

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
DEPARTAMENTO DE LÍNGUAS MODERNAS**

VIVIAN MEURER LAMEIRA

**PROPOSTA DE ANÁLISE LINGUÍSTICA PARA A IDENTIFICAÇÃO DE
FRASEOLOGIAS DE GÊNERO E FRASEOLOGIAS DE GÊNERO DA ÁREA
EXTRAÍDAS DE UM CORPUS DE *ABSTRACTS* DA MEDICINA**

PORTO ALEGRE

2017

VIVIAN MEURER LAMEIRA

**PROPOSTA DE ANÁLISE LINGUÍSTICA PARA A IDENTIFICAÇÃO DE
FRASEOLOGIAS DE GÊNERO E FRASEOLOGIAS DE GÊNERO DA ÁREA
EXTRAÍDAS DE UM CORPUS DE *ABSTRACTS* DA MEDICINA**

Monografia apresentada como requisito parcial
para obtenção do título de Licenciada em
Letras pelo Instituto de Letras da Universidade
Federal do Rio Grande do Sul

Orientadora: Profa. Dra. Ana Eliza Pereira Bocorny

PORTO ALEGRE

2017

CIP - Catalogação na Publicação

Lameira, Vivian Meurer

PROPOSTA DE ANÁLISE LINGUÍSTICA PARA A IDENTIFICAÇÃO DE FRASEOLOGIAS DE GÊNERO E FRASEOLOGIAS DE GÊNERO DA ÁREA EXTRAÍDAS DE UM CORPUS DE ABSTRACTS DA MEDICINA / Vivian Meurer Lameira. -- 2017.

48 f.

Orientador: Ana Eliza Pereira Bocorny.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Letras, Licenciatura em Letras: Língua Portuguesa e Literaturas de Língua Portuguesa, Língua Inglesa e Literaturas de Língua Inglesa, Porto Alegre, BR-RS, 2017.

1. Linguística de Corpus. 2. Inglês para fins acadêmicos. 3. Abstracts. 4. Fraseologia de Gênero.

5. Fraseologia de Gênero da Área. I. Bocorny, Ana

Eliza Pereira, Orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

VIVIAN MEURER LAMEIRA

**PROPOSTA DE ANÁLISE LINGUÍSTICA PARA A IDENTIFICAÇÃO DE
FRASEOLOGIAS DE GÊNERO E FRASEOLOGIAS DE GÊNERO DA ÁREA
EXTRAÍDAS DE UM CORPUS DE *ABSTRACTS* DA MEDICINA**

Monografia apresentada como requisito parcial
para obtenção do título de Licenciada em
Letras pelo Instituto de Letras da Universidade
Federal do Rio Grande do Sul

Porto Alegre, 15 de dezembro de 2017.

BANCA EXAMINADORA

Profª. Dra. Ana Eliza Pereira Bocorny

Doutora em Letras

UFRGS

Ana Paula Seixas Vial

Licenciada em Letras

UFRGS

MSc. Denise von der Heyde Lamberts

Mestra em Letras

Unipampa

AGRADECIMENTOS

Aos meus pais pelo apoio e compreensão durante minha trajetória na faculdade.

À Sandra Lucia Dalcin, por todos os cuidados que tem comigo.

À minha tia Rosalva, por me acolher no seu lar.

À minha madrinha Nísia Sost, por cuidar de mim como se eu fosse sua filha.

Ao meu namorado, por possibilitar eu trabalhar todas as tardes nas escolas de Salvador do Sul e, ao mesmo tempo, estar presente na UFRGS pelas manhãs.

À minha orientadora, Ana Eliza Pereira Bocorny, por me iniciar no mundo da pesquisa e por me acompanhar durante a minha trajetória acadêmica.

Aos amigos que conheci na UFRGS e aos que estiverem do meu lado durante a graduação.

Por último, não menos importante, à UFRGS pela oportunidade de ter aulas com professores que possuem um conhecimento admirável e por me apresentar um mundo de lutas e de persistência.

RESUMO

Abstracts ou resumos acadêmicos são a porta de entrada para trabalhos científicos. Eles influenciam a decisão de ler, ou não, os textos completos que os sucedem. A linguagem é um dos fatores que influenciam tal escolha. O uso de expressões multipalavras típicas do gênero é importante para que a produção linguística dos *abstracts* seja bem-sucedida. Neste contexto, a finalidade deste trabalho é apresentar uma proposta de análise linguística que possibilite a identificação, descrição e análise das unidades ou expressões multipalavras típicas do gênero, chamadas aqui de fraseologia de gênero e fraseologia de gênero da área. A metodologia utilizada apresentou as seguintes etapas: (i) escolha da área de conhecimento a ser observada; (ii) identificação de um periódico de alto fator de impacto da área; (iii) seleção dos *abstracts* estruturados; (iv) observação de exemplos prototípicos; (v) compilação do *corpus* de estudo; (vi) escolha da ferramenta de análise; (vii) extração das fraseologias de gênero e das fraseologias de gênero da área mais frequentes do *corpus*; (viii) identificação dos contextos da fraseologia (de gênero e de gênero da área) mais frequente do movimento retórico, e, por último, (ix) análise dos dados. A partir dos resultados obtidos, concluiu-se que as fraseologias de gênero e de gênero da área, juntamente com as suas coocorrências, não apresentam um padrão léxico-gramatical único. Contudo, a identificação de estruturas frasais usadas por escritores especializados da área podem auxiliar escritores não especializados a escrever de forma mais idiomática.

Palavras-chave: linguística de *corpus*. Inglês para fins acadêmicos. Abstract. Fraseologia de gênero. Fraseologia de gênero da área. Medicina.

ABSTRACT

Abstracts or academic summary are the gateway to scientific papers. It influences the decision to read, or not, the full texts that follow its. Language is one of the factors that influence such choice. The use of multi-word expressions typical of the genre is important for the linguistic production of abstracts be successful. In this context, the purpose of this work is to present a proposal of linguistic analysis that allows the identification, description and analysis of the units or multi-word expressions typical of the genre, here called gender phraseology and gender phraseology of the area. The methodology used presented the following steps: (i) choice of the area of knowledge to be observed; (ii) identification of a high impact factor journal of the area; (iii) selection of structured abstracts; (iv) observation of prototypical examples; (v) compilation of the study corpus; (vi) choice of the analysis tool; (vii) extraction of the most frequent phraseologies of the genre and of the genre phraseologies of the area of the corpus; (viii) identification of the most frequent contexts of the phraseology (genre and genre of the area) of the rhetorical movement and lastly (ix) data analysis. From the results obtained, it was concluded that the genre and genre phraseologies of the area, together with its co-occurrences, do not present a unique lexical-grammatical pattern. However, the identification of phrasal structures used by specialized writers of the area could help non-specialized writers to write more idiomatically.

Key-words: corpus linguistics. English for academic purposes. Abstract. Gender phraseologies. Area gender phraseologies. Medicine.

LISTA DE FIGURAS

Figura 1 - Texto de número 1 (WRIGHT, 2014)	19
Figura 2 – Elementos que tecem o texto.....	23
Figura 3 – Informações do CAE-NEJM.....	24
Figura 4 – Representação dos <i>sub-corpora</i>	25
Figura 5 – Seleção de n-gramas de três a quatro elementos	26
Figura 6 – Frequência das FG e das FGA	27
Figura 7 – Coocorrências anteriores e posteriores da fraseologia de gênero <i>the use of</i>	29
Figura 8 – Coocorrências anteriores e posteriores da fraseologia de gênero <i>the primary outcome was</i>	31
Figura 9 – Coocorrências anteriores e posteriores da fraseologia de gênero <i>as compared with</i>	32
Figura 10 – Coocorrências anteriores e posteriores da fraseologia de gênero <i>was associated with</i>	34
Figura 11 – Coocorrências anteriores e posteriores da fraseologia de gênero da área <i>the treatment of</i>	35
Figura 12 – Coocorrências anteriores e posteriores da fraseologia de gênero da área <i>patients were randomly assigned</i>	37
Figura 13 – Coocorrências anteriores e posteriores da fraseologia de gênero da área <i>among patients with</i>	38
Figura 14 – Coocorrências anteriores e posteriores da FGA <i>in patients with</i>	40

LISTA DE TABELAS

Tabela 1 - Modelo CARS para introduções de artigos de pesquisa	14
Tabela 2 – Lista de nove trigramas e quadrigramas mais frequentes nos <i>sub-corpora</i> estudados	26
Tabela 3 – FG e as FGA mais frequentes nos <i>corpora</i> analisados.....	27
Tabela 4 – Cinco coocorrências anteriores e posteriores da FG <i>the use of</i>	30
Tabela 5 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero <i>the primary outcome was</i>	31
Tabela 6 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero <i>as compared with</i>	33
Tabela 7 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero <i>was associated with</i>	34
Tabela 8 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero da área <i>the treatment of</i>	36
Tabela 9 – Cinco coocorrências anteriores e posteriores da FGA <i>patients were randomly assigned</i>	38
Tabela 10 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero da área <i>among patients with</i>	39
Tabela 11 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero de área <i>in patients with</i>	40
Tabela 12 – Tabela Integradora	42

LISTA DE ABREVIATURAS E SIGLAS

CAE-NEJM	<i>Corpus de abstracts estruturados da NEJM</i>
CARS	<i>Create a Research Space</i>
CRE-NEJM-B	<i>Corpus de Resumos Estruturados da NEJM - Background</i>
CRE-NEJM-C	<i>Corpus de Resumos Estruturados da NEJM - Conclusions</i>
CRE-NEJM-M	<i>Corpus de Resumos Estruturados da NEJM - Methods</i>
CRE-NEJM-R	<i>Corpus de Resumos Estruturados da NEJM - Results</i>
FG	Fraseologias de gênero
FGA	Fraseologias de gênero da área
NEJM	<i>New England Journal of Medicine</i>

SUMÁRIO

1. INTRODUÇÃO	11
2. REFERENCIAL TEÓRICO	13
2.1 Os gêneros	13
2.1.1 Conceção de gênero segundo Bakhtin	13
2.1.2 Conceção de gênero segundo John Swales	14
2.1.3 Conceção de gênero segundo Bhatia.....	15
2.1.4 Conceção de gênero neste trabalho	15
2.2 O gênero abstract	15
2.3 Linguística de <i>Corpus</i>	16
2.3.1 Especificidade do <i>Corpus</i> e representatividade.....	17
2.3.2 Língua como um sistema probabilístico.....	17
2.4 Expressões multipalavra em textos especializados	18
3 METODOLOGIA	22
3.1 Escolha da área de conhecimento	22
3.2 Preferência do periódico	22
3.3 Seleção dos <i>abstracts</i>	22
3.4 Análises de exemplos prototípicos	23
3.5 Compilação do <i>corpus</i>	24
3.6 Ferramenta de análise	24
3.7 Extração das fraseologias de gênero e das fraseologias de gênero da área	25
3.8 Identificações das coocorrências das fraseologias extraídas	28
3.9 Análise dos dados	28
4. ANÁLISE DE DADOS	29
4.1 FG	29
4.1.1 <i>Background</i>	29
4.1.2 <i>Methods</i>	30
4.1.3 <i>Results</i>	32
4.1.4 <i>Conclusions</i>	33
4.2 FGA	35
4.2.1 <i>Background</i>	35
4.2.2 <i>Methods</i>	37

4.2.3 <i>Results</i>	38
4.2.4 <i>Conclusions</i>	39
5. CONSIDERAÇÕES FINAIS	44
REFERÊNCIAS	46

1. INTRODUÇÃO

Abstracts ou resumos acadêmicos são a porta de entrada de vários gêneros acadêmicos (e.g. artigos, monografias, dissertações e teses). Eles influenciam a decisão de ler, ou não, os textos completos que os sucedem. A linguagem é um dos fatores que influenciam tal decisão. O leitor especializado pode deixar de ler um artigo por não reconhecer no *abstract* construções e combinações que lhe pareçam típicas e naturais da sua área de especialidade. A decisão da comunidade acadêmica de ler ou não um artigo pode afetar o impacto que o mesmo terá em estatísticas de produtividade acadêmica.

A identificação de padrões lexicais, entendidos aqui como combinações de palavras que apresentam certa estabilidade e frequência (HYLAND, 2012), que sejam típicos dos gêneros acadêmicos e das áreas de especialidade, passa a ter grande importância pedagógica neste contexto e no âmbito do ensino de linguagens de especialidade. Para ensinar escritores menos proficientes a produzir textos conforme os padrões estabelecidos pelos especialistas das áreas é preciso identificar os padrões recorrentes e, além disso, disponibilizar tais informações de forma clara, intuitiva e amigável.

Os textos especializados comunicam informações específicas por meio de elementos linguísticos (e.g. termos, unidades terminológicas e fraseologias de gênero) e não linguísticos (e.g. gráficos e tabelas). Pesquisas recentes enfatizam a importância de observar o vocabulário não como palavras individuais, mas como unidades poliléxicas (HYLAND, 2012), aqui denominadas expressões multipalavras. Os padrões que estas unidades estabelecem entre si e entre outros elementos do texto se constituem como fios que se entrelaçam em uma trama de sentido. Este trabalho trata da observação de um dos elementos desta trama: os elementos linguísticos multipalavra típicos do gênero *abstract*. Tais unidades são observadas aqui apenas em *abstracts* estruturados da área médica, mais especificamente do periódico *New England Journal of Medicine* (NEJM). A intenção que se tem ao realizar este estudo é a de apresentar uma proposta de análise linguística que possibilite a identificação, descrição e análise das unidades ou expressões multipalavras típicas do gênero, chamadas aqui de fraseologias de gênero (FG) extraídas a partir de um *corpus* de *abstracts*. Ao identificar e descrever tais expressões, pretende-se que estas informações auxiliem na posterior construção de tarefas e recursos pedagógicos que ajudem leitores menos especializados a escreverem de

¹ Neste trabalho, o termo *abstract* será usado para evitar a ambiguidade que o termo resumo acadêmico pode gerar (FREITAS, 2016)

forma mais idiomática. Da mesma forma, considera-se que esta proposta de análise linguística possa ser utilizada para outros gêneros (e.g. artigos e resenhas) e para outras áreas de especialidade (e.g. matemática e geografia).

A motivação para a realização desta pesquisa parte do contexto dentro do qual a mesma se insere, ou seja, um contexto de democratização do acesso às universidades brasileiras, ocorrido nos últimos anos, em nosso país. Os novos perfis de estudantes, professores e pesquisadores, aliado a um crescente processo de internacionalização do ensino superior, impõe a necessidade do desenvolvimento de competências relacionadas à produção e compreensão de diferentes gêneros acadêmicos em inglês, sendo o *abstract* um deles.

A principal iniciativa implementada pelo Governo Federal para alavancar o processo de internacionalização das universidades federais brasileiras foi a criação, em 2012, do Programa Inglês sem Fronteiras² com o objetivo de “propiciar a formação e capacitação de alunos de graduação das instituições de educação superior para os exames linguísticos exigidos para o ingresso nas universidades anglófonas” (BRASIL, 2012). A importância de ampliar o programa para toda a comunidade acadêmica (alunos de graduação e pós-graduação, servidores técnicos e docentes) e de incluir outros idiomas além do inglês, levou à criação do Programa Idiomas sem Fronteiras (IsF)³, em 2014. No momento atual, o programa tem o propósito de “incentivar o aprendizado de línguas, além de propiciar uma mudança abrangente e estruturante no ensino de idiomas estrangeiros nas universidades do País” (BRASIL, 2014). Desta forma, o Programa além de oportunizar o estudo de diferentes idiomas, tem por finalidade promover a internacionalização nas universidades brasileiras.

Para organizar esta proposta, o presente trabalho é composto por cinco seções. Primeiramente, são apresentados os referenciais teórico-metodológicos que embasaram esta pesquisa. Em seguida, relata-se a metodologia utilizada durante o processo de análise. Posteriormente, descrevem-se os dados gerados e a análise realizada. Por último, discutem-se os resultados obtidos e propõem-se parâmetros para o desenvolvimento de análise linguística que poderá ser utilizada para a descrição da língua de outros gêneros acadêmicos e de outras áreas de especialidade.

² Criado através da Portaria Nr. 1.466 do Ministério da Educação, em 18/12/2012.

³ Criado através da Portaria Nr. 973 do Ministério da Educação, em 14/11/2014.

2. REFERENCIAL TEÓRICO

O resultado prático que se pretende atingir com este estudo deriva do encontro e do entrelaçamento de pressupostos teóricos oriundos de diferentes áreas do conhecimento: (i) os estudos sobre gêneros textuais; (ii) os princípios da linguística de *corpus* e (iii) os pressupostos relacionados às expressões multpalavras em textos especializados.

2.1 Os gêneros

Ao investigar a noção de gêneros, encontra-se na literatura uma diversidade denominativa e conceitual. Alguns autores, como Bakhtin (1999), utilizam o termo “gêneros do discurso” e outros autores, como Marcuschi (2003), o termo “gêneros textuais”. Nesta pesquisa, não se fará distinção entre ambos. Quanto à heterogeneidade de definições de gênero, destacam-se brevemente apenas as concepções que são mais relevantes para este estudo, segundo Bakhtin (1992, 1999), Swales (1990) e Bathia (2001). Antes de prosseguir, é necessário ressaltar que este trabalho tem em vista os gêneros secundários, aqueles relativos a situações de comunicação mais complexas, como a escrita (BAKHTIN, 1999). Desta forma, não serão abordados os gêneros primários que são relativos à comunicação verbal (BAKHTIN, 1999).

2.1.1 Concepção de gênero segundo Bakhtin

Bakhtin (1992) define os gêneros do discurso como “tipos relativamente estáveis de enunciados”. Para ele, os gêneros textuais são relativos aos diversos textos que conseguimos identificar pelo fato de eles se repetirem de um modo razoavelmente fixo em determinados contextos (BAKHTIN, 1999). Desta forma, os gêneros possuem certas características estáveis relativas à sua macroestrutura, às escolhas linguísticas e ao público a quem eles se destinam. Foi Bakhtin que passou a entender os gêneros como entidades sócio-discursivas que são criados em distintas esferas da atividade humana, refletindo, desta forma, as condições específicas de cada uma.

2.1.2 Concepção de gênero segundo John Swales

Swales (1990) aponta que “os gêneros são veículos de comunicação para atingir um objetivo” (SWALES, 1990 apud FREITAS, 2016). Tal afirmação é relevante para este trabalho na medida em que o gênero *abstract* serve como porta de entrada (VAN LEEUWEN, 1993) para introduzir o texto que virá em seguida. Um dos traços principais da definição de gênero de acordo com Swales (1990) é que esse possui um propósito comunicativo compartilhado por membros de uma determinada comunidade discursiva. No contexto deste trabalho, a comunidade acadêmica da área médica é a comunidade discursiva na qual o texto alvo é produzido, sendo que o propósito comunicativo do gênero *abstract*, a ser descrito, é de introduzir o texto que o sucede, resumizando o artigo.

Dentre as contribuições de Swales, uma das mais relevantes para este trabalho é a sua investigação sobre a macroestrutura da introdução do gênero artigo acadêmico através da análise dos padrões organizacionais e retóricos, chamados por ele de movimentos. Ao fazer isso, ele desenvolve um modelo analítico para descrever a parte do gênero analisado, denominado *Create a Research Space* (CARS), detalhado na Tabela 1.

Tabela 1 - Modelo CARS para introduções de artigos de pesquisa.

Número	Descrição	Passos
1	Estabelecer o território	Passo 1- Estabelecer a importância da pesquisa e/ou Passo 2- Fazer generalização/ões quanto ao tópico e/ou Passo 3- Revisar a literatura (pesquisa prévia)
2	Estabelecer o nicho	Passo 1A- Contra argumentar ou Passo 1B- Indicar lacuna/s no conhecimento ou Passo 1C- Provocar questionamento ou Passo 1D- Continuar a tradição
3	Ocupar nicho	Passo 1A- Delinear os objetivos ou Passo 1B- Apresentar a pesquisa Passo 2- Apresentar os principais resultados Passo 3- Indicar a estrutura do artigo

Fonte: SWALES, 1990 apud BIASI-RODRIGUES; ARAÚJO; SOUSA, 2009.

Na Tabela 1, é possível identificar os movimentos “estabelecer o território”, “estabelecer o nicho” e “ocupar o nicho”. A noção de movimentos, denominado neste

trabalho como movimentos retóricos, passa a ser importante na medida em que os *abstracts* da NEJM, como descrito adiante, são estruturados, possuindo quatro movimentos retóricos: *background, methods, results e conclusions*.

2.1.3 Concepção de gênero segundo Bhatia

Bhatia (2001) baseou-se no conceito de gêneros de Swales para criar a sua definição. Para ele, gênero é o conjunto de propósitos comunicativos compartilhados. Por consequência, se os propósitos comunicativos mudam, pode haver uma mudança no gênero. Uma das contribuições mais importantes de Bhatia para a proposta deste trabalho é a sua definição de análise de gênero como sendo “o estudo do comportamento linguístico situado em contextos acadêmicos ou profissionais” (BHATIA, 2001) e também a sua proposta de organização do gênero *abstract* (BHATIA, 2001) que muito se assemelha à organização retórica dos *abstracts* estruturados da NEJM, pois possui quatro movimentos retóricos com as mesmas funções comunicativas. A primeira tem o propósito de introduzir o objetivo, a segunda de descrever a metodologia e, por fim, a terceira sumariza os resultados e a quarta apresenta as conclusões.

2.1.4 Concepção de gênero neste trabalho

A concepção de gênero usada nesta pesquisa está alinhada com as três perspectivas dos autores apresentados. De Bakhtin (1992) ressalta-se a sua concepção de gênero do discurso como “tipos relativamente estáveis de enunciados”. De Swales (1990) a ideia de que os gêneros “são veículos de comunicação para atingir um objetivo”. Por fim, de Bathia (1993) o entendimento de análise de gênero como sendo “o estudo do comportamento linguístico situado em contextos acadêmicos ou profissionais”.

2.2 O gênero *abstract*

Com o crescimento das publicações especializadas na segunda metade do século XX, o gênero *abstract* surgiu como uma forma de possibilitar o acesso, de forma sintetizada, às diversas produções acadêmicas (FREITAS, 2016). Segundo Zanella e Heberle (2005) “os

abstracts são resumos escritos na língua inglesa que trazem em seu conteúdo as ideias principais de um trabalho previamente elaborado em um único parágrafo.” Desta forma, eles servem como uma porta de entrada (VAN LEEUWEN, 1993) para os trabalhos científicos, levando o leitor a ler, ou não, o texto completo.

No entanto, a estrutura dos *abstracts* nem sempre é apresentada em um parágrafo único. Alguns, chamados de *abstracts* estruturados, podem ser divididos em parágrafos nomeados de acordo com os movimentos retóricos contidos no texto. Os *abstracts* usados na compilação do *corpus* deste estudo, como mencionado anteriormente, são estruturados e possuem quatro movimentos retóricos: *background*, *methods*, *results* e *conclusions*.

Neste trabalho, portanto, o gênero *abstract* é entendido como uma estrutura que tem nos movimentos retóricos e nas FG importantes elementos constituintes (CORTES, 2013; BIBER, 2002 apud SARDINHA, 2004; HYLAND, 2008) de um discurso coerente e característico do uso de determinadas comunidades em certos contextos especializados. Os movimentos retóricos estão relacionados à macroestrutura do texto, enquanto as fraseologias são como tijolos compondo os blocos macroestruturais e estabelecendo relações estáveis e recorrentes com outros elementos do texto em padrões lexicais.

2.3 Linguística de *Corpus*

A Linguística de *Corpus* parte de uma perspectiva de descrição da língua, seja ela geral ou especializada. Os estudos fundamentados em *corpora* existem desde a antiguidade clássica. “Na Grécia antiga, Alexandre, o Grande, definiu o *Corpus* Helenístico. Na antiguidade e na idade média, produziram-se *corpora* de citações da Bíblia.” (SARDINHA, 2004). Contudo, foi só no século XX que a pesquisa baseada em *corpus* tomou proporções maiores (MCENERY; WILSON, 2010) na medida em que vários estudos passaram a descrever a língua pelo viés de *corpora* (SARDINHA, 2004).

Inicialmente, os *corpora* eram coletados e analisados manualmente, armazenados muitas vezes através de fichas de papel. Foi através de um *corpus* manual, chamado *Survey of English Usage*, que muitos pesquisadores se inspiraram para criar os *corpora* atuais (SARDINHA, 2004). Hoje, a Linguística de *Corpus* está relacionada à tecnologia, que possibilita o armazenamento dos textos e o estudo dos mesmos através de ferramentas computacionais desenvolvidas para a análise de *corpora*. Neste contexto, este estudo considera *corpus* um:

[...] conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (SANCHEZ & CANTOS, 1997 apud SARDINHA, 2004).

Há questionamentos sobre o estatuto da Linguística de *Corpus* ser disciplina ou metodologia (SARDINHA, 2004). Baseando-se na definição de Linguística de *Corpus* como não sendo “somente uma nova metodologia emergente para o estudo da linguagem, mas uma nova empreitada de pesquisa e, na verdade, uma nova abordagem filosófica” (LEECH, 145, p. 106 apud SARDINHA, 2004, p. 37), este trabalho dá preferência ao termo *abordagem baseada em corpus* que foi adotado por Biber (2002 apud SARDINHA, 2004), considerando que se busca a descrição linguística do gênero *abstract* baseada em *corpus*.

2.3.1 Especificidade do *Corpus* e representatividade

Levando em consideração que é improvável compilar um *corpus* contendo toda a língua⁴ para obter a representatividade total do mesmo, Sardinha (2004) sugere restringir o conteúdo e:

[...] delimitar ao máximo a variedade (tipo de texto, por exemplo) incluída no corpus. Uma variedade específica da linguagem demonstra maior padronização e conseqüentemente menor variação no nível do léxico, da gramática, do discurso. Ou seja, apresenta maior grau de fechamento (SARDINHA, 2004, p. 28).

Ao construir um *corpus* específico de um determinado gênero, conteúdo, autor ou periódico, o mesmo passa a ser mais significativo para atingir propósitos particulares (SARDINHA, 2004).

2.3.2 Língua como um sistema probabilístico

A visão da língua como um sistema probabilístico é um dos fundamentos principais da Linguística de *Corpus* (SARDINHA, 2004). Assim, os traços linguísticos não ocorrem todos com a mesma regularidade (SARDINHA, 2004). Por esse motivo, a variação dos traços não é aleatória, pelo contrário, existe “um mapeamento regular entre a frequência maior ou

⁴ Neste trabalho, língua é entendida como língua geral, enquanto a linguagem é entendida como linguagem de especialidade.

menor de um traço e um contexto de ocorrência” (SARDINHA, 2004). Desta forma, falar que os traços não são aleatórios significa dizer que:

[...] a linguagem é padronizada (patterned). A padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente mostra sinais de ser, na verdade, um padrão lexical ou léxico-gramatical (SARDINHA, 2004, p. 31).

A característica da língua como sendo padronizada faz com que seja possível analisar as estruturas fixas que ocorrem em diferentes gêneros. Tais estruturas possuem diversas denominações na literatura. Neste trabalho, trata-se das FG.

2.4 Expressões multipalavra em textos especializados

A comunicação acontece por meio de elementos linguísticos e não linguísticos. Os elementos linguísticos, que usam a palavra como principal meio de comunicação, constituem textos de língua geral que podem ser orais ou escritos. Os elementos não linguísticos, por sua vez, também permitem que a comunicação aconteça, mas por meio de recursos não verbais como imagens, fórmulas, números, gestos e expressões faciais. Estes elementos, combinados, “nos permitem expressar nossos sentimentos, pensamentos, ideias e experiências” (THOMSON, 2008).

Em textos especializados, ou seja, aqueles utilizados em contextos de áreas de especialidade por uma comunidade de sujeitos que compartilham conhecimentos da área em diferentes níveis, também há a convivência de elementos linguísticos e não linguísticos na concretização da comunicação especializada. A observação da Figura 1, um *abstract* estruturado da área médica, nos permite ver estas duas grandes categorias de elementos constituintes do texto. Os elementos linguísticos podem ser exemplificados por itens lexicais monoléxicos (e.g. *between, with, and, was* e *were*) ou sintagmáticos (e.g. *a double-blind e multicenter clinical trial*). Os elementos não linguísticos podem ser exemplificados pelos números, (e.g. 35, 24.4, 75 e 0.95), percentuais (e.g. 73.7%, 15.2% e 95%) e fórmula (e.g. $P=0.35$).

Figura 1 - Texto de número 1 (WRIGHT, 2014).

Traumatic brain injury (TBI) is a major cause of death and disability worldwide. Progesterone has been shown to improve neurologic outcome in multiple experimental models and two early-phase trials involving patients with TBI.

METHODS

We conducted a double-blind, multicenter clinical trial in which patients with severe, moderate-to-severe, or moderate acute TBI (Glasgow Coma Scale score of 4 to 12, on a scale from 3 to 15, with lower scores indicating a lower level of consciousness) were randomly assigned to intravenous progesterone or placebo, with the study treatment initiated within 4 hours after injury and administered for a total of 96 hours. Efficacy was defined as an increase of 10 percentage points in the proportion of patients with a favorable outcome, as determined with the use of the stratified dichotomy of the Extended Glasgow Outcome Scale score at 6 months after injury. Secondary outcomes included mortality and the Disability Rating Scale score.

RESULTS

A total of 882 of the planned sample of 1140 patients underwent randomization before the trial was stopped for futility with respect to the primary outcome. The study groups were similar with regard to baseline characteristics; the median age of the patients were 35 years, 73.7% were men, 15.2% were black, and the mean Injury Severity Score was 24.4 (on a scale from 0 to 75, with higher scores indicating greater severity). The most frequent mechanism of injury was a motor vehicle accident. There was no significant difference between the progesterone group and the placebo group in the proportion of patients with a favorable outcome (relative benefit of progesterone, 0.95; 95% confidence interval [CI], 0.85 to 1.06; P=0.35). Phlebitis or thrombophlebitis was more frequent in the progesterone group than in the placebo group (relative risk, 3.03; CI, 1.96 to 4.66). There were no significant differences in the other prespecified safety outcomes.

CONCLUSIONS

This clinical trial did not show a benefit of progesterone over placebo in the improvement of outcomes in patients with acute TBI. (Funded by the National Institute of Neurological Disorders and Stroke and others; PROTECT III ClinicalTrials.gov number, NCT00822900.)

Legenda:

ELEMENTOS NÃO LINGUÍSTICOS

■ Numerais e Fórmulas

ELEMENTOS LINGUÍSTICOS

■ Palavras gramaticais fora de unidades multipalavras

■ Termos e unidades terminológicas

■ Fraseologia de gênero

■ Fraseologia de gênero da área

■ Verbos fora de unidades multipalavras

Nota-se no texto apresentado na Figura 1 que os elementos linguísticos podem aparecer isolados (e.g. verbo *were* em *the median age of the patients were 35 years, 73.7% were men, 15.2% were black* ou a preposição *or* em *Phlebitis or thrombophlebitis*) conectando outros elementos do texto ou agrupados em unidades multipalavras, numa configuração que, segundo Cortes (2013) e Swales (1990), constituem *building blocks* dos textos especializados. Quando agrupados em unidades multipalavras, tais elementos linguísticos são apresentados na literatura com diferentes nomes: (i) *multiword units*, *multiword expressions*, “*clusters*”, “*chunks*” ou “*bundles*”, em inglês (HYLAND, 2008), (ii) FG, fraseologias de gênero da área (FGA), unidades terminológicas, em português, (iii) *unidades de conocimiento especializado* (CABRÉ; BACH, 2005), em espanhol.

Algumas destas expressões ou unidades multipalavras estão mais vinculados a uma temática, apresentando, desta forma, mais valor terminológico (KILIAN & LOGUERCIO, 2015). Outras são mais típicas do gênero, apresentando, por isso, menor ou nenhum valor terminológico. É importante observar que este tipo de categorização não é absoluta, ou seja, é possível imaginar que as expressões multipalavras, nos textos especializados, possam localizar-se ao longo de um *continuum* (CIAPUSCIO, 1993) de mais especializada para menos ou não especializada, podendo haver, entre as categorias, certa opacidade⁵. Considerando a ideia de *continuum* e de opacidade, durante a categorização dos elementos linguísticos do texto apresentado na Figura 1, foi possível observar uma categoria diferente que diz respeito àqueles padrões típicos do gênero que são mais frequentes em certas áreas de especialidade. Ou seja, uma categoria que apresenta características das FG na medida em que são típicas do gênero, mas que também apresentam conteúdo terminológico, o que, muitas vezes, é evidenciado em um termo da área que faz parte da unidade.

Na constituição do texto, os elementos linguísticos sintagmáticos, referidos neste estudo de forma genérica como expressões ou unidades multipalavras, vão demonstrando preferências lexicais e gramaticais, desta forma, estabelecendo padrões de ocorrência e de coocorrência passíveis de serem observados por meio dos recursos disponibilizados por ferramentas para o estudo de *corpora*.

Neste trabalho, os padrões típicos nominais de uma área de especialidade, ou seja, aqueles que tem maior valor terminológico, serão chamados de unidades terminológicas (e.g. *the planned sample*, *the placebo group* e *clinical trial*). Aqueles típicos do gênero e, portanto,

⁵ Temmerman (2007) quando observa a terminologia de certas áreas de especialidade, afirma que “*Many categories are fuzzy and can not be absolutely classified by logical and ontological means*”. Neste trabalho, a ideia trazida pelo termo “*fuzzy*” foi traduzida como opacidade.

com menor valor terminológico, serão referidos como FG (e.g. *a total of, the primary outcome was e is associated with*). Por fim, os padrões típicos do gênero e mais frequentes em certas áreas de especialidade serão chamados de FGA (e.g. *outcomes in patients with, in the treatment of e patients were randomly assigned*).

Os objetos a serem observados neste trabalho são as FG e as FGA, tendo como objetivo identificar as mais frequentes por movimento retórico nos *abstracts* da revista NEJM. A partir desta identificação, as linhas de concordância serão analisadas com o objetivo de verificar a existência de padrões léxico-gramaticais.

3 METODOLOGIA

A partir dos elementos teóricos apresentados e tendo em vista o objetivo proposto, a metodologia utilizada neste trabalho apresentou as seguintes etapas: (i) escolha da área de conhecimento a ser observada; (ii) identificação de um periódico de alto fator de impacto da área; (iii) seleção dos *abstracts* estruturados; (iv) observação de exemplos prototípicos; (v) compilação do *corpus* de estudo; (vi) escolha da ferramenta de análise; (vii) extração das FG e FGA mais frequentes do *corpus*; (viii) identificação dos contextos da fraseologia (de gênero e de gênero da área) mais frequente do movimento retórico, e, por último, (ix) análise dos dados por meio da construção de tabelas com a inclusão de fórmulas e perguntas que representem os padrões identificados. A seguir, descreve-se com mais detalhes os estágios percorridos durante a pesquisa.

3.1 Escolha da área de conhecimento

Justifica-se a escolha da área médica por a mesma estar bastante alinhada com práticas acadêmicas de âmbito internacional. Portanto, esta possui uma quantidade grande de publicações acadêmicas em inglês redigidas por estudantes e professores brasileiros. Contudo, pretende-se ampliar este estudo para outras áreas de especialidade no futuro.

3.2 Preferência do periódico

O periódico NEJM foi escolhido por ser um dos maiores fatores de impacto na área, de acordo com *Scimago Journal & Country Rank*, representando, por esta razão, um padrão reconhecido pela comunidade de especialistas como referência.

3.3 Seleção dos *abstracts*

Foram selecionados duzentos *abstracts* recentes (2014-2017) do periódico NEJM, redigidos na língua inglesa. Levando-se em conta que a identificação dos movimentos retóricos não é uma tarefa fácil e que a mesma pode resultar em diferentes interpretações

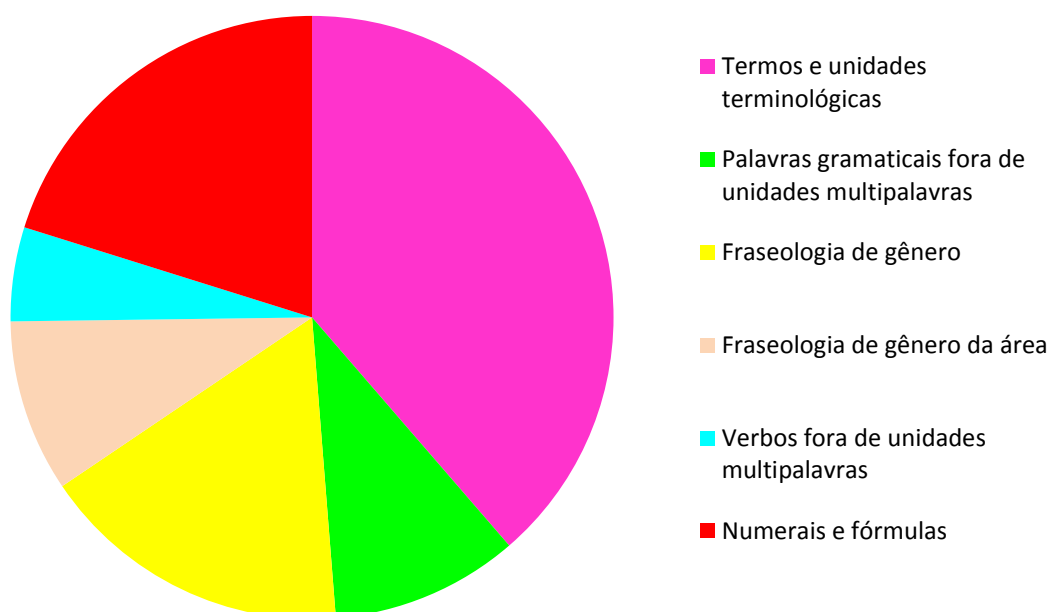
(HYLAND, 2008), optou-se por trabalhar somente com *abstracts* estruturados, pois os movimentos retóricos já foram estabelecidos através de subtítulos nos *corpus* dos mesmos.

3.4 Análises de exemplos prototípicos

Uma análise qualitativa de dez *abstracts* pertencentes ao *corpus* de análise foi feita com o objetivo de identificar os elementos linguísticos e não linguísticos que constituem os textos e a frequência com que eles aparecem. Durante a exploração dos mesmos, mapearam-se os elementos linguísticos e não linguísticos encontrados de diferentes cores, como visto na Figura 1, apresentada anteriormente.

Em seguida, criou-se um gráfico para mostrar visualmente a frequência dos elementos que tecem o texto. Na Figura 2 observa-se que os resumos analisados são escritos, quase que exclusivamente, com a utilização de elementos linguísticos, ou seja, termos e unidades terminológicas, seguido de elementos não linguísticos como os numerais e fórmulas e das FG. Com menos frequência, encontram-se as palavras gramaticais fora de unidades multipalavras, as FGA e verbos fora de unidades multipalavras.

Figura 2 – Elementos que tecem o texto.



Como mencionado anteriormente, ao observar as expressões multipalavras típicas de gênero em estudo, percebeu-se a existência de duas categorias distintas: as FG e as FGA. Enquanto as primeiras são típicas do gênero *abstract*, podendo ser usadas em *abstracts* de qualquer área de conhecimento, as FGA apresentam conteúdo terminológico, podendo, desta forma, ser consideradas típicas da área de especialidade em questão, como afirmam Kilian e Louguercio (2015). A partir deste momento, são estabelecidas duas categorias que são os objetos de análise deste estudo. Inicialmente, tinha-se como objeto observar apenas as FG.

3.5 Compilação do *corpus*

O *corpus* deste estudo, denominado *Corpus de abstracts* estruturados da NEJM (CAE-NEJM), consiste em duzentos resumos acadêmicos estruturados da área médica redigidos em inglês e extraídos do periódico NEJM, publicados entre os anos de 2014-2017. O *corpus* possui 55,832 palavras (*words*), 2,442 sentenças (*sentences*) e os textos estão na língua inglesa, como ilustra a Figura 3.

Figura 3 – Informações do CAE-NEJM.

CAE-NEJM ?

Counts		General info	
Tokens	75,673	Language	English
Words	55,832	Encoding	UTF-8
Sentences	2,442	Compiled	10/26/2017 01:08:38

3.6 Ferramenta de análise

A partir da definição dos objetos de análise e da compilação dos *corpora*, iniciou-se a fase mais quantitativa do estudo com o uso da ferramenta *Sketch Engine*. Esta foi escolhida em função do acesso ser online e gratuito (por um mês) e também por possuir vários tutoriais na internet que facilitam a sua utilização. Outra vantagem é que esta ferramenta armazena o

corpus enquanto a conta do usuário está ativa, não havendo necessidade de carregar os arquivos mais de uma vez. Dois recursos desta ferramenta foram usados: (i) o n-gramas para a identificação das FG e FGA mais frequentes de cada movimento retórico e (ii) o concordanciador para a extração das coocorrências das FG e da FGA.

3.7 Extração das fraseologias de gênero e das fraseologias de gênero da área

Para melhor observar as FG e as FGA por movimento retórico, decidiu-se compilar quatro *sub-corpora* de análise. Cada um foi composto por um movimento retórico diferente pertencente aos *abstracts* do CAE-NEJM. Desta forma, foram criados *sub-corpora* do *background*, do *methods*, do *results* e da *conclusions* que foram denominados sucessivamente de: *Corpus* de Resumos Estruturados da *NEJM - Background* (CRE-NEJM-B), *Corpus* de Resumos Estruturados da *NEJM - Methods* (CRE-NEJM-M), *Corpus* de Resumos Estruturados da *NEJM - Results* (CRE-NEJM-R) e *Corpus* de Resumos Estruturados da *NEJM - Conclusions* (CRE-NEJM-C). Após, os *sub-corpora* foram carregados separadamente no programa *Sketch Engine* (Figura 4) para que fosse possível analisar as FG e as FGA típicas de cada movimento.

Figura 4 – Representação dos *sub-corpora*.

Language	Name	Words
English	CAE-NEJM	55,832
English	CAE-NEJM-B	8,141
English	CAE-NEJM-C	6,962
English	CAE-NEJM-M	15,204
English	CAE-NEJM-R	23,730

Em seguida gerou-se uma lista de palavras (*word list*) para cada *sub-corpora*, selecionando trigramas e quadrigramas com frequência mínima de cinco (Figura 5).

Figura 5 – Seleção de n-gramas de três a quatro elementos.

Word list options ?

Subcorpus: [create new](#) ?

Search attribute:

use n-grams. Value of n: from to ?

hide/nest sub-n-grams

Filter options:

Filter word list by: Regular expression:

Minimum frequency:

Através do recurso n-gramas, identificaram-se as FG e as FGA. Alguns dos elementos identificados não foram considerados relevantes para o trabalho, sendo, por isso, descartados (ver Tabela 2).

Tabela 2 – Lista de nove trigramas e quadrigramas mais frequentes nos *sub-corpora* estudados.

CRE-NEJM-B	CRE-NEJM-M	CRE-NEJM-R	CRE-NEJM-C
IN PATIENTS WITH	THE PRIMARY OUTCOME WAS	AS COMPARED WITH	IN PATIENTS WITH
THE USE OF	WE RANDOMLY ASSIGNED	AMONG PATIENTS WITH	WAS ASSOCIATED WITH
THE TREATMENT OF	PRIMARY END POINT WAS	A MEDIAN FOLLOW-UP	AS COMPARED WITH
IS ASSOCIATED WITH	WE CONDUCTED A	GROUP THAN IN	AMONG PATIENTS WITH
EFFICACY AND SAFETY	THE USE OF	HAZARD RATIO FOR DEATH	THE RISK OF
WE HYPOTHESIZED THAT	WITH THE USE OF	MONTHS IN THE	DID NOT IMPROVE
BEEN SHOWN TO	KILOGRAM OF BODY	PROGRESSION-FREE SURVIVAL WAS	YEARS OF AGE
HUMAN IMMUNODEFICIENCY VIRUS	MG TWICE DAILY	MEDIAN PROGRESSION-FREE SURVIVAL	OF AGE OR

IN A PHASE	WAS A COMPOSITE OF	WAS	A MEDIAN FOLLOW-UP RISK OF DEATH
-------------------	---------------------------	------------	---

■ Fraseologias de gênero
■ Elementos descartados
■ Fraseologias de gênero da área

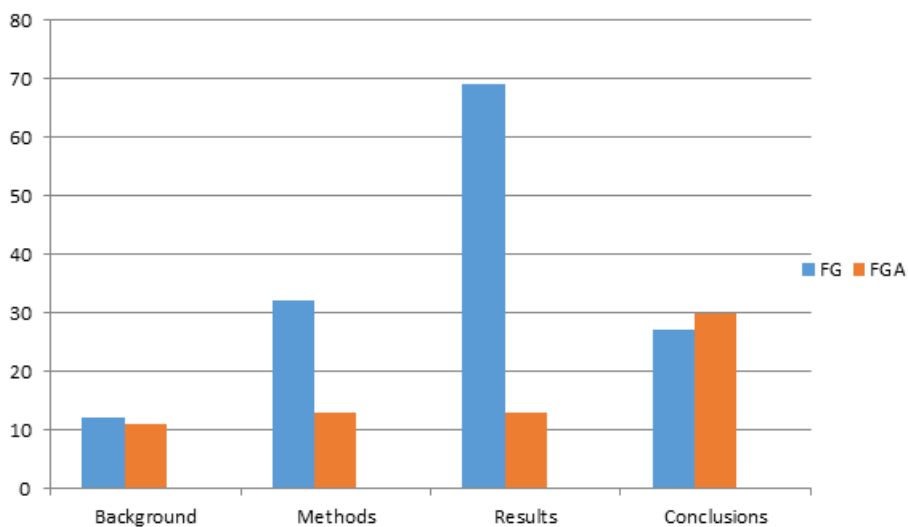
Por fim, foram extraídas as FG e as FGA mais frequentes de cada *sub-corpora* (Tabela 3) e criou-se um gráfico (Figura 6) para ilustrar a diferença das frequências das FG e FGA em cada movimento retórico.

Tabela 3 – FG e as FGA mais frequentes nos *sub-corpora* analisados.

	BACKGROUND	METHODS	RESULTS	CONCLUSIONS
Fraseologias de gênero	<i>the use of</i>	<i>the primary outcome was</i>	<i>as compared with</i>	<i>was associated with</i>
Fraseologias de gênero da área	<i>the treatment of</i>	<i>patients were randomly assigned</i>	<i>among patients with</i>	<i>in patients with</i>

Na Figura 6 é possível observar que a frequência das FG é maior nos movimentos retóricos *background*, *methods* e *results*. As FGA, por outro lado, são mais frequentes no movimento retórico *conclusions*.

Figura 6 – Frequência das FG e das FGA.



3.8 Identificações das coocorrências das fraseologias extraídas

As coocorrências das fraseologias mais frequentes foram identificadas através do modo de pesquisa *concordance*. Esta opção apresenta o contexto em que está inserida cada fraseologia e se refere ao texto que vem à direita e à esquerda das mesmas quando o texto é lido horizontalmente. Para cada fraseologia, foram extraídas cinco coocorrências anteriores e posteriores, organizadas nas Tabelas 4 a 11.

3.9 Análise dos dados

Durante a análise das FG e das FGA e das suas coocorrências, foram elaboradas perguntas para cada movimento retórico com o intuito de instruir os leitores/escritores a refletir sobre os próprios trabalhos e, desta forma, facilitar a redação de seus *abstracts* em possíveis atividades pedagógicas a serem desenvolvidas posteriormente. Além disso, foram criadas tabelas e, a partir da análise dos contextos extraídos, foram construídas fórmulas que representassem os padrões apresentados nas coocorrências. Por último, todas as informações foram colocadas em um quadro integrador (Tabela 12) para sumarizar a presente análise.

4. ANÁLISE DE DADOS

Nesta seção, apresentam-se as tabelas geradas a partir das observações das FG e das FGA, juntamente com as coocorrências extraídas. Ao introduzi-las, são feitas considerações breves das mesmas. As perguntas e fórmulas elaboradas, como mencionado anteriormente, também estão inseridas nas tabelas apresentadas abaixo.

4.1 FG

A seguir, mostram-se os dados referentes aos movimentos retóricos *background*, *methods*, *results* e *conclusions* das FG *the use of*, *the primary outcome was*, *as compared with* e *was associated with*.

4.1.1 Background

Do movimento retórico *background* foram extraídas as linhas de concordância da fraseologia de gênero *the use of*, ocorrida 12x no CRE-NEJM-B, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 7.

Figura 7 – Coocorrências anteriores e posteriores da fraseologia de gênero *the use of*.

Query The, use, of 12 (1,267.29 per million) ⓘ	
file544370...	treatment with oral glucocorticoids despite the use of high-dose inhaled therapy. However, the
file544384...	transplantation. However, emerging data on the use of lenalidomide and low-dose dexamethasone
file544382...	disease (CAD) and are often evaluated with the use of diagnostic testing, although there are
file544370...	toxic effect of chemotherapy. Studies of the use of gonadotropin-releasing hormone (GnRH)
file544374...	has been confirmed radiographically and with the use of current laboratory diagnostic tests are
file544376...	confirmed radiographically and with the use of current laboratory diagnostic tests are
file544381...	ipilimumab-refractory metastatic melanoma. The use of nivolumab in previously untreated patients
file544382...	increases the risk of death and limits the use of inhibitors of the renin - angiotensin -
file544372...	CPR could be increased with the use of a mobile-phone positioning system that could
file544377...	activator (t-PA) alone. Thrombectomy with the use of a stent retriever, in addition to intravenous
file544383...	, we defined five glioma molecular groups with the use of three alterations: mutations in the TERT
file544367...	confirmed radiographically and with the use of current laboratory diagnostic tests are

A Tabela 4 mostra que, das cinco linhas de concordância analisadas, em quatro delas a fraseologia de gênero está no meio da frase. As coocorrências anteriores dividem-se em dois tipos de padrão: (i) Substantivo + Preposição ou (ii) verbo + preposição. Por outro lado, todas

as coocorrências posteriores são seguidas por um substantivo que pode estar modificado ou não.


Tabela 4 – Cinco coocorrências anteriores e posteriores da FG *the use of*.

Pergunta	Coocorrência anterior (Fórmula)	Fraseologia	Coocorrência posterior (Fórmula)
Quais são os dados obtidos até o momento no uso de determinada substância/diagnóstico/terapia?	<i>emerging data on...</i> (ADJ + SUBS + PREP)		<i>lenalidomide and low-dose dexamethasone are...</i> (SUBS+VERB)
O que é frequentemente avaliado com o uso de determinado diagnóstico/terapia/tratamento/cirurgia?	<i>...are often evaluated with</i> (VERB+ PREP)		<i>diagnostic testing...</i> (ADJ+SUBS)
O que mostram os estudos do uso de determinado remédio/terapia/tratamento?	<i>Studies of...</i> (SUBS+PREP)	<i>The/the use of</i>	<i>gonadotropin-releasing hormone (GnRH)...</i> (SUBS)
O que é possível descrever sobre o uso de determinado remédio/terapia/tratamento nos pacientes tratados/não tratados?	.		<i>nivolumab in previously untreated patients...</i> (SUBS+PREP+ADV/ADJ/SUBS)
O que pode aumentar/diminuir com o uso determinado sistema/aparelho/tratamento?	<i>CPR could be increased with</i> (SUBS+VERB+PREP)		<i>a mobile-phone positioning system...</i> (ART+SUBS+ADJ+SUBS)

4.1.2 Methods

Do movimento *methods*, foram extraídas as linhas de concordância da fraseologia de gênero *the primary outcome was*, ocorrida 32x no CRE-NEJM-M, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 8.

Figura 8 – Coocorrências anteriores e posteriores da fraseologia de gênero *the primary outcome was*.

Query **The, primary, outcome, was** 32 (1,719.51 per million) 

Page of 2 [Next](#) | [Last](#)

file541065...	provided to both groups through the quitline. The primary outcome was self-reported continuous abstinence at 1
file541063...	and young adults with Marfan's syndrome. The primary outcome was the rate of aorticroot enlargement, expressed
file541063...	mg or placebo once daily for a maximum of 28 days. The primary outcome was the number of ventilator-free days to day 28.
file541062...	after admission and continued for up to 5 days. The primary outcome was all-cause mortality at 30 days. We performed a
file541062...	antibodies were periodically measured. The primary outcome was the frequency of biopsy-confirmed celiac
file541062...	findings underwent intestinal biopsies. The primary outcome was the prevalence of celiac disease autoimmunity
file541061...	dose, or placebo every 4 weeks for 32 weeks. The primary outcome was the rate of exacerbations. Other outcomes
file541062...	subcutaneously every 4 weeks for 20 weeks. The primary outcome was the degree of reduction in the glucocorticoid
file541061...	daily), in addition to recommended therapy. The primary outcome was a composite of death from cardiovascular
file541059...	, and implementation of best practices. The primary outcome was the cesarean delivery rate in the 1-year
file541058...	genetic testing or newborn examination. The primary outcome was the area under the
file541058...	upfront manual thrombectomy versus PCI alone. The primary outcome was a composite of death from cardiovascular
file541058...	and postoperative transfusions. The primary outcome was the change in Multiple Organ Dysfunction Score
file541057...	according to a protocol-specified algorithm. The primary outcome was the mean change in visual acuity at 1 year. In an
file541057...	PCI with the use of everolimus-eluting stents. The primary outcome was all-cause mortality. Secondary outcomes were
file541056...	assignments and microbiologic test results. The primary outcome was clinical cure 7 to 10 days after the end of
file541055...	were measured against predetermined targets. The primary outcome was the score on the modified Rankin scale (range, 0
file541055...	(hemoglobin level <9 g per deciliter). The primary outcome was a serious infection (sepsis or wound infection
file541055...	unit for organ recovery in the operating room. The primary outcome was delayed graft function in the kidney
file541055...	to 300 mg, or placebo, once weekly for 13 weeks. The primary outcome was the percentage change in APOC3 level from

A Tabela 5 mostra que, das cinco linhas de concordância analisadas, em 100% delas a fraseologia de gênero está em início de frase. Todas as coocorrências posteriores têm como elemento principal um substantivo que pode estar modificado ou não.


Tabela 5 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero *the primary outcome was*.

Pergunta	Coocorrência anterior (fórmula) Exemplo	Fraseologia	Coocorrência posterior (fórmula) exemplo
Qual o resultado principal de sua pesquisa?	-	<i>The primary outcome was</i>	<i>the rate of aorticroot enlargement...</i> (ART+SUBS+PREP+SUBS)
			<i>the number of ventilator-free days to day 28.</i> (ART+SUBS+PREP+SUBS+PREP+NOUN+N°)
			<i>a composite of death from cardiovascular...</i> (ART+SUBS+PREP+SUBS+PREP+ADJ)
			<i>the mean change in visual acuity at 1 year.</i> (ART+ADJ+SUBS+PREP+ADJ+SUBS+PREP+N°+SUBS)
			<i>all-cause mortality at 30 days.</i> (ADV+SUBS+PREP+N°+SUBS)

4.1.3 Results

Do movimento *results*, foram extraídas as linhas de concordância da fraseologia de gênero *as compared with*, ocorrida 69x no CRE-NEJM-R, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 9.

Figura 9 – Coocorrências anteriores e posteriores da fraseologia de gênero *as compared with*.

Query **As, compared, with** 69 (1,858.39 per million) 

Page of 4 [Next](#) | [Last](#)

file541085... showed no treatment effect of progesterone **as compared with** placebo (odds ratio, 0.96; confidence
file541085... disability) was 50.4% with progesterone, **as compared with** 50.5% with placebo. Mortality was similar in
file541084... . Continued treatment with thienopyridine, **as compared with** placebo, reduced the rates of stent thrombosis
file541085... , fatigue, vomiting, and decreased appetite. **As compared with** chemotherapy, crizotinib was associated with
file541084... of 33,915 patients with diabetes (8.0%) died, **as compared with** 4835 of 169,249 controls (2.9%) (adjusted
file541084... hemoglobin level for patients with diabetes **as compared with** controls were 2.36 (95% CI, 1.97 to 2.83) for a
file541083... response in the combination group was 68%, **as compared with** 45% in the control group (P<0.001), including
file541083... of adverse events of grade 3 or higher, **as compared with** vemurafenib and placebo (65% vs. 59%), and
file541083... (1 from sepsis and 1 from pancreatic cancer). **As compared with** no screening, screening with low-dose CT cost
file541083... reductions in the parenteral group, **as compared with** the enteral group, in rates of hypoglycemia (44
file541082... those receiving subcutaneous mepolizumab, **as compared with** those receiving placebo (P< 0.001 for both
file541082... dose was 50% in the mepolizumab group, **as compared with** no reduction in the placebo group (P=0.007).
file541082... dose, patients in the mepolizumab group, **as compared with** those in the placebo group, had a relative
file541082... , 0.82 to 1.29; P=0.81). Prednisolone therapy, **as compared with** placebo, was associated with significant
file541082... . Both prednisolone and *M. indicus pranii*, each **as compared with** placebo, were associated with a significant
file541081... patients) in the thrombus-aspiration group, **as compared with** 5.6% (202 of 3623) in the PCI-only group (hazard
file541080... - dexamethasone than with MPT (70% vs. 78%). **As compared with** MPT, continuous lenalidomide - dexamethasone
file541080... reduced the median time to culture conversion, **as compared with** placebo, from 125 days to 83 days (hazard ratio
file541080... outcome occurred in 3317 participants (3.3%). **As compared with** an estimated sodium excretion of 4.00 to 5.99 g
file541080... sodium excretion of 6.00 g or more per day. **As compared with** the reference range, an estimated sodium

A Tabela 6 mostra que, das cinco linhas de concordância analisadas, em apenas um caso a fraseologia de gênero aparece no início de frase. As coocorrências anteriores e posteriores possuem como elemento principal um substantivo que pode estar modificado ou não.

Tabela 6 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero *as compared with*.

Pergunta	Coocorrência anterior exemplo fórmula	Fraseologia	Coocorrência posterior exemplo fórmula
Qual tratamento demonstrou ter efeito comparado a outro tratamento?	<i>... showed no treatment effect of progesterone</i> (VERB+SUBS+PREP+S UBS)		<i>placebo ...</i> (SUBS)
Qual é a porcentagem entre uma substância/tratamento em comparação a outra substância/tratamento?	<i>... was 50.4% with progesterone,</i> (VERB+N ^o +PREP+SUBS)		<i>50.5% with placebo.</i> (N ^o +PREP+SUBS)
Qual resultado um tratamento que teve continuidade com determinada substância pode ter ao ser comparado com outra substância?	<i>Continued treatment with thienopyridine,</i> (ADJ+SUBS+PREP+S UBS)	<i>As/as compared with</i>	<i>placebo, reduced the rates of stent thrombosis...</i> (SUBS+VERB+ART+SUBS PREP+SUBS+SUBS)
Ao comparar dois tratamentos, qual deles foi associado com resultados positivos/negativos?	-		<i>chemotherapy, crizotinib was associated with...</i> (SUBS+SUBS+VERB+FG)
Qual grupo apresentou redução/aumento em comparação a outro grupo? Quais são os resultados disso?	<i>Reduction in the parenteral group</i> (SUBS+PREP+ART+AD V+SUBS)		<i>the enteral group...</i> (ART+ADJ+SUB)

4.1.4 Conclusions

Do movimento *conclusions*, foram extraídas as linhas de concordância da fraseologia de gênero *was associated with*, ocorrida 27x no CRE-NEJM-C, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 10.

Figura 10 – Coocorrências anteriores e posteriores da fraseologia de gênero *was associated**with*

Query **Was, associated, with** 27 (3,183.96 per million) ⓘ

Page 1 of 2 Go Next | Last

file541102... therapy in helping smokers quit smoking, but it **was associated with** a higher frequency of self-reported adverse
file541100... - ABT-450/r and dasabuvir with ribavirin **was associated with** a low rate of serious adverse events and a high
file541101... .) Monotherapy with an ACE inhibitor **was associated with** blood-pressure control in most patients with
file541101... and cerebrovascular events but **was associated with** an increased risk of bleeding. (Funded by a
file541101... reverse remodeling. Mitral-valve repair **was associated with** a reduced prevalence of moderate or severe
file541100... .) The addition of cobimetinib to vemurafenib **was associated with** a significant improvement in
file541099... , although the later introduction of gluten **was associated with** a delayed onset of disease. A high-risk HLA
file541099... reduced asthma exacerbations and **was associated with** improvements in markers of asthma control. (
file541098... , NCT01691508.) Initial ultrasonography **was associated with** lower cumulative radiation exposure than
file541096... given until disease progression **was associated with** a significant improvement in
file541096... sodium intake between 3 g per day and 6 g per day **was associated with** a lower risk of death and cardiovascular events
file541096... than 1.50 g per day, higher potassium excretion **was associated with** a lower risk of death and cardiovascular events
file541095... therapy in patients with type 1 diabetes **was associated with** a substantial reduction in the long-term risk
file541095... with alcoholic hepatitis. Prednisolone **was associated with** a reduction in 28-day mortality that did not
file541094... NYHA class IV heart failure within 180 days but **was associated with** an increased rate of stroke within 30 days. (
file541093... surgery in patients with rectal cancer **was associated with** rates of locoregional recurrence and
file541094... was similar to that associated with CABG. PCI **was associated with** a higher risk of myocardial infarction (among
file541093...].) We found that treatment with ISIS 304801 **was associated with** significant lowering of triglyceride levels,
file541090... .) Lenvatinib, as compared with placebo, **was associated with** significant improvements in
file541091... of the coronary-sinus reducing device **was associated with** significant improvement in symptoms and

A Tabela 7 mostra que, das cinco linhas de concordância analisadas, em 100% delas a fraseologia de gênero está no meio de frase. As coocorrências anteriores e posteriores possuem como elemento principal um substantivo que pode estar modificado ou não.

Tabela 7 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero *was associated with*.

Pergunta	Coocorrência anterior fórmula exemplo	Fraseologia	Coocorrência posterior fórmula exemplo
Com o que o tratamento realizado pode ser associado em relação a um determinado grupo de pacientes?	<i>Monotherapy with an ACE inhibitor</i> (SUBS+PREP+ART+SUBS)		<i>blood-pressure control in most patients with</i> ... (SUBS+PREP+ADV+SUBS+PREP)
Qual substância em adição a outra foi associada a uma melhora significativa?	<i>The addition of cobimetinib to vemurafenib</i> (ART+SUBS+PREP+SUBS+PREP+SUBS)	<i>was associated with</i>	<i>a significant improvement in ...</i> (ART+ADJ+SUBS+PREP)
Em qual medida certa substância foi associada ao menor risco de morte?	<i>...sodium intake between 3 g per day and 6 g per day</i> (SUBS+SUBS+PREP+NUM+PREP+SUBS+CONJ+NUM+PREP+SUBS)		<i>a lower risk of death...</i> (ART+ADV+SUBS+PREP+SUBS)
Qual procedimento médico foi	<i>PCI</i> (SUBS)		<i>a higher risk of myocardial infarction</i>

associado a um risco maior ao paciente? Qual é risco?		... (ART+ADV+SUBS+PREP+SUBS)
Qual tratamento foi associado a uma significante redução dos fatores de risco à saúde?	<i>We found treatment with ISIS 304801</i> (PRON+VERB+SUBS+PREP+SUBS+NUM)	<i>significant lowering of triglyceride levels...</i> (ADJ+SUBS+PREP+SUBS)


4.2 FGA

A seguir, apresentam-se os dados referentes às FGA *the treatment of, patients were randomly assigned, among patients with e in patients with* dos movimentos retóricos *background, methods, results e conclusions*.

4.2.1 Background

Do movimento *background*, foram extraídas as linhas de concordância da fraseologia de gênero da área *the treatment of*, ocorrida 11x no CRE-NEJM-B, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 11.

Figura 11 – Coocorrências anteriores e posteriores da fraseologia de gênero da área *the treatment of*

Query **The, treatment, of** 11 (1,161.69 per million) 

file544369...	suggest that statins may be beneficial in the treatment of the acute respiratory distress syndrome (ARDS
file544386...	are widely used alternatives to surgery for the treatment of varicose veins, but their comparative
file544380...	and pentoxifylline are both recommended for the treatment of severe alcoholic hepatitis, but uncertainty
file544378...	aflibercept, bevacizumab, and ranibizumab in the treatment of diabetic macular edema are unknown. RESULTS OF
file544376...	, and earlier intervention improve outcomes. The treatment of symptomatic congenital cytomegalovirus (CMV
file544375...	bortezomib was initially approved for the treatment of relapsed mantle-cell lymphoma. We
file544384...	and tipiracil hydrochloride, was effective in the treatment of refractory colorectal cancer. We conducted a
file544385...	concerns about a disproportionate increase in the treatment of youths with low levels of mental health
file544373...	the safety and efficacy of thrombectomy for the treatment of stroke in a trial embedded within a
file544367...	attained FEV1 is less than population norms. The treatment of peripheral artery disease with percutaneous
file544367...	anti - tumor necrosis factor therapies, for the treatment of moderate-to-severe plaque psoriasis.

A Tabela 8 mostra que, das cinco linhas de concordância analisadas, apenas em um caso a fraseologia de gênero da área aparece no início de frase. É possível observar certo padrão nas coocorrências anteriores na medida em que todas elas terminam com preposição. As coocorrências posteriores também apresentam certa estabilidade, pois possuem como elemento principal um substantivo que pode estar modificado ou não.

Tabela 8 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero da área *the treatment of*

Pergunta	Coocorrência anterior exemplo Fórmula	Fraseologia	Coocorrência posterior Exemplo Fórmula
Quais são as alternativas usadas para o tratamento de determinadas doenças?	<i>... are widely used alternatives to surgery for</i> (VERB+ADV+VERB+SUBS+P REP+SUBS+PREP)		<i>varicose veins ...</i> (ADJ+SUBS)
Qual substância é recomendada para o tratamento de determinadas doenças?	<i>... pentoxifylline are both recommended for</i> (SUBS+VERB+ADJ+VERB+P REP)		<i>severe alcoholic hepatitis</i> ... (ADJ+SUBS)
Quais medicamentos são desconhecidos no tratamento de determinadas doenças?	<i>... aflibercept, bevacizumab, and ranibizumas in</i> (SUBS+SUBS+CONJ+SUBS+P REP)	<i>The/the treatment of</i>	<i>diabetic macular edema are unknown.</i> (ADJ+ADJ+SUBS+VERB+ADJ)
Qual a eficácia de um procedimento no tratamento de uma determinada doença?	<i>... the safety and efficacy of thrombectomy for</i> (ART+SUBS+CONJ+SUBS+P REP+SUBS+PREP)		<i>stroke in a trial embedded within a ...</i> (SUBS+PREP+ART+SUBS+ADJ+ADV+ART)
Qual tratamento pode ser feito para determinadas doenças?	-		<i>peripheral artery disease with percutaneous ...</i> (ADJ+SUBS+SUBS+PREP+ADJ)

4.2.2 Methods

Do movimento *methods*, foram extraídas as linhas de concordância da fraseologia de gênero da área *patients were randomly assigned*, ocorrida 13x no CRE-NEJM-M, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 12.

Figura 12 – Coocorrências anteriores e posteriores da fraseologia de gênero da área *patients were randomly assigned*

Query Patients, were, randomly, assigned 13 (698.55 per million) ⓘ	
file541064...	drug (clopidogrel or prasugrel) and aspirin, patients were randomly assigned to continue receiving thienopyridine
file541065...	systemic treatment for advanced disease. Patients were randomly assigned to receive oral crizotinib at a dose of 250 mg
file541064...	protein level of more than 150 mg per liter. Patients were randomly assigned to nasogastric tube feeding within 24 hours
file541058...	assignments in the parent trials, eligible patients were randomly assigned in a 2:1 ratio to receive either evolocumab (140
file541058...	with or without other lipid-lowering therapy. Patients were randomly assigned in a 2:1 ratio to receive alirocumab (150 mg) or
file541057...	analysis in 56 hospitals in England. Patients were randomly assigned to receive either EGDT (a 6-hour resuscitation
file541056...	abscesses underwent incision and drainage. Patients were randomly assigned in a 1:1 ratio to receive either clindamycin or
file541055...	(ISIS 304801 - fibrate cohort). Eligible patients were randomly assigned to receive either ISIS 304801, at doses ranging
file541048...	, randomized, controlled trial in Canada. Patients were randomly assigned to undergo limited occult-cancer screening (
file541047...	interruption of warfarin therapy, patients were randomly assigned to receive bridging anticoagulation therapy
file541047...	coinfecting with HIV-1. Previously untreated patients were randomly assigned in a 2:1 ratio to receive either 12 weeks or 8
file541046...	the Phe508delCFTR mutation. In both studies, patients were randomly assigned to receive either lumacaftor (600 mg once daily
file541046...	plaque psoriasis. A total of 293 patients were randomly assigned to receive guselkumab (5 mg at weeks 0 and 4 and

A Tabela 9 ilustra que, das cinco linhas de concordância analisadas, em apenas um caso a fraseologia de gênero da área aparece no início de frase. As coocorrências anteriores dividem-se em dois tipos: (i) as que apresentam um adjetivo como elemento principal que pode estar modificado ou não e (ii) as que indicam que o elemento central é um substantivo que pode estar modificado ou não. Por outro lado, as coocorrências posteriores apresentam certa estabilidade na medida em que quatro delas iniciam com preposição + verbo.

Tabela 9 – Cinco coocorrências anteriores e posteriores da FGA *patients were randomly assigned*

Pergunta	Coocorrência anterior exemplo FÓRMULA	Fraseologia	Coocorrência posterior exemplo FÓRMULA
Para qual finalidade os pacientes foram alocados aleatoriamente?			<i>to undergo limited occult-cancer screening ...</i> (PREP +VERB+ADJ+ADJ+SUBS+ SUBS)
	<i>Eligible (ADJ)</i>		<i>to receive either ... (PREP+VERB+PRON)</i>
	<i>Previously untreated (ADV+ADJ)</i>	<i>Patients/patients were randomly assigned</i>	<i>in a 2:1 ratio to receive either 12 weeks or 8... (PREP+ART+NUM+SUBS+ PREP+VERB+PRON+NUM +SUBS+CONJ+NUM)</i>
Em que estudos os pacientes foram distribuídos aleatoriamente para receber determinado tipo de tratamento?	<i>In both studies, (PREP+ADJ+SUBS)</i>		<i>to receive either.... (PREP+VERB+PRON)</i>
Quantos pacientes foram distribuídos para receber determinada substância?	<i>A total of 293 (ART+SUBS+PREP+NUMB)</i>		<i>to receive guselkumab ... (PREP+VERB+SUBS)</i>

4.2.3 Results

Do movimento *results*, foram extraídas as linhas de concordância da fraseologia de gênero da área *among patients with*, ocorrida 13x no CRE-NEJM-R, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 13.

Figura 13 – Coocorrências anteriores e posteriores da fraseologia de gênero da área *among patients with*

Query **Among, patients, with** 13 (350.13 per million) ⓘ

file541082...	in the incidence of the primary end point	among patients with	activity-limiting angina but not among those
file541081...	. Colorectal-cancer mortality was increased	among patients with	high-risk adenomas (expected deaths, 209;
file541081...	, 1.16; 95% CI, 1.02 to 1.31), but it was reduced	among patients with	low-risk adenomas (expected deaths, 189;
file541078...	rate was 73.0%; these rates were highest	among patients with	MYD88L265PCXCR4WT (with WT indicating
file541077...	with PCI than with CABG was not significant	among patients with	complete revascularization but was
file541070...), and antipsychotic drugs (from 0.2% to 1.2%).	Among patients with	BRAF wild-type tumors, the rate of confirmed
file541070...	selected as the cutoff from the training group.	Among patients with	a proportion score of at least 50% in the
file541068...	associated with overall survival	among patients with	grade II or III gliomas but not among patients
file541068...	patients with grade II or III gliomas but not	among patients with	grade IV gliomas. The molecular groups were
file541067...	; 98% were receiving antiretroviral therapy.	Among patients with	genotype 1, a sustained virologic response was
file541066...	; 95% CI, 0.41 to 0.76; adjusted hazard ratio	among patients with	a baseline CD4+ count of >=500 cells per cubic
file541066...	; 95% CI, 0.48 to 0.88; adjusted hazard ratio	among patients with	a baseline CD4+ count of >=500 cells per cubic
file541066...	value of the classifier was 91% (95% CI, 75 to 98)	among patients with	a nondiagnostic bronchoscopic examination.

A Tabela 10 demonstra que das cinco linhas de concordância analisadas, em apenas um caso a fraseologia de gênero da área se encontra no início de frase. Entre as coocorrências anteriores, não foi encontrado certa regularidade. Por outro lado, as coocorrências posteriores apresentam certa estabilidade, pois possuem como elemento principal um substantivo que pode estar modificado ou não.

Tabela 10 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero da área *among patients with*

Pergunta	Coocorrência anterior exemplo Fórmula	Fraseologia	Coocorrência posterior exemplo Fórmula
Quais foram os resultados obtidos entre pacientes com determinada doença?	...in the incident of the primary end point (PREP+ART+SUBS+PREP+FG)		activity-limiting angina but not among those... (SUBS+SUBS+CONJ+ADV+PREP+PRON)
Em quais pacientes foi aumentada a mortalidade por determinado tipo de doença?	Colorectal-cancer mortality was increased (ADJ+SUBS+SUBS+VERB)		high-risk adenomas... (ADJ+SUBS+SUBS)
O que foi reduzido em pacientes com determinado tipo de doença?	... but it was reduced (CONJ+PRON+VERB)	Among/among patients with	low-risk adenomas... (ADJ+SUBS+SUBS)
O que pode ser associado a um determinado grupo de pacientes e não a outros?	... associated with overall survival (FG+ADJ+SUBS)		grade II or III gliomas but not among patients grade IV gliomas. (SUBS+CONJ+SUBS+CONJ+ADV+PREP+SUBS+SUBS+SUBS)
Qual a resposta virológica sustentada que se obteve entre um determinado grupo de pacientes?			genotype 1, a sustained virologic response was... (SUBS+ART+ADJ+ADJ+SUBS+VERB)

4.2.4 Conclusions

Do movimento *conclusions*, foram extraídas as linhas de concordância da fraseologia de gênero da área *in patients with*, ocorrida 30x no CRE-NEJM-C, que permitiu observar as coocorrências anteriores e posteriores que estão demonstradas na Figura 14.

Figura 14 – Coocorrências anteriores e posteriores da FGA *in patients with*

Query **In, patients, with** 30 (3,537.74 per million) ⓘ

Page 1 of 2 [Next](#) | [Last](#)

file541102... showed no clinical benefit of progesterone **in patients with** severe TBI. These data stand in contrast to the
 file541102... over placebo in the improvement of outcomes **in patients with** acute TBI. (Funded by the National Institute of
 file541101... pemetrexed-plus-platinum chemotherapy **in patients with** previously untreated advanced ALK-positive
 file541101... Funded by the Frederick Adler Fund and others.) **in patients with** moderate ischemic mitral regurgitation, the
 file541100... , in reducing the rate of infection or death **in patients with** acute pancreatitis at high risk for
 file541100... , crizotinib showed marked antitumor activity **in patients with** advanced ROS1-rearranged NSCLC. ROS1
 file541098... ClinicalTrials.gov number, NCT00047385.) **in patients with** early rheumatoid arthritis who had a remission
 file541109... effects, did not improve clinical outcomes **in patients with** ARDS. (Funded by the U.K. National Institute
 file541099... Controlled Trials number, ISRCTN51995477.) **in patients with** stable coronary artery disease, FFR-guided
 file541098... Controlled Trials number, ISRCTN61576291.) **in patients with** tuberculous pericarditis, neither
 file541099... .) Routine thrombus aspiration before PCI **in patients with** STEMI did not reduce the rate of death from any
 file541098... .) Prehospital administration of ticagrelor **in patients with** acute STEMI appeared to be safe but did not
 file541097... to be useful for detecting somatic mutations **in patients with** brain malformations. High-coverage
 file541096... of Health and others.) Rociletinib was active **in patients with** EGFR-mutated NSCLC associated with the T790M
 file541095...].) AZD9291 was highly active **in patients with** lung cancer with the EGFR T790M mutation who had
 file541095...].) Intensive therapy **in patients with** type 1 diabetes was associated with a
 file541095...].) Pentoxifylline did not improve survival **in patients with** alcoholic hepatitis. Prednisolone was
 file541094...].) **in patients with** STEMI who were undergoing primary PCI, routine
 file541094...].) **in patients with** septic shock who were identified early and
 file541093...].) Laparoscopic surgery **in patients with** rectal cancer was associated with rates of

A Tabela 11 mostra que das cinco linhas de concordância analisadas, somente uma delas está no início de frase. As coocorrências anteriores e posteriores possuem como elemento principal um substantivo que pode estar modificado ou não.

Tabela 11 – Cinco coocorrências anteriores e posteriores da fraseologia de gênero de área *in patients with*

Pergunta	Coocorrência anterior exemplo Fórmula	Fraseologia	Coocorrência posterior exemplo Fórmula
Qual foi o resultado obtido em pacientes com determinado tipo de doença?	... showed no clinical benefit of progesterone (VERB+ADV+ADJ+SUBS +PREP+SUBS)	In/in patients with	severe TBI. (ADJ+SUBS)
O que é possível concluir sobre pacientes com determinado tipo de doença?	.	.	moderate ischemic mitral regurgitation... (ADJ+ADJ+ADJ+SUBS)
Quais foram os resultados clínicos obtidos em pacientes com determinado tipo de doença?	...did not improve clinical outcomes (AUX+ADV+VERB+ADJ +SUBS)	.	ARDS. (SUBS)
Quais os resultados obtidos com determinado tipo de	Routine thrombus aspiration before PCI	.	STEMI did not reduce the rate of death from...

procedimento em pacientes com determinado tipo de doença?	(ADJ+SUBS+SUBS+ADV +SUBS)	(SUBS+AUX+ADV+VERB+ART+SUBS+PREP+SUBS+PREP)
O que pode ser útil para detectar mutações somáticas em pacientes com determinado tipo de doença?	...to be useful for detecting somatic mutations (PREP+VERB+ADJ+PREP+VERB+ADJ+SUBS)	brain malformations. (SUBS+ADJ)

Nesta seção, foram descritos os dados obtidos a partir das observações realizadas. A seguir, é apresentada a tabela integradora (Tabela 12) construída para sintetizar os resultados obtidos até o momento.

Tabela 12 – Tabela Integradora

Movimentos retóricos da NEJM	Pergunta	Função comunicativa	Coocorrência anterior (fórmula)	FG e FGA	Coocorrência posterior (fórmula)
<i>Background</i>	Quais são os dados obtidos até o momento no uso de determinada substância/diagnóstico/terapia?	Descrever os dados obtidos até o momento sobre o uso de determinadas substâncias/diagnóstico/terapia.	<i>emerging data on...</i> (ADJ + SUBS + PREP)	the use of	<i>lenalidomide and low-dose dexamethasone are...</i> (SUBS+VERB)
	Quais são as alternativas usadas para o tratamento de determinadas doenças?	Descrever as alternativas usadas para o tratamento de determinadas doenças.	<i>... are widely used alternatives to surgery for</i> (VERB+ADV+VERB+SUBS+PREP+SUBS+PREP)	the treatment of	<i>varicose veins ...</i> (SUBS)
<i>Methods</i>	Qual o resultado principal de sua pesquisa?	Descrever o resultado principal da pesquisa.	.	the primary outcome was	<i>the rate of aorticroot enlargement...</i> (ART+SUBS+PREP+SUBS)
	Para qual finalidade os pacientes foram alocados aleatoriamente?	Descrever a finalidade para a qual os pacientes terem sido alocados aleatoriamente.	<i>Eligible</i> (ADJ)	patients were randomly assigned	<i>to receive either ...</i> (PREP+VERB+PRON)
<i>Results</i>	Qual tratamento demonstrou ter efeito comparado a outro tratamento?	Descrever o resultado de um tratamento em relação a outro.	<i>... showed no treatment effect of progesterone</i> (VERB+SUBS+PREP+SUBS)	as compared with	<i>placebo ...</i> (NOUN)
	Em quais pacientes foi	Descrever em quais	Colorectal-cancer	among patients	high-risk adenomas...

	aumentada a mortalidade por determinado tipo de doença?	pacientes foi aumentada a mortalidade por determinado tipo de doença.	mortality was increased (ADJ+SUBS+SUBS+VERB)	with	(ADJ+SUBS+SUBS)
<i>Conclusions</i>	Qual tratamento foi associado a uma significativa redução dos fatores de risco à saúde?	Descrever o resultado de um determinado tratamento.	<i>We found treatment with ISIS 304801</i> (PRON+VERB+SUBS+PREP+SUBS+NUM)	was associated with	<i>significant lowering of triglyceride levels...</i> (ADJ+ADV+PREP+SUBS)
	Quais foram os resultados clínicos obtidos em pacientes com determinado tipo de doença?	Descrever os resultados clínicos obtidos em pacientes com determinado tipo de doença.	...did not improve clinical outcomes (AUX+ADV+VERB+ADJ+SUBS)	in patients with	ARDS. (SUBS)

5. CONSIDERAÇÕES FINAIS

A intenção que se teve ao realizar este estudo foi a de apresentar uma proposta de análise linguística que possibilitasse a identificação, descrição e análise das unidades ou expressões multipalavras típicas do gênero *abstract*, e das unidades multipalavras típicas do gênero *abstract* da área médica, aqui representada pelos textos da revista NEJM. Ao expor tal proposta pretendia-se que a mesma pudesse ser ampliada para descrever não apenas os resumos de outros periódicos da área médica, mas também os textos de outras áreas de especialidade e de outros gêneros acadêmicos (e.g. artigos e resenhas), indo ao encontro do que propõe Hyland (2008) quando afirma que:

Numerosos estudos hoje mostram que as características do idioma são específicas de áreas de especialidade e que a melhor maneira de preparar os alunos para seus estudos não é procurar itens didáticos universalmente apropriados, mas proporcionar-lhes uma compreensão das características dos discursos que eles encontrarão em seus cursos específicos (HYLAND, 2008).

Ainda que a amostra tenha sido pequena e que tratasse apenas de um periódico de uma área de especialidade, a metodologia de análise proposta se mostrou bastante produtiva. A partir das etapas metodológicas sugeridas que envolveram: (i) a escolha da área de conhecimento a ser observada; (ii) a identificação de um periódico de alto fator de impacto da área; (iii) a seleção do gênero a ser estudado; (iv) a análise qualitativa de dez textos; (v) a compilação de um corpus de análise; (vi) a definição de uma ferramenta de análise; (vii) a extração das FG e das FGA mais frequentes do *corpus*; (viii) a identificação das coocorrências das FG e FGA e de seus contextos imediatos; (ix) a análise dos dados por meio da construção de tabelas com a inclusão de fórmulas e perguntas que representassem os padrões identificados, foi possível criar um protocolo que poderá ser replicado para amostras maiores de outras áreas de especialidade e de outros gêneros textuais.

Ao longo do estudo, foram sendo percebidas algumas limitações que certamente serão consideradas em outras análises e descrições linguísticas que venham a ser realizadas no futuro. A primeira delas foi o entendimento de que os contextos imediatamente anteriores e posteriores às FG e FGA poderiam ser ampliados, como forma de permitir a observação de padrões que pudessem coocorrer com as FGs e FGAs. Uma outra limitação foi o tempo gasto fazendo a categorização dos padrões sintáticos para a confecção das fórmulas. Uma forma de solucionar tal questão poderia ser o uso de um corpus sintaticamente anotado o que, possivelmente, permitiria a construção automática das fórmulas.

Por fim, a partir dos resultados obtidos nesta pesquisa, é possível perceber que não há um padrão léxico-gramatical único para cada fraseologia identificada. Há padrões que representam estruturas frasais possíveis que são usadas por escritores especializados da área. Tais conclusões têm implicações pedagógicas importantes que apontam para a necessidade de desenvolvimento de tarefas e recursos que permitam o acesso do escritor não especializado aos padrões léxico-gramaticais utilizados pelos escritores especializados, fazendo com que os mesmos possam escrever de forma mais idiomática.

Como já mencionado, estudos futuros poderão focar na ampliação da amostra dentro da área médica e na descrição de outros gêneros acadêmicos. Da mesma forma, o presente estudo poderá informar pesquisas que tenham por objetivo o desenvolvimento de tarefas e de recursos online para a melhoria da produção escrita de gêneros acadêmicos em inglês.

REFERÊNCIAS

- ACOSTA, O. C. **Identificação e tratamento de expressões multipalavras aplicado à recuperação de informação**. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, p. 16, Porto Alegre, 2011.
- BAKHTIN, M. Os gêneros do discurso (1952-1953). In.: **Estética da criação verbal**, v. 4, p. 261-306. Trad. Maria Ermantina Galvão Gomes e Pereira. São Paulo: Martins Fontes, 1992.
- BAKHTIN, M. O discurso de outrem. **Marxismo e filosofia da linguagem**, v. 9, p. 144-154, 1999.
- BHATIA, V. K. Análise de gêneros hoje. Tradução de Benedito Gomes Bezerra. **Revista de Letras**, v. 23, n. 1/2, p. 112-115, 2001.
- BIASI-RODRIGES, B.; ARAÚJO, J. C.; SOUSA, S. C. T. de. (Orgs.). **Gêneros textuais e comunidades discursivas: um diálogo com Jonh Swales**. Belo Horizonte: Autêntica Editora, p. 49-75, 2009.
- BRASIL. MINISTÉRIO DA EDUCAÇÃO. PORTARIA n. 1.466. Diário Oficial da União n 244, 19 de dezembro de 2012, Seção 1, p. 28.
- BRASIL. MINISTÉRIO DA EDUCAÇÃO. PORTARIA n. 973. Diário Oficial da União n 222, 17 de novembro de 2014, Seção 1, p. 11.
- CABRÉ, M. T.; BACH, C. **Coneixement, llenguatge i discurs especialitzat**. Barcelona: IULA, Serie monografies, 2005.
- CIAPUSCIO, G. **Wissenschaft für den Laien: Untersuchungen zu populärwissenschaftlichen Nachrichten aus Argentinien**. Bonn: Romanistischer Verlag, 1993.
- CIAPUSCIO, G. Formulation and reformulation procedures in verbal interactions between experts and (semi-) laypersons. **Discourse Studies**, v. 5, n.2, 207-233, 2003.
- 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.

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.

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

HYLAND, K. As can be seen: Lexical bundles and disciplinary variation. **English for specific purposes**, v. 27, n. 1, p. 4-21, 2008.

KILIAN, C. K.; LOGUERCIO, S. D. Fraseologias de gênero em resumos científicos de Linguística, Engenharia de Materiais e Ciências Econômicas. **TradTerm**, v. 26, p. 241-267, 2015.

MARCUSCHI, L. A. A QUESTÃO DO SUPORTE DOS GÊNEROS TEXTUAIS (parte 1). **DLCV-Língua, Linguística & Literatura**, v. 1, n. 1, 2003.

MCENERY, A. M.; WILSON, A. **Corpus linguistics: an introduction**. 2nd Ed. UK: Edinburgh University Press, 2001.

SWALES, J. M. **Genre analysis: English in academic and research settings**. Cambridge University Press, 1990.

SARDINHA, T. B. **Linguística de Corpus**. São Paulo: Editora Manole. 2004.

TEMMERMAN, R. Approaches to terminology. Now that the dust has settled.... **Synaps**, v. 20, p. 27-36, 2007.

THOMSON, J. J. Turning the Trolley. **Philosophy & Public Affairs**, v. 36, n. 4, p. 359-374, 2008.

VAN LEEUWEN, T. Genre and field in critical discourse analysis. *Discourse & society*, Sage Publications, v. 4, n. 2, p. 193-223, 1993. Disponível em: <http://journals.sagepub.com/doi/abs/10.1177/0957926593004002004>. Acesso em: 15 set. 2016.

WRIGHT, D. W.; YEATTS, S. D.; SILBERGLEIT, R. et al. Very early administration of progesterone for acute traumatic brain injury. **N Engl J Med**, v. 371, n. 25, p. 2457-66, 2014.

ZANELLA, A.; HERBELE, V. Abstract ou *summary*: um estudo do metadiscorso e da organização textual de resumos acadêmicos na área biomédica. **Artigos catarinenses de medicina**, v. 34, n. 4, p.15-22, 2005.