

**O léxico como um recurso linguístico para a produção de significado no texto: um estudo de caso com protocolos de investigação**

---

The lexis as a linguistic resource for the production of meaning in texts:  
a case study with probe protocols

El léxico como un recurso lingüístico para la producción de significado en el texto:  
un estudio de caso con protocolos de investigación

**Giacomo Patrocínio Figueredo**

Universidade Federal de Ouro Preto (UFOP/Brasil)

**Júlia Santos Nunes Rodrigues**

Universidade Federal de Minas Gerais (UFMG/Brasil)

**Francieli Silvéria Oliveira**

Universidade Federal de Minas Gerais (UFMG/Brasil)

**Ricardo José Alves**

Universidade Federal de Minas Gerais (UFMG/Brasil)

**RESUMO**

Este artigo tem por objetivo investigar a complementaridade dos papéis desempenhados pelo o léxico e a gramática na organização do texto. Toma-se como base a Linguística Sistêmico-Funcional, uma vez que esta postula uma relação de interdependência entre léxico e gramática. Pretendendo contribuir com o conhecimento acerca do papel que o léxico desempenha na organização geral da língua, este artigo promove um estudo de caso no qual um *corpus* composto por artigos acadêmicos e matérias jornalísticas a respeito do Diabetes Mellitus foi compilado. Em seguida, foram

---

\* Sobre os autores, ver página 59



aplicados no *corpus* dois protocolos de investigação lexical. Os resultados apontaram que mesmo mantendo o domínio experiencial, construído pelo léxico, as escolhas sistêmicas nos dois processos sociossemióticos foram diferentes. Além disso, para um item ser considerado lexical é necessário que ele possa ser generalizado através do estabelecimento de relações taxonômicas de identificação, classificação e generalização.

**PALAVRAS-CHAVE:** Léxico. Protocolos de investigação. Linguística sistêmico-funcional.

#### **ABSTRACT**

*This paper investigates how lexis is constructed in text through two case studies grounded on Systemic Functional Linguistics (SFL). SFL models the interdependence between grammar and lexis as the former being the most delicate version of the latter. Aiming at investigating complementary properties between grammar and lexis, the paper carries out a case study by compiling a corpus of academic articles and popular science articles about Diabetes Mellitus. Two lexical research protocols were applied to the corpus. The results indicate that the systemic choices in the two socio-semiotic processes were different despite keeping the experiential domain. In addition, the lexical status of a language item depends on its potential to establish taxonomic relations of identification and classification, thus becoming part of a process of generalization.*

**KEYWORDS:** Lexis. Probe protocols. Systemic-functional linguistics.

#### **RESUMEN**

*Este artículo tiene por objetivo investigar la complementariedad de los papeles desempeñados por el léxico y la gramática en la organización del texto. Se toma como base la Lingüística Sistémico-Funcional, ya que esta postula una relación de interdependencia entre léxico y gramática. En este artículo se promueve un estudio de caso en el cual un corpus compuesto por artículos académicos y materias periodísticas acerca de la Diabetes Mellitus fue compilado. A continuación, se aplicaron en el corpus dos protocolos de investigación lexical. Los resultados apuntaron que aun manteniendo el dominio experiencial, construído por el léxico, las elecciones sistémicas en los dos procesos sociosemióticos fueron diferentes. Además, para un ítem ser considerado léxico es necesario que pueda ser generalizado a través del establecimiento de relaciones taxonómicas de identificación, clasificación y generalización.*

**PALABRAS CLAVE:** Léxico. Protocolos de investigación. Lingüística sistémico-funcional.

## **1 Considerações iniciais**

A concepção do sistema linguístico como uma combinação entre paradigmas e estruturas por um lado e, por outro, de um conjunto aberto de itens lexicais se configura como o ponto inicial de investigação deste artigo. Essa visão compartimentalizada e “combinatória” ou de “interdependência”, que é compartilhada por teorias linguísticas de bases epistemológicas distintas

(DI SCIULLO e WILLIAMS, 1987; HALLIDAY, 1977; BIBER *et al.*, 1999), procura, em geral, dar pesos diferentes ora para a organização linguística como um conjunto de 'sistemas fechados' – portanto sistemas e estruturas generalizadas que utilizam o léxico específico como realização (cf. HASAN, 1996) –, ora para a composição de 'conjuntos abertos' – isto é, itens lexicais específicos que formam em sua composição estruturas generalizadas, como aponta Martin (1992):

visto a partir da gramática, o léxico não possui natureza diferente. Este se distingue daquela apenas em termos de generalização: a gramática produz significados mais gerais, ao passo que o léxico produz significados mais específicos. Aqui então o contraste se vê na analogia do “tijolo e argamassa” herdada da gramática tradicional pelos estudos linguísticos formalistas. Ou bem parte-se da argamassa da estrutura e vão-se adicionando os tijolos que são os itens lexicais (abordagem transformacional); ou dá-se início pelo léxico e acrescentam-se a ele as camadas de estrutura (abordagem lexicalista) (MARTIN, 1992, p. 278).

Em que pese averiguar a primazia de uma ou de outra abordagem, a natureza da combinação entre estes dois aspectos fundamentais do sistema linguístico – i.e. a forma como sistemas fechados e conjuntos abertos se combinam na organização linguística – é tomada como um dado e, por conseguinte, se torna o ponto de partida para a investigação linguística. Essa compartimentalização pode dificultar a investigação das propriedades da combinação, uma vez que faz com que ela não seja foco de descrições detalhadas.

Desse modo, a Linguística de *Corpus* postula a combinação como interdependência entre gramática (sistemas fechados) e o léxico (conjuntos abertos). Biber *et al.* (1999) procuram caracterizar a combinação como frequência de co-ocorrência entre determinados sistemas gramaticais e conjuntos lexicais. Assim, identificam-se os tipos de relação entre sistemas e conjuntos sem, no entanto, explicar a natureza dessa interdependência (RÖMER;SCHULZE, 2009).

De modo semelhante, a Linguística Funcional também diferencia o papel que desempenham os sistemas daqueles do léxico; mantendo, assim, a concepção de naturezas distintas que se combinam na totalidade do sistema. Pesquisas de base funcional procuram evidenciar essa distinção caracterizando os aspectos que diferem a natureza dos sistemas fechados daquela dos conjuntos abertos, tanto como generalização/especificação (LEHMANN, 2002), abstração/concretude (BRINTON; TRAUGOTT, 2005) ou funcionalidade/conteúdo (BIDERMAN, 1996). Estas concentram-se mais nos pólos gramatical ou lexical do que na própria relação entre eles.

Para a Linguística Sistemico-Funcional, tradicionalmente confere-se grande peso à organização dos sistemas fechados (HALLIDAY, 2002), a combinação entre estes e os conjuntos abertos tem sua importância mais

diminuída, uma vez que estes são apenas opções mais detalhadas daqueles (HASAN, 1996). Assim, a adoção do ponto de vista prevalente na pesquisa sistêmico-funcional pode dificultar ainda mais o entendimento das propriedades de combinação ou interdependência entre sistemas/estruturas e o léxico.

Nesse sentido, este artigo busca examinar precisamente o papel da combinação entre sistemas fechados e conjuntos abertos na organização do sistema linguístico, privilegiando a descrição de suas propriedades. Como hipótese, postula-se aqui que um maior entendimento das relações pode oferecer um tipo de conhecimento sobre a organização linguística que as descrições compartimentalizadas somente não seriam capazes de oferecer. O artigo parte de um ponto de vista diferente sobre a combinação entre sistemas fechados e conjuntos abertos.

Toma como base, portanto, o conjunto de pesquisas de base sistêmico-funcional. Apesar de tradicionalmente o léxico ser excluído das descrições linguísticas – ou tomado apenas como um aspecto inatingível da gramática (cf. HASAN, 1996), a evolução da pesquisa em Linguística Sistêmico-Funcional, no entanto, tem apontado caminhos alternativos para a inclusão do léxico na descrição do sistema linguístico. Entre eles, ganha destaque aquela apontada por Martin e Rose (2007), a qual entende o léxico como o recurso linguístico que inicia e expande o campo (ou domínio experiencial) de um texto. Igualmente, o processo sociossemiótico do texto determina um conjunto de recursos que estão disponíveis para situações sociais específicas (HALLIDAY MATTHIESSEN, 2014). O campo, por sua vez, pré-determina um certo conjunto de itens lexicais que o realizam linguisticamente. Assim, o léxico de um texto é dependente de um campo e, da mesma forma, é sensível a mudanças nele.

A escolha por essa base teórica se deve ao fato de que a Linguística Sistêmico-Funcional (LSF) modela a língua como um sistema semiótico contextualizado, cujo papel principal é gerar significados que sejam relevantes para o contexto social dos falantes. Com isto, privilegia a língua em uso, bem como sua unidade fundamental – o texto. Em particular, o artigo aproxima-se da visão apresentada por Martin e Rose (2007), no que diz respeito à relação de dependência entre o léxico e o campo/domínio experiencial. Isso posto, partimos de um conjunto de premissas que servem de base para a seção de análise:

- (i) Em primeiro lugar, conceitua-se aqui léxico como o conjunto de recursos lexicais de um sistema linguístico, o qual inclui tanto os itens lexicais quanto suas relações. Desta forma, as relações entre itens lexicais formam, igualmente, parte dos recursos linguísticos do léxico.
- (ii) Dentro do escopo deste artigo, ainda, os recursos lexicais produzem significado no nível do texto e, por este motivo, estão restritos a este. Uma vez que o léxico é um conjunto aberto, a forma de estabelecer contraste e assim produzir significado (cf. o conceito de *valeur*, SAUSSURE, 1996; MARTIN, 2013) é determinar o tamanho do conjunto e a extensão das relações, que coincidem com o limite do texto. Nesse aspecto, vale ressaltar que este artigo se concentra na

forma como a construção e relação entre itens lexicais acontece no texto. Vale ressaltar que, apesar de haver a possibilidade de as relações entre itens lexicais serem intertextuais, estas fogem no presente momento ao escopo desta pesquisa.

(iii) Os recursos lexicais de um determinado texto produzem significados que não podem ser produzidos de outra forma – i.e., não são gerados por sistemas ou funções estruturais.

(iv) Sendo específico de um texto, o léxico contribui para a construção linguística do campo/domínio experiencial desse texto especificamente.

Diante dessas premissas, cabe, então, averiguar sua pertinência, assim perguntando:

- (1) seria possível identificar e descrever a natureza das relações entre itens lexicais de um texto?
- (2) de qual forma o léxico produz significados particularizados restritos a um dado texto?
- (3) o sistema linguístico determina especificidades na produção de significado, dividindo o trabalho de produção de forma que uma parte seja de responsabilidade exclusiva do léxico?
- (4) em que medida o léxico constrói o campo de um texto?

Mediante a necessidade de averiguar as premissas e questões apresentadas acima, este artigo possui como objetivo: investigar a complementaridade dos papéis desempenhados pelo o léxico (conjunto aberto) e a gramática (sistema fechado) na organização do texto, apresentando a forma pela qual o léxico é construído em textos de um tipo específico, com um campo/domínio experiencial e processo sociosemiótico determinados, e contrastá-lo com o léxico construído em textos de outro tipo, de mesmo campo/domínio experiencial, porém de processo sociosemiótico diferente. Mais especificamente, cumprirá o objetivo apresentando um estudo de caso no qual a construção do léxico será observada em um *corpus* de textos de matéria jornalística de popularização da ciência, cujo domínio experiencial é o conhecimento produzido sobre Diabetes Mellitus em português brasileiro. Como forma de distinguir essa construção do léxico, o artigo irá, em seguida, compará-la com um *corpus* de artigos acadêmicos de mesmo domínio experiencial.

Para tanto, o artigo se divide da seguinte forma: a próxima seção apresenta o embasamento teórico sobre o léxico. A seção 3 apresenta a metodologia, que inclui a compilação e análise do *corpus*, bem como os protocolos de investigação lexical utilizados. A seção 4 apresenta os resultados da pesquisa e mostra de que forma o léxico constrói itens e relações lexicais e a forma como estas interagem com sistemas gramaticais. Por fim, o artigo se encerra apresentado as considerações finais.

## 2 O léxico sob a perspectiva da Linguística de *Corpus*

Os estudos em linguística formal objetivam construir gramáticas de linguagens específicas que buscam entender e caracterizar a maneira como a competência linguística foi adquirida pelo falante (CHOMSKY, 1964; 1970). Diante disso, uma Gramática Universal contém uma base que consiste em um componente categorial (*categorial component*) e o léxico, em que este era entendido como uma lista não-estruturada de “palavras”, sendo um componente marginalizado da gramática (CHOMSKY, 1964, 1970; DI SCIULLO; WILLIAMS, 1987). Nesse sentido, como apontam Römer e Schulze (2009), na teoria e descrição linguística, o léxico e a gramática são muitas vezes tratados separadamente.

De modo diverso, sob a perspectiva da Linguística de *Corpus*, haveria, na verdade, uma interdependência entre o léxico e a gramática (RÖMER e SCHULZE, 2009). Essa ideia corrobora a de Quirk *et al.* (1985), que afirma que tanto a gramática quanto a lexicologia abarcam um conjunto indefinido de unidades.

A relação entre gramática e léxico como dependente é também enfatizada no trabalho de Biber *et al.* (1999). Segundo esses autores, estruturas particulares da gramática inglesa tendem a co-ocorrer com classes lexicais restritas, uma vez que há um conjunto de palavras ou frases que são frequentemente usadas em determinadas estruturas sintáticas. Por exemplo, segundo eles, o sujeito de alguns verbos são seres humanos, haveria, com isso, uma maior probabilidade de pronomes de terceira pessoa ou substantivos próprios preencherem o lugar de sujeito. Desse modo, esses aspectos da gramática do inglês que têm associações com as unidades lexicais comprovam, para Biber *et al.* (1999), a relação entre léxico e gramática.

Ainda, Biber *et al.* (1999) acrescentam o fato de essas associações léxico-gramaticais se diferenciarem de forma acentuada dependendo do registro em que são empregadas. Isso é exemplificado por meio de artigos de jornais e conversações: essas associações são bem diferentes em cada um desses registros pelo fato de os falantes, por exemplo, em uma conversa típica terem propósitos comunicativos bem diferentes do que os escritores de jornais.

Biber *et al.* (1999) consideram ainda as diferentes funções das palavras em cada contexto, as quais apresentam-se em três grupos: a palavra lexical, a palavra funcional, e a palavra insertiva. A palavra lexical é aquela que carrega o significado dentro de um texto e compreende substantivos, adjetivos e advérbios. A palavra funcional, por outro lado, é aquela que, dentro de um texto, conecta todas as unidades – sejam elas palavras lexicais sejam elas unidades maiores. Estas são, geralmente, as mais comuns em um texto. Por fim, existe a palavra insertiva, a qual traz conteúdo emotivo ou interacional a um texto (como ‘hm’ ou ‘ugh’). Embora haja debate acerca de sua natureza, e se elas constituem palavras verdadeiramente, sua utilização em textos falados é de grande relevância.

### 2.1 O léxico sob perspectivas funcionalistas

Nos estudos de Brinton e Traugott (2005, p. 9) a respeito dos processos de lexicalização e gramaticalização, o léxico é definido como um conjunto finito de formas armazenadas que possuem diversas possibilidades de combinação. Entre essas formas que abarcam o léxico de uma língua, há categorias tanto gramaticais quanto lexicais. Enquanto as primeiras estão relacionadas a conceitos mais concretos (*concrete concepts*), tais como objetos, ações, qualidades, as segundas dizem respeito a conceitos relacionais (*relational concepts*), mais abstratos, expressos, por exemplo, por afixos, por alterações internas a uma raiz, entre outros. (BRINTON; TRAUGOTT, 2005, p. 12).

Essa ideia corrobora a de Lehmann (2002), quando o autor define que item lexical está relacionado a algo concreto, específico, pertencendo a um inventário de palavras de uma língua; item gramatical, por outro lado, refere-se a algo abstrato, que está em conformidade com as regras da gramática. De forma complementar, em suas reflexões a respeito da lexicalização e gramaticalização, Lehmann (2002) apresenta, com fim de definir esses processos, a diferença entre gramática e léxico. Aqui, a distinção se daria pela regularidade e irregularidade, bem como pela maneira como as estruturas são tratadas: a gramática diz respeito aos sinais que se formam regularmente e que são tratados analiticamente, o léxico, por outro lado, se ocupa daqueles que são formados irregularmente e tratados de maneira holística. Isso significa dizer que, para se identificarem as categorias do léxico, são necessárias estas duas abordagens: a holística (*holistic*) e a analítica (*analytic*).

Na abordagem analítica, consideram-se “cada parte do objeto e a sua contribuição, no que diz respeito a sua natureza e função, para o conjunto, a fim de que, assim, se faça uma representação mental do todo” (LEHMANN, 2002, p. 2). Na abordagem holística, por outro lado, não se consideram as partes: primeiramente compreende-se o todo. Tratar um item de forma holística “significa tratar o objeto como um item de um inventário, um item lexical” (LEHMANN, 2002, p. 3).

Biderman (1996) defende que o léxico é o responsável pela “estocagem da significação e dos conteúdos significantes da linguagem humana” (BIDERMAN, 1996, p. 27); ou seja, é por meio do léxico que os significados da língua são veiculados. A autora associa o desenvolvimento do léxico àquele responsável pela nomeação ocorrente nas línguas, bem como à noção de signo linguístico. Este último equivale a um item lexical, o qual é responsável pela maior parte do significado contido em uma mensagem.

Para além disso, o léxico é composto por lexemas, que são a “unidade abstrata do léxico” (BIDERMAN, 1996, p. 32). Estes podem ter valor lexical (palavras plenas) ou valor gramatical (palavras gramaticais, como morfemas). Os lexemas operam dentro do discurso como lexias, as quais podem ser simples (grafadas como uma unidade só no plano da escrita, como ‘casa’ ou ‘escola’) ou complexas (grafadas como uma sequência de unidades no plano da escrita, porém compreendendo um significado, como ‘dona de casa’ e ‘sala de jantar’).

## 2.2 O Léxico sob a perspectiva da Linguística Sistêmico-Funcional (LSF)

O léxico, nos estudos de viés sistêmico, é definido como o ponto mais delicado da gramática. Isso significa dizer que existe um contínuo lexicogramatical, que compreende tanto o polo da gramática quanto o polo do léxico (MATTHIESSEN, 1995; HALLIDAY; MATTHIESSEN, 2014). Essa ideia corrobora a de Tucker (1998), que, afastando-se de uma abordagem tradicional, a qual entende gramática e léxico como dois componentes separados, defende o fato de tanto a gramática quanto o léxico constroem, juntos, o significado; e é por conta disso que não poderiam ser separados. O léxico é estabelecido por meio da lexicalização, processo por meio do qual o item lexical passa para que um significado seja atribuído a uma palavra ou a um grupo de palavras (MATTHIESSEN *et al.*, 2010). Isso ocorre de acordo com a linguagem e a cultura local, de forma que nem sempre todas as línguas comportam conceitos compatíveis (TUCKER, 1998). Dessa forma, o léxico é mais suscetível a mudanças pela constante evolução de cada cultura, as quais introduzem novos conceitos e fazem com que o léxico se inove mais rapidamente que a gramática (MARTIN; ROSE, 2007).

Desse modo, nota-se o quanto o contexto pode influenciar as mudanças do léxico. Isso é afirmado pelo fato de, ao construir o campo/domínio experiencial, o léxico precisa se adequar às expectativas acerca das configurações do texto, como, por exemplo, o seu assunto e o seu público, de forma que seu caráter seja variável e adaptável conforme o tipo de texto (MARTIN; ROSE, 2007).

Isso, portanto, torna questionável até que ponto gramática e léxico configuram polos de um contínuo, uma vez que eles podem ser concebidos como os responsáveis pela instanciação de diferentes significados, intrínsecos ao tipo de texto em que aparecem. Para além disso, o léxico pode ser concebido ainda como uma variável permanente para a delimitação do campo/domínio experiencial para textos de processos sociosemióticos distintos, já que ele apresenta a capacidade de estabelecer as relações taxonômicas que constituem os textos (MARTIN, 1992).

A próxima seção compreende os passos metodológicos estabelecidos para o desenvolvimento deste trabalho.

### 3 Parâmetros metodológicos

Para alcançar o objetivo desta pesquisa, foi realizado um estudo de caso desenhado para identificar e descrever o comportamento do léxico em contextos de uso real da língua. Para isso, foram compilados cinco artigos acadêmicos a respeito de estudos sobre o Diabetes Mellitus, como também cinco matérias jornalísticas veiculadas pela Revista Superinteressante (Super Interessante, 2018) que abordam os avanços tecnológicos e científicos acerca desse mesmo domínio.

Os artigos acadêmicos compilados foram extraídos de periódicos da área das Ciências da Saúde classificados pela Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) em 2017 como A1 ou A2, publicados há, no máximo, cinco anos (XXX, 2018). Já as matérias coletadas foram publicadas no *site* da revista Superinteressante entre os anos de 2016 e 2018. Após



compilados, os artigos acadêmicos e as matérias foram armazenados e etiquetados de acordo com as regras de etiquetamento e armazenamento do *corpus* CALIBRA (Catálogo da Língua Brasileira) (cf. XXX, 2015a). Apesar de os textos terem sido compilados por completo, para a parte de anotação manual, foram utilizados as introduções dos artigos acadêmicos e o primeiro parágrafo de cada matéria da Revista Superinteressante, visto que são partes que compõem o primeiro macrotema (cf. MARTIN e ROSE, 2007) de cada tipo de texto. Já para a parte semi-automática, foram utilizadas as introduções dos artigos acadêmicos e as matérias de popularização da ciência por completo.

A seleção dos textos pertencentes aos discursos acadêmico e midiático foi orientada pela sua classificação conforme o processo sociosemiótico (cf. HALLIDAY & MATTHESEN, 2014) ao qual cada tipo de texto está vinculado. Os artigos acadêmicos compilados estão no escopo do processo sociosemiótico *explorar*, que comporta textos em que a linguagem é utilizada na criação de novos significados que são postos em negociação com outros membros da comunidade científica. As matérias jornalísticas extraídas estão no escopo do processo sociosemiótico *relatar*, sendo ele constituído por textos que descrevem fenômenos particulares ocorridos no mundo (MATTHESEN, 2013).

Por serem textos que tratam do Diabetes Mellitus, optou-se por utilizar textos científicos que abordassem as recentes descobertas sobre essa condição crônica, bem como textos que reportam e popularizam pesquisas científicas, traçando, assim, o domínio experiencial do Diabetes Mellitus. Partiu-se, portanto, do âmbito privado, composto por pesquisadores e seus pares, para o âmbito público, composto pela população que acessa esse tipo de discurso midiático (FULLER, 1998). Ambos os textos estão vinculados ao modo de produção escrito e monológico e podem ser extraídos semi-automaticamente da internet sem acesso restrito.

O domínio experiencial do Diabetes Mellitus – que compõe os textos desta pesquisa – se caracteriza como uma condição crônica considerada um dos grandes problemas de saúde pública do século XXI (FREITAS *et al.*, 2016). Devido a sua grande penetração social, o domínio experiencial do Diabetes Mellitus vem sendo construído em diferentes tipos de textos, perpassando vários processos sociosemióticos. Portanto, textos sobre o Diabetes Mellitus constituem-se de um domínio experiencial pervasivo, sendo de interesse público e linguístico (cf. XXX, 2014; XXX, 2016; XXX, 2018).

Para o desenvolvimento desta pesquisa, foram utilizados dois protocolos: i) protocolo de investigação lexical: processos sociosemióticos *explorar/relatar* e ii) protocolo de investigação lexical: processo de generalização. Cada protocolo representa um passo metodológico desta pesquisa para verificar o funcionamento do léxico a partir de testes distintos.

Inicialmente, utilizou-se o protocolo de investigação lexical: processos sociosemióticos *explorar/relatar*, a fim de averiguar o funcionamento do léxico no domínio experiencial do Diabetes Mellitus em dois processos sociosemióticos diferentes, *explorar* e *relatar*.

O primeiro parágrafo das cinco matérias jornalísticas, juntamente com seus títulos e linhas finas<sup>1</sup>, e as cinco introduções dos artigos acadêmicos foram segmentados em orações segundo os pressupostos teóricos da LSF. Uma vez que a ordem da oração é a ordem mais abstrata no estrato gramatical e comporta funções das três metafunções sistêmicas, ela foi escolhida para integrar a análise textual do primeiro protocolo de investigação.

Após a segmentação, as orações foram transferidas para duas planilhas eletrônicas, uma para as matérias jornalísticas que chamamos de “popularização da ciência” e outra para os artigos científicos que chamamos de “artigo acadêmico”. Nas planilhas, foram inseridos os sistemas da ordem da oração correspondentes às metafunções interpessoal, textual e ideacional, a saber, os sistemas de modo, tema textual, tema interpessoal, tema ideacional e transitividade. Esses sistemas foram utilizados para a realização da análise textual feita manualmente.

Em seguida, verificou-se o número de ocorrências das categorias analisadas em cada oração com o objetivo de examinar os resultados para cada processo sociosemiótico e, por fim, compará-los observando se as funções lexicogramaticais operariam da mesma forma nos processos explorar e relatar, tendo eles o mesmo domínio experiencial que representaria o mesmo campo que, por sua vez, seria organizado por um conjunto lexical específico de cada texto.

Desta maneira, a mudança no processo sociosemiótico dos textos foi utilizada como ferramenta de verificação do funcionamento do léxico, enquanto o domínio experiencial funcionou como uma variável permanente.

Para o protocolo de investigação lexical: processo de generalização, os mesmos textos foram utilizados. Estes foram salvos em arquivos do bloco de notas distintos e importados para o *software* concordanciador AntConc (LAURENCE ANTHONY, 2018). Em seguida, uma lista de *stopwords* (GITHUB, 2018), selecionada da *internet*, foi inserida no *software*, a fim de apresentar os itens lexicais mais frequentes nos textos de popularização da ciência e nos artigos acadêmicos – ressalta-se aqui o fato de que esses itens lexicais deveriam estar relacionados ao domínio experiencial do Diabetes Mellitus. No total, quarenta itens lexicais foram inseridos em dois quadros distintos, um para o processo sociosemiótico explorar e outro para o processo sociosemiótico relatar, com o objetivo de aplicar o protocolo correspondente à generalização lexical.

Posteriormente, verificou-se a forma como cada item poderia ser generalizado por meio da aplicação de uma rede taxonômica de classe ou composição. Os itens selecionados passaram, então, pelo processo taxonômico e foram, assim, generalizados. Em seguida foram classificados como itens lexicais em conjunto com as relações de classe e composição que produziram no texto.

A seguir, os resultados obtidos a partir da aplicação dos protocolos lexicais são abordados.

#### 4 A aplicação dos protocolos lexicais: o estudo de caso

---

<sup>1</sup> Linha fina corresponde à linha logo abaixo do título e/ou manchete da notícia (XXX, 2018).

Como apontado na metodologia do presente trabalho, dois protocolos de investigação lexical foram desenvolvidos a fim de que a construção lexical do domínio experiencial do Diabetes Mellitus em dois tipos diferentes de textos pudesse ser verificada. Esta seção detalha esses protocolos e apresenta os resultados obtidos a partir do estudo de caso.

#### **4.1 Protocolo de investigação lexical: processos sociossemióticos explorar e relatar**

Neste primeiro protocolo, o objetivo foi identificar o impacto da mudança do processo sociossemiótico no que diz respeito ao léxico e à gramática, uma vez que estes operam juntos para construir textos e produzir significados (HALLIDAY & MATTHIESSEN, 2014). Para isso, os sistemas de modo, tema textual, tema interpessoal, tema ideacional e transitividade foram anotados manualmente para uma amostra de textos do *corpus*.

Em relação ao sistema de modo, a anotação manual das orações segmentadas a partir dos textos localizados no escopo do processo sociossemiótico relatar apresentaram maior variabilidade de realizações desse sistema, uma vez que houve realizações de imperativo: jussivo (3%) e interrogativo (1%). O modo declarativo pode ser considerado o mais frequente, tanto para as orações anotadas sob o processo sociossemiótico relatar (96%) quanto para aquelas anotadas sob o processo sociossemiótico explorar (82%), sendo esse o único modo selecionado para as orações anotadas sob o processo sociossemiótico explorar, já que as outras orações não selecionaram nenhuma opção do sistema de modo, visto que eram orações não-finitas. Esses dados podem indicar que os textos localizados no escopo do processo sociossemiótico relatar são passíveis de apresentarem maior variabilidade de outras opções do modo para além do declarativo.

No que diz respeito ao sistema de transitividade, a maioria das orações segmentadas para os textos dos dois processos sociossemióticos foi anotada como oração material (60% - relatar e 48% - explorar). Em seguida, a maior ocorrência foi de orações relacionais dos tipos identificativas e atributivas.

Para os textos localizados no escopo do processo sociossemiótico explorar, esses dois tipos de orações relacionais apresentaram o mesmo número de ocorrências (17% para cada), enquanto que para os textos localizados no escopo do processo sociossemiótico relatar, as orações relacionais atributivas (14%) foram mais frequentes que as identificativas (8%).

Em sequência, estão as orações mentais e verbais. Para os textos no escopo do processo sociossemiótico explorar as orações mentais (10%) são mais frequentes que as orações verbais (7%), ao passo que para os textos no escopo do processo sociossemiótico relatar, esses dois tipos de orações apareceram emparelhadas com o mesmo número de realizações (5% para cada). Por fim, estão as orações existenciais, que apresentaram realização apenas para os textos localizados no escopo do processo sociossemiótico explorar (1%).

Essas configurações indicam que as matérias de popularização da ciência, textos localizados no escopo do processo sociossemiótico relatar, apresentaram mais orações relacionais atributivas, diferentemente dos artigos acadêmicos, textos localizados no escopo do processo sociossemiótico

explorar, que apresentaram a mesma quantidade de orações relacionais atributivas e identificativas. Isso evidencia que os artigos acadêmicos sob o domínio experiencial do Diabetes Mellitus tendem a construir relações mais específicas, uma vez que aborda os resultados encontrados em uma determinada pesquisa, enquanto as matérias de popularização da ciência sob o mesmo domínio experiencial enfocam as relações mais gerais que podem ser aplicadas a todas as pessoas com Diabetes Mellitus.

Os últimos sistemas anotados para a aplicação do protocolo de investigação lexical: processos sociossemióticos explorar/relatar compreendem os sistemas de tema. Para o primeiro sistema, o de tema textual, a variabilidade de opções selecionadas para as orações dos textos no escopo do processo sociossemiótico explorar foi semelhante às selecionadas para os textos no escopo do processo sociossemiótico relatar, ambos os textos apresentaram realizações de tema textual: relativo (21% – explorar e 8% – relatar), textual: conjuntivo (20% – explorar e 12% – relatar), textual: continuativo (5% – explorar e 2% – relatar) e também orações que não selecionaram tema textual (54% – explorar e 78% – relatar).

Para os textos localizados no escopo do processo sociossemiótico explorar, houve ainda uma oração que apresentou recursividade do tema textual. Em relação às orações anotadas que tiveram alguma realização de tema textual, esses resultados evidenciam que as realizações de tema textual: conjuntivo nos textos do processo sociossemiótico relatar podem aparecer com mais frequência que as realizações de tema textual: relativo, uma vez que o primeiro apresentou maior número de ocorrências que o segundo.

Já em relação ao sistema de tema interpessoal, a maior parte das orações anotadas para os textos localizados no escopo do processo sociossemiótico explorar, bem como a maioria das orações anotadas para os textos localizados no escopo do processo sociossemiótico relatar não selecionaram tema interpessoal (99% – explorar e 96% – relatar). Entre as orações anotadas para os textos localizados no escopo do processo sociossemiótico ‘explorar’, uma apresentou realização de tema interpessoal: avaliação: modo. Entre as orações anotadas para os textos localizados no escopo do processo sociossemiótico ‘relatar’, houve duas que apresentaram realização de tema interpessoal, sendo uma de tema interpessoal: avaliação: modo e outra de tema interpessoal: avaliação: comentário.

Esses dados apontam que as matérias jornalísticas retiradas da revista Superinteressante, textos localizados no escopo do processo sociossemiótico relatar, podem apresentar mais variabilidade de tema interpessoal em comparação com os artigos acadêmicos, textos localizados no escopo do processo sociossemiótico explorar.

O último sistema anotado compreende o de tema ideacional, a maior parte das orações que constituem os textos no âmbito do processo sociossemiótico explorar quanto a maior parte das orações dos textos no âmbito do processo sociossemiótico relatar selecionaram o tema ideacional: *default*: declarativo (64% – explorar e 71% – relatar). Em seguida, apareceram as orações que não selecionaram tema ideacional, por serem orações não-finitas, (18% – explorar e 14% – relatar). Em terceira posição, para as orações que compõem os textos dos dois processos sociossemióticos, está o tema

proeminente: perspectiva: intensificação (13% – explorar e 8% – relatar), responsável por identificar circunstâncias em posição temática.

Ainda, para o processo sociossemiótico explorar, as três últimas opções selecionadas foram, respectivamente, tema proeminente: nuclear: processo (3%), tema proeminente: nuclear: fonte (1%) e tema proeminente: nuclear: participante (1%). Já para o processo sociossemiótico relatar, essas três últimas opções selecionadas foram, respectivamente, tema *default*: imperativo (3%), tema proeminente: nuclear: processo (2%) e tema *default*: interrogativo (2%).

No que compreende o tema ideacional, essas configurações indicam que, para os textos selecionados dentro do escopo desses dois processos sociossemióticos, as diferenças estão localizadas nas opções selecionadas com menor frequência. Houve ainda as distinções entre tema proeminente e o tema *default*, este mais frequente para as orações anotadas sob o escopo dos textos do processo sociossemiótico relatar e aquele mais frequente para as orações anotadas sob o escopo dos textos do processo sociossemiótico explorar.

#### 4.2 Protocolo de investigação lexical: processo de generalização

O segundo protocolo foi responsável pelo processo de generalização em que as cinco matérias jornalísticas sobre o domínio experiencial do Diabetes Mellitus publicadas no *site* da revista Superinteressante, bem como as introduções dos artigos acadêmicos também sobre o domínio experiencial do Diabetes Mellitus publicados em revistas da área das Ciências da Saúde foram importadas em dois momentos diferentes para o *software* concordanciador AntConc. Isso permitiu que os itens lexicais com maior número de ocorrências fossem extraídos semi-automaticamente para cada um dos dois processos sociossemióticos.

Foram selecionados quarenta itens lexicais ao todo, sendo vinte para as matérias jornalísticas da revista Superinteressante e vinte para os artigos acadêmicos publicados em revistas da área das Ciências da Saúde. A identificação dos itens lexicais foi executada no protocolo por meio das operações de classificação e composição, uma vez que apenas itens desse tipo são passíveis de generalização (cf. MARTIN, 1992; HALLIDAY; MATTHIESSEN, 1999). É válido mencionar ainda que, para além do quesito número de ocorrências, os itens lexicais selecionados para a aplicação desse protocolo de generalização deveriam estar no escopo no domínio da saúde, já que o domínio experiencial dos textos dos dois processos sociossemióticos em questão é o Diabetes Mellitus.

Para os textos no escopo do processo sociossemiótico explorar, os itens lexicais com maior número de ocorrências relacionados ao domínio experiencial em questão são apresentados na primeira coluna do Quadro 1 a seguir. Já para os textos no escopo do processo sociossemiótico relatar, os itens lexicais com maior número de ocorrências relacionados ao domínio experiencial do Diabetes Mellitus são detalhados na segunda coluna do mesmo quadro.

**Quadro 1:** Itens lexicais relacionados ao domínio experiencial do Diabetes Mellitus extraídos dos textos no âmbito dos processos sociossemióticos explorar/relatar

Itens lexicais – Processo sociossemiótico Explorar	Número de ocorrências dos itens lexicais - Processo sociossemiótico Explorar	Itens lexicais – Processo sociossemiótico Relatar	Número de ocorrências dos itens lexicais - Processo sociossemiótico Relatar
Diabetes Mellitus	26	Diabetes	62
Doença	23	Doença	17
Autocuidado	20	Insulina	12
Indivíduos	20	Sangue	12
Saúde	17	Açúcar	12
Indígenas	15	Gordura branca	5
Tratamento	14	Células	8
Pés	11	Corpo	8
Atitude	8	Diabéticos	7
Complicações	8	Agulhas	6
Doenças Crônicas	8	Dieta	6
Estilo de vida	7	Injeções	5
Pacientes	7	Pele	5
Cuidados	6	Alimentação	4
Adesão	5	Caféina	4
Faixa etária	5	Droga	4
Diabéticos	4	Hormônio	3
Educação em saúde	4	Implante	3
Pé diabético	4	Bebidas açucaradas	2
Controle metabólico	3	Bebidas adoçadas	2

Fonte: Desenvolvida para fins deste trabalho.

Comparando as colunas do Quadro 1, observa-se que os itens lexicais que foram extraídos para o processo sociossemiótico explorar são em sua maioria diferentes dos itens lexicais que foram extraídos para o processo sociossemiótico relatar. Dentre os vinte itens lexicais selecionados para cada um dos processos sociossemióticos, apenas dois ocorreram em ambos os processos sociossemióticos: ‘doença’ e ‘diabéticos’.

Houve ainda itens lexicais compostos por dois e/ou três *tokens*<sup>2</sup>, constituindo um grupo nominal (cf. XXX, 2007, HALLIDAY & MATTHIESSEN, 2014), como ‘bebidas açucaradas’, ‘bebidas adoçadas’, ‘estilo de vida’, ‘faixa etária’, ‘educação em saúde’ e ‘doenças crônicas’. Isso se deve ao fato de que os *tokens* que formam esses grupos nominais apresentam um significado diferente quando aparecem sozinhos em um texto, e, quando estão juntos, realizam um outro significado, relacionado, neste último caso, ao domínio da Saúde, enfoque deste trabalho.

Por esse motivo, alguns itens formados por dois ou três *tokens* foram contabilizados neste estudo como um item lexical, que pode interagir com a

<sup>2</sup> *Token* é o “resultado do procedimento que consiste em totalizar o número de palavras existentes em um dado texto ou *corpus*” (VIANA, 2011, p. 38)

ordem da palavra e/ou do grupo, uma vez que esses *tokens* juntos operam com uma só função na ordem superior, a ordem do grupo.

Ainda com relação à aplicação desse protocolo de generalização, todos os vinte itens lexicais extraídos para cada um dos processos sociossemióticos em questão foram passíveis de generalização. Para isso, o significado de cada item lexical foi analisado pelo protocolo, averiguando-se sua capacidade de estabelecer relações no texto. A partir disso, aplicou-se uma rede taxonômica de classe ou composição para o item lexical em questão, mediante a busca de significados relacionados no texto, constituindo, assim, a aplicação do protocolo de generalização para os vinte itens lexicais extraídos para cada um dos dois processos sociossemióticos.

A seguir, o Quadro 2 detalha o resultado da aplicação do protocolo de investigação lexical: processo de generalização realizado para os itens lexicais extraídos dos textos no escopo do processo sociossemiótico explorar.

**Quadro 2.** Resultado da aplicação do protocolo de investigação lexical: processo de generalização – processo sociossemiótico explorar

<b>Itens lexicais – Processo sociossemiótico explorar</b>	<b>Aplicação Protocolo de investigação lexical: processo de generalização</b>
Diabetes Mellitus	É um tipo de Doença
Doença	É um tipo de alteração biológica do estado de saúde de um ser
Autocuidado	É um tipo de conjunto de ações ou procedimentos que cada indivíduo faz para manutenção da vida, saúde e bem-estar
Indivíduos	É um tipo de organismo único
Saúde	É um tipo de estado de equilíbrio físico e mental
Indígenas	É um tipo de população que se estabeleceu em um país antes do processo de colonização
Tratamento	É um tipo de ações empregadas para a cura de alguma doença
Estilo de vida'	É um tipo de estratificação da sociedade conforme os padrões de consumo, rotina, hábitos, etc.
Pés	É um tipo de membro inferior
Controle metabólico	É um tipo de monitorização, fiscalização
Atitude	É um tipo de comportamento
Complicações	É um tipo de acontecimento ou processo patológico

	que ocorre durante a evolução de uma doença
Doenças crônicas	É um tipo de doença que persiste por mais de seis meses
Pé diabético	É um tipo de complicação do diabetes
Pacientes	É um tipo de indivíduo que precisa de cuidados médicos
Cuidados	É um tipo de atenção especial
Adesão	É um tipo de aceitação
Faixa etária	É um tipo de intervalo de idades aproximadas
Diabéticos	É um tipo de indivíduo que sofre de diabetes
Educação em saúde	É um tipo de educação na área da saúde

Fonte: Desenvolvida para fins deste trabalho.

Todos os vinte itens lexicais extraídos dos textos do processo sociossemiótico explorar foram passíveis de generalização, mediante a aplicação de uma rede taxonômica de classe ou composição em conjunto com o significado encontrado para cada um dos itens lexicais detalhados no Quadro 2. Outra particularidade dos itens lexicais extraídos dos textos localizados no escopo do processo sociossemiótico explorar diz respeito à quantidade de itens compostos por mais de um *token* interagindo com a ordem da palavra e/ou do grupo nominal, como ‘Diabetes Mellitus’, ‘Estilo de vida’, ‘Doenças Crônicas’, ‘Pé diabético’, ‘Faixa etária’ e ‘Educação em saúde’. Esses dados sugerem que os itens lexicais presentes nos artigos acadêmicos, apresentam significados com maior densidade lexical (cf. URE, 1971) e precisam de mais de um *token* para construir um item.

Os itens lexicais formados por mais de um *token* podem construir sozinhos um significado diferente daquele realizado quando aparecem em conjunto com outro *token*, como, por exemplo, o *token* ‘crônica’ que sozinho pode ter um significado diferente dependendo de onde aparece, podendo a generalização ser realizada por meio da aplicação de uma rede taxonômica de classe – “é um tipo de texto”, todavia quando esse *token* coocorre com ‘doenças’, formando o item lexical ‘doenças crônicas’, o significado construído é distinto do primeiro.

O Quadro 3 apresenta os resultados obtidos a partir da aplicação do protocolo de investigação lexical: processo de generalização para os itens lexicais extraídos dos textos que estão no escopo do processo sociossemiótico relatar.

Quadro 3: Resultado da aplicação do protocolo de investigação lexical: processo de generalização – processo sociossemiótico relatar



Itens lexicais – Processo sociossemiótico relatar	Aplicação Protocolo de investigação lexical: processo de generalização
Diabetes	É um tipo de doença
Doença	É um tipo de alteração biológica do estado de saúde de um ser
Gordura branca	É um tipo substância encontrada no tecido adiposo
Insulina	É tipo de hormônio/remédio
Bebidas açucaradas	É um tipo de líquido com açúcar
Bebidas adoçadas	É um tipo de líquido com adoçante
Células	É um tipo de elemento presente em todo ser vivo
Corpo	É um tipo de conjunto formado por cabeça, tronco e membros
Sangue	É um tipo de líquido vermelho
Açúcar	É um tipo de substância industrializada doce
Cafeína	É um tipo de substância presente no café
Diabéticos	É um tipo de indivíduo que sofre de diabetes
Injeções	É um tipo de aplicação de substâncias
Pele	É um tipo de órgão que cobre o corpo
Agulhas	É um tipo de instrumento perfurante
Alimentação	É um tipo de abastecimento de substâncias necessárias para a vida
Dieta	É um tipo de regime alimentar
Droga	É um tipo de substância química
Hormônio	É um tipo de substância produzida por uma glândula
Implante	É um tipo de material inserido ou retirado do corpo do indivíduo

Fonte: Desenvolvida para fins deste trabalho.

Os resultados da aplicação do processo de generalização apresentados no Quadro 3 mostram que todos os vinte itens lexicais extraídos dos textos no escopo do processo sociossemiótico relatar foram passíveis de generalização, portanto identificados como itens lexicais.

Em comparação com os itens extraídos dos textos do processo sociossemiótico explorar, a variedade de itens lexicais que apresentaram mais de um *token* foi menor para os textos do processo sociossemiótico relatar, uma vez que somente dois itens lexicais dentre os vinte selecionados a partir dos textos do processo sociossemiótico relatar apresentaram essa configuração – ‘bebidas açucaradas’ e ‘bebidas adoçadas’. Isso pode indicar que os textos de popularização da ciência, localizados no escopo do processo sociossemiótico relatar, apresentaram densidade lexical menor que os artigos acadêmicos, localizados no escopo do processo sociossemiótico explorar.

Finalizada a apresentação dos resultados obtidos para cada um dos dois protocolos de investigação lexical, a próxima subseção é responsável por detalhar o impacto desses protocolos no que compreende o estudo do léxico.

#### **4.3 Síntese da construção do léxico em textos localizados no escopo dos processos sociossemióticos explorar e relatar sob o domínio experiencial do Diabetes Mellitus**

Tendo em mente os resultados encontrados mediante a aplicação do primeiro protocolo de investigação lexical, protocolo de investigação lexical: processos sociossemióticos explorar/relatar, que englobou a anotação manual dos sistemas de modo, tema textual, tema interpessoal, tema ideacional e a função do Processo do sistema de transitividade em textos de dois tipos diferentes de processos sociossemióticos com domínios experienciais iguais, é possível afirmar que os sistemas, funções estruturais e classes (protocolo de investigação: processos sociossemióticos explorar/relatar) e as generalizações e especificidades das relações lexicais (protocolo de investigação: processo de generalização) são responsáveis pela construção de um texto, determinando o macro processo sociossemiótico a que esses textos podem estar localizados, uma vez que as opções gramaticais selecionadas para as introduções dos artigos acadêmicos no escopo do processo sociossemiótico explorar e as selecionadas para as matérias de popularização da ciência no escopo do processo sociossemiótico relatar foram instanciadas de maneiras distintas.

Ainda com relação ao léxico, a aplicação do segundo protocolo – protocolo de investigação lexical: processo de generalização, que enfocou a extração semi-automática de itens lexicais presentes nos textos dos dois processos sociossemióticos em razão do processo de generalização – apontou que, para um item ser considerado lexical, é necessário que ele seja passível de ser generalizado mediante o estabelecimento de relações taxonômicas de identificação, classificação e generalização (MARTIN, 1992; MARTIN e ROSE, 2007).

Esse protocolo mostrou ainda indícios de que há itens considerados lexicais que podem ser formados por mais de um *token*. Os resultados do protocolo de investigação lexical: processo de generalização evidenciaram que há duas realizações possíveis para os itens lexicais que englobam essa configuração de mais de um *token*. A primeira realização compreende os itens lexicais formados por um Ente seguido de um Classificador, como, por exemplo, ‘Doenças Crônicas’, em que ‘Doença’ é o Ente e ‘Crônicas’ o Classificador. A segunda realização diz respeito à criação de um Ente novo,

como em 'Educação em Saúde', em que um significado novo é estabelecido a partir da junção de três *tokens* (cf. HALLIDAY e MATTHIESSEN, 1999). A combinação dos dois protocolos mostrou que a utilização do domínio experiencial do Diabetes Mellitus como uma variável permanente possibilitou que as diferenças consideradas gramaticais fossem apontadas para cada um dos tipos de processos sociosemióticos. Além disso, os resultados do segundo protocolo, responsável pelo processo de generalização, evidenciaram que os itens lexicais com maior número de ocorrências para os dois tipos de processos sociosemióticos foram, em sua maior parte, diferentes (cf. Quadro 1), mesmo com a variável do domínio experiencial do Diabetes Mellitus sendo mantida. Esses resultados podem indicar que a construção de taxonomias pelos artigos acadêmicos, no âmbito do processo sociosemiótico explorar, e pelos textos de popularização da ciência, no âmbito do processo sociosemiótico relatar, se dá de maneiras distintas, embora os dois processos sociosemióticos estejam sob o mesmo domínio experiencial do Diabetes Mellitus.

De forma mais detalhada, a taxonomia dos textos do processo sociosemiótico explorar se dá de maneira mais específica por dois motivos principais, um gramatical e outro lexical. O gramatical compreende o número de orações relacionais identificativas que, em comparação com os resultados obtidos para os textos de popularização da ciência, são mais frequentes para os artigos acadêmicos, apontando que as relações taxonômicas estabelecidas nesse último tipo de texto são específicas para a amostra do estudo em questão, não gerais para toda a população com Diabetes Mellitus, como é o caso dos textos de popularização da ciência. Já o motivo lexical diz respeito aos itens lexicais que precisaram de mais de um *token* para conseguirem instanciar um único significado, podendo interagir com a ordem do grupo nominal, o que aumenta a densidade lexical dos artigos acadêmicos em comparação com os textos de popularização da ciência.

## 5 Considerações finais

Tendo como ponto de partida o papel da combinação entre sistemas gramaticais fechados e conjuntos lexicais abertos na organização do sistema linguístico, este artigo buscou oferecer um maior entendimento das relações entre itens lexicais e a forma como estas contribuem para a produção de significado do texto. Partindo de um conjunto de premissas estabelecidas em pesquisas anteriores de base sistêmico-funcional (HALLIDAY e MATTHIESSEN, 1999; MARTIN e ROSE, 2007), o artigo concebeu o léxico como um recurso do sistema linguístico para a produção de significado que desempenha um papel no texto, relacionando-o, assim, ao seu campo/domínio experiencial e limitando-o ao processo sociosemiótico do qual faz parte.

Assim, o artigo objetivou apresentar um estudo de caso no qual a construção do conjunto lexical em um tipo de texto específico e, em seguida, contrastar essa construção com a de outro tipo. O estudo de caso foi realizado em um *corpus* de textos de matéria jornalística de popularização da ciência de domínio experiencial do Diabetes Mellitus, e o contraste feito com artigos acadêmicos de mesmo domínio experiencial.

Para o estudo de caso, foram desenvolvidos dois protocolos de investigação. O primeiro para investigar a constituição linguística dos processos sociosemióticos dos textos analisados; o segundo para investigar as propriedades de construção de relações lexicais, por meio da generalização. Os protocolos se mostraram adequados para o estudo, uma vez que conseguiram identificar, classificar e relacionar os itens lexicais pertencentes ao léxico de cada texto e relacioná-los com o campo/domínio experiencial por eles realizado. Os resultados demonstraram a forma pela qual a construção de taxonomias pelos textos de popularização da ciência e artigos acadêmicos se dá de maneiras distintas, ainda que ambos possuam o mesmo domínio experiencial.

Ganha destaque o fato de que a taxonomia dos textos do processo sociosemiótico explorar se dá de maneira mais específica por dois motivos principais, um gramatical e outro lexical. As relações taxonômicas do processo explorar apresentam maior especificidade (i) gramatical, devido a uma maior ocorrência de orações relacionais identificativas e (ii) lexical, devido a um número maior de itens lexicais com mais *tokens*. A especificidade gramatical nos artigos acadêmicos faz com as relações taxonômicas se restrinjam aos estudos apresentados nos textos em questão (ao passo que estas são generalizadas nos textos de popularização da ciência). A especificidade lexical se deve ao aumento da densidade lexical dos artigos acadêmicos (relativamente à popularização da ciência).

Essas duas motivações revelaram um aspecto muito importante da forma como o sistema linguístico organiza os recursos que são empregados pelos sistemas fechados (no presente estudo representados pelos sistemas gramaticais da oração) e conjuntos abertos (aqui representados pelas relações taxonômicas dos itens lexicais mais frequentes). Elas demonstram como esses dois aspectos do sistema linguístico se relacionam. A especialização do artigo acadêmico se dá por meio de orações identificativas do ponto de vista gramatical e, concomitantemente, pela construção de itens lexicais mais complexos, com maior classificação e composição.

Com isso, o artigo oferece uma contribuição mostrando a integração entre recursos sistêmicos e lexicais, bem como a forma como trabalham em conjunto para a construção do texto e realização linguística do domínio experiencial. Dessa forma, os protocolos poderão ser empregados em estudos futuros que busquem descrever novas propriedades de conjuntos lexicais. Além dos protocolos, os resultados do artigo como um todo apontam ainda caminhos para pesquisas futuras acerca do papel que o léxico desempenha na organização do sistema linguístico. Dentre eles, destaca-se a hipótese de que a frequência repetida de itens lexicais em textos de mesmo tipo – i.e., mesmo domínio experiencial e pertencentes ao mesmo processo sociosemiótico – determinaria significados mais generalizados que extrapolariam o próprio texto, estabelecendo, assim, relações de classificação e composição intertextuais. Igualmente, a frequência de relação entre itens poderia ser investigada a fim de se verificar em que medida essas relações também são expandidas/generalizadas para outros textos do mesmo tipo, bem como para textos de tipos diferentes, pervadindo, portanto, processos sociosemióticos distintos.

**REFERÊNCIAS**

- ANTHONY, Laurence. **Lawrence Anthony Website (AntConc)**. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/index.html>>. Acesso em: 21/10/2018
- BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey, CONRAD, Susan; FINEGAN, Edward. **Longman Grammar of Spoken and Written English**. England: Pearson education, 1999.
- BIDERMAN, Maria Tereza Camargo. Léxico e vocabulário fundamental. **Alfa**, São Paulo, v. 40, p. 27-46, 1996. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/107739/ISSN1981-5794-1996-40-27-46.pdf?sequence=1&isAllowed=y>>. Acesso em: 20/10/2018
- BRINTON, Laurel J. TRAUOGOTT, Elizabeth C. **Lexicalization and Language Change**. Cambridge: Cambridge University Press, 2005. 207 p.
- CHOMSKY, N. **Aspects of the thoery of syntax**. Cambridge: MIT Press, 1964.
- CHOMSKY, Noam. Remarks on nominalization. In: JACOBS, R. A. & ROSENBAUM, P. S., ed. **Readings in English transformational grammar**. Waltham, Mass., Ginn. 1970. p. 184-221.
- DI SCIULLO, A-M; WILLIAMS, E. **On the definition of word**. Cambridge: MIT, Press, 1987.
- FREITAS, G.A.; SOUZA, M. C. C.; LIMA, R. C. Prevalência de diabetes mellitus e fatores associados em mulheres indígenas do Município de Dourados, Mato Grosso do Sul, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 32, n. 8, p. 1-12, 2016.
- GITHUB**. Disponível em: <https://github.com/stopwords-iso/stopwords-pt>. Acesso em: 21/10/2018
- HALLIDAY, M. A .K. 'Text as Semantic Choice in Social Contexts'. In: Van DIJK, T.; PETOFI, J. (Eds.) **Grammars and Descriptions**. Berlin/ New York: de Gruyter. 1977. p. 176-225.
- HALLIDAY, M. A. K. On grammar. In: HALLIDAY , M. A. K. **The collected works of M.A.K. Halliday**. v. 1. London: Continuum, 2002.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. **Construing experience as meaning: a language based approach to cognition**. London: Cassell,, 1999.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. 2014. **An introduction to functional grammar**. London, 2014.
- HASAN, R. **Ways of Saying: Ways of Meaning**. [S.l.]: Cassell, 1996.
- SUPERINTERESSANTE**. Disponível em: <https://super.abril.com.br/> Acesso em: 21/10/2018

- MARTIN, J. R. **English Text System and Structure**. Philadelphia: John Benjamins Publishing Company, 1992.
- LEHMANN, Christian. New reflections on grammaticalization and lexicalization. In: WISCHER, Ilse & DIEWALD, Gabriele (eds.), **New reflections on grammaticalization**. Amsterdam: John Benjamins, 2002, p. 1-18.
- MARTIN, J.; ROSE, D. **Working with discourse: meaning beyond the clause**. 2. ed. London: Continuum, 2007.
- MARTIN, J. **Systemic functional grammar: a next step into the theory – axial relations**. Beijing: Higher Education Press, 2013.
- MATTHIESSEN, C. **Lexicogrammatical Cartography: English Systems**. Tokyo: International Language Sciences Publishers, 1995.
- MATTHIESSEN, C. Applying systemic functional linguistics in healthcare contexts. **Text and Talk**, v. 33, n. 4–5, p. 437–466, 2013.
- MATTHIESSEN, C.; TERUYA, K.; LAM, M. **Key terms in systemic functional linguistics**. London and New York: Continuum, 2010.
- QUIRK, R. LEECH, G; GREENBAUM, S. 1985. **A Comprehensive Grammar of the English Language**. London: Longman, 1985.
- RÖMER, U.; SCHULZE, R. **Exploring the Lexis-Grammar Interface**. Philadelphia: John Benjamins Publishing Company, 2009.
- RÖMER, Ute; SCHULZE, Rainer. Introduction: Zooming in. In: RÖMER, Ute; SCHULZE, Rainer. **Exploring the Lexis-Grammar Interface**. John Benjamins Publishing Company: Amsterdã, 2006.
- SAUSSURE, F. **Curso de Linguística Geral**. São Paulo: Cultrix, 2006.
- TUCKER, G. H. 1998. **The Lexicogrammar of Adjectives A systemic functional approach to lexis**. Londres: Cassel, 1998.
- URE, J. Lexical density and register differentiation. 1971. In: PERREN, G. E.; TRIM, I. L. M. (Ed.). **Applications of linguistics**. Cambridge: Cambridge University Press, 1971. p. 443-452.

*Recebido em julho de 2019.*

*Aprovado em setembro de 2019.*

*Publicado em setembro de 2019.*

## **SOBRE OS AUTORES**

**Giacomo Patrocínio Figueredo** é Doutor em Estudos Linguísticos pela Universidade Federal de Minas Gerais (2011). Pós-doutor em Linguística pela Universidade de Sydney (2016). É professor adjunto da Universidade Federal de Ouro Preto. Tem experiência na área de Linguística, atuando principalmente nos

---

seguintes temas: Tipologia, Estudos Multilíngues, Estudos da Tradução, Teoria Sistêmico-Funcional.

ORCID ID: <http://orcid.org/0000-0002-0967-244X>

E-mail: [giacomojakob@gmail.com](mailto:giacomojakob@gmail.com)

**Júlia Santos Nunes Rodrigues** é doutoranda em Estudos Linguístico do Programa de Pós-graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais. Mestre em Estudos Linguísticos pela Universidade Federal de Minas Gerais. Graduada em Letras pela Universidade Federal de Minas Gerais. Apresenta experiência na área de Linguística Aplicada especialmente no campo dos Estudos da Tradução e nas pesquisas que envolvem a Teoria Sistêmico-Funcional. Tem experiência na área de tradução, adaptação e adequação de questionários da área das Ciências da Saúde e na área de aprendizagem de máquina voltada para a Análise de Sentimentos.

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

E-mail: [julya\\_rodrigues@hotmail.com](mailto:julya_rodrigues@hotmail.com)

**Francieli Silvéria Oliveira** é doutoranda em Estudos Linguísticos do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais (UFMG). Mestre em Linguística Aplicada pela Universidade Federal de Minas Gerais (UFMG) em 2018. Graduada em Letras pela Universidade Federal de Ouro Preto (UFOP) em 2015. Atua na área de Linguística, trabalhando principalmente com os seguintes temas: análise contrastiva, modelagem linguística, variação linguística, sistema de mensagem, perfil metafuncional e Teoria Sistêmico Funcional.

ORCID ID: <http://orcid.org/0000-0002-4895-1919>

E-mail: [franciellysilveria@hotmail.com](mailto:franciellysilveria@hotmail.com)

**Ricardo José Alves** é doutorando em Estudos Linguísticos pela Universidade Federal de Minas Gerais (UFMG) e Mestre em Letras: Estudos da Linguagem pela Universidade Federal de Ouro Preto (UFOP). Graduado em Letras: Licenciatura em Língua Portuguesa (2014) e Bacharelado em Linguística (2016) pela Universidade Federal de Ouro Preto (UFOP). Tem experiência em Linguística, mais especificamente na de Descrição Linguística de base sistêmico-funcional. Atua nos seguintes temas: Estudos Multilíngues, Teoria Sistêmico-Funcional, Ensino de Língua Portuguesa e Revisão de Textos.

E-mail: [alvesricardoj@gmail.com](mailto:alvesricardoj@gmail.com)