

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Faculdade de Letras

Programa de Pós-graduação em Estudos Linguísticos

Júlia Santos Nunes Rodrigues

**O LÉXICO COMO UM RECURSO LINGUÍSTICO PARA A PRODUÇÃO DE
SIGNIFICADO: UMA PROPOSTA DE CARACTERIZAÇÃO SISTÊMICO-
FUNCIONAL**

Belo Horizonte

2022

Júlia Santos Nunes Rodrigues

**O LÉXICO COMO UM RECURSO LINGUÍSTICO PARA A PRODUÇÃO DE
SIGNIFICADO: UMA PROPOSTA DE CARACTERIZAÇÃO SISTÊMICO-
FUNCIONAL**

Versão final

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutora em Linguística Aplicada.

Área de Concentração: Linguística Aplicada

Linha de pesquisa: Estudos da Tradução – 3B

Orientadora: Profa. Dra. Adriana Silvina Pagano

Belo Horizonte

2022

R696l

Rodrigues, Júlia Santos Nunes.

O léxico como um recurso linguístico para a produção de significado [manuscrito] : uma proposta para caracterização sistêmico - funcional / Júlia Santos Nunes Rodrigues – 2022. 149 f., enc.: il., tabs, p&b.

Orientadora: Adriana Silvina Pagano.

Área de concentração: Linguística Aplicada.

Linha de Pesquisa: Estudos da Tradução - 3B.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 110-112.

Anexos: f. 113-149.

1. Tradução e interpretação – Teses. 2. Lexicologia – Teses. 3. Linguística de corpus – Teses. 4. Traduções – Estudo e ensino – Teses. 5. Funcionalismo (Linguística) – Teses. I. Pagano, Adriana Silvina. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 418.02



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

O LÉXICO COMO UM RECURSO LINGUÍSTICO PARA A PRODUÇÃO DE SIGNIFICADO: UMA PROPOSTA DE CARACTERIZAÇÃO SISTÊMICO-FUNCIONAL

JÚLIA SANTOS NUNES RODRIGUES

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA APLICADA, linha de pesquisa Estudos da Tradução.

Aprovada em 31 de março de 2022, pela banca constituída pelos membros:

Prof(a). Adriana Silvina Pagano - Orientadora

UFMG

Prof(a). Giacomo Patrocínio Figueredo

UFOP

Prof(a). Igor Antônio Lourenço da Silva

UFU

Prof(a). Leonardo Pereira Nunes

UFMG

Prof(a). Kícila Ferregueti de Oliveira

UFMG

Belo Horizonte, 31 de março de 2022.



Documento assinado eletronicamente por **Adriana Silvina Pagano, Professora do Magistério Superior**, em 01/04/2022, às 09:55, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Igor Antônio Lourenço da Silva, Usuário Externo**, em 01/04/2022, às 10:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leonardo Pereira Nunes, Professor do Magistério Superior**, em 01/04/2022, às 12:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Kicila Ferregueti de Oliveira, Professora Magistério Superior-Substituta**, em 01/04/2022, às 17:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Giacomo Patrocínio Figueredo, Usuário Externo**, em 05/04/2022, às 09:04, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1324564** e o código CRC **3838E12B**.

Bravery was not an identity as much as a choice.

Hibbert (2019)

AGRADECIMENTOS

À Deus, por ter me concedido saúde e discernimento para que eu chegasse até aqui.

Aos meus pais, não tenho nem palavras para expressar o tamanho da minha gratidão por ter vocês na minha vida. Este doutorado coroa todo esforço material e imaterial que vocês sempre dedicaram a mim. Muito obrigada por tanto!

À Kícila e Bárbara, minhas amigas-irmãs, por vibrarem com as minhas conquistas e me escutarem sempre. Obrigada especialmente à Kícila, por toda convivência fora e dentro da academia, sua amizade-irmandade tornou o caminho até aqui mais leve.

Ao Grupo Clandestino, sobretudo aos seus cofundadores, por me incentivarem a ir além do que eu imaginava ser capaz.

À Professora Adriana Pagano, pela orientação e pelo aprendizado não só neste trabalho, mas desde o início da minha trajetória acadêmica, lá na iniciação científica.

À FAPEMIG, por ter apoiado financeiramente parte desta pesquisa.

Aos colegas do LETRA; principalmente à Fran, com quem pude compartilhar muitas das angústias e alegrias deste doutorado.

RESUMO

Esta tese se insere nos estudos do léxico de base sistêmico-funcional, tendo como finalidade principal identificar os mecanismos de organização e de funcionamento do léxico e do item lexical no escopo do português brasileiro a partir de uma perspectiva de complementariedade entre léxico e gramática (HALLIDAY, MCINTOSH e STREVENS, 1964; MARTIN, 1992; HALLIDAY, 2002; FIGUEREDO, 2007; 2011). Nessa perspectiva, o léxico é visto como um conceito teórico e passa a ser investigado por meio de uma abordagem própria, adotando a gramática como um complemento, não como o ponto de partida. Tem-se aqui como motivação o fato de que as concepções acerca do léxico modular – concebido como um preenchimento de estruturas gramaticais – ou léxico contíguo – entendido como o ponto mais delicado da gramática – carecerem de um aprofundamento maior no que diz respeito aos processos internos sobre a dimensão do léxico, bem como sobre as relações dos itens lexicais para além da gramática. Subsidiada por ferramentas originárias da Linguística de Corpus, por um corpus de artigos acadêmicos sobre “autocuidado em diabetes mellitus” escritos em português brasileiro e publicados na última década, por um corpus de referência para gerar uma lista de itens utilizados no desenvolvimento e na aplicação dos protocolos de investigação lexical e por um corpus de suporte para contrastar os resultados aqui obtidos, a presente pesquisa explora dois processos de transformação lexicais, pelos quais o léxico se integra aos sistemas linguísticos, quando relacionado ao eixo paradigmático no estrato gramatical, e ao contexto através de um processo denominado de co-extensividade, permitindo que o léxico seja tratado como uma das dimensões da linguagem. Especificamente, nesta pesquisa, investigam-se os processos de transformação da gramática e do contexto, incluindo, principalmente, o desenvolvimento e a aplicação de protocolos de investigação lexical. Esses protocolos compreendem os recursos metodológicos utilizados neste estudo para verificar os processos internos do léxico, sobretudo no que diz respeito à emergência de itens lexicais formados por mais de um componente lexical, bem como no que envolve a formação de itens lexicais através da exploração das porções que os constituem. Uma análise mais abrangente dos resultados obtidos no presente estudo revela ainda uma alternativa metodológica para trabalhar com o léxico de forma teórica, que examine, por meio dos protocolos de investigação lexical, como ocorre o funcionamento e a organização dos itens lexicais na língua, o que tende a gerar contribuições para a teoria sistêmico-funcional, assim como para os estudos da tradução no que diz respeito à tradução automática ou semiautomática e aos estudos linguísticos baseados em corpora.

PALAVRAS-CHAVE: Co-extensividade. Item lexical. Relações internas do léxico. Escala de ordens da gramática. Contexto de situação. Linguística de Corpus.

ABSTRACT

This thesis is part of the lexis' studies within the systemic-functional linguistics (SFL). The main purpose of this research is to identify the principles of lexis and lexical item organization and operation in the scope of Brazilian Portuguese under a complementary perspective between lexis and grammar (HALLIDAY, MCINTOSH e STREVENS, 1964; MARTIN, 1992; HALLIDAY, 2002; FIGUEREDO, 2007; 2011). In this perspective, lexis is seen as a theoretical concept and is investigated through its own approach, adopting grammar in a complementary view, not as its origin. Problems related to modular lexis approach – lexis as filling for grammatical structures – or contiguous approach – lexis as the most delicate grammar – motivate this study, in the sense that lexis needs an explanation about how its internal processes work regarding lexis as a dimension, as well as the relations between lexical items beyond the grammatical stratum. Supported by Corpus Linguistics' tools, a corpus of academic papers about “diabetes mellitus self-care”, written in Brazilian Portuguese and published in the last decade, a reference corpus to generate a list of items used in the development and application of the lexical probes, and a supporting corpus to contrast the results obtained under the investigations of lexical probes, the present research explores two transformation processes, through which the lexis is integrated into linguistic systems, when it is related to the paradigmatic axis in the grammatical stratum, and to the context through a process called co-extensiveness, letting lexis to be treated as one of the language dimensions. Specifically, in this research, the processes of grammar and context transformations are investigated, including mainly the development and application of lexical probes. These lexical probes could be seen as methodological resources created within this study to analyze the lexis internal processes, especially, with regard to the emergence of lexical items composed by more than one component and the formation of lexical items through the exploration of the portions that constitute them. A wider analysis of the results obtained in the present research also reveals a methodological alternative to work with lexis in a theoretical approach, which examines, through lexical probes, not only how lexical items are organized but also how they function in a language. These findings can contribute to systemic-functional theory (SFT), to translation studies regarding machine translation, and corpora-based linguistic studies.

Keywords: Co-extensiveness. Lexical item. Lexis internal processes. Grammar rank scale. Context of situation. Corpus linguistics.

LISTA DE QUADROS

QUADRO 1 – Exemplos de co-extensividade entre itens lexicais e a escala de ordens da gramática	31
QUADRO 2 – Lista de itens selecionados para a aplicação dos protocolos de investigação lexical (ProtinLex) – corpus de artigos acadêmicos	60
QUADRO 3 – Agrupamentos formados à direita e à esquerda de cada item da lista de palavras-chave	62
QUADRO 4 – Agrupamentos formados a partir de itens derivados da lista de palavras-chave	67
QUADRO 5 – Itens lexicais co-extensivos à ordem do grupo nominal: resultado default para o processo de mudança do ProtinLex:PTG	82
QUADRO 6 – Componentes lexicais co-extensivos à função do Classificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado default	85
QUADRO 7 – Componentes lexicais co-extensivos à função do Qualificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado default	90
QUADRO 8 – Componentes lexicais co-extensivos à função do Classificador + do Qualificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado default + desvio	94
QUADRO 9 – Componentes lexicais co-extensivos à função do Qualificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado de desvio	98

LISTA DE TABELAS

TABELA 1 – Anotação do sistema de TRANSITIVIDADE: contabilização das funções.. 68

LISTA DE FIGURAS

FIGURA 1 – A representação da propriedade de co-extensividade entre um item lexical e a escala de ordens da gramática.....	30
FIGURA 2 – A co-extensividade do item lexical “diabetes” em relação ao contexto e aos estratos linguísticos da semântica, lexicogramática e fonologia/grafologia.....	32
FIGURA 3 – Captura de tela dos tri-gramas quando "autocuidado" está localizado à esquerda.....	47
FIGURA 4 – Captura de tela dos tri-gramas quando "autocuidado" está localizado à direita.....	47
FIGURA 5 – Captura de tela da planilha eletrônica criada para a anotação do sistema de TRANSITIVIDADE.....	49
FIGURA 6 – Captura de tela da planilha eletrônica criada para isolar os componentes lexicais e estabelecer as relações COESIVAS LEXICAIS.....	51
FIGURA 7 – Representação do processo de mudança do ProtinLex:PTC.....	51
FIGURA 8 – Representação do processo de mudança do ProtinLex:PTC – variáveis do contexto de situação.....	53
FIGURA 9 – Captura de tela da planilha eletrônica criada para o desenvolvimento do ProtinLex:PTG.....	55
FIGURA 10 – Representação do processo de mudança do ProtinLex:PTG.....	56
FIGURA 11 – Representação ProtinLex:PTC – processo de mudança.....	72
FIGURA 12 – Itens lexicais co-extensivos à ordem do grupo nominal: isolamento dos componentes lexicais de "saúde pública".....	77
FIGURA 13 – Representação do ProtinLex:PTG – processo de mudança.....	78

SUMÁRIO

1. INTRODUÇÃO	13
2. OBJETIVOS.....	20
2.1 Objetivo geral	20
2.2 Objetivos específicos	20
3. REVISÃO TEÓRICA	21
3.1 O léxico e a gramática: a perspectiva da contiguidade adotada pela linguística sistêmico-funcional.....	21
3.2 A co-extensividade do item lexical na estratificação do sistema linguístico.....	29
3.3 O léxico e a gramática: a perspectiva modular adotada pela gramática normativa e pela gramática funcionalista no âmbito do português brasileiro	33
3.3.1 Os processos de expansão lexical sob a perspectiva da gramática funcionalista no âmbito do português brasileiro: lexicalização e gramaticalização	35
4. METODOLOGIA	38
4.1 O <i>corpus</i> compilado para o desenvolvimento e a aplicação dos ProtinLexis.....	38
4.2 O <i>corpus</i> de suporte para a aplicação dos ProtinLexis	41
4.3 O desenvolvimento de um protocolo de investigação lexical no escopo da TSF. 42	
4.3.1 Do Protocolo de Investigação Lexical: Processo de Transformação do Contexto - ProtinLex:PTC	45
4.3.2 Do Protocolo de Investigação Lexical: Processo de Transformação da Gramática – ProtinLex:PTG.....	54
5. RESULTADOS E DISCUSSÃO	60
5.1 Da aplicação do ProtinLex:PTC – <i>corpus</i> de artigos acadêmicos	61
5.2 Da aplicação do ProtinLex:PTG – <i>corpus</i> de artigos acadêmicos.....	76
5.2.1 Da aplicação do ProtinLex:PTG – resultado default.....	82
5.2.2 Da aplicação do ProtinLex:PTG – desvio	98
5.2.3 Da aplicação do ProtinLex:PTG – item lexical “relação”	103
6. CONCLUSÃO	105
7. REFERÊNCIAS BIBLIOGRÁFICAS	111
8. ANEXOS.....	114
8.1 Desenvolvimento e aplicação dos ProtinLexis: seleção dos itens a partir da lista de palavras-chave com p-valor = 0.001	114

8.1.1 “%DIFF”	114
8.1.2 “Odds-ratio”	115
8.1.3 “Hardie’s Log-Ratio”	117
8.1.4 “Ratio of relative frequencies”	119
8.2 Desenvolvimento e aplicação dos ProtinLexis: seleção dos itens a partir da lista de palavras-chave com p-valor = 0.0001	121
8.2.1 “%DIFF”	121
8.2.2 “Odds-ratio”	123
8.2.3 “Hardie’s Log-Ratio”	125
8.2.4 “Ratio of relative frequencies”	127
8.4 Representação ProtinLex:PTG – processo de mudança dos itens selecionados para aplicação e desenvolvimento dos ProtinLexis	137
8.4.1 Representação ProtinLex:PTG – Resultado Default	137
8.4.2 Representação ProtinLex:PTG – Resultado Default + Desvio	143
8.4.3 Representação ProtinLex:PTG – Resultado Desvio	146

1. INTRODUÇÃO

No âmbito dos estudos linguísticos, o léxico (*lexis*)¹ é comumente abordado sob duas perspectivas distintas: modular (*modules*) (cf. CHOMSKY, 1965) e contígua (*continuity*) (cf. HALLIDAY, MCINTOSH e STREVENS, 1964; HASAN, 1985, 1987; MARTIN, 1992; HALLIDAY, 2008; HALLIDAY e MATTHIESSEN, 2014).

Na perspectiva modular, disseminada, em primeira instância, pelos formalistas por meio da gramática normativa e, em segunda instância, por alguns estruturalistas brasileiros com viés funcionalista, o léxico é compreendido como um módulo separado do restante da estratificação. Em outras palavras, a gramática responde pela organização estrutural da língua, enquanto que o léxico ocupa essa estrutura por meio dos itens lexicais (*lexical items*).

Na perspectiva contígua, amplamente difundida pela linguística sistêmico-funcional (LSF), léxico e gramática (*grammar*) são concebidos como dois polos de um contínuo (*continuum*), sendo o polo lexical o ponto mais delicado (*delicate*); e o polo gramatical o ponto mais abstrato (*abstract*) desse contínuo. Em outras palavras, o avanço na delicadeza (*delicacy*) de um sistema gramatical (*system*), por meio do processo de descrição (*description*) desse sistema, alcança a especificidade máxima quando chega ao item lexical. A diferença entre um item lexical e um item gramatical, sob essa perspectiva, se deve apenas ao ponto de observação e ao nível de delicadeza a partir dos quais a observação é conduzida.

Ainda no escopo da perspectiva contígua, encontra-se a abordagem chamada de complementariedade (*complementarity*), descrita nos trabalhos de Halliday (1966; 2008). A complementariedade defende que o léxico e a gramática compreendem um estrato (*stratum*) único denominado de lexicogramática (*lexicogrammar*) e que um mesmo fenômeno linguístico pode ser abordado por uma ótica lexical e/ou por uma ótica gramatical. Esse pressuposto reforça o argumento de que o léxico, assim como a gramática, deve apresentar sua própria teoria para que os fenômenos linguísticos possam ser analisados também pelo ponto de vista lexical, tal como já são examinados pelo ponto de vista gramatical. Nesse sentido, quando Halliday (1966, p. 158) aponta que:

¹ Neste trabalho, os termos técnicos referentes à teoria sistêmico-funcional (TSF) aparecem primeiro em português seguido do termo correspondente em inglês entre parênteses. Isso é válido para a primeira instância de uso do termo. Nas demais ocorrências, o termo aparecerá em português. Os termos técnicos da teoria sistêmico-funcional que dizem respeito aos nomes de sistemas são apresentados em VERSALETE. Os termos que indicam função são apresentados com a letra inicial maiúscula. As traduções dos termos técnicos da TSF foram retiradas de Figueredo (2007, 2011) e/ou foram traduzidas para este trabalho.

O léxico pode ser abordado como (a) um fenômeno linguístico que está relacionado à semântica da mesma forma que está relacionado à gramática e (b) o léxico não pode ser abordado apenas sob o viés gramatical, isto é, os padrões lexicais são tratados de maneira distinta dos sistemas gramaticais, o que possibilita que os padrões lexicais sejam examinados para além do princípio da delicadeza ² (HALLIDAY, 1966, p. 158).

O autor concorda com a concepção de Firth (1957; 1962) de que o léxico não é o mesmo que a semântica e deveria ser explorado a partir de suas próprias categorias teóricas. Em outras palavras, o funcionamento do léxico e as relações estabelecidas entre os itens lexicais de uma língua podem ser investigados por meio de análises linguísticas feitas com base no que Firth nomeia como colocação (collocation) – padrões de coocorrência formados entre uma dada palavra de busca e as demais palavras que a cercam no texto/conjunto de textos em análise (cf. FIRTH, 1962, p. 11).

Os estudos que seguem a perspectiva modular não mencionam: (i) as evidências que mostram o mecanismo e/ou o momento específico da produção de significado em que o item lexical pode ser relacionado à estrutura gramatical, isto é, como as estruturas gramaticais são ocupadas pelos itens lexicais; (ii) os pressupostos que sustentam a lexicalização/gramaticalização; (iii) a correlação entre um item lexical e suas propriedades sintáticas (partindo-se do pressuposto de que os itens lexicais possuem, de alguma forma, propriedades gramaticais) e (iii) a possibilidade de um item lexical apresentar significado mesmo estando isolado em relação ao contexto. Quando a perspectiva modular não aborda esses aspectos, a compreensão acerca do léxico e/ou item lexical tende a ficar restrita ao estrato gramatical, sobretudo à correspondência unívoca entre um item lexical e uma palavra (cf. SINCLAIR, 2004; CASTILHO, 2010), o que nem sempre pode ser entendido como um padrão no português brasileiro (ver 5.0).

Os trabalhos que adotam a perspectiva contígua apesar de tratarem o léxico sob uma abordagem complementar, reafirmando a confluência entre o léxico e a gramática como um estrato lexicogramatical e a possibilidade de investigação do léxico a partir de suas próprias categorias teóricas desconsideram: (i) a relação entre o léxico e as variáveis contextuais, (ii) a

² Minha tradução para: “Lexis may be usefully thought of as within linguistic form, and thus standing in the same relation to (lexical) semantics as does grammar to (grammatical) semantics, and (b) as not within grammar, lexical patterns thus being treated as different in kind, and not merely in delicacy, from grammatical patterns.” (HALLIDAY, 1966, p. 158)

relação entre o léxico e os estratos do sistema linguístico e (iii) os mecanismos de organização interna do léxico, a saber: a maior variedade de itens lexicais e a capacidade de generalização e construção de novos significados, a partir da análise de padrões de colocação e da adição de um item lexical a um grupo de itens lexicais (set).

Conforme os pressupostos da teoria sistêmico-funcional (TSF), (MARTIN, 1992; MARTIN e ROSE, 2007), o léxico pode ser visto ainda sob a perspectiva contextual (context) (HASAN, 1985; ROSE, 2006; HAO, 2015). O contexto é responsável por lidar com a variação das situações sociais. Isso sugere que existe uma correlação entre um item lexical e o contexto no qual ele é instanciado, conforme o modo pelo qual os significados são criados, organizados e distribuídos. Quando a produção de significado em contextos sociais é considerada, o aspecto complementar entre o léxico e a organização sistêmica passa a ser motivado pela produção de sentido (meaning-making), o qual inclui a criação, organização e distribuição de significados nos textos segundo os domínios (domains) das atividades humanas e das relações sociais (cf. HAO, 2020), bem como pela produção de itens lexicais emergentes co-extensivos (coextensive) a outras ordens gramaticais. Isso resulta no desenvolvimento de novos significados baseados nas variáveis do contexto de situação (variables of context of situation) (cf. HALLIDAY, 1978), os quais, com o passar do tempo, podem ser integrados à língua, como, por exemplo, o conceito de “autocuidado” (ver 5.0, 5.1 e 5.2).

Essa correlação entre léxico e contexto é também investigada pela Linguística de Corpus, a qual aponta que as unidades lexicais ganham significado quando se relacionam com outras unidades, quer no mesmo texto (i.e., cotexto), quer na sua frequência verificada em diferentes textos (i.e., contexto) (PEARSON, 1998; SINCLAIR, 2004; MOON, 2012). A partir do estudo de Firth (1957), a abordagem do item lexical passa a englobar um caráter probabilístico, identificado pelos conjuntos lexicais (set) formados pelo princípio da colocação (collocation) (cf. HALLIDAY, 1966; 2008). Segundo Halliday (1966), os conjuntos lexicais podem ser entendidos de forma análoga ao sistema (system) (eixo paradigmático) (paradigmatic axis) presente no escopo gramatical; enquanto que a colocação pode ser compreendida de maneira análoga à estrutura (structure) (eixo sintagmático) (syntagmatic axis) também presente no escopo gramatical. Em razão disso, a Linguística de Corpus é vista nesta pesquisa como um recurso metodológico que vai ao encontro da investigação da sistêmico-funcional, teoria, na qual este trabalho se baseia. Esta pesquisa buscou complementar o entendimento acerca do léxico e dos itens lexicais a partir de uma visão que leva em conta o texto em seu contexto. Isso significa dizer que os resultados desta pesquisa partem de uma base em comum, isto é, um *corpus* compilado especificamente para o presente estudo, o qual é composto de artigos

acadêmicos sobre “autocuidado em diabetes mellitus”, escritos em português brasileiro e extraídos de periódicos da área das Ciências da Saúde.

Do ponto de vista dos Estudos da Tradução, campo disciplinar onde o presente trabalho também está localizado, o léxico/item lexical é tratado de forma semelhante à abordagem adotada pela Linguística de Corpus, sobretudo no que diz respeito à tradução automática e/ou semiautomática, sendo visto sob a ótica de classes de palavra (part of speech) e/ou como instâncias que alimentam bancos de dados terminológicos e glossários multilíngues em memórias de tradução, por exemplo (cf. QUAH, 2006, p. 93). Não há, portanto, uma preocupação com a maneira pela qual se dá o funcionamento do léxico e nem como a organização dos itens lexicais acontece nas línguas. Isso faz com que erros de tradução relacionados aos itens lexicais possam ser vistos como questões conflituosas para os tradutores, sejam eles técnicos, literários e/ou audiovisuais, uma vez que nem sempre haverá a correspondência item lexical e palavra entre a língua fonte e a língua alvo, em muitos casos a correspondência será item lexical e grupo nominal, por exemplo.

Essa ausência de estudos que investiguem o léxico e/ou item lexical de forma mais profícua a partir de uma perspectiva linguística pode contribuir para que a tradução automática (machine translation) – mas não apenas – não evolua no que concerne a escolha de itens lexicais mais adequados para um dado tipo de texto, desconsiderando fatores como o contexto de situação onde determinado texto está inserido. Quah (2006, p.183) destaca que a ambiguidade lexical (lexical ambiguity) é um dos desafios relacionados ao léxico e à tradução, bem como à efetividade na produção de textos traduzidos automaticamente que seja capaz de selecionar itens lexicais de acordo com a função que operam na oração, levando em conta que itens lexicais podem ser repetidos ao longo do texto. Um dos exemplos mencionados pelo autor envolve o fato de que a máquina precisa entender que o item “banco (bank)” em “Beth vai ao banco todos os dias”³ pode significar um lugar onde Beth assenta ou uma instituição financeira. A máquina precisa também ter a capacidade de lidar com o significado que determinada construção linguística apresenta para que a escolha dos itens lexicais seja a mais adequada possível. Para esse caso, o exemplo citado pelo autor é “Ele conhece mais mulheres bonitas que Samantha”⁴ (QUAH, 2006, p. 183), cuja interpretação poderia se dar de duas formas, a saber, ele conhece mais mulheres bonitas que a Samantha conhece ou ele conhece mulheres mais bonitas que a própria Samantha. Segundo esse autor, a tradução automática teria que saber lidar com esse tipo de interpretação para decidir quais opções tradutórias seriam ideais de acordo com a

³ Minha tradução para “Beth goes to the bank everyday” (QUAH, 2006, p. 183).

⁴ Minha tradução para: “He knows more beautiful women than Samantha” (QUAH, 2006, p. 183).

interpretação em questão. Nesse sentido, a caracterização do item lexical, por meio da investigação do funcionamento do léxico no português brasileiro, pode auxiliar o avanço dos trabalhos que tratam do desenvolvimento e do aprimoramento de traduções feitas por máquina ou com o auxílio de máquina.

Em função da complementariedade existente entre léxico e gramática, um item lexical está relacionado à gramática, da mesma forma, que está relacionado à semântica, o que torna a interpretação de um item lexical sempre inter-estratos (HALLIDAY, 2008). Isso pode ser verificado pelo papel que o léxico ocupa em conjunto com alguns sistemas da semântica, como, por exemplo, as relações lexicais (lexical relations) de hiponímia (hyponymy) e meronímia (meronymy) encontradas no escopo da COESÃO LEXICAL (LEXICAL COHESION) (HALLIDAY e HASSAN, 1976; MARTIN, 1992; HAO, 2015). É nesse sentido que tanto a organização interna do léxico quanto sua co-extensividade com a dimensão da estratificação se tornam os principais fenômenos investigados nesta pesquisa.

De maneira mais específica, este trabalho admite duas relações adjacentes, a partir da consideração do léxico como conceito teórico e da sua interação com o contexto e a dimensão da estratificação do sistema linguístico. Essas relações incluem: (i) o contexto, através da investigação de itens lexicais co-extensivos à ordem do grupo nominal, que podem emergir nos textos a partir de alterações nas variáveis contextuais – campo (field), sintonia (tenor) e modo (mode) – no processo de transformação do contexto e (ii) o estrato da gramática, por meio da formação de itens lexicais e a co-extensividade às ordens da gramática tais como a palavra (word), o grupo (group) e o morfema (morpheme) – no processo de transformação da gramática. Em função disso, esta pesquisa lança mão de protocolos de investigação lexical, desenvolvidos no âmbito deste trabalho e nomeados como ProtinLex⁵. Os protocolos exploram não só a relação do léxico e/ou item lexical para com o contexto, como também a relação do léxico e/ou item lexical para com os estratos linguísticos da semântica e da gramática. Por isso todos os protocolos de investigação lexical descritos ao longo deste estudo seguem a nomenclatura “ProtinLex: processo de transformação da(o) + contexto ou o nome do estrato linguístico”. Por exemplo, o ProtinLex: processo de transformação da gramática (ProtinLex:PTG) refere-se ao protocolo de investigação lexical responsável por examinar a relação léxico/item lexical e o estrato da gramática. É importante esclarecer ainda que os protocolos de investigação lexical podem ser considerados, a princípio; análogos aos testes (probes) utilizados na descrição

⁵ProtinLex (plural, ProtinLexis) é uma sigla utilizada ao longo deste trabalho para se referir aos protocolos de investigação lexical desenvolvidos no âmbito desta pesquisa de doutorado como forma de examinar a caracterização do léxico/item lexical no português brasileiro.

gramatical no âmbito da TSF, mas, em segunda instância; devem ser concebidos como uma forma de examinar o funcionamento do léxico e do item lexical no português brasileiro, cujo enfoque volta-se para a caracterização léxico/item lexical, diferentemente dos testes que priorizam os sistemas gramaticais.

Nesse sentido, os objetivos principais deste trabalho são: (i) explorar a dicotomia módulo/contiguidade, mostrando os mecanismos pelos quais a dimensão do léxico para o sistema do português brasileiro é co-extensiva aos sistemas gramaticais, e não deve ser concebida como o ponto máximo de delicadeza que um sistema gramatical pode alcançar (cf. HALLIDAY, 1961; HALLIDAY, MCINTOSH e STREVENS, 1964; HALLIDAY, 1966; TUCKER, 1998, p. 10), (ii) investigar a forma pela qual os itens lexicais estão distribuídos no português brasileiro a partir da sua co-extensividade em relação às variáveis do contexto de situação, bem como em relação aos estratos linguísticos (cf. FIGUEREDO, 2007; 2011), ajudando a compreender de que modo o léxico contribui para a produção de significado e (iii) propor uma forma de caracterizar o léxico no português brasileiro que considere a abordagem complementar entre léxico e gramática, sendo capaz de identificar generalizações acerca do funcionamento do léxico e do item lexical nesse sistema linguístico (cf. HALLIDAY, 1966, p. 169).

A partir desses objetivos principais, esta pesquisa visa contribuir ainda para a caracterização do item lexical no português brasileiro com impacto no exame de casos específicos de co-extensividade entre o item lexical e o grupo nominal, principalmente em relação aos grupos nominais metaforizados, como, por exemplo, “intervenções educativas” e “complicações crônicas” (ver 5.1 e 5.2). Faz-se necessário esclarecer que os grupos nominais metaforizados são desenvolvidos a partir da metáfora gramatical, processo no qual “uma classe ou função gramatical passa a funcionar como um outro tipo de classe ou função gramatical” (cf. FERREGUETTI, 2018, p. 51), sendo um recurso utilizado pela linguagem para construir novas camadas de significado por meio de seleções feitas pela gramática. Isso significa dizer que orações não-finitas e encaixadas, por exemplo, que funcionam como Sujeito de uma outra oração, podem ser entendidas como uma forma utilizada pela linguagem para construir a metáfora gramatical. O mesmo vale para as nominalizações, em que instâncias provenientes da classe de palavras verbal geram instâncias localizadas no escopo da classe de palavras nominal e acabam realizando as funções de Participantes (Participants) e/ou Circunstâncias (Circumstances), enquanto que o esperado seria que essas instâncias da classe de palavras verbal realizassem a função do Processo (Process) (HALLIDAY e MATTHIESSEN, 2014, p.709-711).

Adiante, os Objetivos desta pesquisa são detalhados. Em seguida, o capítulo destinado à Revisão Teórica é apresentado. Logo depois, os passos e recursos metodológicos são descritos no capítulo correspondente à Metodologia, seguido do capítulo dos Resultados e Discussão do Resultados, em que os dados obtidos, a partir do desenvolvimento e aplicação dos protocolos de investigação lexical estabelecidos para esta pesquisa, são especificados. Por fim, as Referências Bibliográficas e os Anexos são detalhados.

2. OBJETIVOS

2.1 Objetivo geral

- Caracterizar o léxico e o item lexical enquanto recursos linguísticos para a produção de significado, no âmbito do português brasileiro considerando os pressupostos teóricos da linguística sistêmico-funcional.

2.2 Objetivos específicos

- Explorar a dicotomia módulo/contiguidade, mostrando os mecanismos pelos quais a dimensão do léxico para o sistema do português brasileiro é co-extensiva aos sistemas gramaticais, e não deve ser concebida como o ponto máximo de delicadeza que um sistema gramatical pode alcançar.
- Investigar a forma pela qual os itens lexicais estão distribuídos no português brasileiro a partir da sua co-extensividade em relação às variáveis do contexto de situação, bem como em relação aos estratos linguísticos (cf. FIGUEREDO, 2007; 2011), ajudando a compreender de que modo o léxico contribui para a produção de significado.
- Propor uma forma de caracterizar o léxico no português brasileiro que considere a abordagem complementar entre léxico e gramática, sendo capaz de identificar generalizações acerca do funcionamento do léxico e do item lexical nesse sistema linguístico.

3. REVISÃO TEÓRICA

A definição de léxico, seu processo de formação e sua relação com a gramática foram discutidos pela literatura em vários momentos do desenvolvimento dos estudos linguísticos (cf. HALLIDAY, MCINTOSH e STREVEN, 1964; HALLIDAY, 1966; SINCLAIR, 2004; HALLIDAY, 2008; entre outros). Em função disso, este capítulo apresenta um panorama acerca do léxico no âmbito dos estudos linguísticos, englobando as visões sobre esse objeto de estudo em duas perspectivas principais: a perspectiva modular e a perspectiva contígua. A perspectiva modular é difundida pelos formalistas, por meio da gramática normativa (cf. CHOMSKY, 1965), e por alguns estruturalistas brasileiros, através da gramática funcional (cf. BASÍLIO, 1991; NEVES, 1997; BORBA, 2003; CASTILHO, 2010). A perspectiva contígua, por sua vez, é disseminada pelos sistemicistas, por meio da teoria sistêmico-funcional (TSF) (cf. HALLIDAY, MCINTOSH e STREVEN, 1964; HALLIDAY, 1966; HASAN, 1985, 1987; MARTIN, 1992, HALLIDAY, 2008; HALLIDAY e MATTHIESSEN, 2014). A perspectiva contígua engloba ainda a abordagem complementar (complementarity) preconizada pela linguística sistêmico-funcional com base nos pressupostos teóricos da linguística de corpus, sobretudo no que diz respeito ao conceito de colocação, que segundo os trabalhos de Firth (1957; 1962) e Halliday (1966; 2008), compreende a base da abordagem para se examinar o funcionamento do léxico e do item lexical por meio de categorias próprias que não considerem apenas o princípio da delicadeza, presente no escopo gramatical. Para além da abordagem do léxico sob essas duas perspectivas – modular e contígua – este capítulo aborda ainda a maneira pela qual o léxico e/ou item lexical são enxergados no âmbito dos Estudos da Tradução, uma vez que este trabalho também está localizado nesta área.

3.1 O léxico e a gramática: a perspectiva da contiguidade adotada pela linguística sistêmico-funcional

De acordo com Halliday; McIntosh; Strevens (1964), há dois tipos de escolhas, as gramaticais e as lexicais. As primeiras são entendidas como fechadas e as segundas como abertas. Isso significa dizer que as escolhas gramaticais (grammatical choices) compreendem um sistema fechado (closed system) de possibilidades, que funcionam sob o princípio da agnação (agnation): propriedade do eixo paradigmático (paradigmatic axis) em que uma opção (term) de um sistema (system) está relacionada a outra, sendo que a agnação também pode estar

relacionada ao grau, isto é, as opções de um sistema que possuem a mesma condição de entrada (condition of entry) são opções mais semelhantes entre si (cf. Matthiessen; Teruya; Lam, 2010, p. 49-50). Por exemplo, os modos Interrogativo (Interrogative), Declarativo (Declarative) e Imperativo (Imperative) estão no escopo do sistema de MODO (MOOD) e são considerados semelhantes por fazerem parte do mesmo sistema. Portanto, se o modo Interrogativo é escolhido para uma oração (clause), significa que tal oração será não Declarativa e não Imperativa.

Já as escolhas lexicais (lexical choices) são provenientes de conjuntos abertos (open sets) cujo princípio da agnação pode funcionar de maneira diferente, uma vez que essas escolhas se dão por probabilidade de ocorrência. A escolha de um determinado item lexical pode ser regida pela frequência com que aquele dado item aparece nos textos de um domínio específico ou de um tipo de texto em particular. Halliday, McIntosh e Stevens (1964) argumentam ainda que os itens lexicais são escolhidos a todo momento durante o desenvolvimento do texto, sendo essa relação de escolha de um para um, ou seja, a seleção por um item lexical pode mudar conforme as variáveis do contexto de situação que estão norteando a forma como um dado significado será produzido no texto.

As escolhas gramaticais são apresentadas por meio de sistemas em que uma escolha está em oposição (contrast) e/ou em co-seleção (selection) com outra, sendo essa relação de escolha de um ou outro (oposição) ou de um e outro (co-seleção). Nesse sentido, Halliday, McIntosh e Stevens (1964) apontam que as escolhas lexicais são mais variadas que as gramaticais, por isso a abstração (abstraction) é uma configuração que faz parte da gramática. A abstração diz respeito à capacidade que os itens gramaticais têm de serem localizados em uma categoria gramatical mais geral. Por exemplo, o item “ela” é uma abstração da classe de palavras dos pronomes do tipo pessoal, feminino e singular, sistema de PESSOA: NÃO-INTERLOCUTOR (cf. Figueredo, 2011, p. 191), sendo que essa seria a única abstração possível para tal item, ou seja, não haveria uma abstração acima da “categoria” pronome para o item “ela”. Para o léxico, a “abstração” funciona de maneira diferente, uma vez que os itens lexicais apresentam a capacidade de formar conjunto (sets) ao mesmo tempo que funcionam sob o princípio da abstração (HALLIDAY, 1966). Por exemplo, “anemia”, “diabetes” e “reumatismo” podem ser abstrações de “doenças”, sendo que “doenças” é um item de um conjunto aberto (open set), ou seja, mais itens poderiam ser adicionados a essa abstração sem que o significado dos itens que já estão nesse conjunto lexical fosse alterado.

Isso sugere que tanto as escolhas gramaticais quanto as escolhas lexicais são feitas ao longo do desenvolvimento do texto. Mas as escolhas lexicais são feitas de maneira diferente, dado que se um pronome no feminino e no singular no português brasileiro precisa ser utilizado

em uma oração, a única opção possível para esse sistema é “ela”. No entanto, em conjuntos abertos, isto é, no léxico, as escolhas se dão por probabilidade, ou seja, em um mesmo texto, é possível encontrar: diabetes, condição, doença, condição crônica para se referir ao Diabetes Mellitus, sendo que nenhuma opção anula ou está em oposição a outra.

Ainda em Halliday; McIntosh; Strevens (1964), a relação entre léxico e gramática é estabelecida por meio do conceito de contínuo, no qual em um polo (pole) está a gramática e no polo oposto está o léxico. Entre esses dois polos, localizam-se os itens que não são puramente gramaticais e nem puramente lexicais, como; por exemplo, as palavras da classe dos advérbios (adverbs). Léxico e gramática são entendidos, portanto, como parte do mesmo estrato linguístico – a lexicogramática – regido pelo princípio da delicadeza (HALLIDAY, MCINTOSH e STREVENSON, 1964; HALLIDAY e MATTHIESSEN, 2014). Por esse princípio, as unidades da lexicogramática são descritas através de uma rede de sistemas que é organizada pelo eixo paradigmático – gramática e léxico – e pela estrutura (structure) por meio do eixo sintagmático (syntagmatic) – escala de ordens (rank scale) (oração (clause), grupo/frase (group/phrase), palavra (word) e morfema (morpheme)). O sistema é descrito por meio de subsistemas, quanto mais subsistemas são utilizados na descrição, maior é o nível de delicadeza. No ponto mais delicado do sistema, localiza-se o léxico, no ponto mais abstrato, a gramática, compondo o contínuo lexicogramatical (lexicogrammar continuum).

Na introdução do capítulo “Introduction: a personal perspective” da obra *On Grammar*, Halliday (2002, p. 8) destaca que há um abismo entre os pontos de convergência do polo lexical e do polo gramatical no contínuo lexicogramatical. A perspectiva apresentada no livro “*The linguistic sciences and language teaching*” de Halliday; McIntosh; Strevens (1964) acerca dos conceitos de sistema fechado para se referir à gramática, e conjunto aberto para se referir ao léxico também é reforçada e expandida na obra de 2002.

Na publicação “*Categories of the Theory of grammar*”, Halliday (1961, p. 40-41) aponta que em um sistema fechado: (i) o número de itens é limitado; (ii) cada item é único, ou seja, o item x não será idêntico ao item z e (iii) a cada acréscimo de um novo item, os itens já existentes sofrem modificações no significado. Inicia-se nessa obra a discussão sobre a dissociação entre item lexical e unidades da escala de ordens da gramática (oração, grupo, palavra e morfema), uma vez que o item lexical, no inglês, pode ser co-extensivo a três unidades da escala de ordens da gramática – grupo, palavra e/ou morfema.

Nesse trabalho, Halliday (1961) compara ainda gramática e léxico, quando afirma que a gramática é organizada pela escala de ordens e pelo princípio de abstração, sinalizando que na gramática a relação estabelecida é de um para um, ou seja, em um determinado ponto

da abstração apenas um item gramatical pode ser expoente de uma classe de palavras ou de uma estrutura (structure) na oração. Para o léxico, o autor afirma que esses princípios da escala de ordens e da abstração não são relevantes, sendo a organização lexical relacionada à probabilidade de ocorrência de um dado item lexical em um determinado tipo de texto, o que é confirmado por meio de colocados e conjuntos lexicais. De acordo com Halliday (1961), os colocados são entendidos como a probabilidade de o item x coocorrer com certa frequência com o item y, z e/ou w. Já os conjuntos lexicais compreendem os grupos formados por itens lexicais que apresentam alguma relação entre si, por exemplo, o mesmo domínio ou a alta frequência com que os itens lexicais de um dado conjunto lexical formam colocados.

Ainda no que tange a comparação entre gramática e léxico, Halliday (1961) enfatiza que a gramática é “determinista” e o léxico “probabilístico”⁶. Isso indica que, no âmbito gramatical, é possível distinguir o que é factível do que é não factível de acontecer no sistema, enquanto no âmbito lexical, é possível distinguir o que é mais provável e o que é menos provável de ocorrer em um texto, por exemplo (HALLIDAY, 1963, p. 105).

Em “Lexis as a linguistic level”, Halliday (1966) acrescenta o termo técnico série (serie) aos conceitos de colocados e conjuntos lexicais para a organização do léxico. Para ele, série compreende um conjunto de itens lexicais que apresentam um item constituinte em comum, o qual geralmente está relacionado à morfologia, como, por exemplo a série “alimentar, alimento, alimentação”. Contudo, segundo o autor, uma série pode ser tanto série e conjunto lexical ao mesmo tempo, ou ser apenas série. Em outras palavras, uma série pode ser formada por itens lexicais de domínios diferentes ou que não constituem colocados, a série “lenhador, descritor, computador”, por exemplo, pode ser considerada apenas uma série, porque a relação entre os itens é estabelecida apenas pela presença do morfema ‘or’. Já a série “alimentar, alimento, alimentação” pode ser considerada uma série equivalente a um conjunto lexical, uma vez que todos os itens que a compõem são do mesmo domínio e/ou podem ser encontrados como colocados (collocates) em um texto.

O termo técnico colocação (collocation), por sua vez, foi cunhado primeiramente por Firth (1957) e diz respeito às palavras que coocorrem com uma certa frequência ao redor de uma determinada palavra, formando um colocado. No corpus compilado para a presente pesquisa, “mellitus” sempre coocorreu seguido de “diabetes”. Logo, seguindo a definição de Firth, “diabetes mellitus” pode ser entendido como um colocado para o corpus deste trabalho. Sinclair (2004, p. 28) trata a colocação de maneira semelhante a Firth quando aponta que a

⁶ Minha tradução para: “Another way of drawing the same distinction between grammar and lexis is to say that grammar is “deterministic” by contrast with lexis which is “probabilistic” (HALLIDAY, 1963, p. 105).

colocação corresponde a uma “coocorrência entre palavras que ocorre com certa frequência”⁷. Ele acrescenta ainda que a escolha de palavras de um texto pode ser influenciada pela colocação, ou seja, a escolha de uma palavra está condicionada a escolha da próxima palavra, da palavra seguinte e assim por diante (cf. SINCLAIR, 2004, p. 19). Baseado no conceito de colocação de Firth, Halliday (1961, p. 61) define que a “colocação diz respeito à associação sintagmática dos itens lexicais”⁸, a qual pode ser quantificada pela probabilidade de ocorrência desses itens lexicais. Cada item lexical apresenta uma gama de colocações (range of collocations) prováveis de acontecer, sendo que a delicadeza⁹ dessa gama de colocações aumenta à medida que mais itens lexicais são adicionados a essa gama, ou seja, se um item lexical x coocorre com frequência com o item lexical y, com o z e/ou com o w, a delicadeza do item lexical x aumenta a cada nova coocorrência identificada (cf. HALLIDAY, 1961, p. 61).

Ainda nessa publicação, Halliday (1966) destaca o argumento de Firth (1957) acerca do fato de que o léxico não pode ser visto como análogo ao estrato da semântica. Em outras palavras, o léxico apresenta uma relação com a semântica assim como a gramática também apresenta uma relação com a semântica. Portanto, os padrões lexicais não podem ser considerados somente sob o princípio da delicadeza da gramática, eles precisam ser vistos sob uma ótica diferente, como, por exemplo, sob o nível da colocação (collocational level).

Considerando a probabilidade de coocorrência dos itens lexicais, as pesquisas de Halliday (1966; 2008) destacam que, para além da perspectiva contígua, o léxico pode ser visto de forma complementar à gramática. De acordo o autor, essa complementariedade reside no fato de que tanto as categorias lexicais quanto as categorias gramaticais constituem a linguagem (HALLIDAY, 2008, p. 69). Em outras palavras, para a linguística sistêmico-funcional (LSF), léxico e gramática são entendidos como um único fenômeno, o contínuo/estrato lexicogramatical, sendo que uma mesma questão linguística pode ser abordada por uma visão lexical e/ou por uma visão gramatical (HALLIDAY, 2008, p. 31). Em razão dessa complementariedade, segundo Halliday (2008, p. 32), léxico e gramática podem ser abordados por meio de suas próprias categorias, ou seja, enquanto a gramática é organizada pelos eixos sintagmático (estrutura) e paradigmático (sistema), o léxico seria equivalente ao eixo sintagmático da gramática, quando organizado pelos padrões de colocação e pelos conjuntos

⁷ Minha tradução para: “a frequent co-occurrence of words” (SINCLAIR, 2004, p. 28).

⁸ Minha tradução para: “collocation is the syntagmatic association lexical items” (HALLIDAY, 1961, p. 61).

⁹ Apesar do termo delicadeza mencionado por Halliday (1961) ter a mesma grafia que o termo técnico *delicadeza* (*delicacy*) dos sistemas gramaticais, delicadeza nesse trecho não assume o mesmo significado técnico que a *delicadeza* da gramática. Delicadeza nesse trecho diz respeito à capacidade que um item lexical tem que formar *colocados*, ou seja, quanto maior for a coocorrência de um dado *item lexical* com outros *itens lexicais*, maior será a delicadeza desse *item lexical*.

lexicais (open sets) e equivalente ao eixo paradigmático da gramática quando analisado mediante as probabilidades de coocorrência dos itens lexicais (cf. HALLIDAY, 1966, p. 162). Isso indica que o léxico pode ser abordado de maneira distinta da gramática e que o princípio da delicadeza não é considerado o único possível para a investigação do funcionamento lexical.

Na década de noventa, a ideia de contínuo é reforçada no escopo da TSF por Matthiessen (1995), quando o autor menciona que as opções gramaticais são mais amplas e gerais, enquanto que as opções lexicais são mais específicas e delicadas, sendo que léxico e gramática formam um sistema único, a LEXICOGRAMÁTICA (LEXICOGRAMMAR), que é regida pelo princípio da delicadeza (delicacy). Essa obra destaca ainda que o léxico apresenta a capacidade de mudar de maneira mais rápida que a gramática, possibilitando que ele seja desenvolvido constantemente e tenha modificações recorrentes na sua organização sistêmica. Esse autor aponta também que o léxico pode ser entendido sob a ótica de cada uma das três metafunções (metafunctions). Para a metafunção ideacional (ideational metafunction), o léxico trata da nomeação e organização taxonômica das coisas e dos eventos, sendo, portanto, uma taxonomia para a representação e interpretação da construção da nossa experiência no mundo. Para a metafunção interpessoal (interpersonal metafunction), o léxico lida com a conotação dos significados realizados pelos itens lexicais a partir das relações entre falante/ouvinte, escritor/leitor. Para a metafunção textual (textual metafunction), o léxico é responsável pela criação do texto, como, por exemplo, pelos recursos de coesão (cohesion) e coerência (coherence) (MATTHIESSEN, 1995, p. 110-111).

Tucker (1998), por sua vez, acentua o posicionamento de Matthiessen (1995) no que diz respeito à relação léxico e metafunções quando aponta que os itens lexicais podem variar conforme o “peso” que determinada metafunção exerce em um texto; por exemplo, se o texto é mais voltado para o campo (metafunção ideacional) ou mais para a sintonia (tenor) (metafunção interpessoal). Segundo esse autor, o significado dos itens lexicais escolhidos para um determinado texto, por exemplo, funciona com base nas variáveis do contexto de situação – campo para a metafunção ideacional, sintonia para a metafunção interpessoal e modo (mode) para a metafunção textual. Seguindo os exemplos de Matthiessen (1990), Tucker (1998, p. 26) menciona que a metafunção interpessoal e a textual seriam capazes de contribuir de maneira independente nas escolhas lexicais de um texto, enquanto que a metafunção ideacional apenas contribuiria se estivesse associada às metafunções interpessoal ou textual. A metafunção interpessoal influenciaria nas escolhas lexicais para os tipos de modalidade (modality) e para o sistema de POLARIDADE (POLARITY) e ATITUDE (ATTITUDE). A metafunção textual influenciaria nas escolhas lexicais relacionadas às conjunções (conjunction), substituição (substitution) e

referência (reference). A metafunção ideacional associada à metafunção interpessoal influenciaria nas escolhas lexicais relacionadas ao nível de formalidade (formality) do texto e ao sistema de AFETO (AFFECT), e a metafunção ideacional associada à metafunção textual influenciaria nas escolhas lexicais relacionadas ao sistema de COESÃO LEXICAL¹⁰.

Tucker aponta ainda que o léxico é específico de uma dada língua e cultura. Assim a função social e semiótica de um item lexical somente seria compreendida por meio das interações humanas¹¹ (TUCKER, 1998, p. 2). Essa perspectiva justifica a necessidade de se delimitar o léxico de um dado sistema, no caso da presente pesquisa, o português brasileiro. Em relação à organização lexicogramatical, Tucker defende que a concepção de léxico e gramática juntos, formando um recurso unificado, é uma tentativa de incorporar léxico e gramática que, de acordo com ele, numa visão tradicional, estariam localizados separadamente, sendo a gramática vista como sintaxe e o léxico como vocabulário (TUCKER, 1998, p. 9). Por fim, o autor contesta a visão de que o léxico seria o ponto mais delicado da gramática. Para ele, a diferença entre os itens lexicais não poderia ser capturada pelo aumento da delicadeza dos sistemas gramaticais. Portanto, a organização dos itens lexicais não começa quando a descrição dos sistemas gramaticais atinge a sua delicadeza máxima (TUCKER, 1998, p. 10).

Já Martin e Rose (2007) ressaltam que o léxico é uma parte da linguagem que consegue mudar de maneira mais rápida, suprindo as necessidades de comunicação que surgem ao longo dos tempos (MARTIN; ROSE, 2007, p. 63). Para esses autores, o léxico também é responsável pela construção do campo do texto, isto é, as escolhas lexicais apresentam a capacidade de iniciar ou expandir o domínio de um texto. Martin e Rose (2007) salientam, por fim, que o léxico pode ser analisado pelas relações lexicais de repetição (repetition), sinonímia (synonymy), contraste (contrast), hiponímia e meronímia (MARTIN; ROSE, 2007, p. 81). Essas concepções indicam, portanto, que o léxico pode ser examinado sob a ótica do sistema de IDEAÇÃO (system of IDEATION), visto que para Martin e Rose (2007, p. 75) as relações lexicais (lexical relations), i.e, relações semânticas que podem ser estabelecidas entre determinadas pessoas, coisas, processos, lugares e qualidades que são responsáveis por construir o campo de um texto, estão no escopo desse sistema no âmbito da metafunção ideacional (ideational metafunction).

¹⁰ Quando Tucker (1998) menciona o sistema de COESÃO LEXICAL nesse trecho, ele se refere à perspectiva de Martin (1992) que considera a COESÃO um sistema da IDEAÇÃO (IDEATION), que, por sua vez, está localizada no escopo da metafunção ideacional.

¹¹ Minha tradução para: “Lexis is language and culture specific, and the social and semiotic function of a word can only be acquired through human interaction” (TUCKER, 1998, p. 2).

Na obra “Halliday’s Introduction to Functional Grammar”, Halliday & Matthiessen (2014) destacam novamente o conceito de contínuo lexicogramatical, bem como a ideia de que o léxico é o polo mais delicado da gramática, constituído de grupos abertos, com significados específicos (*specific in meaning*) e colocações (*collocations*), esta última localizada no escopo da COESÃO LEXICAL (*LEXICAL COHESION*). Corroborando a perspectiva apresentada em Halliday e Hasan (1976) acerca do estudo do léxico pela COESÃO LEXICAL (*LEXICAL COHESION*), principalmente no que diz respeito à colocação (*collocation*), Halliday & Matthiessen (2014, p. 648-650) definem colocação como a tendência que determinado item lexical apresenta de coocorrer com outro item lexical e acrescentam que as colocações podem ser associadas a um contexto de situação em específico e/ou a várias funções da linguagem.

Embora as visões sistêmico-funcionais apresentadas anteriormente sejam consideradas válidas no âmbito dos estudos que seguem essa teoria linguística, Halliday; McIntosh; Strevens (1964) e Halliday (1961; 1963; 1966; 2008) ressaltam que o léxico deveria apresentar sua própria teoria com conceitos e categorias próprias. Em outras palavras, o léxico deve ser estudado como um conceito teórico, a partir da sua própria organização interna, deixando que a perspectiva gramatical (eixo paradigmático (sistema) e eixo sintagmático (estrutura)) seja utilizada para lidar com os elementos gramaticais.

Aprofundando o ponto de vista acerca do léxico abordado em Halliday; McIntosh; Strevens (1964) e Halliday (1961; 1963; 1966), em uma publicação mais recente, Fontaine (2017a) contesta a concepção do léxico como o polo mais delicado do contínuo lexicogramatical e defende que a representação lexical carece de descrição no escopo da teoria sistêmico-funcional, dado que o léxico pode ser considerado sob o princípio da instanciação (*instantiation*), ao invés da delicadeza. Considerando essa perspectiva, a autora utiliza o aporte teórico apresentado por HALLIDAY (1991) sobre contexto de cultura (*context of culture*) e contexto de situação (*context of situation*), a saber:

(contexto de) cultura é construído por escolhas feitas nos sistemas linguísticos; enquanto o (contexto de) situação é construído por padrões linguísticos identificados com o uso da língua¹² (HALLIDAY, 1991, p. 282).

¹² Minha tradução para: “The culture is construed by systems of language choices; the situation is construed by patterns of language use” (HALLIDAY, 1991, p. 282).

E esclarece que, sob esse panorama, o léxico pode ser entendido como um potencial de significados (meaning potential); e o item lexical como uma instanciação desse potencial, sendo que o item lexical pode ainda ser visto sob um contexto mais localizado¹³.

Por fim, Fontaine (2017a, p. 14-15) destaca que o léxico (potencial de significados) de uma língua é construído por meio dos usos, sobretudo, através dos usos analisados em conjunto com o(s) respectivo(s) contexto(s); ao passo que o significado de um item lexical pode ser recuperado através de um contexto mais localizado e dos colocados formados por ele, bem como por meio das construções ou padrões identificados para tal item lexical. Nesse sentido, segundo a autora, a TSF deve utilizar as abordagens baseadas em corpus para o avanço dos estudos que tratam da lexicogramática.

3.2 A co-extensividade do item lexical na estratificação do sistema linguístico

Uma das principais dimensões da organização da língua é a estratificação (HALLIDAY, 2002; HALLIDAY e MATTHIESSEN, 2014). Essa dimensão é formada pelo plano da expressão (expression), que engloba os estratos da fonologia (phonology) e da grafologia (graphology), e pelo plano do conteúdo (content), que inclui os estratos da lexicogramática e da semântica. O princípio que organiza os estratos linguísticos é a realização (realization), responsável pela relação conjunta entre um padrão linguístico abstrato com um padrão mais complexo.

O léxico e a gramática compõem um estrato linguístico que se localiza a um nível mais abstrato que os estratos do plano da expressão e a um nível menos abstrato que o estrato da semântica. Esse estrato linguístico é chamado de lexicogramática, cuja constituição se dá em quatro ordens, as quais se relacionam por meio da hierarquia composicional, são elas: morfema, palavra, frase/grupo e oração (HALLIDAY & MATTHIESSEN, 1999, 2014).

Considerando que a presente pesquisa tem como objetivo tratar o léxico como um conceito teórico, a abordagem das unidades da escala de ordens da gramática torna-se um dos focos de investigação deste estudo, uma vez que o trabalho de Halliday (1961) alerta para o fato de que o item lexical no inglês pode ser co-extensivo (coextensive) a três unidades da escala de ordens da gramática – grupo, palavra e/ou morfema. Para além disso, Fontaine (2017a, 2017b) aponta a existência de itens lexicais formados por mais de um componente,¹⁴ os quais, segundo

¹³ Minha tradução para: “Lexis as most local context” (FONTAINE, 2017a, p. 13).

¹⁴Componente deve ser entendido como um item lexical ou como uma porção de um item lexical. Ex.: “diabetes” pode ser compreendido como um item lexical em co-extensividade à ordem da palavra, sendo; portanto, um item lexical formado por um componente. “Diabetes mellitus tipo II”, por sua vez, pode ser entendido como um item

Halliday (1961), podem ser co-extensivos à ordem do grupo (nominal), enquanto que os itens lexicais compostos por um componente apenas podem ser co-extensivos à ordem da palavra.

A co-extensividade é vista como a capacidade que um item lexical pode apresentar de ser extensivo a uma ou mais ordens da escala de ordens da gramática (morfema, palavra, grupo e/ou oração) (HALLIDAY, 1961, p. 59-60). Essa propriedade de ser co-extensivo indica que um item lexical é entendido como um expoente (exponent) do léxico, sendo que esse item lexical pode apropriar-se das propriedades gramaticais de uma das unidades da escala de ordens da gramática, sobretudo, à ordem da palavra e/ou do grupo. A fim de esclarecer como essa propriedade de co-extensividade pode ser estabelecida entre um item lexical e a escala de ordens da gramática, a representação exposta na FIG. 1 a seguir foi desenvolvida.

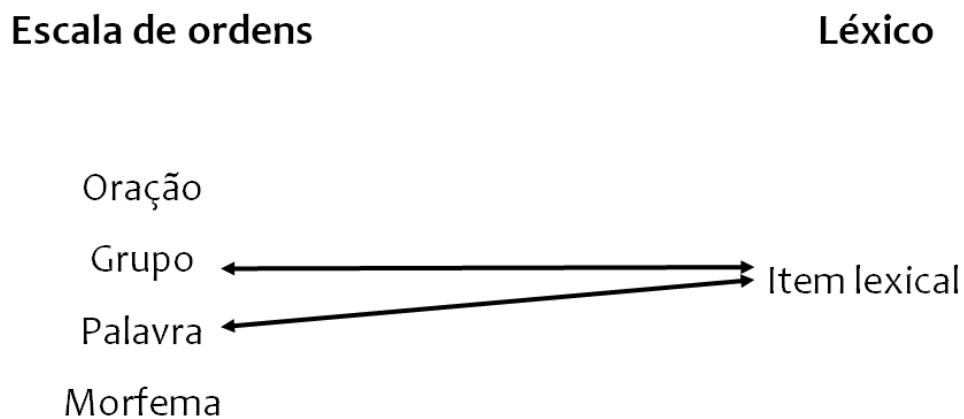


FIGURA 1 – A representação da propriedade de co-extensividade entre um item lexical e a escala de ordens da gramática.

Fonte: Desenvolvida para fins deste trabalho.

A representação da FIG.1 mostra à direita o item lexical, que é compreendido como um expoente do léxico de um sistema linguístico, no caso, o português brasileiro. À esquerda localiza-se a escala de ordens da gramática para o português brasileiro (cf. FIGUEREDO, 2011), formada pela oração, grupo, palavra e morfema. Ao centro da FIG. 1, há duas setas, ambas de pontas duplas, uma representando a propriedade de co-extensividade existente entre o item lexical e o grupo, dado que no português brasileiro há itens lexicais formados por mais de um componente (ver QUADRO 1), e outra representando a propriedade de co-extensividade existente entre o item lexical e a palavra, englobando os itens lexicais constituídos de apenas um componente (ver QUADRO 1). É importante esclarecer que a propriedade de co-

lexical em co-extensividade à ordem do grupo nominal, sendo; portanto, um item lexical formado por quatro componentes, que, se forem separados, apresentarão significados distintos.

extensividade é representada pela seta de pontas duplas para indicar que um item lexical que é co-extensivo às ordens da palavra e/ou do morfema pode coincidir com as funções gramaticais dessas ordens, e as ordens da palavra e/ou do morfema também podem coincidir com um item lexical para delimitar o domínio de um texto, por exemplo (ver 5.2.2).

O QUADRO 1, a seguir, traz exemplos de itens lexicais encontrados no corpus utilizado na presente pesquisa que estão em co-extensividade às ordens da escala gramatical – morfema, palavra, grupo, oração.

QUADRO 1 – Exemplos de co-extensividade entre itens lexicais e a escala de ordens da gramática

escala de ordens da gramática	item lexical	significado semântico
morfema	autocuidado	'auto-' = próprio
palavra	açúcar	= alimento
grupo/frase	pé diabético	= sintoma
oração	se deus quiser	= provavelmente

Fonte: Rodrigues, Figueredo e Oliveira (no prelo)

A primeira coluna do QUADRO 1 traz as unidades da escala de ordens da gramática – morfema, palavra, grupo/frase e oração. A segunda coluna desse quadro mostra a co-extensividade de quatro itens lexicais retirados do corpus da presente pesquisa e a respectiva unidade da escala de ordens da gramática que o precede na coluna anterior. Essa co-extensividade é representada pelo negrito inserido em cada um dos itens lexicais dessa coluna. A terceira coluna identifica o significado semântico extraído a partir do exame do contexto e cotexto onde os itens lexicais da segunda coluna apareceram nas linhas de concordância do corpus compilado para esta pesquisa (ver 4.1). De maneira mais detalhada, ‘auto’ em ‘autocuidado’ está em co-extensividade à ordem do morfema e significa ‘próprio’. ‘Açúcar’ está em co-extensividade à ordem da palavra e pode indicar um ‘alimento’ ou uma ‘substância’, dependendo do tipo de texto em que esse item aparece. ‘Pé diabético’ está em co-extensividade à ordem do grupo e pode compreender um ‘sintoma’ frequente em usuários com a condição do Diabetes Mellitus tipo II. ‘Se deus quiser’ está em co-extensividade à ordem da oração e pode revelar uma incerteza através do significado semântico ‘provavelmente’.

É importante mencionar que a co-extensividade do item lexical no português brasileiro também pode se expandir aos demais estratos linguísticos para além do estrato gramatical, como, por exemplo, o estrato semântico. A co-extensividade entre o item lexical e o estrato da semântica pode ser identificada pelas relações COESIVAS LEXICAIS (LEXICAL COHESION) de hiponímia, antonímia e colocação, por exemplo (ver 5.1, 5.2.1 e 5.2.2). Há ainda no português brasileiro, a co-extensividade entre o item lexical e o contexto, a qual é mediada pelas variáveis do contexto de situação – campo, sintonia, modo – que podem influenciar na emergência de itens lexicais na língua, sobretudo, os itens lexicais formados por mais de um componente lexical (ver 5.1). A FIG. 2 adiante mostra como o item lexical “diabetes”, retirado do corpus utilizado nesta pesquisa, pode estar em co-extensividade ao contexto, bem como aos estratos linguísticos da semântica, gramática, fonologia e grafologia identificados por Halliday & Matthiessen (2014, p. 26).

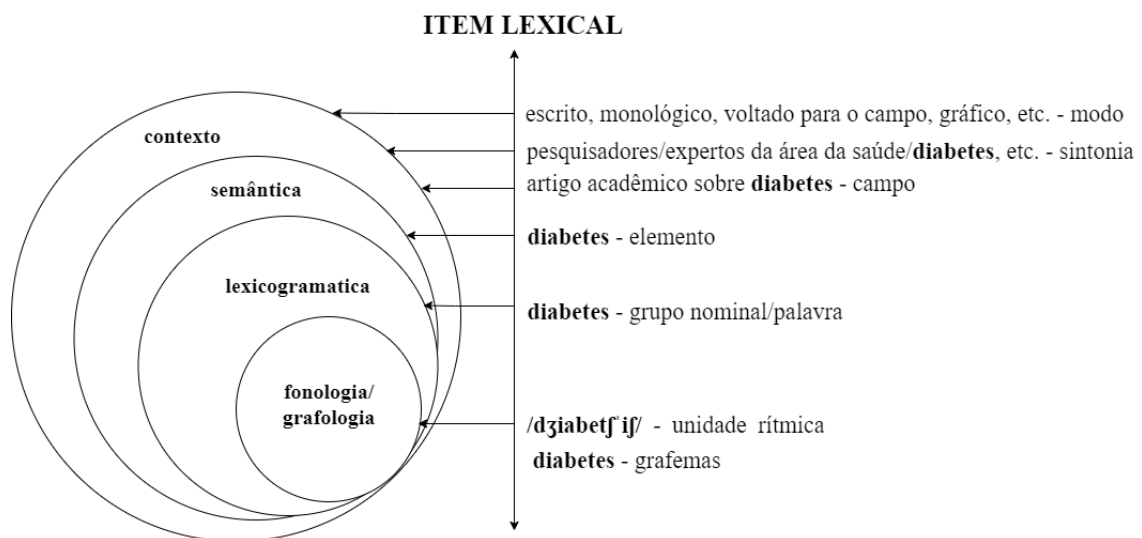


FIGURA 2 – A co-extensividade do item lexical “diabetes” em relação ao contexto e aos estratos linguísticos da semântica, lexicogramática e fonologia/grafologia

Fonte: Adaptada de Halliday & Matthiessen (2014, p. 26)

A FIG. 2 mostra como a língua é organizada por meio dos estratos – fonológico/grafológico, lexicogramatical, semântico e pelo contexto. Há uma escala de ordens para cada um desses estratos. A saber:

Grafologia: sentença (sentence), sub-sentença (sub-sentença), grafema (word) e letra (letter) (cf. HALLIDAY & MATTHIESSEN, 2014, p. 6).

Fonologia: grupo tonal (tone group), pé (foot ou rhythmic unit), sílaba (syllable) e fonema (phoneme) (cf. MATTHIESSEN, TERUYA e LAM, 2010; HALLIDAY & MATTHIESSEN, 2014)

Lexicogramática: oração, grupo/frase preposicional, palavra e morfema (cf. (FIGUEREDO, 2011; HALLIDAY & MATTHIESSEN, 2014)

Semântica: texto (text), paragrafos (parágrafos), sequência (sequence) e figura (figure) (cf. MATTHIESSEN, TERUYA e LAM, 2010; HALLIDAY & MATTHIESSEN, 2014).

Contexto: contexto de situação – campo, sintonia e modo (cf. HALLIDAY, 1978; HALLIDAY & MATTHIESSEN, 2014))

Além disso, essa figura revela que o item lexical “diabetes” pode estar em co-extensividade ao contexto e aos estratos linguísticos da semântica, da lexicogramática, da fonologia e da grafologia. Em relação ao contexto, o item lexical “diabetes” pode identificar o domínio experiencial (variável campo) que é predominante nos artigos acadêmicos (variável campo: processo sociossemiótico), tipo de texto que forma o corpus utilizado nesta pesquisa, bem como apontar a área de expertise dos autores e leitores dos artigos do corpus (variável sintonia) e, pela combinação dos valores identificados pelas variáveis campo e sintonia, auxiliar na configuração sistêmica dos artigos que compõem o corpus (variável modo). No que diz respeito ao estrato da semântica, o item lexical “diabetes” pode ser co-extensivo a um elemento (element) da figura. Para o estrato da lexicogramática, o item “diabetes” pode ser co-extensivo à ordem do grupo e/ou da palavra, podendo ser co-extensivo ainda à função do Participante (Participant) do sistema de TRANSITIVIDADE (TRANSITIVITY). Por fim, em relação aos estratos da fonologia e da grafologia, o item “diabetes” pode ser co-extensivo a um grafema escrito com oito letras, que quando verbalizado, pode ser co-extensivo à unidade rítmica encontrada na FIG. 2.

3.3 O léxico e a gramática: a perspectiva modular adotada pela gramática normativa e pela gramática funcionalista no âmbito do português brasileiro

Uma das formas mais frequentes de abordagem do item lexical é concebê-lo como uma unidade modular, estando assim associada a um conjunto de morfemas (cf. BASÍLIO, 1991; BORBA, 2003; CASTILHO, 2010). Contudo, essa concepção implica na relação direta entre o léxico e um estrato específico – a gramática – e, dentro desse estrato, uma unidade

gramatical particular – a palavra. Dessa maneira, apesar de compartilharem recursos de produção de significado, o item lexical e a palavra operam papéis distintos na organização da língua.

Na obra *Introdução aos Estudos Linguísticos*, Borba (2003, p. 229), lexicógrafo e gramático funcionalista, sugere que toda unidade léxica tem um valor codificado que é seu significado básico. A partir daí o uso propicia várias associações que se incorporam ao signo, passando a fazer parte de suas possibilidades significativas. Para esse autor, uma unidade tem significado lexical quando se refere a alguma coisa do mundo exterior, quer dizer a uma representação linguística da realidade. As unidades lexicais estão agrupadas em conjuntos abertos, ou seja, conjuntos que têm um número ilimitado de membros e podem ser formadas por conjuntos de morfemas relacionados, isto é, as unidades lexicais podem ser vistas como uma composição de morfemas, reafirmando a perspectiva modular do léxico. Além das unidades lexicais, este autor menciona a existência das unidades gramaticais, as quais têm seus valores atrelados a um sistema linguístico e pertencem a conjuntos fechados (BORBA, 2003, p. 230).

Na visão funcional-cognitivista de Castilho (2010) a palavra, ou o item lexical, é uma unidade que associa idealmente um conjunto de morfemas. Este autor divide o léxico sob três perspectivas (CASTILHO, 2010, p. 109):

Léxico mental: seu objeto são matrizes cognitivas armazenadas no cérebro, associadas à sua representação linguística. Estuda-se o léxico mental focalizando a associação entre essas matrizes e as formas e os significados lexicais e gramaticais ativados durante a produção e compreensão da língua.
Léxico dos lexicógrafos: seu objeto é o emparelhamento das formas com seu sentido comum na língua, focalizando os lexemas, as expressões idiomáticas e os morfemas derivacionais como, por exemplo, -dade e -mente.
Léxico dos gramáticos: seu objeto é o emparelhamento das formas com seus sentidos gramaticais, focalizando os lexemas simples, as palavras funcionais e as regras morfológicas de derivação (CASTILHO, 2010, p. 109).

Na obra *Nova Gramática do Português Brasileiro*, Castilho (2010, p. 110) ocupa-se das delimitações acerca dos conceitos de palavra e de item lexical/léxico. Esse autor postula que o léxico é compreendido como “um inventário (i) de categorias e subcategorias cognitivas e (ii) de traços semânticos inerentes”. Esse inventário integra propriedades utilizadas para a criação de palavras, portanto, a palavra é formada a partir do léxico de uma língua (CASTILHO,

2010). O item lexical é distinto da palavra morfológica (+núcleo; +afixos), fonológica (+sílaba tônica; + sílabas átonas) ou gráfica (espaços em branco ao redor de um elemento gráfico, cf. token¹⁵). O item lexical pode constituir-se de mais uma palavra em sua totalidade e pode ser formado por uma única palavra (CASTILHO, 2010).

3.3.1 Os processos de expansão lexical sob a perspectiva da gramática funcionalista no âmbito do português brasileiro: lexicalização e gramaticalização

Alguns gramáticos brasileiros (NEVES, 1997; BASÍLIO, 1991; BORBA, 2003; CASTILHO, 2010; entre outros) apontam que o item lexical se relaciona a outras unidades da gramática, sendo concebido de forma psicológica e individual; além de ser dependente da semântica, do cotexto e do contexto para ganhar sentido, sendo capaz de ser compreendido de maneira coletiva e social. Considerando essas visões, o processo de expansão lexical pode ser mediado por duas perspectivas: a lexicalização e a gramaticalização.

Na obra *Nova Gramática do Português Brasileiro*, Castilho (2010, p. 110) define o léxico como:

um conjunto de categorias cognitivas e traços derivados que são representados nas palavras por meio da lexicalização. O vocabulário é o produto do léxico, compendiado nos dicionários de língua (CASTILHO, 2010, p. 110).

Nesse sentido, “a lexicalização é o processo de criação de palavras, coordenada pelo dispositivo sociocognitivo” (CASTILHO, 2010, p. 113). Segundo esse autor, o processo de lexicalização é compreendido em:

(i) ativação ou lexicalização propriamente dita, que é a escolha de categorias cognitivas e seus traços semânticos, representando-os nas palavras; (ii) a reativação ou relexicalização, entendida como uma nova ativação dessas categorias; (iii) a desativação ou deslexicalização, que é a eliminação ou a substituição de categorias, momento que anuncia a morte da palavra (CASTILHO, 2010, p. 113).

¹⁵*Token* é um conjunto de caracteres entre espaços em branco que, independentemente; de seu significado semântico, é reconhecido e contabilizado pelo computador (VIANA e TAGNIN, 2010).

Nesse trabalho, Castilho (2010, p. 113) postula ainda que a lexicalização percorre os seguinte caminhos: (i) etimologia, quando a lexicalização ocorre na língua-fonte; (ii) neologia, quando a lexicalização ocorre na língua-alvo e (iii) empréstimo, quando a lexicalização ocorre por contato linguístico.

Por outro lado, a perspectiva da gramaticalização adotada pelas gramáticas clássicas compreende um processo no qual um “item lexical é visto como a extensão de um morfema que antes era considerado lexical e passa a ser concebido como gramatical, ou que antes era entendido como menos gramatical e passa a ser compreendido como mais gramatical”¹⁶ (cf. KURYLOWICZ, 1975, p.52; HOPPER e TRAUGOTT, 1993). Para Hopper e Traugott (1993, p. 2), há duas perspectivas para o estudo da gramaticalização, a “histórica”, que estuda as origens das formas gramaticais e as mudanças típicas que afetam essas formas gramaticais; e a “mais sincrônica”, que enfoca nos padrões fluidos do uso linguístico provenientes do processo de gramaticalização.

Em A Gramática Funcional, Neves (1997, p. 115) menciona que a gramaticalização se refere à parte da teoria da linguagem que tem por objetivo a interdependência entre *langue* (língua) e *parole* (palavra), entre o categorial e o menos categorial, entre o fixo e o menos fixo na língua. A autora acrescenta que o estudo da gramaticalização, portanto, põe em evidência a tensão entre a expressão lexical, relativamente livre de restrições, e a codificação morfossintática, mais sujeita a restrições, salientando a indeterminação relativa da língua e o caráter não-discreto de suas categorias.

Ainda com relação à expansão lexical, Borba (2003, p. 160) sugere que por meio do processo de morfologia lexical, que se dá pelo mecanismo básico da derivação, novas palavras são formadas (e.g. saúde, saudar, saudação, saudável, etc.). Borba (2003, p. 164) menciona também um segundo mecanismo no escopo da morfologia lexical capaz de contribuir com a expansão lexical, chamado de composição. Na composição, raízes presas ou livres são combinadas para a obtenção de novas formações, havendo uma aglutinação ou justaposição de componentes (e.g. aguardente, girassol, couve-flor, etc.).

Além dos processos de lexicalização e gramaticalização, Basílio (1991, p. 31) concebe a prefixação como um processo de expansão lexical. Nesse processo, palavras são formadas “através do acréscimo de um prefixo a uma base, respeitando as devidas restrições morfológicas, semânticas, etc”.

¹⁶ Minha tradução para: “Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status” (KURYLOWICZ, 1975, p. 52).

Nesse sentido, a perspectiva modular concebe o léxico e a gramática a partir de visões distintas, as quais não operam juntas, havendo inclusive diferentes abordagens para esses dois aspectos. Isso significa dizer que o léxico de uma língua está descrito nos dicionários e tesouros; e compreende um dos objetos de estudo dos lexicógrafos e terminologistas, ao passo que a gramática de uma língua está compilada nos livros de gramática; e corresponde a um dos objetos de estudo daqueles que lidam com a sintaxe da língua.

Por outro lado, no âmbito da TSF, a perspectiva contígua, adotada pela presente pesquisa, engloba a abordagem complementar entre léxico e gramática e enxerga o léxico como uma categoria teórica que carece de estudos capazes de mostrar como se dá a organização interna do léxico e do item lexical. Diante disso, o próximo capítulo é responsável por apresentar os passos metodológicos desenvolvidos para investigar o funcionamento do léxico/item lexical no português brasileiro.

4. METODOLOGIA

Segundo os pressupostos preconizados pela linguística sistêmico-funcional (HALLIDAY, MCINTOSH e STREVENS, 1964; HALLIDAY, 2002), o desenvolvimento de testes (probes), nesta pesquisa, chamados de protocolos de investigação lexical, podem ser entendidos como uma maneira de aprofundar os estudos acerca do léxico como conceito teórico no âmbito da língua portuguesa do Brasil (FIGUEREDO *et al.* 2019). Neste sentido, torna-se necessário delimitar o conjunto de textos a ser utilizado para a aplicação dos protocolos de investigação lexical, bem como o domínio experiencial desses textos.

A primeira parte desta metodologia explora os dois corpora utilizados neste estudo. O primeiro deles compreende o *corpus* compilado para o desenvolvimento dos protocolos de investigação lexical, sendo um *corpus* formado por artigos acadêmicos sobre autocuidado em diabetes mellitus. O segundo *corpus* recebe, nesta pesquisa, o nome de *corpus* de suporte, e diz respeito a um *corpus* online que foi utilizado como apoio para a aplicação dos protocolos de investigação lexical em conjunto com o primeiro *corpus*.

A segunda parte desta metodologia aborda as explicações acerca do processo de criação dos protocolos de investigação lexical elaborados para o presente trabalho. O primeiro protocolo investiga os mecanismos pelos quais um item lexical no português brasileiro consegue operar como um item lexical formado por mais de um componente lexical, o que possibilita a busca pela co-extensividade entre um item lexical e a ordem do grupo nominal, levando ao exame das variáveis do contexto de situação, responsáveis pelas configurações que constituem os textos do *corpus* compilado, na formação de itens lexicais com mais de um componente lexical. O segundo protocolo verifica a maneira pela qual itens lexicais no português brasileiro são concebidos, ou seja, o que leva um item a ser chamado de um item lexical, não de um item gramatical, por exemplo, averiguando a(s) porção(ões) que forma(m) esse item lexical e as características que os possibilitam ser tratados como parte de um léxico.

4.1 O *corpus* compilado para o desenvolvimento e a aplicação dos ProtinLexis

Os textos selecionados para o desenvolvimento e a aplicação dos protocolos de investigação lexical compreendem artigos acadêmicos que retratam pesquisas desenvolvidas no âmbito do Diabetes Mellitus tipo II. O domínio experiencial (HALLIDAY & MATTHIESSEN, 1999; HAO, 2015) destes textos é, portanto, essa condição crônica, mais especificamente, o autocuidado em Diabetes Mellitus.

É necessário ressaltar a importância de se compilar um corpus cujo o domínio experiencial dos textos seja um denominador comum. No caso desta pesquisa, esse denominador em comum reside no fato de que os artigos que compõem o corpus utilizado apresentam como tema principal o diabetes, sobretudo o autocuidado em relação a essa condição. Essa delimitação do domínio experiencial dos textos do *corpus* torna-se relevante do ponto de vista da homogeneidade do número de itens lexicais utilizados para a aplicação dos protocolos de investigação lexical, uma vez que textos que estão em um mesmo domínio experiencial tendem a apresentar uma repetição de alguns itens lexicais específicos em cotextos diferentes, ou seja, itens lexicais que possuem a mesma grafia aparecem em orações diferentes. Isso possibilita a elaboração de uma lista de palavras-chave com itens lexicais cujo número de ocorrências seja suficientemente adequado para elaboração e aplicação dos protocolos de investigação lexical.

Esses artigos acadêmicos foram extraídos de publicações da área das Ciências da Saúde a partir da busca pela palavra-chave “autocuidado em diabetes mellitus” feita na parte de artigos acadêmicos disponíveis na aba de busca do Google Acadêmico.¹⁷ Foram coletados 40 artigos originalmente escritos em português brasileiro. Esses 40 artigos correspondem à quantidade mínima suficiente para a obtenção de resultados em função dos protocolos de investigação lexical propostos neste estudo. Isso significa dizer que, com menos de 40 artigos, os resultados encontrados não são capazes de mostrar hipóteses ou indicar conclusões acerca da caracterização do item lexical no PB.

É necessário ressaltar a importância de se compilar um corpus cujo o domínio experiencial dos textos seja um denominador comum, no caso desta pesquisa, esse denominador em comum reside no fato de que os artigos que compõem o corpus utilizado apresentam como tema principal o diabetes, sobretudo o autocuidado em relação a essa condição. Essa delimitação do domínio experiencial dos textos do *corpus* torna-se relevante do ponto de vista da homogeneidade do número de itens lexicais utilizados para a aplicação dos protocolos de investigação lexical, uma vez que textos que estão em um mesmo domínio experiencial tendem a apresentar uma repetição de alguns itens lexicais específicos em cotextos diferentes, ou seja, itens lexicais que possuem a mesma grafia aparecem em orações diferentes. Isso possibilita a elaboração de uma lista de palavras-chave com itens lexicais cujo número de ocorrências seja suficientemente adequado para elaboração e aplicação dos protocolos de investigação lexical.

¹⁷ Disponível em: <https://scholar.google.com.br/?hl=pt>

Para que os itens lexicais investigados nesta tese pudessem indicar propriedades lexicais identificadas na última década, os artigos que constituem o *corpus* compilado para o presente estudo foram publicados entre 2010 e 2019, gerando, ao todo, 133.232 *tokens*, um corpus classificado como pequeno-médio, de acordo com a classificação proposta por Sardinha (2003). A escolha por esse tipo de texto para compor o *corpus* se deu pela necessidade de se ter acesso a uma linguagem utilizada entre pares, cujo nível de expertise fosse semelhante, no caso, ambos expertos, ou seja, autor(es) e leitor(es) desse tipo de texto estão no ambiente acadêmico e pertencem a uma determinada área do conhecimento, a área das Ciências da Saúde, na qual a expertise de ambos está localizada. Delimitar a seleção de textos cujos valores identificados para as subvariáveis da variável sintonia – papel institucional, papel social, papel discursivo, papel sociométrico, distância social e valoração – estivessem bem definidos no sistema do português brasileiro constituiu um parâmetro importante para os resultados desta pesquisa pois trabalhando com níveis de expertise semelhantes entre quem produz e quem lê o texto, no caso, autor e leitor de artigos acadêmicos da área das Ciências da Saúde, as subvariáveis da variável sintonia apresentariam resultados semelhantes, ou seja, resultados que não privilegiariam (ou quase não privilegiariam) o papel do autor em detrimento do papel leitor, ou vice-versa. Isso não aconteceria numa reportagem jornalística, por exemplo, em que o autor do texto poderia apresentar maior conhecimento dos acontecimentos reportados do que o leitor que busca a reportagem para se inteirar de um dado acontecimento.

Outra razão que motivou a escolha pelo tipo de texto que compõe o *corpus* compilado para este trabalho compreende o fato de que por se tratar de uma linguagem utilizada por pares considerados expertos na área de domínio dos artigos, esperava-se encontrar uma grande variedade de itens lexicais passíveis de serem utilizados na aplicação dos protocolos de investigação lexical. Uma última razão que justifica a escolha por artigos acadêmicos para o corpus desta pesquisa diz respeito à facilidade de acesso desse tipo de texto sob esse domínio em particular, sobretudo por ser um texto escrito e sem diálogos, o que faz com que uma demanda de tempo e esforço cognitivo, que seriam gastos para transcrição e análise de trocas de turno, pudesse ser alocada para outros aspectos do presente estudo.

Cada artigo foi salvo separadamente em um arquivo no bloco de notas e nomeado com o “Título da Revista” e o “ano de publicação”, por exemplo, “Revista de Atenção Primária à Saúde_2014”. Quando havia dois ou mais artigos de um determinado periódico publicados no mesmo ano, a nomeação se deu da seguinte forma Título da Revista_ano de publicaçãob, a saber, “Revista Brasileira de Enfermagem_2010b”.

É importante destacar que os *abstracts*, quadros, gráficos, anexos, figuras, tabelas e referências de todos os artigos selecionados para o *corpus* foram excluídos do arquivo do bloco de notas, por não apresentarem dados que pudessem contribuir para alcançar o objetivo do trabalho em questão. Isso significa dizer que essas seções não tinham uma grande quantidade de itens lexicais que justificassem o uso delas no desenvolvimento e na aplicação dos protocolos de investigação lexical. Os *abstracts*, por exemplo, contavam com itens lexicais do inglês, fugindo do escopo do presente trabalho que é tratar no PB. Os gráficos e as tabelas apresentavam uma maior quantidade de dados numéricos que itens lexicais. Os quadros e as figuras tinham, muitas vezes, uma diagramação específica que teria que ser modificada quando fosse para o arquivo do bloco de notas, o que poderia gerar confusão na extração e análise dos itens lexicais provenientes desses quadros e figuras. Por fim, os anexos e as referências nem sempre apresentavam informação textual passível de ser entendida de maneira isolada do restante do artigo correspondente, por isso optou-se por descartá-los do *corpus* compilado para a presente pesquisa.

4.2 O *corpus* de suporte para a aplicação dos ProtinLexis

Para a presente pesquisa, um segundo *corpus* se fez necessário para que a aplicação dos protocolos de investigação lexical não ficasse somente restrita aos textos do *corpus* compilado para este estudo. Nesse sentido, um *corpus*, chamado, neste trabalho, de *corpus* de suporte foi utilizado para a busca de exemplos que pudessem ser empregados nos processos de mudança elaborados no âmbito de cada um dos ProtinLexis (ver 4.3.1 e 4.3.2).

Como esse *corpus* de suporte devia estar online para otimizar as buscas e precisava apresentar um panorama geral do português brasileiro, com diversos tipos de textos sob temas diferentes daquele abordado do *corpus* compilado para este estudo, o gerenciador online de análise de textos, o Sketch Engine¹⁸, foi o *corpus* de suporte selecionado para exercer tal função neste trabalho.

A escolha por essa plataforma online foi motivada pela variedade, quantidade e atualidade dos textos escritos em português brasileiro que o site apresenta. No que diz respeito ao PB, o Sketch Engine apresenta três corpora diferentes – OPUS2 Brazilian Portuguese, Brazilian Portuguese Corpus (Corpus Brasileiro) e Open Subtitles 2018 Brazilian Portuguese – ao todo, somam mais de um bilhão de palavras. Para além disso, o Sketch Engine conta com ferramentas que no caso específico do objeto de pesquisa desta tese foram capazes de auxiliar

¹⁸ Disponível em: <https://www.sketchengine.eu/>.

na busca por itens lexicais com certas particularidades, como, por exemplo, itens lexicais formados por mais de um componente lexical.

4.3 O desenvolvimento de um protocolo de investigação lexical no escopo da TSF

O primeiro passo para desenvolver um protocolo de investigação lexical é ter em mente a literatura existente acerca do léxico no âmbito na teoria sistêmico-funcional (TSF), como, por exemplo, os tipos de COESÃO LEXICAL, e; sobretudo, o conceito de colocação (cf. HALLIDAY e HASAN, 1976; HALLIDAY & MATTHIESSEN, 2014), bem como os termos técnicos conjunto lexical, série e colocados apresentados em Halliday (1961), Halliday, McIntosh e Strevens (1964) e Halliday (1966).

Em seguida, ocorre a criação dos protocolos de investigação lexical, a qual se dá por meio do processo de produção de significado particular do item lexical no PB, em contraponto ao chamado item gramatical. Isso significa dizer que a partir dos textos selecionados para o *corpus* utilizado na presente pesquisa, mais especificamente a partir da lista de palavras-chave, que mostra os itens com maior número de ocorrência no *corpus*, uma análise das propriedades desses itens presentes na lista de palavras-chave é feita, distinguindo as propriedades linguísticas que podem ser específicas de itens compreendidos como lexicais daquelas que podem ser específicas de itens entendidos como gramaticais.

Considerando a visão que concebe o léxico como uma dimensão, a dimensão lexical, bem como o pressuposto identificado no trabalho de Halliday (1966) de que o léxico pode ser visto sob uma perspectiva teórica na qual o item lexical funciona como uma instância do léxico e pode ser abordado de uma maneira distinta à gramática, não como o ponto mais delicado no sistema gramatical, o presente trabalho tenta desenvolver uma metodologia que seja capaz de caracterizar o item lexical no português brasileiro, investigando a relação que pode existir entre um item lexical como instância do léxico e a relação da chamada dimensão lexical com a dimensão da estratificação do sistema linguístico sobretudo no que diz respeito ao estrato da gramática, bem como a relação da dimensão lexical com as variáveis contextuais (campo, sintonia e modo). Por isso, dois grupos de perguntas nortearam o desenvolvimento dos protocolos de investigação lexical propostos neste trabalho:

I. Estrato gramatical

- Este item em questão pode ser transformado em outro(s) item(ns) com funções diferentes na ordem da oração – ora funcionando como Participante, ora como

Circunstância, por exemplo – sem que a porção lexical desse(s) item(ns) lexical(is) transformado(s) seja modificada?

- Considerando o arcabouço teórico da LSF, seria possível identificar essas transformações? Se sim, como isso seria feito? Poderia ser por meio da COESÃO LEXICAL e/ou por algum processo de mudança na escala de ordens gramatical, como, por exemplo, um item lexical que estava em co-extensividade com a ordem da palavra passaria a ser co-extensivo à ordem do grupo (nominal)?
- Há coocorrências deste item lexical em questão com outros itens considerados lexicais? Se sim, essa(s) coocorrência(s) ocorre(m) com frequência no corpus a ponto de ser(em) identificada(s) como um *cluster/n-grama* por um *software* concordanciador?
- Se a existência dessas coocorrências for identificada, essas coocorrências são formadas, na sua maioria, por quantos itens considerados lexicais?
- Esses itens lexicais formados por mais de um componente lexical podem estar em co-extensividade às unidades do grupo (nominal) apontada primeiramente por Halliday, McIntosh e Stevens (1964)?

II. Contexto de situação

- As variáveis do contexto de situação são capazes de influenciar na emergência de itens lexicais formados por mais de um componente lexical no corpus, e, conseqüentemente, no português brasileiro?
- Os itens lexicais transformados podem estar relacionados ao domínio experiencial dos textos que constituem o corpus utilizado nesta pesquisa, assim como os itens lexicais selecionados para o desenvolvimento e aplicação dos protocolos de investigação lexical estão atrelados ao domínio experiencial do autocuidado em diabetes mellitus?

É válido explicar que esses dois conjuntos de perguntas foram desenvolvidos mediante a observação de padrões que eram identificados nos itens chamados de lexicais e nos itens denominados de gramaticais. Em outras palavras, a partir dos textos do *corpus* compilado para a presente pesquisa e das ferramentas disponíveis no AntConc, uma análise geral de todos os itens presentes nesses textos do *corpus* foi realizada para que se pudesse verificar a existência de características que poderiam ser entendidas como prototípicas de itens lexicais e/ou prototípicas de itens gramaticais, se sim, quais eram essas características.

Para a aplicação dos protocolos de investigação lexical desenvolvidos para a presente pesquisa, recursos da Linguística de *Corpus* também foram utilizados de maneira metodológica, auxiliando e embasando a seleção dos itens a serem investigados nos protocolos de investigação lexical. Essa metodologia foi seguida no desenvolvimento dos dois protocolos de investigação lexical (ver 4.3.1 e 4.3.2) e será detalhada a seguir.

Após a importação dos 40 artigos sobre autocuidado em Diabetes Mellitus (ver 4.1) para o *software* concordanciador AntConc (ANTHONY, 2019), uma lista de palavras-chave foi gerada pelo *software*. Para tanto, a parte escrita monológica do CALIBRA¹⁹ (cf. FIGUEREDO, PAGANO e FERREGUETTI, 2014), composta por 235 textos, foi utilizada como *corpus* de referência, dado que para gerar uma lista de palavras-chave o *software* concordanciador precisa de um *corpus* mais robusto, chamado de *corpus* de referência, para que possa haver uma comparação interna entre os *tokens* do *corpus* utilizado na pesquisa e os *tokens* do *corpus* de referência, assim; uma lista de palavras-chave é estabelecida pelo *software* concordanciador. É importante mencionar que, antes de gerar essa lista de palavras-chave, um arquivo salvo em .txt²⁰ com uma série de itens denominados como funcionais²¹ foi inserido no AntConc (aba *tool preferences* > *word list* > *word list range*), para que na lista de palavras-chave gerada pelo *software* aparecessem apenas os itens passíveis de serem utilizados no desenvolvimento e aplicação dos protocolos de investigação lexical.

Seguindo os preceitos apontados por Gabrielatos (2018, p.238) acerca das métricas comumente usadas para identificar as palavras-chave de um *corpus*, o p-valor frequentemente empregado nas pesquisas no âmbito da linguística de corpus para estabelecer uma lista de palavras-chave é de no máximo 0.01, sendo que quanto menor é o p-valor, maior tende a ser a quantidade de itens com chavidade (keyness) elevada que aparecem na lista de palavras-chave (GABRIELATOS, 2018, p. 238). Nesse sentido, para gerar a lista de palavras-chave desta pesquisa, vários cenários foram considerados, ou seja, simulações foram feitas no software concordanciador AntConc (ANTHONY, 2019), empregando diferentes métricas para o p-valor – p=0.001 e p=0.0001 – em conjunto com diferentes tipos de testes (Keyword Effect Size Measure, no AntConc) – %DIFF, Odds-ratio, Dice coefficient, Difference Coefficient (relative), Difference of relative frequencies, Hardie’s Log-Ratio, Phi coefficient (Cramer’s V),

¹⁹ Segundo Figueredo, Pagano e Ferregueti (2014, p. 320), o CALIBRA (Catálogo da Língua Brasileira) é um corpus que atualmente conta com cerca de 1 milhão de palavras (*tokens*), compilados com base na tipologia do contexto de cultura de Halliday (1978).

²⁰ Disponível em: <https://github.com/stopwords-iso/stopwords-pt>.

²¹ Itens funcionais são entendidos, nesta pesquisa, como aqueles que estão no escopo dos sistemas fechados, como, por exemplo, preposições, conjunções, artigos e advérbios (cf. HALLIDAY et.al. 1964; 2002).

Ratio of Relative frequencies e Yule's Q coeficiente – e sempre respeitando a quantidade máxima de 100 itens por lista de palavras-chave²². A avaliação desses cenários apontou que as listas de palavras-chave geradas permaneceram com a mesma sequência de itens quando o p-valor foi $p=0.001$ ou $p=0.0001$, havendo mudanças de itens apenas quando os testes de Difference Coefficient (relative), Difference of relative frequencies, Phi coefficient (Cramer's V) e Yule's Q coeficiente foram aplicados. Em razão dessa avaliação, a seleção da lista de palavras-chave se deu por aquela que apresentou maior concordância entre os testes, isto é, a que foi gerada pelos testes de %DIFF, Odds-ratio, Hardie's Log-Ratio e Ratio of Relative frequencies (ver Anexos).

Tendo estabelecido a lista de palavras-chave, um pouco mais de 10% dos itens dessa lista de palavras-chave foi selecionado para ser utilizado no desenvolvimento e aplicação dos protocolos de investigação lexical elaborados para esta pesquisa. É importante explicar que somente uma porcentagem da lista de palavras-chave foi selecionada para a aplicação dos protocolos de investigação lexical em razão das análises manuais previstas no escopo dos protocolos. A escolha por determinados itens da lista de palavras-chave se deu por dois motivos, um primário e outro secundário. O primário compreende o número de ocorrências desses itens na lista de palavras-chave, ou seja, a alta frequência com que eles aparecem na lista de palavras-chave, possibilitando uma variedade de cotextos diferentes para o desenvolvimento e aplicação dos protocolos de investigação lexical. O motivo secundário diz respeito ao fato desses itens estarem localizados no escopo do domínio experiencial – autocuidado em diabetes mellitus – dos textos do corpus compilado para esta pesquisa, o que viabilizaria uma análise semântica, caso fosse necessário.

A seguir, as propriedades dos dois protocolos de investigação lexical desenvolvidos para a presente pesquisa são descritas.

4.3.1 Do Protocolo de Investigação Lexical: Processo de Transformação do Contexto - ProtinLex:PTC

O processo metodológico que engloba o desenvolvimento do ProtinLex:PTC, sigla para Protocolo de Investigação Lexical: Processo de Transformação do Contexto, é formado por duas fases, o desenvolvimento e a aplicação do protocolo. No desenvolvimento,

²² O número de cem itens por lista de palavras-chave compreende a quantidade mínima de itens possíveis de aparecer em uma lista de palavras-chave gerada pelo AntConc. A seleção por esse número se deu em função da presente pesquisa priorizar, em um segundo momento, após a geração da lista de palavras-chave, uma análise manual para o desenvolvimento e a aplicação de protocolos de investigação lexical, utilizados, neste trabalho, como metodologia para a caracterização do léxico e/ou item lexical no PB.

observaram-se as distinções entre os itens considerados gramaticais e aqueles considerados lexicais, isto é, as características que fazem com que determinados itens do PB se localizem no conjunto dos itens gramaticais, *e.g.* itens da classe de palavras dos artigos e das preposições, e outros itens possam ser inseridos no conjunto dos itens lexicais, *e.g.* itens da classe dos substantivos e dos adjetivos. No que diz respeito à aplicação, com exceção no item “mellitus”, os demais itens presentes na lista de palavras-chave gerada anteriormente foram utilizados (ver QUADRO 2). A exclusão do item “mellitus” para o desenvolvimento e a aplicação dos ProtinLexis foi motivada pelo fato de que esse item sempre apareceu seguido do item “diabetes”, indicando que para os textos do *corpus* utilizado nesta pesquisa esses dois itens lexicais não podiam ser tratados de maneira isolada, uma vez que sempre apareciam juntos no *corpus* e instanciavam um único significado – a condição crônica nomeada de diabetes mellitus.

O primeiro passo da aplicação deste protocolo de investigação lexical envolveu a busca pelos agrupamentos (*clusters/n-grams*)²³ formados à direita e à esquerda de cada um dos itens selecionados para a aplicação dos protocolos. Essa busca teve como um dos objetivos investigar a ocorrência no português brasileiro da propriedade de co-extensividade entre itens lexicais formados por mais de um componente lexical e o eixo sintagmático (estrutura) no escopo da dimensão local da linguagem sobretudo a co-extensividade em relação à ordem do grupo (nominal), o que já havia sido observado por Halliday (1961, p. 59-60) para o inglês. A opção pela busca de *clusters/n-grams* como forma de verificar essa co-extensividade se deu em razão desses agrupamentos mostrarem o contexto onde os itens da lista de palavras-chave selecionados para as aplicações dos protocolos de investigação lexical (ver QUADRO 4) se encontravam, ou seja, quais os possíveis itens que apareciam logo antes e/ou logo depois do item de busca e se a coocorrência desses itens com o item de busca em questão poderia estar em co-extensividade com a ordem do grupo (nominal). As FIG. 3 e 4, a seguir, mostram como essa pesquisa foi feita para o item de busca “autocuidado”.

²³ Minha tradução para *clusters/n-grams*. *Clusters/n-grams* diz respeito aos agrupamentos de palavras que ocorrem em sequência repetidas vezes em um corpus (cf. SARDINHA, 2004, p. 88). Nesses agrupamentos podem conter tanto palavras lexicais (conteúdo) quanto gramaticais (funcionais), a depender da extensão da sequência, nesta pesquisa, de duas até cinco palavras.

Corpus Files

- Acta Paulista de Enl
- Acta Paulista de Enl
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Arquivo Ciê
- Revista Baiana de S
- Revista Brasileira d
- Revista Brasileira d
- Revista Brasileira d
- Revista Brasileira d
- Revista Cadernos B
- Revista Ciências e S
- Revista Cinerjis_20
- Revista Cogitare En
- Revista Cogitare En
- Revista da Associaç
- Revista da Escola d
- Revista da Escola d
- Revista da Escola d
- Revista da Escola d
- Revista da Faculdac
- Revista da Rede de
- Revista da Rede de
- Revista da Rede de
- Revista de Atenção
- Revista de Ciência e
- Revista de Ciência e
- Revista de Enferma
- Revista Eletronica E
- Revista Enfermaner

Total No.
40
Files Processed

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

Total No. of Cluster Types 659 **Total No. of Cluster Tokens** 1020

Rank	Freq	Range	Cluster
1	60	15	autocuidado em diabetes
2	35	7	autocuidado com os
3	27	10	autocuidado com o
4	18	2	autocuidado do idoso
5	14	8	autocuidado de pacientes
6	9	3	autocuidado dos pacientes
7	9	6	autocuidado em dm
8	8	4	autocuidado da doença
9	8	7	autocuidado para o
10	7	6	autocuidado de pessoas
11	7	4	autocuidado e controle
12	6	2	autocuidado com diabetes
13	6	4	autocuidado da pessoa
14	6	4	autocuidado das pessoas
15	6	2	autocuidado do paciente
16	6	5	autocuidado do usuário
17	6	4	autocuidado e os
18	6	5	autocuidado em relação
19	6	1	autocuidado às pessoas

Search Term Words Case Regex N-Grams **Cluster Size**
autocuidado Advanced Min. 3 Max. 3

Start Stop Sort **Min. Freq.** 1 **Min. Range** 1

Sort by Invert Order **Search Term Position**
Sort by Freq On Left On Right

FIGURA 3 – Captura de tela dos tri-gramas quando "autocuidado" está localizado à esquerda

Fonte: Desenvolvida para fins deste trabalho.

Corpus Files

- Acta Paulista de Enl
- Acta Paulista de Enl
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Arquivo Ciê
- Revista Baiana de S
- Revista Brasileira d
- Revista Brasileira d
- Revista Brasileira d
- Revista Brasileira d
- Revista Cadernos B
- Revista Ciências e S
- Revista Cinerjis_20
- Revista Cogitare En
- Revista Cogitare En
- Revista da Associaç
- Revista da Escola d
- Revista da Escola d
- Revista da Escola d
- Revista da Escola d
- Revista da Faculdac
- Revista da Rede de
- Revista da Rede de
- Revista da Rede de
- Revista de Atenção
- Revista de Ciência e
- Revista de Ciência e
- Revista de Enferma
- Revista Eletronica E
- Revista Enfermaner

Total No.
40
Files Processed

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

Total No. of Cluster Types 288 **Total No. of Cluster Tokens** 1023

Rank	Freq	Range	Cluster
1	138	33	para o autocuidado
2	96	17	práticas de autocuidado
3	93	17	atividades de autocuidado
4	62	15	ações de autocuidado
5	29	5	capacidade de autocuidado
6	23	7	promoção do autocuidado
7	20	13	prática do autocuidado
8	17	1	capacidades de autocuidado
9	14	6	prática de autocuidado
10	13	10	sobre o autocuidado
11	12	10	relação ao autocuidado
12	11	7	avaliação do autocuidado
13	11	10	realização do autocuidado
14	10	4	gestão do autocuidado
15	10	5	práticas do autocuidado
16	10	7	relacionadas ao autocuidado
17	8	2	acerca do autocuidado
18	8	7	adesão ao autocuidado
19	8	5	comportamentos de autocuidado

Search Term Words Case Regex N-Grams **Cluster Size**
autocuidado Advanced Min. 3 Max. 3

Start Stop Sort **Min. Freq.** 1 **Min. Range** 1

Sort by Invert Order **Search Term Position**
Sort by Freq On Left On Right

FIGURA 4 – Captura de tela dos tri-gramas quando "autocuidado" está localizado à direita

Fonte: Desenvolvida para fins deste trabalho.

As capturas de tela apresentadas nas FIG. 3 e 4 identificam os itens que coocorrem sequencialmente no corpus – logo antes ou logo depois – do item de busca “autocuidado”, considerando a extensão de três itens que podem vir depois (“autocuidado” à esquerda – FIG. 3) ou antes (“autocuidado” à direita – FIG. 4) de “autocuidado”. É fundamental ressaltar que essas buscas representadas pelas capturas de tela das FIG. 3 e 4 foram feitas para cada um dos onze itens da lista de palavras-chave selecionados para o desenvolvimento e aplicação dos protocolos de investigação lexical apresentados neste trabalho, sendo que essas buscas também foram realizadas com extensão de dois gramas (bigramas) e quatro gramas (quadrigramas)²⁴, para além dos três gramas (trigramas) exemplificados nas capturas de tela.

O segundo passo para a aplicação deste protocolo de investigação lexical diz respeito às análises dos *clusters/n-gramas* formados para cada um dos itens de busca conforme a extensão (bigrama, trigrama, quadrigramas, etc.). Em um primeiro momento, essas análises tiveram como propósito examinar a existência de uma possível correspondência entre os *clusters/n-gramas* e os itens lexicais formados por mais um componente lexical. É necessário esclarecer que se essa correspondência fosse constatada para os itens lexicais utilizados no desenvolvimento e na aplicação dos protocolos de investigação elaborados nesta pesquisa, a busca por *clusters/n-gramas*, funcionalidade que está presente em grande parte dos *softwares* concordanceadores e/ou *corpora* online, serviria como uma forma de pesquisar semiautomaticamente por itens lexicais que apresentassem mais de um componente lexical ao mesmo tempo que também evidenciaria a formação de itens lexicais desse tipo (com mais de um componente lexical) para o português brasileiro, os quais podem estar em co-extensividade com a ordem do grupo (nominal), propriedade apontada por Halliday (1961, p. 59-60) para o inglês. Em um segundo momento, para verificar a existência dessa propriedade de co-extensividade entre itens lexicais com mais de um componente lexical e a ordem do grupo (nominal), o contexto, identificado pela ferramenta “concordance” do AntConc (ANTHONY, 2019), foi examinado. A fim de investigar qual a função desses itens lexicais em co-extensividade com a ordem do grupo (nominal) na ordem da oração, algumas orações onde esses itens lexicais estavam instanciados foram selecionadas e anotadas experiencialmente

²⁴ Para alguns itens da lista de palavras-chave utilizada para o desenvolvimento e aplicação dos protocolos apresentados nesta pesquisa, buscas por *clusters/n-gramas* com extensão maior do que quatro gramas, e.g. cinco gramas, seis gramas, foram feitas com o propósito de se investigar a ocorrência de itens lexicais com mais de um componente cuja co-extensividade poderia se dar com grupos nominais metaforizados gramaticalmente em decorrência do tipo de texto que compõe o corpus utilizado neste trabalho – artigos acadêmicos da área das ciências da saúde.

(experiential metafunction) conforme o sistema de TRANSITIVIDADE (system of TRANSITIVITY), o qual lida com as funções de Participante, Processo e Circunstância. Essa anotação na ordem da oração torna-se necessária para investigar se esses itens lexicais formados por mais de um componente lexical e em co-extensividade com a ordem do grupo (nominal) funcionam como elementos permanentes, que, de acordo com Figueredo (2007, p. 141), operam como Participantes; numa perspectiva gramatical (grammatical), sendo responsáveis por construir a experiência de permanência no sistema linguístico ou se funcionam como elementos transitórios, que operam como Processo no âmbito do sistema de TRANSITIVIDADE (cf. FIGUEREDO, 2007, p. 141-144).

Faz-se importante esclarecer dois aspectos fundamentais a respeito desta etapa do ProtinLex:PTC: (i) a cada item lexical formado por mais de um componente lexical cuja extração se deu pela busca dos agrupamentos (*clusters/n-grams*), três orações onde esse item lexical estava instanciado eram selecionadas aleatoriamente para serem anotadas de acordo com o sistema de TRANSITIVIDADE. Por exemplo, “condição de saúde” é um dos itens lexicais formados por mais de um componente lexical que foi identificado a partir da busca pelos *clusters* da palavra-chave “saúde” no *corpus*, assim o sistema de TRANSITIVIDADE de três orações distintas onde esse item lexical estava presente foi anotado em uma planilha eletrônica desenvolvida para tal fim, a FIG 5, a seguir, mostra parte da planilha eletrônica utilizada para essa anotação, (ii) a seleção por três orações para cada um dos itens lexicais formados por mais de um componente se deu em razão da anotação do sistema de TRANSITIVIDADE ser feita manualmente, inviabilizando a seleção de uma quantidade maior de orações em função do dispêndio de esforço cognitivo e de tempo gastos na anotação manual.

	A	B	C	D	E	F	
	clause	process_type	process_lexical_item	participant_1_transitive_role	participant_1_lexical_item	participant_2_transitive_role	participan
1	Em sua teoria, Orem afirma que o engajamento das pessoas nas práticas de autocuidado	verbal	afirma	sayer	Orem	verbiage	o engajam
2	A prática de autocuidado é um longo passo a passo,	relational_identifying_intensive	é	token	a prática de autocuidado	vahe	um longo
3	alguns dos aspectos culturais, sociais e comportamentais que envolvem o estilo de vida e a prática de autocuidado de homens e mulheres com DM	relational_attributive_intensive	envolvem	carrier	alguns dos aspectos culturais, sociais e comportamentais	attribute	o estilo de v autocuida mulhe
4	conscientizar as pessoas com DM2 sobre a importância da mudança do estilo de vida nas atividades de autocuidado	material	conscientizar	actor	[alguémcoesão]	goal	as pess
5	Alguns percebem as atividades de autocuidado como prescrições	mental	percebem	senser	alguns	phenomenon	as atividad
6	Na avaliação das atividades de autocuidado, a atividade que obteve maior média corresponde ao uso da terapia medicamentosa	relational_attributive_intensive	corresponde	carrier	a atividade que obteve maior média	attribute	uso da terap
7							

FIGURA 5 – Captura de tela da planilha eletrônica criada para a anotação do sistema de TRANSITIVIDADE

Fonte: Desenvolvida para fins deste trabalho.

A anotação do sistema de TRANSITIVIDADE dessas orações onde os itens lexicais em co-extensão com a ordem do grupo nominal estão instanciados compreende uma parte fundamental do processo de desenvolvimento do ProtinLex:PTC. Isso se dá em razão do pressuposto teórico da linguística sistêmico-funcional acerca do funcionamento hierárquico da escala de ordens gramatical – oração, grupo/frase preposicional, palavra e morfema – em que uma ordem superior da escala de ordens gramatical é composta pelos elementos da ordem inferior, por exemplo, a estrutura da oração – ordem em que o sistema de TRANSITIVIDADE opera – é composta por grupos, no caso, os itens lexicais formados por mais de um componente lexical em co-extensividade com o grupo nominal. Por isso, para investigar a correspondência entre os itens lexicais e os elementos permanentes da oração tal anotação tornou-se essencial.

Uma vez atestado as correspondências entre *clusters/n-gramas*, itens lexicais formados por mais de um componente lexical e a co-extensividade desses itens lexicais com a ordem do grupo nominal, iniciou-se o último passo metodológico do ProtinLex:PTC que compreendeu uma investigação acerca do contexto dessas correspondências, isto é, se esses itens lexicais formados por mais de um componente lexical poderiam ser entendidos como itens lexicais que emergiram dos textos do *corpus* em função das mudanças nas variáveis do contexto de situação – campo, sintonia e modo – que regem os textos do *corpus* compilado para esta pesquisa. Para verificar essa possibilidade, cada componente desses itens lexicais que estavam em co-extensividade à ordem do grupo (nominal) foi explorado por meio de um processo nomeado, neste trabalho, como mudança. Nesse processo de mudança, os componentes lexicais que formavam esses itens lexicais passavam a funcionar como Participantes de uma oração, isolados dos outros componentes lexicais que formavam os itens lexicais co-extensivos à ordem do grupo nominal, e precisavam estabelecer pelo menos uma relação de COESÃO LEXICAL (hiponímia, meronímia, colocação, sinonímia e/ou antonímia) com outro item lexical. Este item lexical precisava ter sido instanciado a partir de configurações diferentes para as variáveis do contexto de situação, sendo, portanto, um item lexical proveniente de um corpus distinto, nesta pesquisa, o *corpus* de suporte (ver 4.2), daquele de onde foi retirado o item lexical formado por mais de um componente lexical que estava em vias de passar pelo processo de mudança. A FIG 6 mostra a planilha eletrônica onde esse isolamento dos componentes lexicais dos itens lexicais co-extensivos à ordem do grupo nominal foi feito (colunas S e U), bem como exemplos de orações onde as relações de COESÃO LEXICAL foram estabelecidas (colunas T e V) para dois itens lexicais utilizados no desenvolvimento e aplicação dos ProtinLexis (ver Quadro 2).

	S	T	U	V
1	itens lexicais_grupo(nominal)	mudança	itens lexicais_grupo(nominal)	mudança
2	ações educativas	ações são educativas, [não são culturais]	prática educativa	prática é educativa, [não é esportiva]
3		ações educativas e ações culturais são tipos de ações		prática educativa e prática esportiva são tipos de prática
4	ações de autocuidado	ações são de autocuidado, [não são de sustentabilidade]	prática do autocuidado	prática é do autocuidado, [não é do crime]
5		ações de autocuidado e ações de sustentabilidade são tipos de ações		prática do autocuidado e prática do crime são tipos de prática
6	ações de promoção	ações são de promoção, [não são de prevenção]	prática de atividade física	prática é de atividade física, [não é de assédio moral]
7		ações de promoção e ações de prevenção são tipos de ações		prática de atividade física e prática de assédio moral são tipos de prática
8	ações de educação	ações são de educação, [não são de saneamento]	práticas de autocuidado em diabetes	práticas são de autocuidado em diabetes, [não são de educação em saúde]
9		ações de educação e ações de saneamento são tipos de ações		práticas de autocuidado em diabetes e práticas de educação em saúde são tipos de práticas

FIGURA 6 – Captura de tela da planilha eletrônica criada para isolar os componentes lexicais e estabelecer as relações COESIVAS LEXICAIS

Fonte: Desenvolvida para fins deste trabalho

A partir do desenvolvimento dessas relações coesivas lexicais construídas em orações resultantes do processo de mudança em questão, o esquema retratado na FIG. 7, a seguir, foi desenvolvido para representar o processo de mudança do ProtinLex:PTC.

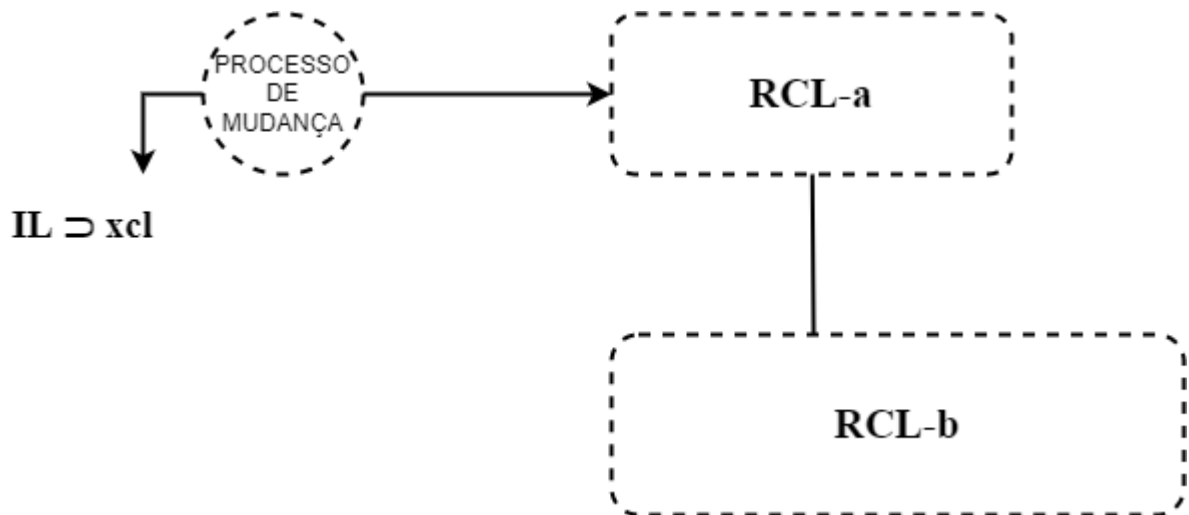


FIGURA 7 – Representação do processo de mudança do ProtinLex:PTC

Fonte: Desenvolvida para fins deste trabalho

A representação detalhada na FIG. 7 insere o item lexical formado por mais de um componente lexical à esquerda da imagem. Este item lexical é indicado pela inscrição $IL \supset xcl$ em que “IL”, escrito em letras maiúscula, significa “item lexical”, em seguida aparece o símbolo matemático “ \supset ” que significa contém, depois tem-se o “x”, em minúsculo, que deve ser modificado de acordo com o número de componentes lexicais que formam o item lexical que está em vias de passar pelo processo de mudança, seguido das letras “cl”, em minúsculo, que significam “componente lexical”. Utilizando o item lexical “profissionais de saúde”, que compreende um dos exemplos de itens lexicais formados por mais de um componente lexical

explorado no desenvolvimento e aplicação deste protocolo de investigação lexical (ver QUADRO 2), a substituição dos elementos dessa inscrição resulta em: profissionais de saúde \supset 2cl. Acima dessa inscrição, aparece o processo nomeado como mudança, o qual é simbolizado por um círculo feito em linhas pontilhadas, designando que tal processo não é linear, podendo ser influenciado pelas mudanças nas variáveis contextuais de uma dada situação. A inscrição IL \supset xCL e o círculo que representa o processo de mudança são ligados por uma seta de pontas dupla a dois retângulos com hastes arredondadas, onde as inscrições RCL-a e RCL-b são inseridas no primeiro e no segundo retângulo respectivamente. Essas inscrições RCL-a e RCL-b dizem respeito à primeira e à segunda relação coesiva lexical estabelecida entre os componentes que formam o item lexical extraído do corpus compilado para esta pesquisa e os componentes que formam o item lexical extraído do corpus de suporte.

Como a metodologia proposta para este ProtinLex salienta que pelo menos uma relação coesiva lexical deve ser estabelecida, a segunda inscrição – RCL-b – localizada no retângulo inferior pode não ser necessária para indicar uma possível influência das variáveis do contexto de situação na formação de itens lexicais com mais de um componente. Nesse caso, apenas o retângulo superior com a primeira relação coesiva lexical (RCL-a) estaria presente na representação. Por outro lado, mais de duas relações coesivas lexicais podem ser necessárias para comprovar essa possível influência, por isso outros retângulos com hastes arredondadas podem ser acrescentados abaixo desse retângulo inferior, aumentando, assim, o número inscrições que seguiriam a ordem alfabética da língua portuguesa, a saber: RCL-c, RCL-d, RCL-e e assim por diante.

Se ao final do processo de mudança do ProtinLex:PTC ficasse comprovado que as variáveis do contexto de situação podem influenciar na emergência de itens lexicais formados por mais de um componente lexical, os semicírculos com a inscrição “contexto de situação” representando essa influência poderiam ser inseridos nas porções inferior/esquerda e superior/direita da representação destacada na FIG. 7, conforme a FIG. 8 demonstra.

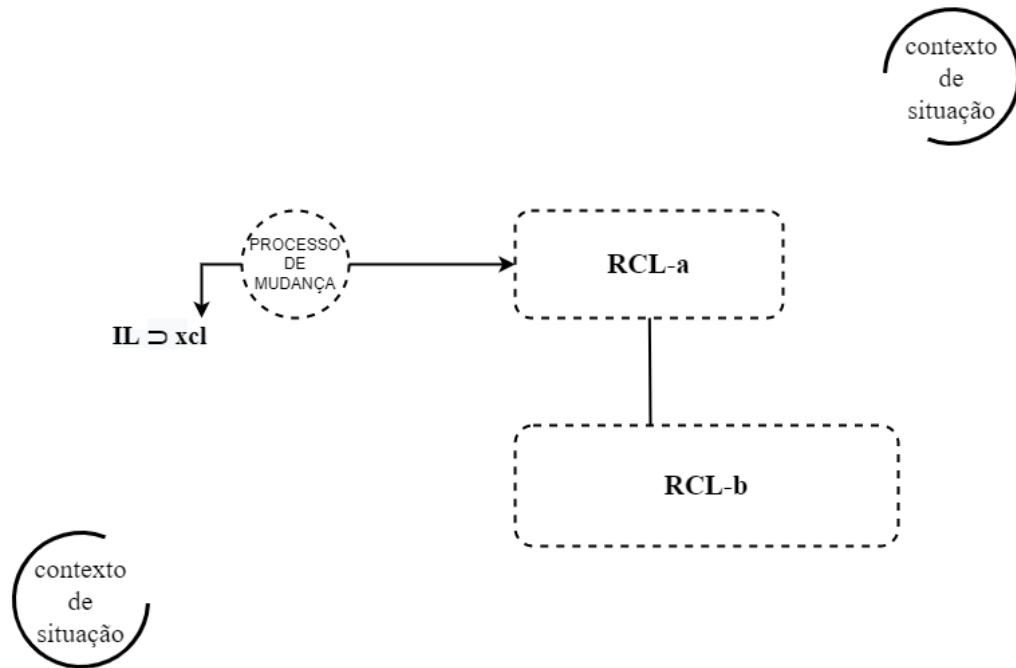


FIGURA 8 – Representação do processo de mudança do ProtinLex:PTC – variáveis do contexto de situação

Fonte: Desenvolvida para fins deste trabalho

A FIG. 8 traz a imagem de como ficaria a representação do processo de mudança do ProtinLex:PTC caso, após o desenvolvimento e aplicação deste protocolo de investigação lexical, ficasse evidente que itens lexicais formados por mais de um componente lexical podem ser influenciados pelas variáveis do contexto de situação que constituem um texto (ou textos de um corpus/corpora) onde esses itens lexicais estão instanciados. Nesse sentido, dois semicírculos com a inscrição “contexto de situação” são inseridos na representação, um na parte inferior e outro na parte superior da imagem. Esses semicírculos indicam que as variáveis do contexto de situação podem influenciar tanto os itens lexicais formados por mais de um componente lexical encontrados no corpus compilado para a presente pesquisa quanto os itens lexicais formados por mais de um componente lexical encontrados no corpus de suporte utilizado neste trabalho, o Sketch Engine. Para apontar uma possível influência das variáveis do contexto de situação no item lexical formado por mais de um componente lexical extraído do corpus compilado para este estudo, um dos semicírculos com a inscrição “contexto de situação” foi introduzido à esquerda inferior da imagem, perto da inscrição $IL \supset xcl$ que representa esse item lexical e seus componentes. Para indicar a possível influência das variáveis do contexto de situação no item lexical formado por mais de um componente lexical retirado do corpus de suporte, o outro semicírculo com a mesma inscrição do anterior foi inserido à direita superior da imagem, perto do retângulo onde estará localizada a primeira relação coesiva

lexical (RCL-a) estabelecida entre os dois itens lexicais formados por mais de um componente lexical, o que foi extraído do corpus compilado para este trabalho e o que foi selecionado a partir do corpus de suporte.

É importante explicar ainda que a escolha por semicírculos para representar a influência das variáveis do contexto de situação nos itens lexicais formados por mais de um componente lexical pode ser justificada pelo fato dessas variáveis contextuais não apresentarem configurações enrijecidas para cada texto (ou conjunto de textos) examinados. Em outras palavras, a forma do semicírculo indica que as variáveis contextuais podem apresentar seleções diferentes para cada situação. Não há, portanto, um significado de total completude na influência dessas variáveis em itens lexicais formados por mais de um componente lexical, pois mesmo que apenas algumas variáveis do contexto de situação sejam modificadas, itens lexicais desse tipo podem emergir no(s) texto(s). Por fim, a abertura do semicírculo direcionada para o processo de mudança do ProtinLex:PTC indica que essa influência das variáveis do contexto de situação nos itens lexicais formados por mais de um componente lexical pode ser percebida ao longo desse processo.

4.3.2 Do Protocolo de Investigação Lexical: Processo de Transformação da Gramática – ProtinLex:PTG

O ProtinLex:PTG, sigla para Protocolo de Investigação Lexical: Processo de Transformação da Gramática, compreende o segundo protocolo de investigação lexical apresentado neste trabalho. O desenvolvimento do ProtinLex:PTG surge a partir de uma análise dos resultados obtidos para o ProtinLex:PTC, na qual verificou-se que o item lexical no português brasileiro parecia ter uma capacidade de se transformar em itens lexicais correlatos.

Nesse sentido, a aplicação do ProtinLex:PTG é organizada em duas fases. A primeira diz respeito à análise dos resultados encontrados na aplicação do ProtinLex:PTC, uma vez que não é possível depreender o contexto sem a realização linguística pelas funções estruturais da gramática. A partir disso, apenas aqueles itens que foram caracterizados como lexicais por meio da aplicação do ProtinLex:PTC estavam aptos a serem utilizados no ProtinLex:PTG.

Considerando esses itens caracterizados como lexicais pelo ProtinLex:PTC, a segunda fase do ProtinLex:PTG é responsável por desenvolver um outro processo de mudança, cujo enfoque está nos itens lexicais formados por mais de um componente lexical abordados no protocolo de investigação lexical anterior. É importante esclarecer que este processo de mudança apresenta uma abordagem diferente daquele desenvolvido para o ProtinLex:PTC. O

processo de mudança em questão investiga a capacidade de transformação gramatical dos itens lexicais formados por mais de um componente lexical, ou seja, a possibilidade que um determinado item lexical no português brasileiro parece ter de ser desmembrado em porções – porção lexical e porção gramatical – bem como de gerar outro(s) item(ns) lexical(is) que pode(m) estar no escopo de uma classe de palavras diferente daquela do item lexical que o originou (ver QUADRO 2). Enquanto que o processo de mudança descrito para o ProtinLex:PTC examinou a capacidade de um item lexical no português brasileiro coocorrer com outro(s) item(ns) lexical(is) em razão da influência das variáveis do contexto de situação na construção de um texto (ou textos) e, por meio disso, conceber um item lexical com um significado único co-extensivo à ordem do grupo (nominal).

Para o processo de mudança do ProtinLex:PTG, uma planilha eletrônica foi elaborada. Na primeira coluna desta planilha, uma oração, retirada de um dos textos que compõe o *corpus* compilado para a presente pesquisa, onde o item lexical a ser transformado estava instanciado, era inserida. Em seguida, na segunda coluna, esse item lexical, que já teve sua co-extensividade com a ordem do grupo nominal comprovada pela aplicação do ProtinLex:PTC, era extraído do restante da oração onde estava instanciado. Por fim, a terceira coluna retrata o processo de mudança em si. Neste processo, esse item lexical que estava em co-extensividade com a ordem do grupo (nominal) e funcionava como um elemento permanente na oração, *i.e.*, Participante ou Circunstância (ver 5.2), tinha seus componentes lexicais isolados, os quais passavam a operar como elementos permanentes, *i.e.*, Participante ou Circunstância (ver 5.2), em orações relacionais e/ou materiais. A FIG. 9 traz a captura de tela que mostra a planilha eletrônica criada para a aplicação do protocolo de investigação lexical em questão.

	A	B	C	D	E	F
	Contexto	Item lexical com mais de um elemento	Oração	Tipo de oração	Item lexical que realiza o processo	Participante 1
1	Toda a <u>prática educativa</u> realizada com os usuários do grupo intervenção foi pautada na abordagem do empoderamento baseada no Protocolo Mudança de Comportamento.	prática educativa	prática educativa é um tipo de prática, [em oposição à prática pedagógica]	relacional intensivo	é	prática educativa
2			essa prática envolve educação	relacional intensivo	envolve	essa prática
3	As principais barreiras e metas relacionadas à <u>prática do autocuidado</u> estão relacionadas aos comportamentos psicossociais e culturais.	prática do autocuidado	prática é do autocuidado, [não do exercício físico]	relacional circunstancial	é	prática
4			essa prática envolve o autocuidado		envolve	essa prática
5	A <u>prática de autocuidado</u> é um longo passo a passo, o qual necessita ser avaliado constantemente e regado com doses diárias de motivação.	prática de autocuidado	prática é de autocuidado, [não de exercício físico]	relacional circunstancial	é	prática
6			essa prática envolve autocuidado	material	envolve	essa prática
7						

FIGURA 9 – Captura de tela da planilha eletrônica criada para o desenvolvimento do ProtinLex:PTG

Fonte: Desenvolvida para fins deste trabalho.

Como resultado deste processo de mudança um dos componentes desse item lexical, que estava em co-extensividade com a ordem do grupo nominal, tinha a sua porção

gramatical modificada e gerava outro(s) item(ns) lexical(is) (ver FIG. 9, coluna C), os quais poderiam estar no escopo de uma classe de palavras diferente daquela que o originou. A FIG. 10, a seguir, foi elaborada para representar o processo de mudança correspondente ao ProtinLex:PTG.

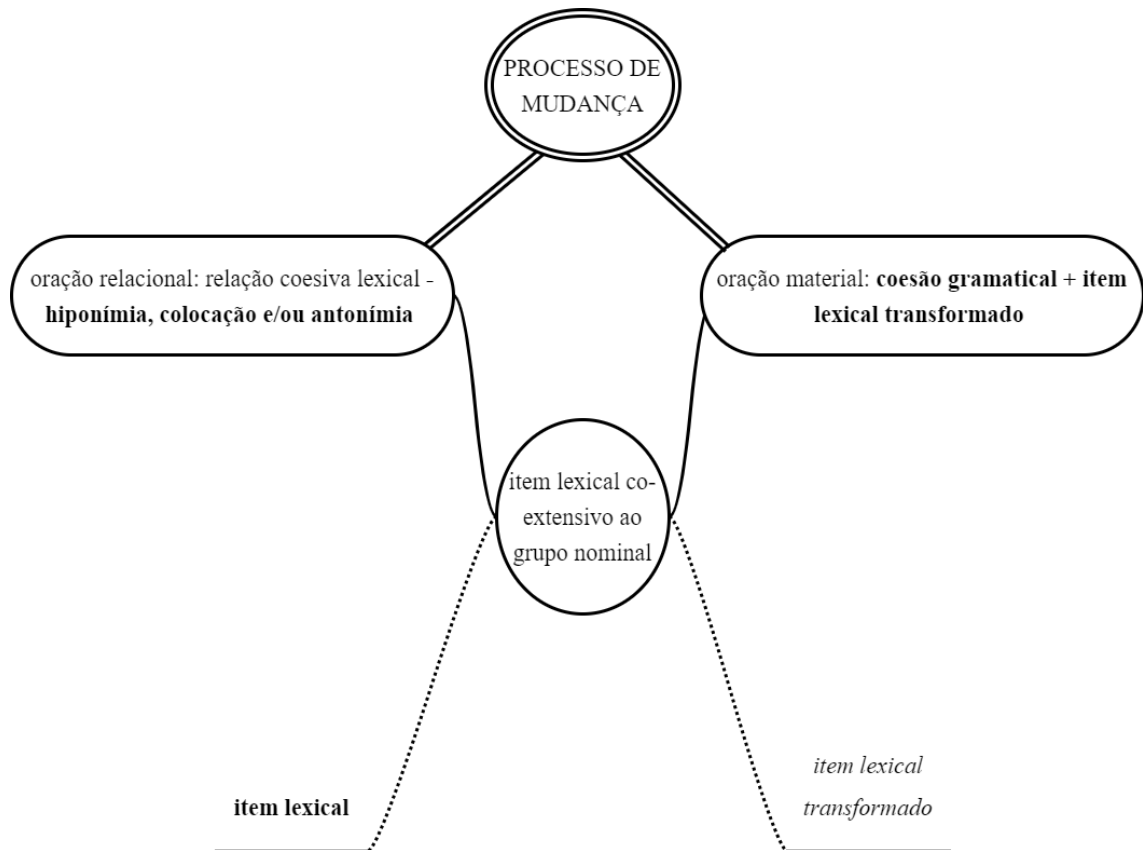


FIGURA 10 – Representação do processo de mudança do ProtinLex:PTG

Fonte: Desenvolvida para fins deste trabalho

O esquema detalhado na FIG. 10 diz respeito ao processo de mudança criado para o ProtinLex:PTG. A representação desse processo de mudança é composta de 3 partes, uma parte superior, outra central e uma inferior.

A parte central delimita-se à apresentação do item lexical com mais de um componente lexical, cuja co-extensividade com a ordem do grupo nominal já precisa ter sido identificada pelo ProtinLex:PTC. É relevante mencionar que o item lexical transformado surge a partir de um dos componentes lexicais que formam o item lexical co-extensivo à ordem do grupo nominal, sendo que este item lexical deve estar inserido dentro do círculo localizado no meio da representação (ver FIG. 10).

Esse círculo conta com ligações tanto na parte superior quanto na parte inferior da representação (ver FIG. 10), indicando que o resultado obtido pela aplicação do ProtinLex:PTC

compreende uma parte inseparável do protocolo de investigação lexical em questão. A ligação com a parte superior é feita por duas linhas contínuas, uma que sai da lateral esquerda do círculo, onde está o item lexical co-extensivo à ordem do grupo nominal, e uma outra que sai da lateral direita desse mesmo círculo. A linha da esquerda encontra o retângulo com arestas arredondadas que está localizado na porção esquerda superior da FIG. 10. É nesse retângulo em que a primeira oração desse processo de mudança é revelada, tal oração é responsável por indicar a(s) relação(ões) de hiponímia, colocação e/ou antonímia estabelecida(s) entre o item lexical posicionado ao centro da representação e um outro item lexical extraído do corpus de suporte (ver 4.2). A linha da direita segue em direção ao outro retângulo com arestas arredondadas inserido na porção direita superior da FIG. 10. É nesse retângulo que a segunda oração do processo de mudança é introduzida. Essa oração compreende uma oração material com dois Participantes. O Participante 1 engloba o Dêitico, que é realizado por uma palavra da classe dos pronomes demonstrativos, e ainda um dos componentes lexicais que aparece na oração relacional, no retângulo da esquerda, como o Participante 2. Já o Participante 2 desta oração material diz respeito ao item lexical transformado, o qual é originado pelo componente lexical, cuja correferência não foi realizada nesta oração material. Em outras palavras, um dos componentes lexicais do item lexical localizado no centro da representação é o responsável por estabelecer a correferencialidade entre itens lexicais da primeira oração com itens lexicais da segunda oração do processo de mudança do ProtinLex:PTG, enquanto que o outro componente lexical é o responsável por gerar o item lexical transformado.

Ainda no que concerne à parte superior da representação exposta na FIG. 10, cada um dos retângulos com arestas arredondadas, dispostos à esquerda e à direita da representação do processo de mudança do ProtinLex:PTG, é ligado por uma linha dupla a uma elipse também desenhada com as mesmas linhas duplas. Essa elipse é ocupada pelo processo de mudança em si e apresenta forma arredondada semelhante aos retângulos para simbolizar o fato de que o processo de mudança deste protocolo de investigação lexical é constituído de, no mínimo, duas orações, dispostas nos retângulos, as quais desempenham a função de apresentar as relações coesivas lexicais e apontar, quando possível, a existência de um item lexical transformado, o qual é proveniente de um dos componentes do item lexical co-extensivo à ordem do grupo nominal posicionado ao centro da representação da FIG. 10.

A parte inferior do esquema que representa o processo de mudança do ProtinLex:PTG trata do componente lexical, que originou o item lexical transformado, bem como o item lexical transformado de fato. O círculo que traz o item lexical formado por mais de um componente lexical está ligado a esses itens lexicais por duas linhas pontilhadas. Uma

que sai da lateral esquerda do círculo e chega ao componente lexical (item lexical), disposto à esquerda inferior da FIG. 10, e outra que sai da lateral direita desse mesmo círculo e encontra o item lexical transformado fixado à direita inferior da FIG. 10. Dessa forma, o componente lexical (item lexical), que gera o item lexical transformado, e o próprio item lexical transformado ficam restritos à parte inferior da imagem criada para representar o processo de mudança deste protocolo de investigação lexical.

É interessante explicar ainda que, em um primeiro momento, a escolha por linhas pontilhadas reflete a necessidade de deixar evidente que o componente lexical (item lexical), que gera o item lexical transformado, e o próprio item lexical transformado tendem a não ser sempre os mesmos itens, uma vez que o item lexical ao qual esses itens estão ligados parece ser passível de ser influenciado pelas mudanças das variáveis do contexto de situação (ver 5.1). Em um segundo momento, estas linhas pontilhadas demonstram que o resultado deste processo de mudança, que lida principalmente com a identificação de um item lexical transformado, pode revelar-se como um resultado com desvios, sinalizando que resultados rotulados como default (padrão) são esperados, mas resultados que desviam desse padrão também podem ser obtidos (ver 5.2.1; 5.2.2 e 5.2.3).

No que diz respeito à formatação do componente lexical (item lexical), que gera o item lexical transformado, e do item lexical transformado, o primeiro deve aparecer em negrito; e o segundo em itálico. Em comum, ambos devem ter suas porções lexicais sublinhadas, como forma de distinguir das porções gramaticais, as quais devem seguir a formatação correspondente de cada item, seja em negrito seja em itálico.

Para finalizar a descrição dos passos metodológicos determinados para o protocolo de investigação lexical em questão, faz-se relevante mencionar dois aspectos que podem interferir no processo de mudança do ProtinLex:PTG. O primeiro deles diz respeito ao número de orações necessárias para que processo de mudança ocorra. Esse número pode variar conforme o item lexical em foco, isto é, há itens lexicais que podem precisar de apenas uma oração para serem transformados e itens lexicais que podem precisar de mais de duas orações para serem transformados por completo. O segundo aspecto está relacionado ao fato de que a propriedade de transformação identificada como uma possível característica para alguns itens lexicais no português brasileiro não está relacionada à etimologia e/ou à propriedade de derivação da gramática normativa e/ou da gramática funcionalista (cf. CHOMSKY, 1965; BASÍLIO, 1991; NEVES, 1997; BORBA, 2003; CASTILHO, 2010). Em outras palavras, a transformação, nesta pesquisa, sobretudo, no que diz respeito ao ProtinLex: PTG tem como objetivo apontar que alguns itens lexicais parecem apresentar a capacidade de

desmembramento em porções – porção lexical e porção gramatical – bem como a capacidade de gerar outros itens lexicais que podem estar no escopo de classes de palavras²⁵ distintas, o que não acontece com os itens gramaticais (funcionais) no PB.

O próximo capítulo explora os resultados obtidos a partir do desenvolvimento e da aplicação dos ProtinLexis e discute o impacto desses resultados para a caracterização do item lexical no PB.

²⁵ No PB, itens considerados funcionais (ou gramaticais) também podem gerar outros itens funcionais. No entanto, esses itens funcionais que foram gerados permanecerão na mesma classe de palavras daquele item funcional que o gerou. Por exemplo, o artigo definido “o” pode gerar o artigo definido “os”, no plural, ambos estão localizados na classe de palavras dos pronomes (pronouns), enquanto que o item lexical “educação” encontra-se na classe de palavras dos substantivos comuns (nominal^noun^common) e pode gerar outros itens lexicais como “educativas” e/ou “educacionais” que estão no escopo da classe de palavras dos adjetivos (nominal^adjective) (ver 5.2.1)

5. RESULTADOS E DISCUSSÃO

Como detalhado na seção anterior, os itens com frequência mais elevada, localizados no início da lista de palavras-chave (um pouco mais de 10% da lista de palavras-chave) foram selecionados para aplicação dos protocolos de investigação lexical. O QUADRO 2, a seguir, mostra quais itens são esses.

QUADRO 2 – Lista de itens selecionados para a aplicação dos protocolos de investigação lexical (ProLinLex) – corpus de artigos acadêmicos

Índice	Número de ocorrências	Itens selecionados a partir da lista de palavras-chave: <i>Corpus</i> de artigos acadêmicos
1	1023	autocuidado
2	952	saúde
3	630	doença
4	362	mellitus
5	329	relação
6	327	usuário
7	288	educação
8	256	complicações
9	225	indivíduos
10	197	pés
11	194	ações
12	194	prática

Fonte: Desenvolvido para fins deste trabalho.

O QUADRO 2 destaca os doze primeiros itens que aparecem na lista de palavras-chave gerada pelo *software* concordanciador a partir da inserção do *corpus* de artigos acadêmicos sobre autocuidado em diabetes mellitus compilado para o presente trabalho. Tendo a parte escrita monológica do CALIBRA sido utilizada como *corpus* de referência. O p-valor estipulado como 0.0001 para os testes %DIFF, Odds-ratio, Hardie's Log-Ratio e Ratio of Relative frequencies, os quais foram unânimes ao apresentar os itens do QUADRO 2 como palavras-chave para o corpus utilizado nesta pesquisa.

Com exceção do item 4, *mellitus*, todos os outros foram utilizados para o desenvolvimento e a aplicação dos protocolos de investigação lexical deste estudo. É importante explicar que apesar do item “mellitus” estar presente na lista de palavras-chave como o quarto item com maior número de ocorrências, enquadrando-se nos parâmetros metodológicos desta pesquisa, a utilização desse item para o desenvolvimento e a aplicação dos protocolos de investigação lexical retratados neste trabalho torna-se inviável, uma vez que, no *corpus* utilizado, tal item sempre coocorre com o item “diabetes”. Em outras palavras, a investigação dos itens detalhados no QUADRO 2 mostra que “mellitus” não é instanciado nos artigos acadêmicos que compõem o *corpus* desta pesquisa como um item com capacidade de existência única, ou seja, capaz de formar *itens lexicais transformados e/ou itens lexicais co-extensivos à ordem do grupo (nominal)* por sempre estar em coocorrência com o item “diabetes”. Em função da exclusão do item “mellitus” das aplicações dos protocolos de investigação lexical desenvolvidos para esta pesquisa, os itens “ações” e “prática”, que apresentam a mesma frequência (cada um com 194 ocorrências), foram selecionados para o desenvolvimento e aplicação dos ProtinLexis.

5.1 Da aplicação do ProtinLex:PTC – *corpus* de artigos acadêmicos

Como mencionado na seção de Metodologia deste trabalho, o ProtinLex:PTC conta com dois objetivos principais. O primeiro deles diz respeito à investigação de itens lexicais formados por mais de um componente lexical. Em função disso, a ferramenta de “cluster/n-gram” do software concordanciador AntConc foi utilizada como forma de pesquisar por itens lexicais que apareciam logo antes ou logo depois dos itens de busca, ou seja, aqueles destacados no QUADRO 2, utilizados para o desenvolvimento e aplicação dos protocolos de investigação lexical deste estudo. Foram feitas buscas por bigramas, trigramas e quadrigramas para cada um dos doze itens. O QUADRO 3, adiante, mostra os agrupamentos com maior número de ocorrências²⁶ para cada item da lista de palavras-chave.

²⁶ Somente agrupamentos com no mínimo de 10 ocorrências (ou próximo a esse número) foram inseridos no referido quadro. A delimitação desse número é justificada por duas razões principais. A primeira pauta-se no fato de que agrupamentos com poucas ocorrências podem ser questionados quanto à existência de um padrão, ou seja, agrupamentos que aparecerem poucas vezes no corpus podem não ser entendidos como itens lexicais formados por mais de um componente lexical, mas, sim, conjuntos de palavras que coocorrem em sequência nos textos que compõem o corpus utilizado no trabalho, não instanciando um novo significado. A segunda razão diz respeito ao dispêndio de tempo gasto na anotação do sistema de TRANSITIVIDADE visto que as orações onde os agrupamentos selecionados estavam realizados foram anotadas manualmente a fim de se investigar a relação agrupamentos e itens lexicais formados por mais de um componente lexical cuja co-extensividade poderia se dar com a ordem do grupo (nominal), o qual tende a funcionar como elemento permanente, *i.e.* Participantes e/ou Circunstâncias, em orações (cf. FIGUEREDO, 2007).

QUADRO 3 – Agrupamentos formados à direita e à esquerda de cada item da lista de palavras-chave

Itens selecionados a partir da lista de palavras-chave	Bigramas/ (Nº de ocorrências)	Trigramas/ (Nº de ocorrências)	Quadrigramas/ (Nº de ocorrências)
autocuidado	-	práticas de autocuidado (96)	autocuidado com os pés (35)
		atividades de autocuidado (92)	educação para o autocuidado (17)
		ações de autocuidado (62)	motivação para o autocuidado (10)
		autocuidado em diabetes (60)	
		capacidade de autocuidado (29)	
		promoção do autocuidado (23)	
		prática do autocuidado (20)	
		avaliação do autocuidado (11)	
		realização do autocuidado (11)	
		gestão do autocuidado (10)	
saúde	saúde pública (19)	profissionais de saúde (83)	unidade básica de saúde (25)
		educação em saúde (56)	
		letramento em saúde (34)	educação para a saúde (18)
		promoção	

		da saúde (28)	
		serviços de saúde (25)	atenção primária à saúde (11)
		equipe de saúde (21)	
		estado de saúde (21)	letramento funcional em saúde (11)
		unidades de saúde (18)	
		condição de saúde (14)	equipe multiprofissional de saúde (10)
		ministério da saúde (13)	
doença	doenças crônicas (37)	controle da doença (50)	-
	doença renal (9)	tempo de doença (19)	
		enfrentamento da doença (17)	
	doenças cardíacas (8)	manejo da doença (16)	
		tratamento da doença (9)	
relação	-	-	em relação à doença (18)
			em relação ao conhecimento (17)
			em relação às variáveis (9)
			em relação ao diabetes (8)
usuário	-	usuários com DM2 (16)	usuários com diabetes mellitus (58)

			usuários do grupo controle (8)
			usuários do grupo intervenção (7)
educação	educação alimentar (5)	educação em saúde (56)	educação para a saúde (18)
		educação em grupo (28)	
		educação em diabetes (20)	
		educação em DM (17)	
	educação individual (5)	ações de educação (10)	educação para o autocuidado (17)
		efeito da educação (8)	
complicações	complicações crônicas (40)	complicações do diabetes (18)	complicações agudas e crônicas (13)
		prevenção de complicações (18)	
	complicações microvasculares (14)	presença de complicações (14)	
indivíduos	-	indivíduos com DM (34)	-
		indivíduos com diabetes (19)	
pés	pé diabético (39)	-	autocuidado com os pés (35)
			cuidados com os pés (32)

ações	ações educativas (27)	ações de autocuidado (62)	-
		ações de promoção (12)	
		ações de educação (10)	
prática	prática educativa (26)	prática do autocuidado (20)	prática de atividade física (27)
		prática de autocuidado (14)	prática de exercícios físicos (13)

Fonte: Desenvolvido para fins deste trabalho.

Antes de explorar as informações do QUADRO 3, é importante esclarecer que os agrupamentos identificados com maior número de ocorrências para o item “usuário” estão no plural (usuários), embora na lista de palavras-chave tal item tenha aparecido no singular. O mesmo acontece para um dos agrupamentos formados com o item “doença” em “doenças cardíacas” e para o item “pés” em “pé diabético”. Mesmo diante dessa diferença no que tange a flexão de número desses itens – usuário e doença – os agrupamentos formados pelo plural de tais itens são vistos como válidos e utilizados no desenvolvimento e na aplicação do protocolo de investigação lexical em questão. A manutenção desses agrupamentos encontra justificativa no fato de que esse tipo de flexão no PB não viabiliza que esses itens mudem de classe de palavras, ou seja, “usuário” ou “usuários” e “doença” ou “doenças” estão no escopo da classe de palavra nominal, que prototipicamente realizam os elementos permanentes na oração, *i.e.* Participantes e/ou Circunstâncias, não prejudicando, assim, a concepção do ProtinLex:PTC.

Os dados do QUADRO 3 mostram que o número de agrupamentos parece ser proporcional ao número de ocorrências do item em questão no corpus. Em outras palavras, quanto maior o número de ocorrências de um dado item, maior será a quantidade de bigramas, trigramas e/ou quadrigramas que esse item será capaz de formar no corpus. Por exemplo, os itens “autocuidado” e “saúde” estão localizados no topo da lista de palavras-chave gerada para o corpus utilizado nesta pesquisa, conseqüentemente o número de *n-gramas* (agrupamentos) formados por esses itens é maior do que aqueles formados para os itens “ações” e “prática”, que tem menor número de ocorrências e estão localizados ao final da lista de palavras-chave selecionada para o desenvolvimento e aplicação dos protocolos de investigação lexical (ver QUADRO 2).

Em relação à presença de bigramas, trigramas e quadrigramas, dos onze itens selecionados, quatro deles – saúde, educação, complicações e prática – apresentam exemplos para essas três configurações de n-gramas. Dentre essas três configurações, a que tem maior número de ocorrências é os trigramas, com exceção dos itens “usuários”, “relação” e “pés”, cuja configuração mais frequente é os quadrigramas. Esse resultado pode indicar uma particularidade do corpus desta pesquisa uma vez que os textos que o constitui são artigos acadêmicos, um tipo de texto localizado no âmbito do processo sociosemiótico explorar (exploring/ campo (field)), cuja linguagem predominante é especializada. No que diz respeito à variável sintonia (tenor), autores e leitores desse tipo de texto são vistos como pares em relação ao nível de conhecimento, portanto, para apresentar novos significados e estabelecer, na área de conhecimento, um novo conceito, itens lexicais formados por mais de um componente lexical costumam ser utilizados, os quais passam a ser referenciados em diversos momentos ao longo dos artigos e, conseqüentemente, podem vir a ser utilizados por outros pesquisadores da área nos próprios trabalhos.

Para além dos agrupamentos apresentados no QUADRO 3, investigou-se ainda se os itens da lista de palavras-chave também formavam agrupamentos a partir de seus itens lexicais derivados. Neste estudo, itens lexicais derivados estão relacionados às classes de palavra, não havendo relação com etimologia. O item presente na lista de palavras-chave deve estar no escopo de uma classe de palavra diferente da classe do seu item derivado. Por exemplo, o item “saúde”, localizado na segunda posição da lista de palavras-chave (ver QUADRO 2), está no escopo da classe de palavra dos substantivos (nouns) > substantivos comuns (common) e deriva o item “saudável”, que está no âmbito da classe de palavra dos substantivos (nouns) > adjetivo (adjective). O resultado dessa investigação apontou que os itens “saúde”, “usuário”, “educação” e “indivíduos”, presentes na lista de palavras-chave e utilizados no desenvolvimento e aplicação dos protocolos de investigação lexical elaborados neste trabalho, apresentaram itens derivados que formaram agrupamentos. O QUADRO 4 mostra quais são os itens derivados, bem como os agrupamentos encontrados e os respectivos números de ocorrências.

QUADRO 4 – Agrupamentos formados a partir de itens derivados da lista de palavras-chave

Itens selecionados a partir da lista de palavras-chave	Item derivado	Bigramas/ (Nº de ocorrências)	Trigramas/ (Nº de ocorrências)	Quadrigramas/ (Nº de ocorrências)
saúde	saudável/ saudáveis	alimentação saudável (33)	-	estilo de vida saudável (12)
		hábitos saudáveis (14)		
usuário	uso	-	uso de insulina (33)	-
			uso de medicamentos (18)	
educação	educativo/as	programa educativo (49)	-	-
		processo educativo (37)		
		práticas educativas (28)		
		ações educativas (27)		
		intervenções educativas (25)		
	educacionais	estratégias educacionais (7)		
indivíduos	individual	intervenção individual (17)	-	-

Fonte: Desenvolvido para fins deste trabalho

A primeira coluna do QUADRO 4 traz os itens da lista de palavras-chave que apresentaram no corpus utilizado nesta pesquisa itens derivados formando agrupamentos, são eles: saúde, usuário, educação e indivíduos. A segunda coluna mostra quais são os itens

derivados que formaram agrupamentos, a saber: “saudável(is)” para “saúde”; “educativo/as” e “educacionais” para “educação”; “uso” para “usuário” e “individual” para “indivíduos” Com exceção do item “educação” que conta com duas derivações diferentes e com o maior número de variedade (seis agrupamentos diferentes) e quantidade (173 somando todas as ocorrências dos agrupamentos formados a partir dos dois itens derivados) de agrupamentos obtidos a partir dos itens derivados, os três itens restantes – saúde, usuário e indivíduos – apresentam um item derivado para cada. As três colunas restantes do QUADRO 4 retratam as configurações encontradas para esses agrupamentos, *i.e.* bigrama, trigrama e quadrigrama, sendo os bigramas a configuração mais frequente para todos os itens derivados a exceção do item “uso” que conta com somente duas ocorrências de trigramas.

Essas observações exploradas mediante a análise dos dados retratados no QUADRO 4 indicam que no português brasileiro itens lexicais parecem ser capazes de derivar outros itens lexicais cuja classe de palavra poderá ser diferente daquela do item lexical que o derivou. Uma vez que os itens lexicais derivados tem seus significados estabelecidos nos textos onde estão instanciados (ver 1.0), esses itens lexicais derivados podem coocorrer com outros itens lexicais, constituindo itens lexicais formados por mais de um componente lexical tal como ocorre com os itens selecionados para o desenvolvimento e aplicação dos protocolos de investigação lexical presentes na lista de palavras-chave (ver QUADRO 2 e QUADRO 3). Nesse sentido, esses itens lexicais derivados formados por mais de um componente lexical também podem estar em co-extensividade com a ordem do grupo nominal, funcionando como elementos permanentes na oração, *i.e.* Participantes e/ou Circunstâncias, por isso uma parcela das orações onde itens lexicais desse tipo estão realizados foram selecionadas para a anotação do sistema de TRANSITIVIDADE. A contabilização dos resultados dessa anotação está detalhada na TABELA 1.

TABELA 1 – Anotação do sistema de TRANSITIVIDADE: contabilização das funções

Itens selecionados a partir da lista de palavras-chave (e itens derivados)		P1	P2	P3 ou PA	Circunstância	Número de orações anotadas
autocuidado		16	16	6	7	45
saúde	saudável (eis)	25	24	5	9	63
doença		7	13	5	3	28
relação		2	2	0	7	11
usuário	uso	11	5	1	1	18

educa ção	educati vas	educacio nal	35	20	2	7	64
complicações			11	5	0	5	21
indivíduos	individual		5	2	0	2	9
pés			5	4	0	3	12
ações			5	4	0	3	12
prática			5	3	0	4	12
TOTAL			127	98	19	51	295

Legenda: P1 = Primeiro Participante, P2 = Segundo Participante, P3 = Terceiro Participante e PA = Participante Adicional.

Fonte: Desenvolvida para fins deste trabalho.

A contabilização das funções do sistema de TRANSITIVIDADE detalhadas na TABELA 1 aponta que 295 orações foram anotadas no total, havendo uma disparidade no número de orações anotadas para cada item extraído da lista de palavras-chave em razão da diferença de itens lexicais formados por mais de um componente lexical para cada um dos itens da lista de palavras-chave (ver QUADRO 3 e QUADRO 4). Em outras palavras, como há itens na lista de palavras-chave que apresentam maior número de coocorrência com outros itens lexicais, constituindo itens lexicais formados por mais de um componente lexical, o número de orações selecionadas para a anotação do sistema de TRANSITIVIDADE onde esses itens estão instanciados é maior. Por isso, foram anotadas 64 orações para o item “educação”, o qual tem, no corpus utilizado nesta pesquisa, maior variação de coocorrências com outros itens lexicais, e nove para “indivíduos” e “individual”, que conta com apenas três coocorrências diferentes – “indivíduos com diabetes”, “indivíduos com DM” e “intervenção individual”. Esse resultado indica que no português brasileiro parece haver itens lexicais com maior capacidade de relacionar-se com outros do mesmo tipo, formando itens lexicais com mais de um componente lexical e instanciando um novo significado no texto em que aparecem.

Outro aspecto que reforça essa propriedade presente em alguns itens lexicais diz respeito ao fato de que essa variedade de itens lexicais formados por mais de um componente lexical a partir de um item em específico parece não estar associada à frequência com esse dado item aparece no corpus. Por exemplo, “autocuidado” com 1.023 ocorrências e “saúde” com 952 ocupam respectivamente a primeira e a segunda posição na lista de palavras-chave (ver QUADRO 2) e possuem menor variedade de itens lexicais formados por mais de um componente lexical que “educação” que tem 288 ocorrências e está localizado na sétima posição dessa lista. Essas observações apontam que a capacidade de um item lexical instanciar

uma maior variedade de outro(s) item(ns) formados por mais de um componente lexical parece estar relacionada às classes de palavra em que esses itens estão localizados. Em outras palavras, quando um item da lista de palavras-chave deriva outro(s) item(ns), isto é, item(ns) que está(ão) no escopo de uma classe de palavras diferente daquela do item que o originou; e esse item derivado também consegue instanciar novos significados por meio da coocorrência com outros itens lexicais, soma-se a variedade de itens lexicais formados por mais de um componente lexical a partir daquele item presente na lista de palavras-chave com a variedade de itens lexicais formados por mais de um componente a partir do item lexical derivado. Considerando uma abordagem sistêmico-funcional, esse resultado sugere que o conceito de colocados (collocation) pode estar relacionado às mudanças na classe de palavras, não à extensão do *corpus* (maior número de textos) e/ou maior número de ocorrências de um item lexical em particular. Itens lexicais que se encontram no escopo de uma determinada classe de palavras e derivam outros itens lexicais que estarão localizados no âmbito de outra classe de palavras apresentam maior possibilidade de instanciar itens lexicais formados por mais de um componente lexical do que itens lexicais que contam com um número elevado de ocorrências, mas não apresentam uma quantidade mínima (ou nenhuma) de itens derivados no *corpus*.

No que tange a co-extensividade entre itens lexicais formados por mais de um componente lexical e a ordem do grupo nominal, o qual prototipicamente realiza um elemento permanente, funcionando como Participantes e/ou Circunstâncias no sistema de TRANSITIVIDADE, os dados detalhados na TABELA 1 revelam que em todas as orações anotadas os itens lexicais formados por mais de um componente lexical podem ser entendidos como co-extensivos à ordem do grupo nominal, pois foram co-extensivos às funções dos Participantes – Participante 1, Participante 2, Participante 3 e Participante Adicional – ou à função da Circunstância. Especificamente, esses itens lexicais formados por mais de um componente lexical são co-extensivos à função do Participante sobretudo o Participante 1, que foi selecionado em 127 (43%) orações como função co-extensiva ao item lexical formado por mais de um componente lexical em questão. Em seguida, aparecem as funções do Participante 2 (33%) e da Circunstância (17,3%).

Esses resultados revelam que no português brasileiro itens lexicais formados por mais de um componente lexical podem ser co-extensivos à ordem do grupo nominal, propriedade também identificada no trabalho de Halliday (1961, p. 59-60) para a língua inglesa. Para além da co-extensividade entre item lexical formado por mais de um componente lexical e o eixo sintagmático (dimensão local), cuja organização se dá pelo princípio da escala de ordens (rank scale) – oração, grupo/frase, palavra e morfema – os resultados da TABELA 1

apontam para o fato de que esses itens lexicais formados por mais de um componente lexical também podem ser co-extensivos ao eixo paradigmático (dimensão local), dado que esses itens lexicais podem ser co-extensivos às funções dos Participantes e Circunstância, que estão no escopo do sistema de TRANSITIVIDADE. Nesse sentido, essas relações de co-extensividade entre item lexical formado por mais de um componente lexical e as dimensões locais da linguagem – eixo sintagmático e eixo paradigmático – reforçam a hipótese de que o léxico pode ser caracterizado como uma dimensão da linguagem.

Como o último objetivo do ProtinLex:PTC foi verificar a influência das variáveis do contexto de situação na emergência dos itens lexicais formados por mais de um item lexical no corpus, esses itens lexicais foram examinados por meio de um processo de mudança em que os componentes lexicais que os formavam passavam a funcionar como Participantes de uma oração. Em outros termos, esses componentes lexicais que em conjunto com os itens lexicais selecionados a partir da lista de palavras-chave ou com os itens lexicais derivados destes últimos eram desmembrados e operavam de maneira isolada em orações relacionais e/ou materiais estabelecidas com base em pelo menos uma relação de COESÃO LEXICAL (hiponímia, meronímia, colocação, sinonímia e/ou antonímia) com outro item lexical, proveniente de outro corpus (ver FIG. 7). A FIG. 11, adiante, traz a representação de como esse processo de mudança ocorreu para os itens lexicais formados por mais de um componente lexical “profissionais de saúde”, “complicações crônicas” e “programa educativo”, este último um item lexical derivado de “educação”.

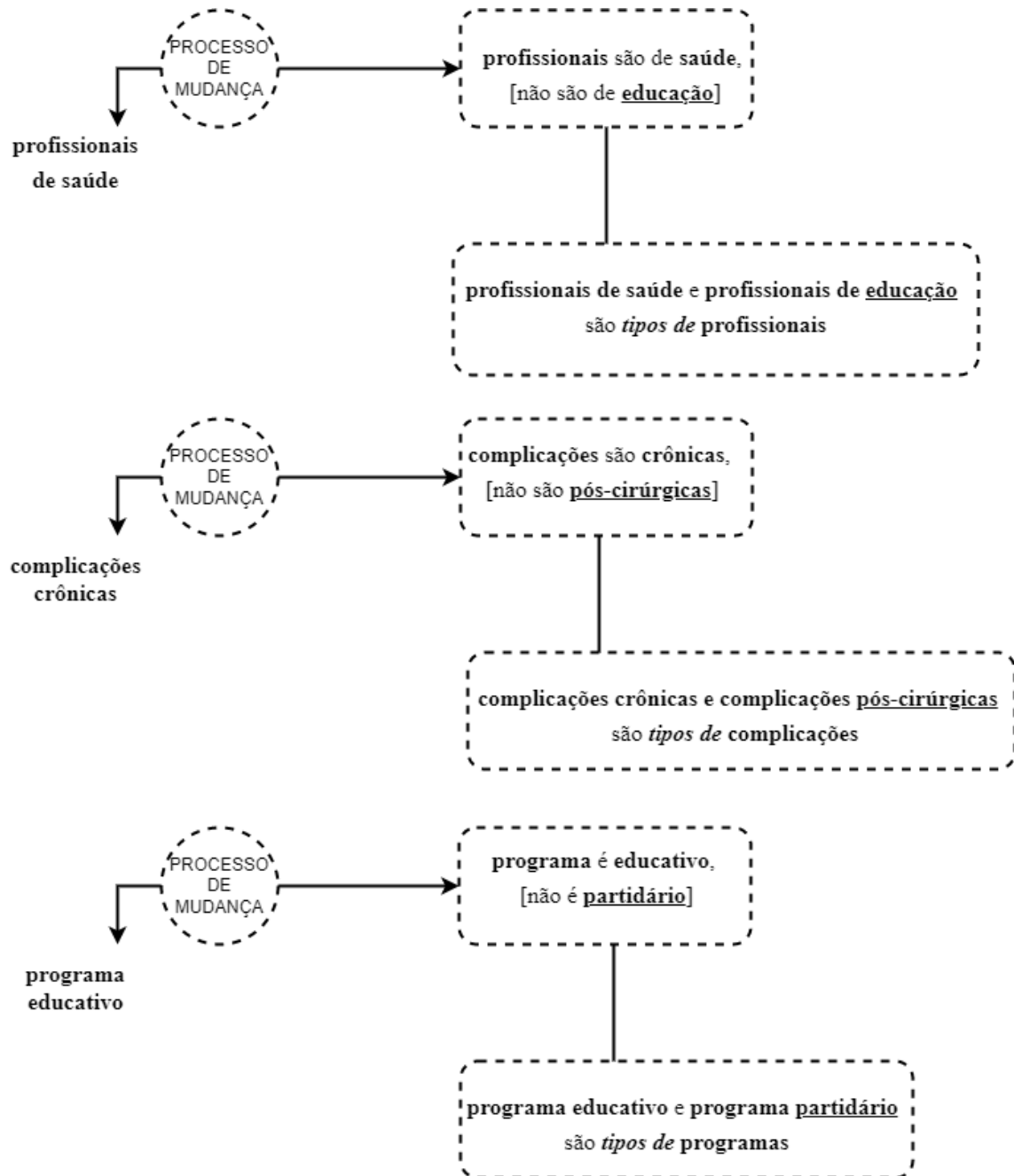


FIGURA 11 – Representação ProtinLex:PTC – processo de mudança

Fonte: Desenvolvida para fins deste trabalho

A FIG. 11 mostra três representações que exemplificam como o processo de mudança do ProtinLex:PTC ocorre. Essas representações trazem à esquerda o item lexical formado por mais de um componente lexical que passará pelo processo nomeado como mudança, o qual é simbolizado por um círculo feito em linhas pontilhadas. Em seguida, o item lexical em questão e o círculo que representa o processo de mudança são ligados por uma seta de pontas dupla a dois retângulos com hastes arredondadas, onde estão as orações resultantes do processo de mudança. Essas orações precisam estabelecer pelo menos uma relação de

COESÃO LEXICAL (hiponímia, meronímia, colocação, sinonímia e/ou antonímia) com outro item lexical. Este item lexical precisa ter sido instanciado a partir de configurações diferentes para as variáveis do contexto de situação, sendo, portanto, um item lexical proveniente de um corpus distinto daquele de onde foi retirado o item lexical formado por mais de um componente lexical que está em vias de passar pelo processo de mudança.

Analisando o processo de mudança detalhado para os itens lexicais formados por mais de um componente lexical representados na FIG. 11 fica claro que duas orações foram necessárias para investigar se as variáveis do contexto de situação podem influenciar na formação dos itens lexicais “profissionais de saúde”, “complicações crônicas” e “programa educativo”. Essa influência se dá em razão da relação de COESÃO LEXICAL do tipo colocação estabelecida na primeira oração resultante do processo de mudança para cada item lexical – localizada no retângulo superior, bem como da relação de COESÃO LEXICAL do tipo hiponímia determinada na segunda oração resultante do processo de mudança para cada item lexical – localizada no retângulo inferior (ver FIG. 11), a saber:

- i) “profissionais de” pode coocorrer com “saúde”, como é o exemplo retirado do corpus utilizado nesta pesquisa; e também pode coocorrer com “educação”, como é o exemplo extraído do corpus de suporte, que funciona como um apoio para a busca de exemplos que pudessem ser empregados no processo de mudança. Nesse sentido, tanto “profissionais de saúde” quanto “profissionais de educação” podem ser concebidos como tipos de profissionais.
- ii) “complicações” pode coocorrer com “crônicas”, como é o exemplo selecionado a partir do corpus deste estudo; e também pode coocorrer com “pós-cirúrgicas”, como visto no corpus de suporte. Nesse sentido, tanto “complicações crônicas” quanto “complicações pós-cirúrgicas” podem ser entendidas como tipos de complicações.
- iii) “programa” pode coocorrer com “educativo”, como acontece nos textos que compõem o corpus utilizado neste trabalho; e também pode coocorrer com “partidário”, como observado no corpus de suporte para o processo de mudança. Nesse sentido, tanto “programa educativo” quanto “programa partidário” podem ser tipos de programas.

Os resultados do processo de mudança dos itens lexicais “profissionais de saúde”, “complicações crônicas” e “programa educativo” revela que um item lexical formado por mais de um componente lexical pode ser desmembrado por meio de um processo de mudança que

tem como resultado duas orações, as quais podem ser relacionais e/ou materiais. Esses componentes lexicais desmembrados estabelecem relações coesivas de colocação (primeira oração) e de hiponímia (segunda oração) com outros itens lexicais extraídos, nesta pesquisa, de um corpus de suporte utilizado apenas para tal finalidade.

O estabelecimento dessas relações coesivas aponta para o fato de que no português brasileiro itens lexicais formados por mais de um componente lexical tendem a ser influenciados pelas seleções feitas para as variáveis do contexto de situação durante a construção de um texto do *corpus* (*corpora*) onde esses itens lexicais estão instanciados. Em outras palavras, quando as variáveis do contexto de situação que constituem os textos do *corpus* utilizado nesta pesquisa são “alteradas” para as variáveis do contexto de situação que configuram os textos do *corpus* de suporte, observa-se que os itens lexicais formados por mais de um componente lexical instanciam novos significados a partir da coocorrência com outros itens lexicais provenientes desse outro *corpus* ao mesmo tempo que a coocorrência que instancia o item lexical formado por mais de um componente lexical encontrada no corpus compilado para este estudo se mantêm. Isso significa dizer que itens lexicais formados por mais de um componente lexical podem ser influenciados pelas configurações das variáveis do contexto de situação que constituem os textos de um *corpus* (ou *corpora*), mas esses itens lexicais são capazes de coexistirem no sistema linguístico, no caso o português brasileiro, e podem ser instanciados em um texto conforme as especificidades de uma dada situação (situation type).

Por exemplo, “programa educativo” refere-se a um item lexical formado por mais de um componente lexical que foi identificado no corpus compilado para esta pesquisa (ver 4.1) e “programa partidário” diz respeito a um item lexical também formado por mais de um componente lexical que foi extraído do corpus de suporte utilizado neste trabalho (ver FIG. 11). Esses itens lexicais coexistem no sistema linguístico do português brasileiro, e embora tenham componentes lexicais análogos, sendo o primeiro componente lexical de ambos “programa”, apresentam significados diferentes e tendem a ser instanciados em um texto (ou textos de um *corpus/corpora*) conforme a configuração das variáveis do contexto de situação que constituem esse texto (ou textos de um *corpus/corpora*).

Esse resultado pode explicar duas propriedades dos itens lexicais que frequentemente são abordadas nos estudos linguísticos (ver 3.0): (i) itens lexicais costumam emergir na língua de forma mais rápida que itens gramaticais e (ii) a existência de um item lexical em uma dada língua não limita e/ou anula a existência de outro item lexical na mesma língua. Em (i), essa emergência pode ser justificada pelas seleções feitas para cada variável do

contexto de situação que constituem os textos construídos pelo português brasileiro, no caso, o que faz com itens lexicais formados por mais de um componente lexical, objeto de análise deste ProtinLex, possam emergir na língua. Em (ii), como itens lexicais formados por mais de um componente lexical são instanciados em um texto em função das configurações das variáveis contextuais que constituem esse texto, um item lexical desse tipo será instanciado em um texto x de acordo com o significado que esse item realizará no texto x, enquanto outro item lexical desse mesmo tipo pode ter um significado diferente do item lexical anterior e será instanciado em um texto y cuja configuração para as variáveis contextuais pode ser distinta daquela que constitui o texto x.

Ainda no que diz respeito aos resultados obtidos para o processo de mudança do ProtinLex:PTC, todos os itens selecionados a partir da lista de palavras-chave do corpus compilado para este estudo e que foram utilizados para o desenvolvimento e aplicação dos protocolos de investigação lexical (ver QUADRO 2) podem ser caracterizados como itens lexicais para o português brasileiro. Esse resultado se dá em função de três aspectos fundamentais investigados neste protocolo de investigação lexical, a saber:

- a) Padrões de coocorrência com outros itens lexicais foram identificados, por meio da busca por agrupamentos (*clusters/n-gramas*), formados à direita e à esquerda de todos os itens da lista de palavras-chave (ver QUADRO 2) utilizados neste ProtinLex. Alguns itens desta lista apresentaram ainda itens lexicais derivados, cuja classe de palavras é diferente daquela do item lexical que os originou. Esses itens lexicais derivados também apresentaram padrões de coocorrência, os quais foram identificados pela busca por agrupamentos, seguindo a mesma metodologia de busca usada para os itens da lista de palavras-chave (ver QUADRO 3 e QUADRO 4).
- b) Todos os agrupamentos detectados anteriormente foram explorados através da seleção de orações onde essas coocorrências estavam instanciadas nos textos do *corpus* compilado. Nessa exploração, as funções do sistema de TRANSITIVIDADE – Participante, Processo e Circunstância – dessas orações foram anotadas manualmente, a fim de verificar-se a co-extensividade entre itens lexicais formados por mais de um componente lexical e a ordem do grupo nominal, a qual geralmente realiza os elementos permanentes, i.e. Participantes e/ou Circunstâncias na ordem da oração (cf. FIGUEREDO, 2007, p. 141-144). Os resultados dessa anotação mostraram que a co-extensividade entre itens lexicais formados por mais de um

componente lexical e a ordem do grupo nominal parece ser uma característica para os itens selecionados a partir da lista de palavras-chave, bem como para os itens derivados uma vez que todos esses itens funcionaram, nas orações anotadas, como Participantes ou Circunstâncias (ver TABELA 1).

c) Todos os itens lexicais formados por mais de um componente lexical extraídos do *corpus* compilado para o presente trabalho apresentaram como resultado do processo de mudança desenvolvido para este protocolo de investigação lexical duas orações, ambas orações relacionais, onde as relações coesivas lexicais entre o item lexical formado por mais de um componente lexical extraído do *corpus* compilado para este estudo e o item lexical do mesmo tipo selecionado a partir do *corpus* de suporte foram estabelecidas, sendo sempre uma primeira relação COESIVA LEXICAL do tipo colocação e a segunda relação COESIVA LEXICAL do tipo hiponímia.

É válido ressaltar, por fim, que as representações do processo de mudança aplicado para todos os itens lexicais formados por mais de um componente lexical extraído do *corpus* compilado para a presente estão detalhadas na seção dos Anexos desta tese. A seguir, os resultados obtidos para o segundo protocolo de investigação lexical desenvolvido para este estudo são abordados.

5.2 Da aplicação do ProtinLex:PTG – *corpus* de artigos acadêmicos

Como apontado na subseção que explora o processo metodológico estabelecido para o ProtinLex:PTG – Protocolo de Investigação Lexical: Processo de Transformação da Gramática (ver 4.3.2), um dos objetivos deste protocolo é investigar uma propriedade que parece estar presente nos itens lexicais em português brasileiro, sobretudo naqueles itens lexicais formados por mais de um componente lexical cuja co-extensividade com a ordem do grupo nominal já estivesse comprovada pela aplicação do ProtinLex:PTC. Por esse motivo, os itens lexicais utilizados para o desenvolvimento e aplicação do protocolo de investigação lexical em questão precisavam ter tido sua co-extensividade com a ordem do grupo nominal atestada pelo ProtinLex anterior. É fundamental esclarecer que essa prerrogativa não foi vista como um impeditivo para o desenvolvimento deste protocolo e nem desta pesquisa uma vez que todos os itens lexicais formados por mais de um componente lexical identificados para o ProtinLex:PTC estavam em co-extensividade com a ordem do grupo nominal (ver 5.1 e TABELA 1). Considerando a simbiose entre esses dois protocolos de investigação lexical, o ProtinLex:PTG

se configura como uma extensão do ProtinLex:PTC, dado que os resultados deste depreendem os resultados que serão obtidos para aquele.

Nesse sentido, a propriedade com a qual o protocolo de investigação lexical em questão lida diz respeito à capacidade de transformação dos itens lexicais formados por mais de um componente lexical. Essa capacidade de transformação envolve a forma como itens lexicais desse tipo no português brasileiro podem ser desmembrado em porções – porção lexical e porção gramatical – bem como podem gerar outro(s) item(ns) lexical(is) que pode(m) estar no escopo de uma classe de palavras diferente daquela do item lexical que o gerou. Em outras palavras, os resultados do ProtinLex:PTG tratam de duas características dos itens lexicais formados por mais de um componente lexical no português brasileiro: (i) itens lexicais podem ser desmembrados em porções que, nesta pesquisa, recebem o nome de “porção gramatical” e “porção lexical” e (ii) itens lexicais formados por mais de um componente lexical podem gerar item(ns) lexical(is) que pode(m) estar no escopo de uma classe de palavras diferente daquela do item lexical que o gerou.

Para investigar essa característica de transformação dos itens lexicais formados por mais de um componente lexical, um processo de mudança foi elaborado partir do exame dos itens lexicais cuja co-extensividade havia sido estabelecida pelo ProtinLex:PTC. Nesse processo de mudança, os componentes que formavam o item lexical em questão eram isolados e, em seguida, passavam a funcionar como Participantes ou Circunstâncias em orações relacionais e/ou materiais resultantes deste processo de mudança. Adiante, a FIG. 12 mostra como os componentes do item lexical “saúde pública” foram isolados.

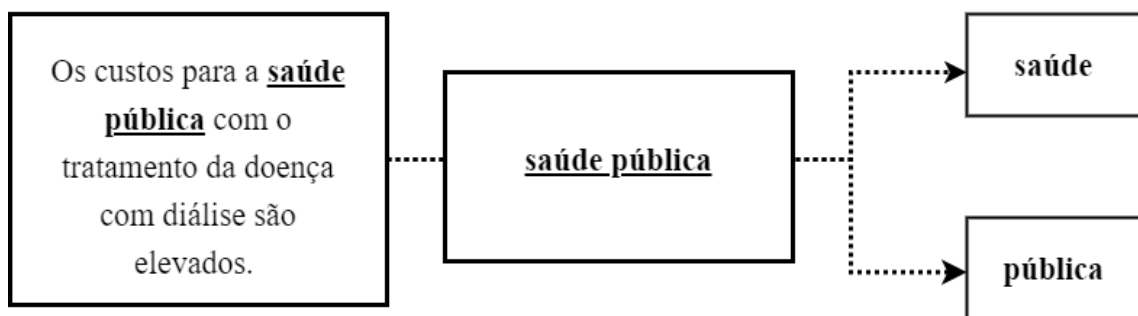


FIGURA 12 – Itens lexicais co-extensivos à ordem do grupo nominal: isolamento dos componentes lexicais de "saúde pública"

Fonte: Desenvolvida para fins deste trabalho.

À esquerda da FIG. 12 tem-se a oração de onde o item lexical “saúde pública” foi retirado. Essa oração foi selecionada partir dos textos que constituem o corpus compilado para esta pesquisa. No centro, o item lexical em questão é, de fato, isolado. À direita, aparecem os

componentes lexicais que formam o item lexical “saúde pública”, o primeiro componente – saúde – está inserido no retângulo superior; e o segundo componente – pública – está localizado no retângulo inferior.

Logo após esse isolamento, os componentes lexicais que formam o item lexical em questão estão prontos para serem utilizados no processo de mudança desenvolvido para este protocolo de investigação lexical. Ainda utilizando como exemplo o item lexical “saúde pública”, a representação detalhada na FIG. 13, à frente, ilustra como se deu o processo de mudança para tal item lexical.

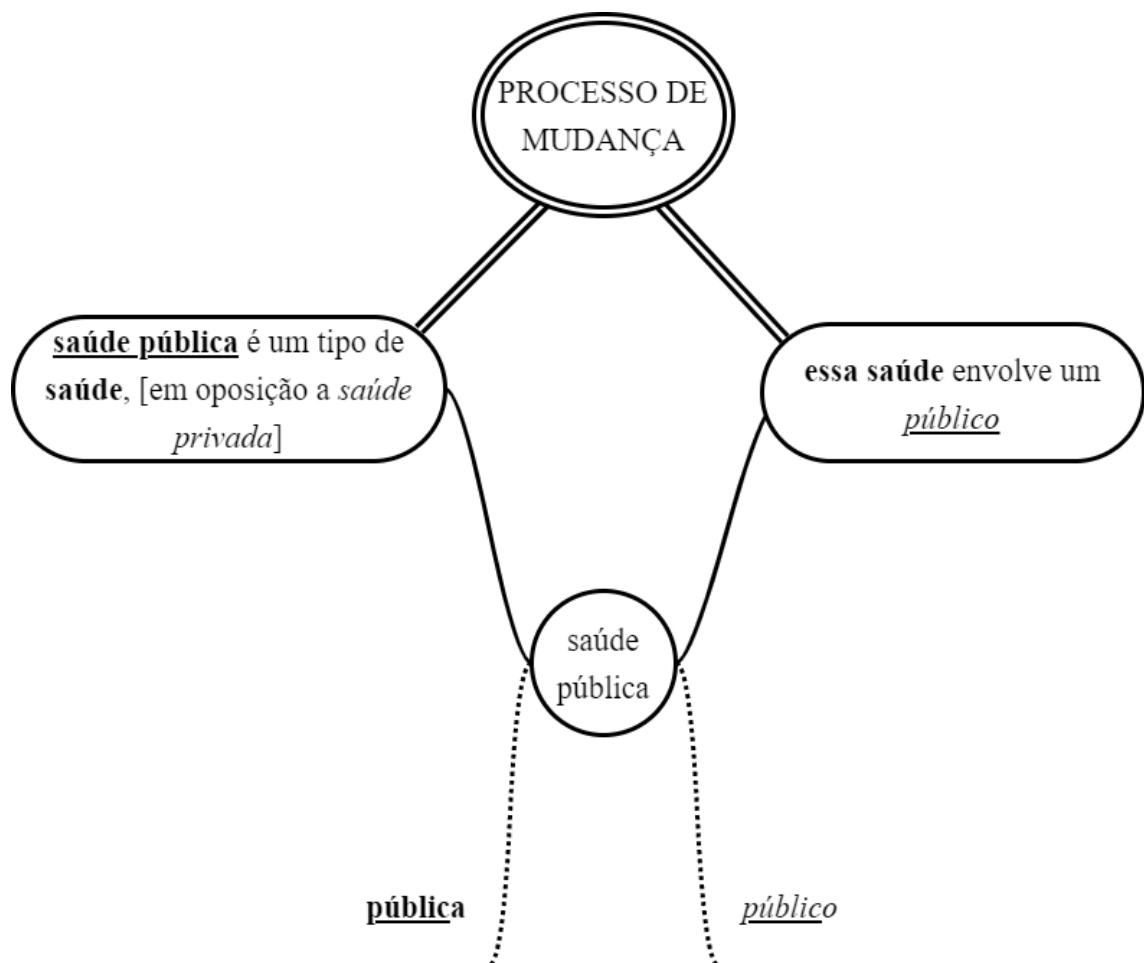
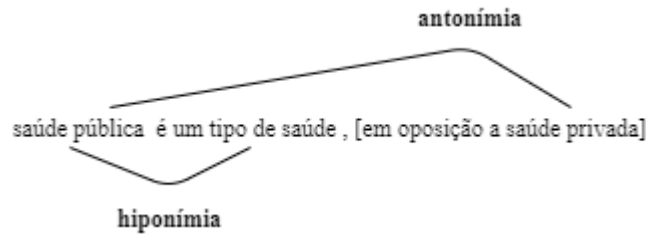


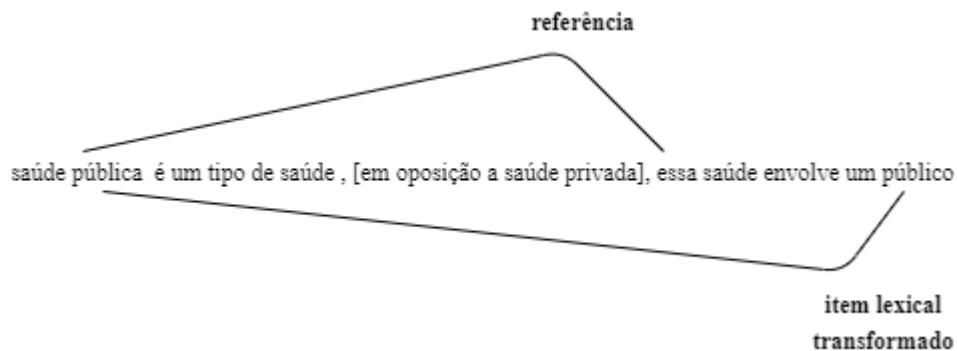
FIGURA 13 – Representação do ProtinLex:PTG – processo de mudança

Fonte: Desenvolvida para fins deste trabalho.

O processo de mudança do item lexical “saúde pública” representado pela FIG. 13 resulta em duas orações, dispostas paralelamente na parte superior do esquema. Localizada à esquerda está a oração que estabelece duas relações de COESÃO LEXICAL com os componentes do item lexical em questão, sendo a primeira relação do tipo hiponímia e a segunda do tipo colocação ou antonímia, a saber:



Inserida à direita está a segunda oração, na qual o item lexical “saúde”, que aparece na oração anterior, é referenciado por meio do determinante “essa” e o item lexical “pública” é transformado em “público”:



O pronome “essa”, utilizado para referenciar a qual saúde a segunda oração se refere, está no escopo da classe de palavras dos pronomes demonstrativos. É interessante ressaltar que essa classe de palavra dos pronomes é constituída por itens concebidos como gramaticais/funcionais. Isso significa dizer que a organização desses itens se dá pela *abstração* que compreende uma capacidade inerente aos itens gramaticais/funcionais de estarem atrelados a uma categoria gramatical mais geral (cf. HALLIDAY, MCINTOSH e STREVENS, 1964). Por exemplo, o item “essa” pode ser entendido como uma *abstração* da classe de palavras dos pronomes do tipo demonstrativos, feminino e singular (cf. Figueredo, 2007, p. 167), não havendo, no português brasileiro, outra abstração possível para essa categoria que não seja “essa”. Essa observação pode reforçar a organização complementar entre gramática e léxico na língua, em que itens gramaticais e itens lexicais operam juntos para a construção de um texto, por exemplo, assumindo uma visão de complementariedade, não de delicadeza em que o léxico é concebido como o ponto mais delicado da gramática, nem de contiguidade em que léxico e gramática compreendem polos do contínuo lexicogramatical (cf. HALLIDAY e MATTHIESSEN, 2014, p. 64-67).

No que diz respeito às funções gramaticais presentes na segunda oração resultante do processo de mudança do ProtinLex:PTG, localizada à direita da FIG. 13, o “essa saúde” é o Participante 1:Ator (Actor) e funciona como Sujeito (Subject) dessa oração material (material

clause), enquanto que o item lexical transformado “público” é o Participante 2:Escopo (Scope) e funciona como Complemento (Complement) dessa oração. É interessante salientar que esse resultado corrobora o que foi atestado no ProtinLex:PTC acerca da co-extensividade entre um item lexical formado por mais de um componente lexical e a ordem do grupo nominal e os elementos permanentes comumente realizados por Participantes e/ou Circunstâncias, funções do sistema de TRANSITIVIDADE cuja anotação se deu como um dos passos metodológicos propostos para o protocolo de investigação lexical anterior.

A parte inferior da representação do processo de mudança do ProtinLex:PTG detalhada na FIG. 13 também apresenta duas divisões. No lado esquerdo que fica abaixo da primeira oração, onde estão realizadas as relações COESIVAS LEXICAIS de hiponímia e de colocação ou antonímia, um dos componentes lexicais que formam o item lexical que passou pelo processo de mudança é inserido. Esse componente lexical que aparece em tal posição na representação em questão deve ser correspondente ao item lexical que gerou o item lexical transformado. No caso da representação da FIG. 13, esse item é “pública” que compreende um dos componentes lexicais que forma o item lexical “saúde pública”. No lado direito que fica abaixo da segunda oração, onde opera a COESÃO GRAMATICAL (GRAMMATICAL COHESION) pelo sistema de REFERÊNCIA (REFERENCE) – essa saúde – e onde o item lexical transformado é instanciado como Participante 2 dessa oração material, o item lexical transformado, que aparece na segunda oração resultante do processo de mudança, é inserido. No caso da representação da FIG. 13, esse item lexical transformado é o item “público”.

Portanto, a parte superior do esquema estabelecido para representar o processo de mudança do ProtinLex:PTG lida com o processo de mudança em si, apresentando as duas orações resultantes desse processo. A parte inferior trata de apresentar o item lexical que originou o item lexical transformado, bem como o próprio item lexical transformado. Por fim, a parte central é ocupada pelo item lexical formado por mais de um componente lexical e interliga a parte superior – processo de mudança – com a parte inferior – item lexical original e item lexical transformado – uma vez que esse item lexical localizado no centro do esquema está relacionado tanto com o processo de mudança, pois é a partir dele que esse processo ocorre, quanto com o(s) item(ns) lexical(is) transformado(s), que é(são) originário(s) de um dos componentes lexicais desse item lexical central.

Contrastando o item lexical que originou o item lexical transformado – pública e público, respectivamente – ambos descritos na parte inferior da FIG. 13, fica claro que o processo de mudança instituído para o ProtinLex:PTG consegue explorar a capacidade de transformação que parece ser inerente a alguns itens lexicais do português brasileiro formados

por mais de um componente lexical. Isso é justificado pela porção lexical que se mantém inalterada entre o item lexical original e o item lexical transformado – *públic* – e pela modificação que acontece apenas entre a porção gramatical do item lexical original para o item lexical transformado – *a* (*pública*) e *o* (*público*).

É essencial explicar que na presente pesquisa um item lexical é concebido como uma junção de duas porções, uma porção lexical e outra porção gramatical. A porção lexical é responsável por estabelecer o significado do item lexical em questão e se mantém inalterada entre o item lexical original e o(s) item(ns) lexical(is) transformado(s). A porção gramatical compreende a parte morfológica do item lexical e é a porção que sofre modificações entre o item lexical original e o(s) item(ns) lexical(is) transformado(s).

Essas modificações fazem com que o item lexical transformado possa estar no escopo de uma classe de palavras diferente daquela em que o item lexical que o originou se encontra. Isso acontece, por exemplo, para o item lexical “saúde pública” detalhado na FIG. 13. “Pública” está no âmbito da classe de palavras nominal no escopo dos adjetivos (adjectives); e “público” está no âmbito da classe de palavras nominal no escopo dos substantivos comuns (noun:common) (cf. HALLIDAY e MATTHIESSEN, 2014, p. 75). Nesse sentido, o processo de mudança desenvolvido para o protocolo de investigação lexical em questão deixa evidente também que o item lexical no português brasileiro é capaz de gerar itens lexicais transformados e, a partir disso, inseri-los em uma classe de palavras distinta da que este item lexical está localizado.

Em decorrência do fato da porção gramatical do item lexical ser a porção modificável na transformação de um determinado item lexical em outro(s) item(ns) lexical(is), a proposição de que o léxico é a parte da linguagem que evolui mais rápido, conseguindo mudar em um curto espaço de tempo para lidar com demandas de comunicação dos falantes de uma dada língua (cf. MARTIN; ROSE, 2007, p. 63) torna-se questionável. Segundo os resultados obtidos para o processo de mudança do ProtinLex:PTG, o que parece ter que mudar de forma mais rápida é a porção gramatical dos itens lexicais para suprimir uma demanda de uma língua em particular, não a porção lexical uma vez que esta permanece inalterada nos itens lexicais transformados. Embora a porção gramatical do item lexical seja a porção responsável por lidar com as mudanças necessárias para gerar um item lexical transformado, a visão de complementariedade entre léxico e gramática pode auxiliar na interpretação do funcionamento lexical a partir do processo de mudança estabelecido para este protocolo de investigação lexical, pois tanto a porção lexical quanto a porção gramatical precisam operar juntas para que um

significado, demandado pelos usuários de uma dada língua, seja instanciado em um texto (ou textos) por meio de um item lexical transformado.

Depois de abordar os resultados obtidos mediante o desenvolvimento e a aplicação do ProtinLex:PTG para o item lexical “saúde pública”, os padrões e as especificidades encontradas para os itens lexicais formados por mais de um componente lexical utilizados na aplicação neste protocolo de investigação lexical são discutidas na próxima subseção.

5.2.1 Da aplicação do ProtinLex:PTG – resultado default

Os resultados obtidos para o item lexical “saúde pública”, detalhados na subseção anterior a partir da descrição da FIG. 13, desenvolvida para representar o processo de mudança do ProtinLex:PTG, configuram como os resultados default (padrão), ou seja, independentemente de qual item lexical estava envolvido no processo de mudança do ProtinLex:PTG, o item lexical transformado era obtido por meio das relações coesivas lexicais, as quais eram instanciadas em orações relacionais e/ou materiais, análogas àquelas estabelecidas para o item lexical “saúde pública”. Esta repetição de resultados para distintos itens lexicais formados por mais de um componente lexical configura para a presente pesquisa um resultado default. Considerando o QUADRO 3 e o QUADRO 4 onde estão os agrupamentos formados à direita e à esquerda dos itens utilizados para desenvolvimento e aplicação dos protocolos de investigação lexical deste estudo, cuja co-extensividade com ao grupo nominal foi identificada através da exploração dos resultados encontrados no ProtinLex:PTC, o QUADRO 5, adiante, revela quais desses itens tiveram resultados semelhantes àqueles do item lexical “saúde pública”.

QUADRO 5 – Itens lexicais co-extensivos à ordem do grupo nominal: resultado default para o processo de mudança do ProtinLex:PTG

Itens selecionados a partir da lista de palavras-chave	Itens lexicais co-extensivos à ordem do grupo nominal com resultado default	
autocuidado	autocuidado em diabetes	
saúde	saúde pública	unidades básicas de saúde*
	atenção primária à saúde*	letramento funcional em saúde*
	equipe multiprofissional de saúde*	alimentação saudável

	hábitos saudáveis	estilo de vida saudável*
doença	doenças crônicas	tempo de doença
	doença renal	enfrentamento da doença
	doenças cardíacas	manejo da doença
	controle da doença	tratamento da doença
usuários	usuários com diabetes mellitus	usuários do grupo intervenção*
	usuários com DM2	uso de insulina
	usuários do grupo controle*	uso de medicamentos
educação	educação alimentar	programa educativo
	educação individual	processo educativo
	educação em grupo	ações educativas
	educação em diabetes	intervenções educativas
	efeito da educação	práticas educativas
	estratégias educacionais	
complicações	complicações crônicas	prevenção de complicações
	complicações microvasculares	presença de complicações
	complicações do diabetes	complicações agudas e crônicas
indivíduos	intervenção individual	
pés	pé diabético	
ações	ações educativas	ações de educação

prática	prática de exercícios físicos	prática de atividade física
	prática educativa	

Legenda: Os itens destacados com o asterisco(*) apresentam particularidades que serão abordadas à frente.

Fonte: Desenvolvido para fins deste trabalho.

Dentre os itens lexicais formados por mais de um componente lexical cuja co-extensividade com a ordem do grupo nominal foi identificada por meio do ProtinLex:PTC, estão os itens lexicais destacados no QUADRO 5 que apresentaram resultado default para o processo de mudança desenvolvido para o ProtinLex:PTG. Isso indica que um dos componentes lexicais desses itens foi capaz de gerar pelo menos um item lexical transformado a partir de relações COESIVAS LEXICAIS do tipo hiponímia e colocação ou antonímia, bem como a partir da COESÃO GRAMATICAL pelo sistema de REFERÊNCIA. É fundamental ressaltar que tanto as relações COESIVAS LEXICAIS quanto as relações COESIVAS GRAMATICAS foram exploradas através de orações relacionais e materiais estabelecidas para o item lexical em foco durante o processo de mudança do ProtinLex:PTG.

Ainda com relação aos itens detalhados no QUADRO 5, o número de itens lexicais que foram capazes de gerar itens lexicais transformados não está relacionado ao número de ocorrências do item selecionado a partir da lista de palavras-chave para ser utilizado no desenvolvimento e aplicação dos protocolos de investigação lexical desta pesquisa. Por exemplo, o item lexical “autocuidado” compreende o primeiro na lista de palavras-chave (ver QUADRO 2) e apresenta 13 itens lexicais cuja co-extensividade com a ordem do grupo nominal foi apontada pelo processo de mudança do ProtinLex:PTC (ver QUADRO 3). Contudo, dentre esses 13 itens lexicais, somente um – autocuidado em diabetes – gerou itens lexicais transformados a partir do processo de mudança do ProtinLex:PTG. Esse resultado revela que o número de itens lexicais transformados parece estar associado mais aos textos²⁷ que formam o corpus deste estudo do que à quantidade de itens lexicais em co-extensividade com a ordem do grupo nominal que esse determinado item apresenta no corpus. Em outras palavras, se os textos que constituem o corpus compilado para a presente pesquisa fossem de outro tipo e/ou abordassem outro tema diferente do “autocuidado em diabetes mellitus”, os itens lexicais transformados poderiam ser outros, mesmo que a metodologia estabelecida para o processo de mudança do ProtinLex:PTG se mantivesse.

²⁷Sobretudo no que diz respeito ao domínio experiencial (domain of experience) e a atividade sócio-semiótica (socio-semiotic activity) dos textos que compõem o corpus compilado.

Para além da relação dos itens lexicais transformados e o corpus compilado para este estudo, uma outra associação parece estar evidente quando os itens retratados no QUADRO 5 são comparados entre si. Essa associação diz respeito ao fato de que a maior parte desses itens conta com pelo menos um componente cuja função pode ser considerada como co-extensiva à função do Classificador²⁸ (Classifier), função que está no escopo do grupo nominal (cf. FIGUEREDO, 2007; HALLIDAY e MATTHIESSEN, 2014) ou, em menor proporção, como co-extensiva à função do Qualificador²⁹, função que também está no âmbito do grupo nominal (cf. FIGUEREDO, 2007; HALLIDAY e MATTHIESSEN, 2014; FERREGUETTI, 2018). A seguir, o QUADRO 6 traz os itens lexicais que apresentam componentes lexicais co-extensivos à função do Classificador, bem como revela os itens lexicais transformados a partir destes componentes, os quais são resultado do processo de mudança do ProtinLex:PTG. Estes itens lexicais transformados estão sublinhados e destacados em negrito na segunda coluna do quadro abaixo.

QUADRO 6 – Componentes lexicais co-extensivos à função do Classificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado default

Itens lexicais com um dos componentes co-extensivos à função do Classificador	Orações resultantes do processo de mudança + Itens lexicais transformados
saúde pública	saúde pública é um tipo de saúde, [em oposição a saúde privada] essa saúde envolve um público .
alimentação saudável	alimentação saudável é um tipo de alimentação, [em oposição a alimentação escolar] essa alimentação envolve saúde .
hábitos saudáveis	hábitos saudáveis são tipos de hábitos, [em oposição aos hábitos culturais] esses hábitos envolvem saúde .

²⁸Segundo Figueredo (2007, p. 204 e 220), o Classificador é uma das funções pertinentes ao grupo nominal utilizada para delimitar o Ente (Thing) em termos de subclasse. Geralmente, é realizado por palavras da classe dos substantivos e/ou adjetivos e relaciona-se à COESÃO LEXICAL do tipo hiponímia.

²⁹De acordo com Figueredo (2007, p. 231), o Qualificador é concebido como uma função que caracteriza um Ente na estrutura do grupo nominal. Esta função geralmente é realizada por uma oração encaixada (bound clause), por uma oração não-finita (non-finite clause) ou por uma frase preposicional (prepositional phrase) que operam dentro do grupo nominal.

doenças crônicas	doenças crônicas são tipos de doença, [em oposição as doenças autoimunes]
	essas doenças envolvem <u>cronicidade</u> .
doença renal	doença renal é um tipo de doença, [em oposição a doença mental]
	essa doença envolve os <u>rins</u> .
doenças cardíacas	doenças cardíacas são tipos de doença, [em oposição a doenças infecciosas]
	essas doenças envolvem <u>cardíacos</u> .
programa educativo	programa educativo é um tipo de programa, [em oposição a programa partidário]
	esse programa envolve <u>educação/educadores</u> .
educação alimentar	educação alimentar é um tipo de educação, [em oposição a educação ambiental]
	essa educação envolve <u>alimentos/alimentação</u> .
processo educativo	processo educativo é um tipo de processo, [em oposição a processo seletivo]
	esse processo envolve <u>educação/educadores</u> .
educação individual	educação individual é um tipo de educação, [em oposição a educação infantil]
	essa educação envolve <u>indivíduos/individualidade</u> .
ações educativas	ações educativas são tipos de ações, [em oposição a ações preventivas]
	essas ações envolvem <u>educação/educadores</u> .
intervenções educativas	intervenções educativas são tipos de intervenções, [em oposição a intervenções artísticas]
	essas intervenções envolvem <u>educação/educadores</u> .
práticas educativas	práticas educativas são tipos de práticas [em oposição as práticas sociais]
	essas práticas envolvem <u>educação/educadores</u> .
estratégias educacionais	estratégias educacionais são tipos de estratégias, [em oposição a estratégias políticas]

	essas estratégias envolvem <u>educação/educadores.</u>
complicações crônicas	complicações crônicas são tipos de complicações, [em oposição a complicações pós-cirúrgicas]
	essas complicações envolvem <u>cronicidade.</u>
complicações microvasculares	complicações microvasculares são tipos de complicações, [em oposição a complicações respiratórias]
	essas complicações envolvem <u>microvascularidade.</u>
complicações agudas e crônicas	complicações agudas e crônicas são tipos de complicações, [em oposição a complicações graves e irreversíveis]
	essas complicações envolvem <u>agudamento e cronicidade.</u>
intervenção individual	intervenção individual é um tipo de intervenção, [em oposição a intervenção federal]
	essa intervenção envolve <u>indivíduos.</u>
pé diabético	pé diabético é um tipo de pé, [em oposição ao pé direito]
	esse pé envolve <u>diabetes.</u>
prática de atividade física	prática de atividade física é um tipo de prática, [em oposição a prática de assédio moral]
	essa prática envolve a <u>ativação do físico.</u>
prática de exercícios físicos	prática de exercício físico é um tipo de prática, [em oposição a prática de assédio moral]
	essa prática envolve a <u>exercitação física.</u>
prática educativa	prática educativa é um tipo de prática, [em oposição a prática esportiva]
	essa prática envolve <u>educação/educadores.</u>

Fonte: Desenvolvido para fins deste trabalho.

Os dados retratados no QUADRO 6 evidenciam que os componentes lexicais cuja co-extensividade com a função do Classificador pode ser verificada correspondem aos mesmos componentes que geraram os itens lexicais transformados. Por exemplo, ao mesmo tempo que o componente “crônica” de “doenças crônicas” pode ser entendido como um componente

lexical cuja função pode ser co-extensiva à função do Classificador, uma vez que “crônica” delimita uma subclasse das doenças existentes no mundo, em oposição, por exemplo, às doenças autoimunes, este componente também gera o item lexical transformado – cronicidade – mediante o processo de mudança aplicado para o item lexical “doenças crônicas”. É preciso salientar que “cronicidade” é o item lexical transformado a partir de “crônicas” em razão da permanência da porção lexical “croni” em conjunto com a alteração da porção gramatical de “cas” para “cidade” em “cronicidade”, sendo este o resultado default esperado para os itens lexicais co-extensivos ao grupo nominal que passaram pelo processo de mudança do ProtinLex:PTG.

Este resultado reforça a proposta examinada ao longo da abordagem do ProtinLex:PTC acerca do fato de que a emergência de itens lexicais formados por mais de um componente lexical é mediada por configurações específicas das variáveis do contexto de situação que constituem o(s) texto(s). Em outras palavras, em um dado item lexical cujo um dos componentes lexicais é co-extensivo à função do Classificador, o significado desse componente parece ser utilizado pelo sistema linguístico, no caso, o português brasileiro, para suprimir uma necessidade daquela língua quando uma nova classificação de determinado item lexical já existente naquele sistema precisa ser instanciada no(s) texto(s) em questão. Em razão das configurações das variáveis contextuais que organizam os textos do corpus utilizado neste trabalho, o componente lexical “individual”, por exemplo, que está em co-extensividade com a função do Classificador, apresenta coocorrência com “educação” no corpus, a fim de solucionar uma demanda do sistema linguístico que parece ser o de estabelecer uma nova subclasse de “educação” para os textos do corpus em que o item “educação individual” está instanciado.

Em relação aos itens lexicais transformados, sublinhados e destacados em negrito na segunda coluna do QUADRO 6 dois aspectos principais não podem deixar de ser explorados. O primeiro deles compreende o fato de que embora os itens lexicais formados por mais de um componente lexical possam ser influenciados pelas configurações das variáveis do contexto de situação (ver 5.1), os itens lexicais transformados parecem não ser influenciados por essas variáveis. Por exemplo, “educativa” em “prática educativa” poderia coocorrer com outro item lexical, como, por exemplo, “medida educativa³⁰”, gerando um item lexical formado por mais de um componente lexical cuja co-extensividade com a ordem do grupo nominal poderia ser identificada, e, mesmo assim, a transformação de “educativa” se mostraria como “educação” e/ou “educadores”. Este exemplo leva ao segundo aspecto que diz respeito ao fato de que o

³⁰Medida educativa é um tipo de medida, [em oposição a medida protetora]. Essa medida envolve educação/educadores.

componente lexical que é co-extensivo à função do Classificador está no escopo da classe de palavras do adjetivo, enquanto que o item lexical transformado gerado a partir deste componente lexical está no escopo da classe de palavras do substantivo, ou seja, a mudança de classe de palavras entre item (componente) lexical e item lexical transformado pode ser reconhecida como uma das características inerentes ao item lexical. Isso significa dizer que o item lexical parece ter uma capacidade de transformação associada à mudança de classe de palavras, ao passo que o item chamado de gramatical não parece ter essa propriedade, pois a “transformação” que este item gera não resulta em um item gramatical transformado, nem em um item gramatical transformado mudado de classe de palavras. A “transformação” evidenciada pelo item gramatical envolve somente as flexões de gênero, número e pessoa³¹ e sem mudança de classe de palavras. Por exemplo, os pronomes possessivos “sua”, no feminino, e “seu”, no masculino, podem ser “transformados” em “suas”, no feminino plural, e em “seus”, no masculino plural, mas continuam no escopo da classe de palavras dos pronomes possessivos.

Essa associação entre itens lexicais transformados e mudança de classe de palavras, que parece tratar-se de uma característica dos itens lexicais destacados na segunda coluna do QUADRO 6, contradiz a ideia de que novos itens lexicais surgem na língua a todo momento, sem evidências acerca do mecanismo e/ou do momento específico da produção de significado em que um item lexical está relacionado à estrutura gramatical (ver 1.0). Tal contradição pode ser investigada pelo número de itens lexicais transformados para cada componente lexical presente nos itens lexicais formados por mais de um componente lexical destacados na primeira coluna do QUADRO 6. A maior parte desses componentes gerou um item lexical transformado apenas – cronicidade, agudamento, ativação, físico, exercitação, física, microvascularidade, cardíacos, rins, saúde e público –, sendo que o número máximo de itens lexicais transformados por cada componente lexical foi 2 – educação/educador, indivíduo/individualidade, alimentos/alimentação e diabetes/diabético. Esse resultado indica que pode haver, no âmbito de um dado sistema linguístico, um potencial de itens lexicais que pode ser atingido por meio das transformações desses itens em conjunto com as associações de classe de palavras. Em outras palavras, os itens lexicais transformados que estão no escopo da classe de palavras dos substantivos e são provenientes de itens (componentes) lexicais que estavam no escopo da classe de palavras dos adjetivos tendem a gerar no máximo dois itens lexicais transformados,

³¹Gênero – feminino/masculino. Ex.: esta/este, essa/esse, a/o, etc.

Número – singular/plural. Ex.: esta/estas, este/estes, essa/essas, esse/esses, a/as, o/os, etc.

Pessoa – 1ª pessoa, 2ª pessoa e 3ª pessoa do singular/1ª pessoa, 2ª pessoa e 3ª pessoa do plural.

Ex.: eu, tu, ela/ele e nós, vos e eles/elas.

contrapondo a visão de que itens lexicais surgem na língua de maneira quase que indiscriminada.

Adiante, o QUADRO 7 apresenta os itens lexicais cujo um dos componentes lexicais é co-extensivo à função do Qualificador, bem como introduz os itens lexicais transformados a partir destes componentes, os quais são resultado do processo de mudança do ProtinLex:PTG. Estes itens lexicais transformados estão sublinhados e destacados em negrito na segunda coluna do quadro abaixo.

QUADRO 7 – Componentes lexicais co-extensivos à função do Qualificador do grupo nominal:
processo de mudança do ProtinLex:PTG – resultado default

Itens lexicais com um dos componentes co-extensivos à função do Qualificador	Orações resultantes do processo de mudança + Itens lexicais transformados
autocuidado em diabetes	autocuidado em diabetes é um tipo de autocuidado, [em oposição a autocuidado em saúde mental]
	esse autocuidado envolve <u>diabético/pessoa com diabetes.</u>
tempo de doença	tempo de doença é um tipo de tempo, [em oposição a tempo de serviço]
	esse tempo envolve <u>doentes.</u>
enfrentamento da doença	enfrentamento da doença é um tipo de enfrentamento, [em oposição a enfrentamento da crise financeira]
	esse enfrentamento envolve <u>doentes.</u>
manejo da doença	manejo da doença é um tipo de manejo, [em oposição a manejo da lavoura]
	esse manejo envolve <u>doentes.</u>
tratamento da doença	tratamento da doença é um tipo de tratamento, [em oposição a tratamento da empresa]
	esse tratamento envolve <u>doentes.</u>
controle da doença	controle da doença é tipo de controle, [em oposição ao controle da inflação]
	esse controle envolve <u>doentes.</u>
uso de medicamentos	uso de medicamentos é um tipo de uso, [em oposição a uso de recursos naturais]

	esse uso envolve <u>medicação.</u>
educação em grupo	educação em grupo é um tipo de educação, [em oposição a educação em casa]
	essa educação envolve <u>agrupar/agrupamento.</u>
educação em diabetes	educação em diabetes é um tipo de educação, [em oposição a educação em ciências]
	essa educação envolve o <u>diabético/pessoa com diabetes.</u>
efeito da educação	efeito da educação é um tipo de efeito, [em oposição a efeito da crise]
	esse efeito envolve <u>educadores</u>
ações de educação	ações de educação são tipos de ações, [em oposição as ações de saneamento]
	essas ações envolvem <u>educadores.</u>
prevenção de complicações	prevenção de complicações são tipos de prevenção, [em oposição a prevenção de doenças]
	essa prevenção envolve <u>complicadores.</u>
presença de complicações	presença de complicações é um tipo de presença, [em oposição a presença de especialistas]
	essa presença envolve <u>complicadores.</u>
complicações do diabetes	complicações do diabetes são tipos de complicações, [em oposição as complicações do parto]
	essas complicações envolvem o <u>diabético/pessoa com diabetes.</u>

Fonte: Desenvolvido para fins deste trabalho.

Para além dos itens lexicais co-extensivos à ordem do grupo nominal, cujo um dos componentes lexicais é co-extensivo à função do Classificador (ver QUADRO 6), o QUADRO 7 traz, na coluna da esquerda, os itens lexicais co-extensivos à ordem do grupo nominal que passaram pelo processo de mudança do ProtinLex:PTG, gerando itens lexicais transformados; sendo que um dos componentes desses itens lexicais está em co-extensividade à função do Qualificador, este realizado por frases preposicionais (prepositional phrases)³². Na coluna da

³²A frase preposicional compreende, em conjunto, com o grupo uma das ordens da escala de ordens da gramática, a qual é formada pelas seguintes ordens: morfema, palavra, grupo/frase preposicional e oração. A frase

direita do QUADRO 7, tem-se as orações com as relações COESIVAS LEXICAIS de hiponímia e de colocação ou antonímia, bem como as orações onde estão construídas as relações de COESÃO GRAMATICAL pelo sistema de REFERÊNCIA e a realização do item lexical transformado. A análise dessas duas colunas demonstra que, mesmo que em menor escala de ocorrências, há itens lexicais formados por mais de um componente lexical, cuja co-extensividade foi identificada por meio da aplicação do ProtinLex:PTC, em que um desses componentes lexicais é co-extensivo à função do Qualificador, função inerente ao grupo nominal (cf. FIGUEREDO, 2007). Assim como aqueles itens lexicais descritos no QUADRO 6 em que o segundo componente compreende aquele que é co-extensivo à função do Classificador, os componentes, que são co-extensivos à função do Qualificador (ver QUADRO 7), compreendem também os segundos componentes, no caso em questão, co-extensivos a uma preposição + um grupo nominal, i.e., um Qualificador. Tal como acontece com os componentes lexicais co-extensivos à função do Classificador (ver QUADRO 6), que são responsáveis por gerar o item lexical transformado, os componentes lexicais co-extensivos à função do Qualificador também geram os itens lexicais transformados, detalhados na coluna da direita do QUADRO 7. O número máximo de itens lexicais transformados gerado por cada componente lexical co-extensivo à função do Qualificador é dois; e isso ocorre nas instâncias em que esse componente é realizado por “em/do diabetes” e “em grupos” que geram, respectivamente, os itens lexicais transformados “diabético”/“pessoa com diabetes” e “agrupar”/“agrupamento”. Os outros componentes geram somente um item lexical transformado, como, por exemplo, “de educação” em “ações de educação”, após o processo de mudança do ProtinLex:PTG, gera o item lexical transformado “educadores”.

Uma das particularidades dos itens lexicais co-extensivos à ordem do grupo nominal, cujo o segundo componente é co-extensivo à função do Qualificador, diz respeito ao fato de que a maioria desses itens lexicais tem como primeiro componente um item lexical que está no escopo da classe de palavras nominal: substantivos comuns, mas é um item proveniente da classe de palavras verbal: verbo lexical. Essa particularidade pode ser constatada nos seguintes itens lexicais³³, a saber:

preposicional é constituída de uma preposição (preposition) + um grupo nominal (nominal group) (cf. FIGUEREDO, 2007; FERREGUETTI, 2018).

³³O desenvolvimento destas orações a partir dos itens lexicais destacados no QUADRO 7 baseia-se nos testes estabelecidos por Ferregueti (2018) para explorar as frases preposicionais com função de Qualificador em grupos nominais extraídos de textos em português.

- **Enfrentamento da doença** – [alguém] **enfrenta** a doença/ a doença é **enfrentada** [por alguém]
- **Manejo da doença** – [alguém] **maneja** a doença/ a doença é **manejada** [por alguém]
- **Tratamento da doença** – [alguém] **trata** a doença/ a doença é **tratada** [por alguém]
- **Controle da doença** – [alguém] **controla** a doença/ a doença é **controlada** [por alguém]
- **Educação em grupos** – em grupos [alguém] **educa** [algo] ou [alguém] **educa** [algo] em grupos/ em grupos [algo] é **educado** [por alguém] ou [algo] é **educado** [por alguém] em grupos
- **Educação em diabetes** – em diabetes [alguém] **educa** [alguém/algo] ou [alguém] **educa** [alguém/algo] em diabetes/ em diabetes [alguém/algo] é **educado** [por alguém] ou [alguém/algo] é **educado** [por alguém] em diabetes
- **Prevenção de complicações** – [algo] **previne** complicações/ complicações são **prevenidas** [por algo]
- **Presença de complicações** – [alguém] **presencia** complicações/ complicações são **presenciadas** [por alguém]
- **Complicações do diabetes** – [algo] **complica** o diabetes/ o diabetes é **complicado** [por algo]

Esse fato pode ser justificado pelas configurações das variáveis do contexto de situação, sobretudo pelo tipo de texto (atividade sócio-semiótica) e pelo tema principal (domínio experiencial) dos textos que formam o corpus utilizado neste trabalho, uma vez que o discurso científico tende a ser construído a partir de uma linguagem especializada, na qual metáforas gramaticais³⁴ são amplamente utilizadas (cf. HALLIDAY e MATTHIESSEN, 2014, p. 709).

Uma das formas utilizadas pela linguagem para construir uma metáfora gramatical compreende a nominalização que é a forma identificada para os itens lexicais descritos acima.

³⁴Halliday e Matthiessen (2014, p.711) estabelecem que a metáfora gramatical (ou metáfora ideacional) é um recurso utilizado pela linguagem para construir novas camadas de significado por meio de seleções feitas pela gramática. Isso significa dizer que orações não-finitas e encaixadas, por exemplo, que funcionam como Sujeito de uma outra oração podem ser entendidas como uma metáfora gramatical. O mesmo acontece para as nominalizações em que instâncias da classe de palavras verbal funcionam em uma oração como Sujeito ou Participante, por exemplo, enquanto que o mais frequente é que palavras dessa classe operem como Processos.

Na nominalização, instâncias que estão no escopo da classe de palavras nominal: substantivo são provenientes de instâncias que estão no escopo da classe de palavras verbal: verbo, ou seja, nesta forma de metáfora gramatical, há uma mudança de classe de palavras³⁵.

Uma situação parecida com a metáfora gramatical que ocorre para os itens lexicais discutidos acima acontece para os itens lexicais transformados, que estão no escopo da classe de palavras nominal: substantivo e funcionam, após o processo de mudança dos ProtinLexis, como Participante. Essa semelhança está no fato de que esses itens lexicais transformados são provenientes de componentes lexicais co-extensivos à função do Classificador ou do Qualificador, sendo que esses componentes estão, respectivamente, no escopo das classes de palavra nominal: adjetivo ou da classe de palavras verbal: preposição e da classe de palavras nominal: substantivo. Em outras palavras, os itens lexicais transformados, que são resultado dos processos de mudança dos protocolos de investigação lexical desenvolvidos para esta pesquisa, operam na oração como Participantes e estão no escopo da classe de palavras dos substantivos, ao passo que quando esses itens lexicais transformados eram componentes dos itens lexicais co-extensivos à ordem do grupo nominal, esses componentes estavam no escopo de classes de palavras diferentes e funcionavam de maneira distinta.

O QUADRO 8, à frente, aborda os itens lexicais co-extensivos à ordem do grupo nominal, que estão sinalizados com um * (asterisco) no QUADRO 5, e apresentam alguns desvios em relação aos resultados default atestados para os itens lexicais detalhados nos QUADROS 6 e 7.

QUADRO 8 – Componentes lexicais co-extensivos à função do Classificador + do Qualificador do grupo nominal: processo de mudança do ProtinLex:PTG – resultado default + desvio

Itens lexicais com um dos componentes co-extensivos à função do Classificador + Qualificador	Orações resultantes do processo de mudança + Itens lexicais transformados
unidades básicas de saúde*	unidades básicas de saúde são tipos de unidades, [em oposição às unidades de vigilância em saúde]
	essas unidades básicas envolvem saúde.

³⁵Há também uma mudança de função, mas que não é abordada no presente trabalho, por ser uma questão que se distancia do problema de pesquisa aqui proposto.

unidades básicas	unidades básicas são tipos de unidades, [em oposição as unidades mistas]
	essas unidades envolvem o <u>básico</u> .
letramento funcional em saúde*	letramento funcional em saúde é um tipos de letramento funcional, [em oposição a letramento funcional em programação]
	esse letramento funcional envolve saúde.
letramento funcional	letramento funcional é um tipo de letramento, [em oposição a letramento digital]
	esse letramento envolve <u>funcionalidade/funcões</u> .
atenção primária à saúde*	atenção primária à saúde é um tipo de atenção primária, [em oposição à atenção primária aos usuários de drogas]
	essa atenção primária envolve saúde.
atenção primária	atenção primária é um tipo de atenção, [em oposição a atenção especial]
	essa atenção envolve o <u>primeiro</u> [atendimento].
equipe multiprofissional de saúde*	equipe multiprofissional de saúde é um tipo de equipe multiprofissional, [em oposição a equipe de professores da UEL]
	essa equipe multiprofissional envolve saúde.
equipe multiprofissional	equipe multiprofissional é um tipo de equipe [em oposição a equipe técnica]
	essa equipe envolve <u>multiprofissionalismo</u> .
estilo de vida saudável	estilo de vida saudável é um tipo de estilo de vida, [em oposição a estilo de vida vegano]
	esse estilo de vida envolve <u>saúde</u> .
estilo de vida	estilo de vida é um tipo de estilo, [em oposição a estilo de jogo]
	esse estilo envolve vida.
usuários do grupo controle	usuários do grupo controle são tipos de usuários, [em oposição aos usuários do transporte coletivo]
	esses usuários envolvem grupo controle.

grupo controle	grupo controle é um tipo de grupo, [em oposição a grupo intervenção]
	esse grupo envolve [um] <u>controlador</u> .
usuários do grupo intervenção	usuários do grupo intervenção são tipos de usuários, [em oposição aos usuários do grupo RH]
	esse usuários envolvem grupo intervenção.
grupo intervenção	grupo intervenção é um tipo de grupo, [em oposição a grupo controle]
	esse grupo envolve [um] <u>interventor</u> .

Fonte: Desenvolvido para fins deste trabalho.

A primeira coluna do QUADRO 8 apresenta os itens lexicais que foram identificados como co-extensivos à ordem do grupo nominal, através da aplicação do processo de mudança desenvolvido para o ProtinLex:PTC, e cujo um dos componentes lexicais gerou um item lexical transformado, por meio do processo de mudança estabelecido para o ProtinLex:PTG. A segunda coluna deste mesmo quadro traz as orações onde as relações coesivas lexicais e gramaticais, resultantes do processo de mudança do ProtinLex:PTG, estão instanciadas em conjunto com o item lexical transformado.

A comparação dos dados deste QUADRO 8 com os dados dos QUADROS 6 e 7 revela que os itens lexicais co-extensivos à ordem do grupo nominal expostos neste último quadro são compostos por mais de dois componentes lexicais, enquanto que os itens lexicais dos dois quadros anteriores eram, na sua maioria, formados por dois componentes, sendo que o segundo componente era o responsável por gerar o item lexical transformado a partir do processo de mudança do ProtinLex:PTG. Todos os itens lexicais dispostos na primeira coluna do QUADRO 8 apresentam um componente lexical co-extensivo à função do Classificador, assim como aqueles especificados no QUADRO 6, e um componente lexical co-extensivo à função do Qualificador, tal como aqueles reproduzidos no QUADRO 7. Ao passo que os demais itens lexicais (ver QUADRO 6 e QUADRO 7) que passaram pelo processo de mudança do ProtinLex:PTG apresentam ora componentes lexicais co-extensivos à função do Classificador, ora componentes lexicais co-extensivos à função do Qualificador.

Nesse sentido, os itens lexicais co-extensivos à ordem do grupo nominal que apresentaram esta configuração de um componente lexical co-extensivo à função do Classificador e um outro componente lexical co-extensivo à função do Qualificador foram investigados sob duas perspectivas do processo de mudança desenvolvido para o

ProtinLex:PTG. Uma dessas perspectivas considerou o componente lexical co-extensivo à função do Qualificador, no caso, realizado por uma palavra da classe verbal: preposição + uma palavra da classe nominal: substantivo comum como o componente que poderia gerar o item lexical transformado mediante o processo de mudança. A outra perspectiva explorou o componente co-extensivo à função do Classificador como o componente que poderia gerar o item lexical transformado por meio do processo de mudança. No QUADRO 8, os itens lexicais que apresentam somente dois componentes são examinados sob essa segunda perspectiva, a qual considera os componentes lexicais que são co-extensivos à função do Classificador.

Comparando os resultados alcançados por essas duas perspectivas, fica evidente que, quando o componente lexical co-extensivo à função do Classificador gera o item lexical transformado, o resultado obtido é aquele nomeado, neste trabalho, como default, em que há a transformação desse componente em um item lexical transformado, encontrado no âmbito de uma classe de palavras diferente e com função distinta do componente que o gerou. Por exemplo, “básica” em “unidades básicas” gera, mediante o processo de mudança do ProtinLex:PTG, o item lexical transformado “básico”, assim como “funcional” em “letramento funcional” gera os itens lexicais transformados “funcionalidades” e “funções”. Por outro lado, quando a perspectiva que considera o componente lexical co-extensivo à função do Qualificador é explorada, itens lexicais transformados aparecem, nas orações resultantes do processo de mudança do ProtinLex:PTG, sob classes de palavras diferentes e funções distintas do componente lexical que o gerou, mas com a grafia inalterada. Por exemplo, o componente lexical “de saúde” em “unidades básicas de saúde” é co-extensivo à função do Qualificador e está no escopo da classe de palavras verbal: preposição + classe de palavras nominal: substantivo comum, gerando, por meio do processo de mudança deste ProtinLex, o item lexical transformado “saúde”, que é co-extensivo à função do Participante na ordem da oração e localizado no escopo da classe de palavras nominal: substantivo comum. Há, portanto, mudanças relacionadas às classes de palavras e às funções, mas a grafia de “saúde” permanece a mesma. Por essa razão, os resultados identificados para os itens lexicais co-extensivos à ordem do grupo nominal cujos resultados são análogos àqueles obtidos para os componentes lexicais co-extensivos à função do Qualificador (ver QUADRO 8) são nomeados, neste estudo, como desvios e estão em oposição aos resultados default.

A próxima subseção aborda os demais itens lexicais co-extensivos à ordem do grupo nominal que apresentaram esse tipo de resultado após o processo de mudança do ProtinLex:PTG e explora as razões pelas quais esse tipo de resultado pode acontecer.

5.2.2 Da aplicação do ProtinLex:PTG – desvio

O QUADRO 9 exibe os itens lexicais co-extensivos à ordem do grupo nominal que após o processo de mudança estabelecido para o protocolo de investigação lexical em questão apresentaram resultados entendidos, no presente trabalho, como sendo de desvio.

QUADRO 9 – Componentes lexicais co-extensivos à função do Qualificador do grupo nominal:
processo de mudança do ProtinLex:PTG – resultado de desvio

Itens lexicais com um dos componentes co-extensivos à função do Qualificador	Orações resultantes do processo de mudança + Itens lexicais transformados – desvio
práticas de autocuidado	práticas de autocuidado são tipos de práticas, [em oposição às práticas administrativas] essas práticas envolvem <u>autocuidado</u> .
atividades de autocuidado	atividades autocuidado são tipos de atividades, [em oposição às atividades de pintura] essas atividades envolvem <u>autocuidado</u> .
ações de autocuidado	ações de autocuidado são tipos de ações, [em oposição às ações do governo] essas ações envolvem <u>autocuidado</u> .
capacidade de autocuidado	capacidade de autocuidado é um tipo de capacidade, [em oposição a capacidade física] essa capacidade envolve <u>autocuidado</u> .
prática do autocuidado	prática do autocuidado é um tipo de prática, [em oposição a prática do exercício físico] essa prática envolve <u>autocuidado</u> .
promoção do autocuidado	promoção do autocuidado é um tipo de promoção, [em oposição a promoção de cidades inteligentes] essa promoção envolve <u>autocuidado</u> .
avaliação do autocuidado	avaliação do autocuidado é um tipo de avaliação, [em oposição a avaliação dos docentes] essa avaliação envolve <u>autocuidado</u> .
realização do autocuidado	realização do autocuidado é um tipo de realização, [em oposição a realização do negócio]

	essa realização envolve <u>autocuidado</u> .
gestão do autocuidado	gestão do autocuidado é um tipo de gestão, [em oposição a gestão do SUS]
	essa gestão envolve <u>autocuidado</u> .
autocuidado com os pés	autocuidado com os pés é um tipo de autocuidado, [em oposição a autocuidado com os cabelos]
	esse autocuidado envolve os <u>pés</u> .
educação para o autocuidado	educação para o autocuidado é um tipo de educação, [em oposição a educação para a cidadania]
	essa educação envolve <u>autocuidado</u> .
motivação para o autocuidado	motivação para o autocuidado é um tipo de motivação, [em oposição a motivação para a serenidade]
	essa motivação envolve o <u>autocuidado</u> .
profissionais de saúde	profissionais de saúde são tipos de profissionais, [em oposição aos profissionais de educação]
	esses profissionais envolvem <u>saúde</u> .
educação em saúde	educação em saúde é um tipo de educação, [em oposição a educação em CTS]
	essa educação envolve <u>saúde</u> .
letramento em saúde	letramento em saúde é um tipo de letramento, [em oposição a letramento em ciência e tecnologia]
	esse letramento envolve <u>saúde</u> .
promoção da saúde	promoção da saúde é um tipo de promoção, [em oposição a promoção da igualdade racial]
	essa promoção envolve <u>saúde</u> .
serviços de saúde	serviços de saúde são tipos de serviço, [em oposição a serviços de inteligência]
	esses serviços envolvem <u>saúde</u> .
equipe de saúde	equipe de saúde é um tipo de equipe, [em oposição a equipe da guarda municipal]
	essa equipe envolve <u>saúde</u> .
estado de saúde	estado de saúde é um tipo de estado, [em oposição a estado de conservação]

	esse estado envolve <u>saúde</u> .
unidades de saúde	unidades de saúde são tipos de unidades, [em oposição às unidades de conservação]
	essas unidades envolvem <u>saúde</u> .
condição de saúde	condição de saúde é um tipo de condição, [em oposição a condição de réu primário]
	essa condição envolve <u>saúde</u> .
ministério da saúde	ministério da saúde é um tipo de ministério, [em oposição ao ministério da agricultura]
	esse ministério envolve <u>saúde</u> .
educação para a saúde	educação para a saúde é um tipo de educação, [em oposição a educação para a sustentabilidade]
	essa educação envolve <u>saúde</u> .
usuários com diabetes mellitus	usuários com diabetes mellitus são tipos de usuários, [em oposição a usuários com autismo]
	esses usuários envolvem <u>diabetes mellitus</u> .
usuários com DM2	usuários com DM2 são tipos de usuários, [em oposição a usuários com câncer]
	esses usuários envolvem <u>DM2</u> .
uso de insulina	uso de insulina é um tipo de uso, [em oposição a uso de tecnologia]
	esse uso envolve <u>insulina</u> .
educação em saúde	educação em saúde é um tipo de educação, [em oposição a educação em direitos humanos]
	essa educação envolve <u>saúde</u> .
educação em dm	educação em dm é um tipo de educação, [em oposição a educação em ciências]
	essa educação envolve <u>dm</u> .
educação do autocuidado	educação do autocuidado é um tipo de educação, [em oposição a educação do campo]
	essa educação envolve <u>autocuidado</u> .
educação do paciente	educação do paciente é um tipo de educação, [em oposição a educação do povo]

	essa educação envolve <u>paciente</u> .
indivíduos com DM	indivíduos com DM são tipos de indivíduos, [em oposição a indivíduos com TDAH]
	esses indivíduos envolvem <u>DM</u> .
indivíduos com diabetes	indivíduos com diabetes são tipos de indivíduos, [em oposição a indivíduos com síndrome de down]
	esses indivíduos envolvem <u>diabetes</u> .
cuidados com os pés	cuidados com os pés são tipos de cuidado, [em oposição aos cuidados com os cabelos]
	esses cuidados envolvem os <u>pés</u> .
ações de promoção	ações de promoção são tipos de ações, [em oposição as ações de prevenção]
	essas ações envolvem <u>promoção</u> .
prática do autocuidado	prática do autocuidado é um tipo de prática, [em oposição à prática do crime]
	essa prática envolve <u>autocuidado</u> .
prática de autocuidado	prática de autocuidado é um tipo de prática, [em oposição à prática de crime]
	essa prática envolve <u>autocuidado</u> .

Fonte: Desenvolvido para fins deste trabalho.

Os itens lexicais co-extensivos à função do grupo nominal que passaram pelo processo de mudança criado para o ProtinLex:PTG e que apresentaram apenas resultados de desvio estão mencionados na primeira coluna do QUADRO 9. Já os itens lexicais transformados, cujo resultado é oriundo de um dos componentes lexicais dos itens da primeira coluna, estão descritos na segunda coluna do mesmo quadro. Diferentemente dos itens lexicais com resultado default, os itens lexicais transformados localizados no QUADRO 9 não apresentam mudança na grafia. Por exemplo, “educadores” que é o item lexical transformado para vários itens lexicais co-extensivos à ordem do grupo nominal que passaram pelo processo de mudança do ProtinLex:PTG, tais como: “programa educativo”, “processo educativo”, “ações educativas”, “intervenções educativas”, “práticas educativas”, “estratégias educacionais”, etc. tem a grafia modificada de “educativa(s)/educativo(s)” ou “educacionais” para “educadores” como um os resultados do processo de mudança do protocolo de investigação lexical em questão. Isso indica que a porção lexical, responsável por realizar o significado do item lexical,

a qual permanece inalterável entre o componente lexical e o item lexical transformado, e a porção gramatical, responsável pelas modificações na morfologia que acontecem entre o componente lexical e o item lexical transformado, não reproduzem o resultado padrão, i.e., default, aquele identificado para a maioria dos itens lexicais que passou pelo processo de mudança do ProtinLex:PTG, sobretudo a porção gramatical que deveria instanciar as variações necessárias para sinalizar a transformação entre o componente lexical e o item lexical transformado.

Entretanto, os itens lexicais transformados descritos na segunda coluna do QUADRO 9 são abordados, neste estudo, como um resultado de desvio, não como um resultado cujas características são completamente diferentes daquelas verificadas para os itens que apresentaram resultados default, porque mesmo que esses itens lexicais transformados permaneçam com a mesma grafia daqueles componentes lexicais que os originaram, as classes de palavras e as funções desses itens lexicais transformados são modificadas ao longo do processo de mudança definido para o ProtinLex:PTG. Quando os componentes lexicais são ainda parte dos itens lexicais co-extensivos à ordem do grupo nominal, eles são co-extensivos à função do Qualificador³⁶ e estão no escopo da classe de palavras verbal: preposição em conjunto com a classe de palavras nominal: substantivo. Após o processo de mudança do ProtinLex:PTG, esses componentes lexicais geram os itens lexicais transformados que são co-extensivos à função do Participante na ordem da oração e estão no escopo da classe de palavras nominal: substantivo, sem a companhia da classe de palavras verbal: preposição.

Nesse sentido, os itens lexicais transformados descritos na segunda coluna do QUADRO 9 podem ser enxergados como oriundos de um processo de mudança cujo resultado é de desvio, uma vez que houve mudanças relacionadas às classes de palavras e às funções, como acontece para os resultados default. No entanto não houve mudanças no que diz respeito à grafia, isto é, mudanças na porção lexical e/ou porção gramatical entre o componente lexical e o item lexical transformado, o que justifica o enquadramento desses resultados na classificação de desvio.

Uma das razões que pode explicar o motivo pelo qual resultados do tipo de desvio foram identificados para a presente pesquisa envolve o fato de que, ao longo do processo de mudança do ProtinLex:PTG, a transformação detectada entre o componente lexical, que gera o item lexical transformado, e o item lexical transformado propriamente dito envolve a exclusão

³⁶Tal co-extensividade também pode se dar com a função do Classificador, mas, como os componentes lexicais detalhados no QUADRO 9 são todos co-extensivos à função do Qualificador, essa outra possibilidade de co-extensividade não precisou ser discutida neste momento.

da co-extensividade à classe de palavras verbal: preposição, que estava incluída em conjunto com a classe de palavras nominal: substantivo na co-extensividade do componente lexical. Em outras palavras, no resultado default, há a modificação total da classe de palavras que envolve a co-extensividade do componente lexical para a classe de palavras que engloba o item lexical transformado, independentemente se esse componente lexical estava em co-extensividade com a função do Classificador ou do Qualificador³⁷, enquanto que, no resultado de desvio, essa mudança compreende a exclusão da co-extensividade com a classe de palavras verbal: preposição e a manutenção da co-extensividade com a classe de palavras nominal: substantivo.

Uma última razão que pode ser vista como uma justificativa para os resultados de desvio diz respeito ao acúmulo de significado verificado nos componentes lexicais que geraram, após o processo de mudança do ProtinLex:PTG, os itens lexicais transformados, detalhados no QUADRO 9. Esse acúmulo de significado abrange a hipótese de que as instâncias lexicais que realizam esses componentes lexicais devem exercer uma função, no português brasileiro, análoga às instâncias provenientes da classe de palavras nominal: substantivo próprio, as quais são utilizadas para nomear as entidades. “Pé”, por exemplo, que é o complemento lexical responsável por gerar o item lexical transformado em “autocuidado com os pés”, pode ser reconhecido como uma instância que tem um acúmulo de significado, visto que, nos textos que constituem o corpus utilizado neste trabalho, pés referem-se ao nome dado a uma parte do corpo. Contudo, para que essa hipótese pudesse ser comprovada de fato, um estudo que adotasse uma visão de cima, considerando de maneira mais intensa o estrato semântico, torna-se necessário, uma vez que a propensão de se trabalhar somente sob essa perspectiva não compreende a abordagem determinada para a presente pesquisa.

Para encerrar esta seção que trata dos resultados obtidos a partir do desenvolvimento e aplicação do ProtinLex:PTG, a subseção seguinte lida com um único item lexical selecionado por meio da lista de palavras-chave (ver QUADRO 2) cujos resultados para esse protocolo de investigação lexical não foram alcançados.

5.2.3 Da aplicação do ProtinLex:PTG – item lexical “relação”

Dentre os itens lexicais selecionados a partir da lista de palavras-chave, gerada pelo software concordanciador AntConc (ANTHONY, 2019) em conjunto com os textos que

³⁷É importante salientar que essa mudança de classe de palavras, após o processo de mudança do ProtinLex:PTG, entre os componentes lexicais co-extensivos à função do Qualificador e os itens lexicais transformados a partir de tais componentes lexicais foi verificada para a maior parte dos itens apresentados no QUADRO 7, não para a totalidade deles como acontece para os componentes lexicais co-extensivos à função do Classificador que estão descritos nos QUADROS 6 e 8.

compõem o corpus compilado para este estudo (ver 4.1 e QUADRO 2), somente o item lexical “relações” não apresenta resultados para a aplicação do ProtinLex:PTG, embora tenha sido definido como um item lexical co-extensivo à ordem do grupo nominal pelas investigações feitas para o ProtinLex:PTC.

Essa ausência de resultados para tal item lexical pode encontrar respaldo na forma como a metodologia do presente trabalho foi desenhada. À medida que os passos metodológicos elaborados para o desenvolvimento e a aplicação dos dois protocolos de investigação lexical enfocados nesta pesquisa tendem a priorizar os mecanismos e/ou momentos específicos da produção de significado em que os itens lexicais podem ser relacionados à estrutura gramatical, ou seja, neste estudo, os itens lexicais são abordados diante das possíveis relações estabelecidas entre si, considerando os itens lexicais que aparecem antes e depois de um item lexical específico, bem como a concepção de itens lexicais formados por mais de um componente lexical; e a relação disso com as classes de palavras e funções na estrutura e no contexto de situação dos textos que constituem o corpus utilizado.

Diante deste enfoque, o item lexical “relações” é caracterizado como um item lexical co-extensivo à ordem do grupo nominal por meio da análise dos resultados obtidos para o ProtinLex:PTC, mas não pode ser caracterizado como um item lexical para o ProtinLex:PTG, por não ter apresentado resultados – default e nem de desvio – para o processo de mudança desenvolvido para o protocolo de investigação lexical em questão. Isso significa dizer que se o item “relações” fosse examinado através de uma ótica que o considerasse de maneira isolada, ele poderia não ter sido caracterizado como um item lexical, tal como o foi pela investigação desenvolvida no primeiro protocolo, o ProtinLex:PTC.

Tendo finalizado as apresentações e discussões acerca dos resultados obtidos para a presente pesquisa, o próximo capítulo destina-se a explorar as considerações finais sobre o estudo realizado nesta tese.

6. CONCLUSÃO

Motivada por questões que as visões modular e contígua levantam na interpretação do léxico, a presente pesquisa ocupou-se da investigação acerca de como o significado de um item lexical não está isolado do restante do sistema, nem restrito apenas ao estrato da gramática e/ou à ordem da palavra e nem afastado do contexto e cotexto onde é instanciado. Tendo em mente a necessidade de uma maior compreensão de como o léxico, na posição de uma das dimensões do sistema linguístico, determina a sua própria organização interna em relação ao restante do sistema linguístico do português brasileiro; e emprega os mecanismos que possibilitam a investigação sobre a propriedade de co-extensividade, sobretudo em relação ao contexto e ao estrato da gramática, dizem respeito aos principais objetos de pesquisa explorados nesta tese.

Tendo como base a abordagem sistêmico-funcional que concebe a organização da linguagem através de dimensões, sendo uma delas a estratificação, que pelo princípio da realização, configura a interdependência entre os estratos linguísticos – contexto, semântica, gramática, fonologia e fonética, esta pesquisa buscou complementar o entendimento acerca do léxico e dos itens lexicais a partir de uma visão que leva em conta o texto em seu contexto. Isso significa dizer que os resultados obtidos apresentam uma base em comum, isto é, um *corpus* compilado especificamente para o presente estudo, o qual é composto de artigos acadêmicos sobre “autocuidado em diabetes mellitus”, escritos em português brasileiro e extraídos de periódicos da área das Ciências da Saúde. Ainda no que diz respeito a essa base, um *corpus* nomeado de *corpus* de suporte se fez essencial para assegurar, por meio de análises contrastivas, a homogeneidade das características identificadas para os itens lexicais que compõem o *corpus* compilado para este trabalho, como também aquelas identificadas para os itens lexicais retirados do *corpus* de suporte.

Nesta pesquisa, essas análises contrastivas foram idealizadas através do desenvolvimento e da aplicação de dois protocolos de investigação lexical, ambos referenciados com a sigla ProtinLex (ProtinLexis, no plural). Um dos protocolos lidou com o contexto, e o outro tratou do estrato da gramática.

O primeiro protocolo, chamado de Protocolo de Investigação Lexical: Processo de Transformação do Contexto – ProtinLex:PTC, abordou: (i) A coocorrência de itens lexicais, que quando aparecem em sequência e repetidas vezes no *corpus*, constituem um agrupamento (*cluster/n-gram*) cujo significado costuma ser divergente daquele verificado para os itens que funcionam como seus componentes lexicais, quando estes são analisados de maneira isolada no

cotexto³⁸ em que aparecem nos textos dos corpora. (ii) A identificação da propriedade de co-extensividade entre um item lexical formado por mais de um componente lexical e a ordem do grupo nominal, por meio da anotação do sistema de TRANSITIVIDADE. Essa anotação apontou que itens lexicais desse tipo podem ser entendidos como co-extensivos aos elementos permanentes no texto, os quais são prototipicamente realizados pela ordem do grupo nominal e pelas funções do Participante e da Circunstância, estas localizadas no escopo do sistema de TRANSITIVIDADE. (iii) A influência das variáveis do contexto de situação – campo, sintonia e modo – na emergência de itens lexicais formados por mais de um componente lexical cuja co-extensividade se dá com a ordem do grupo nominal, uma vez que, mediante os resultados identificados após o processo de mudança estabelecido para o ProtinLex:PTC, observou-se que a formação de itens lexicais desse tipo está relacionada à maneira pela qual os textos, onde esses itens aparecem, foram construídos, sob quais configurações de domínio experiencial, processos sociosemióticos – campo – de papel institucional, papel social, papel discursivo, papel sociométrico, distância social e valoração – sintonia – e de meio, canal, modo retórico, sistemas semióticos participantes, orientação e troca de turnos – modo – esses textos encontram-se.

O segundo protocolo, nomeado de Protocolo de Investigação Lexical: Processo de Transformação da Gramática – ProtinLex:PTG, explorou: (i) A maneira pela qual itens lexicais são formados no português brasileiro mediante a capacidade que alguns itens lexicais apresentam de gerar outros itens, rotulados, nesta pesquisa, de itens lexicais transformados. Esses itens estão, na maior parte das vezes, localizados em classes de palavras diferentes e com a co-extensividade distinta daquela que o item que o gerou está localizado e opera na ordem da oração ou grupo. Por exemplo, o item lexical, que origina o item lexical transformado, se mostrou co-extensivo às funções do Classificador e/ou do Qualificador, ambas localizadas no âmbito das funções inerentes ao grupo nominal (cf. FIGUEREDO, 2007), enquanto que o item lexical transformado propriamente dito mostrou-se co-extensivo aos elementos permanentes, i.e. Participantes e Circunstâncias, nas orações desenvolvidas para o processo de mudança do ProtinLex:PTG. (ii) Através da análise contrastiva dos itens lexicais, que originaram os itens lexicais transformados, e os itens lexicais transformados de fato, o exame das porções que constituem os itens lexicais no português brasileiro foi feito, identificando que os itens lexicais nesse sistema linguístico apresentam duas porções, uma lexical que carrega o significado do item lexical e permanece inalterada entre o item lexical, que gera o item lexical transformado, e o item lexical transformado propriamente dito, e uma gramatical que trata da parte

³⁸Cotexto compreende todos os itens lexicais e/ou gramaticais que aparecem ao redor, seja antes seja depois do item lexical em questão.

morfológica do item lexical, sendo responsável por realizar as modificações necessárias entre o item lexical, que origina o item lexical transformado, e o item lexical transformado propriamente dito.

De maneira mais detalhada, no âmbito da aplicação dos protocolos de investigação lexical, dois processos de mudança, um para o ProtinLex:PTC e outro para o ProtinLex:PTG funcionaram como o mecanismo principal na busca pelas características intrínsecas ao item lexical no português brasileiro. Foi por meio da adoção desses processos de mudança que uma abordagem mais funcionalista se tornou viável, como, por exemplo, a anotação linguística do sistema de TRANSITIVIDADE para verificar a co-extensividade do item lexical formado por mais de um componente com a ordem do grupo nominal, bem como o desenvolvimento de orações relacionais e/ou orações materiais, nas quais relações COESIVAS LEXICAIS de hiponímia, meronímia, antonímia e/ou colocação pudessem ser estabelecidas, possibilitou a investigação acerca: (i) da influência das variáveis contextuais na concepção de itens lexicais com mais de um componente lexical, (ii) da capacidade de transformação de itens lexicais e (iii) do funcionamento das porções que formam os itens lexicais no PB.

Considerando que a análise dos resultados gerados pelo processo de mudança correspondente ao Protocolo de Investigação Lexical: Processo de Transformação do Contexto, desenvolvido, nesta pesquisa, como forma de investigar as características do item lexical no sistema linguístico do português brasileiro, conseguiu apontar que itens lexicais podem operar em conjunto para conceber novos itens lexicais, os quais parecem ser intermediados pelas variáveis do contexto de situação para resolver uma demanda por um significado antes inexistente no escopo de um dado sistema linguístico, no caso o português brasileiro. Esse resultado revela a co-extensividade entre esse item lexical que opera em conjunto com outros itens lexicais, ou seja, itens lexicais formados por mais de um componente lexical com a ordem do grupo nominal e, como consequência, a co-extensividade desse tipo de item lexical com o contexto, uma vez que esses itens lexicais parecem ter a capacidade de expandir a gama de significados de uma língua, por meio da realização frequente de itens lexicais com essa formação nos textos daquele sistema.

A análise dos resultados obtidos pelo processo de mudança equivalente ao Protocolo de Investigação Lexical: Processo de Transformação da Gramática também desenvolvido no âmbito desta pesquisa para explorar as características do item lexical no sistema linguístico do português brasileiro conseguiu evidenciar que o léxico opera em conjunto com a gramática na criação de itens lexicais, os quais, por sua vez, apresentam um porção lexical e outra gramatical e, em função disso, conseguem produzir itens lexicais transformados,

localizados, muitas vezes, sob o escopo de classes de palavras distintas. Esse resultado indica, portanto, que léxico e gramática podem ser explicados como dois polos que trabalham juntos para a construção de novos significados, os quais serão instanciados no sistema linguístico através de novos itens lexicais, i.e. itens lexicais transformados.

Nesse sentido, os resultados deste estudo sugerem que o contexto no português brasileiro assim como a gramática apresentam um papel essencial na emergência de itens lexicais na língua em função das características de co-extensividade com a ordem do grupo nominal e de transformação do item lexical, ambas identificadas neste trabalho. Assim, a presente pesquisa oferece contribuições para a abordagem que enfoca a complementariedade entre léxico e gramática, a qual pode ser modelada pela co-extensividade. Esta pesquisa demonstra ainda que a co-extensividade acerca do item lexical se expande até o contexto de cultura, extrapolando o estrato gramatical e, como consequência, contesta as visões de que o léxico de uma língua não pode estar relacionado às outras dimensões da linguagem e/ou deve ser enxergado como um módulo isolado.

Para além disso, este trabalho acrescenta contribuições no que diz respeito à colocação/colocados (*collocation/collocates*), uma vez que diferentemente do que salientam Firth (1957) que trata a colocação como as palavras que coocorrem com uma dada frequência ao redor de uma palavra em específico, Halliday (1961, p. 61) que, influenciado por Firth, define a colocação como “a associação sintagmática dos itens lexicais”, sendo que tal associação pode ser quantificada pela probabilidade de ocorrência desses itens lexicais, os quais, por sua vez, apresentam uma gama de colocações que são mais prováveis de ocorrer e que são variáveis à medida que novos itens são adicionados à gama já existente, e Sinclair (2004, p. 19) que, também influenciado por Firth, conceitua a colocação como uma “coocorrência entre palavras, a qual deve ocorrer com uma certa frequência”, os resultados da presente pesquisa evidenciam que os colocados podem ser concebidos como itens lexicais que coocorrem com outros itens lexicais, nos textos de uma língua, com tanta frequência que esses itens são capazes de serem instanciados em sequência, construindo um significado antes inexistente e cuja seleção dos itens para concepção do colocado varia conforme o contexto de situação em questão. Isso indica que a formação de colocados não pode ser vista de maneira aleatória, como se essa formação pudesse se dar entre quaisquer itens, o que de acordo com as análises feitas neste trabalho, sobretudo para o ProtinLex:PTC, não pode ser considerado pertinente, visto que as variáveis contextuais de uma dada situação influenciam nessa formação de colocados. Por essa razão, ora o item lexical “educação”, por exemplo, pode formar um

colocado com “alimentar” em “educação alimentar”, ora o mesmo item lexical “educação” pode formar um colocado com “ambiental” em “educação ambiental”.

A abordagem utilizada nesta pesquisa de se analisar o léxico/item lexical no português brasileiro por uma perspectiva colocacional, em que os itens lexicais são formados por mais de um componente lexical, pode ter contribuído para reforçar a visão de Halliday (1961) que já atestava que a organização lexical estava relacionada à probabilidade de ocorrência de um dado item lexical em um determinado tipo de texto, o que segundo esse autor poderia ser corroborado por meio da investigação acerca dos colocados e conjuntos lexicais (ver 3.1). Nessa mesma publicação Halliday afirma que os princípios da escala de ordens e da abstração não são relevantes para a abordagem lexical. No entanto, os resultados do presente estudo apontam que a relação entre item lexical e escala de ordens da gramática – oração, grupo, palavra e morfema – é importante no que se refere aos itens lexicais formados por mais de um componente lexical cuja co-extensividade com a ordem do grupo sobretudo o grupo nominal pode ser identificada pela aplicação dos ProtinLexis desenvolvidos neste trabalho.

Uma última contribuição da presente pesquisa envolve a propriedade de transformação identificada para alguns itens lexicais que geraram itens lexicais transformados no âmbito do processo de mudança estabelecido para o ProtinLex:PTG. Os resultados desse protocolo de investigação lexical relevaram que a maior parte dos itens lexicais selecionados para serem utilizados no desenvolvimento e aplicação dos ProtinLexis gerou itens lexicais transformados, independentemente se esses itens lexicais transformados estavam no escopo do resultado default ou do resultado de desvio (ver 5.2.1 e 5.2.2). Entretanto, a análise desses resultados aponta que essa transformação está associada às classes de palavra, ou seja, parece haver um número limite de itens lexicais transformados a partir de cada componente lexical (item lexical) presente em um item lexical co-extensivo ao grupo nominal. Em outras palavras, embora as concepções acerca do léxico o classifiquem como uma parte da linguagem que consegue evoluir de forma mais rápida e, como consequência disso, seja capaz de apresentar itens lexicais que são mais específicos e delicados em comparação aos itens gramaticais, que são mais amplos e gerais, (MATTHIESSEN, 1995; MARTIN e ROSE, 2007, p. 63), essa evolução se mostra condicionada às classes de palavras, o que indica que cada item lexical que emerge na língua encontra-se no escopo de uma classe de palavras em particular. Por exemplo, o item lexical “saúde” está classe de palavras dos substantivos: comuns e gera o item lexical transformado “saudável”, que está na classe de palavras dos substantivos: adjetivos. Para além das variações relacionadas ao gênero (feminino/masculino), número (singular/plural) e pessoa (primeira, segunda e terceira pessoa), sendo que esta última não se aplica a todos os itens

lexicais, outros itens lexicais transformados que estejam no âmbito da classe de palavras dos substantivos: adjetivos não tendem a surgir na língua. Isso sugere que os itens lexicais no que tange o português brasileiro parecem evoluir de maneira mais rápida que os itens gramaticais, mas essa evolução parece ser feita deliberadamente pelo sistema linguístico.

Embora seja uma pesquisa inicial sobre a caracterização do léxico como um recurso linguístico para a produção de significado no escopo do arcabouço teórico da linguística sistêmico-funcional, o presente estudo aponta que o item lexical no português brasileiro apresenta a propriedade de co-extensividade, identificada primeiramente por Halliday, McIntosh e Strevens (1964) para a língua inglesa, com o estrato gramatical, no que concerne a escala de ordens, principalmente a ordem do grupo (nominal), e as funções do Participante e da Circunstância do sistema de TRANSITIVIDADE. Para além da co-extensividade entre o item lexical e o contexto, à medida que os itens lexicais no português brasileiro podem ser emergidos na língua de acordo com demandas de contextos de situação específicos, no caso deste trabalho, o autocuidado em Diabetes Mellitus. Esses fatos caracterizam as relações existentes entre o item lexical (= dimensões linguísticas do léxico) e o estrato gramatical e o contexto, sendo ambos funcionando como organizações sistêmicas da língua. Esta pesquisa possibilita ainda, no futuro, a investigação acerca da co-extensividade do léxico (item lexical) com o estrato da semântica e a ampliação para outras dimensões da linguagem.

Por fim, no que compreende os Estudos da Tradução, os resultados obtidos no presente estudo podem ser utilizados no aprimoramento de tradutores automáticos e/ou semiautomáticos, uma vez que o entendimento das variáveis do contexto de situação se mostrou pertinente na emergência de itens lexicais formados por mais de um componente lexical. Em outras palavras, se os parâmetros das variáveis contextuais de um conjunto específico de textos de uma dada língua, no caso o português brasileiro, conseguem ser aprendidos pelo(s) tradutor(es) automático(s), espera-se que as escolhas dos itens lexicais numa tradução para o português brasileiro, por exemplo, seja mais precisa de acordo com esses parâmetros contextuais; e que o tradutor automático consiga propor soluções tradutórias que considerem a co-extensividade entre itens lexicais e a ordem do grupo (nominal), deixando de sugerir opções tradutórias para cada componente lexical de forma isolada.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ANTHONY, L. AntConc Homepage. **Laurence Anthony Website**, Tokyo, n. Version 3.5.8, 2019. Disponível em: <<https://www.laurenceanthony.net/software/antconc/>>. Acesso em: 2019.
- BASÍLIO, M. Produtividade, Função e Fronteiras Lexicais. In: NEVES, M. H. D. M. **Descrição do Português II**. Araraquara: Série Encontros, 1991.
- BORBA, F. D. S. **Introdução aos Estudos Linguísticos**. 13^a. ed. Campinas: Pontes Editores, 2003.
- CASTILHO, A. T. D. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2010.
- CHOMSKY, N. **Aspects of the Theory of Syntax**. 1a. ed. Cambridge, Massachusetts: The M.I.T. Press, 1965.
- DRAGSTED, B. Segmentation in Translation and Translation Memory Systems An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process. **Tese de doutorado Copenhagen Business School**, Copenhagen, p. 1-305, 2004.
- DRAGSTED, B. Segmentation in Translation Differences across levels of expertise and difficulty. **Target**, 2005. 49-70.
- FERREGUETTI, K. A FRASE PREPOSICIONAL COM FUNÇÃO DE QUALIFICADOR NO GRUPO NOMINAL: UM ESTUDO DE EQUIVALENTES TEXTUAIS NO PAR LINGUÍSTICO INGLÊS E PORTUGUÊS BRASILEIRO. **Tese de Doutorado. Faculdade de Letras. Universidade Federal de Minas Gerais**, Belo Horizonte, 2018.
- FIGUEREDO, G. Uma descrição sistêmico-funcional do grupo nominal em português brasileiro. **Dissertação de Mestrado. Faculdade de Letras. Universidade Federal de Minas Gerais**, Belo Horizonte, 2007.
- FIGUEREDO, G. Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues. **Tese de Doutorado. Faculdade de Letras. Universidade Federal de Minas Gerais**, Belo Horizonte, 2011.
- FIGUEREDO, G. P. et al. O léxico como um recurso linguístico para a produção de significado no texto: um estudo de caso com protocolos de investigação. **Estudos da Língua(gem)**, Vitória da Conquista, v. 17, n. 3, p. 37-59, 2019.
- FIGUEREDO, G. P.; PAGANO, A. S.; FERREGUETTI, K. Os sistemas textuais de focalização na organização funcional da gramática do Português Brasileiro. **D.E.L.T.A.**, 2014. 309-352.
- FIRTH, J. R. **Papers in Linguistics 1934 - 1951**. London: Oxford University Press, 1957.
- FIRTH, J. R. **Studies in linguistic analysis**. Oxford: Basil Blackwell Oxford, 1962.
- FONTAINE, L. Lexis as most local context: towards an SFL approach to lexicology. **Functional Linguistics**, p. 4-17, 2017a.
- FONTAINE, L. On prepositions and particles: a case for lexical representation in systemic functional linguistics. **Word**, v. 63, n. 2, p. 115-135, 2017b.
- FONTES, C. S. O processo de (re)tradução: um estudo quali-quantitativo baseado em um experimento no par inglês-português. **Tese de Doutorado apresentada ao Programa de Pós-graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais**, Belo Horizonte, 2020. 1-293.
- GABRIELATOS, C. Keyness Analysis: nature, metrics and techniques. In: TAYLOR, C.; MARCHI, A. **Corpus Approaches to Discourse: A critical review**. 1a. ed. New York: Routledge, 2018. Cap. 11, p. 225-258.

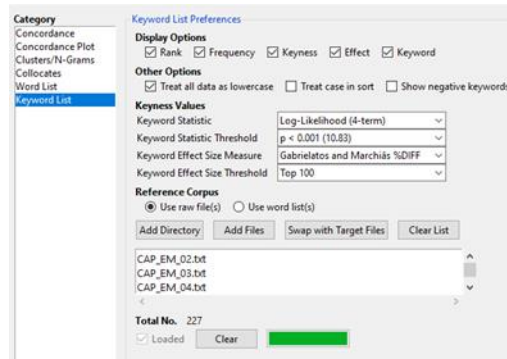
- HALLIDAY, M. A. K. Categories of the Theory of Grammar. In: HALLIDAY, M. A. K. **Collected Works of M.A.K. Halliday**. London-New York: Continuum, v. Volume 1: On Grammar, 1961. Cap. 1, p. 37-88.
- HALLIDAY, M. A. K. Class in relation to the axes of chain and choice in language. **Linguistics**, n. 2, p. 5-15, 1963.
- HALLIDAY, M. A. K. Lexis as a Linguistic Level. In: BAZELL, C. E., et al. **In Memory of J. R. Firth**. [S.l.]: Longman, 1966. p. 148-162.
- HALLIDAY, M. A. K. **Language as social semiotic The social interpretation of language and meaning**. 1a. ed. London: Edward Arnold, 1978.
- HALLIDAY, M. A. K. The notion of "context" in language education. In: WEBSTER, J. **Language and Education, Collected Works of MAK Halliday**. London: Continuum, v. 9, 1991. p. 269-290.
- HALLIDAY, M. A. K. **On Grammar**. London: Continuum, v. I, 2002.
- HALLIDAY, M. A. K. **Complementarities in language**. Beijing: The Commercial Press, 2008.
- HALLIDAY, M. A. K.; HASSAN, R. **Cohesion in English**. London and New York: Longman, 1976.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. **Construing experience as meaning: a language based approach to cognition**. London: Cassell, 1999.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **An Introduction to Functional Grammar**. 3. ed. London: Routledge, 2014.
- HALLIDAY, M. A. K.; MCINTOSH, A.; STREVEENS, P. **The linguist sciences and language teaching**. London: Longmans, 1964.
- HAO, J. Construing biology: An Ideational Perspective. **A thesis submitted at Department of Linguistics University of Sydney**, Sydney, March 2015.
- HAO, J. **Analysing scientific discourse from a systemic functional linguistic perspective. A framework for exploring knowledge-building in biology**. New York: Taylor & Francis, 2020.
- HASAN, R. Lending and borrowing: from grammar to lexis. **Beiträge zur Phonetik und Linguistik**, v. 48, p. 56-67, 1985.
- HASAN, R. The grammarian's dream: lexis as most delicate grammar. In: HALLIDAY, M. A. K.; FAWCETT, R. P. **New Developments in Systemic Linguistics**. London/New York: Frances Pinter, v. I: Theory and Description, 1987.
- HJELMSLEV, L. **Prolegomena to a Theory of Language**. Madison: University of Wisconsin Press, 1963.
- HOPPER, P. J.; TRAUGOTT, J. **Grammaticalization**. 1st. ed. Cambridge: Cambridge University Press, 1993.
- KURYŁOWICZ, J. The Evolution of Grammatical Categories. In: KURYŁOWICZ, J. **Esquisses linguistiques II**. [S.l.]: [s.n.], 1975. p. 38-54.
- MARTIN, J. R. **English Text System and Structure**. Philadelphia: John Benjamins Publishing Company, 1992.
- MARTIN, J. R. Systemic functional grammar: a next step into the theory - axial relations. **Higher Education Press**, Beijing, 2013.
- MARTIN, J. R.; ROSE, D. **Genre Relations Mapping Culture**. 1a. ed. London: Equinox Publishing Ltd, 2008.
- MARTIN, J. R.; WHITE, P. R. R. **The language of evaluation: appraisal in English**. 1st. ed. New York: PALGRAVE MACMILLAN, 2005.

- MARTIN, J.; ROSE, D. **Working with discourse: meaning beyond the clause**. 2. ed. London: Continuum, 2007.
- MATTHIESSEN, C. M. I. M. Lexico(grammaral) choice in text generation. In: PARIS, C. L.; SWARTOUT, W. R.; MANN, W. C. **Natural Language Generation in Artificial Intelligence and Computational Linguistics**. [S.l.]: Kluwer Academic Publishers, 1990. p. 249-292.
- MATTHIESSEN, C. **Lexicogrammatical Cartography: English Systems**. Tokyo: International Language Sciences Publishers, 1995.
- MATTHIESSEN, C.; TERUYA, K.; LAM, M. **Key terms in systemic functional linguistics**. London and New York: Continuum, 2010.
- MOON, R. What can a corpus tell us about lexis? In: O'KEEFFE, A.; MCCARTHY, M. **The Routledge Handbook of Corpus Linguistics**. Abingdon: Routledge, 2010. p. 197-211.
- NEVES, M. H. D. M. **A Gramática Funcional**. 1^a. ed. São Paulo: Livraria Martins Fontes Editora Ltda, 1997.
- PEARSON, J. **Terms in Context**. Philadelphia/Amsterdam: John Benjamins, v. I, 1998.
- QUAH, C. K. **Translation and Technology**. 1a. ed. New York: Palgrave Macmillan, 2006.
- RODRIGUES, J. S. N.; FIGUEREDO, G. P.; OLIVEIRA, F. S. A CO-EXTENSIVIDADE E ORGANIZAÇÃO LEXICAL NO PORTUGUÊS BRASILEIRO: UMA INTRODUÇÃO DESCRITIVA A PARTIR DE UMA ABORDAGEM SISTÊMICO-FUNCIONAL. **Alfa: Revista de Linguística**, no prelo.
- ROSE, D. A systemic functional approach to language evolution. **Cambridge Archaeological Journal**, v. 16, n. 1, p. 73-96, 2006.
- ROSE, D. **Reading to learn: accelerating learning and closing the gap**. Sydney: Reading to Learn, 2019.
- ROSE, D.; MARTIN, J. R. **Learning to write, reading to learn: genre, knowledge and pedagogy in the Sydney school**. [S.l.]: Equinox Publishing, 2012.
- SARDINHA, T. B. **Linguística de Corpus**. Barueri: Manole, 2004.
- SCOTT, M. Problems in investigating keyness, or clearing the undergrowth and marking out trails. In: BONDI, M.; SCOTT, M. **Studies in Corpus Linguistics (SCL)**. Amsterdam / Philadelphia: John Benjamins Publishing Company, v. 41 Keyness in Texts, 2010. p. 43-56.
- SINCLAIR, J. **Trust the text: Language, corpus, and discourse**. London: Routledge, 2004.
- TAVERNIERS, M. Hjelmslev's semiotics model of language: An exegesis. **Semiotica**, 2008. 367-394.
- TUCKER, G. H. **The Lexicogrammar of Adjectives A systemic functional approach to lexis**. Londres: Cassell, 1998.
- VIANA, V.; TAGNÍN, S. **Corpora no Ensino de Línguas Estrangeiras**. São Paulo: HUB Editorial, 2010.

8. ANEXOS

8.1 Desenvolvimento e aplicação dos ProtinLexis: seleção dos itens a partir da lista de palavras-chave com p-valor = 0.001

8.1.1 “%DIFF”



Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	362	+ 1024.4	115178.4869	mellitus
5	329	+ 942.94	209439.3491	relação
6	327	+ 937.2	208165.5537	usuários
7	288	+ 825.31	183326.5427	educação
8	256	+ 733.53	162945.8157	complicações
9	225	+ 644.63	143201.9865	indivíduos
10	197	+ 564.36	125368.8504	pés
11	194	+ 555.76	123458.1572	ações
12	194	+ 555.76	123458.1572	prática
13	185	+ 529.96	117726.0778	enfermagem
14	184	+ 527.09	117089.1801	diagnóstico
15	183	+ 524.22	116452.2823	adesão
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	51807.164	insulina
22	152	+ 435.37	96708.4531	variáveis
23	149	+ 426.78	94797.7599	nível
24	149	+ 426.78	94797.7599	usuário
25	148	+ 423.91	94160.8622	população
26	139	+ 398.12	88428.7828	indivíduo
27	137	+ 392.39	87154.9873	avaliação
28	135	+ 386.66	85881.1919	alimentação
29	133	+ 380.93	84607.3964	análise
30	132	+ 378.06	83970.4987	diabéticos
31	123	+ 352.27	78238.4193	condição
32	116	+ 321.25	36840.0676	educativas
33	116	+ 322.22	73780.1352	importância
34	114	+ 326.49	72506.3398	informações
35	112	+ 320.76	71232.5444	intervenções
36	112	+ 320.76	71232.5444	três
37	111	+ 317.89	70595.6467	idosos
38	111	+ 317.89	70595.6467	realização
39	110	+ 315.03	69958.7489	glicemia
40	110	+ 315.03	69958.7489	mudança

Ordenado por Keyness – Parte I e Parte II

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	362	+ 1024.4	115178.4869	mellitus
5	329	+ 942.94	209439.3491	relação
6	327	+ 937.2	208165.5537	usuários
7	288	+ 825.31	183326.5427	educação
8	256	+ 733.53	162945.8157	complicações
9	225	+ 644.63	143201.9865	indivíduos
10	197	+ 564.36	125368.8504	pés
11	194	+ 555.76	123458.1572	ações
12	194	+ 555.76	123458.1572	prática
13	185	+ 529.96	117726.0778	enfermagem
14	184	+ 527.09	117089.1801	diagnóstico
15	183	+ 524.22	116452.2823	adesão
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Search Term Words Case Regex Hit Location Search Only 0 Reference Corpus Loaded

Sort by Invert Order Sort by Keyness

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	51807.164	insulina
22	152	+ 435.37	96708.4531	variáveis
23	149	+ 426.78	94797.7599	nível
24	149	+ 426.78	94797.7599	usuário
25	148	+ 423.91	94160.8622	população
26	139	+ 398.12	88428.7828	indivíduo
27	137	+ 392.39	87154.9873	avaliação
28	135	+ 386.66	85881.1919	alimentação
29	133	+ 380.93	84607.3964	análise
30	132	+ 378.06	83970.4987	diabéticos
31	123	+ 352.27	78238.4193	condição
32	116	+ 332.22	73780.1352	importância
33	114	+ 326.49	72506.3398	informações
34	116	+ 321.25	36840.0676	educativas
35	112	+ 320.76	71232.5444	intervenção
36	112	+ 320.76	71232.5444	três
37	111	+ 317.89	70595.6467	idosos
38	111	+ 317.89	70595.6467	realização
39	110	+ 315.03	69958.7489	glicemia
40	110	+ 315.03	69958.7489	mudança

Search Term Words Case Regex Hit Location Search Only 0 Reference Corpus Loaded

Sort by Invert Order Sort by Keyness

Ordenado por Effect – Parte I e Parte II

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	329	+ 942.94	209439.3491	relação
5	327	+ 937.2	208165.5537	usuários
6	288	+ 825.31	183326.5427	educação
7	256	+ 733.53	162945.8157	complicações
8	225	+ 644.63	143201.9865	indivíduos
9	197	+ 564.36	125368.8504	pés
10	194	+ 555.76	123458.1572	ações
11	194	+ 555.76	123458.1572	prática
12	185	+ 529.96	117726.0778	enfermagem
13	184	+ 527.09	117089.1801	diagnóstico
14	183	+ 524.22	116452.2823	adesão
15	362	+ 1024.4	115178.4869	mellitus
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Search Term Words Case Regex Hit Location Search Only 0 Reference Corpus Loaded

Sort by Invert Order Sort by Effect

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	152	+ 435.37	96708.4531	variáveis
22	149	+ 426.78	94797.7599	nível
23	149	+ 426.78	94797.7599	usuário
24	148	+ 423.91	94160.8622	população
25	139	+ 398.12	88428.7828	indivíduo
26	137	+ 392.39	87154.9873	avaliação
27	135	+ 386.66	85881.1919	alimentação
28	133	+ 380.93	84607.3964	análise
29	132	+ 378.06	83970.4987	diabéticos
30	123	+ 352.27	78238.4193	condição
31	116	+ 332.22	73780.1352	importância
32	114	+ 326.49	72506.3398	informações
33	112	+ 320.76	71232.5444	intervenção
34	112	+ 320.76	71232.5444	três
35	111	+ 317.89	70595.6467	idosos
36	111	+ 317.89	70595.6467	realização
37	110	+ 315.03	69958.7489	glicemia
38	110	+ 315.03	69958.7489	mudança
39	109	+ 312.16	69321.8512	após
40	104	+ 297.84	66137.3626	número

Search Term Words Case Regex Hit Location Search Only 0 Reference Corpus Loaded

Sort by Invert Order Sort by Effect

8.1.2 “Odds-ratio”

Tool Preferences

Category: Keyword List

Display Options: Rank Frequency Keyness Effect Keyword

Other Options: Treat all data as lowercase Treat case in sort Show negative keywords

Keyword Values: Keyword Statistic: Log-Likelihood (4-term); Keyword Statistic Threshold: p < 0.001 (10.83); Keyword Effect Size Measure: Odds-ratio; Keyword Effect Size Threshold: Top 100

Reference Corpus: Use raw file(s) Use word list(s)

Add Directory Add Files Swap with Target Files Clear List

CAP_EM_02.txt
CAP_EM_03.txt
CAP_EM_04.txt

Total No. 227 Loaded Clear

Apply Cancel

Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	362	+ 1024.4	1158.3183	mellitus
5	329	+ 942.94	2104.5386	relação
6	327	+ 937.2	2091.6895	usuários
7	288	+ 825.31	1841.2694	educação
8	256	+ 733.53	1635.9898	complicações
9	225	+ 644.63	1437.2912	indivíduos
10	197	+ 564.36	1257.9617	pés
11	194	+ 555.76	1238.7557	ações
12	194	+ 555.76	1238.7557	prática
13	185	+ 529.96	1181.1469	enfermagem
14	184	+ 527.09	1174.7467	diagnóstico
15	183	+ 524.22	1168.3468	adesão
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	520.1894	insulina
22	152	+ 435.37	970.032	variáveis
23	149	+ 426.78	950.8489	nível
24	149	+ 426.78	950.8489	usuário
25	148	+ 423.91	944.4548	população
26	139	+ 398.12	886.9161	indivíduo
27	137	+ 392.39	874.1316	avaliação
28	135	+ 386.66	861.3478	alimentação
29	133	+ 380.93	848.5646	análise
30	132	+ 378.06	842.1733	diabéticos
31	123	+ 352.27	784.6589	condição
32	116	+ 321.25	369.966	educativas
33	116	+ 332.22	739.935	importância
34	114	+ 326.49	727.1583	informações
35	112	+ 320.76	714.3822	intervenções
36	112	+ 320.76	714.3822	três
37	111	+ 317.89	707.9944	idosos
38	111	+ 317.89	707.9944	realização
39	110	+ 315.03	701.6068	glicemia
40	110	+ 315.03	701.6068	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	362	+ 1024.4	1158.3183	mellitus
5	329	+ 942.94	2104.5386	relação
6	327	+ 937.2	2091.6895	usuários
7	288	+ 825.31	1841.2694	educação
8	256	+ 733.53	1635.9898	complicações
9	225	+ 644.63	1437.2912	indivíduos
10	197	+ 564.36	1257.9617	pés
11	194	+ 555.76	1238.7557	ações
12	194	+ 555.76	1238.7557	prática
13	185	+ 529.96	1181.1469	enfermagem
14	184	+ 527.09	1174.7467	diagnóstico
15	183	+ 524.22	1168.3468	adesão
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	520.1894	insulina
22	152	+ 435.37	970.032	variáveis
23	149	+ 426.78	950.8489	nível
24	149	+ 426.78	950.8489	usuário
25	148	+ 423.91	944.4548	população
26	139	+ 398.12	886.9161	indivíduo
27	137	+ 392.39	874.1316	avaliação
28	135	+ 386.66	861.3478	alimentação
29	133	+ 380.93	848.5646	análise
30	132	+ 378.06	842.1733	diabéticos
31	123	+ 352.27	784.6589	condição
32	116	+ 332.22	739.935	importância
33	114	+ 326.49	727.1583	informações
34	116	+ 321.25	369.966	educativas
35	112	+ 320.76	714.3822	intervenções
36	112	+ 320.76	714.3822	três
37	111	+ 317.89	707.9944	idosos
38	111	+ 317.89	707.9944	realização
39	110	+ 315.03	701.6068	glicemia
40	110	+ 315.03	701.6068	mudança

Ordenado por Effect – Parte I e Parte II

Corpus Files

Acta Paulista de Ent
 Revista Acta Paulist
 Revista Acta Paulist
 Revista Acta Paulist
 Revista Acta Paulist
 Revista Arquivo Ciê
 Revista Baiana de S
 Revista Brasileira di
 Revista Brasileira di
 Revista Brasileira di
 Revista Cadernos B
 Revista Ciências e S
 Revista Cinergis_20
 Revista Cogitare En
 Revista Cogitare En
 Revista da Associaç
 Revista da Escola d
 Revista da Escola d
 Revista da Escola d
 Revista da Escola d
 Revista da Faculdad
 Revista da Rede de
 Revista da Rede de
 Revista da Rede de
 Revista de Atenção
 Revista de Ciência
 Revista de Ciência
 Revista de Enferma
 Revista Electronica E
 Revista Enfermaner

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	329	+ 942.94	2104.5386	relação
5	327	+ 937.2	2091.6895	usuários
6	288	+ 825.31	1841.2694	educação
7	256	+ 733.53	1635.9898	complicações
8	225	+ 644.63	1437.2912	indivíduos
9	197	+ 564.36	1257.9617	pés
10	194	+ 555.76	1238.7557	ações
11	194	+ 555.76	1238.7557	prática
12	185	+ 529.96	1181.1469	enfermagem
13	184	+ 527.09	1174.7467	diagnóstico
14	183	+ 524.22	1168.3468	adesão
15	362	+ 1024.4	1158.3183	mellitus
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Search Term Words Case Regex Hit Location Search Only: 0
 Start Stop Sort Reference Corpus Loaded
 Sort by Invert Order
 Sort by Effect

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	152	+ 435.37	970.032	variáveis
22	149	+ 426.78	950.8489	nível
23	149	+ 426.78	950.8489	usuário
24	148	+ 423.91	944.4548	população
25	139	+ 398.12	886.9161	indivíduo
26	137	+ 392.39	874.1316	avaliação
27	135	+ 386.66	861.3478	alimentação
28	133	+ 380.93	848.5646	análise
29	132	+ 378.06	842.1733	diabéticos
30	123	+ 352.27	784.6589	condição
31	116	+ 332.22	739.935	importância
32	114	+ 326.49	727.1583	informações
33	112	+ 320.76	714.3822	intervenções
34	112	+ 320.76	714.3822	três
35	111	+ 317.89	707.9944	idosos
36	111	+ 317.89	707.9944	realização
37	110	+ 315.03	701.6068	glicemia
38	110	+ 315.03	701.6068	mudança
39	109	+ 312.16	695.2194	após
40	104	+ 297.84	663.2847	número

Search Term Words Case Regex Hit Location Search Only: 0
 Start Stop Sort Reference Corpus Loaded
 Sort by Invert Order
 Sort by Effect

8.1.3 “Hardie’s Log-Ratio”

Tool Preferences

Category

- Concordance
- Concordance Plot
- Clusters/N-Grams
- Collocates
- Word List
- Keyword List**

Keyword List Preferences

Display Options

Rank Frequency Keyness Effect Keyword

Other Options

Treat all data as lowercase Treat case in sort Show negative keywords

Keyness Values

Keyword Statistic: Log-Likelihood (4-term)
 Keyword Statistic Threshold: $p < 0.001$ (10.83)
 Keyword Effect Size Measure: Hardie's Log Ratio
 Keyword Effect Size Threshold: Top 100

Reference Corpus

Use raw file(s) Use word list(s)

Add Directory Add Files Swap with Target Files Clear List

CAP_EM_02.txt
 CAP_EM_03.txt
 CAP_EM_04.txt

Total No. 227
 Loaded Clear

Apply Cancel

Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	12.0588	autocuidado
2	952	+ 2153.25	11.955	saúde
3	630	+ 1423.12	11.3594	doença
4	362	+ 803.85	9.5601	mellitus
5	329	+ 742.29	10.4222	relação
6	327	+ 737.78	10.4134	usuários
7	288	+ 649.68	10.2302	educação
8	256	+ 577.42	10.0602	complicações
9	225	+ 507.44	9.874	indivíduos
10	197	+ 444.24	9.6823	pés
11	194	+ 437.47	9.6601	ações
12	194	+ 437.47	9.6601	prática
13	185	+ 417.16	9.5916	enfermagem
14	184	+ 414.9	9.5838	diagnóstico
15	183	+ 412.65	9.5759	adesão
16	175	+ 394.6	9.5114	intervenção
17	170	+ 383.31	9.4696	média
18	168	+ 378.8	9.4525	práticas
19	166	+ 374.29	9.4353	física
20	165	+ 372.03	9.4265	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	8.409	insulina
22	152	+ 342.7	9.3082	variáveis
23	149	+ 335.94	9.2794	nível
24	149	+ 335.94	9.2794	usuário
25	148	+ 333.68	9.2697	população
26	139	+ 313.38	9.1792	indivíduo
27	137	+ 308.87	9.1583	avaliação
28	135	+ 304.35	9.137	alimentação
29	133	+ 299.84	9.1155	análise
30	132	+ 297.59	9.1046	diabéticos
31	123	+ 277.29	9.0027	condição
32	116	+ 250.77	7.9182	educativas
33	116	+ 261.5	8.9182	importância
34	114	+ 256.99	8.8931	informações
35	112	+ 252.48	8.8676	intervenções
36	112	+ 252.48	8.8676	três
37	111	+ 250.22	8.8546	idosos
38	111	+ 250.22	8.8546	realização
39	110	+ 247.97	8.8416	glicemia
40	110	+ 247.97	8.8416	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	12.0588	autocuidado
2	952	+ 2153.25	11.955	saúde
3	630	+ 1423.12	11.3594	doença
4	362	+ 803.85	9.5601	mellitus
5	329	+ 742.29	10.4222	relação
6	327	+ 737.78	10.4134	usuários
7	288	+ 649.68	10.2302	educação
8	256	+ 577.42	10.0602	complicações
9	225	+ 507.44	9.874	indivíduos
10	197	+ 444.24	9.6823	pés
11	194	+ 437.47	9.6601	ações
12	194	+ 437.47	9.6601	prática
13	185	+ 417.16	9.5916	enfermagem
14	184	+ 414.9	9.5838	diagnóstico
15	183	+ 412.65	9.5759	adesão
16	175	+ 394.6	9.5114	intervenção
17	170	+ 383.31	9.4696	média
18	168	+ 378.8	9.4525	práticas
19	166	+ 374.29	9.4353	física
20	165	+ 372.03	9.4265	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	8.409	insulina
22	152	+ 342.7	9.3082	variáveis
23	149	+ 335.94	9.2794	nível
24	149	+ 335.94	9.2794	usuário
25	148	+ 333.68	9.2697	população
26	139	+ 313.38	9.1792	indivíduo
27	137	+ 308.87	9.1583	avaliação
28	135	+ 304.35	9.137	alimentação
29	133	+ 299.84	9.1155	análise
30	132	+ 297.59	9.1046	diabéticos
31	123	+ 277.29	9.0027	condição
32	116	+ 261.5	8.9182	importância
33	114	+ 256.99	8.8931	informações
34	112	+ 252.48	8.8676	intervenções
35	112	+ 252.48	8.8676	três
36	116	+ 250.77	7.9182	educativas
37	111	+ 250.22	8.8546	idosos
38	111	+ 250.22	8.8546	realização
39	110	+ 247.97	8.8416	glicemia
40	110	+ 247.97	8.8416	mudança

Ordenado por Effect – Parte I e Parte II

The image displays two screenshots of the Concordance tool interface. The left screenshot shows a list of keywords sorted by effect size, with 'autocuidado' at the top. The right screenshot shows the same list sorted by frequency, with 'variáveis' at the top. Both screenshots show the search interface with 'Search Term' and 'Reference Corpus' options.

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	12.0588	autocuidado
2	952	+ 2153.25	11.955	saúde
3	630	+ 1423.12	11.3594	doença
4	329	+ 742.29	10.4222	relação
5	327	+ 737.78	10.4134	usuários
6	288	+ 649.68	10.2302	educação
7	256	+ 577.42	10.0602	complicações
8	225	+ 507.44	9.874	indivíduos
9	197	+ 444.24	9.6823	pés
10	194	+ 437.47	9.6601	ações
11	194	+ 437.47	9.6601	prática
12	185	+ 417.16	9.5916	enfermagem
13	184	+ 414.9	9.5838	diagnóstico
14	183	+ 412.65	9.5759	adesão
15	362	+ 803.85	9.5601	mellitus
16	175	+ 394.6	9.5114	intervenção
17	170	+ 383.31	9.4696	média
18	168	+ 378.8	9.4525	práticas
19	166	+ 374.29	9.4353	física
20	165	+ 372.03	9.4265	além

8.1.4 “Ratio of relative frequencies”

The image shows the 'Tool Preferences' dialog box, specifically the 'Keyword List Preferences' tab. The 'Display Options' section has 'Rank', 'Frequency', 'Keyness', 'Effect', and 'Keyword' checked. The 'Other Options' section has 'Treat all data as lowercase' checked. The 'Keyness Values' section has 'Keyword Statistic' set to 'Log-Likelihood (4-term)', 'Keyword Statistic Threshold' set to 'p < 0.001 (10.83)', 'Keyword Effect Size Measure' set to 'Ratio of relative frequencies', and 'Keyword Effect Size Threshold' set to 'Top 100'. The 'Reference Corpus' section has 'Use raw file(s)' selected. A list of files is shown below, including 'CAP_EM_02.txt', 'CAP_EM_03.txt', and 'CAP_EM_04.txt'. The 'Total No.' is 227.

Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	362	+ 803.85	754.8631	mellitus
5	329	+ 742.29	1372.0993	relação
6	327	+ 737.78	1363.7583	usuários
7	288	+ 649.68	1201.1082	educação
8	256	+ 577.42	1067.6517	complicações
9	225	+ 507.44	938.3658	indivíduos
10	197	+ 444.24	821.5914	pés
11	194	+ 437.47	809.0798	ações
12	194	+ 437.47	809.0798	prática
13	185	+ 417.16	771.5452	enfermagem
14	184	+ 414.9	767.3747	diagnóstico
15	183	+ 412.65	763.2042	adesão
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	339.8969	insulina
22	152	+ 342.7	633.9182	variáveis
23	149	+ 335.94	621.4067	nível
24	149	+ 335.94	621.4067	usuário
25	148	+ 333.68	617.2362	população
26	139	+ 313.38	579.7015	indivíduo
27	137	+ 308.87	571.3605	avaliação
28	135	+ 304.35	563.0195	alimentação
29	133	+ 299.84	554.6784	análise
30	132	+ 297.59	550.5079	diabéticos
31	123	+ 277.29	512.9733	condição
32	116	+ 250.77	241.8898	educativas
33	116	+ 261.5	483.7797	importância
34	114	+ 256.99	475.4387	informações
35	112	+ 252.48	467.0976	intervenções
36	112	+ 252.48	467.0976	três
37	111	+ 250.22	462.9271	idosos
38	111	+ 250.22	462.9271	realização
39	110	+ 247.97	458.7566	glicemia
40	110	+ 247.97	458.7566	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	362	+ 803.85	754.8631	mellitus
5	329	+ 742.29	1372.0993	relação
6	327	+ 737.78	1363.7583	usuários
7	288	+ 649.68	1201.1082	educação
8	256	+ 577.42	1067.6517	complicações
9	225	+ 507.44	938.3658	indivíduos
10	197	+ 444.24	821.5914	pés
11	194	+ 437.47	809.0798	ações
12	194	+ 437.47	809.0798	prática
13	185	+ 417.16	771.5452	enfermagem
14	184	+ 414.9	767.3747	diagnóstico
15	183	+ 412.65	763.2042	adesão
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	339.8969	insulina
22	152	+ 342.7	633.9182	variáveis
23	149	+ 335.94	621.4067	nível
24	149	+ 335.94	621.4067	usuário
25	148	+ 333.68	617.2362	população
26	139	+ 313.38	579.7015	indivíduo
27	137	+ 308.87	571.3605	avaliação
28	135	+ 304.35	563.0195	alimentação
29	133	+ 299.84	554.6784	análise
30	132	+ 297.59	550.5079	diabéticos
31	123	+ 277.29	512.9733	condição
32	116	+ 261.5	483.7797	importância
33	114	+ 256.99	475.4387	informações
34	112	+ 252.48	467.0976	intervenções
35	112	+ 252.48	467.0976	três
36	116	+ 250.77	241.8898	educativas
37	111	+ 250.22	462.9271	idosos
38	111	+ 250.22	462.9271	realização
39	110	+ 247.97	458.7566	glicemia
40	110	+ 247.97	458.7566	mudança

Ordenado por Effect – Parte I e Parte II

Keyword List (Left Screenshot):

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	329	+ 742.29	1372.0993	relação
5	327	+ 737.78	1363.7583	usuários
6	288	+ 649.68	1201.1082	educação
7	256	+ 577.42	1067.6517	complicações
8	225	+ 507.44	938.3658	indivíduos
9	197	+ 444.24	821.5914	pés
10	194	+ 437.47	809.0798	ações
11	194	+ 437.47	809.0798	prática
12	185	+ 417.16	771.5452	enfermagem
13	184	+ 414.9	767.3747	diagnóstico
14	183	+ 412.65	763.2042	adesão
15	362	+ 803.85	754.8631	mellitus
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

Keyword List (Right Screenshot):

Rank	Freq	Keyness	Effect	Keyword
21	152	+ 342.7	653.9182	variáveis
22	149	+ 335.94	621.4067	nível
23	149	+ 335.94	621.4067	usuário
24	148	+ 333.68	617.2362	população
25	139	+ 313.38	579.7015	indivíduo
26	137	+ 308.87	571.3605	avaliação
27	135	+ 304.35	563.0195	alimentação
28	133	+ 299.84	554.6784	análise
29	132	+ 297.59	550.5079	diabéticos
30	123	+ 277.29	512.9733	condição
31	116	+ 261.5	483.7797	importância
32	114	+ 256.99	475.4387	informações
33	112	+ 252.48	467.0976	intervenções
34	112	+ 252.48	467.0976	três
35	111	+ 250.22	462.9271	idosos
36	111	+ 250.22	462.9271	realização
37	110	+ 247.97	458.7566	glicemia
38	110	+ 247.97	458.7566	mudança
39	109	+ 245.71	454.5861	após
40	104	+ 234.44	433.7335	número

8.2 Desenvolvimento e aplicação dos ProtinLexis: seleção dos itens a partir da lista de palavras-chave com p-valor = 0.0001

8.2.1 “%DIFF”

Keyword List Preferences

Display Options

Rank Frequency Keyness Effect Keyword

Other Options

Treat all data as lowercase Treat case in sort Show negative keywords

Keyness Values

Keyword Statistic: Log-Likelihood (4-term)

Keyword Statistic Threshold: p < 0.0001 (15.13)

Keyword Effect Size Measure: Gabrielatos and Marchiás %DIFF

Keyword Effect Size Threshold: Top 100

Reference Corpus

Use raw file(s) Use word list(s)

Add Directory Add Files Swap with Target Files Clear List

CAP_EM_02.txt
CAP_EM_03.txt
CAP_EM_04.txt

Total No. 227

Loaded Clear

Apply Cancel

Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	362	+ 1024.4	115178.4869	mellitus
5	329	+ 942.94	209439.3491	relação
6	327	+ 937.2	208165.5537	usuários
7	288	+ 825.31	183326.5427	educação
8	256	+ 733.53	162945.8157	complicações
9	225	+ 644.63	143201.9865	indivíduos
10	197	+ 564.36	125368.8504	pés
11	194	+ 555.76	123458.1572	ações
12	194	+ 555.76	123458.1572	prática
13	185	+ 529.96	117726.0778	enfermagem
14	184	+ 527.09	117089.1801	diagnóstico
15	183	+ 524.22	116452.2823	adesão
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	51807.164	insulina
22	152	+ 435.37	96708.4531	variáveis
23	149	+ 426.78	94797.7599	nível
24	149	+ 426.78	94797.7599	usuário
25	148	+ 423.91	94160.8622	população
26	139	+ 398.12	88428.7828	indivíduo
27	137	+ 392.39	87154.9873	avaliação
28	135	+ 386.66	85881.1919	alimentação
29	133	+ 380.93	84607.3964	análise
30	132	+ 378.06	83970.4987	diabéticos
31	123	+ 352.27	78238.4193	condição
32	116	+ 321.25	36840.0676	educativas
33	116	+ 332.22	73780.1352	importância
34	114	+ 326.49	72506.3398	informações
35	112	+ 320.76	71232.5444	intervenções
36	112	+ 320.76	71232.5444	três
37	111	+ 317.89	70595.6467	idosos
38	111	+ 317.89	70595.6467	realização
39	110	+ 315.03	69958.7489	glicemia
40	110	+ 315.03	69958.7489	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	362	+ 1024.4	115178.4869	mellitus
5	329	+ 942.94	209439.3491	relação
6	327	+ 937.2	208165.5537	usuários
7	288	+ 825.31	183326.5427	educação
8	256	+ 733.53	162945.8157	complicações
9	225	+ 644.63	143201.9865	indivíduos
10	197	+ 564.36	125368.8504	pés
11	194	+ 555.76	123458.1572	ações
12	194	+ 555.76	123458.1572	prática
13	185	+ 529.96	117726.0778	enfermagem
14	184	+ 527.09	117089.1801	diagnóstico
15	183	+ 524.22	116452.2823	adesão
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	51807.164	insulina
22	152	+ 435.37	96708.4531	variáveis
23	149	+ 426.78	94797.7599	nível
24	149	+ 426.78	94797.7599	usuário
25	148	+ 423.91	94160.8622	população
26	139	+ 398.12	88428.7828	indivíduo
27	137	+ 392.39	87154.9873	avaliação
28	135	+ 386.66	85881.1919	alimentação
29	133	+ 380.93	84607.3964	análise
30	132	+ 378.06	83970.4987	diabéticos
31	123	+ 352.27	78238.4193	condição
32	116	+ 332.22	73780.1352	importância
33	114	+ 326.49	72506.3398	informações
34	116	+ 321.25	36840.0676	educativas
35	112	+ 320.76	71232.5444	intervenções
36	112	+ 320.76	71232.5444	três
37	111	+ 317.89	70595.6467	idosos
38	111	+ 317.89	70595.6467	realização
39	110	+ 315.03	69958.7489	glicemia
40	110	+ 315.03	69958.7489	mudança

Ordenado por Effect – Parte I e Parte II

Corpus Files

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	651446.3652	autocuidado
2	952	+ 2734.51	606226.6272	saúde
3	630	+ 1807.55	401145.5621	doença
4	329	+ 942.94	209439.3491	relação
5	327	+ 937.2	208165.5537	usuários
6	288	+ 825.31	183326.5427	educação
7	256	+ 733.53	162945.8157	complicações
8	225	+ 644.63	143201.9865	indivíduos
9	197	+ 564.36	125368.8504	pés
10	194	+ 555.76	123458.1572	ações
11	194	+ 555.76	123458.1572	prática
12	185	+ 529.96	117726.0778	enfermagem
13	184	+ 527.09	117089.1801	diagnóstico
14	183	+ 524.22	116452.2823	adesão
15	362	+ 1024.4	115178.4869	mellitus
16	175	+ 501.29	111357.1006	intervenção
17	170	+ 486.96	108172.612	média
18	168	+ 481.23	106898.8166	práticas
19	166	+ 475.5	105625.0211	física
20	165	+ 472.63	104988.1234	além

Search Term: Words Case Regex Advanced Search Only: 0

Reference Corpus: Loaded

Sort by: Invert Order Sort by Effect

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	152	+ 435.37	96708.4531	variáveis
22	149	+ 426.78	94797.7599	nível
23	149	+ 426.78	94797.7599	usuário
24	148	+ 423.91	94160.8622	população
25	139	+ 398.12	88428.7828	indivíduo
26	137	+ 392.39	87154.9873	avaliação
27	135	+ 386.66	85881.1919	alimentação
28	133	+ 380.93	84607.3964	análise
29	132	+ 378.06	83970.4987	diabéticos
30	123	+ 352.27	78238.4193	condição
31	116	+ 332.22	73780.1352	importância
32	114	+ 326.49	72506.3398	informações
33	112	+ 320.76	71232.5444	intervenção
34	112	+ 320.76	71232.5444	três
35	111	+ 317.89	70595.6467	idosos
36	111	+ 317.89	70595.6467	realização
37	110	+ 315.03	69958.7489	glicemia
38	110	+ 315.03	69958.7489	mudança
39	109	+ 312.16	69321.8512	após
40	104	+ 297.84	66137.3626	número

Search Term: Words Case Regex Advanced Search Only: 0

Reference Corpus: Loaded

Sort by: Invert Order Sort by Effect

8.2.2 “Odds-ratio”

Tool Preferences

Keyword List Preferences

Display Options

Rank Frequency Keyness Effect Keyword

Other Options

Treat all data as lowercase Treat case in sort Show negative keywords

Keyness Values

Keyword Statistic: Log-Likelihood (4-term)

Keyword Statistic Threshold: p < 0.0001 (15.13)

Keyword Effect Size Measure: Odds-ratio

Keyword Effect Size Threshold: Top 100

Reference Corpus

Use raw file(s) Use word list(s)

Add Directory | Add Files | Swap with Target Files | Clear List

CAP_EM_02.txt
CAP_EM_03.txt
CAP_EM_04.txt

Total No. 227 Loaded

Ordenado por Frequência – Parte I e Parte II

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	362	+ 1024.4	1158.3183	mellitus
5	329	+ 942.94	2104.5386	relação
6	327	+ 937.2	2091.6895	usuários
7	288	+ 825.31	1841.2694	educação
8	256	+ 733.53	1635.9898	complicações
9	225	+ 644.63	1437.2912	indivíduos
10	197	+ 564.36	1257.9617	pés
11	194	+ 555.76	1238.7557	ações
12	194	+ 555.76	1238.7557	prática
13	185	+ 529.96	1181.1469	enfermagem
14	184	+ 527.09	1174.7467	diagnóstico
15	183	+ 524.22	1168.3468	adesão
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Search Term Words Case Regex Hit Location Search Only 0

Start Stop Sort Reference Corpus Loaded

Sort by Invert Order Sort by Freq

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	520.1894	insulina
22	152	+ 435.37	970.032	variáveis
23	149	+ 426.78	950.8489	nível
24	149	+ 426.78	950.8489	usuário
25	148	+ 423.91	944.4548	população
26	139	+ 398.12	886.9161	indivíduo
27	137	+ 392.39	874.1316	avaliação
28	135	+ 386.66	861.3478	alimentação
29	133	+ 380.93	848.5646	análise
30	132	+ 378.06	842.1733	diabéticos
31	123	+ 352.27	784.6589	condição
32	116	+ 321.25	369.966	educativas
33	116	+ 332.22	739.935	importância
34	114	+ 326.49	727.1583	informações
35	112	+ 320.76	714.3822	intervenções
36	112	+ 320.76	714.3822	três
37	111	+ 317.89	707.9944	idosos
38	111	+ 317.89	707.9944	realização
39	110	+ 315.03	701.6068	glicemia
40	110	+ 315.03	701.6068	mudança

Search Term Words Case Regex Hit Location Search Only 0

Start Stop Sort Reference Corpus Loaded

Sort by Invert Order Sort by Freq

Ordenado por Keyness – Parte I e Parte II

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	362	+ 1024.4	1158.3183	mellitus
5	329	+ 942.94	2104.5386	relação
6	327	+ 937.2	2091.6895	usuários
7	288	+ 825.31	1841.2694	educação
8	256	+ 733.53	1635.9898	complicações
9	225	+ 644.63	1437.2912	indivíduos
10	197	+ 564.36	1257.9617	pés
11	194	+ 555.76	1238.7557	ações
12	194	+ 555.76	1238.7557	prática
13	185	+ 529.96	1181.1469	enfermagem
14	184	+ 527.09	1174.7467	diagnóstico
15	183	+ 524.22	1168.3468	adesão
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Search Term Words Case Regex Hit Location Search Only 0

Start Stop Sort Reference Corpus Loaded

Sort by Invert Order Sort by Keyness

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

Keyword Types: 100 Keyword Tokens: 13211 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 455.25	520.1894	insulina
22	152	+ 435.37	970.032	variáveis
23	149	+ 426.78	950.8489	nível
24	149	+ 426.78	950.8489	usuário
25	148	+ 423.91	944.4548	população
26	139	+ 398.12	886.9161	indivíduo
27	137	+ 392.39	874.1316	avaliação
28	135	+ 386.66	861.3478	alimentação
29	133	+ 380.93	848.5646	análise
30	132	+ 378.06	842.1733	diabéticos
31	123	+ 352.27	784.6589	condição
32	116	+ 332.22	739.935	importância
33	114	+ 326.49	727.1583	informações
34	116	+ 321.25	369.966	educativas
35	112	+ 320.76	714.3822	intervenções
36	112	+ 320.76	714.3822	três
37	111	+ 317.89	707.9944	idosos
38	111	+ 317.89	707.9944	realização
39	110	+ 315.03	701.6068	glicemia
40	110	+ 315.03	701.6068	mudança

Search Term Words Case Regex Hit Location Search Only 0

Start Stop Sort Reference Corpus Loaded

Sort by Invert Order Sort by Keyness

Ordenado por Effect – Parte I e Parte II

Corpus Files

- Acta Paulista de Enl
- Acta Paulista de Enl
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Acta Paulist
- Revista Arquivo Cié
- Revista Baiana de S
- Revista Brasileira di
- Revista Brasileira di
- Revista Brasileira di
- Revista Brasileira di
- Revista Cadermos B
- Revista Ciências e S
- Revista Cimergis_20
- Revista Cogitare En
- Revista Cogitare En
- Revista da Associaç
- Revista da Escola d
- Revista da Escola d
- Revista da Escola d
- Revista da Faculdad
- Revista da Rede de
- Revista da Rede de
- Revista da Rede de
- Revista de Atenção
- Revista de Ciência
- Revista de Enferma
- Revista Eletrônica E
- Revista Enfermaner

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2939.19	6604.7046	autocuidado
2	952	+ 2734.51	6140.4764	saúde
3	630	+ 1807.55	4046.1234	doença
4	329	+ 942.94	2104.5386	relação
5	327	+ 937.2	2091.6895	usuários
6	288	+ 825.31	1841.2694	educação
7	256	+ 733.53	1635.9898	complicações
8	225	+ 644.63	1437.2912	indivíduos
9	197	+ 564.36	1257.9617	pés
10	194	+ 555.76	1238.7557	ações
11	194	+ 555.76	1238.7557	prática
12	185	+ 529.96	1181.1469	enfermagem
13	184	+ 527.09	1174.7467	diagnóstico
14	183	+ 524.22	1168.3468	adesão
15	362	+ 1024.4	1158.3183	melitius
16	175	+ 501.29	1117.1532	intervenção
17	170	+ 486.96	1085.1627	média
18	168	+ 481.23	1072.3677	práticas
19	166	+ 475.5	1059.5733	física
20	165	+ 472.63	1053.1764	além

Keyword Types: 100 **Keyword Tokens:** 13211 **Search Hits:** 0

Search Term Words Case Regex Advanced Search Only 0 Reference Corpus Loaded

Total No. 40 Files Processed

Rank	Freq	Keyness	Effect	Keyword
21	152	+ 435.37	970.032	variáveis
22	149	+ 426.78	950.8489	nível
23	149	+ 426.78	950.8489	usuário
24	148	+ 423.91	944.4548	população
25	139	+ 398.12	886.9161	indivíduo
26	137	+ 392.39	874.1316	avaliação
27	135	+ 386.66	861.3478	alimentação
28	133	+ 380.93	848.5646	análise
29	132	+ 378.06	842.1733	diabéticos
30	123	+ 352.27	784.6589	condição
31	116	+ 332.22	739.935	importância
32	114	+ 326.49	727.1583	informações
33	112	+ 320.76	714.3822	intervenções
34	112	+ 320.76	714.3822	três
35	111	+ 317.89	707.9944	idosos
36	111	+ 317.89	707.9944	realização
37	110	+ 315.03	701.6068	glicemia
38	110	+ 315.03	701.6068	mudança
39	109	+ 312.16	695.2194	após
40	104	+ 297.84	663.2847	número

Keyword Types: 100 **Keyword Tokens:** 13211 **Search Hits:** 0

Search Term Words Case Regex Advanced Search Only 0 Reference Corpus Loaded

Sort by Invert Order

8.2.3 “Hardie’s Log-Ratio”

Tool Preferences

Category

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

Keyword List Preferences

Display Options

- Rank
- Frequency
- Keyness
- Effect
- Keyword

Other Options

- Treat all data as lowercase
- Treat case in sort
- Show negative keywords

Keyness Values

Keyword Statistic: Log-Likelihood (4-term)

Keyword Statistic Threshold: p < 0.0001 (15,13)

Keyword Effect Size Measure: **Hardie's Log Ratio**

Keyword Effect Size Threshold: Top 100

Reference Corpus

Use raw file(s) Use word list(s)

Add Directory Add Files Swap with Target Files Clear List

CAP_EM_02.txt
CAP_EM_03.txt
CAP_EM_04.txt

Total No. 227

Loaded Clear

Apply Cancel

Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	12.0588	autocuidado
2	952	+ 2153.25	11.955	saúde
3	630	+ 1423.12	11.3594	doença
4	362	+ 803.85	9.5601	mellitus
5	329	+ 742.29	10.4222	relação
6	327	+ 737.78	10.4134	usuários
7	288	+ 649.68	10.2302	educação
8	256	+ 577.42	10.0602	complicações
9	225	+ 507.44	9.874	indivíduos
10	197	+ 444.24	9.6823	pés
11	194	+ 437.47	9.6601	ações
12	194	+ 437.47	9.6601	prática
13	185	+ 417.16	9.5916	enfermagem
14	184	+ 414.9	9.5838	diagnóstico
15	183	+ 412.65	9.5759	adesão
16	175	+ 394.6	9.5114	intervenção
17	170	+ 383.31	9.4696	média
18	168	+ 378.8	9.4525	práticas
19	166	+ 374.29	9.4353	física
20	165	+ 372.03	9.4265	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	8.409	insulina
22	152	+ 342.7	9.3082	variáveis
23	149	+ 335.94	9.2794	nível
24	149	+ 335.94	9.2794	usuário
25	148	+ 333.68	9.2697	população
26	139	+ 313.38	9.1792	indivíduo
27	137	+ 308.87	9.1583	avaliação
28	135	+ 304.35	9.137	alimentação
29	133	+ 299.84	9.1155	análise
30	132	+ 297.59	9.1046	diabéticos
31	123	+ 277.29	9.0027	condição
32	116	+ 250.77	7.9182	educativas
33	116	+ 261.5	8.9182	importância
34	114	+ 256.99	8.8931	informações
35	112	+ 252.48	8.8676	intervenções
36	112	+ 252.48	8.8676	três
37	111	+ 250.22	8.8546	idosos
38	111	+ 250.22	8.8546	realização
39	110	+ 247.97	8.8416	glicemia
40	110	+ 247.97	8.8416	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	12.0588	autocuidado
2	952	+ 2153.25	11.955	saúde
3	630	+ 1423.12	11.3594	doença
4	362	+ 803.85	9.5601	mellitus
5	329	+ 742.29	10.4222	relação
6	327	+ 737.78	10.4134	usuários
7	288	+ 649.68	10.2302	educação
8	256	+ 577.42	10.0602	complicações
9	225	+ 507.44	9.874	indivíduos
10	197	+ 444.24	9.6823	pés
11	194	+ 437.47	9.6601	ações
12	194	+ 437.47	9.6601	prática
13	185	+ 417.16	9.5916	enfermagem
14	184	+ 414.9	9.5838	diagnóstico
15	183	+ 412.65	9.5759	adesão
16	175	+ 394.6	9.5114	intervenção
17	170	+ 383.31	9.4696	média
18	168	+ 378.8	9.4525	práticas
19	166	+ 374.29	9.4353	física
20	165	+ 372.03	9.4265	além

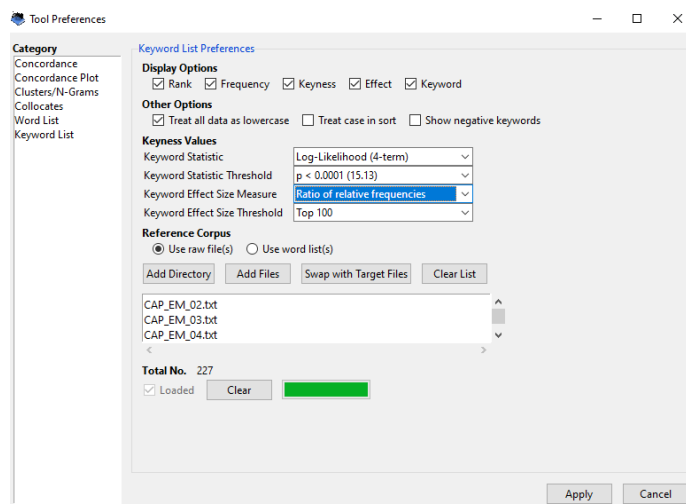
Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	8.409	insulina
22	152	+ 342.7	9.3082	variáveis
23	149	+ 335.94	9.2794	nível
24	149	+ 335.94	9.2794	usuário
25	148	+ 333.68	9.2697	população
26	139	+ 313.38	9.1792	indivíduo
27	137	+ 308.87	9.1583	avaliação
28	135	+ 304.35	9.137	alimentação
29	133	+ 299.84	9.1155	análise
30	132	+ 297.59	9.1046	diabéticos
31	123	+ 277.29	9.0027	condição
32	116	+ 261.5	8.9182	importância
33	114	+ 256.99	8.8931	informações
34	112	+ 252.48	8.8676	intervenções
35	112	+ 252.48	8.8676	três
36	116	+ 250.77	7.9182	educativas
37	111	+ 250.22	8.8546	idosos
38	111	+ 250.22	8.8546	realização
39	110	+ 247.97	8.8416	glicemia
40	110	+ 247.97	8.8416	mudança

Ordenado por Effect – Parte I e Parte II

Concordance		Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Keyword Types: 100		Keyword Tokens: 13211		Search Hits: 0				
Rank	Freq	Keyness	Effect	Keyword				
1	1023	+ 2314.49	12.0588	autocuidado				
2	952	+ 2153.25	11.955	saúde				
3	630	+ 1423.12	11.3594	doença				
4	329	+ 742.29	10.4222	relação				
5	327	+ 737.78	10.4134	usuários				
6	288	+ 649.68	10.2302	educação				
7	256	+ 577.42	10.0602	complicações				
8	225	+ 507.44	9.874	indivíduos				
9	197	+ 444.24	9.6823	pés				
10	194	+ 437.47	9.6601	ações				
11	194	+ 437.47	9.6601	prática				
12	185	+ 417.16	9.5916	enfermagem				
13	184	+ 414.9	9.5838	diagnóstico				
14	183	+ 412.65	9.5759	adesão				
15	362	+ 803.85	9.5601	mellitus				
16	175	+ 394.6	9.5114	intervenção				
17	170	+ 383.31	9.4696	média				
18	168	+ 378.8	9.4525	práticas				
19	166	+ 374.29	9.4353	física				
20	165	+ 372.03	9.4265	além				

Concordance		Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Keyword Types: 100		Keyword Tokens: 13211		Search Hits: 0				
Rank	Freq	Keyness	Effect	Keyword				
21	152	+ 342.7	9.3082	variáveis				
22	149	+ 335.94	9.2794	nível				
23	149	+ 335.94	9.2794	usuário				
24	148	+ 333.68	9.2697	população				
25	139	+ 313.38	9.1792	indivíduo				
26	137	+ 308.87	9.1583	avaliação				
27	135	+ 304.35	9.137	alimentação				
28	133	+ 299.84	9.1155	análise				
29	132	+ 297.59	9.1046	diabéticos				
30	123	+ 277.29	9.0027	condição				
31	116	+ 261.5	8.9182	importância				
32	114	+ 256.99	8.8931	informações				
33	112	+ 252.48	8.8676	intervenções				
34	112	+ 252.48	8.8676	três				
35	111	+ 250.22	8.8546	idosos				
36	111	+ 250.22	8.8546	realização				
37	110	+ 247.97	8.8416	glicemia				
38	110	+ 247.97	8.8416	mudança				
39	109	+ 245.71	8.8284	após				
40	104	+ 234.44	8.7607	número				

8.2.4 “Ratio of relative frequencies”



Ordenado por Frequência – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	362	+ 803.85	754.8631	mellitus
5	329	+ 742.29	1372.0993	relação
6	327	+ 737.78	1363.7583	usuários
7	288	+ 649.68	1201.1082	educação
8	256	+ 577.42	1067.6517	complicações
9	225	+ 507.44	938.3658	indivíduos
10	197	+ 444.24	821.5914	pés
11	194	+ 437.47	809.0798	ações
12	194	+ 437.47	809.0798	prática
13	185	+ 417.16	771.5452	enfermagem
14	184	+ 414.9	767.3747	diagnóstico
15	183	+ 412.65	763.2042	adesão
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	339.8969	insulina
22	152	+ 342.7	633.9182	variáveis
23	149	+ 335.94	621.4067	nível
24	149	+ 335.94	621.4067	usuário
25	148	+ 333.68	617.2362	população
26	139	+ 313.38	579.7015	indivíduo
27	137	+ 308.87	571.3605	avaliação
28	135	+ 304.35	563.0195	alimentação
29	133	+ 299.84	554.6784	análise
30	132	+ 297.59	550.5079	diabéticos
31	123	+ 277.29	512.9733	condição
32	116	+ 250.77	241.8898	educativas
33	116	+ 261.5	483.7797	importância
34	114	+ 256.99	475.4387	informações
35	112	+ 252.48	467.0976	intervenção
36	112	+ 252.48	467.0976	três
37	111	+ 250.22	462.9271	idosos
38	111	+ 250.22	462.9271	realização
39	110	+ 247.97	458.7566	glicemia
40	110	+ 247.97	458.7566	mudança

Ordenado por Keyness – Parte I e Parte II

Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	362	+ 803.85	754.8631	mellitus
5	329	+ 742.29	1372.0993	relação
6	327	+ 737.78	1363.7583	usuários
7	288	+ 649.68	1201.1082	educação
8	256	+ 577.42	1067.6517	complicações
9	225	+ 507.44	938.3658	indivíduos
10	197	+ 444.24	821.5914	pés
11	194	+ 437.47	809.0798	ações
12	194	+ 437.47	809.0798	prática
13	185	+ 417.16	771.5452	enfermagem
14	184	+ 414.9	767.3747	diagnóstico
15	183	+ 412.65	763.2042	adesão
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

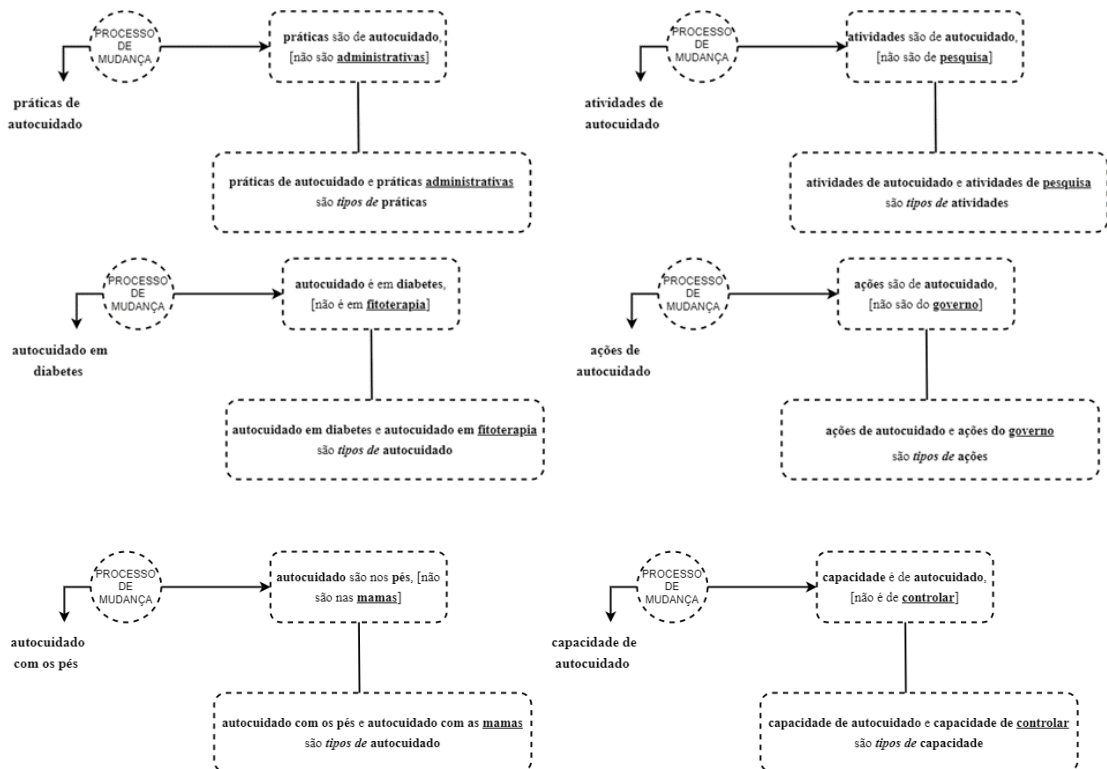
Rank	Freq	Keyness	Effect	Keyword
21	163	+ 356.11	339.8969	insulina
22	152	+ 342.7	633.9182	variáveis
23	149	+ 335.94	621.4067	nível
24	149	+ 335.94	621.4067	usuário
25	148	+ 333.68	617.2362	população
26	139	+ 313.38	579.7015	indivíduo
27	137	+ 308.87	571.3605	avaliação
28	135	+ 304.35	563.0195	alimentação
29	133	+ 299.84	554.6784	análise
30	132	+ 297.59	550.5079	diabéticos
31	123	+ 277.29	512.9733	condição
32	116	+ 261.5	483.7797	importância
33	114	+ 256.99	475.4387	informações
34	112	+ 252.48	467.0976	intervenção
35	112	+ 252.48	467.0976	três
36	116	+ 250.77	241.8898	educativas
37	111	+ 250.22	462.9271	idosos
38	111	+ 250.22	462.9271	realização
39	110	+ 247.97	458.7566	glicemia
40	110	+ 247.97	458.7566	mudança

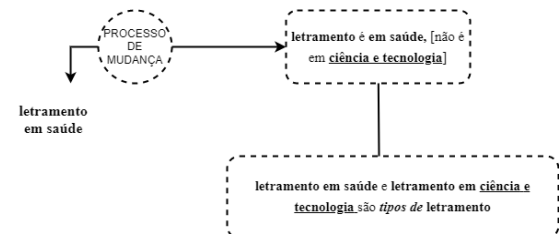
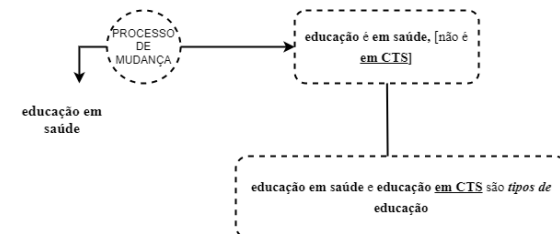
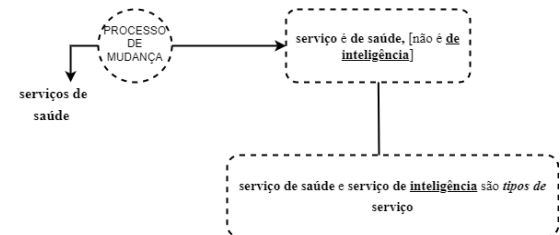
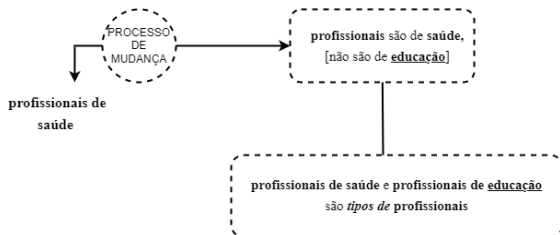
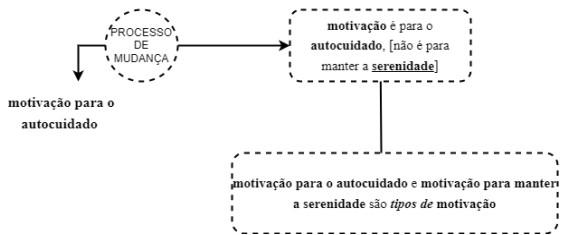
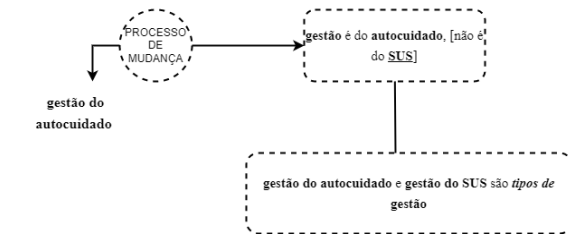
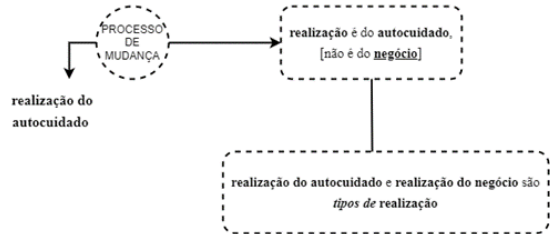
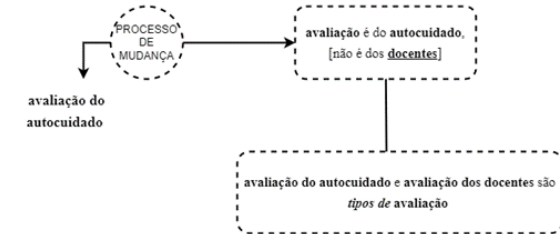
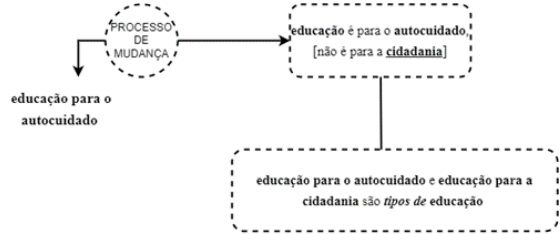
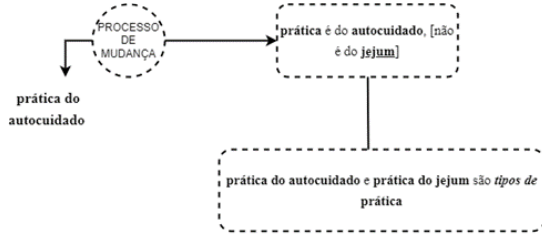
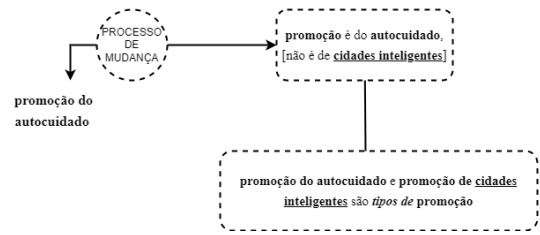
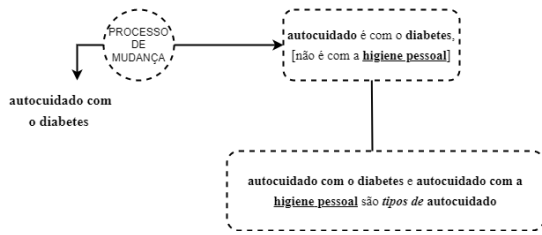
Ordenado por Effect – Parte I e Parte II

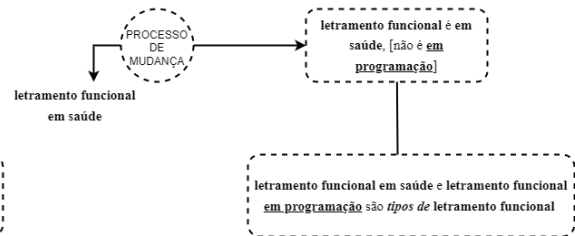
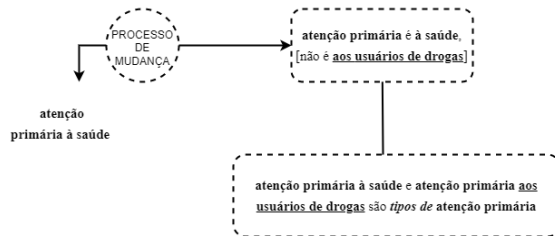
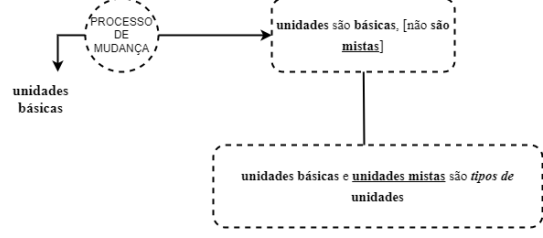
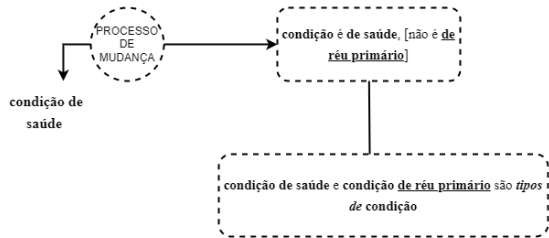
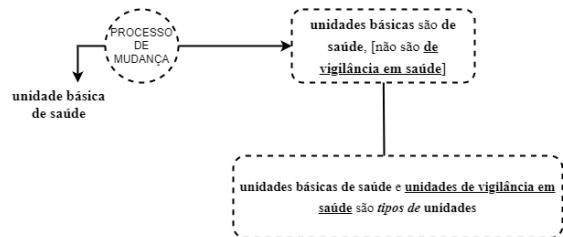
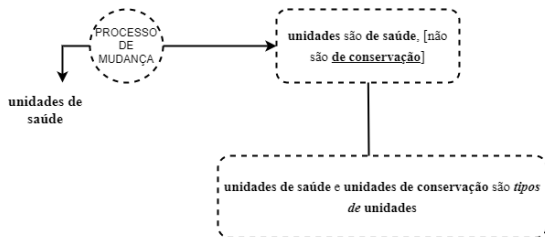
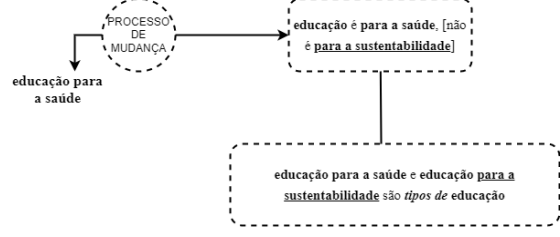
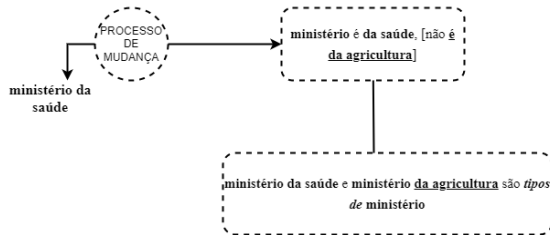
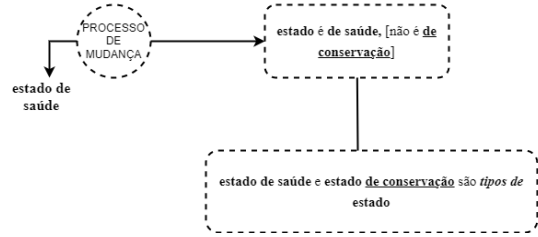
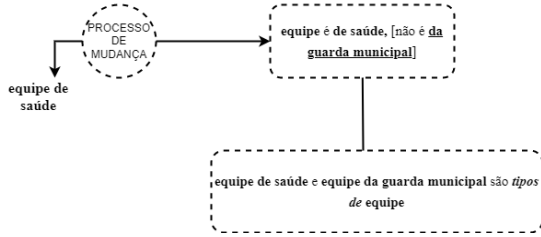
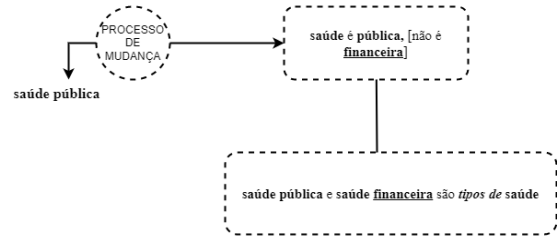
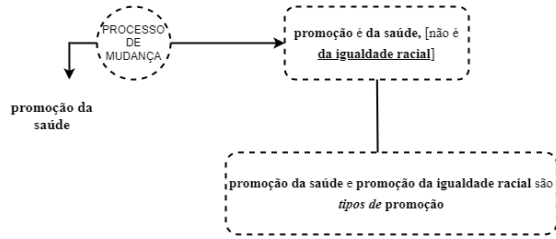
Rank	Freq	Keyness	Effect	Keyword
1	1023	+ 2314.49	4266.4364	autocuidado
2	952	+ 2153.25	3970.3299	saúde
3	630	+ 1423.12	2627.4242	doença
4	329	+ 742.29	1372.0993	relação
5	327	+ 737.78	1363.7583	usuários
6	288	+ 649.68	1201.1082	educação
7	256	+ 577.42	1067.6517	complicações
8	225	+ 507.44	938.3658	indivíduos
9	197	+ 444.24	821.5914	pés
10	194	+ 437.47	809.0798	ações
11	194	+ 437.47	809.0798	prática
12	185	+ 417.16	771.5452	enfermagem
13	184	+ 414.9	767.3747	diagnóstico
14	183	+ 412.65	763.2042	adesão
15	362	+ 803.85	754.8631	mellitus
16	175	+ 394.6	729.8401	intervenção
17	170	+ 383.31	708.9875	média
18	168	+ 378.8	700.6464	práticas
19	166	+ 374.29	692.3054	física
20	165	+ 372.03	688.1349	além

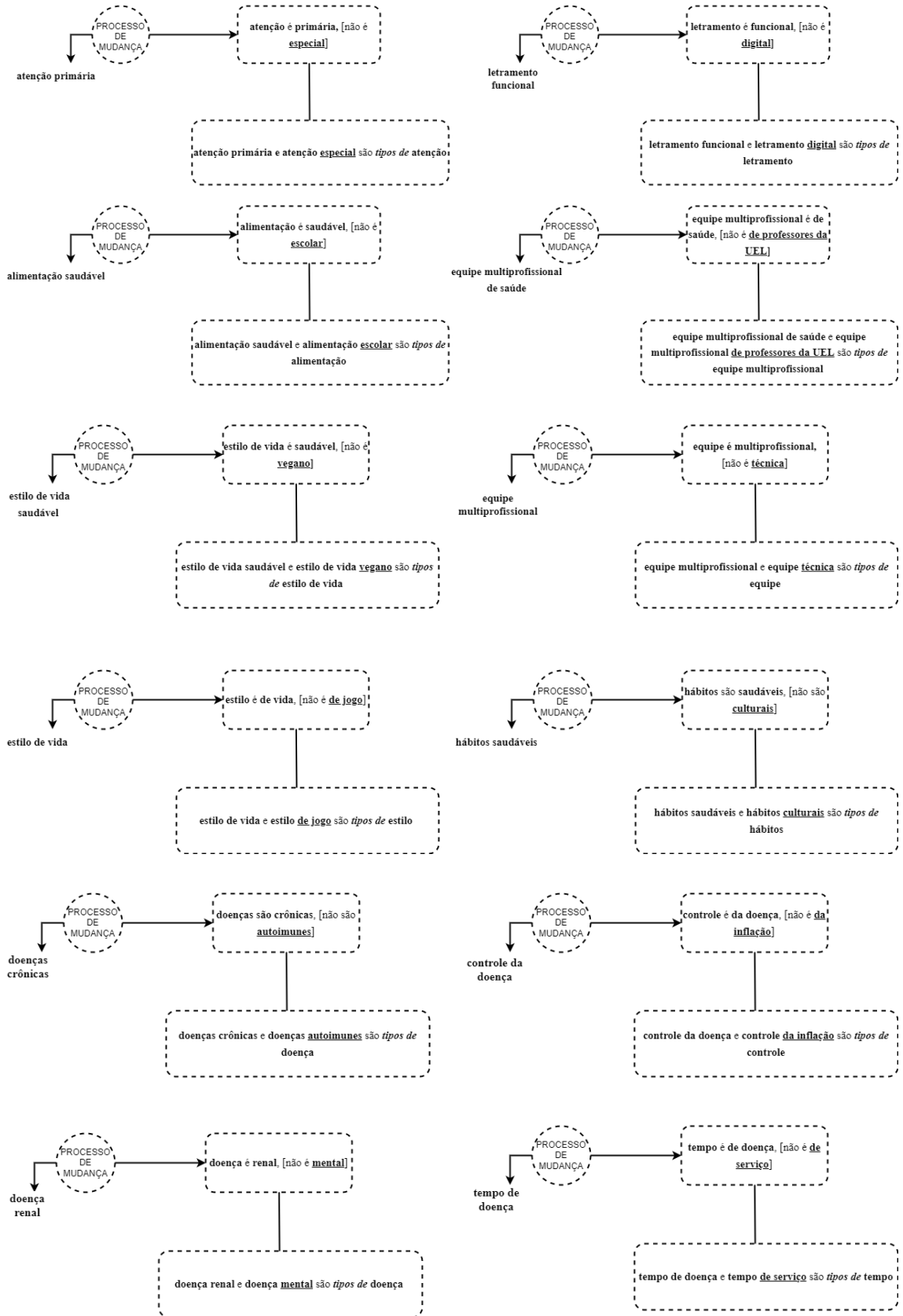
Rank	Freq	Keyness	Effect	Keyword
21	152	+ 342.7	633.9182	variáveis
22	149	+ 335.94	621.4067	nível
23	149	+ 335.94	621.4067	usuário
24	148	+ 333.68	617.2362	população
25	139	+ 313.38	579.7015	indivíduo
26	137	+ 308.87	571.3605	avaliação
27	135	+ 304.35	563.0195	alimentação
28	133	+ 299.84	554.6784	análise
29	132	+ 297.59	550.5079	diabéticos
30	123	+ 277.29	512.9733	condição
31	116	+ 261.5	483.7797	importância
32	114	+ 256.99	475.4387	informações
33	112	+ 252.48	467.0976	intervenções
34	112	+ 252.48	467.0976	três
35	111	+ 250.22	462.9271	idosos
36	111	+ 250.22	462.9271	realização
37	110	+ 247.97	458.7566	glicemia
38	110	+ 247.97	458.7566	mudança
39	109	+ 245.71	454.5861	após
40	104	+ 234.44	433.7335	número

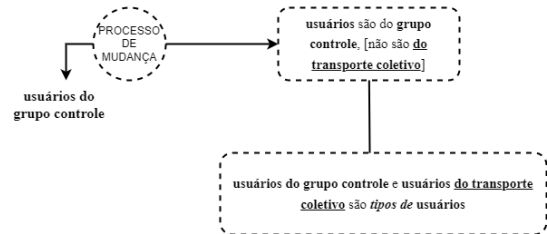
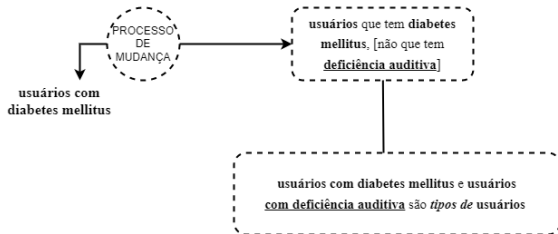
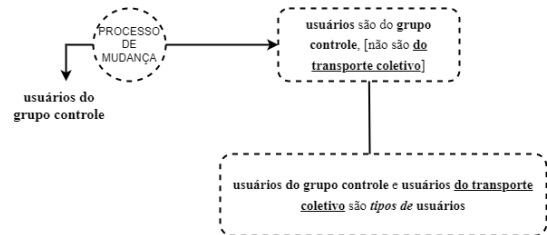
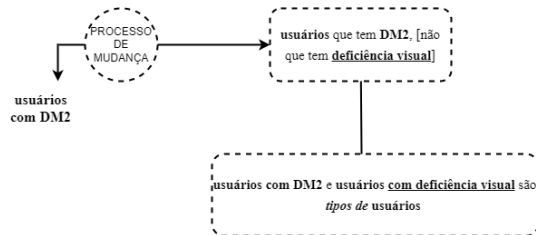
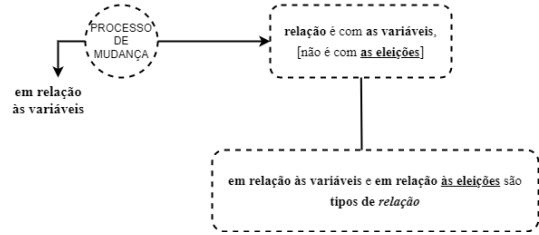
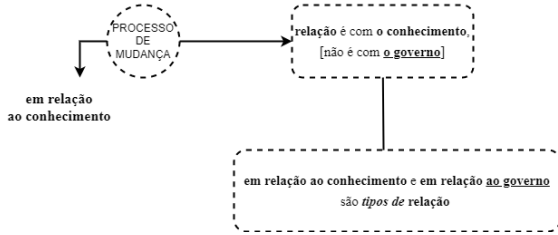
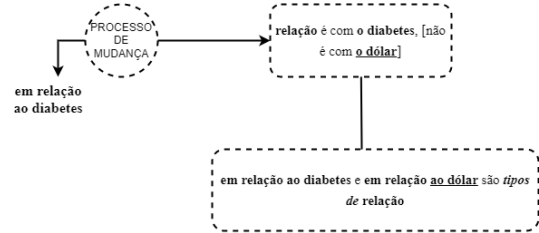
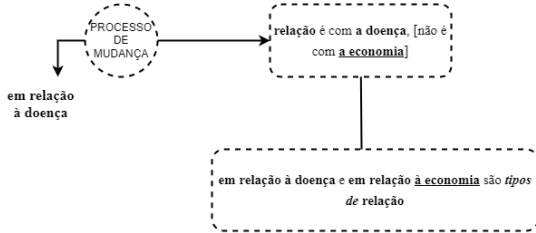
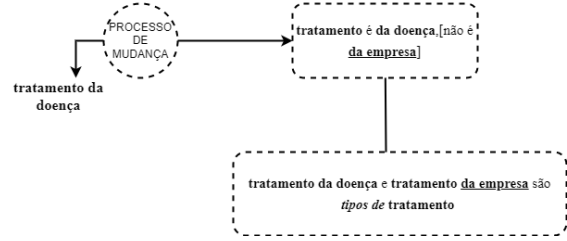
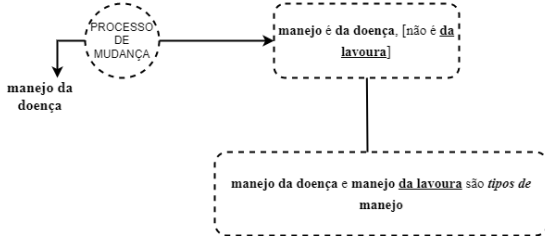
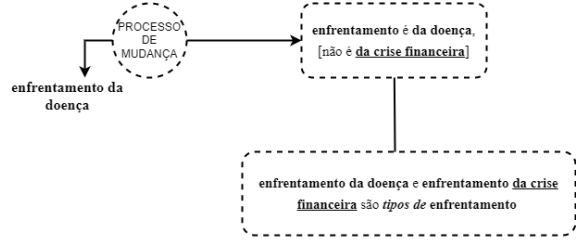
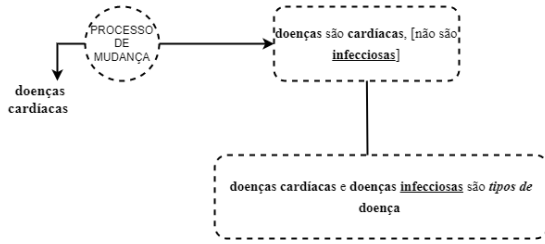
8.3 Representação ProtinLex:PTC – processo de mudança dos itens selecionados para aplicação e desenvolvimento dos ProtinLexis

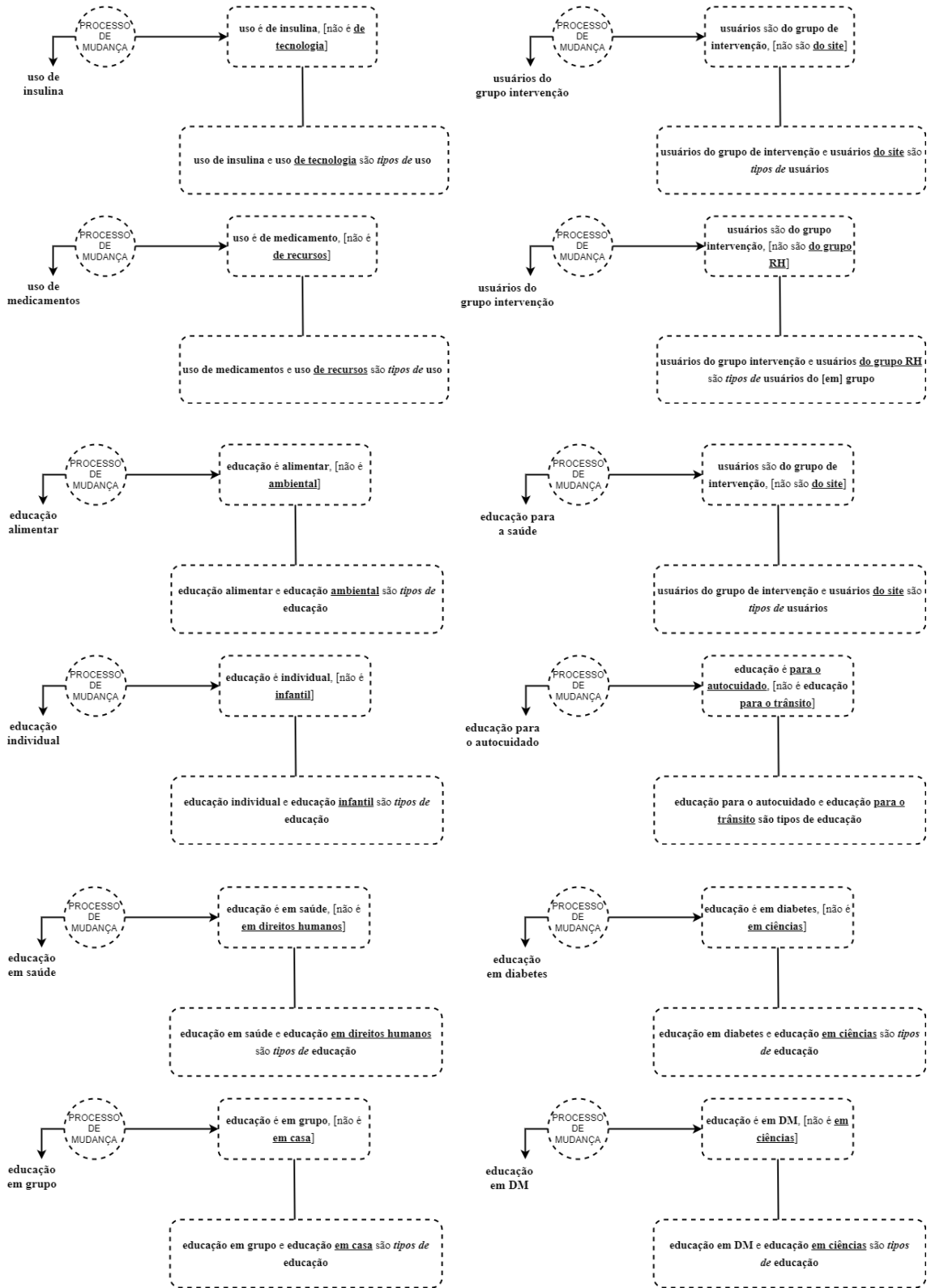


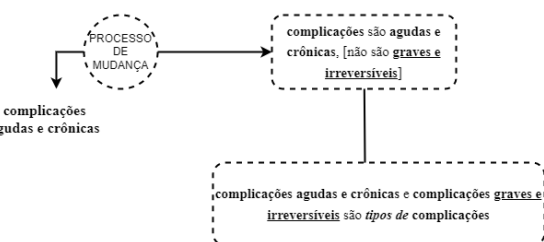
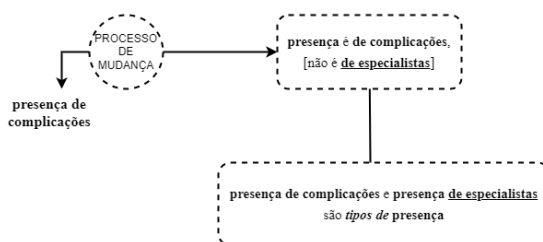
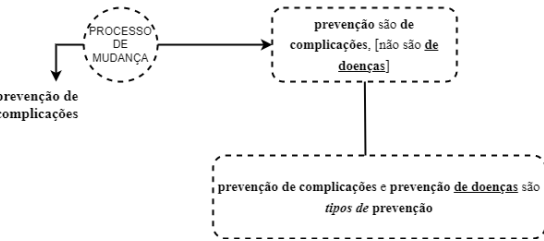
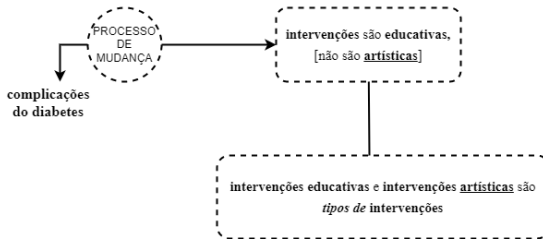
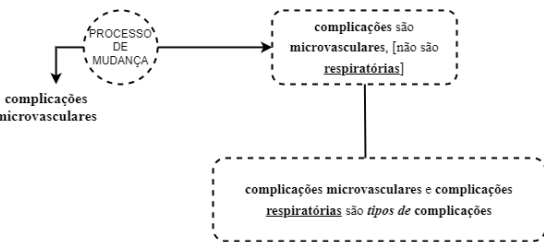
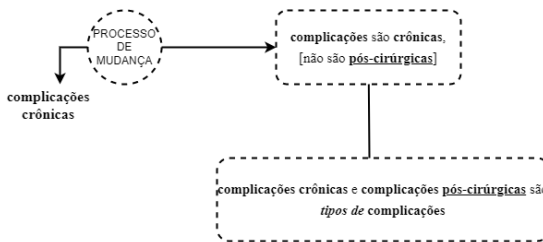
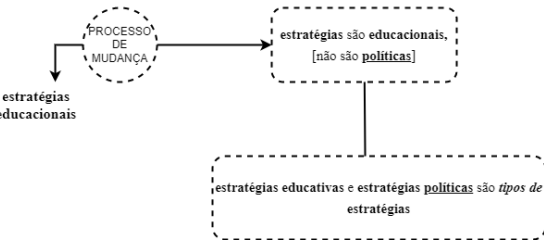
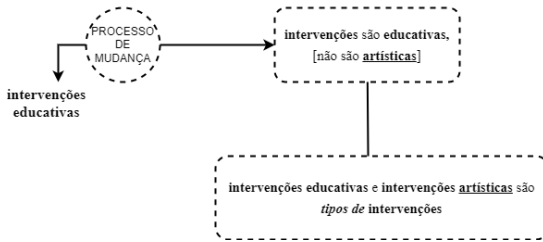
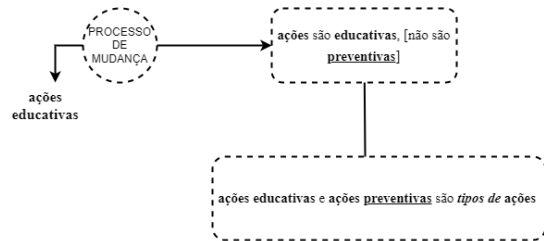
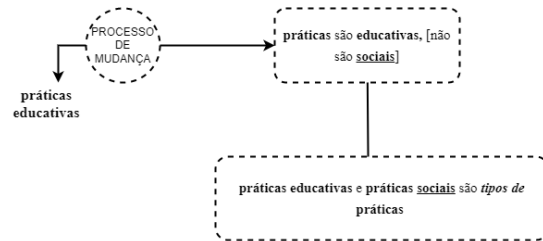
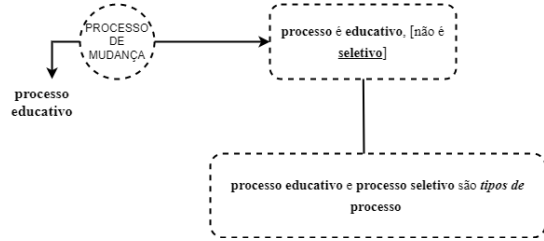
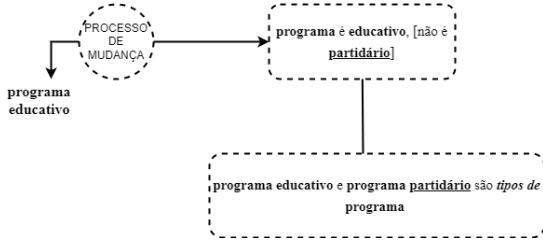
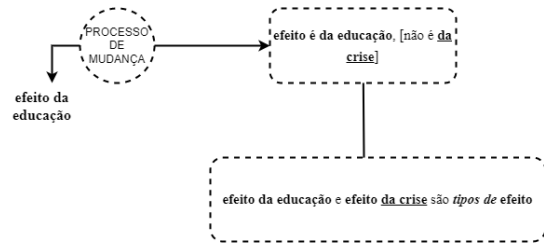
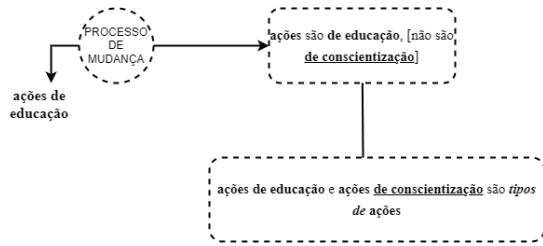


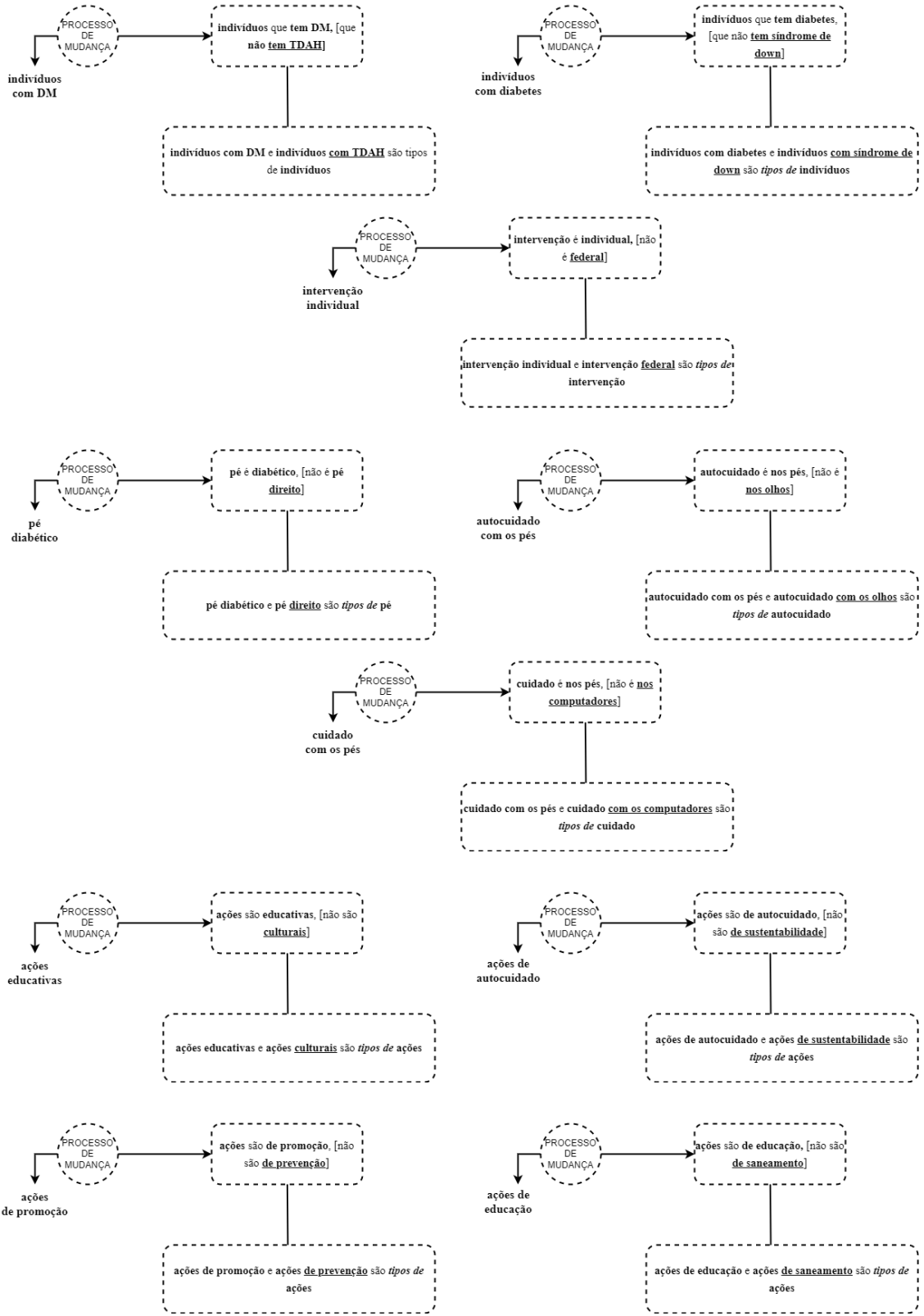


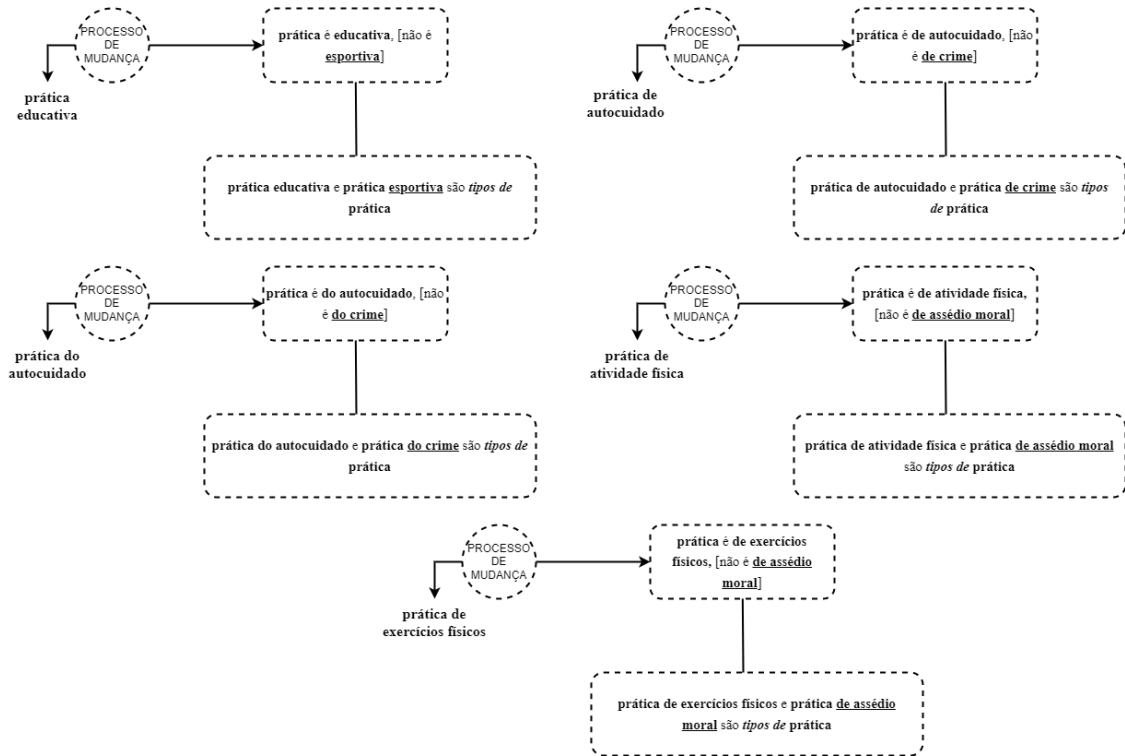






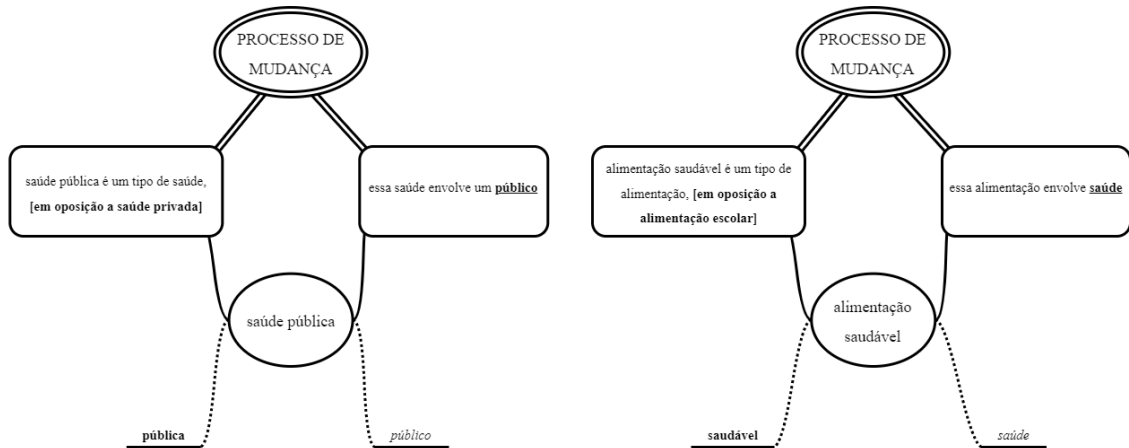


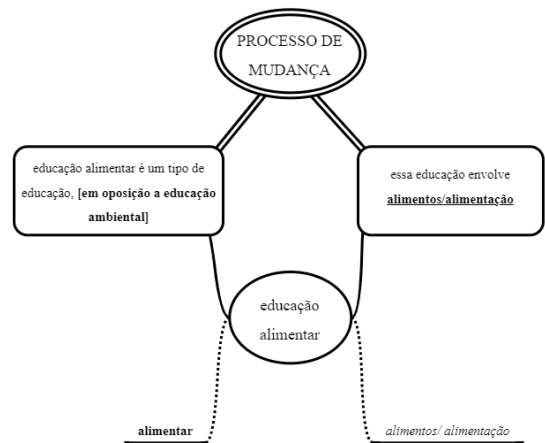
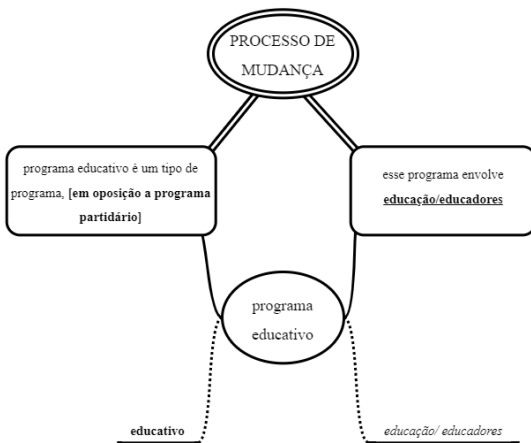
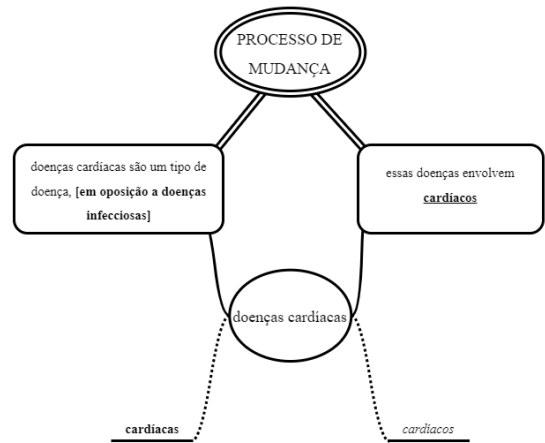
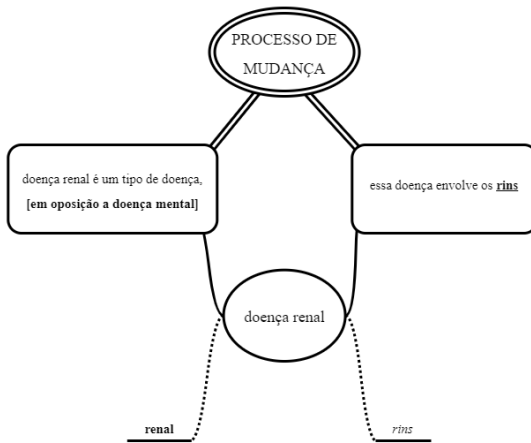
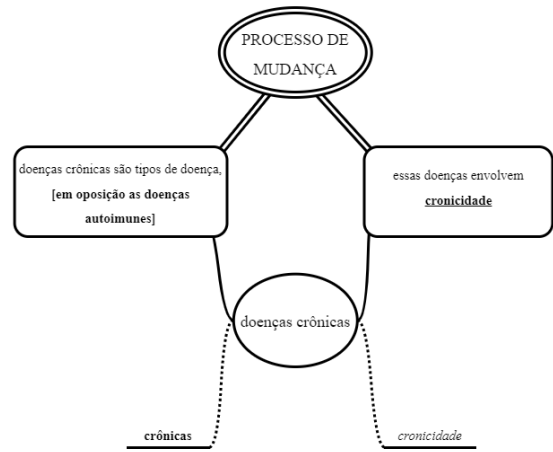
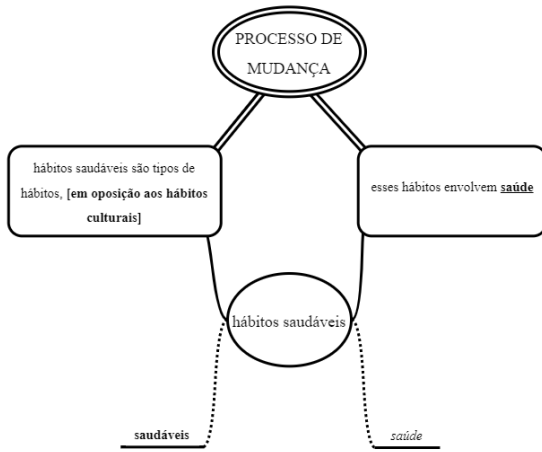


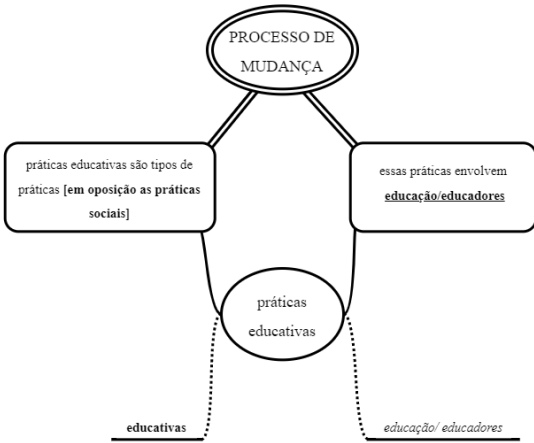
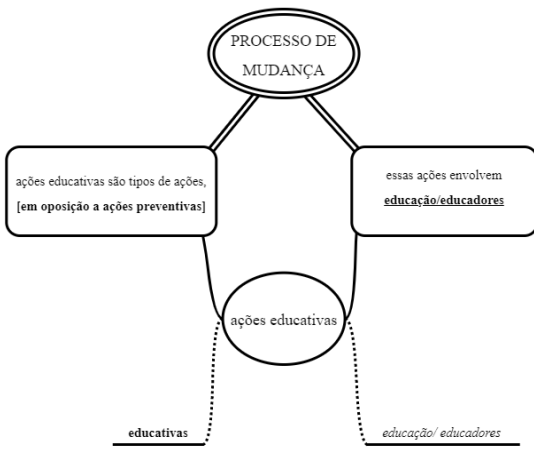
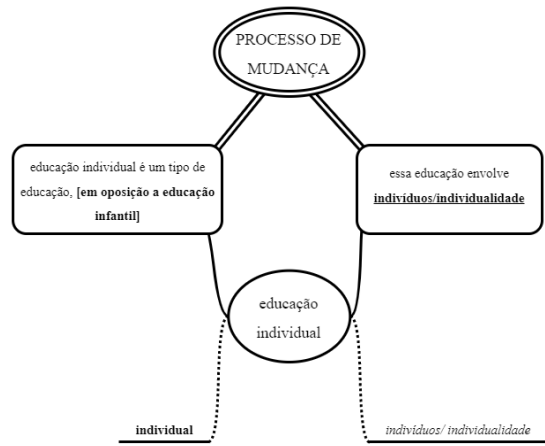
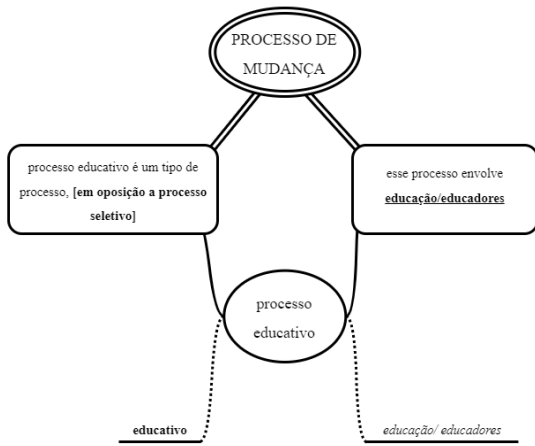


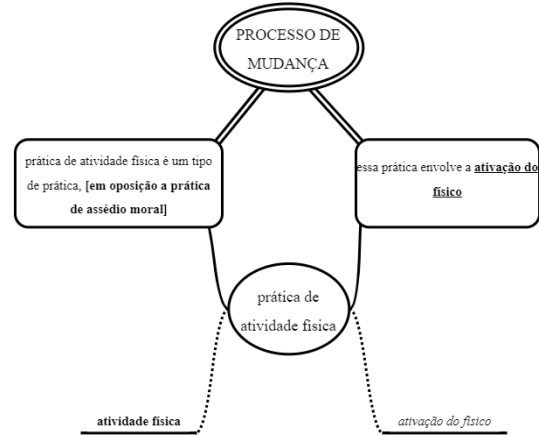
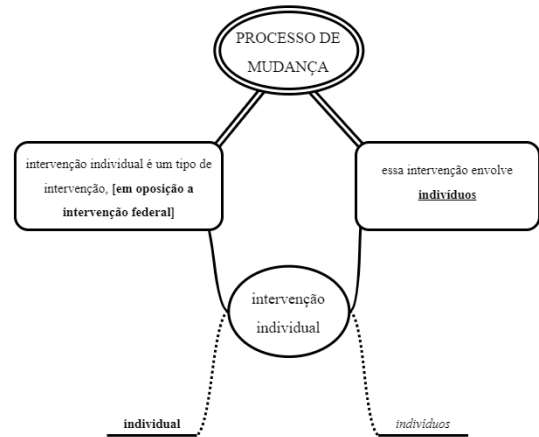
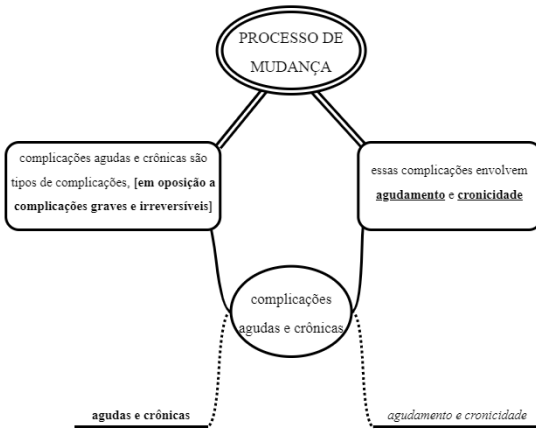
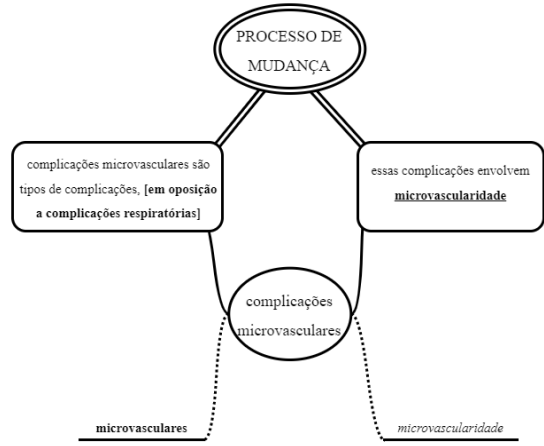
8.4 Representação ProtinLex:PTG – processo de mudança dos itens selecionados para aplicação e desenvolvimento dos ProtinLexis

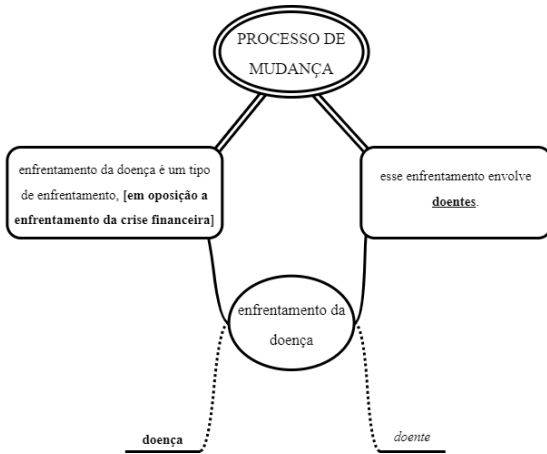
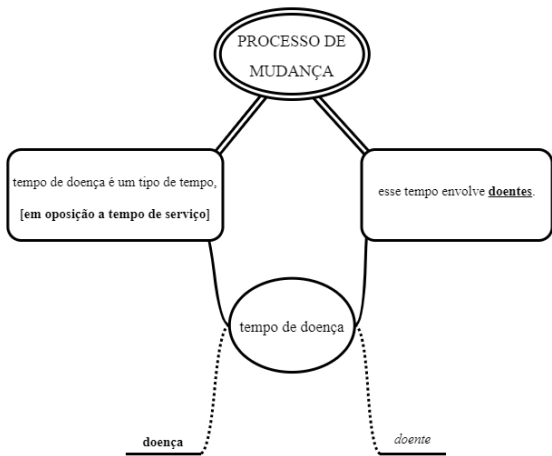
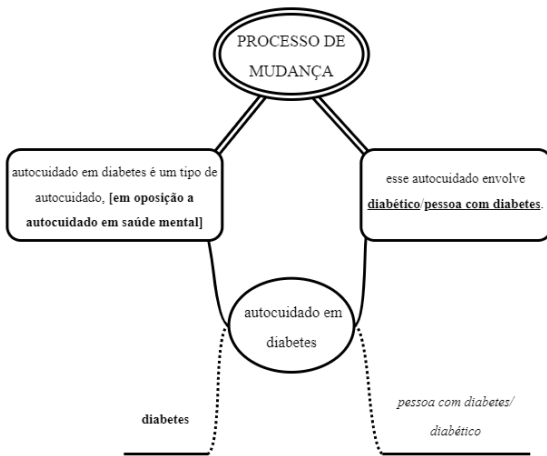
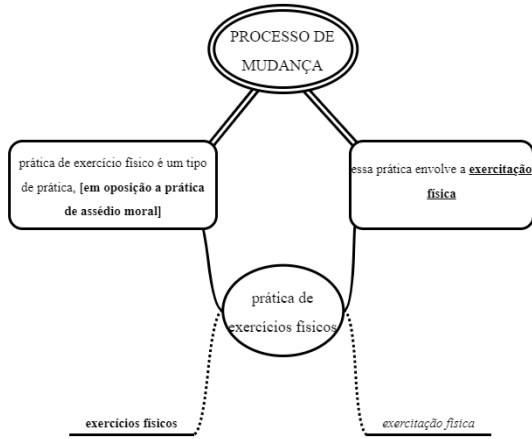
8.4.1 Representação ProtinLex:PTG – Resultado Default

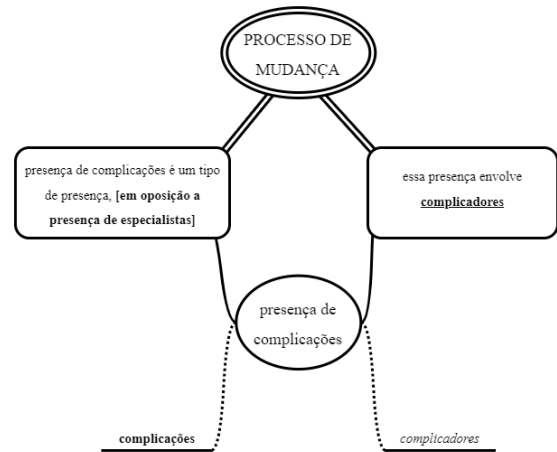
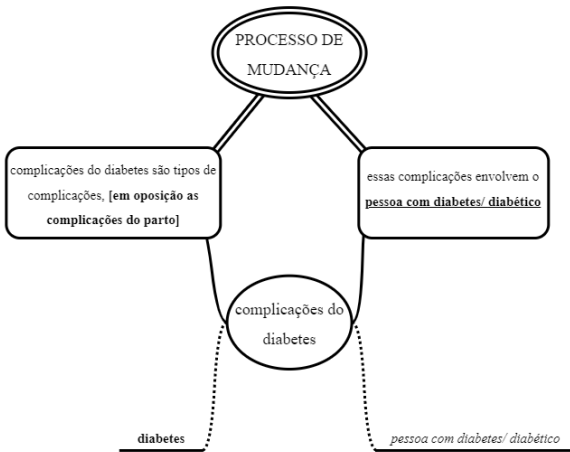
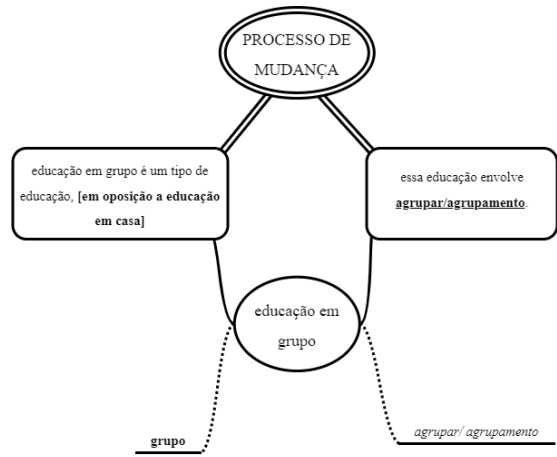
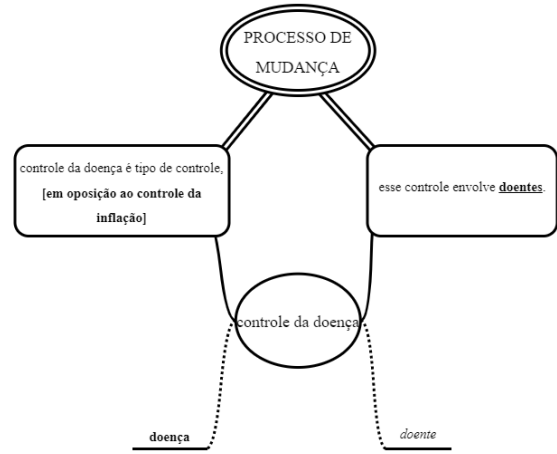
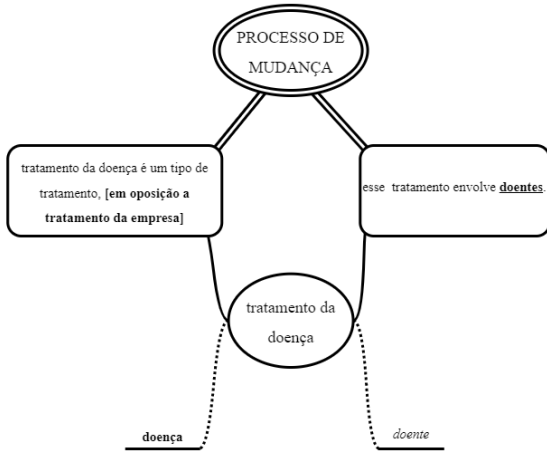


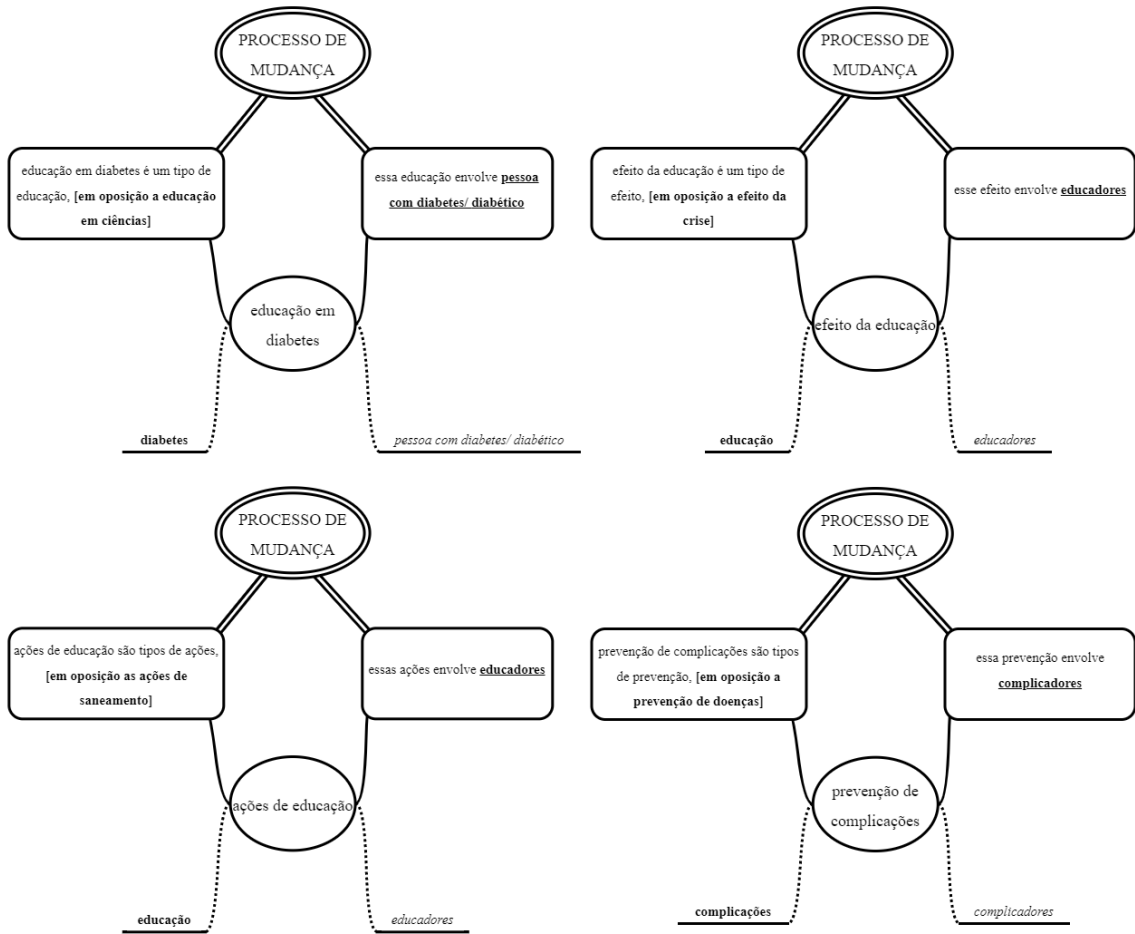




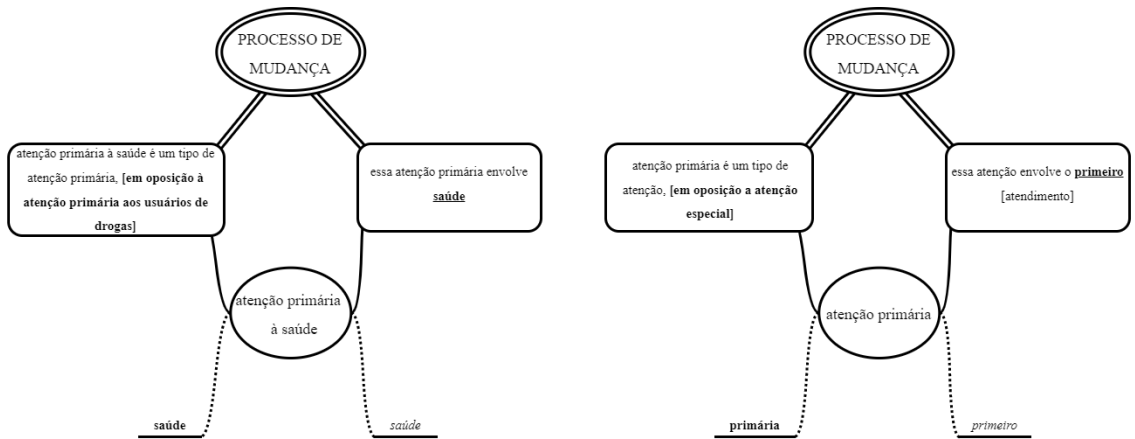


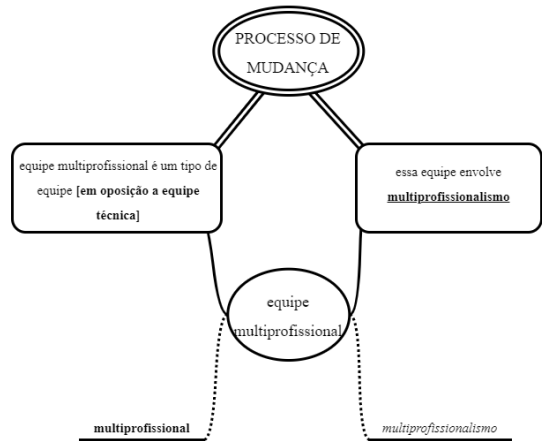
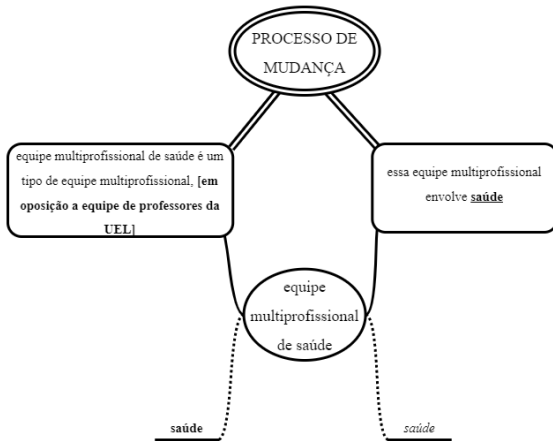
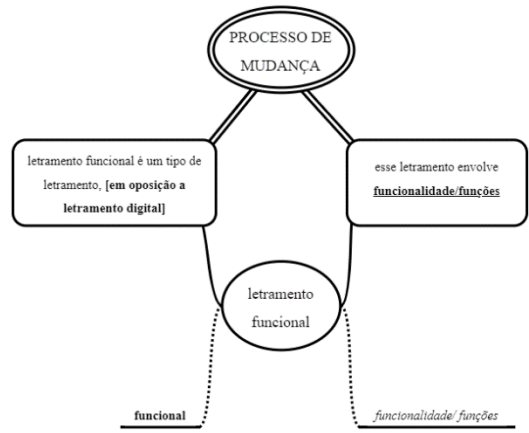
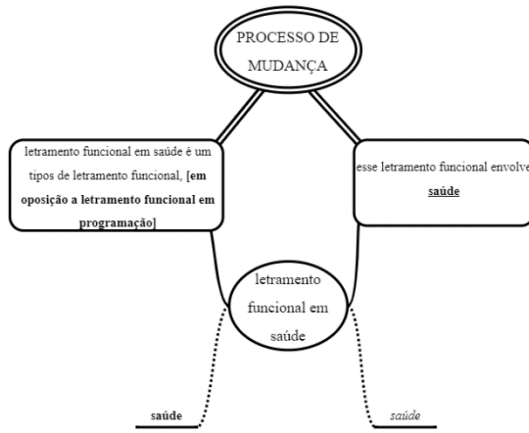


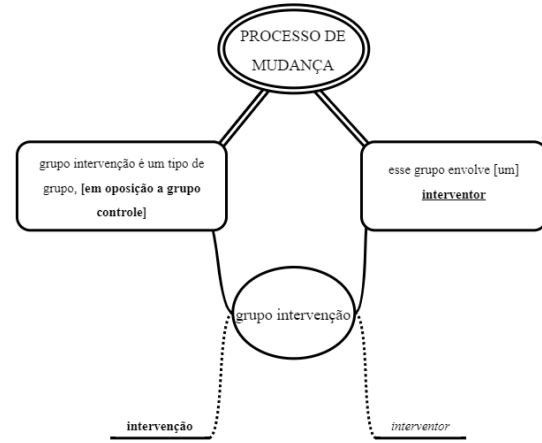
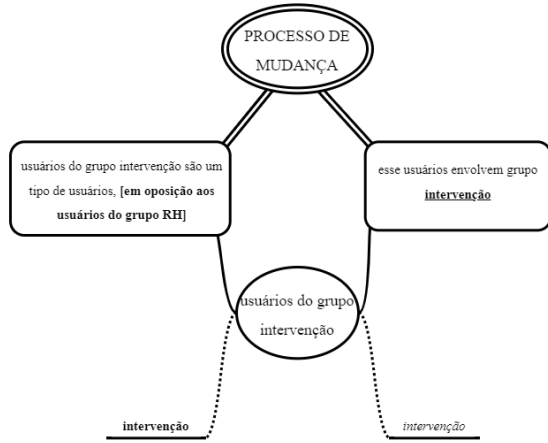
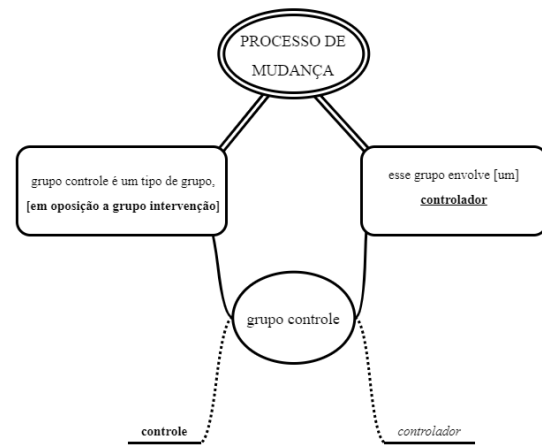
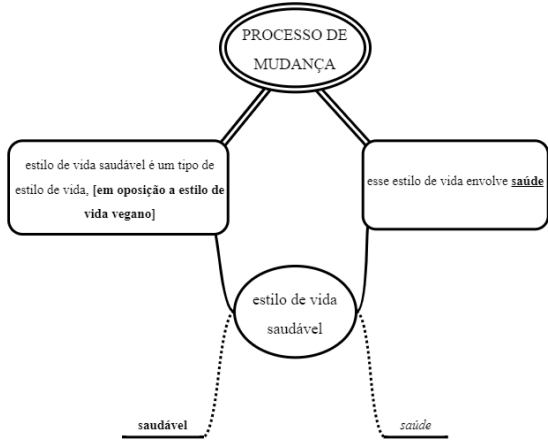




8.4.2 Representação ProtinLex:PTG – Resultado Default + Desvio







8.4.3 Representação ProtinLex:PTG – Resultado Desvio

