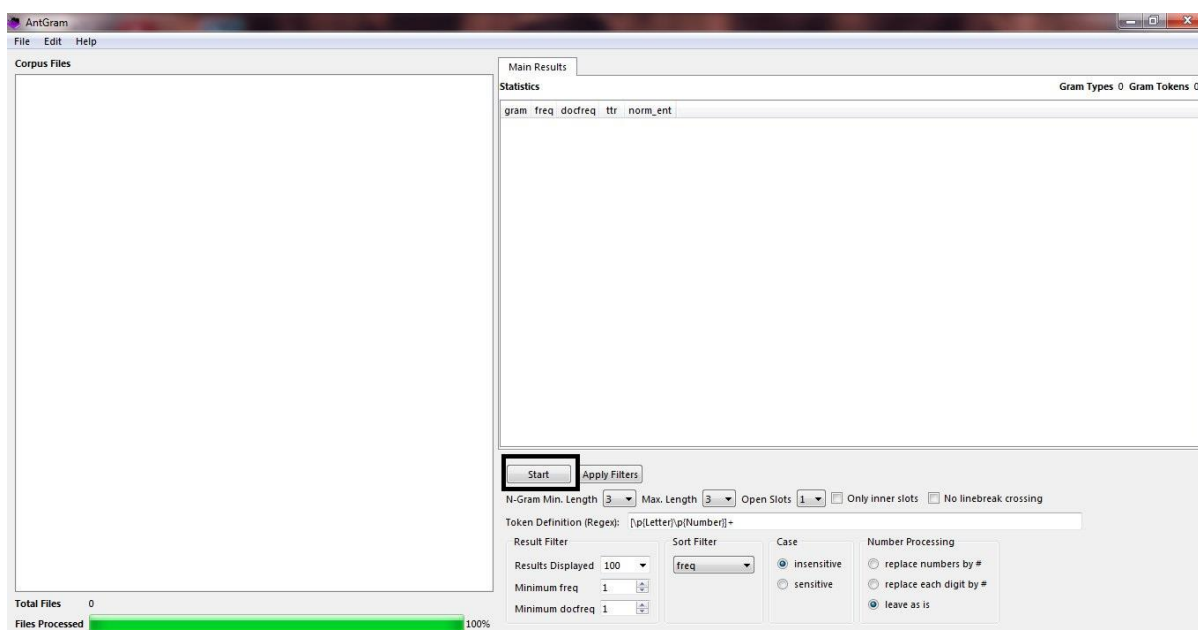
Fonte: *AntGram* (Version 1.3.0) (2021).

No recurso *Sort Filter*, é possível escolher a ordem em que os resultados serão exibidos. Nesse sentido, as ordens podem obedecer às seguintes classificações: (i) segundo a ordem alfabética (*alphabetical*), (ii) segundo a frequência (*freq*) (*n-gram* mais frequente ao menos frequente), (iii) segundo o *n-gram* com valor mais alto de dispersão (*docfreq*) para o menor, (iv) segundo o valor de razão de *tipo-token* (*TTR*) ou, por fim, (v) segundo o valor normalizado de entropia (*norm_ent*) (do maior valor de entropia normalizado para o menor) (ANTHONY, 2021).

Os dois últimos recursos para geração dos resultados disponíveis na ferramenta são as opções *Case* e *Number Processing* (cf. Figura 5). Em *Case*, é possível definir o tratamento que a ferramenta dará às letras maiúsculas ao processar o *corpus*: se desconsiderará a presença das letras maiúsculas (*insensitive*) ou se as considerará (*sensitive*) (ANTHONY, 2021). Já o recurso *Number Processing* diz respeito ao processamento de números pela ferramenta onde é possível escolher que os números permaneçam, que sejam substituídos por # (por exemplo: 100 - #) ou se cada dígito deve ser substituído por # (por exemplo: 100 - ###) (ANTHONY, 2021).

Depois de definidas as configurações, é necessário apertar “*Start*” para que a ferramenta gere os resultados de acordo com as opções escolhidas em cada um dos recursos (cf. Figura 5). De posse dos resultados, é possível criar uma lista *p-frame* ou até mesmo um perfil de *slot p-frame* clicando duas vezes em uma das entradas do *p-frame* (ANTHONY, 2021).

Figura 5 – *Layout* da ferramenta *AntGram* (Version 1.3.0) para extração de estruturas lexicais.



Fonte: *AntGram* (Version 1.3.0) (2021).

Por fim, a ferramenta também possibilita que os resultados encontrados sejam salvos em arquivos no formato *txt*. A esse respeito, há duas formas de salvar os dados: (a) a forma por meio da qual são salvos apenas os resultados que estão sendo apresentados na tela e realizados a partir dos

filtros aplicados (*Save display results as*); (b) a forma por meio da qual são salvos todos os dados encontrados pela ferramenta, inclusive aqueles ocultos devido aos filtros (*Save all results as*) (ANTHONY, 2021).

A Figura 6 mostra um exemplo de extração de *phrase frames* utilizando os seguintes critérios: a) aparecer em no mínimo cinco textos do *corpus*; b) ocorrer no mínimo 10 vezes no *corpus*; c) obter duas lacunas e d) conter entre quatro e seis palavras de comprimento. Essa busca apresentou os resultados mostrados na Figura 6.

Figura 6 – Resultados da experiência-teste realizada na ferramenta *AntGram (Version 1.3.0)*.

The screenshot shows the AntGram software interface. On the left, there is a list of corpus files. The main window displays a table of search results for the query '* the * of: slot 1' and '* the * of: slot 2'. The table has columns for 'gram', 'freq', 'doc_freq', 'ttr_s1', 'norm_ent_s1', 'ttr_s2', and 'norm_ent_s2'. Below the table, there are configuration options for N-Gram Min. Length, Max. Length, Open Slots, and various filters like 'Results Displayed', 'Minimum freq', and 'Case'.

	gram	freq	doc_freq	ttr_s1	norm_ent_s1	ttr_s2	norm_ent_s2
1	* the * of	125	28	0,62	0,89	0,7	0,96
2	the * of *	125	28	0,7	0,96	0,62	0,9
3	** of the	118	28	0,68	0,9	0,73	0,97
4	* of the *	118	28	0,73	0,97	0,75	0,97
5	of the **	116	28	0,76	0,97	0,74	0,96
6	the ** of	60	24	0,77	0,96	0,82	0,97
7	** in the	52	21	0,81	0,96	0,88	0,98
8	* in the *	52	21	0,88	0,98	0,69	0,95
9	in the **	52	21	0,69	0,95	0,65	0,87
10	the ** the	39	21	0,9	0,99	0,49	0,69
11	** to the	31	15	0,81	0,98	0,84	0,98
12	* to the *	31	15	0,84	0,98	0,97	1
13	to the **	31	15	0,97	1	0,81	0,96
14	* we * that	30	16	0,87	0,98	0,33	0,93

Fonte: *AntGram (Version 1.3.0)* (2021).

Como resultado, obtivemos 30 *phrase frames*³⁶. Para cada *phrase frame*, é possível verificar nas Figuras 7 e 8, nas páginas seguintes, as palavras que completam suas lacunas nas opções ao lado de “*Main results*”. A possibilidade de verificar as palavras que completam as lacunas é especialmente

³⁶ Cabe esclarecer que as unidades multipalavras designadas por Anthony (2021) como *phrase frames*, são denominadas, neste estudo, como *estruturas lexicais* (ELs).

importante para este estudo. A observação dessas palavras (*slot fillers*) se constitui em uma das etapas metodológicas descritas neste estudo para a identificação da função retórica das ELs (cf. seções 2.4.3).

Figura 7 – Palavras que completam a lacuna 1.

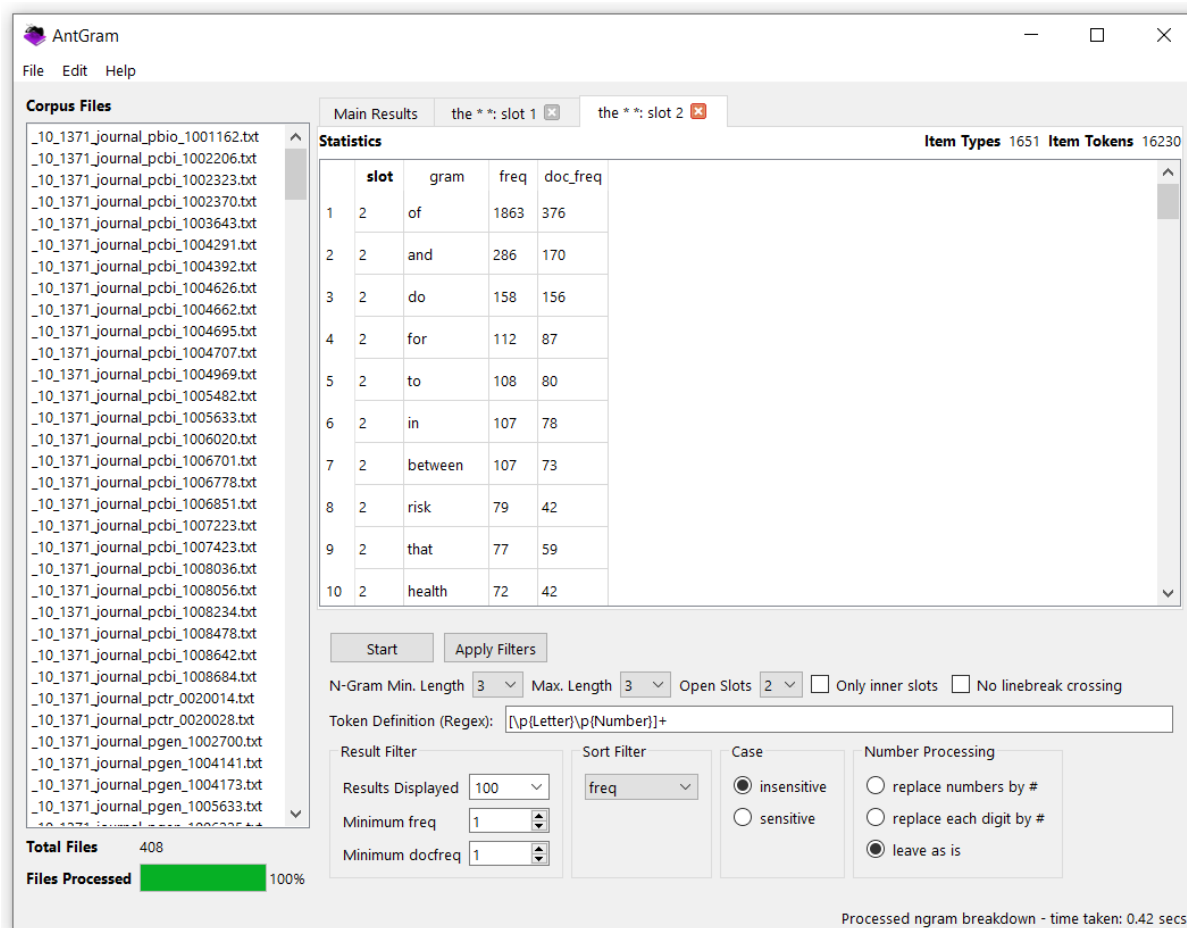
The screenshot shows the AntGram software interface. On the left, a list of corpus files is visible. The main window displays search results for the query 'the **: slot 1'. The 'Statistics' panel shows the following data:

slot	gram	freq	doc_freq
1	researchers	331	158
2	risk	174	90
3	study	116	87
4	most	109	87
5	first	105	64
6	association	81	47
7	primary	70	44
8	disease	70	46
9	number	67	45
10	clinical	55	44

Below the table, there are search filters: N-Gram Min. Length (3), Max. Length (3), Open Slots (2), and Token Definition (Regex): `[\\p{Letter}\\p{Number}]+`. The interface also shows 'Total Files: 408' and 'Files Processed: 100%'. At the bottom right, it indicates 'Processed ngram breakdown - time taken: 0.42 secs'.

Fonte: *AntGram* (Version 1.3.0) (2021).

Figura 8 – Palavras que completam a lacuna 2.



Fonte: *AntGram* (Version 1.3.0) (2021).

Ao final, caso necessário, é possível salvar as informações no computador em formato txt. A ferramenta permite, portanto, a extração de dados de maneira rápida e simples. Nesse sentido, cabe ainda ressaltar que, em 2021, várias funcionalidades do *AntGram* foram migradas para o *AntConc 4.0*, outra ferramenta desenvolvida por Anthony (2021) para a análise textual e a Linguística de *Corpus*, e descrita a seguir.

1.2.5.2 *AntConc 4.0.10*

O *AntConc (Portable)* (4.1.0), lançado por Laurence Anthony (2022), corresponde a um programa utilizado para a realização de pesquisas em Linguística de *Corpus*, bem como para a inserção de métodos de *corpus* e para o auxílio no ensino-aprendizagem de idiomas baseado em dados (ANTHONY, 2022). O programa é executável nos mais diferentes sistemas operacionais e emprega, de forma implícita, o SQLite como banco de dados (ANTHONY, 2022).

O *AntConc* apresenta, em sua interface, um total de oito ferramentas para o manejo dos dados, quais sejam: *KWIC (Key-Word-In-Context) Tool*, *Plot Tool*, *File Tool*, *Cluster Tool*, *N-Gram Tool*, *Collocate Tool*, *Word List* e *Keyword List*. Tais recursos são brevemente descritos a seguir, a partir das informações apresentadas por Anthony (2022):

1) *KWIC*: apresenta os resultados a partir de um modelo de concordância, ou seja, permite a visualização de palavras e frases comuns em um determinado *corpus* de textos.

2) *Plot Tool*: organiza os resultados de concordância em uma linha vertical, o que possibilita a visualização da posição dessas concordâncias nos textos individuais do *corpus*.

3) *File Tool*: exibe o conteúdo de textos individuais, permitindo ao usuário uma observação mais detalhada dos resultados.

4) *Cluster Tool*: designa os padrões de palavras próximas segundo a circunstância de pesquisa.

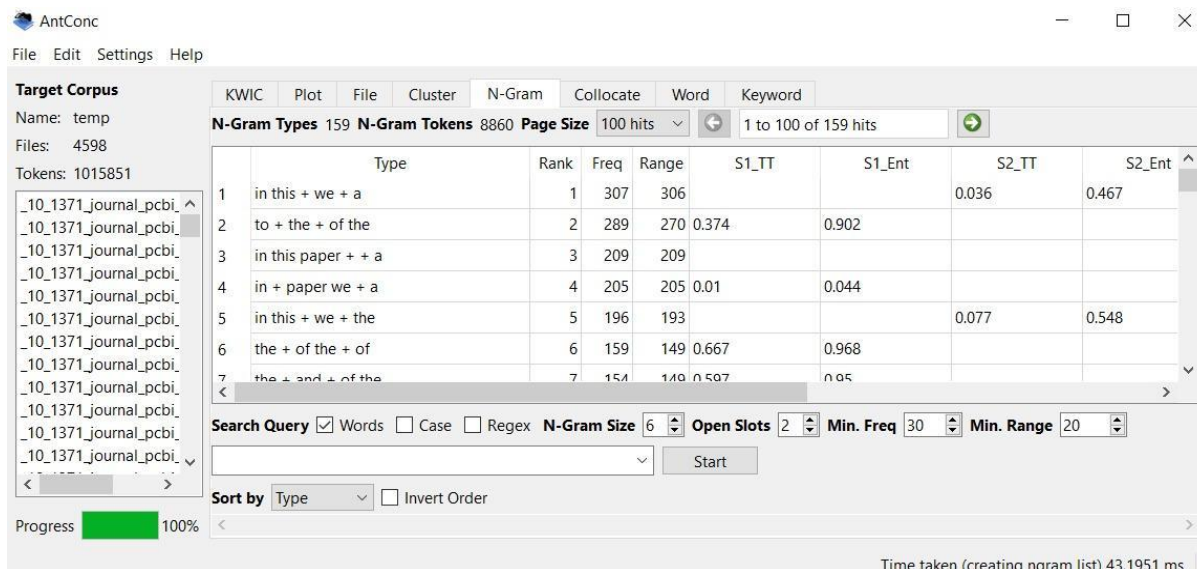
5) *Collocate Tool*: exibe as palavras que co-ocorrem com outras palavras no *corpus*.

6) *Word List*: contabiliza todas as palavras do *corpus* e as organiza de forma ordenada, facilitando a observação de quais palavras são mais frequentes no *corpus* analisado.

7) *Keyword List*: tem por função mostrar as palavras com uma frequência atípica no *corpus* de destino quando comparadas com *corpus* de referência através de uma medida estatística. Conforme Anthony (2022), alterando-se as configurações, também é possível averiguar as palavras com uma pequena frequência no *corpus* de destino considerando-se o *corpus* de referência.

Por fim, a oitava ferramenta é aquela utilizada neste trabalho, a *N-Gram Tool* (cf. Figura 9). Essa ferramenta pesquisa, em todo o *corpus*, os *clusters* de tamanho 'N', isto é, do tamanho configurado (por exemplo, *clusters* de 2 ou 3 palavras). Sendo assim, a ferramenta exibe ao usuário as expressões mais comuns no *corpus* analisado a partir das configurações realizadas baseadas nos critérios estabelecidos no estudo.

Figura 9 – Ferramenta *N-Gram Tool* do programa *Antconc 4.1.0*.



Fonte: *AntConc 4.1.0* (2022).

Como é possível observar na Figura 9, na página seguinte, a *N-Gram Tool* possibilita a seleção dos seguintes parâmetros: (i) no *N-Gram Size*, o comprimento da estrutura desejada, ou seja, quantas palavras deve ter a estrutura a ser pesquisada; (ii) no *Open Slots*, o número de lacunas que se deseja na estrutura a ser extraída; (iii) no *Min. Freq*, a frequência mínima necessária para que a estrutura seja extraída; (iv) no *Min. Range*, a dispersão, ou seja, o número de textos nos quais a EL deve aparecer para que seja extraída. .

Na Figura 9, nas ELs que aparecem na coluna *Type* é possível perceber que os sinais de mais (+) representam as lacunas (*slots*) nas estruturas lexicais. Em outras palavras, ao clicar-se em cada um desses itens (1, 2, 3, 4, 5) da estrutura, será possível visualizar uma lista de linhas de concordância nas quais aparecem as palavras que podem preencher as lacunas.

1.3 Linguística de *Corpus*

A Linguística de *Corpus* dedica-se à coleta e à análise de *corpora*, “conjuntos de dados linguísticos textuais coletados criteriosamente” (SARDINHA, 2004, p. 3), com o intuito de descrever uma língua ou uma variedade linguística. Desse modo, empenha-se em explorar a língua através de evidências empíricas obtidas por computador (SARDINHA, 2004). Caracteriza-se, pois, como uma abordagem empírica de estudo da língua que realiza generalizações lexicais, gramaticais e semânticas a partir da observação de dados autênticos (STUBBS, 2004).

McEnery e Wilson (1996) referem-se a Linguística de *Corpus* como o estudo da língua a partir de exemplos oriundos do uso desta na vida real. Nessa perspectiva, a Linguística de *Corpus* também se define como um conjunto de procedimentos utilizados para se estudarem dados linguísticos em grande escala (MCENERY e HARDIE, 2012). Na mesma linha, Biber et al. (1998) consideram a Linguística de *Corpus* como uma abordagem de pesquisa. Os autores afirmam, ainda, que a Linguística de *Corpus* pode (também) ser considerada como uma abordagem metodológica, tendo-se em vista a sua característica empírica, o conjunto de textos compilados segundo critérios pré-estabelecidos e a demanda de técnicas quantitativas e qualitativas de análise (BIBER et al., 1998). Isso posto, Biber et al. (1998) pontuam que, através da Linguística de *Corpus*, podem-se obter resultados profusos e válidos que não seriam possíveis de outra forma.

Além disso, essa abordagem de pesquisa linguística permite observar a sistematização da variação da linguagem e possibilita descrevê-la por meio de métodos empíricos e quantitativos (BIBER, 2015). Dessa forma, o autor afirma que, através da qualidade dos estudos de Linguística de *Corpus*, foi possível que linguistas ampliassem também as suas perspectivas e indagações de pesquisa.

1.3.1 O histórico da Linguística de *Corpus*

Ao fim dos anos 1950, Noam Chomsky provocou uma quebra no paradigma da linguística da época com seu livro *Estruturas sintáticas*: o empirismo deu lugar ao racionalismo, inaugurando a linguística gerativa (SARDINHA, 2004). Essa ruptura ocorreu justamente na época em que os estudos baseados em *corpus* eram questionados devido ao volume de textos analisados e à qualidade de processamento, realizado de forma manual.

A concepção de *corpus* foi desenvolvida durante a década de 1960. Nesse período, Johnson, no âmbito dos estudos do léxico em língua inglesa, observou frases de grandes estudiosos para analisar o significado das palavras em uso (BONELLI, 2010). Nesse mesmo período, segundo Bonelli (2010), linguistas de campo registravam línguas, geralmente faladas, em diversas comunidades pelo mundo. A partir dos anos 1980, com o advento dos computadores, uma série de estudos descritivos da língua tornaram-se possíveis tendo como fundamentação os princípios da Linguística de *Corpus*.

1.3.2 Definição de *corpus*

Sabe-se que a Linguística de *Corpus* estuda e analisa conjuntos de dados linguísticos textuais (SARDINHA, 2004) que se configuram em um *corpus* (singular) ou vários *corpora* (plural), todavia nem toda coletânea de dados pode ser chamada de *corpus*.

Sanchez (1995) define um *corpus* a partir de sua origem, de seu objetivo, de sua composição, de sua formatação, de sua representatividade e de sua extensão. Assim, conforme o autor e a partir do advento dos computadores, um *corpus* pode ser caracterizado como (i) uma coleção de dados linguísticos (do uso oral ou escrito da língua, ou de ambos), (ii) sistematizados deliberada e representativamente em relação à totalidade do uso linguístico ou de algum de seus domínios, (iii) organizados de maneira que possam ser processados por computador para viabilizar resultados pertinentes e diversos para estudo e descrição (SANCHEZ, 1995).

Para compreender-se a representatividade de um *corpus*, é preciso conceber a linguagem como um sistema probabilístico no qual algumas ocorrências são mais frequentes que as outras (SARDINHA, 2004). Ainda segundo Sardinha (2004), em relação ao léxico e ao sentido das palavras, é possível distinguir, no caso do primeiro, as palavras de maior e menor frequência e, no caso do segundo, os sentidos mais ou menos frequentes. Quanto maior o *corpus*, maior será a probabilidade de ocorrerem palavras de baixa frequência e, também, de aparecerem sentidos mais raros, mesmo em palavras que ocorrem em alta frequência em um *corpus* (SARDINHA, 2004).

Nessa direção, essencialmente, um *corpus* é considerado como representativo da linguagem, seja de uma língua ou de uma variedade dela (SARDINHA, 2004). De acordo com Sardinha (2004), a extensão do *corpus* é uma propriedade naturalmente relacionada à representatividade, isto é, quanto maior o *corpus*, mais representativo ele é.

Biber (2010) menciona duas características que considera fundamentais para legitimar-se a representatividade do *corpus*: tamanho e composição. O tamanho, para representar com qualidade a distribuição dos mecanismos linguísticos e a composição, pois a seleção dos textos para a formação do *corpus* deve ter o objetivo de representar os registros da porção da linguagem que é centro do estudo (BIBER, 2010).

A extensão do *corpus* configura-se, pois, como um elemento importante na representatividade deste conjunto de textos. Nesse âmbito, de acordo com Sardinha (2004) pode-se estabelecer três perspectivas: a impressionística, a histórica e a estatística. A perspectiva impressionística considera observações provenientes da criação e da exploração de *corpora* (SARDINHA, 2004). Já a perspectiva histórica acompanha e se refere aos *corpora* realmente utilizados pela comunidade. Por fim, a perspectiva estatística assenta-se em teorias estatísticas (SARDINHA, 2004).

Já a especificidade do *corpus* deve-se à compreensão de que não é possível abordar toda a linguagem existente de um idioma, o que caracterizaria a representatividade total de um *corpus*. Desse modo, conforme Sardinha (2004), pode-se optar por circunscrever o *corpus* a um autor ou a uma variedade da língua, como, por exemplo, um gênero textual. Essa última possibilita que o *corpus* seja ainda mais específico: “[...] uma variedade específica da linguagem demonstra maior padronização e conseqüente menor variação no nível do léxico, da gramática, do discurso” (SARDINHA, 2004, p. 28).

Por fim, a adequação do *corpus* também integra os aspectos fundamentais na composição do *corpus*. Nesse sentido, é importante compreender que, mesmo configurado como representativo, um *corpus* tem limites definidos.

Sendo assim, “Além de representativo, o *corpus* deve ser adequado aos interesses do pesquisador, que deve ter uma questão a investigar para a qual necessite de um *corpus* específico” (SARDINHA, 2004, p. 29). Dessa maneira, o *corpus* e a questão de pesquisa devem estar alinhados com os objetivos da pesquisa.

1.3.3 Abordagens e metodologias analíticas utilizadas na Linguística de *Corpus*

Biber (2012) diferencia estudos baseados em *corpus* (*corpus-based studies*) de estudos direcionados por *corpus* (*corpus-driven studies*). O primeiro tipo de estudos tem como principal objetivo de pesquisa “analisar os padrões sistemáticos de variação e uso das características linguísticas pré-definidas” (BIBER, 2012, p. 162), enquanto o segundo tipo de estudos busca uma perspectiva de pesquisa na qual “as próprias construções linguísticas emergem da análise de um *corpus*” (BIBER, 2012, p. 162). Assim, a abordagem direcionada por *corpus* (*corpus-driven*) difere da pesquisa baseada em *corpus* (*corpus-based*), na medida em que faz “o mínimo de afirmações a priori em relação às características linguísticas que devem ser empregadas na análise de um *corpus*” (BIBER, 2012, p. 162).

Biber et al. (2007) descreve duas perspectivas principais na realização de pesquisas que tratam de características linguísticas de textos e discursos: a primeira, chamada de *bottom-up*, diz respeito aos estudos realizados com *corpora* relativos a “distribuição e funções de elementos linguísticos”; a outra, chamada de *top-down*, de caráter discursivo, tem como foco “a organização interna dos textos”. Enquanto a primeira perspectiva é quantitativa e, em estudos recentes, realizada com grandes volumes de textos compilados em *corpora*, a segunda é geralmente qualitativa e realizada com um pequeno número de textos. Surpreendentemente, como afirmam Biber et al. (2007, p. 10), “poucos estudos

tentaram combinar essas duas perspectivas de pesquisa”³⁷. Como já mencionado (cf. introdução), o pequeno número de estudos que realizam tal combinação é igualmente destacado por Moreno e Swales (2018) e por Gray et al. (2020).

Biber et al. (2007) descreve em detalhes as etapas analíticas das abordagens *top-down* e *bottom-up*. Preliminarmente, é preciso ressaltar que, para obterem-se descrições gerais com base em *corpus* da estrutura do discurso, é necessário observar algumas etapas, como a determinação das categorias comunicativas, a segmentação, a classificação, a análise linguística das unidades, a descrição linguística das categorias do discurso, a estrutura do texto e as tendências de organização do discurso (BIBER, 2007). Desse modo, os pesquisadores podem optar por realizar essas etapas de cima para baixo (*top down*) ou de baixo para cima (*bottom up*) que, por sua vez, apresentam uma ordem diferente das etapas de análise.

Dessa maneira, na abordagem *top-down*, inicialmente, é desenvolvida a estrutura de análise (cf. Figura 10), ou seja, os tipos de unidade de discurso são fixados antes da análise do *corpus* e, conseqüentemente, todo o estudo é realizado considerando-se os termos fixados (BIBER, 2007).

Figura 10 – Etapas analíticas da abordagem *top-down*.

Required step in the analysis	Realization in this approach
1. Communicative/Functional Categories	Develop the analytical framework: determine set of possible functional types of discourse units, that is, the major communicative functions that discourse units can serve in corpus
2. Segmentation	Segment each text into discourse units (applying the analytical framework from Step 1)
3. Classification	Identify the functional type of each discourse unit in each text of the corpus (applying the analytical framework from Step 1)
4. Linguistic analysis of each unit	Analyze the lexical/grammatical characteristics of each discourse unit in each text of the corpus
5. Linguistic description of discourse categories	Describe the typical linguistic characteristics of each functional category, based on analysis of all discourse units of a particular functional type in the corpus
6. Text structure	Analyze complete texts as sequences of discourse units shifting among the different functional types
7. Discourse organizational tendencies	Describe the general patterns of discourse organization across all texts in the corpus

Fonte: Biber et al. (2007, p. 13).

³⁷ No original: “[...] few studies have attempted to combine these two research perspectives” (BIBER et al., 2007, p. 10).

Já na abordagem *bottom-up*, não há categorias *a priori*. Logo, a análise do *corpus* ocorre primeiro (cf. Figura 11), sendo a partir dessa análise que se identificam os padrões do *corpus* (BIBER et al., 2007).

Figura 11 – Etapas analíticas da abordagem *bottom-up*.

Required step in the analysis	Realization in this approach
1. Segmentation	Segment each text in the corpus into discourse units, based on shifts in vocabulary or other linguistic features
2. Linguistic analysis of each unit	Analyze the full range of lexical / grammatical characteristics of each discourse unit in each text of the corpus
3. Classification	Identify the set of discourse units types that emerge from the corpus analysis, based on linguistic criteria; that is, group all discourse units in the corpus into linguistically-defined categories or 'types'
4. Linguistic description of discourse categories	Describe the typical linguistic characteristics of each discourse category, based on analysis of all discourse units of a particular type in the corpus
5. Communicative/functional categories	Describe the functional bases of each discourse category, based on post-hoc analysis of the discourse units identified as belonging to a particular type
6. Text structure	Analyze complete texts as sequences of discourse units shifting among the different functional types
7. Discourse organizational tendencies	Describe the general patterns of discourse organization across all texts in the corpus

Fonte: Biber et al. (2007, p. 14).

Como já mencionado (cf. seção 1.1), tanto na pesquisa quanto no ensino, são raros os exemplos de estudos que buscam a convergência dos dois paradigmas analíticos descritos. A integração das abordagens *bottom-up* e *top-down* constitui-se em um modelo pedagógico interessante para o ensino de redação acadêmica, na medida em que possibilita o exame do discurso baseado em gênero e a investigação dos elementos linguísticos dos textos a partir do *corpus* (CHARLES, 2007).

Nesse sentido, as abordagens *top-down* e *bottom-up*, embora envolvam metodologias diferentes, complementam-se (CHARLES, 2007). Assim, conforme Charles (2007), através da combinação das abordagens, é possível que os alunos relacionem propósitos comunicativos a escolhas léxico-gramaticais específicas. Outrossim, a combinação das duas abordagens é igualmente exemplificada nas ferramentas *Academic Phrasebank* (MORLEY, 2014) e *AWSuM* (MIZUMOTO, 2017).

1.3.4 Linguística de *Corpus* e o Ensino de Línguas Estrangeiras

A extração, a análise e a descrição de dados por meio da Linguística de *Corpus* têm como objetivo não apenas auxiliar estudos de língua ou de variedade linguística (SARDINHA, 2004), mas também contribuir para a elaboração de aplicações pedagógicas utilizadas no ensino de diferentes gêneros do discurso (BOCORNÝ; WELP, 2021), especialmente no âmbito das linguagens para fins específicos e para fins acadêmicos. É importante lembrar que materiais didáticos produzidos em massa, em um modelo único para todos (*one-size fits all*), não atendem às especificidades das áreas e, tampouco, às necessidades dos usuários finais que atuam em variados contextos acadêmicos e profissionais. Nesse sentido, a possibilidade de extrair dados linguísticos típicos de um determinado gênero, usado de forma recorrente em uma área de especialidade, enriquece, de forma significativa, as aplicações pedagógicas desenvolvidas.

O uso de dados linguísticos extraídos de *corpora* para o desenvolvimento de aplicações pedagógicas, entretanto, é relativamente recente. Dos anos 1930 à metade dos anos 1950, nos Estados Unidos, houve um movimento chamado *Reading Movement*, cujo propósito era desenvolver materiais para ensinar a leitura através do auxílio de uma lista de palavras frequentes no *corpus* (SARDINHA, 2004). Entretanto, com o advento da Linguística Gerativa, a partir dos estudos de Chomsky (1965, 1966), como já mencionado, a descrição da linguagem por meio de *corpora* foi deixada de lado no que se refere ao ensino de línguas (MCENERY, 2003; SARDINHA, 2004).

Entre meados de 1980 e início de 1990, o uso da LC no ensino de línguas voltou a emergir. Relata-se que Tim Johns já usava *corpora* com seus alunos na Universidade de Birmingham por volta de 1986 (SCOTT, 2010). Em 1991, Johns apresenta o termo *data-driven learning* (DDL), conhecido como “aprendizagem direcionada/movida por dados” em português (SCOTT, 2010).

O *data-driven learning* (DDL), ou aprendizagem direcionada por dados (ADD), consiste em uma metodologia que utiliza, no ensino de línguas, dados extraídos de *corpora* (LEE; SWALES, 2006). Thomas (2015) sinaliza que tal metodologia oportuniza um contato mais direto com a língua na aprendizagem direta desta, sem a utilização de outros recursos, como livros didáticos, gramáticas e dicionários.

Flowerdew (2012) pontua que a aplicação pedagógica por meio de *corpus* tem sido muito importante nos últimos anos e, de modo geral, pode ser considerada a partir de dois tipos: aplicações indiretas (*indirect DDL*) e aplicações diretas (*direct DDL*). Nesse primeiro tipo, conforme Flowerdew (2012), enquadram-se pesquisas de *corpora* que oferecem dados fraseológicos e teóricos para construção de dicionários e gramáticas, por exemplo. Por sua vez, as aplicações diretas exigem

esforços dos alunos e dos professores que precisarão identificar padrões a partir da interação com o *corpus* (FLOWERDEW, 2012).

Na abordagem ADD, o professor fornece os dados linguísticos necessários para que o aluno consiga analisá-los e chegar a conclusões sozinho. Geralmente, a aprendizagem direcionada por dados é realizada por meio de atividades organizadas que possibilitam aos alunos examinarem e elucidarem os dados, identificando padrões léxico-gramaticais, ligados, por exemplo, a frequências e a colocações (CROSTHWAITE et al., 2019). Nos anos 1990, o concordanciador era considerado por Johns (1991) como a ferramenta digital mais relevante para a aplicação da ADD, em função da possibilidade “de recuperar do texto todos os contextos de um determinado item (morfema, palavra ou frase) e imprimi-los de uma forma que facilite a leitura e a comparação rápidas”³⁸ (JOHNS, 1991, p. 2).

Durante três anos, ao trabalhar com seus alunos, a maioria pós-graduandos de diferentes áreas que precisavam melhorar o inglês para razões instrumentais, Johns (1991) chegou a três conclusões quanto à aprendizagem direcionada por dados. Em primeiro lugar, o autor concluiu que o uso do concordanciador estimula a investigação e contribui para que o aluno, ao observar e analisar os dados, desenvolva a capacidade de explicar certas regras e padrões da língua que, muitas vezes, sequer o professor sabe *a priori* (JOHNS, 1991, p. 2-3). Em segundo lugar, o pesquisador concluiu que o papel do professor muda à proporção que este precisa aprender a instruir e a ajudar os alunos durante o processo da pesquisa desenvolvida (JOHNS, 1991, p. 3). Por fim, o estudioso concluiu que a ADD pode desenvolver uma “conscientização gramatical” (RUTHERFORD, 1987), ao passo que as descobertas da gramática por parte do aluno, através do estudo de contextos autênticos, são colocadas no centro da aprendizagem (JOHNS, 1991).

Por volta do mesmo período, John Sinclair liderava o projeto *Cobuild*, o qual utilizava a concordância como a principal fonte para descrever o significado de palavras (SARDINHA, 2011). Essa fase inicial da LC no ensino de línguas foi denominada *classroom concordancing*, porque fundamentada “no uso das concordâncias (listagens de palavras de um *corpus*, ladeadas por porções dos textos em que ocorrem, feitas por computador) como instrumento de ensino” (SARDINHA, 2011, p. 302).

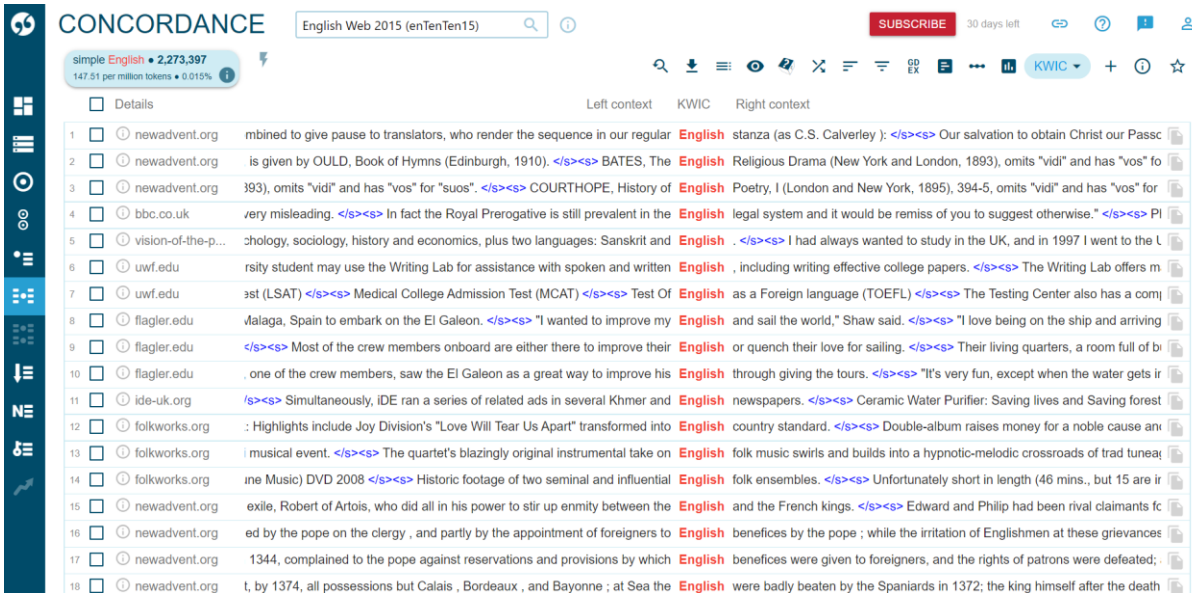
Atualmente, outras formas de visualização dos dados linguísticos extraídos dos *corpora* estão disponíveis. O *Sketch Engine*³⁹, uma ferramenta desenvolvida para explorar o funcionamento da linguagem, permite que os resultados de uma busca sejam apresentados, para além de linhas de concordância, em forma de tabela (*word sketch*), em uma nuvem de palavras ou em um gráfico. A

³⁸ No original: “[...] is able to recover from text all the contexts for a particular item (morpheme, word or phrase) and to print them out in a way which facilitates rapid scanning and comparison” (JOHNS, 1991, p. 2).

³⁹ Disponível em: <https://www.sketchengine.eu/>. Acesso em: 16 jun. 2022.

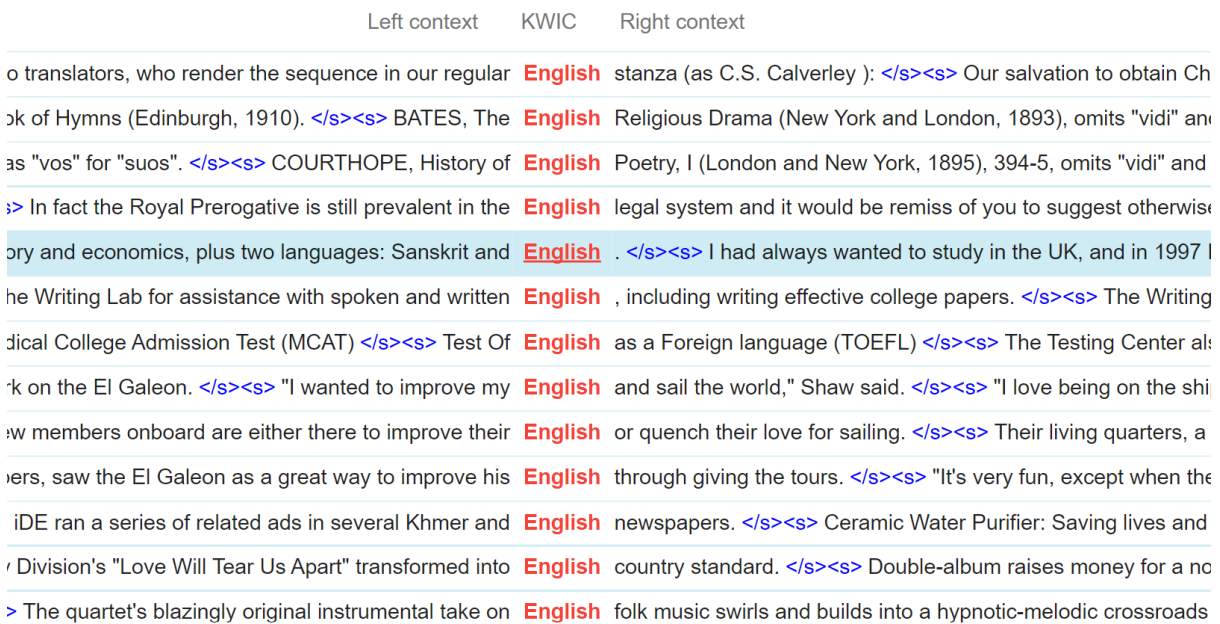
título de exemplo, utilizamos o *corpus English Web 2015 (enTenTen15)*, disponibilizado pelo programa, e selecionamos a funcionalidade *concordance* para analisarmos as linhas de concordância da palavra *English*. Nessa opção, é possível analisar os contextos de coocorrência com “*English*”, como mostram as Figuras 12 e 13.

Figura 12 – Linhas de concordância da palavra *English*.



Fonte: Site *Sketchengine* (2022).

Figura 13 – Linhas de concordância aproximadas da palavra *English*.



Fonte: Site *Sketchengine* (2022).

Em seguida, utilizando a mesma palavra, selecionamos a opção *word sketch*. Nessa funcionalidade, é possível analisar quais advérbios, substantivos, verbos, preposições, adjetivos, dentre outros aspectos gramaticais e lexicais, são utilizados ao redor de *English*, como exemplifica a Figura 14.

Figura 14 – *Word Sketch* da palavra *English*.



Fonte: Site *Sketchengine* (2022).

Por fim, o desenvolvimento da tecnologia permitiu, na segunda metade do século XX, avanços nas pesquisas relacionadas à LC e aos estudos sobre a redação de gêneros acadêmicos (MIZUMOTO, 2017). Nos últimos anos, cresceu o número de trabalhos acerca da escrita acadêmica com base e direcionadas por *corpus*, *corpus-based* e *corpus-driven*, respectivamente. Os dados linguísticos resultantes de alguns desses estudos foram utilizados para a construção de recursos *on-line* que auxiliam na redação de diferentes gêneros em diferentes línguas, como o modelo de escrita acadêmica baseado na *Web*, o qual, criado por Sun (2007), assiste a falantes não nativos na escrita de pesquisas sobre gênero (MIZUMOTO, 2017). Supomos que um dos empregos mais recentes da Linguística de *Corpus* no âmbito do ensino-aprendizagem de línguas estrangeiras tem sido no desenvolvimento de aplicações *on-line* que podem ser usadas tanto em sala de aula, com a supervisão de um professor, quanto por aprendizes de forma autônoma, sem a supervisão de um professor, como sugerem Bocorny et al. (2021).

Neste capítulo, apresentou-se o aporte teórico e situou-se o presente trabalho frente a outros encontrados na literatura. Na primeira seção, abordou-se a área dos Gêneros do Discurso (GD), mais especificamente aqueles utilizados em contexto acadêmico. Na segunda seção, à luz da revisão de teorias e conceitos relativos à linguagem formulaica, descreveram-se as ELs, objeto deste estudo. Na terceira seção, revisitou-se a Linguística de *Corpus*.

2 PROCEDIMENTOS METODOLÓGICOS

Este capítulo delinea os procedimentos metodológicos adotados para a coleta e para a análise de dados, com o objetivo de responder às questões de pesquisa propostas. Inicialmente, caracterizamos a investigação realizada. Em seguida, descrevemos os *corpora* e as ferramentas utilizadas. Na sequência, retomamos os objetivos e as questões do estudo ora apresentado. Por fim, detalhamos os procedimentos metodológicos implementados.

2.1 Caracterização da pesquisa

O trabalho aqui apresentado caracteriza-se como um estudo empírico e indutivo, que busca descrever a linguagem em uso a partir da análise de três *corpora* especializados, com base em princípios oriundos de duas áreas de estudos: os Estudos sobre Gêneros do Discurso e a Linguística de *Corpus* (LC). O modelo proposto neste estudo, portanto, busca estabelecer uma relação entre função e forma (MORENO, SWALES, 2018) por meio da convergência de dois paradigmas analíticos, como sugerido por Moreno e Swales (2018) e Gray et al., (2020), de forma a mapear as relações entre função retórica e elementos linguísticos.

2.2 Descrição dos *corpora* e das ferramentas usadas no estudo

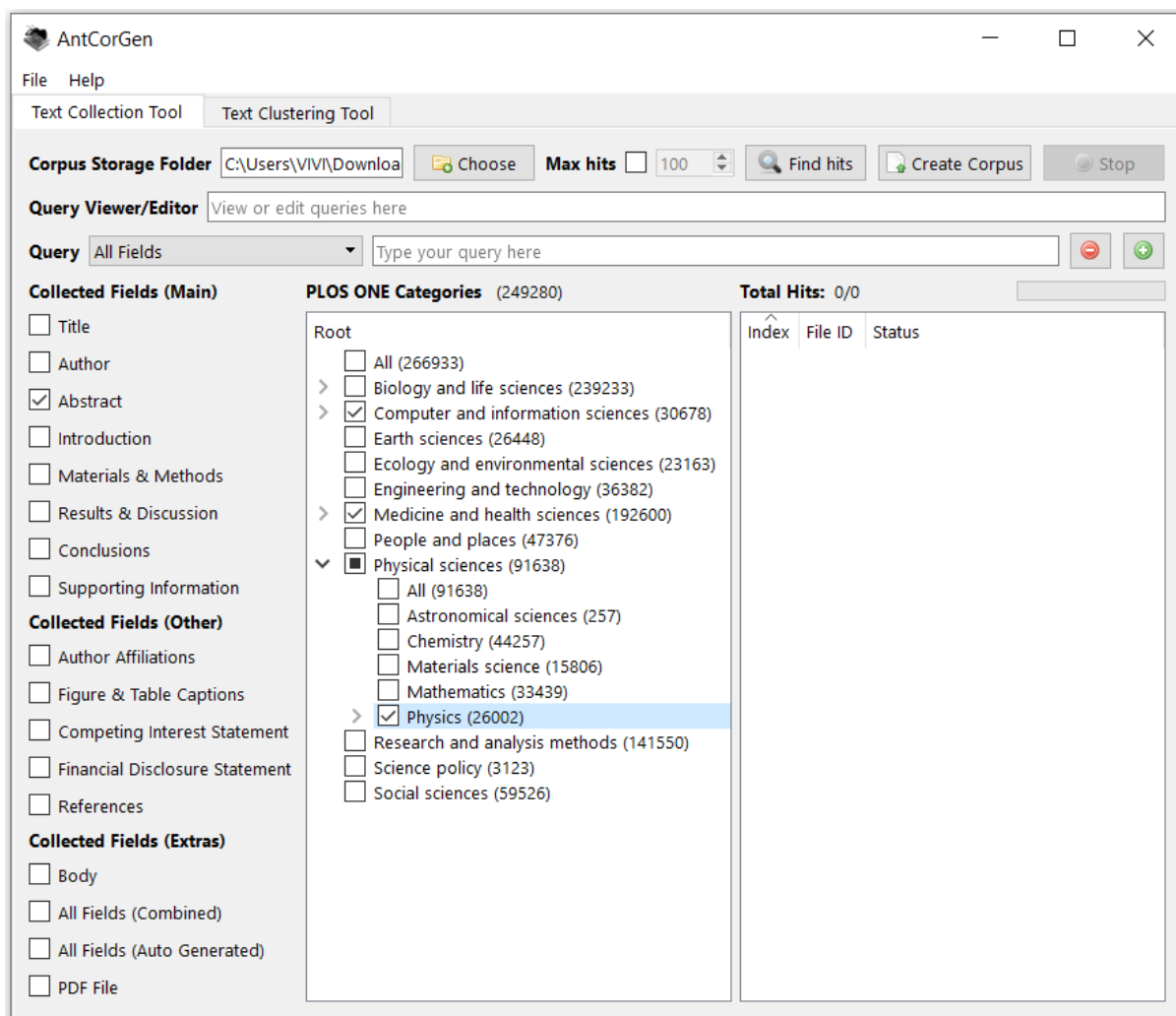
Os dados utilizados nesta pesquisa foram coletados em três *corpora* de estudo, cada qual com aproximadamente um milhão de palavras, compostos por resumos de artigos acadêmicos de três áreas de especialidade: (1) Ciências da Computação e da Informação, (2) Física e (3) Medicina e Ciências da Saúde. Os artigos dos quais os resumos acadêmicos foram extraídos são textos escritos em língua inglesa e publicados, entre 2003 e 2019, em periódicos internacionais de livre acesso, revisados por pares.

A compilação dos *corpora* foi realizada com o auxílio da ferramenta *AntCorGen*⁴⁰ (cf. ANTHONY, 2019), representada pela Figura 15, na página seguinte. Tal ferramenta possibilita a compilação de textos de diferentes seções (*abstract, introduction, conclusions* etc.) e áreas (*computer*

⁴⁰ Disponível em: <https://www.laurenceanthony.net/software/antgram/>. Acesso em: 15 jun. 2022.

and information sciences, earth sciences, social sciences) de artigos científicos publicados na plataforma *PLOS ONE*⁴¹.

Figura 15 – Página inicial do programa *Antcorgen*.



Fonte: Laurence Anthony's Website (2022).

Dessa forma, selecionamos a opção do *Collected Fields (Main)*: *Abstract* e compilamos, separadamente, três *corpora* de aproximadamente 1 milhão de palavras⁴² das áreas *Computer and Information Sciences*”, *Medicine and Health Sciences* e *Physics*.

Tendo em vista os objetivos desta pesquisa, os critérios considerados para a compilação dos três *corpora* de estudo foram os seguintes: (i) o registro e o gênero dos textos; (ii) o meio, o idioma e

⁴¹ *PLOS ONE* é uma revista científica, gratuita e disponível apenas em formato digital, que aborda pesquisas das áreas da Engenharia, Medicina e Ciências Sociais e Humanas. Disponível em: <https://journals.plos.org/plosone/>. Acesso em: 15 jun. 2022.

⁴² Para que fosse possível compilar tais *corpora* de 1 milhão de palavras, primeiramente, 100 documentos (*hits*) foram armazenados no computador. Em seguida, fizemos o *upload* dos textos para o programa *Sketch Engine* (SE) para conferirmos o número de palavras desses arquivos. Em seguida, calculamos, por meio de uma regra de três, quantos documentos (*hits*) precisavam ser compilados para chegarmos a aproximadamente 1 milhão de palavras. Por fim, para darmos continuidade à análise desta pesquisa, os três *corpora* foram inseridos no SE.

o ano (ou período) da publicação; (iii) os domínios aos quais os textos estavam relacionados; e (iv) as seções dos artigos. O Quadro 6 apresenta as características de *design do corpus* de estudo utilizado nesta pesquisa.

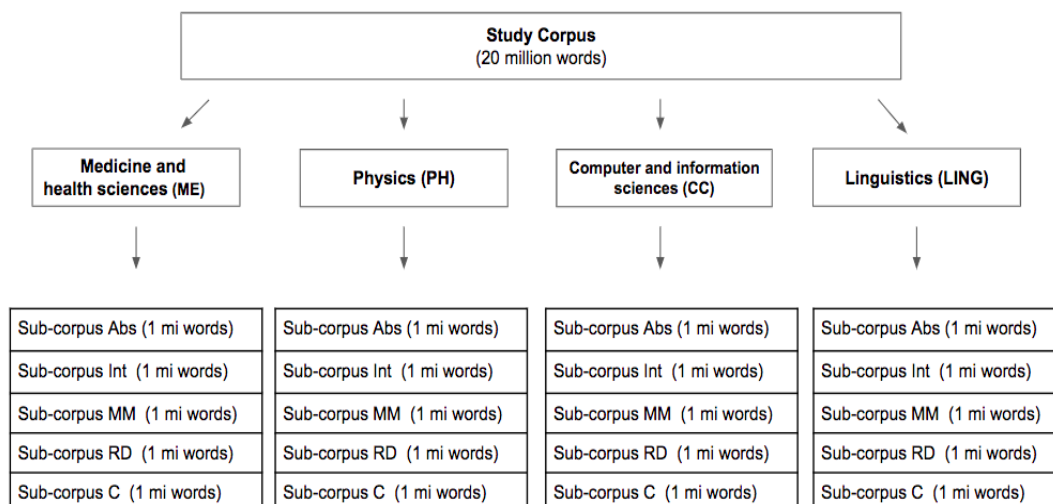
Quadro 6 – Critérios de compilação dos *corpora* de estudo.

REGISTRO ⁴³	Acadêmico
GÊNERO	Resumos acadêmicos
MEIO DE PUBLICAÇÃO	Periódicos revisados por pares e de acesso aberto da plataforma <i>PLOS ONE</i>
LÍNGUA DE PUBLICAÇÃO	Inglês
ANO (OU PERÍODO) DE PUBLICAÇÃO	de 2003 a 2019
DOMÍNIOS	Ciências da Computação e da Informação
	Física
	Medicina e Ciências da Saúde

Fonte: Elaborado pela autora.

Os *corpora* de estudo utilizados nesta pesquisa são parte do *Corpus of Discipline and Section-Specific Academic English (CODISSAE)* (BOCORNY et al., 2021), um *corpus* de 20 milhões de palavras, compilado com a ferramenta *AntCorGen* e que contém 20 *sub-corpora*, de 1 milhão de palavras cada, de resumos e das seções de artigos de pesquisa de quatro domínios: Ciências da Computação e da Informação; Física; Medicina e Ciências da Saúde; Linguística. A Figura 16, na página seguinte, mostra a composição do *CODISSAE*.

⁴³ Biber et al. (2007, p. 8) definem *registro* (quando diferenciado de *gênero*) como a linguagem associada a uma área do conhecimento ou a um domínio, como o registro jurídico ou o registro acadêmico. O termo *gênero*, por sua vez, quando contrastado com *registro*, é usado para referir-se a um tipo de mensagem com uma estrutura interna convencionalizada, como um artigo da área de Biologia ou um memorando de negócios.

Figura 16 – Composição do *corpus* CODISSAE.

Fonte: Elaborada pela autora.

Os *corpora* usados neste estudo, *sub-corpora* do CODISSAE, foram nomeados conforme códigos (cf. Quadro 7), compostos de caracteres alfa-numéricos, que representam as seguintes informações sobre o *corpus* e sobre os textos que o compõem: tipo de *corpus*, período de publicação, gênero, domínio, tamanho do *corpus*.

Quadro 7 – Códigos alfa-numéricos utilizados para nomear os *corpora* do CODISSAE.

TIPO DE CORPUS	<i>Corpus</i> de estudo	SC
	<i>Corpus</i> de referência	RC
PERÍODO DE PUBLICAÇÃO DOS TEXTOS QUE COMPÕEM O CORPUS	De 2003 até 2019	0319
GÊNERO	Resumo acadêmico	Abs
	Artigo	Ar
DOMÍNIO	Ciências da Computação e da Informação	CS
	Física	PHY
	Medicina e Ciências da Saúde	ME
TAMANHO DO CORPUS	1 milhão de palavras	1mi
		SC0319AbsCS1mi

Fonte: Adaptado de Guedes (2017).

Conforme as convenções descritas, os três *corpora* usados neste estudo foram chamados de **SC0319AbsCS1mi**, **SC0319AbsPHY1mi** e **SC0319AbsME1mi**. A Tabela 1, a seguir, mostra os três *corpora* em números de *tokens*, *types*, frases e textos, bem como a média de palavras por seção do *corpus* do estudo.

Tabela 1 – *Corpora* de estudo em números.

DOMÍNIO	CORPUS DE ESTUDO	NÚMERO TOTAL DE PALAVRAS (TOKENS ⁴⁴)	NÚMERO TOTAL DE PALAVRAS SEM REPETIÇÃO (TYPES)	FRASES	TEXTOS	MÉDIA DE PALAVRAS POR RESUMO
Ciências da Computação e da Informação	SC0319AbsCS1mi	1,000,040	45,838	42,371	4,598	217
Física	SC0319AbsPHY1mi	1,000,006	54,679	41,729	4,690	213
Medicina e Ciências da Saúde	SC0319AbsME1mi	1,000,006	42,742	47,913	4,039	247

Fonte: Elaborada pela autora.

Os *corpora* de estudo utilizados nesta pesquisa estão disponíveis em <https://drive.google.com/drive/folders/1pnDDXgftNzJOE58j96kAbj9LpxyLP6CH?usp=sharing>.

Por fim, para dar continuidade à análise desta pesquisa, os três *corpora* foram inseridos no SE.

⁴⁴ A ferramenta SE denomina *token* a menor unidade existente em um *corpus*. Assim, diferentes formas das palavras e sinais de pontuação são contabilizados como *tokens* distintos. Palavras unidas por apóstrofo e hífen são contabilizadas separadamente.

2.3 Objetivos e questões de pesquisa

O presente estudo tem como objetivo principal fazer a análise funcional de ELs recorrentes nos resumos acadêmicos de três áreas do conhecimento: (1) Ciências da Computação e da Informação; (2) Física; (3) Medicina e Ciências da Saúde. Em especial, busca-se identificar e analisar os referidos blocos formulaicos extraídos dos três *corpora*, em uma abordagem dirigida por *corpus* (*data-driven approach*), bem como relacionar tais unidades sintagmáticas à estrutura e às funções retóricas do gênero-alvo, com vistas ao uso pedagógico de tais dados em cursos de escrita de resumos acadêmicos das áreas de especialidade que são foco deste estudo. Assim, os objetivos específicos deste trabalho baseiam-se nas seguintes questões de pesquisa:

1. Quais seções constituem os resumos estruturados das áreas-alvo?
2. Quais ELs são recorrentes nos resumos das áreas-alvo?
3. Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?

Para tanto, assim como proposto por Gray *et al.* (2020), partimos da convergência de dois paradigmas analíticos: a análise de movimentos retóricos (*move analysis*) e a extração de ELs, de forma a mapear a relação entre os referidos elementos linguísticos que realizam funções retóricas expressas nas diferentes seções dos resumos das três áreas estudadas. Busca-se, assim, estabelecer uma relação entre **função** e **forma**, conforme proposto por Moreno e Swales (2018).

2.4 Procedimentos analíticos

Uma vez que os *corpora* de estudo foram compilados e preparados para análise, uma série de procedimentos metodológicos foram empregados para responder às questões de pesquisa. O Quadro 5 alinha cada questão de pesquisa com os dados usados e as principais etapas metodológicas realizadas para abordar as questões. O Quadro 8 apresenta, de forma resumida, as etapas metodológicas realizadas no estudo, relacionando-as às perguntas de pesquisa e aos dados a partir dos quais as perguntas de pesquisa foram respondidas.

Quadro 8 – Perguntas de pesquisa e etapas metodológicas realizadas.

PERGUNTA DE PESQUISA	DADOS A PARTIR DOS QUAIS AS PERGUNTAS DE PESQUISA FORAM RESPONDIDAS	ETAPAS METODOLÓGICAS REALIZADAS
<p>Pergunta de pesquisa 1: Quais seções constituem os resumos estruturados das áreas-alvo? (identificação da estrutura dos resumos)</p>	<p>Etiquetas de seção dos resumos estruturados de cada área</p>	<ol style="list-style-type: none"> 1) Ferramenta SE. 2) Seleção do <i>corpus</i> de estudo. 3) Aplicação da sintaxe de busca: <code><s> [1,3] [word=""]</code>. 4) Organização dos resultados por ordem alfabética. 5) Categorização a partir das seções propostas por Swales e Feak (2009). 6) Inclusão de seções não existentes em Swales e Feak (2009). 7) Criação de tabela e gráfico.
<p>Pergunta de pesquisa 2: Quais ELs são recorrentes nos resumos das áreas-alvo? (identificação das ELs recorrentes)</p>	<p>Todas as ELs extraídas dos <i>corpora</i> analisados, conforme os critérios de tamanho, número de lacunas, frequência mínima e dispersão estabelecidos no estudo</p>	<ol style="list-style-type: none"> 1) Ferramenta <i>AntConc</i>. 2) Inserção dos parâmetros de extração na ferramenta. 3) Extração de todas as estruturas lexicais. 4) Seleção das 50 mais frequentes.
<p>Pergunta de pesquisa 3: Quais funções retóricas são linguisticamente realizadas por ELs recorrentes nos resumos acadêmicos das áreas-alvo? (análise da associação entre ELs e movimentos retóricos)</p>	<p>As 50 ELs mais frequentes em cada um dos <i>corpora</i> analisados</p>	<ol style="list-style-type: none"> 1) Observação dos elementos lexicais que não variam (<i>frame</i>) nas ELs. 2) Observação das palavras que preenchem as lacunas (<i>slot fillers</i>). 3) Observação das estruturas lexicais em seu contexto.

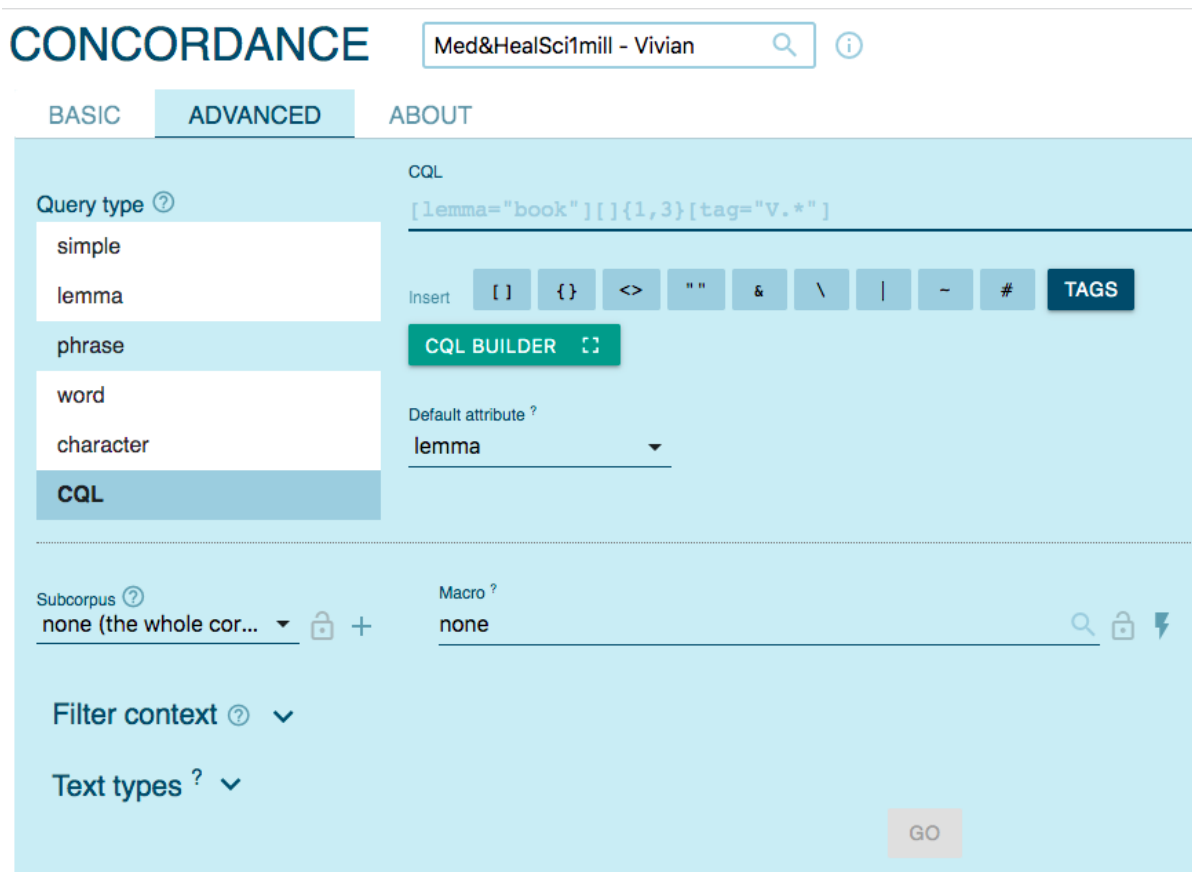
Fonte: Elaborado pela autora, com base em Geluso (2019).

2.4.1 Pergunta de pesquisa 1: Quais seções retóricas constituem os resumos estruturados das áreas-alvo?

Na etapa metodológica relativa à identificação das seções retóricas que constituem os resumos estruturados das áreas-alvo a partir dos *corpora* compilados, usou-se como referência a metodologia utilizada por Freitas e Bocorny (2021) para a análise de resumos estruturados da área de Epidemiologia. Nesse estudo, a identificação da estrutura retórica se deu a partir da extração das palavras que iniciam as seções do resumo, aqui chamadas de *etiquetas de seção*.

Para a identificação das etiquetas de seção, usou-se uma sintaxe de busca construída por meio do *CQL Builder* da ferramenta SE: `<s> [1,3] [word=":"]`, como mostram as Figuras 17 e 18.

Figura 17 – Localização da funcionalidade *CQL Builder* na ferramenta *Sketch Engine*.



Fonte: Ferramenta *Sketchengine* (2022).

Figura 18 – Sintaxe de busca usada para extrair as etiquetas de seção.

The screenshot shows the Sketch Engine CQL builder interface. At the top, the logo for Sketch Engine is on the left, and the title 'CQL builder' is in the center. Below the title, the current query is displayed: 'CQL: <s> [] {1,3} [word=":"]'. The main area shows a visual representation of the query components: a blue box containing '<s>' (labeled 'structure'), a plus sign, a blue box containing '[] {1,3}' (labeled 'distance'), a plus sign, a blue box containing '[word=":"]' (labeled 'normal token'), and another plus sign followed by a plus icon in a blue box. Below this visual representation, a note states: 'Your query may include a sentence boundary. Adding 'within <s/>' ensures the result is found inside the same sentence.' To the right of this note is a red button labeled 'USE THIS CQL'. Below the note is a section titled 'result example' with a dropdown arrow. The example shows a list of text snippets, each followed by a red 'AuthorSummary:' label and a corresponding summary. The snippets are: 'hanistic explanation for the long recognized link between inflammation and cancer progression.', 'ng mathematical modeling and in silico-guided design of solid tumor immunotherapy in the clinic.', 'of drug combinations and provide a new way to effectively utilize existing drugs for new purposes.', 'and help guide the choice of treatment strategies that best inhibit diverse tumor cell populations.', 'can also be generalized for other infectious diseases, computer viruses, or ecological networks.', 'interventions should be more carefully studied, for both influenza and other infectious diseases.', 'enovel respiratory pathogens can be rapidly assessed and compared with previous pandemics.', 'ics of metastasis development is dependent on spatial interactions between metastatic lesions.', 'nd estimated the numbers of required metastatic foci to obtain the observed metastatic volumes.', 'ant advantage over radiofrequency ablation (RFA) in terms of systematic removal of a hepatic segment.'

Fonte: Ferramenta *Sketchengine* (2022).

Tal sintaxe permite a identificação de um a três substantivos em início de frase, seguidos de dois pontos (:), a exemplo de *Background;*, *Method;*, *Results;*, como mostra a Figura 19, na página seguinte.

Figura 19 – Resultado da busca feita com a sintaxe: <s> [1,3] [word=“:”]

Left context	KWIC	Right context
neticum. </s><s>	Background:	Mycolactones ai
product. </s><s>	Conclusion/Significance:	This study demc
ane core. </s><s>	Background:	Animals that cor
filtering? </s><s>	Principal Findings:	Combining beh
eurones. </s><s>	Conclusions:	There is no indic
samples. </s><s>	Background and aims:	In contrast to se
nditions. </s><s>	Methods:	In a greenhouse
ssessed. </s><s>	Key results:	The influence of
ranches. </s><s>	Conclusions:	Clonal integratio
munities. </s><s>	Background and Objective:	Light and lasers
RASER. </s><s>	Results:	Specific peaks c
tion rate. </s><s>	Conclusion:	Considering the
limidiata. </s><s>	Background:	The autofluores
F-7 cells. </s><s>	Conclusions/Significance:	We have demon
i system. </s><s>	Background:	Pleckstrin homo
manner. </s><s>	Principal Findings:	To study the str
ding site. </s><s>	Conclusions:	Our results sugg
id fishes. </s><s>	Background:	The study on the
in detail. </s><s>	Conclusions/Significance:	S-Ni0.1MgAl-La
i models. </s><s>	Background:	Binding of two 9
de chain. </s><s>	Conclusions/Significance:	Remarkably high
hologies. </s><s>	These techniques are:	widefield, apotor
hanisms. </s><s>	Background:	Fitness in birds l

Fonte: Ferramenta *Sketchengine* (2022).

A partir da identificação de todas as etiquetas de seção de cada *corpora*, criou-se uma tabela na qual as etiquetas foram inseridas. Iniciou-se, então, o processo de categorização das etiquetas de seção, tendo-se como base o modelo de estrutura retórica de resumos acadêmicos proposto por Swales e Feak (2009) (cf. seção 3.1). Neste momento, seções não existentes no modelo proposto por Swales e Feak (2009) foram incluídas.

2.4.2 Pergunta de pesquisa 2: Quais ELs são recorrentes nos resumos das áreas-alvo?

Na etapa metodológica relativa à identificação das ELs recorrentes nos resumos das áreas-alvo, partimos dos dados linguísticos em uso em uma abordagem baseada na frequência para a identificação dos elementos linguísticos analisados. Assim, a partir da observação dos dados e da identificação de regularidades, realizamos categorizações. Não partimos, portanto,

das categorias à priori, ainda que tenhamos usado o padrão retórico proposto por Swales e Feak (2009) como referência.

2.4.2.1 Critérios para a extração das ELs

De acordo com Cortes (2013), quanto maior a sequência de palavras, menor é o número das suas repetições. Com base nesse pressuposto, os critérios para a extração das ELs dos *corpora* utilizados na presente dissertação foram os seguintes: (i) ser uma sequência de seis palavras, (ii) ter duas lacunas variáveis, (iii) ocorrer pelo menos 30 vezes por milhão de palavras (pmp) nos *corpora* analisados, (iv) constar em pelo menos 20 textos nos *corpora* analisados. A Tabela 2 resume os critérios estabelecidos para a extração dos dados.

Tabela 2 – Critérios para a extração das ELs.

Número de palavras	6
Número de lacunas	2
Frequência mínima	30
Dispersão	20 textos ou mais

Fonte: Elaborada pela autora.

2.4.2.2 Ferramenta utilizada para a extração dos dados

A ferramenta utilizada para a extração dos dados foi o *AntConc 4.0.10* (ANTHONY, 2022), descrito anteriormente (cf. seção 1.2.5.2). Para a etapa de identificação das ELs mais recorrentes nos resumos de cada área-alvo, foram extraídas todas as ELs que cumprissem os critérios estabelecidos.

2.4.3 Pergunta de pesquisa 3: Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?

Como mencionado na seção 1.2.3, para a realização da análise funcional retórica proposta neste estudo, lançamos mão de uma abordagem indutiva semelhante à adotada por Biber et al. (2004). Assim, a partir da identificação das 50 ELs mais recorrentes em cada uma

das áreas estudadas e da organização das ELs em tabelas (cf. Anexos 4, 5 e 6), os procedimentos metodológicos adotados para a efetivação dessa análise foram os seguintes: (i) observação dos elementos lexicais que não variam (*frame*) nas ELs; (ii) observação das palavras que preenchem as lacunas (*slot fillers*); (iii) observação das estruturas lexicais em seu contexto. Uma vez identificadas as funções retóricas, as ELs foram agrupadas conforme as funções retóricas que realizavam.

A Tabela 3 exemplifica a forma como as 50 ELs mais recorrentes da área no *corpus* das Ciências da Computação e da Informação foram organizadas em tabelas. Os Anexos 7, 8 e 9, apresentam as 50 ELs mais recorrentes das áreas de Ciências da Computação e da Informação, de Física e de Medicina e Ciências da Saúde, respectivamente.

Tabela 3 – 50 ELs do *corpus* das Ciências da Computação e da Informação.

Número	EL	Freq. Abs.	Dispersão
1	in this + we + a	307	306
2	to + the + of the	289	270
3	in this paper + + a	209	209
4	in + paper we + a	205	205
5	in this + we + the	196	193
6	the + of the + of	159	149
7	the + and + of the	154	149
8	in + + we propose a	132	132
9	in this + + propose a	122	122
10	in this paper + + the	120	119
11	we + the + of the	116	112
12	the + of this + is	115	113
13	used to + the + of	107	102
14	of the + + of the	106	99
15	in + paper + propose a	96	96
16	the + of the + and	96	96
17	and + the + of the	94	92

18	in + paper we + the	94	93
19	we + that the + of	94	89
20	the + results + that the	87	85
21	the + of + in the	84	84
22	based on the + + of	74	72
23	in this + we + an	70	70
24	this + we + a novel	69	69
25	of the + + in the	69	69
26	in the + + of the	66	60
27	order to + the + of	66	65
28	this study + to + the	66	65
29	can be + to + the	65	64
30	the + of this + was	61	61
31	and the + + of the	61	60
32	and + + as well as	60	60
33	this + we + a new	60	60
34	as well as + + of	60	58
35	in terms of + + and	60	60
36	the + and the + of	59	59
37	aim of this + + to	59	59
38	the + of the + in	58	58
39	in this study + + the	58	58
40	the + of the + is	57	54
41	of the + + and the	56	54
42	to the + + of the	56	55
43	in + + we present a	55	55
44	the + + in this paper	55	55
45	to + the + of a	55	54

46	on the + + of the	55	55
47	in the + of + and	54	53
48	the + + show that the	53	53
49	results + that the + of	52	51
50	in this + + present a	52	52

Fonte: Elaborada pela autora.

Foram, então, observados os elementos lexicais que não variam (*frame*) nas ELs e as palavras que preenchem as lacunas (*slot fillers*). Para cada palavra que preencheu a lacuna 1 (L1), foram identificadas 3 palavras na lacuna 2 (L2). Dessa forma, enquanto na L1 temos 3 palavras preenchedoras, podemos ter até 9 palavras preenchedoras na L2. Em alguns casos, para algumas palavras da L1, não havia 3 palavras relacionadas na L2. Por exemplo, para a estrutura lexical “to + the + of the”, extraída do *corpus* SC0319AbsCS1mi, obteve-se o resultado apresentado no Quadro 9, a seguir:

Quadro 9 – Resultado das lacunas da estrutura lexical “to + the + of the”.

to + the + of the	to	L1	the	L2	of the
	-	a. evaluate	-	a. effect/ influence/ details	-
		b. assess		b. impact/ accuracy/ quality	
		c. demonstrate		c. effectiveness/ feasibility/ advantages	

Fonte: Elaborado pela autora.

A partir do Quadro 9, observa-se que na L1 foram inseridas 3 opções de palavras, enquanto na L2 há 9 palavras. Tais quantidades de palavras foram decididas ao longo da análise das linhas de concordância, pois percebeu-se que as palavras da L1 se repetiam em grandes quantidades, enquanto as palavras da L2 variavam mais. Como exemplo, nota-se que, no Quadro 9, a letra (a) tem como preenchedora o verbo “evaluate” na L1, ao passo que na L2 há três palavras que estão associadas ao verbo “evaluate”. Sendo assim, esta é uma maneira de ampliar as opções na hora de completar a estrutura lexical “to + the + of the”.

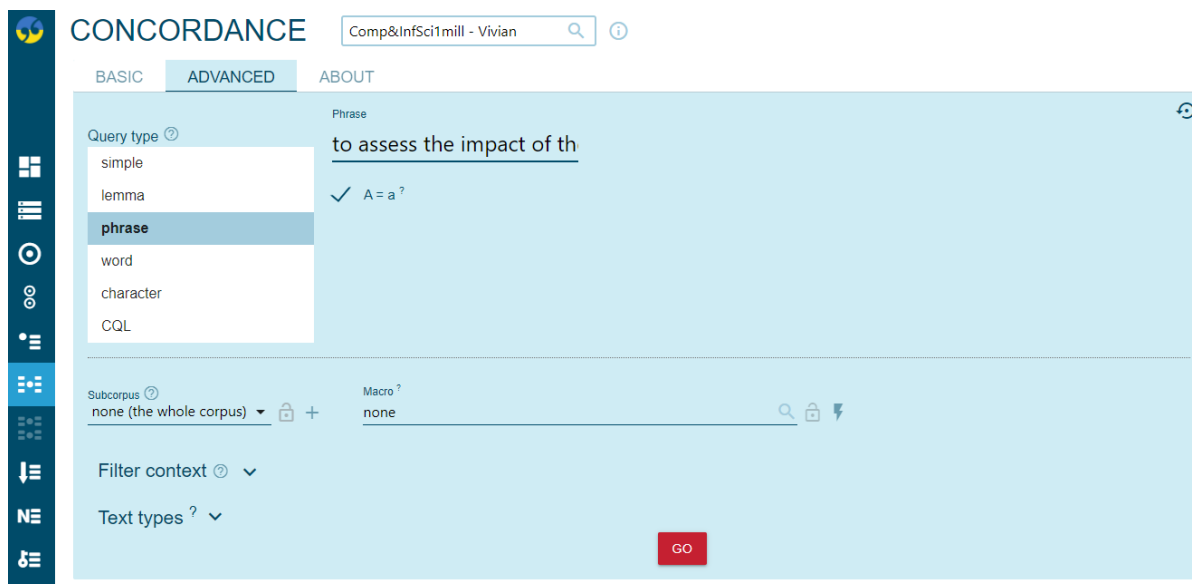
2.4.4 Análise dos dados para a identificação da função retórica das ELs

Para a identificação da função retórica das ELs, como já dito, utilizou-se a ferramenta SE. Antes de iniciar tal etapa metodológica, definiram-se os critérios a seguir: (i) analisar as opções a, b e c da L1 com, no mínimo, uma das opções a, b e c da L2; (ii) analisar as linhas de concordância para identificar a função retórica da estrutura lexical; (iii) caso só fosse identificada uma função retórica nas três linhas de concordância analisadas, apenas um exemplo seria exposto no quadro de resultados; (iv) caso fosse identificada mais de uma função retórica, mais de um exemplo seria incluído para a estrutura lexical.

Após definidos os critérios, seguiram-se os seguintes passos no programa SE:

- 1) selecionar o *corpus*;
- 2) selecionar a opção *Concordance*;
- 3) selecionar a opção *Advanced*;
- 4) selecionar a opção *phrase*;
- 5) redigir a EL (ver exemplo na Figura 20):

Figura 20 – Exemplo de onde redigir a EL no SE.



Fonte: Ferramenta *Sketchengine* (2022).

- 6) selecionar “Go”;
- 7) e, por fim, analisar as linhas de concordância geradas pelo programa, como exemplifica a Figura 21, a seguir.

Figura 21 – Linhas de concordância do SE.

The screenshot displays the CONCORDANCE web interface. At the top, the search bar contains 'Comp&InfSci1mill - Vivian'. A notification bar indicates the search phrase 'to assess the impact of the' with 3 results, a frequency of 2.63 per million tokens, and a percentage of 0.00026%. Below the notification, a toolbar includes icons for search, download, list, eye, cursor, zoom, and filters, along with a 'KWIC' dropdown menu. The main content area shows three search results, each with a checkbox, a document ID, and a snippet of text where the search phrase is highlighted in red.

Checkbox	Document ID	Text Snippet
<input type="checkbox"/>	doc#2168	mmunity structure within the schools, and to assess the impact of the physical environment of schools on proxim
<input type="checkbox"/>	doc#4025	orkshop checklist scores were compared to assess the impact of the workshop on the computational reproductit
<input type="checkbox"/>	doc#4041	pplying machine learning techniques, and to assess the impact of the approach on model complexity and perform

Fonte: Ferramenta *Sketchengine* (2022).

3 RESULTADOS

Este capítulo, dividido em três seções, apresenta os resultados e faz a análise dos dados obtidos à luz de pesquisas e de teorias nas quais o presente estudo se baseia. A seção 1 trata dos resultados relacionados à primeira questão de pesquisa (quais seções constituem os resumos estruturados das áreas-alvo?). Nessa seção, pode-se ter uma visão geral da estrutura retórica do gênero em cada uma das áreas estudadas. Já a seção 2 inicia com dados quantitativos referentes ao número total de ELs extraídas de cada um dos *corpora* do estudo, conforme os critérios de extração estabelecidos (cf. seção 2.4.2.1). Fica contemplada, nessa seção, nossa segunda questão de pesquisa (quais ELs são recorrentes nos resumos das áreas-alvo?). Ainda na seção 2, são mostrados dados relativos à razão entre *n-gram types* e *n-gram tokens*, que dizem respeito à variedade lexical das ELs (*N-gram Type-Token Ratio*) nos *corpora* estudados. Por fim, na seção 3, examinamos a natureza da relação entre as seções retóricas dos resumos estudados e as 50 ELs mais frequentes em cada *corpus*, para respondermos a nossa terceira questão de pesquisa (quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?).

3.1 Pergunta de pesquisa 1: Quais seções constituem os resumos estruturados das áreas-alvo?

A questão de pesquisa 1 visa identificar as seções mais frequentes em resumos estruturados de cada uma das três áreas-alvo. Conhecer a composição dos resumos estruturados que fazem parte da amostra de resumos de cada domínio estudado é importante, pois permite mapear a estrutura dos resumos e indicar as seções mais recorrentes em cada área.

Para respondermos à primeira questão de pesquisa, inicialmente foram identificadas, contabilizadas e categorizadas as palavras que serviam como etiquetas de seção (por exemplo, *Background, Aim, Methods, Results*) nos resumos estruturados de cada área. Para tanto, utilizamos a sintaxe de busca <s> [1,3] [word=""] e a ferramenta *Sketch Engine* (cf. seção 2.4). A Tabela 4, na página seguinte, mostra o número total de resumos, o número de resumos estruturados e o percentual de resumos estruturados nos três *corpora* estudados. Os dados revelam que as áreas com maiores percentual de resumos estruturados são as áreas de Medicina

e Ciências da Saúde (78,50 %)⁴⁵, seguida das Ciências da Computação e da Informação (9,98 %) e da Física (7,35 %). Tal dado não surpreende, pois tradicionalmente os artigos de pesquisa vinculados à área de Medicina e Ciências da Saúde costumam usar resumos estruturados, o que é confirmado pelos dados de nosso estudo.

Tabela 4 – Percentual de resumos estruturados nos corpora.

	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
Número total de resumos no corpus	4.598	4.690	4.039
Número de resumos estruturados no corpus	≈ 459	≈ 345	≈ 3.171
Percentual de resumos estruturados no corpus	≈ 9,98%	≈ 7,35%	≈ 78,50%

Fonte: Elaborada pela autora.

Na sequência, e ainda sobre a pergunta de pesquisa 1, temos as Tabelas 5, 7 e 9, que apresentam todas as etiquetas de seção identificadas nos resumos estruturados dos três corpora analisados.

A Tabela 5 apresenta as etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação. Há uma variedade de 71 etiquetas de seção diferentes. Tal fato revela uma grande diversidade denominativa, que pode significar pouca convencionalidade com relação à terminologia usada para designar as seções dos resumos estruturados da área.

Tabela 5 – Identificação das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.

Área: Ciências da Computação e da Informação		
Nome do corpus: SC0319AbsCS1mi		
Número de resumos acadêmicos: 4.598		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO	Percentual de resumos estruturados desse corpus que têm a seção
BACKGROUND, CONTEXT	Background:	
	Backgrounds:	

⁴⁵ O número de resumos estruturados de cada área foi identificado a partir da contagem do número de ocorrências da etiqueta de seção mais frequente. O percentual de resumos estruturados foi contabilizado dividindo-se o número de resumos estruturados multiplicado por 100 pelo número total de resumos do corpus. É importante registrar que tanto o número de resumos estruturados quanto o percentual de resumos estruturados são números aproximados (≈).

	Context:	9,98%
	Rationale/background:	
	Rationale:	
	Introduction:	
AIM OBJECTIVE PURPOSE	Aim:	3,71%
	Aims:	
	Background & Aims:	
	Background & Objectives:	
	Background and aim:	
	Background and objective:	
	Background and objectives:	
	Background and purpose:	
	Motivation:	
	Objective:	
	Objectives:	
	Purpose/objective:	
	Purpose:	
	Data and methods:	
	Data collection:	
	Data extraction:	
	Data source:	
	Data sources:	
	Data synthesis:	
	Design/Methods:	
MATERIALS, METHODS	Design:	
	Material and methods:	
	Materials and methods:	
	Method:	
	Methodology:	
	Methods/Design:	
	Methods:	
	Methods and Design:	

	Methods and findings:	
	Methods and materials:	
	Methods and results:	
	Participants:	
	Setting:	
	Significance:	
	Software availability:	
	Standard analyses comprise:	
	Standard settings:	
	Study design:	
	Subjects:	
<i>FINDINGS, RESULTS</i>	Findings:	
	Principal findings:	
	Result & Discussion:	
	Result:	8,46%
	Results:	
	Results and discussion:	
	Secondary outcomes:	
<i>CONCLUSION, DISCUSSION</i>	Conclusion/Significance:	
	Conclusion:	
	Conclusion and recommendation:	
	Conclusion and Significance:	
	Conclusions/implications:	
	Conclusions/Significance:	
	Conclusions:	
	Conclusions and relevance:	12,09%
	Conclusions and Significance:	
	Discussion:	
	Discussion and conclusion:	
	Discussion and conclusions:	
	Impact:	
	Implementation:	
	Implication:	
Implications:		

	Importance:	
	Limitations:	
<i>AUTHOR SUMMARY</i>		3,15%

Fonte: Elaborada pela autora.

Ainda sobre a pergunta de pesquisa 1, observamos, na Tabela 5, que as três seções mais recorrentes entre os resumos estruturados são, respectivamente, *CONCLUSION, DISCUSSION* (12,09%); *BACKGROUND, CONTEXT* (9,98%); e *MATERIALS, METHODS* (9%). Tais seções são seguidas pelas seções *FINDINGS, RESULTS* e *AIM, OBJECTIVE, PURPOSE*, as quais ocorrem em 8,46% e 3,71% dos resumos estruturados do *corpus*, sendo, respectivamente, as duas seções menos frequentes da área. Possivelmente, a baixa incidência da seção *AIM, OBJECTIVE, PURPOSE*, nos resumos estruturados de Ciências da Computação e da Informação, deve-se ao fato de o conteúdo da referida seção ser veiculado na seção *BACKGROUND, CONTEXT*. As observações feitas com base na Tabela 5 apresentam-se sintetizadas na Tabela 6.

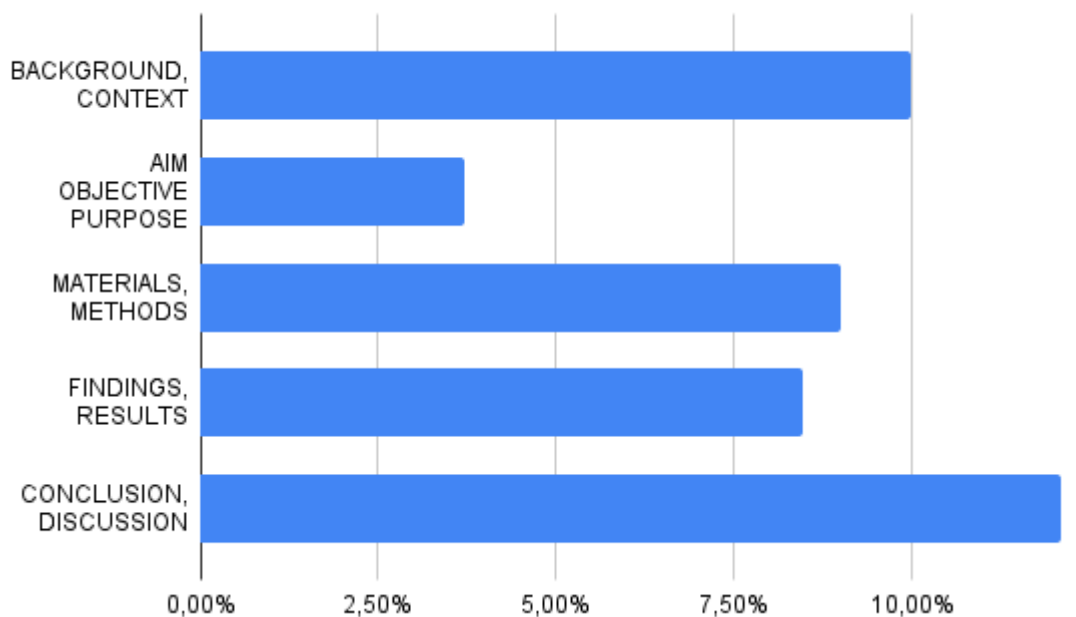
Tabela 6 – Identificação das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.

CATEGORIAS DE SEÇÃO	Ciências da Computação e da Informação	Número total de resumos do <i>corpus</i>
<i>BACKGROUND, CONTEXT</i>	9,98%	4.598
<i>AIM OBJECTIVE PURPOSE</i>	3,71%	
<i>MATERIALS, METHODS</i>	9%	
<i>FINDINGS, RESULTS</i>	8,46%	
<i>CONCLUSION, DISCUSSION</i>	12,09%	

Fonte: Elaborada pela autora.

As observações feitas com base na Tabela 6 podem ser mais bem visualizadas por meio do Gráfico 1.

Gráfico 1 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação.



Fonte: Elaborado pela autora.

Os dados apresentados na Tabela 7, a seguir, mostram as etiquetas de seção dos resumos estruturados da área de Física. Percebe-se uma tendência semelhante àquela verificada na área de Ciências da Computação e da Informação, ou seja, uma variedade significativa de etiquetas de seção diferentes (56). Ainda assim, a diversidade denominativa da área de Física é menor do que a da área de Ciências da Computação e da Informação, o que pode significar uma maior convencionalidade com relação à terminologia usada para designar as seções dos resumos estruturados da área.

Tabela 7 – Identificação das etiquetas de seção dos resumos estruturados da área de Física.

Área: Física Nome do corpus: SC0319AbsPHY1mi Número de resumos acadêmicos: 4.690		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO	Percentual de resumos estruturados deste corpus que têm a seção
BACKGROUND, CONTEXT	Background:	4,37%
	Context:	
	Introduction:	
AIM OBJECTIVE PURPOSE	Aim(s):	3,83%
	Aim of study:	
	Background and aims:	
	Background and objective:	
	Background and objectives:	
	Background and Purpose:	
	Objective(s):	
	Objectives and Motivation:	
	Purpose:	
	Study Objectives:	
MATERIALS, METHODS	Basic Procedures:	6,03%
	Design:	
	Experimental Design:	
	Interpretation(s):	
	Main Outcome Measures:	
	Material and Method:	
	Materials and Methods:	

	Method(s):	
	Method and Material:	
	Methods and materials:	
	Participants:	
	Patients:	
	Patients and methods:	
	Procedure:	
	Research methodology:	
	Setting(s):	
	Study design:	
	Study sample:	
	Subjects and methods:	
	Aim and methods:	
	Background and Methods:	
	Objective and Methods:	
	Finding(s):	6,22%
<i>FINDINGS, RESULTS</i>	Main Findings:	
	Key results:	
	Main results:	
	Principal Findings:	
	Result(s):	
	Results and limitations:	
	Objectives and Results:	
	Measurements and Results:	
	Clinical Significance:	7,35%
<i>CONCLUSION, DISCUSSION</i>	Clinical Relevance:	
	Conclusion/Significance:	
	Conclusions/Significance:	
	Conclusion(s):	
	Conclusion and Impact:	
	Conclusion and Relevance:	
	Conclusion and significance:	
	Discussion:	
	Discussion and Conclusion:	

	Discussion and Conclusions: Principal Conclusions:	
<i>AUTHOR SUMMARY</i> ⁴⁶		0,34%

Fonte: Elaborada pela autora.

Ainda sobre a pergunta de pesquisa 1, observamos, na Tabela 7, que as três seções mais recorrentes entre os resumos estruturados são, respectivamente, *CONCLUSION, DISCUSSION* (7,35%); *FINDINGS, RESULTS* (6,22%); e *MATERIALS, METHODS* (6,03%). Tais seções são seguidas por *BACKGROUND, CONTEXT* e *AIM, OBJECTIVE, PURPOSE*, as quais ocorrem em 4,37%, e 3,83% dos resumos estruturados do *corpus*, sendo, respectivamente, as duas seções menos frequentes da área. Possivelmente, a baixa incidência da seção *AIM, OBJECTIVE, PURPOSE*, nos resumos estruturados da Física, deve-se também ao fato de a referida seção ser expressa no interior da seção *BACKGROUND, CONTEXT*. As observações feitas com base na Tabela 7 apresentam-se sintetizadas na Tabela 8.

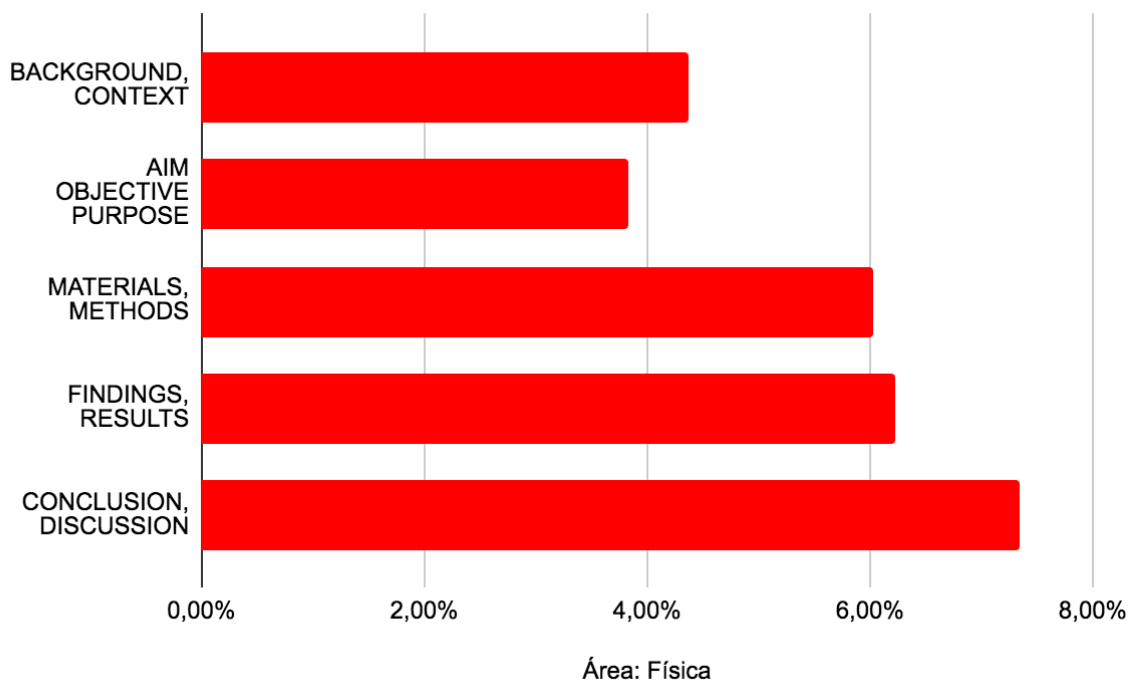
Tabela 8 – Identificação das etiquetas de seção dos resumos estruturados da área de Física.

CATEGORIAS DE SEÇÃO	Física	Número total de resumos do <i>corpus</i>
<i>BACKGROUND, CONTEXT</i>	4,37%	4.690
<i>AIM OBJECTIVE PURPOSE</i>	3,83%	
<i>MATERIALS, METHODS</i>	6,03%	
<i>FINDINGS, RESULTS</i>	6,22%	
<i>CONCLUSION, DISCUSSION</i>	7,35%	

Fonte: Elaborada pela autora.

As observações feitas com base na Tabela 8 podem ser mais bem visualizadas por meio do Gráfico 2, na página seguinte.

⁴⁶ Ao revisar a literatura, contatamos que *AUTHOR SUMMARY*, segundo Breeze (2016), é um novo gênero científico apresentado em alguns periódicos como um “resumo não técnico destinado a estudantes” (BREEZE, 2016, p. 50). Por essa razão, *AUTHOR SUMMARY* não foi incluído na análise.

Gráfico 2 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Física.

Fonte: Elaborado pela autora.

Por fim, a Tabela 9, a seguir, apresenta as etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde. Considerando as três áreas analisadas, a Medicina e Ciências da Saúde tem a menor variedade de etiquetas de seção diferentes (38). Pode-se considerar que esta seja a área com maior convencionalidade com relação à terminologia usada para designar as seções dos resumos estruturados da área.

Tabela 9 – Identificação das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.

Área: Medicina e Ciências da Saúde Nome do corpus: SC0319AbsME1mi Número de resumos acadêmicos: 4.039		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO	Percentual de resumos estruturados deste corpus que tem a seção
BACKGROUND, CONTEXT	Background:	50,92%
	Introduction:	
AIM OBJECTIVE PURPOSE	AIM:	34,61%
	Background and aim:	

	Background and objective:	
	Background and purpose:	
	Objective:	
	Purpose:	
	Hypothesis:	
<i>MATERIALS, METHODS</i>	Design:	73.92%
	Design and Methods:	
	Design and Setting:	
	Exclusion criteria were:	
	Experimental design:	
	Exposure:	
	Inclusion criteria:	
	Material and Method:	
	Measurements:	
	Method:	
	Methodology:	
	Methods and materials:	
	Data sources:	
	Participants:	
	Patients:	
	Procedure:	
	Setting:	
	Study design:	
<i>FINDINGS, RESULTS</i>	Findings:	72.44%
	Principal findings:	
	Result:	
	Main results:	
	Data synthesis:	
<i>CONCLUSION, DISCUSSION</i>	Conclusion:	78.50%
	Discussion:	
	Impact:	
	Implications:	
	Interpretation:	

	Limitation:	
<i>AUTHOR SUMMARY</i>		0,22%

Fonte: Elaborada pela autora.

Ainda sobre a pergunta de pesquisa 1, observamos, na Tabela 9, que as três seções mais recorrentes entre os resumos estruturados são, respectivamente, *CONCLUSION, DISCUSSION* (78,50%); *MATERIALS, METHODS* (73,92%); e *FINDINGS, RESULTS* (72,44%). Tais seções são seguidas por *BACKGROUND, CONTEXT* e *AIM, OBJECTIVE, PURPOSE*, as quais ocorrem em 50,92% e 34,61% dos resumos estruturados do *corpus*, sendo, respectivamente, as três seções menos frequentes da área. Possivelmente, a menor incidência da seção *AIM, OBJECTIVE, PURPOSE*, nos resumos estruturados da Medicina e das Ciências da Saúde, deva-se, também, ao fato de a referida seção ser expressa no interior da seção *BACKGROUND, CONTEXT*. As observações feitas com base na Tabela 9 apresentam-se sintetizadas na Tabela 10.

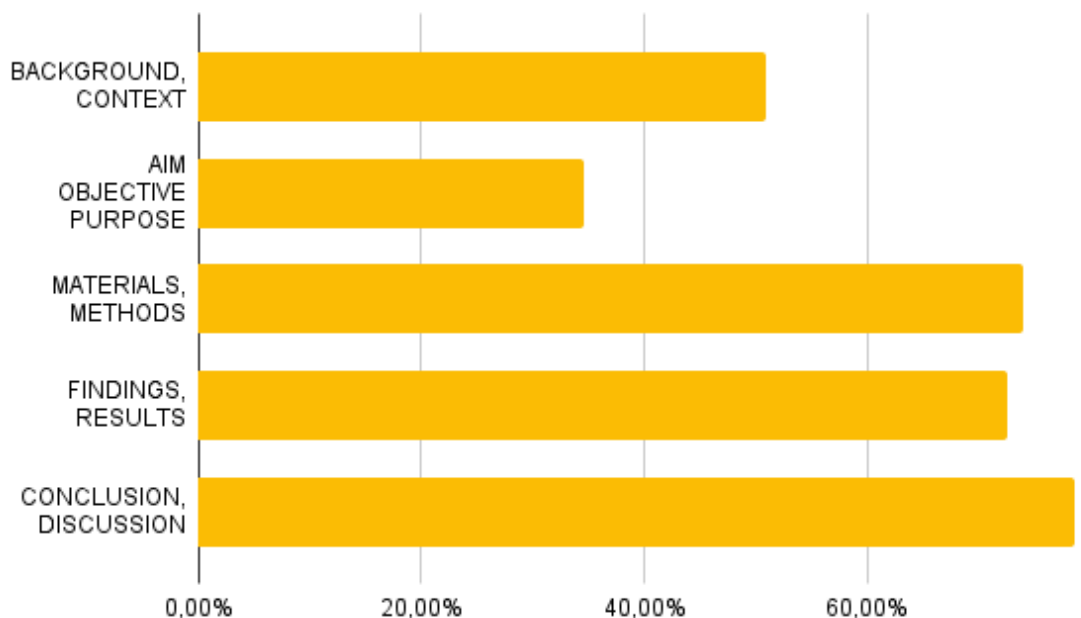
Tabela 10 – Identificação das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.

CATEGORIAS DE SEÇÃO	Medicina e Ciências da Saúde	Número total de resumos do <i>corpus</i>
<i>BACKGROUND, CONTEXT</i>	50,92%	4.039
<i>AIM OBJECTIVE PURPOSE</i>	34,61%	
<i>MATERIALS, METHODS</i>	73,92%	
<i>FINDINGS, RESULTS</i>	72,44%	
<i>CONCLUSION, DISCUSSION</i>	78,50%	

Fonte: Elaborada pela autora.

As observações feitas com base na Tabela 10 podem ser mais bem visualizadas no Gráfico 3, na página seguinte.

Gráfico 3 – Porcentagem das etiquetas de seção dos resumos estruturados da área de Medicina e Ciências da Saúde.



Fonte: Elaborado pela autora.

A integração dos dados obtidos, referentes às seções que constituem os resumos estruturados das três áreas-alvo, aponta, como pode ser observado na Tabela 11, que a área de Medicina e Ciências da Saúde é a que apresenta maior volume de resumos estruturados, seguida da área de Ciências da Computação e da Informação e da área de Física. Em todas as áreas, a seção *CONCLUSION, DISCUSSION* é a mais recorrente, o que pode indicar uma etiqueta de seção com menor variação. Também é interessante notar que, nas três áreas, o percentual de etiquetas relativas a *AIM, OBJECTIVE, PURPOSE* é o mais baixo, quando comparado com as demais etiquetas de seção. Tais dados confirmam as observações feitas por Freitas e Bocorny (2021, p. 3) no sentido da seção *BACKGROUND* ser uma “seção guarda-chuva, na qual podem ser encontradas generalizações sobre o tema, lacunas a serem preenchidas pela investigação e, às vezes, seu objetivo”⁴⁷. Também é interessante notar que a tendência não parece ser apenas da área de Medicina e Ciências da Saúde, mas também da área de Ciências da Computação e da Informação e da área de Física. Uma outra observação interessante é que estudos no âmbito da análise dos movimentos retóricos frequentemente trazem os percentuais de ocorrência dos elementos retóricos analisados. Ao trazerem tais dados, sugerem que os movimentos e passos com menor percentual de ocorrência sejam considerados elementos opcionais. A Tabela 11, a seguir, integra os dados obtidos nas Tabelas 6, 8 e 10.

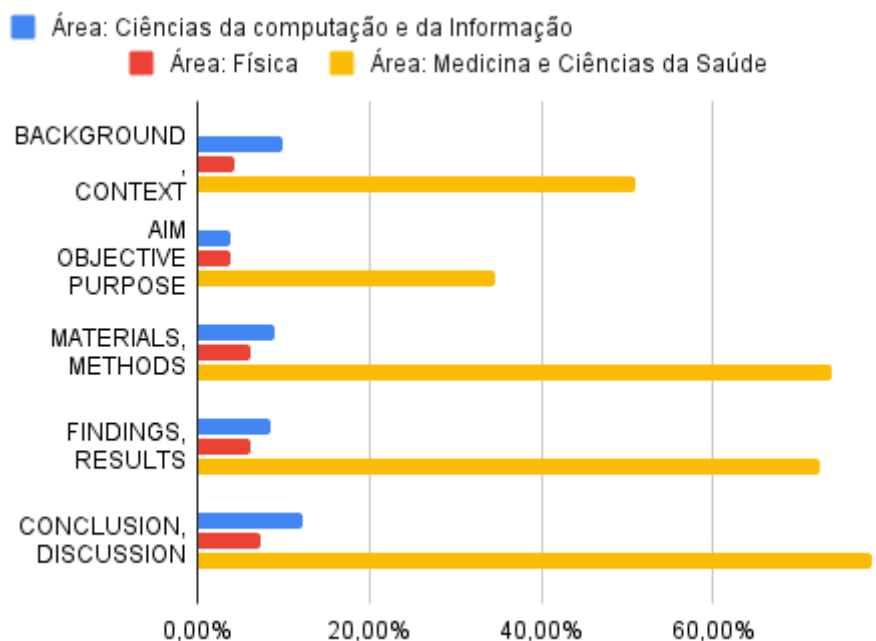
⁴⁷ No original: “[...] umbrella section, in which generalizations about the topic, gaps to be filled by the investigation, and sometimes its aim can be found” (FREITAS; BOCORNY, 2021, p. 3).

Tabela 11 – Seções que constituem os resumos estruturados das áreas-alvo.

	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
<i>BACKGROUND, CONTEXT</i>	9,98%	4,37%	50,92%
<i>AIM OBJECTIVE PURPOSE</i>	3,71%	3,83%	34,61%
<i>MATERIALS, METHODS</i>	9%	6,03%	73,92%
<i>FINDINGS, RESULTS</i>	8,46%	6,22%	72,44%
<i>CONCLUSION, DISCUSSION</i>	12,09%	7,35%	78,50%

Fonte: Elaborada pela autora.

Tais constatações ficam bastante claras quando observamos o Gráfico 4, a seguir. Há um volume muito maior de resumos estruturados na área de Medicina e Ciências da Saúde do que nas demais áreas estudadas.

Gráfico 4 – Porcentagem das seções que constituem os resumos estruturados das áreas-alvo.

Fonte: Elaborado pela autora.

Após a análise dos dados referentes a todas as etiquetas de seção extraídas, fizemos a exclusão das etiquetas que ocorriam menos de 10 vezes. O Quadro 10, a seguir, apresenta todas as etiquetas que aparecem 10 vezes ou mais nos *corpora*. Os Anexos 1, 2 e 3, apresentam

informações mais detalhadas sobre as etiquetas de seção dos resumos estruturados das três áreas de estudo que ocorrem 10 ou mais vezes.

Quadro 10 – Identificação das etiquetas de seção dos resumos estruturados das três áreas de estudo que ocorrem 10 ou mais vezes.

Categorias de seção	Estrutura retórica proposta por Swales e Feak (2009)	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
BACKGROUND, CONTEXT	Contextualização/introdução/situação	Background: (409x)	Background: (178x)	Background: (1.796x)
		Introduction: (45x)	Introduction: (26x)	Introduction: (261x)
AIM OBJECTIVE PURPOSE	Apresentar a pesquisa/objetivo	Aim: (10x)	Aim(s): (15x)	Aim: (146x)
		Motivation: (15x)	Objective(s): (80x)	Background and aim: (70x)
		Objective(s): (104x)	Purpose: (69x)	Background and objective: (26x)
		Purpose: (24x)		Background and purpose: (45x)
				Objective: (829x)
				Purpose: (272x)
MATERIALS, METHODS	Metodologia/materiais/ sujeitos/ procedimentos	Materials and methods: (30x)	Materials and Methods: (32x)	Design: (87x)
		Method: (29x)	Method(s): (208x)	Material and Method: (187x)
		Methodology: (34x)		Measurements: (10x)
		Methods: (231x)		Method: (2,463x)
		Methods and findings: (36x)		Methodology: (40x)
		Methods and results: (11x)		Methods and materials: (22x)
				Data sources: (15x)
		Participants: (37x)		
		Setting: (52x)		
			Study design: (34x)	
FINDINGS, RESULTS	Resultados/descobertas	Results: (361x)	Result(s): (266x)	Findings: (82x)
				Result: (2,833x)

CONCLUSION, DISCUSSION	Discussão/conclusões/im- plicações/ recomendações	Conclusion/Si- gnificance: (10x)	Conclusions/Si- gnificance: (29x)	Conclusion: (3.071x)
		Conclusion: (157x)	Conclusion(s): (290x)	Discussion: (55x)
		Conclusions/S- ignificance: (88x)	Discussion: (10x)	Interpretation: (25x)
		Conclusions: (266x)		
		Discussion: (21x)		
		Conclusion/Si- gnificance: (10x)		

Fonte: Elaborado pela autora.

De forma a identificar as etiquetas de seção mais convencionais em cada uma das áreas estudadas, mantivemos apenas a etiqueta de seção mais frequente. Pode-se notar que, quando excluídas as etiquetas de seção menos frequentes, o padrão retórico resultante equivale ao proposto por Swales e Feak (2009), como mostra o Quadro 11, a seguir.

Quadro 11 – Comparação entre a estrutura retórica proposta por Swales e Feak (2009) e a identificada neste estudo para cada uma das áreas-alvo, por meio das etiquetas de seção.

Categorias de seção	Estrutura retórica proposta por Swales e Feak (2009)	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
BACKGROUND, CONTEXT	Contextualização/ introdução/situação	Background: (409x)	Background: (178x)	Background: (1.796x)
AIM OBJECTIVE PURPOSE	Apresentar a pesquisa/objetivo	Objective(s): (104x)	Objective(s): (80x)	Objective(s): (829x)
MATERIALS, METHODS	Metodologia/materiais/ sujeitos/procedimentos	Method(s): (231x)	Method(s): (208x)	Method(s): (2.463x)
FINDINGS, RESULTS	Resultados/ descobertas	Result(s): (361x)	Result(s): (266x)	Result(s): (2.833x)
CONCLUSION, DISCUSSION	Discussão/conclusões/ implicações/ recomendações	Conclusion(s): (157x)	Conclusion(s): (290x)	Conclusion(s): (3.071x)

Fonte: Elaborado pela autora.

3.2 Pergunta de pesquisa 2: Quais ELs são recorrentes nos resumos das áreas-alvo?

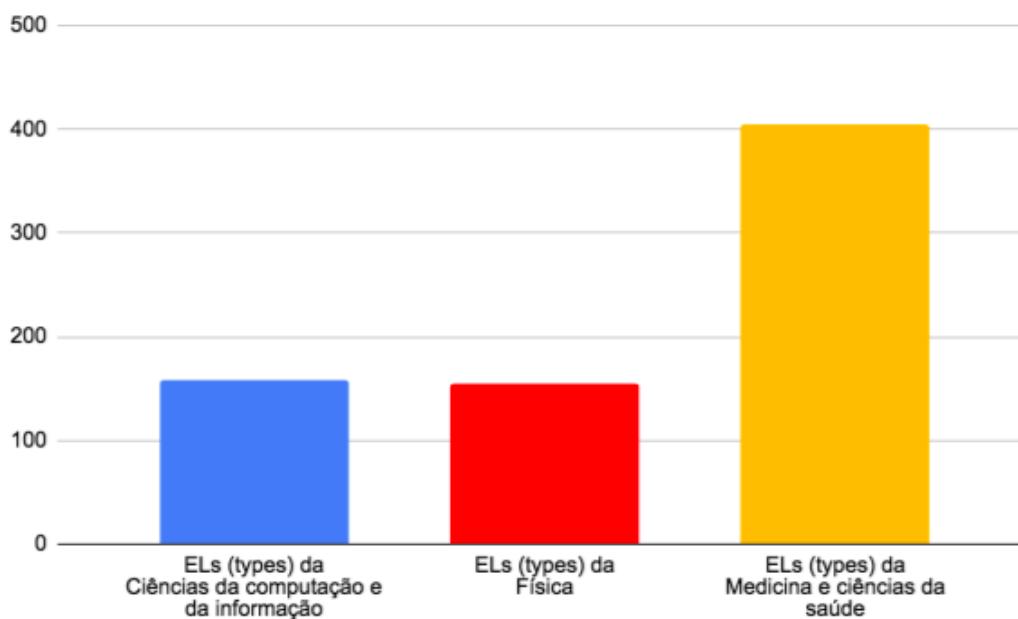
Nesta seção, são apresentados e discutidos os dados quantitativos referentes à nossa segunda questão de pesquisa (quais ELs são recorrentes nos resumos das áreas-alvo?). Para responder a essa questão de pesquisa, um total de 717 ELs foram extraídas dos *corpora* de estudo, conforme os critérios de extração estabelecidos (cf. seção 2.4.2.1). A aplicação dos mesmos parâmetros de extração resultou em quantidades diferentes de ELs em cada área. Enquanto na Medicina e nas Ciências da Saúde temos 404 ELs diferentes, na Física e na Ciências da Computação e da Informação, temos 154 e 159 ELs, respectivamente, como pode ser observado na Tabela 12. As listas completas de ELs extraídas de cada um dos *corpora* são apresentadas nos Anexos 10, 11 e 12.

Tabela 12 – Número de *N-gram types* (sem repetição), *N-gram tokens* (com repetição).

	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
Número de ELs (sem repetição – <i>types</i>)	159	154	404
Número de ELs (com repetição – <i>tokens</i>)	8.860	8.509	23.498

Fonte: Elaborada pela autora.

As diferenças na quantidade de ELs extraídas de cada uma das áreas estudadas ficam ainda mais evidentes quando observamos o Gráfico 5, na página seguinte.

Gráfico 5 – ELs recorrentes nos resumos das áreas-alvo.

Fonte: Elaborado pela autora.

Após extrair as 50 ELs mais frequentes de cada *corpora* e compará-las, é possível observar (cf. Anexo 13) que apenas 6 unidades ocorrem nas três áreas do conhecimento, enquanto 35 são comuns a duas áreas, como mostra a Tabela 13, na página seguinte. A grande maioria, 64 das ELs, ocorrem em apenas um *corpus*. Tais dados confirmam observações feitas em estudos anteriores (BIBER, 2006; HYLAND, 2008; CORTES, 2004, 2013; MIZUMOTO, 2017; GRAY et al., 2020), os quais indicam a existência de variação no uso de elementos linguísticos, em especial pacotes lexicais e ELs, em textos acadêmicos produzidos em diferentes áreas do conhecimento. Resultados obtidos por Hyland (2008, p. 20) (“menos da metade dos 50 pacotes lexicais mais frequentes ocorrendo nas listas das diferentes áreas estudadas”⁴⁸), semelhantes aos deste estudo, indicam que pesquisadores de diferentes áreas “recorrem a diferentes recursos para desenvolver seus argumentos, estabelecer sua credibilidade e persuadir seus leitores”⁴⁹.

⁴⁸ No original: “[...] less than half of the top 50 bundles in each list occurring in any other list” (HYLAND, 2008, p. 20).

⁴⁹ No original: “[...] draw on different resources to develop their arguments, establish their credibility and persuade their readers” (HYLAND, 2008, p. 20).

Tabela 13 – ELs comuns às áreas do conhecimento estudadas.

Ciências da Computação e da Informação			Física			Medicina e Ciências da Saúde		
EL	Freq. Abs.	Dispersão	EL	Freq. Abs.	Dispersão	EL	Freq. Abs.	Dispersão
ELs que ocorrem nos três corpora								
to + the + of the	289	270	to + the + of the	250	236	to + the + of the	85	85
the + of this + is	115	113	the + of this + is	67	67	the + of this + is	88	88
used to + the + of	107	102	used to + the + of	102	98	used to + the + of	95	93
this study + to + the	66	65	this study + to + the	102	102	this study + to + the	434	434
the + of this + was	61	61	the + of this + was	140	140	the + of this + was	358	357
aim of this + + to	59	59	aim of this + + to	89	89	aim of this + + to	256	256
ELs que ocorrem em dois corpora								
in this + we + a	307	306	in this + we + a	101	101	-	-	-
in this + we + the	196	193	in this + we + the	143	143	-	-	-
the + of the + of	159	149	the + of the + of	127	117	-	-	-
the + and + of the	154	149	the + and + of the	145	139	-	-	-
we + the + of the	116	112	we + the + of the	65	65	-	-	-
of the + + of the	106	99	of the + + of the	152	145	-	-	-
the + of the + and	96	96	the + of the + and	139	128	-	-	-
and + the + of the	94	92	and + the + of the	79	78	-	-	-
we + that the + of	94	89	we + that the + of	74	74	-	-	-
the + results + that the	87	85	the + results + that the	59	56	-	-	-
of the + + in the	69	69	of the + + in the	97	90	-	-	-
in the + + of the	66	60	in the + + of the	115	111	-	-	-
can be + to + the	65	64	can be + to + the	64	63	-	-	-
and the + + of the	61	60	and the + + of the	133	125	-	-	-

and + + as well as	60	60	and + + as well as	58	56	-	-	-
the + and the + of	59	59	the + and the + of	76	72	-	-	-
the + of the + in	58	58	the + of the + in	63	62	-	-	-
in this study + + the	58	58	in this study + + the	81	81	-	-	-
the + of the + is	57	54	the + of the + is	57	56	-	-	-
of the + + and the	56	54	of the + + and the	97	93	-	-	-
to the + + of the	56	55	to the + + of the	106	104	-	-	-
on the + + of the	55	55	on the + + of the	101	99	-	-	-
results + that the + of	52	51	results + that the + of	59	58	-	-	-
-	-	-	the + of + study was	121	121	the + of + study was	392	391
-	-	-	the + + this study was	113	113	the + + this study was	347	346
-	-	-	aim of + study + to	77	77	aim of + study + to	290	290
-	-	-	aim of + + was to	71	71	aim of + + was to	252	252
-	-	-	purpose of this + + to	70	70	purpose of this + + to	98	98
-	-	-	aim + this study + to	68	68	aim + this study + to	243	243
-	-	-	this + was to + the	67	67	this + was to + the	225	225
-	-	-	the aim of + + was	65	65	the aim of + + was	238	238
-	-	-	aim + this + was to	63	63	aim + this + was to	216	216
-	-	-	aim + + study was to	60	60	aim + + study was to	242	242
-	-	-	this study was + + the	59	58	this study was + + the	221	221
-	-	-	the aim + this + was	58	58	the aim + this + was	206	206

ELs que ocorrem em um *corpus*

-	-	-	-	-	-	-	this ++ to evaluate the	97	97
in this paper + + a	209	209	-	-	-	-	-	-	-
in + paper we + a	205	205	-	-	-	-	-	-	-
in ++ we propose a	132	132	-	-	-	-	-	-	-
in this + + propose a	122	122	-	-	-	-	-	-	-
in this paper + + the	120	119	-	-	-	-	-	-	-
in + paper + propose a	96	96	-	-	-	-	-	-	-
in + paper we + the	94	93	-	-	-	-	-	-	-
the + of + in the	84	84	-	-	-	-	-	-	-
based on the ++ of	74	72	-	-	-	-	-	-	-
in this + we + an	70	70	-	-	-	-	-	-	-
this + we + a novel	69	69	-	-	-	-	-	-	-
order to + the + of	66	65	-	-	-	-	-	-	-
this + we + a new	60	60	-	-	-	-	-	-	-
as well as ++ of	60	58	-	-	-	-	-	-	-
in terms of ++ and	60	60	-	-	-	-	-	-	-
in ++ we present a	55	55	-	-	-	-	-	-	-
the ++ in this paper	55	55	-	-	-	-	-	-	-
to + the + of a	55	54	-	-	-	-	-	-	-
in the + of + and	54	53	-	-	-	-	-	-	-
the ++ show that the	53	53	-	-	-	-	-	-	-

in this + + present a	52	52	-	-	-	-	-	-
-	-	-	that the + + of the	79	76	-	-	-
-	-	-	this study + the + of	71	71	-	-	-
-	-	-	in + study we + the	70	70	-	-	-
-	-	-	here we + the + of	70	69	-	-	-
-	-	-	the effect of ++ on	69	67	-	-	-
-	-	-	the + of + of the	66	60	-	-	-
-	-	-	between the ++ and the	64	62	-	-	-
-	-	-	of the + + on the	62	60	-	-	-
-	-	-	of ++ as well as	57	56	-	-	-
-	-	-	-	-	-	the aim + + study was	229	229
-	-	-	-	-	-	associate d with + + risk of	197	164
-	-	-	-	-	-	was to + the + of	167	165
-	-	-	-	-	-	associate d with a + + of	156	131
-	-	-	-	-	-	this + aimed to + the	126	126
-	-	-	-	-	-	this study aimed + + the	123	123
-	-	-	-	-	-	the associati on between + + and	115	103
-	-	-	-	-	-	associate d with + increased + of	111	98
-	-	-	-	-	-	purpose of + study + to	110	110
-	-	-	-	-	-	aimed to + the + of	108	108

-	-	-	-	-	-	associate d with an + + of	108	96
-	-	-	-	-	-	associate d + an increased + of	103	91
-	-	-	-	-	-	associate d + a + risk of	100	83
-	-	-	-	-	-	objective of + study + to	99	99
-	-	-	-	-	-	of + + in patients with	97	90
-	-	-	-	-	-	this + + to evaluate the	97	97
-	-	-	-	-	-	the + of + study is	95	95
-	-	-	-	-	-	purpose + this study + to	94	94
-	-	-	-	-	-	this study + + evaluate the	94	94
-	-	-	-	-	-	of + study + to evaluate	92	92
-	-	-	-	-	-	associate d + + increased risk of	91	81
-	-	-	-	-	-	was associate d with + + of	90	79
-	-	-	-	-	-	purpose of + + was to	88	88
-	-	-	-	-	-	objective of this + + to	88	88
-	-	-	-	-	-	the purpose of + + was	87	87
-	-	-	-	-	-	associate d + an + risk of	86	77
-	-	-	-	-	-	hazard + + confiden ce	84	84

						interval ci		
-	-	-	-	-	-	purpose + + study was to	84	84
-	-	-	-	-	-	objective + this study + to	84	84
-	-	-	-	-	-	of this + + to evaluate	84	84
-	-	-	-	-	-	hazard ratio + + interval ci	83	83
-	-	-	-	-	-	the purpose + + study was	83	83

Fonte: Elaborada pela autora.

Na mesma linha das observações feitas por Hyland (2008) para pacotes lexicais, os resultados descritos nesta seção têm implicações importantes para o ensino de IFA. Em especial, destacamos a constatação de que ELs ocorrem com frequências diferentes em cada área de especialidade. Sendo assim, como afirma Hyland (2008), é importante que o ponto de partida de um curso de IFA seja o contexto de atuação dos alunos. Tal fato tem importantes consequências pedagógicas e reforça o nosso entendimento de que aplicações pedagógicas (por exemplo, livros didáticos, programas de cursos, tarefas) não devem seguir um modelo de *one size fits all*, como é o caso de muitos livros didáticos para cursos de IFA.

Portanto, são importantes instrumentos pedagógicos listas de pacotes lexicais e de ELs extraídas de *corpora* especializados – sobretudo aquelas que, como as geradas neste estudo (cf. Anexos 10, 11 e 12), são específicas de seções (*section specific*) e de disciplinas (*discipline specific*). Tais listas podem ser usadas, por exemplo, por professores de IFA (em sala de aula ou para desenvolver materiais) e pelos próprios pesquisadores (como obras de referência ao escreverem seus textos acadêmicos).

De acordo com Hyland (2008), vários estudos mostram que “a melhor maneira de preparar os alunos (universitários) para seus estudos não é ensinar elementos universais, mas sim oferecer a esses alunos um entendimento das características dos discursos que eles irão encontrar em seus cursos”⁵⁰ (HYLAND, 2008, p. 20).

⁵⁰ No original: “[...] the best way to prepare students for their studies is not to search for universally appropriate teaching items, but to provide them with an understanding of the features of the discourses they will encounter in their particular courses” (HYLAND, 2008, p. 20).

3.3 Pergunta de pesquisa 3: Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?

Para responder à pergunta de pesquisa 3 (quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?), selecionamos uma amostra com as 50 ELs mais frequentes de cada área. A observação da amostra de 150 ELs foi realizada em três etapas, conforme os procedimentos metodológicos já descritos (cf. seção 2.4.3): (i) a observação dos elementos lexicais que não variam (*frame*) nas ELs; (ii) a observação das palavras que completam as lacunas (*slot fillers*) das ELs; e (iii) a observação das ELs em contexto.

A observação da amostra de 150 ELs, no que tange às funções retóricas que realizam nos resumos acadêmicos, indica a existência de duas grandes categorias de unidades multipalavras descontínuas: (i) as ELs transparentes e (ii) as ELs opacas. A categoria por nós denominada ELs transparentes representa o grupo de ELs que tem a sua função retórica mais facilmente identificável a partir da observação: (i) dos elementos fixos que constituem a sua estrutura (por exemplo, *the aim of + + was*); (ii) das palavras que preenchem as lacunas variáveis (por exemplo, em *we + that the + of*, a observação das palavras que preenchem a lacuna 01 a. *conclude*, b. *demonstrate*, c. *found*); e (iii) dos contextos de ocorrência (por exemplo, *Results: In the first phase of the study, with over 56 months of recruitment using traditional sources, 35 women were enrolled in the study, resulting in a mean rate of ±0.62 recruits/month.*).

Por sua vez, a categoria designada neste estudo como ELs opacas reúne ELs que não têm sua função retórica facilmente identificável por meio da observação dos elementos descritos anteriormente (elementos fixos, palavras que preenchem as lacunas variáveis e contextos de ocorrência). Alguns exemplos de ELs que categorizamos como “opacas” são *to + the + of the* e *the + of the + of*, na área de Ciências da Computação e da Informação; *of the + + of the* e *to + the + of the*, na área de Física; e *to + the + of the*, na área de Medicina e Ciências da Saúde.

A Tabela 14, na página seguinte, mostra o número de ELs identificadas como “transparentes” nas áreas estudadas: 33 (66%) na área de Medicina e Ciências da Saúde; 32 (64 %) na área de Física; e 24 (48%) na área de Ciências da Computação e da Informação, como mostra a Tabela 14.

Tabela 14 – Número de ELs identificadas como “transparentes” nas áreas estudadas.

	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
Número total de ELs extraídas (sem repetição – types)	159	154	404
Amostra usada na análise das funções retóricas das ELs	50	50	50
Número de ELs transparentes	24	32	33
Número de ELs opacas	26	18	17

Fonte: Elaborada pela autora.

As tabelas onde as ELs foram organizadas e categorizadas (cf. Apêndices 4, 5 e 6) e o Quadro 12, que resume dados dos Apêndices 4, 5 e 6, foram codificadas por cores, para facilitar-se a identificação das categorias que representam seções dos resumos estruturados. Os códigos usados e suas respectivas etiquetas foram os seguintes:

- lavanda, para *BACKGROUND*, *CONTEXT*;
- lavanda, para etiquetas da seção *AIM*, *OBJECTIVE*, *PURPOSE*;
- verde, para *MATERIALS*, *METHODS*;
- rosa, para *FINDINGS*, *RESULTS*;
- laranja, para *CONCLUSION*, *DISCUSSION* e

Em função das ELs que expressam o objetivo de um estudo ocorrerem com frequência na seção *BACKGROUND*, *CONTEXT* optou-se por colocar as seções *AIM*, *OBJECTIVE*, *PURPOSE* e *BACKGROUND*, *CONTEXT* em uma mesma categoria (cor lavanda).

De uma forma geral, a análise da amostra de 94 ELs transparentes mostrou que a grande maioria destas ELs realizam a função retórica de apresentar o objetivo do estudo, como mostra o Quadro 12, a seguir.

Quadro 12 – Funções retóricas realizadas pelas ELs transparentes.

Funções retóricas realizadas	Ciências da Computação e da Informação	Física	Medicina e Ciências da Saúde
---	---	---------------	---

<p>BACKGROUND, CONTEXT</p> <p>AIM, OBJECTIVE, PURPOSE</p> <p>Apresentar o objetivo do estudo</p>	1. in this + we + a	4. In this + we + the	1. this study + to + the
	3. in this paper ++ a	3. the + and + of the	2. the + of + study was
	4. in + paper we + a	4. In this + we + the	3. the + of this + was
	50. in this ++ present a	20. in this study ++ the	4. the ++ this study was
	43. in ++ we present a	16. in this + we + a	5. aim of + study + to
	15. in + paper + propose a	5. the + of this + was	6. aim of this ++ to
	8. in ++ we propose a	9. the + of + study was	7. aim of ++ was to
	9. in this ++ propose a	34. this + was to + the	8. aim + this study + to
	24. this + we + a novel	19. aim of this ++ to	9. aim ++ study was to
	23. in this + we + an	23. aim of + study + to	10. the aim of ++ was
	5. in this + we + the	26. aim of ++ was to	11. the aim ++ study was
	10. in this paper ++ the	27. this study + the + of	12. this + was to + the
	18. in + paper we + the	32. aim + this study + to	13. this study was ++ the
	23. in this + we + an	33. the + of this + is	24. aimed to + the + of
	33. this + we + a new	37. the aim of ++ was	31. this ++ to evaluate the
	39. in this study ++ the	40. aim + this + was to	1. this study + to + the
	12. the + of this + is	43. aim ++ study was to	14. aim + this + was to
	30. the + of this + was	46. this study was ++ the	15. the aim + this + was
	37. aim of this ++ to	48. the aim + this + was	32. the + of + study is
	28. this study + to + the	50. the + of the + is	39. the + of this + is
	39. in this study ++ the	30. purpose of this ++ to	23. purpose of + study + to to
		13. this study + to + the	29. purpose of this ++ to
		28. in + study we + the	35. this study ++ evaluate the
		25. we + that the + of	36. of + study + to evaluate
		29. here we + the + of	40. purpose of ++ was to
		36. we + the + of the	42. the purpose of ++ was
			46. purpose ++ study was to
			48. of this ++ to evaluate

			50. the purpose + + study was
			28. objective of + study + to
			41. objective of this + + to
			47. objective + this study + to
			19. this + aimed to + the
			20. this study aimed + + the
MATERIALS, METHODS	1. in this + we + a		
	3. in this paper + + a		
Metodologia/ materiais/ sujeitos/procedimentos	4. in + paper we + a		
	50. in this + + present a		
	11. we + the + of the		
FINDINGS, RESULTS	20. the + results + that the	44. the + results + that the	
	48. the + + show that the		
Resultados/ descobertas	49. results + that the + of		
CONCLUSION, DISCUSSION	44. the + + in this paper	25. we + that the + of	
		45. results + that the + of	
Discussão/ conclusões/ implicações/ recomendações		44. the + results + that the	

Fonte: Elaborado pela autora.

A observação dos elementos lexicais que não variam (*frame*) nas ELs revelou que 24 ELs extraídas apresentam, em sua composição, palavras lexicais que indicam a função retórica que as referidas estruturas realizam linguisticamente nos resumos. Tais ELs apresentam em sua composição as palavras lexicais *aim (ed)* (17 ocorrências), *purpose* (7 ocorrências), *objective* (3 ocorrências) e *results* (2 ocorrências), como é possível observar nas ELs em negrito no Quadro 12. Assim, as ELs transparentes identificadas nesta etapa do estudo realizam linguisticamente as funções retóricas de apresentar os objetivos do estudo, descrever a metodologia, relatar resultados e apresentar as conclusões e/ou discutir resultados.

A observação das palavras que completam as lacunas (*slot fillers*) das ELs e dos contextos nos quais as ELs ocorrem permitiu a identificação das demais 65 ELs classificadas

como transparentes. Observou-se que essas 65 ELs exercem as funções retóricas semelhantes às 24 citadas no parágrafo anterior, quais sejam: i) apresentar os objetivos do estudo; ii) descrever a metodologia; iii) relatar os resultados e iv) apresentar as conclusões e/ou as discussões.

Assim, há ELs que, na construção do discurso acadêmico, assumem o papel de realizar linguisticamente funções retóricas específicas (por exemplo, apresentar o objetivo de um estudo). Tais ELs, chamadas neste estudo de ELs funcionais retóricas, estão mais vinculadas a uma unidade retórica (seção, movimento ou passo). É importante mencionar, no entanto, que há ELs funcionais retóricas que expressam uma mesma função (por exemplo, apresentar o objetivo de um estudo) em diferentes seções do resumo. É o caso, por exemplo, de *in this + we + a*, na área de Ciências da Computação e da Informação, que ocorre expressando linguisticamente a função de apresentar o objetivo de um estudo tanto na seção *BACKGROUND, CONTEXT* quanto na seção *AIM, OBJECTIVE, PURPOSE*.

Da mesma forma, como já observado em Freitas e Bocorny (2021), a presença de palavras lexicais (por exemplo, *results* e *aim*) na EL não indica, necessariamente, que tais ELs ocorram, exclusivamente, na seção *FINDINGS, RESULTS* e *AIM, OBJECTIVE, PURPOSE*. A EL *results + that the + of*, por exemplo, ocorre realizando a mesma função de apresentar as conclusões e/ou discutir resultados, tanto na seção *CONCLUSIONS, DISCUSSION* (por exemplo, “*Conclusions: Our results indicate that the notion of scientific impact is a multi-dimensional construct that can not be adequately measured by any single indicator, although some measures are more suitable than others.*”) quanto na seção *FINDINGS, RESULTS* (por exemplo, “*Results : Results showed that the majority of patients' PHQ-8 and Item 9 scores displayed strong temporal correlations.*”).

A observação das ELs transparentes mais recorrentes nas áreas estudadas também revelou que algumas unidades contêm, em sua estrutura (*frame*) ou nas palavras que preenchem as lacunas (*slot fillers*), palavras lexicais muito associadas às áreas de especialidade. Essa observação é mais recorrente na área da Medicina e Ciências da Saúde, com ELs como *of + + in patients with, associated + a + risk of, associated + + increased risk of, associated + a + risk of, of + + in patients with*, onde as palavras lexicais *patients* e *risk* são vinculadas à área de especialidade em questão. Neste estudo, chamamos essas ELs de *terminológicas*.

Logo, enquanto há ELs retóricas que assumem funções retóricas na construção do discurso acadêmico, realizando linguisticamente funções como apresentar o objetivo de um estudo e que, por essa razão, estão mais vinculadas a um movimento ou seção retórica, há ELs terminológicas, que estão mais vinculadas às áreas de especialidade do que a funções retóricas

na construção do discurso acadêmico. Tais ELs realizam a importante função de referir, por exemplo, termos, procedimentos e conceitos consagrados e convencionais das áreas especializadas. Tais ELs terminológicas são menos frequentes do que as ELs retóricas, sendo, por isso, necessários pontos de corte mais baixos para que possamos identificá-las no *corpus*. Também é possível imaginar que essas ELs sejam mais recorrentes em áreas de especialidade em que o discurso acadêmico esteja mais consagrado, ou seja, em que especialistas usem as mesmas construções formulaicas para designar, por exemplo, termos, procedimentos e conceitos. Tal hipótese é confirmada pelos dados obtidos neste estudo. Como pode ser observado na Tabela 13, as ELs terminológicas, destacadas em vermelho, ocorrem apenas na área de Medicina e Ciências da Saúde.

Quanto às ELs opacas, ainda que não sejam o foco deste estudo, podemos supor que elas, por não estarem aparentemente vinculadas a funções retóricas, transitam mais livremente através das seções e movimentos retóricos (“*across rhetorical moves and sections*”). O Quadro 13, a seguir, apresenta, de forma sintética, a tipologia e a caracterização das ELs funcionais, conforme observações realizadas neste estudo.

Quadro 13 – Tipologia e caracterização das ELs funcionais.

Tipo	Subtipo	Características			
		Composição	Função	Localização	Frequência
Transparentes	Retóricas	Apresentam palavras lexicais que indicam função retórica (por exemplo, <i>aim, purpose, results</i>), tanto na composição da sua estrutura (<i>frame</i>) quanto nas palavras que preenchem as lacunas (<i>slot fillers</i>).	Realizam linguisticamente uma ou mais de uma função retórica expressa nas unidades retóricas (seção, movimento, passo) dos gêneros em estudo.	Geralmente estão mais vinculadas a uma unidade retórica (seção, movimento, passo). Podem ocorrer, com menor frequência, em mais de uma unidade retórica.	Ocorrem com maior frequência em certas funções retóricas (por exemplo, expressar o objetivo de um estudo).
	Terminológicas	Apresentam palavras lexicais que indicam vinculação a uma área de especialidade (por exemplo, <i>risk, hazard patients</i>), tanto na composição da sua estrutura (<i>frame</i>) quanto nas palavras que	Expressam elementos (por exemplo, conceitos, procedimentos, metodologias) relacionados a uma área de especialidade.	Geralmente ocorrem em mais de uma unidade retórica (seção, movimento, passo).	São menos frequentes que as ELs retóricas. Em função disso, para que sua identificação no <i>corpus</i> seja possível, deve-se considerar pontos de corte mais baixos

		preenchem as lacunas (<i>slot fillers</i>)			
Opacas	Não são objeto deste estudo				

Fonte: Elaborado pela autora.

Este capítulo, dividido em três seções, apresentou os resultados e fez a análise dos dados obtidos à luz de pesquisas e de teorias nas quais o presente estudo se baseia. A seção 1 tratou dos resultados relacionadas à primeira questão de pesquisa (quais seções constituem os resumos estruturados das áreas-alvo?). Nessa seção, foi possível ter uma visão geral da estrutura retórica do gênero em cada uma das áreas estudadas. Já a seção 2 iniciou com dados quantitativos referentes ao número total de ELs extraídas de cada um dos *corpora* do estudo, conforme os critérios de extração estabelecidos (cf. seção 2.4.2.1). Ficou contemplada, nessa seção, nossa segunda questão de pesquisa (quais ELs são recorrentes nos resumos das áreas-alvo?). Por fim, na seção 3, examinamos a natureza da relação entre as seções retóricas dos resumos estudados e as 50 ELs mais frequentes em cada *corpus*, para respondermos a nossa terceira questão de pesquisa (quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?).

CONSIDERAÇÕES FINAIS

A principal motivação de Swales para desenvolver seu modelo analítico de textos (*move analysis*), conforme Moreno e Swales (2018, p. 40), foi "ajudar estudantes avançados, que não tinham o inglês como primeira língua, a melhorarem sua leitura e escrita de artigos de pesquisa em inglês"⁵¹ (MORENO; SWALES, 2018, p. 40). Da mesma forma, uma das principais motivações para a realização da presente pesquisa foi a possibilidade de seu uso pedagógico.

A partir do exposto, este estudo teve por **objetivo geral** realizar a análise funcional retórica das ELs em resumos acadêmicos de três áreas do conhecimento: (i) Ciências da Computação e da Informação, (ii) Física e (iii) Medicina e Ciências da Saúde. Os **objetivos específicos** do estudo foram os seguintes:

1. Identificar o padrão retórico de resumos estruturados das áreas-alvo.
2. Identificar as ELs recorrentes nos resumos das áreas-alvo.
3. Relacionar as ELs identificadas às funções retóricas realizadas nas seções retóricas dos resumos das áreas-alvo.

Em linhas gerais, as observações feitas neste estudo são resumidas nos parágrafos que seguem.

Pergunta 1: Quais seções constituem os resumos estruturados das áreas-alvo?

Esta etapa de nosso estudo mapeou a estrutura retórica de resumos acadêmicos de três áreas de especialidade (Ciências da Computação e da Informação; Física; Medicina e Ciências da Saúde), com base em aproximadamente 3.975 resumos estruturados identificados (≈ 459 da área de Ciências da Computação e da Informação; ≈ 345 da área de Física; ≈ 3.171 da área de Medicina e Ciências da Saúde), os quais fazem parte dos três *corpora* compilados para esta pesquisa. A partir de princípios oriundos da área de estudos dos Gêneros do Discurso, de teorias e conceitos da linguagem formulaica e da Linguística de *Corpus* (cf. capítulo 1), esta etapa do trabalho identificou as seções retóricas mais recorrentes nos resumos estruturados (cf. capítulo

⁵¹ No original: “[...] to help advanced students for whom English is not their first language to improve their reading and writing of RAs in English” (MORENO; SWALES, 2018, p. 40).

2) das áreas estudadas como forma de ter uma indicação do padrão retórico convencional de cada domínio.

Assim, à luz dos dados obtidos, um esquema retórico foi proposto, o que resultou em um mapeamento indicativo da estrutura retórica dos resumos considerada convencional para cada área. Os resultados mostram que, nas três áreas, as seções *CONCLUSION*, *DISCUSSION* são as mais recorrentes e, por consequência, as mais convencionais em todas as áreas estudadas. Em contrapartida, o percentual de etiquetas relativas a *AIM*, *OBJECTIVE*, *PURPOSE* é o mais baixo nas áreas. Tais dados confirmam as observações de Freitas e Bocorny (2021, p. 3) sobre o fato de a seção *BACKGROUND* apresentar, muitas vezes, além da generalização acerca do tema, estruturas lexicais com a função de apresentar o objetivo da pesquisa. De uma forma geral, no entanto, foi possível constatar que as seções que são convencionais nos resumos estruturados das áreas-alvo são as mesmas elencadas por Swales e Feak (2009): contextualização/introdução/situação (*BACKGROUND*), apresentar a pesquisa/ objetivo (*AIM*, *OBJECTIVE*, *PURPOSE*), Metodologia/ materiais/ sujeitos/ procedimentos (*MATERIALS*, *METHODS*), resultados/descobertas e discussão/ conclusões/ implicações/ recomendações (*CONCLUSION*, *DISCUSSION*).

Pergunta 2: Quais ELs são mais recorrentes nos resumos das áreas-alvo?

Esta etapa de nosso estudo extraiu as ELs mais frequentes de resumos acadêmicos de três áreas de especialidade (Ciências da Computação e da Informação; Física; Medicina e Ciências da Saúde). Os critérios para a extração das ELs dos *corpora* foram: (i) ser uma sequência de seis palavras, (ii) ter duas lacunas variáveis, (iii) ocorrer pelo menos 30 vezes por milhão de palavras (pmp) nos *corpora* analisados, (iv) constar em pelo menos 20 textos nos *corpora* analisados.

Um total de 717 ELs (*n-gram types*) foram extraídas dos *corpora* em estudo, sendo 159 da área de Ciências da Computação e da Informação, 154 da área de Física e 404 da área de Medicina e Ciências da Saúde. Em seguida, as 50 ELs mais frequentes de cada *corpora* foram selecionadas para serem analisadas nos passos seguintes da pesquisa. Ao comparar as 50 ELs de cada área, foi possível observar que apenas 6 unidades ocorreram nas três áreas do conhecimento, enquanto 35 são comuns a duas áreas. A grande maioria, 64 das ELs, ocorrem em apenas um *corpus*. Esses dados corroboram os resultados de trabalhos anteriores (BIBER, 2006; HYLAND, 2008; CORTES, 2004, 2013; MIZUMOTO, 2017; GRAY et al., 2020) que

indicam a existência de variação no uso de elementos linguísticos, em especial pacotes lexicais e ELs, em textos acadêmicos produzidos em diferentes áreas do conhecimento.

Pergunta 3: Quais ELs realizam funções retóricas expressas nos resumos acadêmicos das áreas-alvo?

Esta etapa de nosso estudo observou as 150 ELs em três subetapas, conforme os procedimentos metodológicos já descritos (cf. seção 2.4.3): (i) a observação dos elementos lexicais que não variam (*frame*) nas ELs; (ii) a observação das palavras que completam as lacunas (*slot fillers*) das ELs; e (iii) a observação das ELs em contexto.

A observação da amostra de 150 ELs, no que tange às funções retóricas que realizam nos resumos acadêmicos, indicou a existência de duas grandes categorias de unidades multipalavras descontínuas: (i) as ELs transparentes e (ii) as ELs opacas. A categoria por nós denominada ELs transparentes representa o grupo de ELs que têm a sua função retórica mais facilmente identificável a partir da observação: (i) dos elementos fixos que constituem a sua estrutura (por exemplo, *the aim of + + was*); (ii) das palavras que preenchem as lacunas variáveis (por exemplo, em *we + that the + of*, a observação das palavras que preenchem a lacuna 01 a. *conclude*, b. *demonstrate*, c. *found*); e (iii) dos contextos de ocorrência (por exemplo, *Results: In the first phase of the study, with over 56 months of recruitment using traditional sources, 35 women were enrolled in the study, resulting in a mean rate of ± 0.62 recruits/month.*).

IMPLICAÇÕES PEDAGÓGICAS

As recomendações pedagógicas decorrentes deste estudo incluem, por exemplo, incentivar professores de IFA que estejam desenvolvendo cursos sobre o gênero *resumo acadêmico* a utilizarem a mesma metodologia descrita por nós, no sentido de descreverem o padrão retórico de resumos estruturados a partir da identificação das etiquetas de seção. De igual forma, a utilização de seções como unidade retórica, em detrimento dos movimentos e dos passos, parece-nos uma abordagem de mais fácil entendimento por parte de professores (principalmente os iniciantes) e de alunos de IFA. Relacionar as ELs e exemplos associados a cada seção retórica, conforme o modelo de combinação de forma e de função proposto por Gray et al. (2020), é um modelo pedagógico adequado a cursos de escrita acadêmica.

Como proposta para auxiliar os professores a relacionarem as ELs a cada seção retórica identificadas nos *corpora* deste estudo, criou-se um modelo de resumo acadêmico da área de Física (cf. Anexo 14) que pode servir como exemplo para as demais áreas analisadas no presente trabalho e para futuras pesquisas.

Por fim, esperamos que este estudo sirva para o aprimoramento da formação de pesquisadores de alto nível em nosso país e leve a um aumento no impacto dos artigos produzidos no Brasil nas áreas eleitas. Pesquisas futuras buscarão ampliar, para outras áreas do conhecimento e para outros gêneros acadêmicos, a descrição linguística e a construção de recursos pedagógicos propostos neste estudo.

LIMITAÇÕES DO ESTUDO

Acreditamos que uma das principais limitações deste estudo é o fato de não ter sido feita, de forma automática, a identificação da estrutura retórica e o estabelecimento de sua relação com elementos linguísticos. Pesquisas semelhantes ao estudo aqui proposto (YOON; CASAL, 2020; GRAY et al., 2020) utilizam a etiquetagem retórica (*rhetorical move-step coding*) para identificar a estrutura retórica dos gêneros estudados e para estabelecerem a relação dos elementos linguísticos com suas respectivas funções e unidades retóricas (seções, movimentos e passos). A partir da criação de códigos que permitem a etiquetagem retórica (*rhetorical move-step coding*), os processos mencionados demandam bem menos tempo do que as análises manuais utilizadas neste estudo. O uso de códigos para a etiquetagem retórica, no entanto, exige conhecimentos de programação que não temos. A esse propósito, vale pontuar que tal limitação também é comum entre professores de línguas, tanto iniciantes quanto experientes. Estudos futuros que busquem criar etiquetadores retóricos automáticos seriam bem-vindos e facilitariam o processo de análise da associação entre ELs e movimentos retóricos. Sabe-se que a ferramenta *AntMover* (ANTHONY, 2003), um analisador automático da estrutura retórica de textos, é uma tentativa nessa direção, mesmo que ainda demande revisão após a extração automática.

REFERÊNCIAS

- AMOSOVA, N. N. **The English phraseology**. Leningrad: Higher School Publishers, 1963.
- ANTHONY, L. **AntCorGen (Version 1.1.2) [Computer Software]**. Tokyo, Japan: Waseda University, 2019. Disponível em: [https:// www.laurenceanthony.net/software](https://www.laurenceanthony.net/software). Acesso em: 25 out. 2019.
- ANTHONY, L. **AntGram (Version 1.3.0) [Computer Software]**. Tokyo, Japan: Waseda University, 2021. Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 20 jul. 2022.
- ANTHONY, L. AntMover. **Software Retrievable from <http://antpc1.ice.ous.ac.jp>**, 2003.
- ANTHONY, L. Resources for researching vocabulary. In: Webb, S. (Ed.), **The Routledge handbook of vocabulary studies**. Routledge, 2019, p. 561-590.
- BAKHTIN, M. **Estética da Criação Verbal**. 2.ed. São Paulo: Martins Fontes, 1997.
- BAKHTIN, M. **Marxismo e filosofia da linguagem**. São Paulo: Hucitec, 1999.
- BALLY, C. Tratado de estilística francesa. **Paris-Genebra: Livraria Geog & Cie**, 1951.
- BERBER SARDINHA, T. Como usar a Linguística de Corpus no Ensino de Língua Estrangeira. Por uma Linguística de Corpus educacional brasileira. In: VIANA, V e TAGNIN, S. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2011.
- BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Editora Manole. 2004.
- BIBER, D. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. **International journal of corpus linguistics**, v. 14, n. 3, p. 275-311, 2009.
- BIBER, D. Corpus-based and corpus-driven analyses of language variation and use. In: BIBER, D. **The Oxford handbook of linguistic analysis**. 2ed. Oxford: Oxford University Press, 2015.
- BIBER, D. **University Language – A corpus-based study of spoken and written registers**. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.
- BIBER, D.; CONNOR, U.; UPTON, T. **Discourse on the move: Using corpus analysis to describe discourse structure**. Amsterdam: John Benjamins, 2007.
- BIBER, D.; CONRAD, S. Lexical bundles in conversation and academic prose. In: HASSELGARD, H.; OKSEFJELL, S. (Eds.). **Out of corpora: studies in honour of Stig Johansson**. Amsterdam; Atlanta: Rodopi, 1999. p. 181-190.
- BIBER, D.; CONRAD, S. **Register, genre, and style**. Cambridge: Cambridge University Press, 2009.

BIBER, D; CONRAD, S; CORTES, V. If you look at...: Lexical bundles in university teaching and textbooks. **Applied linguistics**, v. 25, n. 3, p. 371-405, 2004.

BIBER, D. What can a corpus tell us about registers and genres? In: O'KEEFFE, A.; MCCARTHY, M. (Ed.). **The Routledge Handbook of corpus Linguistics**. London, New York: Routledge, 2010. p. 241-254.

BIBER, D. Corpus-based and corpus-driven analyses of language variation and use. In: HEINE, Bernd; NARROG, Heiko (Ed.). **The Oxford handbook of linguistic analysis**. Oxford: Oxford University Press, 2012.

BOCORNY, A. E. P.; REBECHI, R.; REPPEN, R.; DELFINO, M. C. N.; LAMEIRA, V. M. Writing scientific articles with the support of key lexical bundles: a corpus-driven study in the area of health sciences. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 37, n. 1, 2021.

BOCORNY, A.; WELP, A. O desenho de tarefas pedagógicas para o ensino de Inglês para Fins Acadêmicos: conquistas e desafios da Linguística de Corpus. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 29, n. 2, 2021. Disponível em: <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/17484>. Acesso em: 07 fev. 2021.

BONELLI, E. T. Theoretical overview of the evolution of corpus linguistics. In: **The Routledge handbook of Corpus Linguistics**, Routledge, 2010.

BREEZE, R. **Tracing the development of an emergent part-genre**: The author summary. *English for Specific Purposes*, v. 42, p. 50-65, 2016.

CASAL, J. E.; KESSLER, M. Form and rhetorical function of phrase-frames in promotional writing: A corpus-and genre-based analysis. **System**, v. 95, 2020.

CHARLES, M. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. **Journal of English for Academic Purposes**, v. 6, p. 289-302, 2007.

CHEN, M.; FLOWERDEW, J. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. **International Journal of Corpus Linguistics**, v. 23, n. 3, p. 335-369, 2018.

CHENG, W.; GREAVES, C.; WARREN, M. From n-gram to skipgram to concgram. **International Journal of Corpus Linguistics**, v. 11, n. 4, p. 411-433, 2006.

CORPAS PASTOR, G. **Manual de fraseología española**. Madrid: Gredos, 1997.

CORTES, V. **Lexical bundles in published and student disciplinary writing**: Examples from history and biology. *English for specific purposes*, v. 23, n. 4, p. 397-423, 2004.

CORTES, V. Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In: CORTES, V.; CSOMAY, E.; BIBER, D (Eds.).

Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber.

Amsterdam: John Benjamins, 2015, p. 197-216.

CORTES, V. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. **Journal of English for academic purposes**, v. 12, n. 1, p. 33-43, 2013.

COTOS, E.; LINK, S.; HUFFMAN, S. Effects of DDL technology on genre learning. **Language Learning & Technology**, v. 21, n. 3, p. 104-130, 2017.

COWIE, A. P. (ed.). **Phraseology: theory, analysis, and applications**. Oxford: Oxford University Press, 1998.

CROSTHWAITE, P. (Ed.). **Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners**. London: Routledge, 2019.

DAYRELL, C. Textual Patterns. In: SCHUSTER, E.; LEVKIWITZ, H.; OLIVEIRA JR, O. N. **Writing Scientific Papers in English Successfully: Your Complete Roadmap**. São Carlos: Editora Compacta. 2014.

EEG-OLOFSSON, M.; ALTENBERG, B. Discontinuous recurrent work combinations in the London-Lund Corpus. In: FRIES, Udo; TOTTIE, Gunnel; SCHNEIDER, Peter (Ed.). **Creating and using English language corpora: Papers from the fourteenth international conference on English language research on computerized corpora**. Atlanta: Rodopi, 1994.

ERMAN, B.; WARREN, B. The idiom principle and the open choice principle. **Text & Talk**, v. 20, n. 1, p. 29-62, 2000.

EVERT, S. **The Statistics of Word Cooccurrences: Word Pairs and Collocations**. Tese (Doutorado em Filosofia), Institut für maschinelle Sprachverarbeitung, University of Stuttgart, 2004.

FIRTH, J. R. **A Synopsis of Linguistic Theory 1930-1955**. Oxford: University Press, 1957.

FLOWERDEW, L. Applying Corpus Linguistics in Teaching Arenas. In: **Corpora and Language Education**. Research and Practice in Applied Linguistics. London: Palgrave Macmillan, 2012.

FREITAS, A. L. P. de. **Proficiência escrita em inglês especializado: estudo de corpus de abstracts em Medicina Nutrição e Farmácia**. Tese (Doutorado em Estudos da Linguagem) Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras, Porto Alegre, 2016.

FREITAS, A. L. P.; BOCORNY, A. E. P. How to write medical abstracts? The rhetorical structure and phrases used in Epidemiology. **Brazilian Journal of Sexually Transmitted Diseases**, 33, 1-6, 2021.

- GELUSO, J. **Frequency, semantic, and functional characteristics of discontinuous formulaic language: A learner corpus study.** Tese (Doutorado em Filosofia), Iowa State University, 2019.
- GRANGER, S.; PAQUOT, M. Disentangling the phraseological web. In: GRANGER, S.; MEUNIER, F. **Phraseology: an interdisciplinary perspective.** Amsterdam / Philadelphia: John Benjamins, 2008.
- GRAY, B.; BIBER, D. Lexical frames in academic prose and conversation. **International Journal of Corpus Linguistics**, v. 18, n. 1, p. 109-136, 2013.
- GRAY, B.; COTOS, E.; SMITH, J. Combining rhetorical move analysis with multidimensional analysis. Research writing across disciplines. In: RÖMER, U.; CORTES, V.; FRIGINAL, E. **Advances in Corpus-based Research on Academic Writing: Effects of discipline, register, and writer expertise.** Amsterdam: John Benjamins, p. 137-168, 2020.
- GRONDONA, C. B. **Eliminação de pacotes lexicais relacionados ao tópico e de pacotes lexicais em contexto de sobreposição: uma proposta metodológica para os estudos da linguística de corpus.** Dissertação (Mestrado em Estudos Linguísticos), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.
- GUEDES, A. de S. **Verbos do inglês acadêmico escrito e suas colocações: um estudo baseado em um corpus de aprendizes brasileiros de inglês.** 2017. 200f. Tese (Doutorado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.
- HERBST, T.; FAULHABER, S.; UHRIG, P. (Orgs.). **The phraseological view of language: A tribute to John Sinclair.** Berlin/New York, NY: Walter de Gruyter, 2011.
- HUNSTON, S. **Corpora in applied linguistics.** Cambridge, UK: Cambridge University Press, 2002. 10.1017/CBO9781139524773
- HYLAND, K. As can be seen: Lexical bundles and disciplinary variation. **English for specific purposes**, v. 27, n. 1, p. 4-21, 2008.
- HYLAND, K. Bundles in academic discourse. **Annual review of applied linguistics**, v. 32, p. 150-169, 2012.
- HYLAND, K.; WONG, L. L. C. (Orgs.). **Specialised English: New directions in ESP and EAP research and practice.** London: Routledge, 2019.
- JENKINS, J. **English as a lingua franca in the international university: The politics of academic English language policy.** Londres: Routledge, 2013.
- JOHNS, T. It is presented initially: linear dislocation & inter-language strategies in Brazilian academic abstracts in English and Portuguese. **Ilha do Desterro**, Florianópolis, n. 27, p. 9-32, 1992.

KANOKSILAPATHAM, B. Rhetorical moves in biochemistry research articles. In: BIBER, D.; CONNOR, U.; UPTON, T. A. (Orgs.). **Discourse on the move: using corpus analysis to describe discourse structure**. Amsterdam: John Benjamins, 2007, p. 73-119.

KILGARRIFF, Adam et al. Itri-04-08 the sketch engine. **Information Technology**, v. 105, n. 116, p. 105-116, 2004.

LEE, D.; SWALES, J. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. **English for Specific Purposes**, v. 25, n. 1, p. 56-75, 2006. <https://doi.org/10.1016/j.esp.2005.02.010>

LU, X.; YOON, J.; KISSELEV, O. Matching phrase-frames to rhetorical moves in social science research article introductions. **English for Specific Purposes**, v. 61, p. 63-83, 2021.

MAURANEN, A. **Exploring ELF in academia**. Cambridge: Cambridge University Press, 2011.

MAURANEN, A.; HYNNINEN, N.; RANTA, E. English as an Academic Lingua Franca: The ELFA Project. **English for Specific Purposes**, v. 29, p. 183-190, 2010.

MAURANEN, A.; HYNNINEN, N.; RANTA, E. English as the academic lingua franca. In: **The Routledge Handbook of English for Academic Purposes**. London: Routledge, 2016, p. 68-79.

MCENERY, T.; WILSON, A. **Corpus Linguistics**. Edinburgh: Edinburgh University Press, 1996.

MCENERY, T.; HARDIE, A. **Corpus Linguistics: method, theory and practice**. Edinburgh: Cambridge University Press, 2012.

MCENERY, T.; WILSON, A. Corpus linguistics. **The Oxford handbook of computational linguistic**. United States: Oxford University Press, p. 448-463, 2003.

MIZUMOTO, A; HAMATANI, S; IMAO, Y. Applying the bundle–move connection approach to the development of an online writing support tool for research articles. **Language Learning**, v. 67, n. 4, p. 885-921, 2017.

MORENO, A. I.; SWALES, J. M. Strengthening move analysis methodology towards bridging the function-form gap. **English for Specific Purposes**, v. 50, p. 40-63, 2018.

MORLEY, J. (2014): **Academic Phrasebank: A compendium of commonly used phrasal elements in academic English in PDF format**. Manchester: University of Manchester, 2014.

NESSSELHAUF, N. What are collocations? In: ALLERTON, D. J.; NESSELHAUF, N.; SKANDERA, P. (Orgs.). **Phraseological Units: Basic Concepts and their Application**. Basel: Schwabe Verlag, p. 1-21, 2004.

NESSSELHAUF, N. **Collocations in a learner corpus**. Amsterdam/Philadelphia: John Benjamins Publishing, 2005.

- PAWLEY, A.; SYDER, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: RICHARDS, J. C.; SCHMIDT, R. W. (Orgs.). **Language and communication**. Longman: New York, 1983, p. 191-226.
- RENOUF, A.; SINCLAIR, J. Collocational frameworks in English. In: AIJMER, K.; ALTENBERG, B. (Orgs.). **English corpus linguistics**. Londres: Longman, p. 128-143, 1991.
- RUTHERFORD, William E. The meaning of grammatical consciousness raising. **World Englishes**, v. 6, n. 3, p. 209-216, 1987.
- SANCHEZ, A. Definición e historia de los corpus. In: A. SANCHEZ et al. (Org.). **CUMBRE Corpus Lingüístico de Español Contemporáneo**. Madrid: SGEL, 1995.
- SARDINHA, T. B. **Lingüística de Corpus**. São Paulo: Editora Manole. 2004.
- SCOTT, M. Aprendizagem direcionada por dados: uma homenagem a Tim Johns (1936-2009). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010, p. 7-11.
- SINCLAIR, J. M. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.
- STUBBS, M. An example of frequent English phraseology: Distributions, structures and functions. In: FACCHINETTI, R. (Org.). **Corpus Linguistics 25 Years On**. Amsterdam: Rodopi, 2007, p. 89-105.
- STUBBS, M. Language Corpora. In: DAVIES, A. & ELDER, C. **The handbook of applied linguistics**. Oxford: Blackwell Publishing, 2004. P. 106-132.
- SWALES, J. **Aspects of article introductions** (Aston ESP Research Report No. 1). Birmingham: Language Studies Unit: University of Aston, 1981.
- SWALES, J. **Genre analysis: English in academic and research settings**. Cambridge: Cambridge University Press, 1990.
- SWALES, J. M.; FEAK, C. B. **Academic writing for graduate students: A course for non-native speakers of English**. Ann Arbor: University of Michigan Press, 1994.
- SWALES, J. M. **Research genres: Exploration and applications**. Cambridge: Cambridge University Press, 2004.
- SWALES, J. M.; FEAK, C. **Abstracts and the writing of abstracts**. Michigan, University of Michigan Press, 2009.
- TANKÓ, G. Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. **Journal of English for Academic Purposes**, v. 27, p. 42-55, 2017.
- THOMAS, J. Deriving extended collocations from full text for student analysis and synthesis. In: LÉNKO-SZYMÁNSKA, A.; BOULTON, A. (Orgs.). **Multiple Affordances of Language Corpora for Data-driven Learning**. Amsterdam: Philadelphia, PA: John Benjamins, 2015, p. 85-108.

VAN LEEUWEN, T. Genre and field in critical discourse analysis. **Discourse & society**, Sage Publications, v. 4, n. 2, p. 193-223, 1993.

VINOGRADOV, V. V. Ob osnovnyx tipach frazeologičeskix edijic v sovremennom ruskom jazyke. **VV Vinogradov, Izbrannyje trudy: leksikologija i leksikografija**. Moskva: Nauka, 1947.

WRAY, A. **Formulaic language and the lexicon**. Cambridge: Cambridge University Press, 2002.

YOON, J.; CASAL, J. E. P-frames and rhetorical moves in applied linguistics conference abstracts. In: RÖMER, U.; CORTES, V.; FRIGINAL, E. (Orgs.). **Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise**. Amsterdam, the Netherlands: John Benjamins Publishing Company, 2020, p. 281-305.

APÊNDICES

APÊNDICE 1

Identificação das etiquetas de seção dos resumos estruturados da área de Ciências da Computação e da Informação que ocorrem mais de 10 vezes.

Área: Ciências da Computação e da Informação Nome do corpus: SC0319AbsCS1mi Número de resumos acadêmicos: 4.598		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO (Frequência absoluta da palavra como título de seção)	Percentual de resumos estruturados desse corpus que tem a seção
BACKGROUND, CONTEXT	Background: (409x)	9,87%
	Introduction: (45x)	
AIM OBJECTIVE PURPOSE	Aim: (10x)	3,32%
	Motivation: (15x)	
	Objective(s): (104x)	
	Purpose: (24x)	
MATERIALS, METHODS	Materials and methods: (30x)	8,06%
	Method: (29x)	
	Methodology: (34x)	
	Methods: (231x)	
	Methods and findings: (36x)	
	Methods and results: (11x)	
FINDINGS, RESULTS	Results: (361x)	7,85%
CONCLUSION, DISCUSSION	Conclusion/Significance: (10x)	11,78%
	Conclusion: (157x)	
	Conclusions/Significance: (88x)	
	Conclusions: (266x)	
	Discussion: (21x)	
AUTHOR SUMMARY		3,15%

Fonte: Elaborada pela autora.

APÊNDICE 2

Identificação das etiquetas de seção dos resumos estruturados da área da Física que ocorrem mais de 10 vezes.

Área: Física. Nome do <i>corpus</i>: SC0319AbsPHY1mi Número de resumos acadêmicos: 4.690.		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO (Frequência absoluta da palavra como título de seção)	Percentual de resumos estruturados deste <i>corpus</i> que têm a seção
BACKGROUND, CONTEXT	Background: (178x)	4,34%
	Introduction: (26x)	
AIM OBJECTIVE PURPOSE	Aim(s): (15x)	3,49%
	Objective(s): (80x)	
	Purpose: (69x)	
MATERIALS, METHODS	Materials and Methods: (32x)	5,11%
	Method(s): (208x)	
FINDINGS, RESULTS	Result(s): (266x)	5,67%
CONCLUSION, DISCUSSION	Conclusions/Significance: (29x)	7,01%
	Conclusion(s): (290x)	
	Discussion: (10x)	
AUTHOR SUMMARY		0,34%

Fonte: Elaborada pela autora.

APÊNDICE 3

Identificação das etiquetas de seção dos resumos estruturados da área da Medicina e Ciências da Saúde que ocorrem mais de 10 vezes.

<p>Área: Medicina e Ciências da Saúde Nome do <i>corpus</i>: 4.039 Número de resumos acadêmicos: SC0319AbsME1mi</p>		
CATEGORIAS DE SEÇÃO	ETIQUETAS DE SEÇÃO (Frequência absoluta da palavra como título de seção)	Percentual de resumos estruturados deste <i>corpus</i> que tem a seção
BACKGROUND, CONTEXT	Background: (1.796x)	50,92%
	Introduction: (261x)	
AIM OBJECTIVE PURPOSE	Aim: (146x)	34,36%
	Background and aim: (70x)	
	Background and objective: (26x)	
	Background and purpose: (45x)	
	Objective: (829x)	
	Purpose: (272x)	
MATERIALS, METHODS	Design: (87x)	72,96%
	Material and Method: (187x)	
	Measurements: (10x)	
	Method: (2,463x)	
	Methodology: (40x)	
	Methods and materials: (22x)	
	Data sources: (15x)	
	Participants: (37x)	
	Setting: (52x)	
	Study design: (34x)	
FINDINGS, RESULTS	Findings: (82x)	72,17%
	Result: (2,833x)	
CONCLUSION, DISCUSSION	Conclusion: (3.071x)	78,01%
	Discussion: (55x)	
	Interpretation: (25x)	
AUTHOR SUMMARY		0,22%

Fonte: Elaborada pela autora.

APÊNDICE 4

Estruturas lexicais da área de Ciências da Computação e da Informação e exemplos.

<p>Área: Ciências da computação e da informação Corpus: SC0319AbsCS1mi</p>					
<p>LEGENDA:</p> <p>Quando as palavras estiverem em negrinho na cor preta elas pertencem a uma só estrutura lexical (quando houver só uma como exemplo), ou pertence a todas (quando houver mais de uma como exemplo).</p>					
Seções identificadas nos resumos estruturados	Padrão retórico sugerido por Swales e Feak (2009)	Estruturas lexicais identificadas	LACUNA 1	LACUNA 2	Exemplo
<p><i>BACKGROUND CONTEXT</i></p> <p><i>AIM OBJECTIVE PURPOSE</i></p>					
		2. to + the + of the	<p>a. evaluate</p> <p>b. assess</p> <p>c. demonstrate</p>	<p>a. performance/ quality/ capability</p> <p>b. impact/ accuracy/ quality</p> <p>c. effectiveness/ feasibility/ advantages</p>	<p>“In plants, efforts are concentrated on various regions of the plastid genome, but no agreement exists as to what kinds of regions are ideal, though most researchers agree that more than one region is necessary. One reason for this discrepancy is differences in the tests that are used to evaluate the performance of the proposed regions. Most tests have been made in a floristic setting, where the genetic distance and therefore the level of variation of the regions between taxa is large, or in a limited set of congeneric species.”</p>

		6. the + of the + of	a. analysis b. prediction c. effect	a. properties/ dynamics/ structure b. direction/ existence/ response c. number/ parameters/ level	“By means of a lumped Markov chain model of a random walker, we propose two novel ways of inferring the lumped markov transition matrix. Furthermore, some useful results are proposed based on the analysis of the properties of the lumped Markov process.”
		11. we + the + of the	a. demonstrate b. investigate c. compared	a. usefulness/value/ applicability b. behavior/ impact/ characteristics c. performance/ characteristics/ results	“In this paper we introduce a new medical technology paradigm centered on wireless technology and cloud computing that was designed to overcome the problems of increasing health technology costs. We demonstrate the value of the concept with an example; the design of a wireless, distributed network and central (cloud) computing enabled three-dimensional (3-D) ultrasound system.”
		16. the + of the + and	a. complexity b. performance c. study	a. series/model b. spectral/ methodology/ AFTSMC c. structure/ accuracy	“Recent developments in network theory have allowed for the study of the structure and function of the human brain in terms of a network of interconnected components.”
		27. order to + the + of	a. improve b. assess c. explore	a. performance/ accuracy/ efficiency b. fit/ effects/ performance c. selection/ potentialities/ role	“With the rapid development of 3D somatosensory technology, human behavior recognition has become an important research field. Human behavior feature analysis has evolved from traditional 2D features to 3D features. In order to improve the performance of human activity recognition, a human behavior recognition method is proposed, which is based on a hybrid texture-edge local pattern coding feature extraction and integration of RGB and depth videos information.”

		29. can be + to + the	a. used b. applied c. employed	a. estimate/ simulate/ predict b. understand/ discriminate/ predict c. explore/ control/ improve	“Recent studies have demonstrated that a variety of neuroimaging and neurophysiological techniques such as functional magnetic resonance imaging (MRI), diffusion MRI and electroencephalography/magnetoencephalography can be employed to explore the topological organization of human brain networks. ”
		35. in terms of + + and	a. their b. both c. the	a. efficiency/ inputs/ identification b. sensitivity/ backend/ accuracy c. sensitivity/ computational/ relevance	“The identification of phenotype descriptions in the scientific literature, case reports and patient records is a rewarding task for bio-medical text mining. Any progress will support knowledge discovery and linkage to other resources. However because of their wide variation a number of challenges still remain in terms of their identification and semantic normalisation before they can be fully exploited for research purposes.”
		36. the + and the + of	a. structure b. reliability c. information	a. degree/ communities/ status b. timeliness/ accuracy c. amount	“Here we develop an approach and propose a quantity (measure) which is simple enough to be widely applicable, reveals a number of universal features of the organization of real-world networks and, as we demonstrate, is capable of capturing the essential features of the structure and the degree of hierarchy in a complex network.”
		41. of the + + and the	a. proposed b. data c. discrimination	a. circuit/ controller/ method b. generated/ used c. approach/ method	“Spatial analysis extracts meaning and insights from spatially referenced data, where the results are highly dependent on the quality of the data used and the manipulations on the data when preparing it for analysis.”

		47. in the + of + and	a. context b. evolution c. field	a. genetics/ epidemics/ modeling b. PINs/ species/ culture c. microdrive/ SBM/ bibliometrics	“ In the context of genetics and breeding research on multiple phenotypic traits, reconstructing the directional or causal structure between phenotypic traits is a prerequisite for quantifying the effects of genetic interventions on the traits.”
MR 2 - Apresentar a pesquisa/objetivo	1. in this + we + a	a. paper	a. introduce/ propose/ present	“ In this paper, we present a weighted model for undirected and weighted networks based on the mutual information of local network structures, where link weights are applied to further enhance the distinguishable extent of candidate links.”	
	3. in this paper + + a	a. we	a. present/ introduce/ describe		
	4. in + paper we + a	a. this	a. present/ introduce/ developed		
	50. in this + + present a	a. paper, b. article, c. study,	a. we b. we c. we		
		22. based on the + + of	a. topological b. combined c. mutual	a. distribution/ properties/ analysis b. dataset/ analysis c. information	
		34. as well as + + of	a. the b. a c. in	a. presence/ frequency/ distribution b. catalog/ combination/ selection c. areas/ analyses/ terms	“ In this paper, we present a versatile approach to addressing this issue, based on multiple interpretations of sentiment labels that encode information regarding the polarity, subjectivity, and ambiguity of a text, as well as the presence of sarcasm or a mixture of sentiments”
MR 2 - Apresentar a pesquisa/objetivo	43. in + + we present a	a. this b. this	a. paper/ article/work b. contribution		

		15. in + paper + propose a	a. this	a. we	<p>“In this paper we propose a novel distributed solution called Random Walk with Network Division (RWND) for the detection of node replication attack in static WSNs which is based on claimer-reporter-witness framework and combines a simple random walk with network division.”</p>
		8. in + + we propose a	a. this	a. paper/ work	
		9. in this + + propose a	a. paper, b. work, c. study,	a. we b. we c. we	
		24. this + we + a novel	a. paper, b. article, c. work,	a. propose/ present/ introduce b. present/ describe/ propose c. introduce/ present/ examine	
		23. in this + we + an	a. paper, b. study, c. paper,	a. present/ propose/ provide b. present/ applied/ developed c. present/ propose/ introduce	
		25. of the + + in the	a. data b. current c. information	a. delivery/ distribution/ generated b. knowledge/ work c. encoded/ processing	<p>“In this study, we developed an ontology representing clinical features, treatments, genetic factors, environmental factors, and other aspects of the current knowledge in the domain of PE.”</p>
	MR 2 - Apresentar a pesquisa/objetivo	5. in this + we + the	a. paper, b. study	a. propose/ demonstrate/ explore b. examined/ determined/ analyse	
		10. in this paper + + the	a. paper,	a. propose/ demonstrate/ explore	
		18. in + paper we + the	a. this	a. explore, evaluate, motivate	<p>“In this paper we explore the accuracy of using self-reported estimates when compared with actual smartphone use.”</p>

		23. in this + we + an	a. paper, b. study, c. paper,	a. present/ propose/ provide b. present/ applied/ developed c. present/ propose/ introduce	“ In this study, we developed an ontology representing clinical features, treatments, genetic factors, environmental factors, and other aspects of the current knowledge in the domain of PE. We call this pre-eclampsia ontology "PEO".”
		33. this + we + a new	a. paper b. study, c. problem	a. propose/ introduce/ present b. developed/ apply/ describe c. propose/ design	“ In this study, we explored the phylogenetic positions of the genus Longgenacris and the species L. rufiantennus and the relationships among F. viridifemorata group based on the 658-base fragment of the mitochondrial gene cytochrome c oxidase subunit I (COI) barcode and the complete sequences of the internal transcribed spacer regions (ITS1 and ITS2) of the nuclear ribosomal DNA.”
		31. and the + + of the	a. average b. complete c. completion	a. flow/value b. sequences c. time	
	MR 2 - Apresentar a pesquisa/objetivo	39. in this study + + the	a. we b. based c. I	a.investigate/ explore/ modeled b. on c. extend	“ In this study, based on the analytical models of EDCA, a game theoretic approach called GTXOP is proposed to determine TXOP dynamically (i.e. according to the dynamisms of WLAN networks and the number of nodes in the network).”
		29. can be + to + the	a. used b. applied c. employed	a. estimate/ simulate/ predict b. understand/ discriminate/ predict c. explore/ control/ improve	“This research presents a new approach that can be used to estimate the most appropriate scale at which to aggregate point data to areas.”
	MR 2 - Apresentar a pesquisa/objetivo	12. the + of this + is	a. aim b. purpose c. objective	a. paper/study/ work b. paper/ study/research c. review/ work/ method	“ The aim of this study is to measure the precision and accuracy of the timing mechanisms of some of the most popular

		30. the + of this + was	a. aim b. objective c. purpose	a. project/trial/study b. study/ research/ c. study/project/ investigation	software packages used in a typical laboratory scenario in order to assess whether presentation times configured by researchers do not differ from measured times more than what is expected due to the hardware limitations.”
		37. aim of this + + to	a. study b. work c. research	a. study b. work c. research	
		7. the + and + of the	a. effectiveness b. feasibility c. precision	a. efficiency/ adaptability/ convenience b. effectiveness/ availability/ reliability c. accuracy/ recall/ growth	
	MR 2 - Apresentar a pesquisa/objetivo	28. this study + to + the	a. was b. aims c. aimed	a. infer/ assess/ evaluate b. investigate/ assess/ understand c. examine/ determine/ investigate	“Objectives: This study aimed to examine the potential of combining routine tuberculosis (TB) surveillance and demographic and socioeconomic variables into the Geographic Information System (GIS) to describe the geographical distribution of TB notified incidence in relation to distances to health services as well as local demographic and socioeconomic factors, including population density, urban/rural status, and household poverty rates in Nam Dinh, Vietnam”
		14. of the + + of the	a. degree b. network c. main	a. centrality/ distribution/ structure b. structure/ topology c. functions/ surprises	“Using graph theory-based network analysis, we investigated the regional changes of the degree centrality of the rsFC network.”
		21. the + of + in the	a. number b. role c. degree	a. nodes/ edges/ events b. participants/ place/ yeasts c. feedback/ frustration/	“In this study, based on the analytical models of EDCA, a game theoretic approach called GTXOP is proposed to determine TXOP

				structure	dynamically (i.e. according to the dynamisms of WLAN networks and the number of nodes in the network).”
		38. the + of the + in	a. performance b. validity c. position	a. device/ CV/ algorithm c. survey/ idea/ simulator c. boundary/ pixel	“We evaluate the performance of the device in the digital assessment of estrogen receptor-alpha (ER) expression in breast cancer samples.”
	MR 2 - Apresentar a pesquisa/objetivo	39. in this study ++ the	a. we b. based c. I	a. investigate/ explore/ modeled b. on c. extend	“ In this study, based on the analytical models of EDCA, a game theoretic approach called GTXOP is proposed to determine TXOP dynamically (i.e. according to the dynamisms of WLAN networks and the number of nodes in the network).”
		45. to + the + of a	a. measure b. evaluate c. determine	a. randomness/ sensitivity/ trustworthiness b. effect/ efficacy/ merits c. accuracy/ association	“This work presents a nonparametric statistical test, S-maup, to measure the sensitivity of a spatially intensive variable to the effects of the Modifiable Areal Unit Problem (MAUP).”
MATERIALS, METHODS					
	MR 3 - Metodologia/ materiais/ sujeitos/ procedimentos	1. in this + we + a	1) a. paper	1) a. introduce/ propose/ present	“Methods: In this paper we present a new, alignment-free method for detecting conserved regions in protein sequences called NADDA (No-Alignment Domain Detection Algorithm)”
		3. in this paper ++ a	3) a. we	3) a. present/ introduce/ describe	
		4. in + paper we + a	4) a. this	4) a. present/ introduce/ developed	
		50. in this ++ present a	a. paper, b. article, c. study,	a. we b. we c. we	

		2. to + the + of the	a. evaluate b. assess c. demonstrate	a. performance/ quality/ capability b. impact/ accuracy/ quality c. effectiveness/ feasibility/ advantages	“The interview data was analyzed using a thematic analysis approach, and the pre and post workshop checklist scores were compared to assess the impact of the workshop on the computational reproducibility of the researchers' workflows.”
	MR 3 - Metodologia/ materiais/ sujeitos/ procedimentos	11. we + the + of the	a. demonstrate b. investigate c. compared	a. usefulness/value/ applicability b. behavior/ impact/ characteristicsc. performance/ characteristics/ results	“ We compared the performance of the GPU implementation with those of multi-core CPU implementations with 16 CPU threads, single-thread C/C++ implementation and single-thread R implementation.”
		13. used to + the + of	a. evaluate b. analyze c. assess	a. significance/ performance/ fit b. ability/ effect/ effect c. diversity/ reliability/ accuracy	“Methodology: Morphological data based on the examination of living specimens, scanning electron and light microscopic images, and morphometric analyses were used to assess the diversity of Bugula along the coastal areas of southern, northeastern, and southeastern Brazil.”
		16. the + of the + and	a. complexity b. performance c. study	a. series/model b. spectral/ methodology/ AFTSMC c. structure/ accuracy	“Specifically, we compared the performance of the spectral and information-theoretic Granger pipelines as well as standard Granger causality in their relations to the observed phase differences for frequencies at which the signals become synchronized to each other.”
		17. and + the + of the	a. compare b. evaluate c. improve	a. performance/ accuracy b. performance/ accuracy c. accuracy/ performance/ resistance	“To test our method, we use four species and build cross-species training sets with two of them and evaluate the performance of the remaining ones.”
		22. based on the + + of	a. topological b. combined c. mutual	a. distribution/ properties/ analysis b. dataset/ analysis c. information	“We formed cell graphs based on the topological distribution of the tissue cell nuclei and extracted the corresponding graph features.”

		31. and the + + of the	a. average b. complete c. completion	a. flow/value b. sequences c. time	“The measure we introduce is based on a generalization of the m-reach centrality, which we first extend to directed/partially directed graphs. Then, we define the global reaching centrality (GRC), which is the difference between the maximum and the average value of the generalized reach centralities over the network.”
		34. as well as + + of	a. the b. a c. in	a. presence/ frequency/ distribution b. catalog/ combination/ selection c. areas/ analyses/ terms	“We tackle this Missing Data Problem using Link-Prediction techniques from social network research, network generation techniques from the area of Social Simulation, as well as a combination of both ”
		35. in terms of + + and	a. their b. both c. the	a. efficiency/ inputs/ identification b. sensitivity/ backend/ accuracy c. sensitivity/ computational/ relevance	“The performance of the algorithm was assessed in terms of the sensitivity and positive predictive value (PPV) for detecting compressions and in terms of its accuracy through the analysis of measurement error.”
		42. to the + + of the	a. current b. existing c. static	a. state b. state c. character/ nature	“Due to the static character of the challenge data, we tackled the problem via a sparse Gaussian Markov Random Field, which relates network topology with the covariance inverse generated by the gene measurements.”
		46. on the + + of the	a. specific b. topological c. overall	a. complexity/ constraints/ shape b. distribution/ organization/ properties c. flow/ performance	“We formed cell graphs based on the topological distribution of the tissue cell nuclei and extracted the corresponding graph features.”
FINDINGS, RESULTS					

		16. the + of the + and	a. complexity b. performance c. study	a. series/model b. spectral/ methodology/ AFTSMC c. structure/ accuracy	“In addition, the denoising and decomposition strategies decrease the complexity of the series and reduce the difficulties of the forecasting.”
		19. we + that the + of	a. show b. find c. demonstrate	a. topology/ existence/ use b. period/ hubs/ emotions c. performance/ number/ protocol	“Results: We created a workflow known as Illinois Mayo Taxon Organization from RNA Dataset Operations (IM-TORNADO) for processing non-overlapping reads while retaining maximal information content. Using synthetic mock datasets, we show that the use of both reads produced answers with greater correlation to those from full length 16S rDNA when looking at taxonomy, phylogeny, and beta-diversity.”
	MR 4 - Resultados/ descubertas	20. the + results + that the	a. experimental b. simulation c. experimental	a. show/ showed/ demonstrate b. show/ demonstrate/ indicate c. demonstrated/ indicate/ prove	“Results: The experimental results show that the earthquake disaster casualty prediction model based on ACPSO-ELM algorithm has better stability and prediction accuracy.”
	MR 4 - Resultados/ descubertas	48. the + + show that the	a. experimental b. simulation c. obtained	a. results b. results c. results	
		25. of the + + in the	a. data b. current c. information	a. delivery/ distribution/ generated b. knowledge/ work c. encoded/ processing	“In networks that are close to critical the intrinsic noise required for the tuning is smaller and thus also has the smallest effect in terms of the information processing in the system.”
		26. in the + + of the	a. first b. second c. central	a. half/ phase/ round b. part/step/years c. part/position	“Results: In the first phase of the study , with over 56 months of recruitment using traditional sources, 35 women were enrolled in the study, resulting in a mean rate of ± 0.62 recruits/month.”
		36. the + and the + of	a. structure b. reliability	a. degree/ communities/ status b. timeliness/ accuracy	“It was found that there was a relationship between the awareness of the information and

			c. information	c. amount	the amount of publicity generated around the terminology.”
		40. the + of the + is	a. performance b. efficiency c. accuracy	a. ensemble/ system/ engine b. algorithm/ scheme/ model c. system/ model	“The results indicate that the performance of the engine is limited by the number of events coming in rather than the queries being processed.”
	MR 4 - Resultados/ descubiertas	49. results + that the + of	a. show b. suggest c. indicate	a. combination/ resolutions/ family b. initiation/ combination/ number c. use/ notion/ performance	“Our results show that the combination of the dynamic nature of S-dFasArt classifier with a supervisory module are very robust and suggest that this methodology could be very useful to take into account emotional states in robot-assisted environments and help to enhance and better understand human-robot interactions.”
CONCLUSION, DISCUSSION					
		7. the + and + of the	a. effectiveness b. feasibility c. precision	a. efficiency/ adaptability/ convenience b. effectiveness/ availability/ reliability c. accuracy/ recall/ growth	“Conclusions: The effectiveness and efficiency of the label ranking algorithms proposed in this study make them be of practical value for infection source identification.”
		29. can be + to + the	a. used b. applied c. employed	a. estimate/ simulate/ predict b. understand/ discriminate/ predict c. explore/ control/ improve	“Our approach can be applied to understand the impact of large, correlated disruptions in financial systems, ecosystems and other systems with a complex interaction structure between heterogeneous components.”
		32. and + + as well as	a. gene b. network c. data	a. entities/ expression b. coverage/ measures c. management	“Integration of project and data management as well as workflow resources in one place present clear advantages over existing solutions.”

		42. to the ++ of the	a. current b. existing c. static	a. state b. state c. character/ nature	“Excellent features of the experimental results demonstrate that the proposed steganographic approach is greatly proficient for preserving the visual information of an image. An improvement in the imperceptibility as much as 28% is achieved compared to the existing state of the art methods. ”
		44. the ++ in this paper	a. method b. approach c. findings	a. proposed b. presented c. presented	“Conclusion: The method proposed in this paper provides evidence of being a promising adjunct to existing automated ADE detection and analysis approaches.”
		46. on the ++ of the	a. specific b. topological c. overall	a. complexity/ constraints/ shape b. distribution/ organization/ properties c. flow/ performance	“Our results suggest that the buffered gossip algorithm can solve the rendezvous problem faster than the uniform gossip algorithm; however, the relative performance between these two solutions depends on the specific constraints of the problem and the parameters of the buffered gossip algorithm.”

Fonte: Elaborada pela autora.

APÊNDICE 5

Estruturas lexicais da área de Física e exemplos.

Disciplina: Física Corpus: SC0319AbsPHY1mi					
LEGENDA:					
Quando as palavras estiverem em negrito na cor preta elas pertencem a uma só estrutura lexical (quando houver só uma como exemplo), ou pertence a todas (quando houver mais de uma como exemplo).					
Seções identificadas nos resumos estruturados	Padrão retórico sugerido por Swales e Feak (2009)	Estruturas lexicais identificadas	LACUNA 1	LACUNA 2	Exemplo
BACKGROUND CONTEXT AIM OBJECTIVE PURPOSE					
		2. of the + + of the	a. biomechanical b. different c. mechanical	a. properties/changes b. areas/ bands/ elements c. properties	“However, the relationship between the activation of the different areas of the atria and the characteristics of the BSPM and P-wave signals are still far from being completely understood.”
	MR 2 - Apresentar a pesquisa/objetivo	4. In this + we + the	d. study, e. work, f. paper,	d. report/ propose/ determined e. probed/ describe/ report f. consider/ explore/ investigate	“ In this paper we consider the process of the second harmonic generation in a gradient waveguide, taking into account diffraction and relatively weak temporal dispersion.”
		10. in the + + of the	a. first b upper c. southern	a. step/trial b. body/edges/ part/plate c. part	“Citrate synthase from <i>M. sedula</i> (MsCS) is an enzyme involved in the first step of the incomplete TCA/glyoxylate cycle by converting oxaloacetate and acetyl-CoA into

					citrate and coenzyme A.”
		14. used to + the + of	a. study b. evaluate c. alculate	a. role/ effects/changes b. performance/ effect/ properties c. number/ deformation/ movement	“The CHARMM force field in combination with the FACTS implicit solvation model is used to study the role of long-distance interactions between the TCR and pMHC.”
		31. the effect of + + on	a. these b. elevated c. mixing	a. approximations b. pressure c. time	“In the past, the effect of these approximations on macromolecular structure and stability has been evaluated mostly through quantitative studies of small-molecule systems or qualitative observations of short-timescale simulations of biological macromolecules.”
		38. between the + + and the	a. CAT b. DNA c. applied	a. dimensions b. lesion c. force	“In an admittance-controlled haptic device, input forces are used to calculate the movement of the device. Although developers try to minimize delays, there will always be delays between the applied force and the corresponding movement in such systems, which might affect what the user of the device perceives.”
		8. the + of the + of	a. position b. analysis c. understanding	a. center b. effect/ improvement/ distributions c. mechanism/transport/functions	“A long-standing hypothesis links enzymatic activity to the protein oligomeric state, hence the study of the solution behavior in general and the oligomer state in particular of glutaminase C is important for the understanding of the mechanism of protein activation and inhibition.”
		41. the + of the + in	a. accuracy b. approach c. structure	a. simulation/ measurement b. loop c. LD/protein	“There is also uncertainty regarding the accuracy of the various crystal structures available, let alone the structure of the protein in a physiologically relevant environment.”
	MR 2 - Apresentar a	3. the + and + of the	a. effectiveness	a. accuracy/ capability/ feasibility	“In this study, we evaluate the accuracy and

	pesquisa/objetivo		b. accuracy c. feasibility	b. consistency/ efficiency/ proficiency c. effectiveness/ capability/ validity	consistency of the temperature measurement and control settings for different brands of fourth-generation ENDS.”
		4. In this + we + the	a. study, b. work, c. paper,	a. report/ propose/ determined b. probed/ describe/ report c. consider/ explore/ investigate	“ In this study, we propose the detailed mechanism of PRMT1-catalyzed asymmetric dimethylation, which increases insight on the small-molecule effectors design, and enables further investigations into the physiological function of this family.”
		20. in this study + + the	a. we b. to c. and	a. describe/ demonstrate/ examined b. determine/ simulate c. evaluate	“ In this study, we determine the variation in UV transmission in a relatively homogenous sample of young adults.”
		16. in this + we + a	a. paper b. study c. work	a. present/ propose/ introduce b. propose/ present/ developed/ c. propose/ report/ develop	“ In this paper, we propose a new TSV test architecture to transfer the combined test output of the test result and the specific defect type to the pad during the single test-clock-period.”
		17. of the + + and the	a. test b. proposed c. pulse	a. distance/ result b. method/ approach/ model c. pressure	
	MR 2 - Apresentar a pesquisa/objetivo	5. the + of this + was	a. aim b. purpose	a.study/work/ investigation b. study/ work/ experiment	“ The aim of this study was to investigate the variation of the miRNA expression levels in normal renal cortical tissue after 177Lu-octreotate administration, a radiopharmaceutical used for treatment of neuroendocrine cancers.” “ The aim of this paper is to develop, calibrate, and validate a novel passive hybrid FE-MB open-access simulation model of a ligamentous LSS using ArtiSynth.”
		9. the + of + study was	a. aim b. purpose c. objective	a. this b. this c. this	
		34. this + was to + the	a. study b. work c. investigation	a. investigate/ compare/ describe b. evaluate/ test/ describe c. establish/ reproduce	

		19. aim of this + + to	a. paper b. study c. work	a. is b. is/was c. is/was
		23. aim of + study + to	a. our b. the	a. was/ is b. was/ is
		26. aim of + + was to	a. our b. the c. this	a. study/work b. study c. study/work
		27. this study + the + of	a. demonstrate b. investigated c. determines	a. importance/ addition/ use b. adjustment/ changes/ feasibility c. impact
		32. aim + this study + to	a. of	a. is/was
		33. the + of this + is	a. aim b. goal c. purpose	a. paper/study/ work b. work/ study/ paper c. study/ analysis/ work
		37. the aim of + + was	a. study b. work c. investigation	a. was b. was c. was
		40. aim + this + was to	a. of	a. study/ work/ project

		43. aim + + study was to	a. of	a. this	
		46. this study was + + the	a. to	a. determine/ investigate/ compare	
		48. the aim + this + was	a. of	a. study/ work/ investigation	
		50. the + of the + is	a. aim	a. paper/study	
		30. purpose of this + + to	a. study b. work c. experiment	a. was/is b. was/is c. was	“Objective: The purpose of this work was to use sweating thermal manikin and mathematical modeling techniques to quantify the tradeoff between increased BA protection, the accompanying mass, and thermal effects on human performance.”
		13. this study + to + the	a. was b. is c. aimed	a. determine/understand/ investigate b. assess/ investigate/ compare c. characterize/ compare/ assess	“The aim of this study was to determine the role played by vigilance on the anaerobic performance recorded during a Wingate test performed at the bathyphase (nadir) of the circadian rhythmicity.”
		28. in + study we + the	a. our b. this	a. investigated b. scrutinize/ illustrate/ describe	“ In this study we illustrate the potential for calling rates to be used as a tool for determining habitat quality by using an Antarctic pack-ice seal, the leopard seal, as a model.”
		31. the effect of + + on	a. these b. elevated c. mixing	a. approximations b. pressure c. time	“This paper reports the effect of elevated pressure on the invasive phenotype of patterned three-dimensional (3D) aggregates of MDA-MB-231 human breast cancer cells.”
		7. and the + + of the	a. absolute b. dynamic c. relative	a. value b. response/ responses c. position/ stiffness/ superiority	“The Fourier spectrum and response spectrum characteristics in the horizontal site are first investigated in this study, and the dynamic responses of the inclined layered sites are then

					studied and compared to the corresponding responses of the horizontal layered site.”
		8. the + of the + of	a. position b. analysis c. understanding	a. center b. effect/ improvement/ distributions c. mechanism/transport/functions	“This study was focused on 1) the calculation of the pulse pressure in an annular turbulent flow field based on the finite volume method, and 2) the analysis of the effect of pulse pressure on borehole stability.”
		15. on the + + of the	a. mechanical b. physical c. two	a. properties/ outputs/ property b. properties/ features c. components/ particles/ sides	“To search for the answer, this paper makes an analysis on the mechanical properties of the two shapes of the cross-sections.”
		22. that the + + of the	a. structural b. absolute c. relevant	a. plasticity/ properties/ stability b. value c. component	“This paper describes the mathematical properties of the linear model and the correlation evaluation metric that may lead to a misinterpretation of the results of this type of decoders. Firstly, the use of a linear regression model to adjust the two temporal signals (EEG and velocity profiles) implies that the relevant component of the signal used for decoding (EEG) has to be in the same frequency range as the signal to be decoded (velocity profiles).”
		24. the + and the + of	a. beginning b. accident c. alloys	a. end b. effects c. performance	“After the 2011 Fukushima Daiichi nuclear power plant accident, little information has been available on individual doses from external exposure among residents living in radioactively contaminated areas near the nuclear plant; in the present study we evaluated yearly changes in the doses from external exposure after the accident and the effects of decontamination on external exposure.”
	MR 2 - Apresentar a pesquisa/objetivo	25. we + that the + of	a. conclude b. demonstrate c. found	a. effect/ use/ site/ type b. sum/ effects/ anisotropy c. addition/ directionality/ distribution	“ We demonstrate that the effects of stabilizing mutations on the rigidity \Leftrightarrow flexibility equilibrium within the native state ensemble manifest themselves through

					enthalpy-entropy compensation as the protein structure adjusts to restore the global balance between the two.”
		29. here we + the + of	a. investigate b. describe c. show	a. effect/ process/ role b. structure/ accumulation/ characterization c. implementation/ action/ potential	“ Here we investigate the effect of crystal packing on side-chain conformational dynamics with crystal and solution molecular dynamics simulations using Cyanovirin-N as a model system.”
		36. we + the + of the	a. demonstrate b. evaluated c. investigated	a. utility/ efficacy/ capability b. outcome/ effect/ sensitivity c. interaction	“In the present work we investigated the interaction of the chemotherapeutic drug chlorambucil (CBL) with human GSTA1-1 (hGSTA1-1) using kinetic analysis, protein crystallography and molecular dynamics.”
MATERIALS, METHODS					
		1. to + the + of the	a. investigate b. evaluate c. determine	a. effect/ influence/ details b. accuracy/ performance/ effect c. influence/ correctness/ direction	“ To evaluate the accuracy of the measurement method in terms of particle density and interrogation voxel size, the gauge strain and one of the two cameras for Tomo-PIV were used as a video-extensometer in the tensile test.”
		2. of the + + of the	a. biomechanical b. different c. mechanical	a. properties/changes b. areas/ bands/ elements c. properties	“The evaluation of the biomechanical properties of the new EXPN with regard to torsional, compressive and bending stability as well as direct comparison to the biomechanical properties of conventional Steinmann (STMN)- and interlocking (ILN) nails was carried out with different biomechanical test arrangements.”
		6. the + of the + and	a. end b. effect c. results	a. movement/ bout/ measurements b. inlet/ magnitude/ broadness/ number/ nertia/ thermophoresis c. modeling/ simulation	“In this method, the effect of the number and thickness of soil layers, pore water pressure, arbitrary roof profile, and ground additional load is considered simultaneously.”

		8. the + of the + of	a. position b. analysis c. understanding	a. center b. effect/ improvement/ distributions c. mechanism/transport/functions	“Ground reaction forces and position changes of both feet were measured using instrumented shoes and used to estimate the position of the center of mass (CoM) .”
		12. to the + + of the	a. dynamic b. lateral c. different	a. behavior/mobility/pressure b. wall/organization c. models/ characteristics	“Owing to the dynamic behavior of the measurement system (radar and tripod), an accelerometer has been installed on the radar head to filter out the movement component of the measuring instrument from the measurement of the David's displacement.”
		14. used to + the + of	a. study b. evaluate c. alculate	a. role/ effects/changes b. performance/ effect/ properties c. number/ deformation/ movement	“A dataset of 15 congeneric inhibitors of the ACK1 protein was used to evaluate the performance of AFE set up protocols that varied in the steps taken to prepare input files (using previously docked and best scored poses, manual selection of poses, manual placement of binding site water molecules).”
		21. and + the + of the	a. compare b at c. by	a. composition/ performance/ values b. end c. intensity/ slope	“We combine spectrographic examinations, spectral analyses and automated classifications to identify different call types, and compare the composition of the vocal repertoire with that of other red deer subspecies.”
		24. the + and the + of	a. beginning b. accident c. alloys	a. end b. effects c. performance	“Intraocular pressure was measured at the beginning and the end of the measurements and ocular perfusion pressure was calculated as $\frac{2}{3} \times \text{mean arterial pressure} - \text{intraocular pressure}$.”
		31. the effect of + + on	a. these b. elevated c. mixing	a. approximations b. pressure c. time	“Quantitative and qualitative interactions between the effect of mixing time on drug detachment and the type of drug, the carrier size fraction and the flow rate were measured, which could be explained with the three processes mentioned.”

		35. the + of + of the	a. rate b. centre c. order	a. release/ change b. gravity c. magnitude	“The position of the centre of gravity of the human body was measured in dynamic conditions at 100 Hz.”
		38. between the + + and the	a. CAT b. DNA c. applied	a. dimensions b. lesion c. force	“A priori criterion for construct validity was that 75% of the correlations between the CAT dimensions and the subscales of the other questionnaires were as expected.”
		42. of the + + on the	a. CA b. channel c. compaction	a. diffusivity b. electrons c. parameters	“The blood compartment is modeled as a uniform compartment. The different blocks of the simulation are validated and compared to classical models. The impact of the CA diffusivity on the permeability and blood volume estimates is evaluated.”
FINDINGS, RESULTS					
		6. the + of the + and	a. end b. effect c. results	a. movement/ bout/ measurements b. inlet/ magnitude/ broadness/ number/ nertia/ thermophoresis c. modeling/ simulation	“ The results of the simulation and their comparison with the experiment involving human subjects are reported and discussed.”
		10. in the + + of the	a. first b upper c. southern	a. step/trial b. body/edges/ part/plate c. part	“During right and left turns, it significantly changes in the upper body of the dummy. ”
		18. of the + + in the	a. ASA b. CI c. DE	a. molecules b. particles c. loop	“Simulations at high temperature, could not re-activate the mobility of the loop in the thermophile.”
		21. and + the + of the	a. compare b at c. by	a. composition/ performance/ values b. end c. intensity/ slope	“Zygote mortality after one week averaged 50% and at the end of the second week, embryonic survival was higher under lower light and temperature conditions, which also determined

					the growth of larger embryos.”
		35. the + of + of the	a. rate b. centre c. order	a. release/ change b. gravity c. magnitude	“Therefore, these mutants increase the dissociation of the nucleotide cofactor, thereby increasing the rate of release of the product and producing the rate-limiting step in the hydride transfer.”
	MR 4 - Resultados/ descobertas	44. the + results + that the	a. experimental b. research c. simulation	a. demonstrated/ demonstrate/ /revealed b. indicated/ showed/ indicate c. showed/ demonstrated/ show	“ The experimental results showed that the proposed RTCI-DFSA algorithm performed better than the traditional frame slotted ALOHA (FSA) and dynamic frame slotted ALOHA (DFSA) algorithms.”
		47. and + + as well as	a. concentration b. temperature c. COOH	a. distribution/profiles b. fields/ profiles c. content	“The statistical analysis showed that the parameters calculated from QMS (areas of m/z = 16, 17, 18, 44), DSC (MaxDSC) and TG (ΔM) at R1, R2 and R3 correlated with selected chemical properties of the samples, such as N, O and COOH content as well as E2/E6 and E2/E4 indexes.”
CONCLUSION, DISCUSSION					
		3. the + and + of the	a. effectiveness b. accuracy c. feasibility	a. accuracy/ capability/ feasibility b. consistency/ efficiency/ proficiency c. effectiveness/ capability/ validity	“Simulation results verify the feasibility and effectiveness of the proposed control algorithm.”
		7. and the + + of the	a. absolute b. dynamic c. relative	a. value b. response/ responses c. position/ stiffness/ superiority	“The increase in permeability parameter leads to the decrease in both the temperature and the absolute value of the surface friction coefficient, and the increase in both the velocity and the absolute value of the surface temperature gradient.”

		10. in the + + of the	a. first b upper c. southern	a. step/trial b. body/edges/ part/plate c. part	“Our results suggest that the interface between the coastal current and inshore waters may constitute a major ecological barrier to larval dispersal in the southern part of the region that may only be overcome by rare, strong wind-forced events.”
		17. of the + + and the	a. test b. proposed c. pulse	a. distance/ result b. method/ approach/ model c. pressure	“The Lyapunov theory proves the system robust stability, semi-physical simulation verifies the validity of the proposed method and the superiority compared with the traditional method.”
	MR 5 - Conclusões	25. we + that the + of	a. conclude b. demonstrate c. found	a. effect/ use/ site/ type b. sum/ effects/ anisotropy c. addition/ directionality/ distribution	“ We conclude that the use of green light PPG could be of particular benefit in ambulatory monitoring where motion artifacts are a significant issue.”
		35. the + of + of the	a. rate b. centre c. order	a. release/ change b. gravity c. magnitude	“Moreover, this value is consistent with the order of magnitude of the volume occupied by the hydrated sugar-phosphate backbone of the DNA molecule, which is the part of the DNA molecule responsible for strand breaks.”
		39. can be + to + the	a. used b. applied c. expected	a. study/ distinguish/ /probe b. describe/ characterize/ screen c. affect/yield	“Based on these results, we conclude that synthetic lipid bilayers can be used to study the effects of ultrasound on cell membranes and membrane proteins.”
	MR 5 - Conclusões	45. results + that the + of	a. suggest b. show c. demonstrated	a. role/ flexibility/ expression b. influence/ effectiveness/ use c. impression/ size/ photolysis	“ Our results demonstrated that the size of outflow portal was the critical factor in determining irrigated flow rate, with a larger inflow portal and a smaller outflow portal leading to higher intra-articular pressure.”

		22. that the + + of the	a. structural b. absolute c. relevant	a. plasticity/ properties/ stability b. value c. component	“Together, these results underline that the structural plasticity of the N-terminal capping helix might represent a structural determinant for TPR domain structural versatility in cargo binding.”
		27. this study + the + of	g. demonstrate h. investigated i. determines	g. importance/ addition/ use h. adjustment/ changes/ feasibility i. impact	“Results of this study demonstrate the importance of taking integrated and place-based approaches when quantifying and managing for the coastal protection services supplied by ecosystems.”
	MR 5 - Discussão/conclusões/ implicações	44. the + results + that the	a. experimental b. research c. simulation	a. demonstrated/ demonstrate/ /revealed b. indicated/ showed/ indicate c. showed/ demonstrated/ show	“ The research results indicated that the horizontal position of the dispenser, the hole number, and the diameter had significant effects on the oil/water separation efficiency, and that the longitudinal position of the dispenser and the position of the weir plate had insignificant effects on the oil/water separation efficiency.”
	MR 5 - Discussão/conclusões/ implicações	45. results + that the + of	a. suggest b. show c. demonstrated	a. role/ flexibility/ expression b. influence/ effectiveness/ use c. impression/ size/ photolysis	“Overall, these results suggest that the role of polar phytoplankton in the utilization of labile dissolved organic matter may be significant under certain conditions and further confirm the relevance of solar radiation in regulating heterotrophy in the pelagic ocean.”

Fonte: Elaborado pela autora.

APÊNDICE 6

Estruturas lexicais da área de Medicina e Ciências da Saúde e exemplos.

Disciplina: Medicina e ciências da saúde Corpus: SC0319AbsME1mi					
LEGENDA:					
Quando as palavras estiverem em negrinho na cor preta elas pertencem a uma só estrutura lexical (quando houver só uma como exemplo), ou pertence a todas (quando houver mais de uma como exemplo).					
Seções identificadas nos resumos estruturados	Padrão retórico sugerido por Swales e Feak (2009)	Estruturas lexicais identificadas	LACUNA 1	LACUNA 2	Exemplo
BACKGROUND CONTEXT					
AIM OBJECTIVE PURPOSE					
		22. associated with + increased + of	a. an	a. risk/likelihood/incidence	“Despite evidence that older age, male sex and deprivation are associated with an increased incidence of colorectal cancer, uptake of bowel cancer screening varies across demographic groups.”
		25. associated with an + + of	a. increased	a. frequency/risk/incidence	
		26. associated + an increased + of	a. with	a. frequency/risk/incidence	
		30. of + + in patients with	a. disease b. femoral	a. recurrence/flares b. fractures	“Introduction: Correct risk assessment of disease recurrence in patients with early breast cancer

			c. adjuvant	c. chemotherapy/therapy	is critically important to detect patients who may be spared adjuvant chemotherapy.”
		30. of + + in patients with	a. disease b. femoral c. adjuvant	a. recurrence/flare b. fractures c. chemotherapy/therapy	“Most studies which analyzed the risk of femoral fractures in patients with idiopathic PD were performed either in specialized centers or excluded very frail patients.”
	MR 2 - Apresentar a pesquisa/objetivo	1. this study + to + the	a. was	a. determine/describe/ investigate	“ The aim of this study was to determine the incidence and associates of bacterial infection severe enough to require hospitalization in well-characterized community-based patients with type 2 diabetes.”
	MR 1 - Apresentar a pesquisa/objetivo	2. the + of + study was	a. aim	a. this	
	MR 1 - Apresentar a pesquisa/objetivo	3. the + of this + was	a. aim	a. study/ research/work	
	MR 1 - Apresentar a pesquisa/objetivo	4. the + + this study was	a. aim	a. of	
	MR 1 - Apresentar a pesquisa/objetivo	5. aim of + study + to	a. this	a. was/is	
	MR 1 - Apresentar a	6. aim of this + + to	a. study	a. was	

	pesquisa/objetivo				
MR 1 - Apresentar a pesquisa/objetivo	7. aim of + + was to	a. this	a. study		
MR 1 - Apresentar a pesquisa/objetivo	8. aim + this study + to	a. of	a. is/was		
MR 1 - Apresentar a pesquisa/objetivo	9. aim + + study was to	a. of	a. this		
MR 1 - Apresentar a pesquisa/objetivo	10. the aim of + + was	a. this	a. study/trial/research		
MR 1 - Apresentar a pesquisa/objetivo	11. the aim + + study was	a. of	a. this		
MR 1 - Apresentar a pesquisa/objetivo	12. this + was to + the	a. study	a. determine/describe/investigate		

	MR 1 - Apresentar a pesquisa/objetivo	13. this study was + + the	a. to	a. determine/describe/investigate	
	MR 1 - Apresentar a pesquisa/objetivo	24. aimed to + the + of	a. evaluate b. assess c. determine	a. effectiveness/impacts/effect b. effect/risk/feasibility c. percentage/performance/ incidence	“ This study aimed to evaluate the impact of the number of neoadjuvant chemotherapy (NAC) and postoperative adjuvant chemotherapy (POAC) cycles on the survival of patients with advanced ovarian cancer undergoing NAC/IDS/POAC.”
	MR 1 - Apresentar a pesquisa/objetivo	31. this + + to evaluate the	a. study b. article	a. was/aimed/is b. is	
		17. was to + the + of	a. evaluate b. investigate c. assess	a. effect/accuracy b. risk/usefulness/ability c. impact/knowledge/frequency	
	MR 2 - Apresentar a pesquisa/objetivo	14. aim + this + was to	a. of	a. this	“ The aim of this paper is to evaluate if a programme that aimed to improve maternal health, and mainly focusing on strengthening intrapartum care, also would reduce the number of stillbirths, and to estimate if there are other indicators that explains high stillbirth rates.”
	MR 1 - Apresentar a pesquisa/objetivo	15. the aim + this + was	a. of	a. study/research/paper	
	MR 1 - Apresentar a pesquisa/objetivo	32. the + of + study is	a. aim b. purpose c. objective	a. this/the/our b. the/this c. this/the	

MR 1 - Apresentar a pesquisa/objetivo	39. the + of this + is	a. aim b. purpose c. objective	a. study/review/paper b. study/article c. study/work	<p>“The purpose of this study was to evaluate the thyroid cancer rate and analyze recent changes in the extent of thyroid cancer surgeries in Korea.”</p>
MR 1 - Apresentar a pesquisa/objetivo	23. purpose of + study + to	a. this b. ous	a. was/is b. was/is	
MR 1 - Apresentar a pesquisa/objetivo	29. purpose of this + + to	a. study b. investigation c. work	a. was/is b. was c. was	
MR 1 - Apresentar a pesquisa/objetivo	35. this study + + evaluate the	a. was b. aimed c. is	a. to b. to c. to	
MR 1 - Apresentar a pesquisa/objetivo	36. of + study + to evaluate	a. this b. the c. our	a. was/is b. was/were/is c. is/was	
MR 1 - Apresentar a pesquisa/objetivo	40. purpose of + + was to	a. this b. our c. the	a. study/investigation/work b. study c. study	
MR 1 - Apresentar a pesquisa/objetivo	42. the purpose of + + was	a. this b. our c. the	a. study/work/investigation b. study c. study	
MR 1 -	46. purpose + + study was to	a. of b. of	a. this b. our	

	Apresentar a pesquisa/objetivo				
	MR 1 - Apresentar a pesquisa/objetivo	48. of this + + to evaluate	a. study	a. was/ was/were	
	MR 1 - Apresentar a pesquisa/objetivo	50. the purpose + + study was	a. of	a. this/our/the	
	MR 1 - Apresentar a pesquisa/objetivo	28. objective of + study + to	a. this b. the c. our	a. was/is b. was/is c. was	“ The objective of this study was to evaluate patterns and timing of risk of cancer in patients with metal-on-metal total hip replacements (THR).”
	MR 1 - Apresentar a pesquisa/objetivo	41. objective of this + + to	a. study b. analysis c. article	a. was/is b. was c. is	
	MR 1 - Apresentar a pesquisa/objetivo	47. objective + this study + to	a. of	a. is/was	
	MR 1 - Apresentar a pesquisa/objetivo	18. associated with a + + of	a. higher b. lower c. reduced	a. risk/rate/HR b. risk/proportion/incidence c. risk/hazard/dosage	
		21. the association between + + and	a. the b. PPI c. depressive	a. DFI/DII b. intake/use c. symptoms/disorders	“We aimed to examine the association between depressive symptoms and mental stress-induced myocardial ischemia in young survivors of an MI.”

		22. associated with + increased + of	a. an	a. risk/likelihood/incidence	“Objective: To determine whether latent tuberculosis infection risk factors are associated with an increased likelihood of latent tuberculosis infection testing in the US private healthcare sector.”
	MR 2 - Apresentar a pesquisa/objetivo	19. this + aimed to + the	a. study	a. investigate/establish/evaluate	“ This study aimed to evaluate the impact of the number of neoadjuvant chemotherapy (NAC) and postoperative adjuvant chemotherapy (POAC) cycles on the survival of patients with advanced ovarian cancer undergoing NAC/IDS/POAC.”
	MR 1 - Apresentar a pesquisa/objetivo	20. this study aimed + + the	a. to	a. investigate/establish/evaluate	
		44. to + the + of the	a. assess b. evaluate c. compare	a. impact/ performance/ effects b. impact/ applicability/ effect c. effectiveness/ performance/ utility	
MATERIALS, METHODS					
		21. the association between + + and	a. the b. PPI c. depressive	a. DFI/DII b. intake/use c. symptoms/disorders	“Cox proportional hazard models were used to estimate hazard ratios (HR) and 95% confidence intervals (CI) for the association between the DII and incident cardiovascular disease (myocardial infarction, stroke or cardiovascular death).”
		30. of + + in patients with	a. disease b. femoral c. adjuvant	a. recurrence/flare b. fractures c. chemotherapy/therapy	“Methods: To evaluate the role of adjuvant therapy in patients with thymic carcinoma, we retrospectively reviewed the records of patients with thymic carcinoma who were diagnosed and treated between 2004 and 2014.”
		33. used to + the + of	a. assess b. estimate c. evaluate	a. risk/ association/ quality b. effects/ impacts/ number c. role/ impacts/ association	“Cox proportional hazards regression model was used to assess the risk of developing CVD with adjustment for age, gender and co-morbid

					conditions.”
		44. to + the + of the	a. assess b. evaluate c. compare	a. impact/ performance/ effects b. impact/ applicability/ effect c. effectiveness/ performance/ utility	“Methods: Using English data from 2006/07 to 2010/11, we ran multilevel logit models to assess the impact of the QOF review on the risk of care home placement following emergency admission to acute hospital.”
FINDINGS, RESULTS					
		22. associated with + increased + of	a. an	a. risk/likelihood/incidence	“Results: The results showed that factors associated with an increased risk of cholangiocarcinoma included cholangitis, cholelithiasis, cholecystitis, cirrhosis of liver, alcoholic liver disease, chronic non-alcoholic liver disease, hepatitis B, hepatitis C, diabetes, chronic pancreatitis, inflammatory bowel disease, and peptic ulcer.”
		27. associated + a + risk of	a. with	a. higher/lower/reduced	
		37. associated + + increased risk of	a. with	an/the/a	
		38. was associated with + + of	a. increased b. higher c. lower	a. risk/incidence/proportion b. number/odds/rates c. risk/rates/likelihood	
		43. associated + an + risk of	a. with	a. increased/ elevated	
		45. hazard + + confidence interval ci	a. ratio(s)	a. (HR), 6.139; 95%	“Results: The Non-Traditional Risk Factor Index, made up of health deficits unrelated to CHD, was independently associated with incident CHD events over 10 years after controlling for age, sex,

					and the Traditional Risk Factor Index [adjusted {adj.} Hazard Ratio {HR} = 1.31; Confidence Interval {CI} 1.14–1.51]. ”
		49. hazard ratio + + interval ci	a. [HR] 0.58, 95%	a. confidence	
CONCLUSION DISCUSSION					
		16. associated with + + risk of	a. a b. an	a. higher/lower/reduced b. increased	“Preadmission use of antipsychotics was associated with a higher risk of severe stroke, a longer duration of hospital stay, and a higher post-stroke mortality, even after adjustment for known confounders.”
		18. associated with a + + of	a. higher b. lower c. reduced	a. risk/rate/HR b. risk/proportion/incidence c. risk/hazard/dosage	
		22. associated with + increased + of	a. an	a. risk/likelihood/incidence	“Conclusions: IBS is associated with an increased risk of dementia, and this effect is obvious only in patients who are ≥ 50 years old.”
		30. of + + in patients with	a. disease b. femoral c. adjuvant	a. recurrence/flares b. fractures c. chemotherapy/therapy	“Conclusion: The findings confirm the increased risk of femoral fractures in patients with idiopathic PD.”

Fonte: Elaborado pela autora.

APÊNDICE 7

50 ELs mais frequentes do *corpus* de Ciências da Computação e da Informação, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão
1	in this + we + a	307	306
2	to + the + of the	289	270
3	in this paper + + a	209	209
4	in + paper we + a	205	205
5	in this + we + the	196	193
6	the + of the + of	159	149
7	the + and + of the	154	149
8	in + + we propose a	132	132
9	in this + + propose a	122	122
10	in this paper + + the	120	119

11	we + the + of the	116	112
12	the + of this + is	115	113
13	used to + the + of	107	102
14	of the + + of the	106	99
15	in + paper + propose a	96	96
16	the + of the + and	96	96
17	and + the + of the	94	92
18	in + paper we + the	94	93
19	we + that the + of	94	89
20	the + results + that the	87	85
21	the + of + in the	84	84
22	based on the + + of	74	72

23	in this + we + an	70	70
24	this + we + a novel	69	69
25	of the + + in the	69	69
26	in the + + of the	66	60
27	order to + the + of	66	65
28	this study + to + the	66	65
29	can be + to + the	65	64
30	the + of this + was	61	61
31	and the + + of the	61	60
32	and + + as well as	60	60
33	this + we + a new	60	60
34	as well as + + of	60	58

35	in terms of + + and	60	60
36	the + and the + of	59	59
37	aim of this + + to	59	59
38	the + of the + in	58	58
39	in this study + + the	58	58
40	the + of the + is	57	54
41	of the + + and the	56	54
42	to the + + of the	56	55
43	in + + we present a	55	55
44	the + + in this paper	55	55
45	to + the + of a	55	54
46	on the + + of the	55	55

47	in the + of + and	54	53
48	the + + show that the	53	53
49	results + that the + of	52	51
50	in this + + present a	52	52

Fonte: Elaborado pela autora.

APÊNDICE 8

50 ELs mais frequentes do *corpus* da área de Física, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão
1	to + the + of the	250	236
2	of the + + of the	152	145
3	the + and + of the	145	139
4	in this + we + the	143	143
5	the + of this + was	140	140
6	the + of the + and	139	128
7	and the + + of the	133	125
8	the + of the + of	127	117
9	the + of + study was	121	121
10	in the + + of the	115	111

11	the + + this study was	113	113
12	to the + + of the	106	104
13	this study + to + the	102	102
14	used to + the + of	102	98
15	on the + + of the	101	99
16	in this + we + a	101	101
17	of the + + and the	97	93
18	of the + + in the	97	90
19	aim of this + + to	89	89
20	in this study + + the	81	81
21	and + the + of the	79	78
22	that the + + of the	79	76

23	aim of + study + to	77	77
24	the + and the + of	76	72
25	we + that the + of	74	74
26	aim of + + was to	71	71
27	this study + the + of	71	71
28	in + study we + the	70	70
29	here we + the + of	70	69
30	purpose of this + + to	70	70
31	the effect of + + on	69	67
32	aim + this study + to	68	68
33	the + of this + is	67	67
34	this + was to + the	67	67

35	the + of + of the	66	60
36	we + the + of the	65	65
37	the aim of + + was	65	65
38	between the + + and the	64	62
39	can be + to + the	64	63
40	aim + this + was to	63	63
41	the + of the + in	63	62
42	of the + + on the	62	60
43	aim + + study was to	60	60
44	the + results + that the	59	56
45	results + that the + of	59	58
46	this study was + + the	59	58

47	and + + as well as	58	56
48	the aim + this + was	58	58
49	of + + as well as	57	56
50	the + of the + is	57	56

Fonte: Elaborado pela autora.

APÊNDICE 9

50 ELs mais frequentes do *corpus* da área de Medicina e Ciências da Saúde, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão
1	this study + to + the	434	434
2	the + of + study was	392	391
3	the + of this + was	358	357
4	the + + this study was	347	346
5	aim of + study + to	290	290
6	aim of this + + to	256	256
7	aim of + + was to	252	252
8	aim + this study + to	243	243
9	aim + + study was to	242	242
10	the aim of + + was	238	238

11	the aim + + study was	229	229
12	this + was to + the	225	225
13	this study was + + the	221	221
14	aim + this + was to	216	216
15	the aim + this + was	206	206
16	associated with + + risk of	197	164
17	was to + the + of	167	165
18	associated with a + + of	156	131
19	this + aimed to + the	126	126
20	this study aimed + + the	123	123
21	the association between + + and	115	103
22	associated with + increased + of	111	98

23	purpose of + study + to	110	110
24	aimed to + the + of	108	108
25	associated with an + + of	108	96
26	associated + an increased + of	103	91
27	associated + a + risk of	100	83
28	objective of + study + to	99	99
29	purpose of this + + to	98	98
30	of + + in patients with	97	90
31	this + + to evaluate the	97	97
32	the + of + study is	95	95
33	used to + the + of	95	93
34	purpose + this study + to	94	94

35	this study + + evaluate the	94	94
36	of + study + to evaluate	92	92
37	associated + + increased risk of	91	81
38	was associated with + + of	90	79
39	the + of this + is	88	88
40	purpose of + + was to	88	88
41	objective of this + + to	88	88
42	the purpose of + + was	87	87
43	associated + an + risk of	86	77
44	to + the + of the	85	85
45	hazard + + confidence interval ci	84	84
46	purpose + + study was to	84	84

47	objective + this study + to	84	84
48	of this + + to evaluate	84	84
49	hazard ratio + + interval ci	83	83
50	the purpose + + study was	83	83

Fonte: Elaborado pela autora.

APÊNDICE 10

ELs do *corpus* de Ciências da Computação e da Informação, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão
1	in this + we + a	307	306
2	to + the + of the	289	270
3	in this paper + + a	209	209
4	in + paper we + a	205	205
5	in this + we + the	196	193
6	the + of the + of	159	149
7	the + and + of the	154	149
8	in + + we propose a	132	132
9	in this + + propose a	122	122
10	in this paper + + the	120	119

11	we + the + of the	116	112
12	the + of this + is	115	113
13	used to + the + of	107	102
14	of the + + of the	106	99
15	in + paper + propose a	96	96
16	the + of the + and	96	96
17	and + the + of the	94	92
18	in + paper we + the	94	93
19	we + that the + of	94	89
20	the + results + that the	87	85
21	the + of + in the	84	84
22	based on the + + of	74	72

23	in this + we + an	70	70
24	this + we + a novel	69	69
25	of the + + in the	69	69
26	in the + + of the	66	60
27	order to + the + of	66	65
28	this study + to + the	66	65
29	can be + to + the	65	64
30	the + of this + was	61	61
31	and the + + of the	61	60
32	and + + as well as	60	60
33	this + we + a new	60	60
34	as well as + + of	60	58

35	in terms of + + and	60	60
36	the + and the + of	59	59
37	aim of this + + to	59	59
38	the + of the + in	58	58
39	in this study + + the	58	58
40	the + of the + is	57	54
41	of the + + and the	56	54
42	to the + + of the	56	55
43	in + + we present a	55	55
44	the + + in this paper	55	55
45	to + the + of a	55	54
46	on the + + of the	55	55

47	in the + of + and	54	53
48	the + + show that the	53	53
49	results + that the + of	52	51
50	in this + + present a	52	52
51	it is + to + the	52	51
52	based on the + + the	52	50
53	the + of + study was	51	51
54	the experimental + + that the	51	49
55	in this study + + a	51	51
56	the experimental results + + the	51	49
57	of + + in this paper	50	50
58	the + of the + method	50	47

59	the + of + and the	49	48
60	in + study we + the	49	49
61	this paper + the + of	49	49
62	in + study we + a	48	48
63	this study + the + of	48	48
64	the + + this study was	47	47
65	the + of + of the	47	47
66	the + of the + network	47	45
67	the number of + + the	47	44
68	in + paper we + an	46	46
69	based on + + of the	46	44
70	can be used + + the	46	45

71	due to the + + of	46	46
72	in this paper + + an	46	46
73	the + + in terms of	45	43
74	the + of + proposed method	45	43
75	the + results show + the	44	44
76	can + used to + the	43	42
77	this paper + + a new	43	43
78	this paper + + a novel	43	43
79	paper we + the + of	43	43
80	this paper we + + new	43	43
81	the + + of the network	42	41
82	the + + the proposed method	42	40

83	this paper we + + novel	42	42
84	the + of the + the	41	41
85	the + + this paper is	40	40
86	aim of + + is to	40	40
87	aim of + study + to	40	40
88	the relationship between + + and	40	34
89	the + + as well as	39	38
90	the + + of the proposed	39	37
91	in + paper + present a	39	39
92	the + of + paper is	39	39
93	of + + as well as	38	37
94	the + + the number of	38	35

95	of + and the + of	38	38
96	objective of this + + to	38	38
97	based + the + of the	37	35
98	the + of + study is	37	37
99	that the + + of the	37	36
100	experimental results + that + proposed	37	37
101	have been + to + the	37	37
102	here we + the + of	37	36
103	of the + and + of	37	35
104	experimental + + that the proposed	36	36
105	this + is to + the	36	36
106	purpose of this + + to	36	36

107	results + the + of the	35	35
108	as + as the + of	35	35
109	the + of the + model	35	31
110	experimental results + + the proposed	35	35
111	as well + the + of	35	35
112	the aim of + + is	35	35
113	under the receiver + + curve	35	34
114	and + + in this paper	34	33
115	the + + that the proposed	34	33
116	in + work we + a	34	34
117	to + the performance + the	34	32
118	and to + the + of	34	33

119	study we + the + of	34	34
120	in this work + + a	34	34
121	in this work + + the	34	34
122	this + + propose a new	33	33
123	aim + this + is to	33	33
124	aim + this study + to	33	33
125	area + the receiver + characteristic	33	32
126	is + on the + of	33	33
127	is + to the + of	33	33
128	the + of the + to	33	32
129	of the + + on the	33	33
130	area under + receiver + characteristic	33	32

131	in this + we + on	33	33
132	in this + we + to	33	33
133	area under the + + characteristic	33	32
134	the + + this study is	32	32
135	to + + performance of the	32	30
136	under + receiver + characteristic curve	32	31
137	and + in this + we	32	32
138	this + we propose + new	32	32
139	under the + + characteristic curve	32	31
140	is to + the + of	32	32
141	we show that + + of	32	30
142	in + + in this paper	31	31

143	of + + in this study	31	31
144	of + + we show that	31	31
145	the + + in this study	31	29
146	of + in this + we	31	31
147	this + we propose + novel	31	31
148	in the + and + of	31	31
149	on the + and + of	31	31
150	this + + propose a novel	30	30
151	the + of + in a	30	29
152	main + of this + is	30	30
153	results + that the + method	30	30
154	the + in the + of	30	29

155	the + in this + we	30	30
156	the + of a + of	30	30
157	aim of + + was to	30	30
158	the impact of + + on	30	29
159	we propose a + + for	30	29

APÊNDICE 11

ELs do *corpus* da área de Física, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão	Número
1	to + the + of the	1	250	236
2	of the + + of the	2	152	145
3	the + and + of the	3	145	139
4	in this + we + the	4	143	143
5	the + of this + was	5	140	140
6	the + of the + and	6	139	128
7	and the + + of the	7	133	125
8	the + of the + of	8	127	117
9	the + of + study was	9	121	121
10	in the + + of the	10	115	111

11	the + + this study was	11	113	113
12	to the + + of the	12	106	104
13	this study + to + the	13	102	102
14	used to + the + of	13	102	98
15	on the + + of the	15	101	99
16	in this + we + a	15	101	101
17	of the + + and the	17	97	93
18	of the + + in the	17	97	90
19	aim of this + + to	19	89	89
20	in this study + + the	20	81	81
21	and + the + of the	21	79	78
22	that the + + of the	21	79	76

23	aim of + study + to	23	77	77
24	the + and the + of	24	76	72
25	we + that the + of	25	74	74
26	aim of + + was to	26	71	71
27	this study + the + of	26	71	71
28	in + study we + the	28	70	70
29	here we + the + of	28	70	69
30	purpose of this + + to	28	70	70
31	the effect of + + on	31	69	67
32	aim + this study + to	32	68	68
33	the + of this + is	33	67	67
34	this + was to + the	33	67	67

35	the + of + of the	35	66	60
36	we + the + of the	36	65	65
37	the aim of + + was	36	65	65
38	between the + + and the	38	64	62
39	can be + to + the	38	64	63
40	aim + this + was to	40	63	63
41	the + of the + in	40	63	62
42	of the + + on the	42	62	60
43	aim + + study was to	43	60	60
44	the + results + that the	44	59	56
45	results + that the + of	44	59	58
46	this study was + + the	44	59	58

47	and + + as well as	47	58	56
48	the aim + this + was	47	58	58
49	of + + as well as	49	57	56
50	the + of the + is	49	57	56
51	purpose of + study + to	49	57	57
52	was to + the + of	49	57	56
53	based on the + + of	49	57	57
54	in the + of + and	54	56	54
55	purpose + this study + to	55	55	55
56	the + of + and the	56	54	51
57	the + of + in the	56	54	50
58	the aim + + study was	56	54	54

59	based on the + + the	59	53	53
60	for the + + of the	60	52	52
61	purpose of + + was to	60	52	52
62	order to + the + of	60	52	51
63	as well as + + of	60	52	51
64	the effects of + + on	64	51	49
65	purpose + this + was to	65	50	50
66	the + of the + the	65	50	49
67	the purpose of + + was	65	50	50
68	can be used + + the	68	49	49
69	the + of + on the	69	48	47
70	the purpose + this + was	69	48	48

71	is + to the + of	71	47	47
72	it is + to + the	71	47	45
73	can + used to + the	73	46	46
74	the + of the + was	73	46	45
75	purpose + + study was to	75	45	45
76	in this paper + + a	75	45	45
77	the + + in this study	77	44	42
78	of the + + to the	77	44	42
79	the + in the + of	79	43	41
80	the purpose + + study was	79	43	43
81	in this + the + of	79	43	43
82	the influence of + + on	79	43	40

83	in + paper we + a	83	42	42
84	the + of the + study	83	42	42
85	due to the + + of	83	42	41
86	we + the + of a	86	41	40
87	the + + the present study	87	40	40
88	is + in the + of	87	40	39
89	in this work + + the	87	40	40
90	play + + role in the	90	39	39
91	the + + as well as	90	39	38
92	of + and the + of	90	39	35
93	the + of the + to	90	39	37
94	the effects of + + and	90	39	38

95	based on + + of the	95	38	37
96	the + + based on the	96	37	35
97	in + the + of the	96	37	37
98	the + of + study is	96	37	37
99	this + is to + the	96	37	37
100	effect of + + on the	96	37	36
101	study we + the + of	96	37	37
102	the impact of + + on	96	37	36
103	the relationship between + + and	96	37	35
104	by + the + of the	104	36	36
105	the + and + properties of	104	36	36
106	in + work we + the	104	36	36

107	of + in the + of	104	36	36
108	at the + + of the	104	36	35
109	of the + study + to	104	36	36
110	to investigate the + + of	104	36	34
111	the + + this study is	111	35	35
112	the + the + of the	111	35	34
113	to + the + of a	111	35	33
114	the + between the + and	111	35	33
115	of this + + to investigate	111	35	35
116	with the + + of the	111	35	35
117	in the + and + of	111	35	34
118	of + study + to investigate	118	34	34

119	the + of + present study	118	34	34
120	is + for the + of	118	34	32
121	of the + and + of	118	34	34
122	in this paper + + the	118	34	34
123	the + + show that the	123	33	32
124	for the + and + of	123	33	33
125	is to + the + of	123	33	33
126	model is + to + the	123	33	32
127	in terms of + + and	123	33	33
128	of the present + + to	123	33	33
129	of the + + was to	129	32	32
130	in the + of + the	129	32	31

131	of the + the + of	129	32	32
132	on the + of + and	129	32	32
133	in this study + + a	129	32	32
134	objective of this + + to	129	32	32
135	as + as the + of	135	31	31
136	we + the crystal + of	135	31	31
137	aim of + + is to	135	31	31
138	as well + the + of	135	31	31
139	on the + and + of	135	31	31
140	this paper + the + of	135	31	31
141	in order to + + the	135	31	31
142	in this study + + of	135	31	31

143	in + + we propose a	143	30	30
144	of + + was to investigate	143	30	30
145	the + + in order to	143	30	29
146	and + in the + of	143	30	30
147	be + for the + of	143	30	30
148	in + study we + a	143	30	30
149	by the + + of the	143	30	30
150	in the + + and the	143	30	29
151	has been + to + the	143	30	29
152	method is + to + the	143	30	30
153	due to the + + the	143	30	30
154	of this study + + investigate	143	30	30

APÊNDICE 12

ELs do *corpus* da área de Medicina e Ciências da Saúde, ordenadas por sua frequência absoluta no *corpus*.

Número	EL	Frequência absoluta	Dispersão
1	this study + to + the	434	434
2	the + of + study was	392	391
3	the + of this + was	358	357
4	the + + this study was	347	346
5	aim of + study + to	290	290
6	aim of this + + to	256	256
7	aim of + + was to	252	252
8	aim + this study + to	243	243
9	aim + + study was to	242	242
10	the aim of + + was	238	238

11	the aim + + study was	229	229
12	this + was to + the	225	225
13	this study was + + the	221	221
14	aim + this + was to	216	216
15	the aim + this + was	206	206
16	associated with + + risk of	197	164
17	was to + the + of	167	165
18	associated with a + + of	156	131
19	this + aimed to + the	126	126
20	this study aimed + + the	123	123
21	the association between + + and	115	103
22	associated with + increased + of	111	98

23	purpose of + study + to	110	110
24	aimed to + the + of	108	108
25	associated with an + + of	108	96
26	associated + an increased + of	103	91
27	associated + a + risk of	100	83
28	objective of + study + to	99	99
29	purpose of this + + to	98	98
30	of + + in patients with	97	90
31	this + + to evaluate the	97	97
32	the + of + study is	95	95
33	used to + the + of	95	93
34	purpose + this study + to	94	94

35	this study + + evaluate the	94	94
36	of + study + to evaluate	92	92
37	associated + + increased risk of	91	81
38	was associated with + + of	90	79
39	the + of this + is	88	88
40	purpose of + + was to	88	88
41	objective of this + + to	88	88
42	the purpose of + + was	87	87
43	associated + an + risk of	86	77
44	to + the + of the	85	85
45	hazard + + confidence interval ci	84	84
46	purpose + + study was to	84	84

47	objective + this study + to	84	84
48	of this + + to evaluate	84	84
49	hazard ratio + + interval ci	83	83
50	the purpose + + study was	83	83
51	hazard ratio + confidence + ci	82	82
52	of this study + + evaluate	81	81
53	objective of + + was to	80	80
54	objective + + study was to	78	78
55	purpose + this + was to	78	78
56	analysis was + to + the	78	73
57	the purpose + this + was	77	77
58	the + + this study is	76	76

59	of + + was to evaluate	74	74
60	of + study + to investigate	74	74
61	of + study was + evaluate	73	73
62	the objective of + + was	72	72
63	the objective + + study was	71	71
64	to evaluate the + + of	70	69
65	hazard + hr + interval ci	69	69
66	hazard + hr confidence + ci	69	69
67	hr ci + + hr ci	69	56
68	hazard ratio hr + + ci	69	69
69	objective + this + was to	68	68
70	was associated with + + risk	68	58

71	this + + to investigate the	66	66
72	the + of + in the	66	62
73	in the + group + in	66	53
74	of this + was + evaluate	66	66
75	or ci + + or ci	65	50
76	this study + the + of	65	63
77	there + no significant + in	64	61
78	associated with + higher + of	64	56
79	odds + + confidence interval ci	63	63
80	the + group + in the	63	51
81	of this + + to investigate	63	63
82	of + + study was to	62	62

83	of + + was to investigate	62	62
84	odds ratio + + interval ci	62	62
85	models were + to + the	62	62
86	the objective + this + was	62	62
87	of this study + + investigate	62	62
88	of + study + to assess	61	61
89	of + study was + investigate	61	61
90	ratio + confidence interval + p	61	61
91	this study + + investigate the	61	61
92	ratio + + interval ci p	60	60
93	results + total of + were	60	60
94	objective to + the + of	60	60

95	odds ratio + confidence + ci	60	60
96	results a + of + were	60	60
97	used to + the + between	60	59
98	results a total + + were	60	60
99	the relationship between + + and	60	56
100	survival + + overall survival os	59	58
101	the + + of patients with	59	56
102	ratio + confidence + ci p	59	59
103	the + of the + of	59	55
104	survival + and + survival os	58	57
105	survival + and overall + os	58	57
106	associated + a higher + of	57	52

107	group + in the + group	57	44
108	the + national health + research	57	56
109	were associated with + + of	57	55
110	the + + health insurance research	56	55
111	the + national + insurance research	56	55
112	odds + or + interval ci	55	55
113	the + of the + study	55	53
114	free survival + + overall survival	55	54
115	to assess the + + of	55	55
116	free + + and overall survival	54	53
117	this + + to assess the	54	54
118	our + was to + the	54	54

119	logistic regression + + used to	54	54
120	free survival + and + survival	54	53
121	of the + study + to	54	54
122	of this + was + investigate	54	54
123	of + study + to determine	53	53
124	taiwan + health + research database	53	53
125	this + was + evaluate the	53	53
126	odds + or confidence + ci	53	53
127	taiwan national + + research database	53	53
128	in the + group + the	53	45
129	there was + significant + in	53	53
130	is associated with + + of	53	52

131	odds ratio or + + ci	53	53
132	taiwan national health + + database	53	53
133	taiwan + + insurance research database	52	52
134	taiwan + health insurance + database	52	52
135	this + is to + the	52	52
136	aim of + + is to	52	52
137	of the + + was to	52	52
138	of this + + to assess	52	52
139	this study + + assess the	52	52
140	taiwan national + insurance + database	52	52
141	models were used + + the	52	52

142	this study is + + the	52	52
143	this + + to determine the	51	51
144	study + to + the association	51	51
145	the impact of + + on	51	48
146	models + used to + the	50	50
147	the + epidemiology and + results	50	50
148	regression was + to + the	50	50
149	the surveillance + and + results	50	50
150	the taiwan + health + research	50	50
151	the aim of + + is	50	50
152	the national health + + database	50	50
153	the surveillance epidemiology + + results	50	50

154	the taiwan national + + research	50	50
155	non + + lung cancer nslc	49	49
156	the + + and end results	49	49
157	was + + in patients with	49	46
158	non + cell + cancer nslc	49	49
159	the + epidemiology + end results	49	49
160	non + cell lung + nslc	49	49
161	the + health insurance + database	49	49
162	non small + + cancer nslc	49	49
163	the surveillance + + end results	49	49
164	the taiwan + + insurance research	49	49
165	logistic regression + were + to	49	49

166	non small + lung + nslc	49	49
167	the national + insurance + database	49	49
168	non small cell + + nslc	49	49
169	the + health + research database	48	48
170	analysis + used to + the	48	47
171	there + no significant + between	48	45
172	of this + + to determine	48	48
173	the national + + research database	48	48
174	study was + to + the	48	48
175	of this study + + assess	48	48
176	aim + + study is to	47	47

177	from + + national health insurance	47	47
178	of + + was to assess	47	47
179	the + + insurance research database	47	47
180	the + + than in the	47	45
181	the + of + in patients	47	40
182	significant + in the + of	47	44
183	in the + + than in	47	45
184	this study + + determine the	47	47
185	in patients with + + and	47	44
186	of the present + + to	47	47
187	the + + the present study	46	44
188	a + of + were included	46	46

189	of + study was + assess	46	46
190	there was + + difference in	46	46
191	a total of + + included	46	46
192	there was no + + in	46	46
193	of + + was to determine	45	45
194	results + + of patients were	45	45
195	the + of + present study	45	43
196	to + the + between the	45	42
197	of + and the + of	45	44
198	of + present study + to	45	45
199	a total + + were included	45	45
200	the aim + + study is	45	45

201	aimed to + the + and	45	45
202	was to + the + between	45	45
203	of this study + + determine	45	45
204	the effect of + + on	45	42
205	we conducted a + + study	45	45
206	and + + in patients with	44	43
207	survival + and + free survival	44	43
208	to + the + value of	44	44
209	model was + to + the	44	44
210	present study + to + the	44	44
211	analysis was used + + the	44	43
212	surveillance epidemiology and + + seer	44	44

213	the aim of + + study	44	44
214	a + + of patients with	43	42
215	surveillance + and + results seer	43	43
216	we + a + cohort study	43	43
217	of + study was + determine	43	43
218	surveillance + and end + seer	43	43
219	the + of the + was	43	43
220	surveillance epidemiology + + results seer	43	43
221	patients were + into + groups	43	42
222	surveillance epidemiology + end + seer	43	43
223	was to + the + and	43	43
224	and + + were associated with	42	39

225	surveillance + + end results seer	42	42
226	of + present + was to	42	42
227	results + total + patients were	42	42
228	was + evaluate the + of	42	42
229	aim of + + study was	42	42
230	in the + + and in	42	38
231	in the + + in the	42	39
232	results a + + patients were	42	42
233	in this study + + the	42	42
234	median follow up + + months	42	42
235	was to evaluate + + of	42	42
236	there + + significant difference in	41	41

237	was + in the + group	41	33
238	with a + + of years	41	40
239	logistic regression + was + to	41	41
240	patients with a + + of	41	40
241	was associated with + + in	41	41
242	aim + this + is to	40	40
243	the + and + of the	40	40
244	the + of + and the	40	38
245	there + no + difference in	40	40
246	was + to + the association	40	39
247	significantly + in the + group	40	33
248	to + the prognostic + of	40	40

249	from the + + health insurance	40	40
250	in this + we + the	40	40
251	objective the + of + study	40	40
252	there was + significant + between	40	36
253	a total of + + were	40	39
254	is associated with + + risk	40	38
255	associated + + higher risk of	39	35
256	this + + to examine the	39	39
257	of + study + to examine	39	39
258	aimed to + the + between	39	39
259	analyses were + to + the	39	38
260	of this + was + determine	39	39

261	a total of + + patients	39	39
262	the risk of + + and	39	35
263	and + + as well as	38	37
264	the + + and in the	38	34
265	of + were + in the	38	37
266	in + study we + the	38	38
267	regression + used to + the	38	38
268	risk of + + patients with	38	31
269	the aim + this + is	38	38
270	the risk of + + in	38	34
271	and + + were and respectively	37	35
272	in + + of patients with	37	36

273	p + + hr ci p	37	27
274	with + + age of years	37	36
275	model + used to + the	37	37
276	ci p + + hr ci	37	27
277	in the + + group and	37	28
278	of this + + to examine	37	37
279	ratio hr + + ci p	37	37
280	is to + the + of	37	37
281	of this + was + assess	37	37
282	ratio hr + interval + p	37	37
283	the aim + the + study	37	37
284	aim of the + + was	37	37

285	significantly associated with + + of	37	34
286	is + + risk factor for	36	33
287	taiwan + + health insurance research	36	36
288	used + + the association between	36	35
289	adjusted + ratio + confidence interval	36	36
290	taiwan + national + insurance research	36	36
291	s + health insurance + database	36	36
292	taiwan + national health + research	36	36
293	had a + + risk of	36	33
294	taiwan s + + insurance research	36	36

295	this study + + examine the	36	36
296	used to + + association between	36	35
297	purpose to + the + of	36	36
298	taiwan s + health + research	36	36
299	was significantly + in + with	36	34
300	ratio hr confidence + + p	36	36
301	taiwan s national + + research	36	36
302	was an independent + + for	36	35
303	s + + insurance research database	35	35
304	aim + the + study was	35	35
305	s + health + research database	35	35
306	ci p + + ci p	35	30

307	p and + + ci p	35	32
308	s national + + research database	35	35
309	was an + + factor for	35	34
310	was associated + + risk of	35	31
311	s national + insurance + database	35	35
312	this study + to + whether	35	35
313	with a + age + years	35	34
314	ci p and + + ci	35	32
315	regression was used + + the	35	35
316	s national health + + database	35	35
317	there was no + + between	35	32
318	to determine the + + of	35	35

319	during + + follow up of	34	32
320	for + + in patients with	34	33
321	this + was + investigate the	34	34
322	was + with + risk of	34	30
323	a + follow up + of	34	33
324	the + follow up + was	34	33
325	of the + + of the	34	29
326	survival os + + free survival	34	34
327	and to + the + of	34	34
328	associated with a + + risk	34	31
329	cox proportional hazards + + used	34	34
330	hr ci p + + hr	34	26
331	in patients with + + disease	34	30

332	model was used + + the	34	34
333	of patients with + + and	34	33
334	of this study + + examine	34	34
335	overall survival os + + free	34	34
336	patients in the + + group	34	26
337	were associated with + + risk	34	30
338	a + + up of years	33	32
339	methods + + of patients with	33	33
340	a + follow + of years	33	32
341	a + follow up + years	33	32
342	the + of the + and	33	32
343	was + with a + risk	33	28

344	proportional hazards + + used to	33	33
345	with a + + risk of	33	30
346	from the + national + insurance	33	33
347	had a significantly + + of	33	31
348	logistic regression analysis + + to	33	33
349	a + + up of months	32	32
350	objective + + of this study	32	32
351	of + + was to examine	32	32
352	risk + + in patients with	32	24
353	a + follow + of months	32	32
354	was + independent + factor for	32	31
355	a + follow up + months	32	32

356	cox proportional + + was used	32	32
357	objective the + + this study	32	32
358	the aim + + present study	32	32
359	we conducted + + cohort study	32	32
360	this study + to + a	32	32
361	total of + were + in	32	32
362	was associated + a + risk	32	27
363	survival os and + + survival	32	32
364	there were no + + in	32	31
365	a + + patients were included	31	31
366	aim + + present study was	31	31
367	of + + a total of	31	31

368	of + + included in the	31	31
369	proportional + + was used to	31	31
370	ratios + + confidence intervals ci	31	31
371	using + + national health insurance	31	30
372	of + study + to identify	31	31
373	from + national health + research	31	31
374	of + study was + examine	31	31
375	in the + + of the	31	30
376	of the + + in the	31	30
377	aim of + present + was	31	31
378	patients were + in + study	31	31
379	results a + of + with	31	31
380	be associated with + + of	31	31

381	from the taiwan + + insurance	31	31
382	from + + health insurance research	30	30
383	ratio + + confidence interval ci	30	29
384	the + + the study was	30	30
385	from + national + insurance research	30	30
386	from + taiwan + health insurance	30	30
387	study + to + the risk	30	30
388	the + of + on the	30	30
389	to + the + of a	30	30
390	were + to + the association	30	30
391	a + of patients + included	30	30
392	aim + the present + was	30	30
393	logistic + analysis was + to	30	30

394	primary + was the + of	30	30
395	results + total of + with	30	30
396	we + a retrospective + of	30	30
397	of this + + to identify	30	30
398	a total + patients + included	30	30
399	overall survival + and + free	30	30
400	performed to + the + of	30	30
401	risk of + in + with	30	24
402	was associated + increased + of	30	28
403	results a total + + with	30	30
404	was to investigate + + of	30	30

Fonte: Elaborado pela autora.

APÊNDICE 13

Estruturas Lexicais que aparecem em um dos *corpora*, em dois *corpora* ou em três *corpora*.

Ciências da computação e da informação			Física			Medicina e ciências da saúde		
EL	Freq. Abs.	Dispersão	EL	Freq. Abs.	Dispersão	EL	Freq. Abs.	Dispersão
ELs que ocorrem nos três <i>corpora</i>								
to + the + of the	289	270	to + the + of the	250	236	to + the + of the	85	85
the + of this + is	115	113	the + of this + is	67	67	the + of this + is	88	88
used to + the + of	107	102	used to + the + of	102	98	used to + the + of	95	93
this study + to + the	66	65	this study + to + the	102	102	this study + to + the	434	434
the + of this + was	61	61	the + of this + was	140	140	the + of this + was	358	357
aim of this + + to	59	59	aim of this + + to	89	89	aim of this + + to	256	256
ELs que ocorrem em dois <i>corpora</i>								
in this + we + a	307	306	in this + we + a	101	101	-	-	-

in this + we + the	196	193	in this + we + the	143	143	-	-	-
the + of the + of	159	149	the + of the + of	127	117	-	-	-
the + and + of the	154	149	the + and + of the	145	139	-	-	-
we + the + of the	116	112	we + the + of the	65	65	-	-	-
of the + + of the	106	99	of the + + of the	152	145	-	-	-
the + of the + and	96	96	the + of the + and	139	128	-	-	-
and + the + of the	94	92	and + the + of the	79	78	-	-	-
we + that the + of	94	89	we + that the + of	74	74	-	-	-
the + results + that the	87	85	the + results + that the	59	56	-	-	-
of the + + in the	69	69	of the + + in the	97	90	-	-	-
in the + + of the	66	60	in the + + of the	115	111	-	-	-
can be + to + the	65	64	can be + to + the	64	63	-	-	-

and the + + of the	61	60	and the + + of the	133	125	-	-	-
and + + as well as	60	60	and + + as well as	58	56	-	-	-
the + and the + of	59	59	the + and the + of	76	72	-	-	-
the + of the + in	58	58	the + of the + in	63	62	-	-	-
in this study + + the	58	58	in this study + + the	81	81	-	-	-
the + of the + is	57	54	the + of the + is	57	56	-	-	-
of the + + and the	56	54	of the + + and the	97	93	-	-	-
to the + + of the	56	55	to the + + of the	106	104	-	-	-
on the + + of the	55	55	on the + + of the	101	99	-	-	-
results + that the + of	52	51	results + that the + of	59	58	-	-	-
-	-	-	the + of + study was	121	121	the + of + study was	392	391
-	-	-	the + + this study was	113	113	the + + this study was	347	346

-	-	-	aim of + study + to	77	77	aim of + study + to	290	290
-	-	-	aim of + + was to	71	71	aim of + + was to	252	252
-	-	-	purpose of this + + to	70	70	purpose of this + + to	98	98
-	-	-	aim + this study + to	68	68	aim + this study + to	243	243
-	-	-	this + was to + the	67	67	this + was to + the	225	225
-	-	-	the aim of + + was	65	65	the aim of + + was	238	238
-	-	-	aim + this + was to	63	63	aim + this + was to	216	216
-	-	-	aim + + study was to	60	60	aim + + study was to	242	242
-	-	-	this study was + + the	59	58	this study was + + the	221	221
-	-	-	the aim + this + was	58	58	the aim + this + was	206	206
ELs que ocorrem em um <i>corpus</i>								
-	-	-	-	-	-	this + + to evaluate the	97	97

in this paper + + a	209	209	-	-	-	-	-	-
in + paper we + a	205	205	-	-	-	-	-	-
in + + we propose a	132	132	-	-	-	-	-	-
in this + + propose a	122	122	-	-	-	-	-	-
in this paper + + the	120	119	-	-	-	-	-	-
in + paper + propose a	96	96	-	-	-	-	-	-
in + paper we + the	94	93	-	-	-	-	-	-
the + of + in the	84	84	-	-	-	-	-	-
based on the + + of	74	72	-	-	-	-	-	-
in this + we + an	70	70	-	-	-	-	-	-
this + we + a novel	69	69	-	-	-	-	-	-
order to + the + of	66	65	-	-	-	-	-	-

this + we + a new	60	60	-	-	-	-	-	-
as well as + + of	60	58	-	-	-	-	-	-
in terms of + + and	60	60	-	-	-	-	-	-
in + + we present a	55	55	-	-	-	-	-	-
the + + in this paper	55	55	-	-	-	-	-	-
to + the + of a	55	54	-	-	-	-	-	-
in the + of + and	54	53	-	-	-	-	-	-
the + + show that the	53	53	-	-	-	-	-	-
in this + + present a	52	52	-	-	-	-	-	-
-	-	-	that the + + of the	79	76	-	-	-
-	-	-	this study + the + of	71	71	-	-	-
-	-	-	in + study we + the	70	70	-	-	-

-	-	-	here we + the + of	70	69	-	-	-
-	-	-	the effect of ++ on	69	67	-	-	-
-	-	-	the + of + of the	66	60	-	-	-
-	-	-	between the ++ and the	64	62	-	-	-
-	-	-	of the ++ on the	62	60	-	-	-
-	-	-	of ++ as well as	57	56	-	-	-
-	-	-	-	-	-	the aim ++ study was	229	229
-	-	-	-	-	-	associated with ++ risk of	197	164
-	-	-	-	-	-	was to + the + of	167	165
-	-	-	-	-	-	associated with a ++ of	156	131
-	-	-	-	-	-	this + aimed to + the	126	126
-	-	-	-	-	-	this study aimed ++ the	123	123

-	-	-	-	-	-	the association between + + and	115	103
-	-	-	-	-	-	associated with + increased + of	111	98
-	-	-	-	-	-	purpose of + study + to	110	110
-	-	-	-	-	-	aimed to + the + of	108	108
-	-	-	-	-	-	associated with an + + of	108	96
-	-	-	-	-	-	associated + an increased + of	103	91
-	-	-	-	-	-	associated + a + risk of	100	83
-	-	-	-	-	-	objective of + study + to	99	99
-	-	-	-	-	-	of + + in patients with	97	90
-	-	-	-	-	-	this + + to evaluate the	97	97
-	-	-	-	-	-	the + of + study is	95	95
-	-	-	-	-	-	purpose + this study + to	94	94

-	-	-	-	-	-	this study ++ evaluate the	94	94
-	-	-	-	-	-	of + study + to evaluate	92	92
-	-	-	-	-	-	associated ++ increased risk of	91	81
-	-	-	-	-	-	was associated with ++ of	90	79
-	-	-	-	-	-	purpose of ++ was to	88	88
-	-	-	-	-	-	objective of this ++ to	88	88
-	-	-	-	-	-	the purpose of ++ was	87	87
-	-	-	-	-	-	associated + an + risk of	86	77
-	-	-	-	-	-	hazard ++ confidence interval ci	84	84
-	-	-	-	-	-	purpose ++ study was to	84	84
-	-	-	-	-	-	objective + this study + to	84	84
-	-	-	-	-	-	of this ++ to evaluate	84	84

-	-	-	-	-	-	hazard ratio + + interval ci	83	83
-	-	-	-	-	-	the purpose + + study was	83	83

Fonte: Elaborado pela autora.

APÊNDICE 14

Exemplo de resumo acadêmico da área da Física.

MODELO DE RESUMO ACADÊMICO DA ÁREA DA FÍSICA

SEÇÃO 1: apresentando a pesquisa

Exemplo:

(...) Here, we **(investigate) the (effect/ process/ role)** ou **(describe) the (structure/ accumulation/ characterization)** ou **(show) the (implementation/ action/ potential)** of **(DESCREVA O QUE FOI/ESTÁ SENDO INVESTIGADO/ DESCRITO OU APRESENTADO) micropillar arrays comprising pillars of different modulus, spacing, diameter, and aspect ratio due to the sliding of a silicone cast of a human finger.** The (aim/ purpose/ objective) (of) this (paper/ study/ work) (was/is) to **(determine) the (influence/ correctness/ direction)** ou **(investigate) the (effect/ influence/ details)** ou **(evaluate) the (accuracy/ performance/ effect)** of **(DESCREVA O QUE FOI/ESTÁ SENDO DETERMINADO/INVESTIGADO OU AVALIADO) (pumping coordination in branched network structures on the regulation of lymph flow).** (...)

SEÇÃO 2: apresentação da metodologia

(...) **The evaluation of the (biomechanical properties) ou (different elements) ou (mechanical properties) of the new EXPN with regard to torsional, compressive and bending stability as well as direct comparison to the biomechanical properties of conventional Steinmann (STMN)- and interlocking (ILN) nails was carried out with different biomechanical test arrangements.**

SEÇÃO 3: apresentação dos resultados

Exemplo: (...) **In healthy persons at age 20, 40, and 60, respectively, the rate of (release/ change) of the oxygen photoconsumption by lipofuscin increases by 1.3, 1.7, and 2.4 fold during SW-AF-imaging as compared to diffuse sunlight. In patients with STGD1 below the age of 30, this rate was 3.3-fold higher compared to age-matched controls during either sunlight or SW-AF imaging.**

SEÇÃO 4: apresentação das conclusões

Exemplo:

(...) **Results (suggest) that the (role/ flexibility/ expression) ou (show) that the (influence/ effectiveness/ use) ou (demonstrated) that the (impression/ size/ photolysis) ou The (experimental) results (demonstrated/ demonstrate/ revealed) ou The (research) results (indicated/ showed/ indicate) ou The (simulation) results (showed/ demonstrated/ show) (DESCREVA O QUE OS RESULTADOS SUGEREM/APRESENTAM OU DEMONSTRAM) of patients with STGD1 is generally at increased risk of photo-oxidative stress, while exposure during SW-AF-imaging amplifies this risk. These theoretical results have not yet been verified with in-vivo data due to a lack of sufficiently sensitive in-vivo measurement techniques.**

Fonte: Elaborado pela autora.