

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS HUMANAS E NATURAIS
DEPARTAMENTO DE LÍNGUAS E LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM ESTUDOS
LINGUÍSTICOS**

VANGEVALDO CARDOSO DOS SANTOS

**DESCRIÇÃO DE EXPRESSÕES FIXAS DO PORTUGUÊS BRASILEIRO
PARA PROCESSAMENTO AUTOMÁTICO DE LINGUAGEM NATURAL**

**VITÓRIA
2011**

VANGEVALDO CARDOSO DOS SANTOS

**DESCRIÇÃO DE EXPRESSÕES FIXAS DO PORTUGUÊS BRASILEIRO
PARA PROCESSAMENTO AUTOMÁTICO DE LINGUAGEM NATURAL**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Estudos Linguísticos do Centro de Ciências Humanas e Naturais, da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Estudos Linguísticos na área de concentração em Estudos Analítico-Descritivos da Linguagem.

Aprovada em 31 de março de 2011.

COMISSÃO EXAMINADORA:

Prof^a. Dr^a. Aucione Das Dores Smarsaro – PPGEL – UFES
Orientadora

Prof. Dr. Oto Araújo Vale – UFSCar
Membro Titular Externo da Comissão Examinadora

Prof. Dr. Alessandro Meireles – PPGEL - UFES
Membro Titular Interno da Comissão Examinadora

Prof. Dr. Orivaldo Tavares – Dept^o. Informática - UFES
Membro Suplente Interno da Comissão Examinadora

Dados internacionais de catalogação

Dedico esta dissertação à minha família de coração, Dona Laís, Marcelo e Mônica Fiorotti, pelo excesso de amor e confiança que creditam a mim. Isso me faz uma pessoa afortunada e feliz para seguir adiante.

Agradecimentos

De início, quero agradecer à Prof^a. Dr^a. Aucione Smarsaro, pois com suas fantásticas orientações e apontamentos, me apresentou o então desconhecido mundo da pesquisa, em que pude conhecer e me apaixonar pela linguística computacional, direcionando meus estudos para esse árduo e magnífico campo.

Ao professor Dr. Eric Laporte, pelas ótimas sugestões, questionamentos e apontamentos durante a qualificação, que me fizeram rever e melhorar a pesquisa.

Ao Marcelo Fiorotti, amigo e irmão do peito, pela paciência e excelentes leituras corretivas e sugestivas que iniciou na graduação e se estendeu a esta pesquisa de Pós-Graduação, meu muito obrigado.

Aos meus professores e colegas de estudo que, em todos os descontraídos encontros, me fez rever e melhorar alguns conceitos sobre linguagem e, com isso, me ajudou a entender um pouco sobre o que é aprendizagem.

Ao amigo Carlos Roberto de Souza Rodrigues, também pesquisador da área, pelas ótimas contribuições para o exame de qualificação.

A todos os meus amigos, em especial, a Luciano Debortolli, pela força e companheirismo em momentos difíceis.

À Fundação de Apoio à Ciência e Tecnologia do Espírito Santo (FAPES), pelo suporte financeiro através da bolsa de mestrado para que eu pudesse ter mais tempo e dedicação a esta pesquisa.

Um especial agradecimento aos meus irmãos Gigi, Rubens e Eder. Amo muito vocês. Sei que representamos uma família.

Enquanto não houver uma ligação física entre o cérebro e a máquina, o texto continuará nos oferecendo o caminho mais direto entre a mente e o mundo exterior.

(Paul Saffo)

Estrategista, professor da Stanford University e fundador do Board Científico da Samsung

Resumo

Esta pesquisa propõe um estudo das expressões fixas do português do Brasil, em especial, aquelas com nomes de *partes do corpo*, identificadas em dicionários e textos jornalísticos. A descrição e a formalização dessas expressões fixas são de suma importância para o Processamento Automático de Linguagem Natural. Selecionou-se um *corpus* com 558 expressões fixas, 351 são de estruturas argumentais **verbais** e 207 de estruturas argumentais **substantivais**. Essas expressões são analisadas a partir de critérios linguísticos formais, para se identificar o comportamento morfossintático-semântico de cada uma delas e, a partir dessas propriedades estruturais, observadas em contexto de uso, propõe-se uma codificação para formalizá-las, com intuito de serem inseridas em um ambiente computacional, ampliando os estudos descritivos sobre o léxico do português e corrigindo falhas que possam prejudicar a qualidade de um texto processado automaticamente. A representação formal dessas expressões fixas é essencial para que se possa incluí-las em um dicionário eletrônico e, assim, utilizá-las pelo software Unitex ou por qualquer outro analisador morfossintático.

Palavras-chave: Expressões fixas; Linguística computacional; Léxico-gramática.

Abstract

This research proposes a study of fixed expressions of Portuguese in Brazil, especially those with names of body parts, identified in newspaper texts and dictionaries. The description and formalization of these fixed expressions are of utmost importance to Automatic Processing of Natural Language. We selected a corpus with 558 fixed expressions, 351 are of verbal argument structures and 207 nouns argument structures. These expressions are analyzed from formal linguistic criteria to identify the morphosyntactic-semantic behavior of each one and, from these structural properties observed in the context of use, we propose an encoding to formalize them, with the aim to be inserted into a computing environment, expanding the descriptive studies on the lexicon of Portuguese and correcting deficiencies that might impair the quality of a text processed automatically. The formal representation of these fixed expressions is essential so that we can include them in an electronic dictionary, and thus use them at Unitex software or any other morphosyntactic analyzer.

Keywords: fixed expressions, Computational Linguistics, Lexicon-grammar.

Simbologia

PLN	Processamento Automático de Linguagem Natural.
PB	Português do Brasil.
C _{1pc}	Complemento fixo direto com nome de parte do corpo.
N	Substantivo.
N ₁	Substantivo pertencente ao complemento livre após o complemento fixo.
N ₀	Sujeito livre em posição argumental no enunciado.
V	Verbo pertencente à estrutura da expressão fixa.
+	Representa a ocorrência de determinada propriedade nas tábuas.
-	Representa a não ocorrência de determinada propriedade nas tábuas.
< >	Representa a palavra que é possível de flexão, que no caso são os verbos.
C=:Npc	Referente a todas às expressões fixas definidas como partes do corpo, que pode ser de humano ou animal, mas sempre se refere ao ser humano como, por exemplo, a expressão fixa abaixo cair nas garras :

Eder **caiu nas garras** de Marcelo.

A palavra **garra**, que é parte de corpo de um animal, nessa expressão fixa foi usada como sendo parte de corpo de um humano, e por ser convencionalizada como uma expressão, não há nenhum problema quanto à sua aceitabilidade.

Nhum	Substantivo humano, que pode ser representado tanto pelo sujeito livre no enunciado como pelo complemento livre, após o complemento fixo.
N-hum	Substantivo não humano, que pode ser representado tanto pelo sujeito livre como pelo complemento livre, após o complemento fixo no enunciado.
Prep	Preposição
Art	Artigo
Adj	Adjetivo.
N1	Primeiro substantivo de uma estrutura argumental NprepN.
N2	Segundo substantivo de uma estrutura argumental NprepN.
N0	Expressa um sujeito livre em posição argumental no enunciado presente nas estruturas argumentais substantivais NAdj e NprepN.
+pc	Parte do corpo
+ser	Indica a expressão fixa, de estrutura argumental NAdj, funciona como uma totalidade com relação ao N0 do enunciado.
+ter	Indica a expressão fixa, de estrutura argumental NAdj, funciona apenas como uma parte de corpo para o N0.
+Pred	Indica que a expressão fixa apresenta Predicatividade do adjetivo.
-Pred	Indica que a expressão NÃO apresentar Predicatividade do adjetivo.
+AdjC	Aceita coordenar o adjetivo da expressão fixa com uma estrutura argumental NAdj.
-AdjC	Não aceita coordenar o adjetivo da expressão fixa com uma estrutura argumental NAdj.

+EAdj	Há possibilidade de elidir o adjetivo de uma expressão fixa NAdj.
-EAdj	Não há possibilidade de elidir o adjetivo de uma expressão fixa NAdj.
+N1Pron	Indica que é possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas.
-N1Pron	Indica que NÃO é possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas.
+n	A expressão fixa permite variar em número.
-n	A expressão fixa não permite variar em número.
:fs	Estrutura da expressão fixa substantival é feminina no singular.
:fp	Estrutura da expressão fixa substantival feminina no plural.
:ms	Estrutura da expressão fixa substantival masculina no singular.
:mp	Estrutura da expressão fixa substantival masculina no plural.
+1	Expressão fixa, com propriedade N1deN2 ser N1.
-1	Expressão fixa não tem a propriedade N1deN2 ser N1.
+2	Expressão fixa, com propriedade N1deN2 ser N2.
-2	Expressão fixa não tem a propriedade N1deN2 ser N2.
-1-2	Expressão fixa com propriedade N1deN2 não ser nem N1 nem N2.
+N1Pron	Indica que é possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas.
-N1Pron	Indica que NÃO é possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas.
+N1E	Possibilidade de elidir N1.
-N1E	Não há possibilidade de elidir N1.

- +N2E Possibilidade de elidir N2.
- N2E Não há possibilidade de elidir N2.
- +v Possibilidade de variação em número da expressão fixa, com estrutura argumental NprepN.
- v Não há possibilidade de variação em número da expressão fixa, com estrutura argumental NprepN.
- * Representa uma sequência inaceitável em consonância ao sentido da expressão fixa em relação às restrições impostas por essa expressão. Alguns exemplos de expressões passam por transformações no enunciado para o julgamento de aceitabilidade por falantes nativos, como uma expressão fixa. Quando determinada sequência vier acompanhada desse símbolo, pode não haver nenhum problema estrutural como um grupo nominal livre, mas há problema de aceitabilidade como uma expressão fixa.
- AT Referente ao Jornal *A Tribuna*, veículo de comunicação local escolhido como fonte de pesquisa para se constatar o uso das expressões fixas na modalidade escrita.

Lista de tabelas, figuras e anexos

Tabela 1 - Tabela representativa das expressões fixas, de estrutura argumental verbal	25
Tabela 2 - Tabela representativa das expressões fixas, de estrutura argumental substantival	26
Tabela 3 - Etapas de construção de um sistema de PLN	40
Tabela 4 - Evolução dos sistemas de PLN	40
Tabela 5 - Representação da estrutura argumental verbal	80
Tabela 6 - Representação da estrutura argumental substantival	113
Tabela 7 - Classificação das expressões fixas	151
Tabela 8 - Representação das classes verbais	176
Figura 1 – Grafo usado para ajudar na resolução de ambiguidade	145
Figura 2 – Grafo representando uma ambiguidade	145
Figura 3 – Grafo representativo dos traços morfológicos de uma sequência	146
Figura 4 – Fragmento da Tábua PB-C1, representando a estrutura argumental $N_0 V C_{1pc}$	152
Figura 5 – Fragmento da Tábua PB-C1PN, representando a estrutura argumental $N_0 V C_{1pc} Prep N$	153
Figura 6 – Fragmento da Tábua PB-CA1, representando a estrutura argumental $N_0 V Art C_{1pc}$	154

Figura 7 – Fragmento da Tábua PB-CA1, representando a estrutura argumental $N_0 V \text{ Art } (C \text{ de } N)_{1pc}$	156
Figura 8 – Fragmento da Tábua PB-CAP1PN, representando a estrutura argumental $N_0 V \text{ Art } C_{1pc} \text{ Prep } N$	158
Figura 9 – Fragmento da Tábua PB-CP1, representando a estrutura argumental $N_0 V \text{ Prep } C_{1pc}$	159
Figura 10 – Fragmento da Tábua PB-CPN, representando a estrutura argumental $N_0 V \text{ Prep } (C \text{ Prep } N)_{1pc}$	160
Figura 11 – Fragmento da Tábua PB-CP1PN, representando a estrutura argumental $N_0 V \text{ Prep } C_{1pc} \text{ Prep } N_1$	161
Figura 12 – Fragmento da Tábua PB-CNP2, representando a estrutura argumental $N_0 V N_1 \text{ Prep } C_{1pc}$	162
Anexo I – Lista das tábuas das expressões fixas verbais formalizadas	189
Anexo II – Lista das expressões fixas substantivais formalizadas, com estrutura argumental $N \text{ Adj}$	225
Anexo III – Lista das expressões fixas substantivais formalizadas, com estrutura argumental $N \text{ de } N$	229

Sumário

1	INTRODUÇÃO	19
1.1	Apresentação do tema	20
2	LÉXICO-GRAMÁTICA	28
2.1	Pressupostos teóricos	28
2.2	Base teórica	28
2.3	Distribuição sintática dos elementos lexicais	30
2.4	Formalização	30
2.5	Resultados	31
3	O LÉXICO	34
3.1	Conceito e abordagem	34
3.2	A aquisição do léxico por humanos	35
3.3	A aquisição do léxico por computador	37
3.3.1	Processamento de linguagem natural (PLN)	39
3.4	Itens lexicais – diferentes tipos	42
3.4.1	Itens lexicais complexos	42
3.5	O que é uma palavra para o PLN	44
3.6	O léxico do português brasileiro	47

3.7	A relevância do léxico para o PLN	49
3.8	A relevância de dicionário eletrônico de expressões fixas para PLN	49
3.9	Léxico computacional ou dicionário eletrônico	51
4	ABORDAGEM SOBRE AS EXPRESSÕES FIXAS	54
4.1	Abordagens em literaturas diversas	54
4.2	Composicionalidade / não-composicionalidade	64
4.3	O uso do hífen nas expressões fixas	72
4.4	Abordagem estabelecida pela teoria do Léxico-gramática	73
5	CRITÉRIO DE IDENTIFICAÇÃO DAS EXPRESSÕES FIXAS	79
5.1	Expressões fixas de estruturas argumentais verbais	79
5.1.1	Critérios utilizados referentes às propriedades morfossintáticas e semânticas para a identificação das expressões fixas verbais	81
5.1.2	VartN	82
5.1.2.1	Distribuição sintática dos itens lexicais	82
5.1.2.2	Substituição do verbo por outro verbo	84
5.1.2.3	Determinante obrigatório	85
5.1.2.4	Negação obrigatória	87
5.1.2.5	Inserção de um elemento lexical	89
5.1.2.6	Flexão em número de N	90
5.1.3	VprepN	91
5.1.3.1	Distribuição sintática dos itens lexicais	92

5.1.3.2	Negação obrigatória	94
5.1.3.3	Substituição do verbo por outro verbo	95
5.1.3.4	Inserção lexical	97
5.1.3.5	Preposição obrigatória	98
5.1.3.6	Variação em número do N	99
5.1.3.6.1	Plural obrigatório de N	100
5.1.4	Estrutura argumental VN	100
5.1.4.1	Distribuição sintática dos itens lexicais	103
5.1.4.2	Substituição do verbo com outros verbos	105
5.1.4.3	Inserção lexical	106
5.1.4.4	Negação obrigatória	108
5.1.4.5	Ausência de determinante ou preposição	109
5.1.4.6	Variação em número de N	110
5.2	Expressões fixas de estruturas argumentais substantivais	113
5.2.1	Critérios de identificação das expressões fixas	113
5.2.2	Estrutura argumental NAdj	113
5.2.2.1	Perda da predicatividade do adjetivo	115
5.2.2.2	Coordenação do adjetivo com outro adjetivo	117
5.2.2.3	Elisão do adjetivo	118
5.2.2.4	Ruptura paradigmática	120
5.2.2.5	Variação em número	121
5.2.3	Estrutura argumental NprepN	123
5.2.3.1	Coordenação de grupos nominais	124

5.2.3.2	Elisão de um dos elementos	126
5.2.3.3	Inserção de um elemento lexical	128
5.2.3.4	Ruptura paradigmática	130
5.2.3.5	Variação em número	131
6	A AMBIGUIDADE	133
6.1	Conceituando	133
6.2	O problema da ambiguidade para o PLN	136
6.3	Resolução de ambiguidade	138
6.3.1	Resolução de ambiguidade nas expressões fixas não-composicionais	141
7	FORMALIZAÇÃO DAS EXPRESSÕES	149
7.1	Formalização das propriedades das expressões fixas verbais	149
7.1.1	Tábuas – alguns comentários	150
7.2	Codificação das descrições das propriedades das expressões fixas substantivais de estruturas argumentais NAdj e NdeN	163
7.2.1	Propriedades das estruturas argumentais NAdj	163
7.2.2	Estrutura argumental NdeN	169
8	CONCLUSÕES E CONTRIBUIÇÕES	173
9	REFERÊNCIAS	179

Introdução

Desde a invenção do computador na década de 40, o homem tenta, com muito esforço, desenvolver programas aptos para processar informações de maneira eficiente. Com o passar do tempo, os computadores tornaram-se ferramentas indispensáveis na vida das pessoas, pois estão presentes em quase todos os setores das atividades humanas. Por isso, de acordo com Youssef e Fernandez (1988, p. 11), o desenvolvimento da atividade humana exigiu sempre que o homem criasse dispositivos de registros e processamento da informação como forma de interferir no, adaptar-se ao e controlar o meio em que vive. Assim, segundo Silva (2006, p. 103), o homem passa por um desafio, que é criar meios para tornar a comunicação homem-máquina mais natural e intuitiva. Desse desafio nasce, portanto, a Linguística Computacional¹.

Na Linguística Computacional, o homem procura fazer uma interação entre a linguística e a informática². Para Othero (2006, p. 341), a Linguística computacional é a parte da ciência linguística que se preocupa com o tratamento computacional da linguagem. A Linguística Computacional surge como uma área responsável em investigar a possibilidade de um tratamento computacional tanto da linguagem como das línguas naturais.

De acordo com Othero (2006, p. 342), a Linguística Computacional envolve diferentes áreas de pesquisa tradicionalmente conhecidas em Linguística Teórica e Aplicada, como por exemplo, a Sintaxe, a Semântica, a Fonética, a

¹ O nome Linguística Computacional foi cunhado em 1967 por David Hays. Antes da Linguística Computacional se especializar enquanto uma disciplina que focalizava alguns aspectos do estudo computacional das línguas naturais, focalizava essencialmente o estudo das linguagens formais e das linguagens de programação (FERNANDES *apud* SILVA, 2005, p. 129).

² Segundo Youssef e Fernandez (1988, 25-26), a palavra informática surgiu na década de 60 como uma nova ciência, e esse termo informática foi criado pela junção dos vocábulos informação e automática.

Fonologia, a Pragmática etc.; e todo esse conhecimento é utilizado para tentar processar as línguas naturais e dominar o conhecimento linguístico envolvido no domínio de uma língua natural.

Na tentativa de se fazer uma comunicação entre o homem e a máquina, por meio da elaboração de programas que tenham a capacidade de processar e manipular, com eficiência, cada vez mais, diversos tipos de informações dadas pelo homem, nasce, assim, o Processamento Automático de Linguagem Natural (PLN). Em Othero (2006, p. 343), o PLN pode ser entendido como uma subárea da Linguística Computacional, e se preocupa diretamente com o estudo voltado para a construção de *softwares*. Também cabe à área do PLN a construção de programas capazes de interpretar e/ou gerar informações em linguagens naturais.

Sendo assim, por causa dos vários aspectos da língua natural, sempre foi um grande desafio, tanto para os linguistas como para os informatas, criar softwares eficientes que trabalhem com a língua natural para uma língua artificial, capaz de ser executável em um ambiente computacional.

1.1 Apresentação do tema

Esta pesquisa apresenta um estudo descritivo e formalizado de expressões fixas do português brasileiro, constituído de **estrutura argumental Verbal e substantival**, com *nomes de partes do corpo*, para PLN.

Segundo Makkai (1978), as expressões fixas são numerosas em qualquer língua, e a fixidez é uma característica universal das línguas naturais (GROSS, 1988). As expressões fixas formam palavras novas e constituem-se em estruturas lexicais bastante produtivas para a expansão e o funcionamento delas na linguagem, servindo-nos para comunicação cotidiana. Essa fixidez se dá a partir do momento em que os significados das palavras não podem ser entendidos pela soma dos seus constituintes. Por exemplo, em Eder **bateu perna** no shopping, deve ser interpretado como: Eder **andou muito**. Chegou-se a esse entendimento porque o significado da unidade lexical **bater perna** não foi deduzido da soma dos valores das palavras que compõem tal unidade,

constituindo, assim, uma expressão fixa. Assim sendo, não se tem em uso Eder *bateu braço*.

Para Laporte (2009), as unidades de multi-palavras fixas ou palavras compostas são expressões feitas tipograficamente de várias palavras, mas que linguisticamente requerem a descrição de uma entrada lexical específica, como em ***keep in mind*** (ter em mente) ou ***hit the jackpot*** (tirar a sorte grande). Ainda segundo o autor, informações sobre tais palavras são fundamentais. Gross (1986) também salienta a importância informacional das palavras para o PLN, pois para o autor, as aplicações em linguagem de processamento mais elaboradas do que sistemas atuais exigem uma grande cobertura informacional das palavras descritas.

Em consonância com Laporte, a importância das informações sobre as expressões fixas se dá porque cada expressão fixa possui suas peculiaridades. De acordo com Tagnin (1989), é necessário conhecer as combinações lexicais que não estejam baseadas em relações de significados. Nesse sentido, a expressão ***olho-de-sogra*** não pode ser interpretada por relações de significados. Se isso acontecer, tem-se um problema na interpretação, pois sua soma é constituída por uma combinação lexical metafórica, como afirma Basílio (2004). Para a autora, nas expressões fixas ocorre uma situação de nomeação metafórica quando a descrição de um objeto, ao invés de caracterizá-lo por critérios objetivos, estabelece para este uma descrição em termos de propriedades transferidas em termos associativos. De acordo com Vale (2001, p. 3-4),

[...] essas expressões são um problema de peso quando olhadas com mais vagar. No aspecto sintático, as expressões cristalizadas³ são frases aparentemente normais, semelhantes às frases "comuns" (i.e. sem expressões cristalizadas), com raras diferenças formais. No aspecto semântico, elas são, em geral, imediatamente interpretadas pelos falantes com seu significado idiomático, e não pela soma dos sentidos literais das palavras que as compõem.

Os enunciados abaixo vêm corroborar com as idéias de Tagnin, Basílio e Vale:

- (1) Eder teve uma tremenda *dor de cabeça* após a bebedeira com os amigos.

³ O autor caracteriza as expressões fixas como expressões cristalizadas.

- (2) Eder teve uma tremenda **dor de cabeça** para resolver a prova de matemática.

No exemplo (1), a interpretação para *dor de cabeça* é literal, Eder realmente teve dor de cabeça, que na medicina pode-se chamar de cefaléia. Já no exemplo (2), a interpretação é puramente metafórica. Qualquer falante nativo brasileiro, que tem dificuldades em resolver questões de matemática, entenderia que se trata de uma preocupação excessiva, uma dificuldade com a prova de matemática, o que a caracteriza como uma expressão fixa. Segundo Smarsaro (2004), essas características são extremamente relevantes. É a partir dessas observações que este estudo vai se atentar para melhor representá-las. Isso porque, para o PLN, há uma grande necessidade em se distinguir diferenças entre um grupo nominal que é livre de uma expressão fixa. É fundamental entender que a expressão fixa é sempre interpretada como um único item lexical.

Tagnin (1989, p. 45), acrescenta que quando uma forma linguística foi convencionalizada entende-se que tal forma linguística passou a ter um significado distinto do significado de seus constituintes. Por isso, entende-se por convenção tudo aquilo que é aceito por todos em uma determinada língua sempre com o mesmo sentido. Então, a expressão fixa **pé-de-meia** é semanticamente convencionalizada, porque se entende que o significado global refere unicamente à economia de uma reserva para eventuais necessidades.

Garrão (2001, p. 2) argumenta que dentre os vários problemas linguísticos com os quais um programa de tradução se depara, há uma questão particularmente relevante, que é a de reconhecimento e gerações de expressões cristalizadas. Por isso, a descrição das estruturas e de suas propriedades morfossintáticas e semânticas são condições para que elas possam ser inseridas em um dicionário eletrônico.

São essas peculiaridades que reclamam por um estudo descritivo mais pormenorizado, tendo em vista que, para se atingir a interpretação semântica dessas expressões é preciso investigar os comportamentos morfológicos, sintáticos e semânticos delas a partir de situações de uso na língua. Esse

estudo é imprescindível para desfazer problemas de interpretação, como por exemplo, em estruturas idênticas como o caso de *pé de meia* e ***pé-de-meia***, para não gerar falhas ao PLN.

O estudo dessas expressões se faz necessário porque um dos grandes problemas para o avanço no desenvolvimento de programas na área de PLN é a ausência de pesquisas com descrição do português do Brasil para inserção em um ambiente computacional, tendo em vista, principalmente, alta produtividade das expressões no léxico das línguas naturais (GROSS, 1996) e o uso frequente delas no dia-a-dia (SEHELLART, 1998).

Para o estudo das expressões fixas, parte-se da seguinte hipótese: é possível definir critérios formais os quais nos apontam que, no léxico do português brasileiro, as palavras designando *nomes de partes do corpo* originam dezenas de expressões fixas, formando expressões usuais em quantidades surpreendentes. A descrição de estruturas lexicais de palavras como ***mão, boca, olhos, cabeça e pé*** são de grande relevância para se elaborar um dicionário eletrônico de expressões fixas. A partir do estudo descritivo e formalizado dessas expressões, propõe-se uma codificação das propriedades estruturais, possibilitando o seu uso em um banco de dados para processamento automático.

Com este fim, esta pesquisa está ancorado na teoria do Léxico-gramática, desenvolvida pelo linguista Maurice Gross (1975, 1976, 1981, 1982, 1984, 1986a, 1986b, 1986c, 1988, 1989, 19994; BOONS; GUILLET; LECLÈRE, 1976) na década de 60, juntamente com a equipe do *Laboratoire d'Automatique Documentaire et Linguistique* (LADL). A base teórica que fundamenta o Léxico-gramática é o Distribucionalismo, oriundo da Gramática Transformacional de Harris (1964, 1968, 1976, 1991) e Gross (1975, 1977, 1981, 1988).

Dessa forma, nesta pesquisa, o estudo das expressões fixas se baseia na metodologia da teoria do Léxico-gramática, de Maurice Gross (1975). A metodologia do Léxico-gramática foi elaborada numa perspectiva de tratamento automatizado da língua e se propõe estabelecer um inventário de informações linguísticas explícitas, precisas e exaustivas (Smarsaro, 2004, p. 360). Cada expressão será examinada em um contexto de uso para se avaliar a

distribuição sintática dos itens lexicais, por meio das transformações das frases, conforme orienta a teoria do Léxico-gramática. É importante ressaltar que as qualidades das transformações nas frases são fundamentais para explicar o sentido de cada item lexical a partir de cada transformações sintáticas.

A análise das estruturas lexicais candidatas a expressões fixas por meio dos critérios sintático-semânticos, se baseia no julgamento e aceitabilidade dos falantes nativos. Conforme salienta Smarsaro (2004, p. 36), durante a análise dessas expressões, são detectadas ambiguidades lexicais relativas às propriedades gramaticais. As entradas seguidas por esse processo de análise serão classificadas e codificadas de modo a especificar as propriedades gramaticais e morfológicas essenciais: número, variações eventuais e formas verbais flexionadas.

Através da codificação das expressões fixas, é possível inseri-las em um software como o Unitex⁴ para o tratamento de textos em língua natural.

Inicialmente, fez-se um levantamento de 1.500 expressões fixas com *nomes de partes do corpo*. Para a recolha dessas expressões, utilizou-se o Dicionário Houaiss da Língua Portuguesa (impresso e digital), o Dicionário Enriqueça seu vocabulário, o Pequeno Vocabulário da Língua Portuguesa e alguns jornais, tais como: A Tribuna. Apesar de a coleta do *corpus* em dicionários ter sido relevante para a seleção, não foi suficiente, já que a quantidade de expressões fixas encontradas não corresponde à diversidade verificada no uso da língua. Por isso, a pesquisa também contou com ajuda da competência linguística de alguns falantes nativos da cidade de Vitória, ES, por meio de exemplos em uso, como por exemplo, ***meter a testa, boca doce e cintura de pilão***.

⁴ “O Unitex é um conjunto de programas que possibilitam a tratamento de textos em língua natural utilizando recursos linguísticos. Esses recursos encontram-se sob a forma de dicionários eletrônicos, gramáticas e tábuas de Léxico-gramática e tem origem nos trabalhos desenvolvidos pelo Linguista Maurice Gross no Laboratoire d’Automatique et Linguistique (LADL). A criação desse software tem como objetivo descrever as palavras simples e compostas de uma língua, associando a cada uma um lema e uma série de códigos gramaticais, semânticos e flexionais” (PAUMIER, 2002, p. 2). Diferente dos outros dicionários usuais, o dicionário proposto no Unitex se define como um programa de processamento de texto com dados linguísticos.

Os exemplos utilizados nesta pesquisa foram selecionados e organizados de duas maneiras:

- a) exemplos fabricados, em que não há indicação de fonte;
- b) exemplos de ocorrência retirados do jornal A Tribuna (AT-21/08/05), representado entre parênteses após o enunciado pela sigla AT, seguido da data da ocorrência.

As expressões fixas em discussão estão em **negrito** e *itálico*. Isso foi feito tanto nos exemplos fabricados quanto nos exemplos retirados do Jornal pesquisado (AT). Já as sequências livres estão representadas apenas em *itálico*. A finalidade é tão somente para melhor diferenciar as sequências livres das sequências fixas, que se caracterizam como expressões fixas.

A grande diversidade de formas encontradas resultou em uma variada distribuição estrutural. As expressões fixas com *nomes de partes do corpo* foram classificadas com base nos elementos que as constituem como formações verbais (***abrir a cabeça***), substantivais (***cara-de-pau***), adjetivais (***frio na barriga***) e locucionais (***a olho nu***). Dentre as várias classes encontradas, foi necessário fazer um recorte para atender os objetivos propostos nesta pesquisa, tendo como meta de análise o estudo de 351 expressões com estrutura argumental **verbal** e 207 com estrutura argumental **substantival**, por dois motivos:

- 1- alta produtividade de expressões fixas no português;
- 2- recorrência substancial das estruturas argumentais verbais e substantivais em relação às outras estruturas.

As estruturas:

Estrutura argumental verbal	Exemplo	Nº ocorrências
V+art+N	<i>Esfriar a cabeça</i>	175
V+prep+N	<i>Entrar de cabeça</i>	112
V+N	<i>Bater-boca</i>	64
Total		351

Tabela 1. Tabela representativa das expressões fixas, de estrutura argumental verbal.

Estrutura argumental substantiva	Exemplo	Nº ocorrências
N+Adj	<i>barba-azul</i>	99
N+prep+N	<i>Olho de sogra</i>	108
Total		207

Tabela 2. Tabela representativa das expressões fixas, de estrutura argumental substantiva.

As expressões selecionadas são analisadas através de uma descrição mais pormenorizada das suas propriedades, obedecendo alguns critérios sintático-semânticos para identificá-las como expressões fixas, pois, a partir do estudo das expressões fixas, com nomes de *partes do corpo*, ancorado em critérios formais, torna-se possível fazer um quadro descritivo e representativo das estruturas argumentais verbais e substantivas.

Para processamento automático de qualquer estrutura linguística é necessário que se faça antes uma descrição e, a partir daí, a formalização. Essa formalização viabiliza a possibilidade de implementações e aplicações computacionais. Portanto, o mérito da pesquisa consiste não só em contribuir para os estudos linguísticos, como também é fundamental para a realização de trabalhos na área de PLN, amenizando problemas de ambiguidade lexical e estrutural, preenchendo lacunas de sentido que possam tornar os textos imprecisos e incoerentes.

O objetivo desta pesquisa é, portanto, descrever e formalizar as expressões fixas selecionadas. Esse intento pode ser resumido em:

- a) observar quais são os fatores que interferem e determinam o uso dessas sequências como uma expressão fixa;
- b) observar restrições quanto às propriedades sintáticas e semânticas das expressões fixas dentro de um contexto oracional;
- c) desfazer possíveis ambiguidade entre as mesmas estruturas fixas e livres;
- e) codificar as expressões identificadas como fixas para que possam ser inseridas em um dicionário eletrônico, com intuito de utilizá-lo por sistemas de PLN.

Esta dissertação está organizada da seguinte forma: no Capítulo 2, aborda-se a Teoria do Léxico-gramática como sustentação teórico-metodológica para a descrição e formalização das expressões fixas. No Capítulo 3, apresenta-se um estudo sobre o léxico e sua importância para sistemas computacionais. O Capítulo 4 apresenta uma revisão de literatura sobre as expressões fixas e como são abordadas. No Capítulo 5, descreve-se critérios com propriedades morfológicas, sintáticas e semânticas para identificação das expressões fixas. No Capítulo 6, discute-se o problema da ambiguidade recorrente entre as expressões fixas e as unidades lexicais livres com a mesma estrutura e como desfazer essa ambiguidade. No Capítulo 7, apresenta-se uma codificação, conforme a Teoria do Léxico-gramática, das expressões fixas analisadas e descritas, a fim de inseri-las em um dicionário eletrônico. No Capítulo 8, apresenta-se conclusões desta pesquisa, bem como sua relevância para o PLN.

Léxico-gramática:

Revolução nos estudos da Linguística Computacional

2.1 Pressupostos teóricos

O estudo das expressões fixas, com nomes de *partes do corpo*, tem como base a Teoria do Léxico-gramática. Desde a década de 1960, linguistas vêm pesquisando recursos linguísticos para sistema de processamento de linguagem natural. O linguista Maurice Gross (1975, 1981, 1994), juntamente com a equipe do *Laboratoire d'Automatique Documentaire et Linguistique* (LADL), deu um grande passo ao desenvolver a Teoria do Léxico-gramática, cujo propósito é designar uma metodologia e uma prática efetiva da descrição manual sintático-semântica do léxico numa determinada língua.

Para Smarsaro (2004, p. 28), o Léxico-gramática

[...] tem como meta investigar os procedimentos lexicais e gramaticais que levam ao reconhecimento de padrões de palavras nos quais está baseado o processo de entendimento para representação das propriedades de sequências linguísticas computacionalmente.

A metodologia do Léxico-gramática é baseada nas ciências experimentais (LAMIROY, 2003), que têm como enfoque a coleta dos fatos, conferindo com a realidade dos usos linguísticos, tanto do ponto de vista quantitativo (com relação à descrição metódica do léxico) quanto do ponto de vista qualitativo (quando se leva em consideração os cuidados metodológicos).

2.2 Base teórica

A base teórica que fundamenta o Léxico-gramática é o Distribucionalismo, oriundo da Gramática Transformacional do linguista Zellig Harris (1964, 1976)

que, de acordo com (GROSS, 1991, p.2), a construção do Léxico-gramática teve como ponto de partida a Teoria da Gramática Transformacional de Harris.

Com base no referencial teórico de Harris e, adicionando a exigência de proporcionar uma cobertura significativa da linguagem, levou ao desenvolvimento de:

- uma extensa lista de transformações;
- um léxico de verbos (ou sentenças elementares), para que estas transformações fossem aplicadas.

Segundo Smarsaro (2004, p. 26), na Gramática Transformacional de Harris, o objeto central da sintaxe são as relações entre as frases. Essa gramática procura representar frases diretamente observáveis, capaz de serem julgadas como aceitáveis ou não pelos falantes de uma língua.

De acordo com a Teoria do Léxico-gramática, é por meio da distribuição sintática⁵ que se podem aplicar um método e uma prática estável e real de descrição formal dos itens lexicais numa frase em uma determinada língua. Pois é a partir dessa prática que se tem a noção de transformação e as implicações que dela decorrem, o que se permite o julgamento de aceitabilidade⁶ ou não de frases como fonte básica para o conhecimento linguístico. A unidade de significado é a frase, pois as palavras só adquirem um sentido preciso dentro dela. Quando se insere um item lexical ou uma palavra dentro de uma frase, tem-se a vantagem de manusear uma sequência passível de ser julgada aceitável ou inaceitável. Na frase é imprescindível que haja uma inter-relação entre o léxico e a sintaxe. O Léxico-gramática também lida com procedimentos empíricos para mostrar regularidades no uso da língua.

Em Laporte (2008, p. 36), para explorar a interação entre o léxico e a sintaxe, é necessário que se combine sistematicamente as entradas lexicais com todas as estruturas de frases observadas. Deve-se também observar e analisar se as sequências geradas são aceitáveis e quais suas particularidades distribucionais

⁵ Distribuição sintática corresponde a um método de análise linguística, por meio do qual os elementos da sentença em estudo são reutilizados em novas sentenças, com o intuito de verificar a ocorrência ou não da manutenção do sentido veiculado por cada elemento no exemplo fonte.

⁶ A aceitabilidade é calcada nos conhecimentos obtidos a partir de dialetos ou de socioletos atribuídos a determinadas comunidades linguísticas brasileiras.

e semânticas. Para o autor, a qualidade dos resultados depende muito da capacidade linguística do pesquisador em julgar aceitável uma dada sequência. Para obter resultados extremamente favoráveis, Laporte afirma que no Léxico-gramática é recomendável que o linguista aplique o método somente em sua própria língua materna.

2.3 Distribuição sintática dos elementos lexicais

Pode-se afirmar que a distribuição sintática é um tópico muito importante na Teoria do Léxico-gramática. Pois com a distribuição sintática, observa-se como cada elemento lexical se comporta nas frases. Para avaliar a fixidez de uma expressão, usando a Teoria do Léxico-gramática, faz-se necessário observar, em um enunciado, as distribuições sintáticas dos componentes dessa expressão e, a partir daí, verificar se essa distribuição conserva ou não o sentido da expressão.

Tem-se abaixo, como demonstração, a distribuição sintática dos elementos lexicais da expressão fixa ***perder a cabeça***:

- (1) Eder ***perdeu a cabeça*** na disputa pela bola.
- (1a) *A cabeça de Eder foi encontrada decapitada.
- (1b) *A cabeça de Eder está perdida.
- (1c) *Eder perdeu a cabeça e um braço na disputa pela bola.

A distribuição sintática é um critério que ajuda a mostrar se os elementos lexicais de uma expressão fixa mantêm o mesmo sentido ou não. Assim, a distribuição acima evidencia que quando são distribuídos sintaticamente os elementos da expressão fixa ***perder a cabeça***, perde-se o seu sentido porque os elementos que compõem tal expressão são interpretados como um único item lexical.

2.4 Formalização

O Léxico-gramática também exige que o *corpus* analisado e descrito seja formalizado. Os resultados da descrição dos itens lexicais de um *corpus* devem ser formalizados o suficiente para que possam ser inseridos em um ambiente

computacional por linguístas ou informatas⁷. Para Laporte (2008, p. 4) os resultados suficientemente formais permitem:

- uma verificação pela confrontação com a realidade do uso;
- uma aplicação ao tratamento automático das linguas.

O modelo conceitual determina que a partir dos resultados da descrição se criem tábuas chamadas de tabelas ou matrizes. Essas tabelas são “matrizes binárias que descrevem as propriedades de certas palavras (PAUMIER, 2006, p. 2)”. São elas que cruzam os itens lexicais com as propriedades morfossintático-semânticas, e os resultados obtidos constituem uma base de informações sintático-semânticas (GROSS, 1975).

2.5 Resultados

Desde a criação do método do Léxico-morfossintático-semântico, os resultados obtidos através deste método, converteu-se em uma excelente base de informações sintático-semânticas. Essa qualidade informacional só foi possível devido ao desempenho dos pesquisadores em observar a riqueza dos fenômenos linguísticos das línguas estudadas, avaliando um grande volume de itens, bem como suas propriedades e fazendo as devidas formalizações, para serem usadas em PLN.

No laboratório LADL já se desenvolveu várias pesquisas que produziram uma grande base de dados lexicográficos para várias línguas. Só em Francês (GROSS, 2006) já foram descritos mais de 75.000 itens. Segundo Gross (1994, p. 18), uma das principais vantagens do Léxico-gramática é permitir comparações entre as línguas nos níveis lexicais e sintáticos; e grandes segmentos do Léxico-gramática já foram construídos para o italiano (M. Martinelli, 1981; A. Elia, 1984; E. d'Agostino, 2001), para português europeu (J. Malaca Casteleiro, 1981; E. Macedo, 1984; RANCHHOD, 1990), para o espanhol (B. Lamiroy, 1983; C. Subirats, 1987; L. Masso-Pellat, 1990), que já permite comparações das línguas românicas.

⁷ Autores como Silva (2006, p. 117), preferem chamar de engenheiros da computação, outros preferem chamar de cientistas da computação. São profissionais que atuam na área das Ciências da Computação, dedicando-se aos estudos dos algoritmos, possibilitando suas aplicações e complementações como softwares para serem executados em computadores eletrônicos.

Gross também afirma que muitas classes de construções também foram descritas para o Inglês (M. Salkoff, 1983; P. Freckleton, 1985; P. Machonis, 1988), para o alemão (T. Treigg, 1977; F. Caroli, 1984), para o árabe (M. Chade, 1988; M. El hannach, 1988), para o coreano (Hong Chai-Song, 1984).

Há, ainda, de acrescentar outros trabalhos mais recentes do que os supracitados por Gross, que são: no português europeu (BAPTISTA, 1995; CARVALHO, 2001; RANCHHOD *et al*, 2004; BAPTISTA; FERNANDES, 2005; FERNANDES, 2007), no italiano (D'AGOSTINO *et al*, 2004), e no grego moderno (KYRIACOPOULOU, 2004).

Também existem estudos em outras línguas que ainda não têm a mesma representação como nas línguas citadas, mas que há um forte desempenho como, por exemplo, no holandês, no norueguês, no tailandês e, mais recentemente, no português brasileiro (ainda com poucas pesquisas), algumas ancoradas na Teoria do Léxico-gramática.

Para Gross (1994, p.18), todos esses trabalhos confirmam a possibilidade de construir Léxico-gramáticas de línguas variadas em uma forma comparável.

O estudo das expressões fixas, ancorado no quadro do Léxico-gramática, só começou a partir da década de 1980 (GROSS, 1982, 1986), (VALE, 2001; DANLOS, 1981; RANCHHOD, 1990, 2003) através de um programa metódico de descrição. Gross (1982) *apud* Vale (2001) estabelece para as expressões fixas verbais uma classificação que as separa de acordo com a estrutura interna: os constituintes fixos e os constituintes livres. Gross observa também que as expressões fixas são bem mais numerosas do que se supõe nos estudos linguísticos, que as tratam geralmente como exceções.

Segundo Vale (2001, p. 71), usando a metodologia do Léxico-gramática como trabalho,

[...] já foram descritas expressões fixas do árabe (BENKADDOUR, 1987), do coreano (LIM, 2000), do espanhol (MASSÓ, 1989), do grego (FOTOPOULOU, 1993), do inglês (MACHONIS, 1985) e do italiano (VIETRI, 1985), entre outros trabalhos.

Todo esse interesse dos linguistas em estudar as expressões fixas se dá pelo simples fato de que já é possível aferir *quantitativamente* a importância destas

expressões em diferentes gêneros textuais (FERNANDES; BAPTISTA, 2007) inseridas em quase todas as línguas naturais.

Portanto, nesta dissertação o objetivo é descrever as expressões fixas, com nomes de *partes do corpo*, para a elaboração de um dicionário eletrônico. Para tanto, o léxico é dividido em classes gramaticais e cada expressão fixa é descrita, levando-se em conta as propriedades morfossintático-semânticas em contexto de uso. Esses dados são apresentados num modelo de tábua do Léxico-gramática (cf. p. 149), para que se possa constatar as propriedades de cada estrutura.

O léxico

3.1 Conceito e abordagem

O léxico é geralmente definido como um conjunto de palavras ou um acervo de palavras, ou até mesmo como o vocabulário de uma determinada língua. Para Basílio (2004, p. 7), a visão do léxico como um conjunto arbitrário de palavras ou itens lexicais é recorrente não apenas nas abordagens tradicionais, mas também no que se chama de senso comum. A autora ainda afirma que o léxico apresenta um alto teor de regularidade e é um componente fundamental da organização linguística, tanto do ponto de vista semântico e gramatical, quanto do ponto de vista textual e estilístico.

Em Perini (1996, p. 343), o léxico é composto por uma extensa lista em que se armazena a informação idiossincrática (não reduzível a regras gerais) de uma determinada língua, e é formado de milhares de itens lexicais, e cada um desses itens possui informações que contêm características fonológicas, morfológicas, sintáticas e semânticas de uma palavra ou lexema⁸, de um morfema, ou ainda de uma expressão idiomática⁹.

Trask (2004, p. 155) busca explicação também na estrutura das palavras e apresenta o léxico como um conjunto de recursos lexicais, que incluem os morfemas da língua e mais os processos disponíveis na língua para construir palavras com esses recursos. Como exemplo, o autor cita a existência em português de verbos como *compensar* e *oxidar* e dos afixos formadores de palavras *-ável-(a)nte* e *anti-*. Para Trask (2004, p. 155),

⁸ Para Perini (1996, p. 345), lexema é visto como um conjunto de palavras que diferem apenas quanto a morfemas flexionais.

⁹ Assim como vários linguistas e gramáticos, Perini (1996, p. 347) define as expressões fixas como expressões idiomáticas.

qualquer falante podia chegar às palavras *compensável* (como em *cheque compensavel*) e *anti-oxidante [sic]* (como em *substância anti-oxidante [sic]*), bem antes que elas fossem dicionarizadas, usando-as com a expectativa de ser imediatamente compreendido por seu interlocutor, mesmo que os dois não as tivessem nunca encontrado antes.

Seja como for, o léxico não é algo estável, ele está sempre em movimento e sempre mudando. Esta mutabilidade se dá porque, em quase todas as línguas naturais, determinadas palavras se tornam obsoletas, caindo em desuso, ao mesmo tempo em que outras vão sendo criadas e inseridas no léxico, além daquelas que mudam de sentido com o passar do tempo.

Como exemplo, tem-se a palavra **Vândalo**, que segundo Houaiss (2009, versão digital), a princípio, fazia referência ao povo germânico, mas com o passar do tempo, teve seu sentido estendido para designar tanto uma pessoa que destrói ou estraga bens públicos, coisas belas e históricas, quanto que ou aquele que não tem cuidado, esmero, zelo, tudo estraga. Também pode ser usado para que ou aquele cuja ação ou omissão traz prejuízos à civilização, à arte, à cultura. No entanto, esse processo no léxico ocorre de forma tão gradual, que praticamente não é percebido pelas pessoas.

As definições supra citadas resumem o léxico como um conjunto de palavras que os indivíduos utilizam para se comunicarem tanto oralmente como por escrito. Pode-se também dizer que, quando o sistema léxico de uma língua revela a experiência cultural acumulada por uma sociedade através do tempo, esse léxico pode ser considerado como um patrimônio cultural, de importância reconhecida de uma comunidade linguística por meio de sua história, transformando-se em um acervo que pode ser transmitido de geração a geração.

3.2 A aquisição do léxico por humanos

Segundo Biderman (2001, p. 10), o falante não é um indivíduo livre quando fala ou escreve. Sua liberdade no uso da linguagem é circunscrita, por ser a língua um fenômeno social. A autora também afirma que os condicionamentos da fala e da estrutura da língua impõem ao indivíduo um complexo mecanismo de automação no exercício da linguagem. De fato, quando estamos aprendendo os primeiros passos da aquisição da linguagem, na verdade estamos herdando

inconscientemente uma instituição, com um sistema pronto e convencionalizado, que é a língua.

O falante de uma língua adquire um determinado vocabulário. Esse vocabulário pode ser ativo, representado pelo uso das palavras, e um vocabulário passivo, representado pelas palavras que compreende, mas normalmente não usa (TRASK, 2004, p. 155). Esse vocabulário, que é toda a soma de palavras disponível ao indivíduo, pode ser comumente chamado de léxico. É com esse inventário de palavras disponíveis na língua que as pessoas formam seu vocabulário para a comunicação cotidiana. Desta forma, uma pessoa caracteriza o seu vocabulário pela maneira pessoal como ela emprega o léxico.

Para Basílio (2004, p. 9), as línguas existem para que possamos falar uns com os outros. A autora ainda acrescenta que

o objeto da nossa comunicação é o mundo, mais precisamente nosso mundo: as coisas, pessoas, lugares, idéias etc. e suas relações, sejam essas naturais ou artificiais, concretas ou abstratas, reais ou imaginárias. Naturalmente, é necessário primeiro identificar as coisas de que queremos falar e, portanto, designar pessoas, lugares, acontecimentos etc. sobre os quais vamos nos expressar. Assim, a língua é ao mesmo tempo um sistema de classificação e um sistema de comunicação (BASÍLIO, 2004, p. 9).

Nesse sentido, a autora afirma que o léxico tem uma total relação com a dupla função da língua, que é o sistema de classificação e o sistema de comunicação. Segundo Basílio (2004, p. 9), o léxico é uma espécie de banco de dados previamente classificados, um depósito de elementos de designação, o qual fornece unidades básicas para a construção dos enunciados.

Em Rabuske (1995, p. 117), o entendimento da linguagem natural é difícil, pois requer conhecimentos de linguística e domínio do discurso. O autor acrescenta ainda que estes conhecimentos, em boa dose, o ser humano adquire naturalmente, desde a tenra idade. À medida que a pessoa se desenvolve, tem-se a oportunidade de acumular mais e mais conhecimentos e experiências que vão enriquecendo a possibilidade de comunicação. Por isso, quanto maior for o léxico de uma pessoa, maior é a possibilidade de escolha de palavras mais adequadas para a sua comunicação.

3.3 A aquisição do léxico por computador

A linguística é uma ciência que se preocupa em estudar a linguagem humana; e segundo Câmara Jr (2002, p. 159),

[...] é uma ciência que observa e interpreta os fenômenos linguísticos em uma dada língua, numa família ou blocos de línguas e nas línguas em geral, para depreender os princípios fundamentais que regem a organização e o funcionamento da faculdade da linguagem entre os homens.

O interesse e o estudo dos fenômenos linguísticos é bem recente se comparado com outras ciências como a Filosofia ou a Psicologia. Desde que o homem se voltou a aprofundar os estudos dos fenômenos da linguagem humana, a linguística tornou-se uma ciência autônoma. Então, os linguistas começaram a estudar os esforços que as pessoas fazem, através da linguagem, para se comunicarem.

A partir disso, alguns linguistas direcionaram o estudo do léxico na linguagem para a área computacional, com intuito de desenvolver técnicas modernas, tentando “associar conhecimentos linguísticos a conhecimentos computacionais numa atividade interdisciplinar” (BIDERMAN, 2001, p. 79). A história da computação sempre procurou manter à tona a preocupação com a qualidade da interface homem-máquina (RABUSKE, 1999). Com isso, pode-se utilizar o computador como ferramenta para manipulação da linguagem humana. Dessa interação, nasce a Linguística computacional ou Processamento de linguagem natural.

Desde a década de 1960 que os linguistas computacionais se esforçam para desenvolver, através de muita pesquisa, sistemas de modelo formal de línguas naturais, com capacidade de reconhecer e produzir informações apresentadas em linguagem natural.

No entanto, para um computador manipular o léxico de uma língua, é necessário que o pesquisador trabalhe muito sobre as nuances existentes na língua estudada e que fazem parte da estrutura de um léxico, como as redes semânticas, os campos lexicais e a polissemia. Assim, torna-se possível criar ferramentas computacionais qualificadas para processar bons textos. Mas por

ser o léxico um sistema aberto, que está sempre em expansão, traz grandes dificuldades para os pesquisadores da área.

A necessidade de aproximar o computador ao homem não é recente. Segundo Quental e Dias (2004, p. 7), com o advento dos computadores na vida moderna, umas das maiores ambições do homem sempre foi desenvolver uma tecnologia que pudesse, em maior ou menor grau, simular o comportamento humano inteligente. Com inúmeras pesquisas na área, houve um grande avanço no desenvolvimento de *softwares* voltados para o tratamento das línguas naturais, com intuito de compreender e produzir textos orais e escritos.

Todo esse interesse faz com que mais e mais pesquisadores se debrucem sobre teorias linguísticas e, principalmente sobre léxicos, com objetivo de elaborar descrições de línguas naturais para criar programas aptos a serem usados em ambientes computacionais. Para Smarsaro (2000, p. 77), com o aperfeiçoamento dessa máquina (o computador) tão poderosa, a humanidade vem usufruindo trabalhos que antes só poderiam ser desenvolvidos pelo homem, de maneira muito eficiente e mais rápida. É justamente essa rapidez e capacidade de armazenamento de informações que o homem moderno busca.

Para que o computador seja eficiente ao manipular recursos linguísticos, é imprescindível que antes o pesquisador descreva e formalize o léxico de uma determinada língua. A formalização é o principal componente para um ambiente computacional. Pois

não se pode esquecer que a linguagem em computação é altamente formalizada, temos que dar fórmulas para explicar os fatos da língua se a queremos inserir num ambiente computacional. Para isso, antes, é preciso reconhecer as generalizações, regularidades e irregularidades de um reconhecimento linguístico. Só podemos incluir num sistema computacional informações objetivas, ou seja, regras formalizadas. Em outras palavras, o computador só pode lidar com conteúdos formalizados, com informações precisas, através de regras bem explícitas, pois qualquer nível de ambiguidade dificulta o funcionamento de qualquer programa que utilize uma linguagem natural. Assim sendo, a linguagem deve ser o mais determinística possível, caracterizando todos os símbolos linguísticos através de seus traços sintáticos, morfológicos, semânticos e pragmáticos (SMARSARO, p. 81).

Por isso, é importantíssimo um trabalho bem descrito e altamente formalizado para que o computador manipule o léxico o mais próximo possível dos

humanos. Isso porque, para o computador, é de suma importância explicitar, formalizando muito do que é reconhecimento implícito de uma pessoa (BIDERMAN, 2001).

3.3.1 Processamento de Linguagem Natural (PLN)

O estudo do PLN é bastante amplo, pois pode abarcar várias áreas do conhecimento. Segundo Nunes *et al* (1999, p. 5), atualmente, o PLN apresenta-se como um campo de estudos bastante heterogêneo e fragmentado, acumulando uma vasta literatura e agregando pesquisadores das mais variadas especialidades, com formação, embasamento teórico e interesses diversos. Essa heterogeneidade envolve uma diversidade de métodos empregados por pesquisadores em detrimento a cada área de interesse. Por isso, devido à amplitude que tem o PLN, faz-se necessário, nesta pesquisa, traçar apenas um recorte sobre a área.

Para Silva (2005, p. 132-133), a construção do corpo de conhecimentos necessários para a implementação de sistemas de PLN, com grau de sofisticação, exige seleção, organização, representação e codificação de uma variedade de informações na complexa tarefa de criar um simulacro computacional da competência e do desempenho linguístico humano. Sendo assim, através da interação homem-máquina, no PLN, estuda-se os problemas da geração e compreensão automática de línguas humanas naturais. Sistemas de geração de linguagem natural convertem informação de bancos de dados de computadores em linguagem normalmente compreensível ao ser humano, e sistemas de compreensão de linguagem natural convertem ocorrências de linguagem humana em representações mais formais, mais facilmente manipuláveis por programas de computador.

Silva (2005, p. 122) afirma que no estudo do PLN, faz-se necessário realizar, no mínimo, três etapas investigativas: a explicitação dos conhecimentos e habilidades linguísticas, a representação formal desses conhecimentos e habilidades e a construção do programa de computador que codifica essa representação. O autor demonstra mais claramente essas etapas através da tabela abaixo:

Tarefas	Resultados
Explicitação do conhecimento	Descrições linguísticas precisas
Representação formal do conhecimento	Representação linguístico-computacional
Codificação computacional do conhecimento	Implementação computacional

Tabela 3. Etapas de construção de um sistema de PLN.

Nunes *et al* (1999, p. 8) apresenta, na tabela abaixo, um breve histórico evolutivo do PLN:

<p>Década de 50: A Tradução automática</p> <ul style="list-style-type: none"> ▪ sistematização computacional das classes de palavras da gramática tradicional ▪ identificação computacional de poucos tipos de constituintes oracionais. <p>Década de 60: Novas aplicações e criação de formalismo</p> <ul style="list-style-type: none"> ▪ primeiros tratamentos computacionais das gramáticas livres de contexto ▪ criação dos primeiros analisadores sintáticos ▪ primeiras formalizações do significado em termos de redes semânticas <p>Década de 70: Consolidação dos estudos do PLN</p> <ul style="list-style-type: none"> ▪ implementação de parcelas das primeiras gramáticas e analisadores sintáticos ▪ busca de formalização de fatores pragmáticos e discursivos <p>Década de 80: Sofisticação dos sistemas</p> <ul style="list-style-type: none"> ▪ desenvolvimento de teorias linguísticas motivadas pelos estudos do PLN <p>Década de 90: Sistemas baseados em "representações do conhecimento"</p> <ul style="list-style-type: none"> ▪ desenvolvimento de projetos de sistemas de PLN complexos que buscam a integração dos vários tipos de conhecimentos linguísticos e extralinguísticos e das estratégias de inferência envolvidos nos processos de produção, manipulação e interpretação de objetos linguísticos
--

Tabela 4. Evolução dos sistemas de PLN

O seu estudo e suas aplicações são voltados para algumas áreas como:

- linguística de corpus;
- análise sintática (*parsing*) em língua natural;
- recuperação de informação;

- tradução automática;
- reconhecimento de fala;
- síntese de voz;
- sistema de busca;
- correção automática em processadores de texto;
- extração de informações de textos;
- geração de linguagem natural;
- interpretação de linguagem natural.

As diferentes áreas que abarcam o PLN faz com que esse sistema seja bem expressivo e bastante usado na Linguística Computacional. Nesta pesquisa, restringe-se ao estudo das descrições de expressões fixas para implementação na forma de dicionário eletrônico.

A elaboração de um banco de dados no formato de um dicionário eletrônico de qualquer língua natural exige uma boa cobertura descritiva do léxico dessa língua. Segundo Smarsaro (2004, p. 16), o tratamento automático requer das línguas uma descrição sistemática o mais completa possível, pois a insuficiência de dados linguísticos pode gerar falha no processamento automático. Isso porque quando se tenta descrever o léxico de uma língua, se depara com questões de níveis morfológicos, sintáticos e semânticos comum a quase todas as línguas naturais. Portanto, deve-se analisar morfológicamente os traços pertinentes de cada item lexical na sentença; levando-se em consideração a coordenação sintática, os atributos e suas funções sintáticas; e os aspectos semânticos, se o significado da palavra ou das palavras que compõem o item lexical relacionam-se entre si ou se aponta para outro significado. Smarsaro (2004, p. 16) afirma ainda que cada nível deve associar descrições lexicais bastante apropriadas.

Sem essas descrições coerentes dos dados linguísticos, o processamento automático fica bastante comprometido e sujeito a falhas. A qualidade dos

softwares (PLN) depende da qualidade das descrições linguísticas de um léxico, que devem ser explícitas, sistemáticas e coerentes.

3.4 Itens lexicais – diferentes tipos

De acordo com o Dicionário Houaiss digital (2009), item lexical é entendido como cada unidade que entra na constituição do léxico de uma língua, podendo ser formada de uma palavra (*janela, sol, homem* etc.) ou de mais palavras (*guarda-roupa, copo-de-leite* etc.). Esse dicionário é um pouco superficial quando se trata do conceito sobre item lexical. Pode-se acrescentar que os itens lexicais transmitem um significado elementar por serem como unidades mínimas de significação. Nesse sentido, são entendidos como unidades naturais para se aprender uma nova língua ou fazer tradução para outra língua.

Perini (1996, p. 349) argumenta que o estudo do léxico é prejudicado por um fator bastante incômodo, que é a inexistência de um critério seguro para permitir, em certos casos, saber onde termina um item léxico e onde começa outro. Para o autor, esse problema surge quando se têm casos de homonímia ou de polissemia, pois não se sabe bem como distinguir casos como a palavra *folha*. O autor então faz um questionamento ao perguntar se *folha* é um item lexical com dois significados (polissemia) ou de dois itens léxicos que se pronunciam e escrevem da mesma maneira (homonímia); e afirma que como resultado, a estrutura do léxico fica até certo ponto indefinida.

Pode-se dizer também que o conjunto completo de itens lexicais forma o léxico de uma língua na medida em que esse léxico é constituído não apenas por palavras simples, que são os itens lexicais simples, mas também por combinações de palavras que armazenamos em nosso léxico mental, denominados como itens lexicais complexos.

3.4.1 Itens lexicais complexos

Os itens lexicais complexos têm várias denominações, tais como:

- locuções;
- nomes compostos;

- frasesmas;
- expressões idiomáticas;
- expressões cristalizadas;
- expressões fixas;
- expressões lexicalizadas;
- provérbios;
- ditos populares.

Para Lewis (2007a), há diferentes tipos de itens lexicais complexos, que frequentemente incluem:

- nomes compostos, como "*couve-flor*", "*homem de negócios*";
- locuções adverbiais, como "*em todo caso*", "*às vezes*";
- colocações ou palavras de parcerias, como "*amor cego*", "*serviços comunitários*", "*absolutamente convencido*";
- frases institucionalizadas, como "*Vamos ver*", "*Se eu fosse você*", "*Aceita café?*";
- expressões verbais idiomáticas, como "*jogar fora*" ou "*deitar fora*";
- moldes frásticos, como "*Não é tão... como você pensa*", "*O problema é*";
- moldes textuais, como "*Nessa dissertação pretendemos... Primeiro... Segundo... Enfim...*";
- multi-palavras, como "*a propósito*", "*de cabeça para baixo*";
- armações e cabeças de sentença, como "*que não é tão... como se pensa o fato*", "*sugestão*", "*perigo*", "*problema era...*", e até quadros de texto, como "*por exemplo*", "*neste artigo nós exploramos...*"; "*primeiro lugar... em segundo lugar...*"; "*último...*".

As colocações são os fenômenos facilmente observáveis em que co-ocorrem certas palavras nos textos com maior frequência aleatória (Lewis, 1997b, p. 8).

Lewis (2007a) afirma também que a linguagem é composta de pedaços significativos que, quando combinados, produzem textos coerentes, contínuos, e apenas uma minoria das frases faladas são criações inteiramente novas. Nos itens lexicais complexos, há todo um esforço em não tentar quebrar estes itens em partes pequenas, mas fazer um esforço consciente de ver esses itens de uma forma global, por se tratar de expressões totalmente institucionalizadas.

Dentre as denominações supra citadas sobre itens lexicais complexos, as que se enquadram na nossa pesquisa são as do tipo expressões fixas. Isso porque nosso objeto de interesse são os itens lexicais não-composicionais¹⁰, não analisáveis por parte, mas por sua totalidade.

As unidades lexicais que formam as expressões fixas são constituídas por palavras que podem pertencer a qualquer classe de palavras, pois o significado global é diferente das palavras que os compõem e muitas vezes assume o comportamento diferente dos seus componentes, podendo funcionar em outras classes. Nem sempre há uma relação recíproca entre os componentes, pois os seus significados podem não ser previsíveis.

3.5 O que é uma palavra para o PLN

Basílio (2004, p.13) advoga que é comum se definir o léxico como um conjunto de palavras de uma língua. A autora põe em questão a pergunta: mas, o que é uma palavra?

Ao pesquisar sobre o assunto, percebe-se que o conceito sobre palavra é uma questão em estudo que ainda não se esgotou entre linguistas e gramáticos. Para Câmara Jr. (2002, p. 87), palavras são vocábulos providos de significação externa, concentrada no radical; são vocábulos providos de semantema; a palavra é sempre uma forma livre e, pois um lexema. No Houaiss (2003, p. 385), palavra é uma unidade mínima da língua, com som e sentido. Pode ser também uma representação gráfica. Já para Carone (1989, p. 32), uma palavra

¹⁰ A não-composicionalidade é discutida no item 4.2 do Capítulo 4.

pode ser constituída de uma forma livre mínima (leal); de duas formas livres mínimas (couve-flor); de uma forma livre e uma ou mais presas (leal-dade, infeliz-mente) e apenas de formas presas (re-abert-ur-a). Para Trask (2004, p. 218, grifo do autor),

palavra é uma unidade linguística tipicamente maior que o **morfema**, mas menor que um **sintagma**. O termo **palavra** poderia parecer familiar e suficientemente transparente, mas, na realidade, há pelo menos quatro maneiras de definir a palavra, essas maneiras não se equivalem: a **palavra ortográfica** é algo que se escreve com espaços brancos de ambos os lados, mas sem espaços brancos em seu interior. O interesse lingüístico [sic] das palavras ortográficas é mínimo. A **palavra fonológica** é algo que se pronuncia como uma unidade. O **item lexical** ou **lexema** é uma palavra do dicionário, uma unidade para a qual se espera que exista uma entrada própria no dicionário. A **forma gramatical de palavra** ou **palavra morfossintática** é qualquer uma das formas que um item lexical pode assumir para fins gramaticais.

Dadas as múltiplas dimensões que palavra pode abarcar, Carone (1989, p. 33) afirma ainda que é difícil chegar ao conceito de palavra, realidade que todo falante intui. Na modalidade escrita da língua é mais fácil perceber o que é uma palavra, pois em um texto, ela pode ser delimitada por espaços em ambos os lados. Basílio (2004, p. 13) dá como exemplo a frase:

– *João viajou ontem*

Para a autora, ninguém teria problema em reconhecer as três palavras, e que graficamente, palavra pode ser definida como sequência de caracteres que aparecem entre espaços e/ou pontuação e que corresponde a uma dada sequência de sons que formam uma palavra na língua.

Ainda em Basílio (2004, p.14), uma palavra pode apresentar diferentes formas por causa da sua flexão. Se usarmos a palavra *ganhar* para exemplificar o que a autora disse, teremos:

– *Ganhou – ganho – ganharia – ganhará*

Perceberemos que há quatro formas do verbo *ganhar*. Por causa da flexão, as quatro formas desse verbo são consideradas palavras distintas dentro dos enunciados abaixo:

– Paulo *ganhou* um presente.

- Eu *ganho* um presente.
- Eu disse que ela *ganharia* um presente.
- Eder *ganhará* um presente.

Pode-se perceber que nesses enunciados há quatro palavras, cada uma correspondendo a uma das formas do verbo *ganhar*. Segundo Basílio (2004, p. 14), um dos enfoques que temos para palavra é o de “unidade de que se compõe o enunciado”; e que o outro enfoque é o que considera a palavra “como uma unidade estrutural que congrega diversas formas”. Por isso, o verbo *ganhar* pode ser considerado como uma palavra ou unidade estrutural que congrega as diversas formas de conjugação.

Basílio (2004, p. 16) também fala de palavra fonológica. Para a autora,

A palavra também pode ser entendida como uma unidade fonológica. Por um lado, podemos pensar na palavra como uma sequência fônica que ocorre entre pausas potenciais. Por isso, na estrutura do português as palavras apresentam um padrão acentual baseado em tonicidade e duração. Chamamos de vocábulo fonológico o lado fonológico da palavra.

Na oralidade, distinguir palavras individuais é extremamente difícil. As palavras podem ser pronunciadas de uma única vez, principalmente as mais curtas. Há também as palavras polissílabas que podem ser quebradas facilmente pelos falantes.

Bloomfield (1926) propõe o conceito de “Mínimas formas livres” em que as palavras são tidas como a menor unidade significativa da fala e, por si só, podem constituir um enunciado, o que não acontece com as formas presas. Por isso, agrupa fonemas (unidades sonoras) em lexemas (unidades de significado). Para Basílio (2004), o conceito de Bloomfield é interessante, pois distingue palavras de frases, sintagmas e afixos, mas quando se pensa em palavras compostas (formadas por duas ou mais palavras) apresenta problemas, porque fica difícil sustentar ao mesmo tempo em que palavras não podem ser subdivididas em formas livres.

Toda essa divergência sobre o conceito de palavra talvez se deva ao fato de que uma palavra, enquanto unidade lexical, apresenta um conjunto de propriedades morfológicas, sintáticas, semânticas e pragmáticas que lhe atribui

diferentes graus de relevância. Mas para o PLN, tanto uma palavra simples, como **casa**, ou uma combinação fixa de palavras simples, como a expressão fixa **boca-de-fumo**, deve ser analisada como um único item lexical e assim computada.

Assim, da mesma maneira que uma palavra simples, a expressão fixa também deve listada no léxico de uma língua. Isso é necessário porque, quando se quer um resultado significativo na descrição para a aquisição do léxico por um programa computacional, deve-se levar em conta a importância e a relevância das expressões fixas para o léxico. Pois, deixando de fora do léxico de uma língua o reconhecimento de unidades lexicais formadas por mais de uma palavra, pode-se restringir ou dificultar a comunicação numa determinada situação de uso da língua. As expressões fixas já fazem parte do conhecimento linguístico do falante nativo e sua exclusão pode causar graves falhas a qualquer ambiente computacional que tenta manipular uma língua natural. As expressões fixas, portanto, são itens lexicais que dão uma abrangência maior ao léxico do português. Sendo assim, é de grande relevância a implementação delas em um léxico computacional de qualquer programa que lide com PLN.

3.6 O léxico do português brasileiro

Como acontece em quase todas as línguas, o léxico do português do Brasil está sempre se modificando, tendo em vista os processos de formação de palavras bastante produtivos, como a derivação e a composição, que permite o surgimento de novas palavras a cada momento, que com o passar do tempo se integram ao léxico da língua devido à sua aceitabilidade por parte dos falantes.

No português brasileiro, observa-se uma grande variação linguística, as pessoas formam novas palavras com muita criatividade e improvisação. Mas, apesar de algumas palavras novas não serem utilizadas em um contexto de linguagem formal, isso não se impede o seu uso, principalmente pelos falantes nativos.

Na língua portuguesa, o léxico “apresenta um alto teor de regularidade e é um componente fundamental na organização linguística” (BASILIO, 2004, p. 7), e as palavras novas constituem-se em estruturas lexicais produtivas para a

expansão e o funcionamento delas na linguagem, servindo para nossa comunicação. Para Basílio (2004, p. 9), o léxico é uma espécie de banco de dados previamente classificados, um depósito de elementos de designação, o qual fornece unidades básicas para a construção dos enunciados.

Como o léxico está sempre em expansão, há a necessidade de se aproveitar o léxico já existente para se formar novas palavras ou novos itens lexicais e, assim, evitar sobrecarga para nossa memória. Ou seja, as pessoas utilizam o léxico como um sistema dinâmico para formar novos itens lexicais a partir de itens já existentes. Por exemplo, para se dizer que alguém agiu insensatamente, pode-se optar pelo uso do verbo **perder** + o substantivo **cabeça**, que são palavras já existentes, formando a expressão fixa **perder a cabeça**. Não foi necessário criar uma nova palavra, mas sim, recorrer ao léxico, juntar palavras distintas e mediante uma nova sequência, dar a sequência um novo sentido, formando um item lexical.

Devido à grande produtividade do léxico, os dicionários não dão conta de classificar todas as palavras. Para classificá-las é necessário que antes seja comprovado o seu uso em vários segmentos sociais, e só então serão dicionarizadas. Mas nesse período em que se busca essa comprovação, várias outras palavras são criadas, e isso faz com que os dicionários estejam sempre desatualizados. É interessante observar que, se por um lado existem várias palavras não dicionarizadas que são bastante usadas, por outro, existem aquelas dicionarizadas que são totalmente desconhecidas pela maioria das pessoas e que provavelmente jamais serão utilizadas.

Nesse sentido, “o léxico português, basicamente de origem latina, tem ampliado seu acervo por meio dos processos de derivação e composição” (ALVES, 2004, p. 5). Contudo, apesar da língua portuguesa ser de origem latina, o léxico português também sofreu influências dos franceses, árabes, africanos, ingleses, tupis, espanhóis e tantos outros. Mas para Carone (1986, p. 36) os recursos mais atuantes são internos ao sistema, sempre prontos para entrar em processo e desencadear a formação de novas palavras.

3.7 A relevância do léxico para o PLN

Atualmente, o léxico não pode ser mais considerado como uma simples lista de palavras, como nas décadas de 1960 e 1970. Supõe-se que deve estar contidas num léxico quase todas as informações morfológicas, sintáticas, semânticas e fonológicas de uma língua (TIBERIUS, 1999, *apud* MUNIZ, 2004, p. 5). Para Vieira e Lima (2001), o léxico é uma estrutura fundamental para a maioria dos sistemas e aplicações. Já em Biderman (2001, p. 90-91),

em virtudes de todas as pesquisas referidas em PLN, TA¹¹, BDL¹², Lexicografia Computacional [...], o léxico veio ocupar um lugar central nas pesquisas e teorias sobre linguagem. nos primeiros tempos o léxico foi tido como um gargalo no desenvolvimento da linguística computacional, por causa da magnitude do patrimônio vocabular que constitui o acervo herdado de uma língua moderna de civilização, além do obstáculo adicional criado pelo volume incessante de neologismo que se incorporam continuamente ao léxico. Entretanto, isso é essencial nas teorias sobre a linguagem, o léxico constitui um repositório de informação sobre o mundo, uma vez que o conhecimento está condensado em palavras.

De acordo com Smarsaro (2004, p. 16), com relação ao tratamento do léxico, se os dicionários utilizados pelos sistemas de processamentos não forem adequados, quer do ponto de vista da sua cobertura lexical, quer do ponto de vista da formalização e sistematização da informação lingüística, isso afetará não só a análise lexical de um determinado texto, mas também de todas as fases de processamento subsequentes. Se uma palavra não for reconhecida e não for corretamente identificada, a análise sintática da frase ou da estrutura em que ela se encontra não poderá ser feita. Logo, o léxico é uma estrutura de dados contendo os itens lexicais e as informações correspondentes a estes itens, o que o torna um componente essencial nas operações de PLN. Sendo assim, o conhecimento sobre o léxico deve ser o mais abrangente possível, porque através desse léxico é possível dar um passo a mais no processamento da língua natural em um ambiente computacional.

3.8 A relevância de dicionário eletrônico de expressões fixas para PLN

Sabe-se que as expressões fixas estão presentes e são bastante numerosas nas línguas naturais. Jackendoff (1997) acredita que existem estimadamente

¹¹ Tradução automática.

¹² Bases de dados lexicais.

mais de 25.000 expressões fixas na língua inglesa; Gross (1982) também estima esse mesmo número de expressões fixas para o francês. Ainda não há um estudo que comprove a real quantidade de expressões fixas ao redor do mundo, nem tampouco no Brasil. Entretanto, sabe-se que o seu uso é bastante regular, pois as pessoas *lançam mão* das expressões fixas de maneira muito corriqueira para efetuar uma comunicação de forma expressiva, ou até mesmo de uma forma pitoresca nos discursos cotidianos.

Sabe-se também que os dicionários tradicionais têm importância como um bem cultural; visto que, muitas vezes, se transformam em bens de consumo pelas pessoas. Por isso, é pertinente investir em pesquisas criteriosas para elaborar, não somente os dicionários tradicionais de uso, como também os dicionários de variadas obras lexicográficas como, por exemplo, os dicionários eletrônicos formados por expressões fixas.

Ranchhod (2001, p. 17) evidencia que as expressões fixas se prendem com a falta de composicionalidade lexical, sintática e semântica de toda a expressão. Para a autora, as informações de natureza mais sintático-semântica também têm de ser tratadas nos dicionários eletrônicos. Logo, os dicionários eletrônicos devem ser desde o início concebidos para receber cumulativamente não só informações adicionais sobre palavras, mas também sobre combinações de palavras, isto é, sobre comportamento das propriedades sintáticas e semânticas dessas combinações.

Para se obter sucesso na elaboração de um Léxico-gramática do português, é preciso descrever sistematicamente e formalizar não só as estruturas sintáticas consideradas livres, mas também as estruturas sintáticas com restrições sintáticas como, por exemplo, as expressões fixas. Já há diversos estudos dessas estruturas seguindo a metodologia do Léxico-gramática. Dentre eles, têm-se as expressões nominais (BAPTISTA, 1995), adjetivais (CARVALHO, 2001), adverbiais (RANCHHOD, 1991), frases fixas (RANCHHOD, 1993). No português brasileiro, existem alguns estudos, como os das expressões fixas (VALE, 2001) e dos nomes compostos (SMARSARO, 2004).

Silberztein (1997, *apud* SMARSARO, p. 21) chama a atenção para a importância de ferramentas computacionais. Segundo o autor, para confrontar

o dicionário com um texto e associar as palavras do texto às informações linguísticas do dicionário, é preciso ter ferramentas de análise lexical. E a constatação da ampla ocorrência das expressões fixas na comunicação cotidiana, ou até mesmo na literatura, respalda um estudo pormenorizado para que os estudiosos da área possam analisá-las cientificamente, verificando suas construções, evidenciando seus elementos lexicais constituintes, até que seja possível criar ferramentas computacionais para elaborar um dicionário eletrônico dessas expressões.

Partindo do princípio de que quanto maior for a cobertura lexical de uma língua natural, maior será a qualidade dos programas que processam essa língua, o dicionário eletrônico de expressões fixas se mostra como mais um recurso computacional de grande relevância para programas que trabalham com PLN.

3.9 Léxico computacional ou dicionário eletrônico

O dicionário eletrônico é um léxico computacional que contém informações referentes aos itens lexicais de uma língua, que pode abarcar tanto palavras simples, como *mão*, *boca*, *perna*, *olho* quanto expressões fixas com específica significação, como ***dar as caras***, ***ficar na mão***, ***bater cabeça***, ***mão de vaca***, ***perna de pau***, ***olho gordo***. Para Muniz (2004, p. 5), entre as informações associadas aos itens lexicais destacam-se as referentes à categoria gramatical do item. Deve-se destacar também os valores morfossintático-semânticos. Isso é necessário porque os itens lexicais possuem propriedades de natureza distinta, refletindo o comportamento que eles adquirem quando combinados entre si na atividade comunicativa (NUNES *et al*, 1999, p. 17).

O dicionário eletrônico é concebido para ser usado, sem a intervenção humana, por programas informáticos em várias operações de processamento de linguagem natural (SMARSARO, 2004, p. 18). A sua finalidade é o reconhecimento das unidades lexicais simples e complexas num texto a ser automaticamente indexado. Nele, faz-se análise de um texto com objetivo de extrair informações ou até mesmo traduzir um texto para outra língua.

Segundo Ranchhod (1998), a informação linguística contida nos dicionários eletrônicos é determinada, formalizada e codificada à mão por uma equipe de linguistas. Esse tipo de dicionário é também elaborado com intuito específico

de ser usado em análise automática de textos. Ele deve conter informações linguísticas codificadas e formatadas. A partir disso, torna-se acessível aos programas de análise lexical e semântica. Não pode haver lacunas, tanto na parte lexical quanto gramatical, e todas as informações linguísticas têm de estar coerentes e estruturadas. A codificação da informação é um requisito imprescindível para dicionário eletrônico.

Frequentemente ocorre certa confusão entre dicionário eletrônico e dicionário informatizado como, por exemplo, o Dicionário digital Houaiss. Diferentemente dos dicionários eletrônicos, em que suas informações são em forma de códigos, destinadas aos profissionais da linguística computacional, os dicionários informatizados, que também têm as informações apresentadas tradicionalmente em formatos manuais, ou seja, em papel, são elaborados para serem utilizados por humanos. Segundo Biderman (2001, p. 91), os dicionários tradicionais foram elaborados para as pessoas que possuem um grande volume de conhecimentos implícitos sobre a estruturação do léxico, e tais conhecimentos não estão explicitados nesses dicionários. Biderman então afirma que por isso não se pode fazer uma mera transposição de um dicionário tradicional para a forma eletrônica porque um sistema de PLN não saberia processá-lo. Ainda, por serem manuseados pelos humanos, esses dicionários tradicionais não estão sujeitos às imposições dos dicionários eletrônicos.

Segundo Gross (1989, *apud* SMARSARO, 2004, p. 19), a ambiguidade do termo informatização levou a um mal-entendido entre essas duas categorias de dicionário; e que as informações contidas em cada dicionário não têm nada em comum. Pois os dicionários eletrônicos são voltados para atuantes da linguística computacional; enquanto que no dicionário informatizado, os textos são destinados ao grande público.

Nos dicionários informatizados se omitem muitas informações que poderiam ser evidentes pelo utilizador, no caso, o humano, ou são apenas implicitamente referidas. Mas o conhecimento implícito ou explícito dos humanos que consultam este tipo de dicionário ajuda a contornar no todo ou em parte tais lacunas (MUNIZ, 2004, p. 8). Por isso, espera-se que os usuários que os consultam já possuam conhecimentos linguísticos suficientes para manuseá-

los. Logo, os dicionários informatizados não podem ser usados como os dicionários eletrônicos, porque nos dicionários informatizados há muitas informações implícitas, caso contrário, seria necessário explicitar essas informações implícitas nas definições do léxico e isso obrigaria a reescrever completamente o conteúdo de praticamente todas as entradas nesses dicionários.

Em síntese, todo sistema computacional, que objetiva processar uma língua natural, tem o léxico como um componente fundamental. E para o PLN, o léxico computacional ou dicionário eletrônico entra como uma base de dados linguísticos, com informações morfossintático-semânticas detalhadas, totalmente descritas e formalizadas para automatização dos processos linguísticos como, por exemplo, a tradução automática de textos. Por isso, a inserção do léxico em ambiente computacional implica profundo conhecimento lexical e gramatical de uma língua.

Abordagem sobre as expressões fixas

4.1 Abordagem em literaturas diversas

Atualmente, o estudo das expressões fixas passou a ter grande importância não só para a linguística computacional, mas também como suporte para a aprendizagem da língua em sala de aula, tanto para os nativos como para os estrangeiros que queiram aprender determinada língua. O interesse pelas expressões fixas se dá pelo simples fato de elas serem muito recorrentes em qualquer língua, pois são facilmente encontradas e identificadas em contextos diversos. Também são bem aceitas pelas pessoas por serem totalmente institucionalizadas e terem um valor pragmático, já que fazem parte do conhecimento de mundo de cada um. Além disso, são constituídas de temas gramaticais como *verbos*, *substantivos*, *advérbios*, *adjetivos*, *conjunções*, *artigos*, *preposições*, que refletem realmente o uso desses temas no mundo real. Isso quer dizer que, através dessas sentenças institucionalizadas, é possível traçar um caminho não só para aplicá-las no ensino-aprendizagem de uma língua materna ou estrangeira, como também para o PLN.

Cowie (1988, p. 136) argumenta que a existência de unidades lexicais de uma língua como o inglês, atende as necessidades dos falantes nativos; e a fusão generalizada de expressões parece satisfazer as necessidades comunicativas do indivíduo num dado momento, que são posteriormente reutilizadas.

Se por um lado não se pode mais negar a existência e a importância das expressões fixas na formação e ampliação do léxico, por outro não se encontra um estudo mais pormenorizado delas, nem em dicionários nem em livros

didáticos nem na maioria das gramáticas. É como se as expressões fixas não fizessem parte do léxico na proporção que existem e fazem.

Embora já exista uma importante bibliografia sobre as expressões fixas pelos linguistas, principalmente os da área computacional, a noção de fixidez ('*figée*', em francês; '*frozen*', em inglês) está longe de ser consensual, tanto do ponto de vista analítico e conceitual como terminológico (GROSS, apud RANCHHOD, 2003, p. 1). Talvez por isso, assim como em outras línguas, existem também no português alguns termos para designá-las, como *fraseologias*, *fraseolexemas*, *expressões cristalizadas* ou *expressões idiomáticas*.

Borba (1971, p. 85) trata as expressões fixas como *idiotismo*. Para o autor, o termo já pouco usado para designar os elementos linguísticos particulares de uma língua é resultante da análise *sui generis* da experiência; e cita a palavra *saudade* como exemplo. Câmara Jr. (2002, p. 142) também usa o termo *idiotismo* e argumenta que as construções vocabulares e frasais não se prestam a uma análise, satisfatória na base dos valores atuais da língua, porque resultaram de fenômenos de analogia e atração, e só se explicam à luz da história da língua. De acordo com Câmara Jr., são especialmente dignos de nota os *idiotismos locucionais*, cuja significação não decorre das dos vocábulos componentes e da sua articulação sintática. O autor cita como exemplo as expressões ***dar as da vila-diogo*** e ***chorar pitanga***¹³. Explica que o termo *idiotismo* provém de *idiota*, e seu sentido inicial veio da palavra *Idiotes*, que quer dizer *particular, individual*. O autor também denomina as expressões fixas como *idiomatismo*, que é derivado de *idioma*, como sendo da mesma família léxica.

Trask (2004, p. 141-142) não cita a palavra *idiotismo* para designar as expressões fixas, apenas cita *idiomatismo* ou *expressão idiomática*, designando que em uma *expressão idiomática*, o significado não pode ser derivado de análise das palavras que a constituem. Trask cita como exemplo a expressão em inglês *let the cat out of the bag*, que literalmente significa *deixar que o gato saia da sacola*. Mas para os falantes da língua inglesa significa *revelar publicamente alguma coisa que se esperava que fosse mantida em*

¹³ Grifo nosso

segredo. Para chegar a esse entendimento, é preciso que se aprenda esse significado separadamente. Trask considera isso uma característica de *idiomatismo*, e acrescenta ainda que um fato linguisticamente fascinante em matéria de *idiomatismo* é a possibilidade de aplicar em alguns deles os processos sintáticos correntes da língua. *Receber alguém com quatro pedras na mão* pode parecer em sentença como *fui recebida pelas irmãs dele com quatro pedras na mão*, em que o *idiomatismo* foi quebrado e suas partes distribuídas pela sentença, e nem por isso o significado idiomático se perdeu. Para o autor, essas descobertas criam problemas interessantes, tanto para a sintaxe como para a psicolinguística. No entanto, isso é também um problema para a linguística computacional ao manipular uma língua natural.

Bechara (1988, p. 333) considera as expressões fixas como *anomalias da linguagem*, chamando-as, como alguns autores, de *idiotismos* ou *expressões idiomáticas*. Para o autor, são *anomalias* porque não podem ser analisadas ou porque estão em choque com os princípios gerais da gramática, apesar de serem bem aceitas no falar culto; e cita como exemplo a expressão *é que*, o *infinitivo* e a preposição em *o bom do pároco*.

Almeida (1980, p. 480, §786) também fala de *idiotismo* ou *expressão idiomática*. Para o autor, *idiomatismo* se refere ao termo ou dicção existente numa língua e não há correspondente em outras línguas. Almeida segue afirmando que *idiotismo* são frases e modismos que se afastam dos princípios gerais da sintaxe, mesmo sendo consagrados pelo uso de pessoas cultas e geralmente adotados na boa linguagem. O autor cita como *idiotismo* o *infinitivo flexionado*; o emprego da locução *é que*, em “eu *é que* fiz isso”; o emprego da preposição *de* nas expressões “pobre *do* homem”, “o bom *do* velhinho”; a anteposição do artigo aos possessivos, como em “o meu livro”, “as nossas relações”.

Almeida (1980, p. 325, §354) ainda considera como *idiomatismo* as frases ou orações as locuções adverbiais compostas por duas ou mais palavras que exprimem circunstâncias de *lugar*, *negação*, *modo*, *intensidade*, *tempo*, *dúvida*, *afirmação* e *modo*. O autor elenca uma lista de locuções adverbiais, como *a mão tenente*, *a pé quedo*, *a todo pulso*, *a unha de cavalo*, *de mão beijada*,

entre outras. Também sugere que o aluno deva estudar e recordar tais locuções adverbiais, porque o conhecimento e aplicação delas irão influenciar muito na boa linguagem. Entretanto, não explica como será o estudo delas pelo aluno.

Já Torres (1964, p. 214) chama de *barbarismo* o emprego de palavras ou frases estranhas à língua, quer em sua forma, quer em sua ideia, ou seja, para o autor, é um vício de linguagem por se caracterizar como deturpações que a língua sofre em sua pronúncia e escrita por ignorância ou descuido de quem o fala. O autor então classifica alguns itens como *barbarismo*, entre eles os *Galicismos fraseológicos*, como *guardar o leite por estar de cama*, *chefe-de-obra por obra-prima*, *golpe-de-vista* no lugar de *olhadela* etc.

Ignácio (2003, p. 51) salienta que há casos em que seria problemático classificar determinado complemento como objeto direto, pois não seria possível nem a *pronominalização* nem *apassivação*, como

[...] o caso dos complementos que fazem parte de uma estrutura formada com base num verbo *suporte*, ou seja, o verbo tem por complemento uma forma nominal que, com ele, formam um conjunto unitário que pode ser substituído por um verbo da mesma raiz do nome que forma o conjunto ou que tenha significado correlato. Ex.: dar uma surra = surrar; fazer um apelo = apelar; dar um chute = chutar. Às vezes não é possível substituir o conjunto “verbo + nome” por um verbo da mesma raiz, mas pode-se ter uma forma verbal equivalente. Ex.: *levar um tombo* = cair. Outras vezes, seria até previsível uma forma verbal da mesma raiz do nome, no entanto ela ainda não está consagrada pelo uso. Ex.: *dar um pontapé*, deveria responder a “*pontapear*”; como essa forma ainda não existe, a expressão se substitui por *chutar*. Então, nesses casos, o melhor seria considerar o conjunto todo como um único núcleo verbal que seria classificado como “intransitivo” (*levar um tombo*) ou “transitivo indireto” (*dar uma surra em alguém*).

Segundo Ignácio, nas *expressões idiomáticas*, os complementos verbais também não se submetem às regras supracitadas, e um dos testes para se identificar a *expressão idiomática* é tentar pronominalizar o complemento do verbo. Em *piquei a mula* não pode corresponder a **piquei-a*, porque o significado da expressão é diferente do significado de cada uma das partes. Também seria impossível a *apassivação*, como **a mula foi picada por mim*. Ignácio até tenta explicar um dos testes para a identificação das expressões fixas, chamadas por ele de *idiomáticas*. Entretanto, existem algumas

expressões fixas que são perfeitamente apassivadas, como a expressão ***empurrar com a barriga***:

- (1) Eder ***empurrou com a barriga*** a reforma da casa.
- (2) A reforma da casa foi ***empurrada com a barriga*** por Eder.

Inácio não atentou para o fato de que cada expressão é um caso, por isso, não se pode englobá-las em uma única regra para sua identificação. Isso traria falhas gravíssimas para o PLN.

Perini (1996, p. 347) reconhece que o léxico precisa incluir certas expressões idiomáticas fixas, como *bater as botas*, *a olhos vistos* e outras; e afirma que as expressões citadas não são propriamente palavras, porque em *bater as botas*, pode-se flexionar a primeira parte, no caso o verbo: *bateu as botas*, *bateram as botas*. O autor também afirma que as expressões idiomáticas não podem ser consideradas frases ou sintagmas normais, porque

[...] na fala, nunca podem ser interrompidas por hesitações, sem destruir o efeito de expressão idiomática. Assim, se alguém disser:

Cidinha bateu – ééé... – as botas.

A mensagem transmitida não será de que “Cidinha morreu”, mas de que realmente bateu umas botas (para tirar a poeira, talvez).

Depois, essas expressões se põem de elementos fixos. Não podemos sequer mudar certas flexões; a frase:

Cidinha bateu a bota.

Novamente quer dizer que ela esteve limpando o calçado, e não que morreu.

Finalmente, em certos casos, como em *a olhos vistos*, a própria estrutura da expressão é peculiar e não corresponde exatamente à estrutura de um sintagma (igualmente, se interpretarmos a expressão literalmente, o resultado pode ser anômalo) (PERINI, 1996, P. 347).

Perini conclui que tais expressões não são estruturas montadas pela sintaxe e interpretadas pela semântica, mas que são verdadeiros itens compostos, listados separadamente no léxico. Para Vale (2001, p. 56), Perini, de certa maneira, reconhece que o problema existe, por isso não pode ser colocado de lado. Perini também argumenta que, aquilo que pode ser sistematizado em um conjunto de regras deve ser tratado nas gramáticas e a parte idiossincrática das línguas nos dicionários.

Para Sandmann (1991, p. 29-30), depois que a palavra é formada e se estabelece dentro do estoque mais ou menos comum de unidades lexicais utilizadas pelos falantes de uma língua, ela geralmente sofre um processo de *lexicalização* ou *idiomatização*. Para o autor, esse processo também pode ser chamado de *desmotivação*, pois o todo não é mais a simples soma das partes. Então, a palavra *amável* não é mais ‘*aquele que pode ser amado*’ ou ‘*que é digno de amor*’, ela passa a ter a denominação de aquele que é ‘*gentil, polido*’. Também acrescenta que *bóia-fria* não é a ‘*comida fria*’, mas a pessoa ‘*que come comida fria*’. Mas, se observar o item lexical *bóia-fria* no Houaiss (2009, versão digital), trata-se de:

- trabalhador rural itinerante que se ocupa em tarefas temporárias sem vínculo empregatício;
- empregado que come no local de trabalho a *bóia* (comida) que traz de casa (tal como os *bóias-frias* rurais).

Têm-se então sentidos diferentes do que Sandmann considera. No Houaiss, não há relação entre o item lexical *bóia-fria* e a comida. Há referência apenas a uma profissão e, a partir disso, tem-se uma derivação por extensão de sentido entre a pessoa que traz de casa a própria comida e a come no local de trabalho.

Numa outra obra sua, Sandmann (1988, p. 96-97) fala em *formações exocêntricas* que, em Stein (1977, apud SANDMANN, 1988, p. 96.), “são formações de palavras, onde o objeto denotado está fora dos significados dos constituintes. A formação *cara-pálida* não é um rosto pálido como sugere os constituintes, e sim, alguém com um rosto pálido”¹⁴.

Para Sandmann, a formação dessas unidades lexicais baseia-se no princípio da contiguidade, no relacionamento externo e espacial entre dois fenômenos extralinguísticos, ou seja, de dois referentes. A contiguidade é vista então como uma ponte para a expansão do significado de rosto pálido para a unidade lexical *cara-pálida*. Ainda conforme o autor, esse processo é uma das modalidades de metonímia que, juntamente com a metáfora, é um capítulo “à

¹⁴ Tradução nossa.

parte” na formação de palavras, e cita alguns exemplos de formações exocêntricas, como abaixo:

Sem teto para (o) *sem-teto*; (o) meio-campo ‘parte do campo’ para (o) *meio-campo* ‘jogador’; (a) bóia fria para (o) *bóia-fria*; (o) centro esquerdo para (a) *centro-esquerda* ‘ala de partido, partido’.

No entanto, no uso, pode-se um pouco mais além do que o autor aponta, pois segundo Houaiss (2009, versão digital), em ***cara-pálida***, pode-se também ser:

- 1- designação dada às pessoas de pele branca pelos índios norte-americanos. Essa designação foi bem popularizada nos filmes e livros de faroeste;
- 2- no Brasil, expressão usada também como *interlocutório pessoal* para expressar estranhamento ou oposição entre pessoas. Ex.: A quem você pensa que engana, ***cara-pálida***?

Elson e Pickett (1978, p. 39-40) ressaltam que, se por um lado o significado de um morfema é definido pelo contexto no qual ocorre, por outro, considerando o morfema em questão e seu contexto, é a sequência dos morfemas que realmente transmite o significado. Ainda para os autores, são as sequências de morfemas, no caso, as palavras, as locuções, as orações e os períodos que geralmente têm um significado que pode provir dos significados dos elementos em comum, como em *menino* + *-s* dá o significado de mais de um menino; *meninos* + *grandes* dão o sentido de mais de um menino, sendo todos grandes etc. Mas quando uma sequência de morfemas tem um significado que não é inferido do significado das palavras que formam o conjunto, recebe o nome de expressão idiomática. Para os autores, a expressão idiomática pode ser constituída a partir de apenas uma palavra (composta), de uma expressão ou de uma oração, e uma mesma sequência de morfemas pode representar tanto uma expressão idiomática como uma forma livre. As palavras *red tape* podem formar uma expressão idiomática quando usada com sentido de burocracia, mas quando se tratar de apenas uma fita vermelha, usada por uma costureira, tem-se, então, uma expressão não idiomática. Na visão dos autores,

Uma sequência de morfemas em uma língua estrangeira não deveria ser considerada como expressão idiomática simplesmente porque a tradução nos soa estranha. As expressões idiomáticas são definidas dentro da própria língua e podem ou não coincidir com uma expressão idiomática, em outra língua. Portanto, as seguintes orações espanholas usadas em tabuletas não são expressões idiomáticas, apesar de soar estranha a tradução literal em inglês ou em português *Empuje usted* “Push you”, “Empurre você”. No hay paso “There is not pass”, “Não há passo” (significando que a estrada ou a rua está interditada). No estacionarse “Not to park yourself” (Not parking); “Não se estacionar” (Proibido estacionar).

Elson e Pickett finalizam afirmando que apesar de se encontrar expressões idiomáticas idênticas ou semelhantes em duas línguas, quer por um acaso quer por tradução, como *second hand* em inglês ou *segunda mão* em português (com sentido de algo que não é novo), o estudante de uma língua deve ter muito cuidado em tentar traduzir literalmente as expressões idiomáticas para outra língua. Por isso, não se deve levar em consideração apenas as transferências das palavras traduzidas.

Em Carone (1986, p. 14-15), os objetos linguísticos são construídos em diferentes níveis. No 1º nível está o *morfema*, considerado como a menor unidade significativa; no 2º nível, o *vocábulo*, por ser uma unidade construída de um ou mais morfemas. No 3º nível, tem-se a *lexia*, a qual pode ter a conformação de um vocábulo, mas o que nela se considera não é sua estrutura, e sim, o seu comportamento dentro de uma unidade maior. Para a autora, o vocábulo *médicos* pode ser central em uma situação, como *médicos bons*, mas que também pode ter um papel marginal em outra, como em *cuidados médicos*. Segundo Carone, ao fazer uma nova distribuição dentro do sintagma, os mesmos vocábulos trocaram seus papéis, concluindo que *lexia* é uma unidade de comportamento. Carone (1986, p. 14) também diz que

No entanto, nem sempre ela se confunde formalmente com o vocábulo; pode ser *composta* de dois ou mais vocábulos (água-de-cheiro, madressilva), constituindo unidades já cristalizadas pelo uso. Pode também ser *complexa*, quando construções sintáticas normalmente não-cristalizadas se imobilizam em uma situação específica. A frase volitiva “deus nos acuda!”, por exemplo, pode imobilizar-se, assumindo o comportamento léxico de um substantivo: “Foi um deus-nos-acuda”.

Por isso, a autora afirma que são também *lexias complexas* as construções sintáticas, e por fazerem parte do uso em memória coletiva se cristalizam, caracterizando-se como frases feitas e ditos populares.

De um modo geral, após analisar alguns autores sobre a abordagem das expressões fixas, percebe-se que há uma grande lacuna a ser preenchida. Para Vale (2001, p. 53-54), a não-abordagem delas nas gramáticas tradicionais nos mostra que foram totalmente deixadas de lado por considerá-las sem importância. Isso acontece porque as gramáticas tradicionais tem sido o campo da normatividade, o que sempre aconselha a não usá-las. Assim, Vale cita como Martins (1997, apud VALE, 2001, p. 54) e Cipro Neto (FSP 06/01/200 apud VALE, 2001, p. 54) execram a expressão fixa *correr atrás do prejuízo*:

CORRER ATRÁS DO PREJUÍZO. Além de constituir modismo, é expressão incorreta. Corre-se atrás do empate, da vitória, da vantagem, do título, do lucro e nunca "do prejuízo". (MARTINS, 1997, p.82)

Já vi muita gente boa defender a legitimidade dessa construção ("correr atrás do prejuízo"), com o argumento de que o uso lhe dá razão. O estranho é que ninguém diz que corre atrás do fracasso, do insucesso, da tristeza. O que se diz é que o time corre atrás da medalha, da vitória, da classificação. Por que diabos, então, correr atrás do prejuízo?

Vale ressalta que nos dois casos, os autores não se conformam com o fato de que construções desse tipo não seguem a lógica, e que na expressão em questão, o seu significado não está expresso por seus componentes. Assim, *correr atrás do prejuízo* não significa que o sujeito esteja buscando o prejuízo, mas, ao contrário, busca se livrar dele. O que define a expressão, muito mais do que o seu componente, é, evidentemente, o uso (VALE, 2001, p. 54).

Duarte (2000, p. 26-27) também acredita que *correr atrás do prejuízo* é uma incoerência da fala, pois "só louco corre atrás do prejuízo. Do prejuízo eu fujo. Eu corro atrás do lucro". Assim como essa expressão, Duarte também considera as expressões fixas *correr risco de vida*, *crise do desemprego* e *tirar a pressão* como casos incoerentes ou absurdos. Entretanto, o significado dessas expressões fixas e de tantas outras, como ***encher a cabeça***, ***enganar o estômago***, ***fazer uma boquinha***, ***molhar a garganta***, não se analisam pelo significado de cada item que as compõem, mas pela convenção, transmitindo os seus significados através do uso.

Para Marques, Mattos e Taille (1986, p. 48), expressões cristalizadas como *Maria-vai-com-as-outras* ou *bom gosto* e outras, trazem grandes problemas para as gramáticas pedagógicas tradicionais, porque nelas, o conceito de

palavra, baseado na morfologia, está fundado mais na aparência do que em critérios implícitos, formalizáveis, de significação e referência. De acordo com os autores,

[...] a grande revolução da lingüística moderna foi abandonar a descrição fragmentadora, que pulverizava a língua em elementos isolados, estáticos, e passar a vê-la como uma atividade criadora, como um sistema dinâmico, composto de elementos também dinâmicos e inter-relacionados (MARQUES; MATTOS; TAILLE, 1986, p. 48).

Em seu estudo, Gross (1982, apud BIDERMAN, 1999, p. 91) afirma que os linguistas geralmente atribuíram as expressões cristalizadas um caráter de exceção, de anomalia lingüística, e não propuseram alternativas para tratar cientificamente essa questão.

Contudo, para Bidermann (1999, p. 91), trata-se de um fenômeno de grande envergadura e muito frequente. Ainda, o estudo dessas combinações lexicais ou fraseológicas suscita muitos problemas teóricos e coloca em causa os papéis atribuídos tradicionalmente à sintaxe e ao léxico.

A falta de critérios para identificar e definir as expressões fixas, consideradas por muitos estudiosos da área como construções complexas, faz com que alguns autores não tenham tanto interesse lingüístico para que possam estudá-las, ou até mesmo dar certa atenção a elas. Segundo Ranchhod (2003, p. 2), as divergências terminológicas e a ausência de critérios de análise adequados, elevaram as expressões fixas como objetos lingüísticos excepcionais, não integráveis na gramática das línguas, por não poderem ser objetos de regras gerais.

Mesmo consideradas por muitos como objetos lingüísticos anômalos, ultimamente, tem-se verificado um enorme interesse sobre as expressões fixas com intuito de descrever suas propriedades lingüísticas e formalizá-las para serem usadas em PLN. Esse interesse se dá porque, em quase todos os textos, é possível encontrar tais itens lexicais, conforme se vê no decorrer desta pesquisa. Pressupõe, então, que a existência delas é fato, e estão no uso da língua, portanto, não há como ignorá-las nem deixá-las de lado.

A contribuição da pesquisa em questão se coaduna com este objetivo, na medida em que se descreve as propriedades das expressões fixas com nomes de *partes do corpo*, com as estruturas verbais e substantivais do português brasileiro.

4.2 Composicionalidade / não-composicionalidade

Conforme Gross (1986), alguns grupos de palavras são semanticamente e / ou sintaticamente não-composicionais, e a falta de composicionalidade resulta da constatação da restrição lexical em uma determinada sequência, pois o seu sentido será sempre homogêneo. Para Gross, a não-composicionalidade impede muitas variações dos componentes de uma sequência fixa, como nos exemplos abaixo:

- (3) Choveu cães e gatos.
- (3a) *choveu muitos cães e gatos.
- (3b) *Choveu grandes felinos e cães
- (3c) *choveu caninos e felinos

Gross (1986, p. 5) também afirma que os termos compostos formam uma parte essencial de um Léxico-gramática, pois eles forçam tanto o linguista quanto o especialista em computador a adotar uma visão muito mais abstrata da linguagem e, semanticamente, por definição, os enunciados compostos não podem ser decompostos em expressões simples. Por isso, em certo sentido, é preciso reconhecer que o significado não tem muito a ver com as palavras que formam as unidades compostas.

Assim como Gross, Laporte, Nakamura e Voyatzi (2008, p. 1) consideram que, nas expressões fixas, as combinações não obedecem a regras de produção sintática e semântica de composicionalidade; e qualquer combinação de elementos linguísticos, que é lícito na linguagem, mesmo não sendo representadas sintático e semanticamente nas gramáticas, serão armazenadas no léxico. Como exemplo, têm-se as expressões fixas ***crescer o olho, destacar na pele, pegar barriga, olho gordo e cabeça-de-vento.***

Para Freckleton *et al* (*apud* LAPORTE; NAKAMURA; VOYATZI, 2008, p. 1), a composicionalidade sintático-semântica é geralmente definida como uma combinação de elementos linguísticos de composição porque seu significado é calculado pela soma dos seus elementos. Laporte, Nakamura e Voyatzi também compactuam com essa concepção, no entanto, acrescentam que a possibilidade de calcular o significado das frases a partir de seus elementos é somente de interesse se esta for a melhor solução de armazenar as mesmas frases em léxicos, ou seja, se elas contam com as regras gramaticais com generalidades suficientes. Conclui-se, portanto, que uma composição é uma combinação de elementos linguísticos se somente o seu significado puder ser computado a partir de seus elementos por uma gramática.

Segundo Laporte, Nakamura e Voyatzi (2008, p. 2), uma expressão é tida como fixa quando essa expressão é composta de várias palavras e alguns ou todos os seus elementos são congelados em conjunto. Sendo assim, a combinação dos elementos que compõem a expressão nem sempre obedecem a regras da produtividade sintática e da composicionalidade semântica. Por exemplo, na sequência abaixo **olho gordo**:

(4) Eder tem um **olho gordo**.

(4a) Eder tem um **olho (gordo + *obeso + *cheio + *corpulento + *magro)**.

A falta de composicionalidade presente na distribuição do adjetivo **gordo** na sequência **olho gordo** do enunciado em (4a) se caracteriza como uma expressão fixa porque os seus elementos são congelados em conjunto.

De acordo com Smarsaro (2004, p. 78), uma sequência é composicional quando, normalmente, cada elemento pode ser substituído por uma dezena de outros. Assim como para Gross, Laporte, Nakamura e Voyatzi, para Smarsaro, a noção de composicionalidade também tem a ver com a possibilidade de deduzir o significado de uma sequência a partir dos significados dos seus componentes e, no caso da composicionalidade de sequências linguísticas, trata-se de um processo que pode ser associado a uma construção sintática, e a partir daí, é possível aplicar a exemplos variados, como abaixo:

(5) Eder comprou *carne de boi* para o almoço.

(5a) Eder comprou *carne de (boi + vaca + carneiro + porco + coelho + frango)* para o almoço.

Como se pode ver em (5), na sequência *carne de boi*, é totalmente possível deduzir o significado a partir dos seus componentes. Também é possível aplicar a exemplos variados, como em (5a). Trata-se, portanto, de uma sequência livre, que aceita composicionalidade semântica.

Portanto, quando uma sequência é composicional, se entende que essa sequência não deve ser considerada como uma expressão fixa. Em uma sequência composicional há tanto transparência semântica quanto produtividade, e não deve haver opacidade. Vale (2001, p. 72) entende que transparência é uma proximidade do cálculo do significado total da expressão por seus componentes, já opacidade seria a total impossibilidade desse cálculo como, por exemplo, as sequências abaixo *forrar o sofá* e ***forrar o estômago***:

(6) Eder *forrou o sofá* antes de sair de casa.

(7) Eder ***forrou o estômago*** antes de sair de casa.

Em (6), a sequência *forrar o sofá* é entendido como uma sequência composicional porque há produtividade e transparência semântica:

(6a) Eder *forrou o (sofá + mesa + cadeira)* antes de sair de casa.

(6b) Eder (*forrou + revestiu + cobriu*) *o sofá* antes de sair de casa.

Dessa forma, por causa da transparência semântica e da produtividade, a sequência *forrar o sofá* não se caracteriza como uma expressão fixa. Já em (7), a sequência ***forrar o estômago*** não demonstra transparência semântica nem aceita produtividade:

(7a) Eder ***forrou o (estômago + *intestino + *fígado)*** antes de sair de casa.

- (7b) Eder (**forrou** + ***revestiu** + ***cobriu**) **o estômago** antes de sair de casa.

O bloqueio na tentativa de uma produtividade e a falta de transparência semântica para o substantivo **estômago** e o verbo **forrar** demonstra que se trata de uma expressão fixa. Como não há possibilidade de composicionalidade como acontece em (6a) e (6b), a sequência **forrar o estômago** se caracteriza como uma sequência não-composicional. Por isso, segundo Smarsaro (2004, p. 41), o grande problema que se observa em torno da composição é o estabelecimento de critérios para sua identificação. Quando se pensa em processamento automático, esta é uma questão de grande relevância, porque para a autora, uma palavra composta deve ser analisada computacionalmente como um bloco, para que se possam construir representações adequadas de estrutura sintática e semântica das frases em que se encontra. Logo, a partir da composição por justaposição de **pé-de-vento**, tem-se um novo item lexical, com significados e comportamentos sintáticos diferentes dos substantivos **pé** e **vento**, se analisados como substantivos livres, que podem ser verificados no enunciado abaixo:

- (8) Eder enfrentou um **pé-de-vento** para chegar aqui.

Em (8), o significado não pode ser deduzido a partir das propriedades sintáticas dos constituintes **pé** e **vento**. A palavra **pé**, empregada separadamente, pode se posicionar com o verbo **cortar** em diversos contextos. Já a expressão fixa **pé-de-vento** não, como se observa abaixo:

- (9) Eder cortou o **pé** após uma longa caminhada.

- (9a) *Eder cortou o **pé-de-vento** após uma longa caminhada.

Baptista (1995, p. 8) salienta que é preciso desenvolver critérios linguísticos que permitam identificar com rigor, de modo operativo e reproduzível as unidades não-composicionais; listar esses elementos; descrever o seu comportamento morfossintático e semântico e formalizar essas informações linguísticas, de modo que possam ser utilizadas por sistemas automáticos de texto

Portanto, em uma análise automática de texto, o estudo dessas particularidades não-composicionais, formadas por itens lexicais com categorias gramaticais diversas, é de suma importância.

A identificação de itens lexicais não-composicionais esbarra em conceitos pouco esclarecedores por linguistas e gramáticos. Segundo Baptista (1994, p. 5-6), tradicionalmente, a identificação dos nomes compostos é feita, sobretudo, com base em critérios de natureza semântica. Para Grevisse (1969, *apud* BAPTISTA, 1995, p. 6), o nome composto evoca na mente, não as imagens distintas relativas a qualquer dos elementos correspondentes, mas uma imagem única.

Basílio (1999, p. 15) menciona que no estudo do português, as abordagens tradicionais avultam controvérsias sobre várias questões, entre elas, sobre a natureza composicional. Para Basílio, tais problemas, naturalmente, encontram sua expressão mais viva na lexicografia, e que decisões fundamentais devem ser tomadas a cada passo.

Ainda em Basílio (2004, p. 29-30), a composição envolve a junção de uma base a outra. Também não há elementos fixos, não há funções predeterminadas no nível dos elementos. Por isso, a autora defende a idéia de que o que caracteriza e define a função de processo de composição é sua estrutura, de tal modo que, das bases que se juntam para formar uma palavra, cada uma tem seu papel definido pela estrutura. Como exemplo, a autora cita composições do tipo substantivo + substantivo: **sofá-cama**, **peixe-espada** e **couve-flor**, em que o primeiro substantivo funciona como núcleo da construção e o segundo como modificador ou especificador. Já em composições de substantivo + adjetivo: **obra-prima**, **livre-arbítrio**, **caixa-alta** e **belas-artes**, o núcleo é o substantivo e o modificador é o adjetivo, seja qual for a ordem de ocorrência. Logo, para Basílio, a composição é um processo de formação de palavras que utiliza estruturas sintáticas para fins lexicais. Acrescenta ainda que os mecanismos ou as estruturas que são normalmente utilizados na formação de enunciados passam a ser utilizados na função de denominar e/ou caracterizar seres, que é uma função fundamental do léxico.

Baptista (1994, p. 45) não pensa como Basílio com relação à formação composicional do tipo substantivo + substantivo. Para Baptista, do ponto de vista semântico, alguns nomes compostos do tipo substantivo + substantivo parecem ter um significado composicional como, por exemplo, **água-mel**, em que o composto designa uma bebida realmente preparada com água e mel. Entretanto, outros são semanticamente opacos. No composto **água-pé** o significado global dos substantivos **água** e **pé** em pouco a ver com o significado individual de cada uma das palavras simples que formam o composto. Não há, portanto, nem modificador nem especificador, como afirma Basílio.

Já Bechara (1988, p. 175) considera a composição como a criação de uma palavra nova composta por meio de duas ou mais outras cuja significação depende das que encerram os seus componentes. Por esse conceito de composição, tem-se, então, **bicho-de-pé**, **cintura de pilão**, **cabelo duro**, **olho no olho**. Entretanto, o autor não se atentou que nos compostos **cabeça-de-porco**, **pé d'água**, **unha-de-fome**, **cara-metade**, **braço da lei** e tantas outras, não há dependência de sentidos. Já que **cabeça-de-porco** não se refere nem a uma **cabeça** e nem a um **porco**, mas a **cortiço**; assim como não há referências entre **pé** e **água**, **unha** e **fome**, **cara** e **metade**, **braço** e **lei**.

Bidermann (1999, p. 89-90) apresenta um estudo diferente para as formações composicionais. Para a autora, no plano da língua, o termo *lexema* se refere à unidade abstrata do léxico. Já as manifestações discursivas dos *lexemas* devem ser referidas tecnicamente como *lexias*. Então, as *lexias* se repartem em duas categorias, que são as *lexias simples*, graficamente constituídas de uma sequência gráfica, separadas por dois brancos, como **cesta**, **guarda**, **dona** e **mãe**; e as *lexias complexas*, formadas por várias unidades separadas por brancos, mas não ligadas por hífen, como **cesta básica** e **dona de casa**.

Bidermann então chama de *lexias compostas*, as unidades ligadas por hífen, como **guarda-roupa** e **mãe-de-santo**. Para distinguir entre uma *lexia complexa* e uma *sequência discursiva variável*, a autora aplica dois testes:

1º) teste da substituição.

Em uma sequência como a saudação *bom dia* ou *boa noite* se pode substituir o primeiro vocábulo por outro adjetivo mais ou menos sinônimo, dizendo: **ótimo dia** e **ótima noite**. Já na sequência **bater as botas**, não tem como substituir o termo **botas** por **sapatos**. Por conseguinte, tudo leva a crer que **bom dia**, **boa noite** e **bater as botas** já estão lexicalizados no nível do sistema lexical.

2º teste da inserção.

Numa sequência como **dor de cabeça** não diremos *dor* “terrível” *de cabeça*, mas “terrível” *dor de cabeça*, ou *dor de cabeça* “terrível”. Também não se diria *capa* “bonita” *de chuva*, mas *capa de chuva* “bonita”. Tampouco se poderia inserir um advérbio entre os constituintes de **mercado negro**, resultando: *mercado* “muito” *negro*. Portanto, o teste de inserção demonstra que estão lexicalizados **dor de cabeça**, **capa de chuva** e **mercado negro**.

A autora acrescenta que no exemplo acima não se pode inserir nada na sequência **bater as botas**, dizendo, por exemplo: Carlos *bateu rapidamente as botas*.

Apesar de a autora argumentar que no teste 2 que não é possível inserir um elemento lexical em sequências lexicalizadas, no entanto, esse teste não demonstra ser tão seguro assim, pois em diversas sequências lexicalizadas é perfeitamente possível inserir elementos como em:

- **Trazer no coração**

- (10) Ele, meu pai, que sempre acreditou em mim, ficava revoltado quando ia ao estádio e via a torcida rubro-negro me vaiando. **Trago-o no coração** a cada gol (Adriano, da Seleção brasileira) (AT - 03/06/05).

- **Dar as costas**

- (11) Rodrigues disse que foi expulso sem direito a defesa e que o comando da igreja **deu-lhe as costas** (AT - 07/07/05).

- **Entrar na cabeça**

(12) A presidente do PT do Ceará, Sônia Braga, disse “assombrada”: isso não **entra na minha cabeça** (AT - 21/08/05).

▪ **Destacar na pele**

(13) Depois de se **destacar em “Cabocla” na pele** da romântica Tina, Maria Flor vai estrear no horário nobre da Globo (AT - 17/07/05).

▪ **Cortar na carne**

(14) “**Cortaremos na própria carne** se necessário (o Presidente Lula sobre o mensalão)” (AT - 08/06/05).

Como se nota, o estudo das expressões fixas é bem mais complexo do que se imagina. Sendo assim, em apenas um ou dois testes, como é proposto pela autora, não é possível dar uma posição exata da identificação de todas as expressões fixas.

As expressões fixas são bastante produtivas porque podem ser formadas por combinações de palavras em variados contextos. Isso deduz que determinados testes são seguros para umas e não são tão seguros assim para outras. Logo, é preciso entender que cada caso é um caso. Pois, como não é possível estabelecer padrões nem criar regras quanto à sua produção, as expressões fixas se apresentam como um fenômeno idiossincrático. E isso se revela um problema, o que ocasiona uma falta de consenso ao seu tratamento teórico por linguistas e gramáticos.

A partir da questão do reconhecimento da composicionalidade e da não-composicionalidade, levando-se em conta a não-produtividade nas mais diversificadas sequências de um léxico - como o caso das expressões fixas - pode-se perfeitamente estudá-las e descrevê-las pensando em um tratamento automático para serem usadas em um ambiente computacional.

Portanto, à luz da Teoria do Léxico-gramática e do trabalho desenvolvido por pesquisadores como Gross (1986), Baptista (1994), Vale (2001), Ranchhod (2003), Smarsaro (2004) e Laporte (2008), para esta pesquisa também se propõe descrever as propriedades estruturais das expressões fixas, do tipo

verbal e substantival, a partir dos critérios utilizados, principalmente por Smarsaro e Baptista.

Por isso, nesta pesquisa, a noção e o conceito de não-composicionalidade, de que não há produtividade nem transparência semântica em uma sequência não-composicional, é de grande relevância, porque as expressões fixas aqui estudadas são também de natureza não-composicional.

4.3 O uso do hífen nas expressões fixas

O emprego do hífen sempre gerou dúvidas na língua portuguesa, principalmente quando se fala em noções e regras gerais. Isso porque as regras sempre trazem exceções, sem contar com as novas palavras que surgem de momentos em momentos, trazendo cada vez mais incertezas.

Bechara (1988, p. 65) afirma que, seguindo os princípios do Pequeno Vocabulário Ortográfico da Língua Portuguesa (PVOLP), deve-se empregar o hífen nas palavras compostas em que os elementos, com a sua acentuação própria, não conservam, considerados isoladamente, a sua significação, mas o conjunto constitui uma unidade semântica. Nesse sentido, temos então, como unidades lexicais, ***guarda-costas***, ***pé-de-meia***, ***mão-de-obra***, entre outras.

Bidermann (1999, p. 91) também defende o uso do hífen nos compostos. Para a autora, as unidades complexas deveriam ser grafadas mais apropriadamente por hífen, como fim-de-semana, jogo-do-bicho, indicando sua natureza de unidade.

Mas será que as expressões fixas ***unha afiada***, ***sangue novo***, ***língua solta***, ***frio na barriga***, ***olho da rua***, ***corpo de bombeiros***, ***nervos de aço*** e tantas outras, perderiam a natureza de expressão fixa pela ausência do hífen?

Já Rocha Lima (1997, p. 227) se posiciona afirmando que na composição, por não haver grafia uniforme para as palavras compostas, a ausência ou presença do hífen não serve de critério para a identificação composicional porque o sentido de composição “tem seu critério no espírito”.

Segundo Carone (1986, p. 38), convém notar o papel do hífen nos compostos, como representação gráfica de que ocorreu a cristalização. De acordo com a autora, dispensa-se o hífen quando ocorre a incorporação dos componentes em um só vocábulo fonológico, como *pontapé*, *madressilva* e *vaivém*. Carone também afirma que é lícito supor que o mesmo possa ocorrer com os que ainda mantêm sua integridade fonético-fonológica; fator decisivo é frequência do emprego da palavra, pois também na língua o uso produz desgaste.

Ranchhod e Carvalho (2003, p. 212-213) argumentam que os nomes grafados com hífen são normalmente identificados pelos lexicógrafos e registrados nos dicionários. Para as autoras, muitos deles, entretanto, não têm tal grafia, mesmo porque esta não seria um critério linguístico adequado para caracterizar uma sequência de palavras como unidade lexical. Pois, na identificação dos compostos há que utilizar um conjunto de critérios linguísticos que vão desde a verificação do comportamento morfológico dos seus constituintes até a análise das propriedades sintáticas e semânticas.

Pode-se, então, perceber que tanto linguistas quanto gramáticos têm idéias divergentes sobre o emprego do hífen como fator determinante para a identificação das unidades fixas. Por isso, apesar de no *corpus* desta pesquisa perceber uma ocorrência bem maior de hífen nas unidades fixas de estruturas substantivais do que nas estruturas verbais, não é isso que determina se as unidades fixas, com ou sem hífen, representam um novo item lexical, com significado distinto dos constituintes. Mesmo que existam algumas regras para as unidades fixas com hífen na forma escrita, “os usuários do português já perceberam que o uso do hífen é uma imposição que não tem nenhuma relação com o sentido de composição” (SMARSARO, 2004, p. 53). Logo, todas as regras sobre as unidades fixas que levam e que não levam hífen “*caem por terra*” quando essas unidades são usadas na modalidade falada, já que na oralidade não há nenhuma identificação desse “*traço-de-união*”.

4.4 Abordagem estabelecida pela teoria do Léxico-gramática

Nesta pesquisa, o estudo das expressões fixas se apóia no âmbito da Teoria do Léxico-gramática. As expressões fixas possuem propriedades linguísticas, que apenas o estudo da semântica não é suficiente para descrevê-las. Assim,

o estudo morfossintático-semântico das expressões fixas por meio da Teoria do Léxico-gramática, torna-se possível formular métodos de investigação linguística satisfatórios para o uso em um ambiente computacional. Para Smarsaro (2004, p. 53), no método de descrição do Léxico-gramática, a identificação e a descrição das unidades lexicais vão além das dimensões do léxico, porque leva em conta também a adequação linguística.

O estudo descritivo e transformacional do Léxico-gramática foi inspirado na teoria americana de Harris nos anos sessenta. Segundo Lacavalla (2007, p. 64), O ponto de partida da descrição dos fenômenos naturais da linguagem é representado pelo acúmulo de um número importante de dados linguísticos, a partir do qual é realizado um estudo exaustivo nas transformações das sentenças. No Léxico-gramática, o princípio da aceitabilidade das propriedades descritas é um fator muito importante, pois por ser um procedimento empírico, leva em conta o uso e a aceitabilidade pelos falantes nativos, e isso ajuda a entender a real produtividade lexical da língua. Mas para tanto, por meio de operações transformacionais, cada sequência deve ser testada para se julgar a sua aceitabilidade ou não.

Segundo Gross (1989), *apud* Lacavalla (2007, p. 65), a análise do Léxico-gramática é fundamental porque certamente há razões para localizar o nível de significação da sentença como um todo e não apenas nas palavras que a compõe, e um dos motivos de força maior é a existência de muitas frases fixas como "**pegar o touro pelos chifres**", fica claro que o significado do verbo e seus complementos não estão envolvidos na interpretação. De fato, para identificação e descrição de uma expressão fixa, seja ela de estrutura verbal ou substantival, é preciso estabelecer critérios linguísticos que, de acordo com Ranchhod (1990) e Baptista (1994) *apud* Smarsaro (2004, p. 53), vão desde a análise do seu comportamento morfológico até a verificação total ou parcial perda de composicionalidade lexical, sintática e semântica. E é através de critérios linguísticos que se pode distinguir uma sequência fixa de uma sequência livre, como nos exemplos abaixo:

Carne-de-sol

(15) Eder preparou ***carne-de-sol*** para o almoço.

(15a) Eder preparou (**carne-de-sol** + * **de lua** + ***de satélite**) para o almoço.

Carne-de-sol = representa um tipo de carne (fixa).

Carne de boi

(16) Eder preparou *carne de boi* para o almoço.

(16a) Eder preparou (*carne de boi* + *de frango* + *de carneiro* + *de porco*) para o almoço.

Carne de boi = representa a carne de um animal (livre).

Ou uma sequência ambígua, que tanto pode ser livre como fixa:

Cruzar os braços

(17) Eder **cruzou os braços para** o problema da empresa.

(17a) Eder **cruzou (os braços e *as pernas)** para o problema da empresa.

Cruzar os braços = demonstrar descontentamento a algo, como fazer greve ou ficar inativo voluntariamente (fixa).

Cruzar os braços

(18) Eder *cruzou os braços* para meditar.

(18a) Eder *cruzou (os braços e as pernas)* para meditar.

Cruzar os braços = pôr-se em cruz (livre).

No método de descrição do Léxico-gramática é possível aplicar vários testes para avaliar a aceitabilidade de uma expressão como fixa. Apesar das expressões fixas terem uma forma e um conteúdo parecidos com uma sequência livre de palavras, entretanto apresentam particularidades que vão desde uma distribuição única até a restrição total dos seus elementos lexicais. Para Smarsaro (2004, p. 54), as restrições podem ser avaliadas, se forem consideradas as relações que ligam a sequência em frases, a partir de um

conjunto de critérios que possam identificar as propriedades morfossintáticas e semânticas. Mas, para a autora, nenhum critério, por si só é suficiente. Pois é na intersecção de vários critérios que se define a composicionalidade de uma dada combinação, que será tanto mais fixa, quanto mais restrições se observarem em relação às propriedades sintáticas que caracterizam um grupo nominal livre formado pela mesma sequência interna de categorias gramaticais.

A partir da definição, pautada na Teoria do Léxico-gramática por Gross (1986), de que uma expressão fixa é não-composicional, é possível perceber também que determinadas expressões fixas ora são totalmente opacas ora são parcialmente opacas, quando um dos seus componentes é transparente.

Em:

(19) Eder precisa **esfriar a cabeça** para tomar a decisão correta.

Esfriar a cabeça = acalmar (opacidade total).

Ou seja, o significado de **esfriar a cabeça** em (19), que é **acalmar**, não tem relação com o verbo **esfriar** nem com o substantivo **cabeça**, o que caracteriza uma opacidade total do sentido com relação as partes que compõem o sentido nessa expressão fixa.

Em:

(20) Eder comprou **carne-seca** para a feijoada.

Carne-seca = tipo de carne (transparência parcial).

Já **carne-seca**, o significado, apesar de não ter relação com o adjetivo **seco**, tem uma total relação com o substantivo **carne**, pois trata-se de um tipo de **carne**. Isso faz com que a expressão fixa **carne-seca** seja parcialmente opaca, ou seja, com transparência parcial de um dos seus elementos, sem que, necessariamente, seja uma sequência composicional. Por isso, Smarsaro (2004, p. 56-56) advoga que a composicionalidade de uma sequência depende de suas propriedades sintáticas.

Embora uma sequência apresente transparência parcial, ela deve ser considerada como fixa, porque apresenta uma distribuição restrita, como um

item lexical autônomo, mesmo que essa sequência apresente uma variação na escrita, mas o seu sentido global sempre será mantido, como se vê nos exemplos abaixo, a partir do Houaiss (2009, versão informatizado):

Carne-seca: Culinária. Regionalismo: Brasil. Charque feito com bastante sal e ressecado geralmente em estufa.

Carne de sol: Regionalismo: Norte do Brasil, Nordeste do Brasil. Carne bovina levemente salgada e seca ao sol ou ao vento

Carne de vento: Regionalismo: Sul do Brasil, Centro-Oeste do Brasil. Charque de vento

Carne do sertão: Regionalismo: Brasil. Charque

Carne do sul: Regionalismo: Ceará. Charque

Apesar das variações regionais na escrita para **carne-seca**, o sentido é sempre o mesmo: um tipo de **carne**, também conhecida por **charque**, que no Houaiss, têm-se as seguintes definições:

Charque: Regionalismo: Brasil.

- 1 Rubrica: alimentação.
Carne bovina cortada em mantas, salgada e seca ao sol ou por processos afins, inclusive com utilização hoje de produtos químicos; jabá.
- 2 Derivação: por metonímia. Rubrica: culinária. Preparação ou prato feito com essa carne

Para **charque**, Houaiss apresenta as seguintes variações: **carne de ceará**, **carne do ceará**, **carne do sertão**, **carne do sul**, **carne-seca**, **carne-velha**, **ceará**, **iabá**, **jabá**, **sambamba**, **sumaca**.

Assim como **carne-seca**, que mantém uma parte do significado do primeiro nome **carne**, existem outras tantas expressões fixas que também apresentam certa composicionalidade semântica como, por exemplo, **dente de leite** e

cabeça de cebola. Portanto, usar o critério de não-composicionalidade como um único critério de identificação de uma unidade lexical como fixa, pode deixar de fora muitas expressões fixas, como afirma Baptista (1994, p. 7), para os quais é possível reconstituir, pelo menos parcialmente, o seu significado global a partir do significado dos seus elementos constituintes.

Critérios de identificação de expressões fixas

No português do Brasil, encontra-se sequências usuais que podem ser diferenciadas em dois tipos: aquelas chamadas sequências livres ou transparentes, por apresentarem uma ampla produtividade distribucional; e aquelas chamadas de sequências fixas ou opacas, “por possuírem características que não são distribucionalmente produtivas nem são interpretáveis composicionalmente” (RANCHHOD, p. 8).

Neste capítulo, as expressões fixas, formadas por *nomes de partes do corpo*, serão avaliadas por um conjunto de critérios formais, analisando-se as propriedades dos elementos que constituem essa sequência fixa. Segundo Smarsaro (2004, p. 58), a aplicação dos critérios será feita, levando-se em conta as distribuições sintáticas dos componentes de cada sequência e a interpretação linguística também. Isso porque, só um nativo ou um grande pesquisador da área, naquela língua, é que pode julgar se uma sequência é uma expressão fixa ou não (livre).

Como a pesquisa tem como objetivo o estudo de duas estruturas argumentais, a verbal e a substantival, primeiro será analisada a estrutura argumental verbal, logo em seguida, será a estrutura argumental substantival.

5.1 Expressões fixas de estruturas argumentais verbais

O quadro abaixo representa a divisão da estrutura argumental verbal a partir de um *corpus* com 351 expressões fixas:

VERBAL	EXEMPLOS	EFETIVO
VartN	Abaixar a cabeça	175
VprepN	Falar pelos cotovelos	112
VN	Abrir mão	64
TOTAL		351

Tabela 5. Representação da estrutura verbal.

Para Vale (2001, p. 18), uma das características que diferencia as expressões fixas verbais das não verbais é o fato de que nas expressões fixas verbais sempre há a presença de um verbo que pode se flexionar e que está atrelado à estrutura da expressão fixa. O verbo paradigmaticamente tende sempre a concordar com o sujeito contido na frase. Podemos perceber claramente essa afirmação quando dizemos que “eles **bateram o pé** e não cederam à chantagem do patrão”. Observa-se que o verbo **bater** concorda com o sujeito, enquanto que o substantivo **pé** permanece no singular, independentemente de o sujeito da frase estar ou não no singular. Isso ocorre com vários outros exemplos como, por exemplo, **Abaixar a cabeça, perder a cabeça, queimar a língua, bater boca, ficar de olho** entre outras.

Muitas vezes, as expressões verbais podem ser substituídas por um verbo de igual valor semântico:

- (1) Eder **abriu mão** da herança.
= Eder **desistiu** da herança.
- (2) Eder **deu no pé** diante da situação.
= Eder **correu** diante da situação.
- (3) Eder **apunhalou pelas costas** seu melhor amigo.
= Eder **traiu** seu melhor amigo.

Segundo Carone (1986, p. 38), em algumas nas frases verbais, o verbo presta-se à formação de compostos, mas o resultado nem sempre é um verbo. Não dizemos ele **estraga-prazeres**, ele **leva-e-traz** – porque o composto é um nome e só poderia elevá-lo ao status de predicado com uma translação operada pelo verbo **ser**:

(4) Ele é um **estraga-prazeres**, um **leva-e-traz**.

Ou então, destrói-se o composto, dizendo:

(5) Ele *estraga prazeres*, ele *leva e traz*.

Devido a essas particularidades, não se pode incluir as expressões fixas a uma única categoria. Cada caso é um caso, isso requer extrema atenção na hora da descrição, para que assim, possa ser feita uma descrição satisfatória delas.

5.1.1 Critérios utilizados referentes às propriedades morfosintáticas e semânticas para a identificação das expressões fixas verbais

Para mostrar a fixidez das expressões de estruturas argumentais verbais VartN, VprepN e VN, faz-se necessário submetê-las a alguns critérios cabíveis às estruturas das expressões, que são:

- a) distribuição sintática dos itens lexicais;
- b) inserção lexical;
- c) negação obrigatória;
- d) determinante obrigatório;
- e) ausência de determinante;
- f) preposição obrigatória;
- g) coordenação do verbo com outro verbo;
- h) variação em número do N.

5.1.2 VartN

As expressões fixas formadas por VartN se mostram altamente numerosas. Pode-se, abaixo, exemplificar algumas delas:

Abrir a cabeça, amarrar a cara, bater o pé, confundir a cabeça, conseguir uma boca, cortar o coração, crescer o olho, cruzar os braços, dar as caras, dobrar a língua, encher a cara, enfeitar a testa, enganar o estômago, fazer a cabeça, fechar a cara, lavar as mãos, meter o nariz, mostrar a unhas, passar a perna, perder a cabeça, quebrar a cara, soltar a língua, torcer o nariz, virar a cabeça.

A partir de vários exemplos, foi possível, então, submeter as expressões fixas aos seguintes critérios:

- a) distribuição sintática dos itens lexicais;
- b) inserção lexical;
- c) negação obrigatória;
- d) determinante obrigatório;
- e) substituição do verbo por outro verbo;
- f) variação em número de N.

5.1.2.1 Distribuição sintática dos itens lexicais

De acordo com Harris (1952), a partir da distribuição sintática dos elementos em uma sequência, tem-se a noção de transformação. Por isso, para avaliar a fixidez de uma expressão, faz-se necessário observar as distribuições sintáticas dos seus componentes em uma dada sequência e, a partir daí, analisar se essa distribuição conserva ou não o sentido da expressão.

Tendo, a distribuição sintática como um tópico extremamente importante na teoria do léxico-gramática, e, levando-se em consideração que é a partir da distribuição que podemos observar como cada elemento lexical se comporta nas frases, se a unidade de significado da expressão é mantida, se foi alterada

ou se houve total perda do seu sentido, é usado, então, o critério da distribuição sintática, em alguns exemplos, para verificar o que acontece com o significado das expressões em uma dada sequência.

Em **esquentar a cabeça**:

- (6) Eder **esquentou a cabeça** durante a avaliação em um concurso.
- (6a) *a cabeça de Eder Esquentou muito e ferveu
- (6b) *Eder queimou a mão quando a encostou na cabeça quente
- (6c) *a cabeça foi esquentada em fogo baixo
- (6d) *Eder esquentou a cabeça e os pés por causa do frio
- (6e) *após um longo tempo quente, a cabeça esfriou

Podemos sentir frio nos **pés, braços, orelhas, pernas** e ter a necessidade de esquentá-los. Inserir o substantivo **cabeça** na lista dessas partes do corpo que podem ser literalmente aquecidas causa grande estranhamente, isso porque o verbo **esquentar**, como na frase, mas a um estado que esse sujeito se encontra, como uma preocupação ou até mesmo uma aflição por causa do concurso. A distribuição sintática dos itens acima comprova a restrição da sentença, caracterizando-a como uma expressão fixa.

Usando a expressão fixa **bater o pé** em uma distribuição sintática no enunciado (7), tem-se a seguinte conclusão:

- (7) Eder **bateu o pé** e não tomou a vitamina
- (7a) *Eder bateu o pé e palmas
- (7b) *o pé de Eder não parou de bater por um longo tempo
- (7c) *Eder bateu o pé na pedra e se machucou
- (7d) *o pé de Eder está muito batido
- (7e) *Eder bateu o pé no liquidificador

O verbo **bater**, na forma simples, funciona, na maioria das vezes, como um verbo transitivo direto, com o sentido de alguém bater um objeto em alguém ou alguém bater em algum lugar. Através da distribuição sintática, pode-se observar que quando verbo **bater** se junta com o substantivo **pé**, forma-se uma expressão fixa. Essa fixidez interna da expressão **bater o pé** impede que o verbo **bater** seja usado com outras intenções que não seja a de manifestar uma oposição ou até mesmo agir de maneira insistente.

5.1.2.2 Substituição do verbo por outro verbo

Em uma sequência livre, é facilmente possível substituir o verbo por outros sem causar prejuízo para o sentido no enunciado:

(8) Eder *cortou a cabeça* do porco.

(8a) Eder (*cortou + amputou + decepou*) *a cabeça* do porco.

Em (8a), o verbo *cortar* pode ser substituído pelos verbos *amputar* e *decepar* sem perda de sentido para o enunciado. Pode-se também perceber no enunciado abaixo:

(9) Eder *livrou o porco* do chiqueiro.

(9a) Eder (*livrou + libertou + tirou + soltou*) *o porco* do chiqueiro.

Entretanto, quando se trata de expressões fixas, a fixidez impede que se faça tal substituição, como se pode ver abaixo:

(10) Eder **cortou as asas** do funcionário que só chegava atrasado.

(10a) Eder (**cortou + *amputou + *decepou**) *as asas* do funcionário que só chegava atrasado.

(11) Eder **livrou a cara** do amigo de um grande vexame.

(11a) Eder (**livrou + *libertou + *tirou + *soltou**) *a cara* do amigo de um grande vexame.

Tanto em **cortar as asas** como em **livrar a cara** não aceitam substituir seus respectivos verbos nas frases acima. Isso acontece porque em (10a), o verbo

cortar junta-se ao substantivo **asa** para formar uma expressão com um novo sentido, que é reprimir alguém. Da mesma forma, se substituir o verbo **livrar**, da sequência **livrar a cara**, como na frase (11a), tem-se a perda de seu sentido, que é sair de uma situação embaraçosa. A impossibilidade de substituição dos verbos, em (10a) e (11a), confirmam que essas sequências são realmente expressões fixas.

Apesar de a grande maioria das expressões fixas de estruturas argumentais VartN não aceitarem substituir o verbo por outro, como seu viu acima, algumas admitem a substituição do verbo, como abaixo:

(12) Eder **amarrou a cara** para Marcelo.

(12a) Eder (**amarrou + fechou + *prendeu + *tapou**) **a cara** para Marcelo.

Em (12a), a expressão fixa **amarrar a cara**, admite a troca do verbo **amarrar** pelo verbo **fechar** sem comprometer o sentido da expressão, que é demonstrar zanga ou desagrado. No entanto, nessa expressão fixa, a possibilidade da troca dos verbos **amarrar** e **fechar** por outros verbos sinônimos, como os verbos **prender** e **tapar** presentes em (12a), desfaz totalmente o sentido da expressão. Mas pode-se encontrar expressão fixa que admite a troca do verbo por outro sinônimo:

(13) Eder **fechou os ouvidos** para as fofocas do trabalho.

(13a) Eder (**fechou + tapou**) **os ouvidos** para as fofocas do trabalho.

A troca do verbo fechar pelo verbo sinônimo tapar, em (13a), não prejudica o sentido metafórico da expressão fixa **fechar os ouvidos**.

5.1.2.3 Determinante obrigatório

Nas expressões verbais, de estrutura argumental VartN, a sua fixidez se dá a partir da junção de alguns determinantes entre V e N, que podem ser:

a) **Determinante indefinido:**

(14) A irmã de Eder mal casou e já **carrega uma barriga**.

Essa expressão fixa é mais conhecida no estado do Maranhão. Trata-se de uma expressão regional, que significa estar **grávida**. Se trocar o determinante indefinido feminino **uma** por outros determinantes, tem-se, então, a quebra do seu sentido, e, por conseguinte, a perda de sua fixidez:

(14a) A irmã de Eder mal casou e já **carrega** (*uma + *a + *aquela + *esta + *sua*) **barriga**.

Entretanto, em algumas expressões fixas, como **conseguir uma boca**, há a possibilidade da troca do determinante indefinido por outros determinantes sem comprometer sua fixidez:

(15) Eder **conseguiu uma boca** na empresa do amigo.

(15a) Será que Eder **conseguirá** (*uma + aquela + outra + alguma*) **boca** na empresa do amigo?

Em (15a), a troca do determinante indefinido **uma** pelos determinantes **aquela**, **outra**, **alguma** não desfaz o seu sentido, que pode ser entendido como uma oportunidade de emprego.

Isso também acontece com esta expressão fixa **Tirar uma pestana**:

(16) Eder **tirou uma pestana** no trabalho.

(16a) Eder **tirou** (*aquela + outra + algumas + diversas*) **pestana(s)** no trabalho.

b) Determinante definido:

Na sequência **dobrar a calça**, como o enunciado abaixo:

(17) Eder fez Marcelo **dobrar a calça** diante de todo mundo.

Percebe-se que é totalmente possível a variação do determinante definido feminino por variados determinantes:

(17a) Eder fez Marcelo **dobrar** (*uma + esta + aquela + minha + sua + a maioria das + ambas + algumas + diversas*) **calça(s)** diante de todo mundo.

Essa flexibilidade na troca do determinante *a*, como em (17a), por outros determinantes só é possível porque a sequência *dobrar a calça* não caracteriza como uma expressão fixa. A falta de fixidez nessa sequência também permite que seja trocado o verbo *dobrar* por outros como *rasgar*, *sujar*, *amarrotar*, *lavar*, ou até mesmo trocar o substantivo *calça* por *camisa*, *bermuda*, *meia*, *cueca* sem nenhum problema. Pode-se dizer, então, que a sequência *dobrar a calça* é amplamente produtiva, por isso, não se percebe nenhuma fixidez nela. Mas quando se trata de uma expressão fixa, esta vasta possibilidade de substituição não é possível, como no exemplo abaixo:

(18) Eder fez Marcelo ***dobrar a língua*** diante de todo o mundo.

(18a) Eder fez Marcelo ***dobrar (a + *uma + *esta + *aquela + *a maioria das + *ambas + *alguma + *diversas) língua(s)*** diante de todo o mundo.

Nessa expressão, há possibilidade em aceitar certos determinantes possessivos como ***minha***, ***sua*** e ***nossa*** sem a perda de sentido:

(18b) Eder fez Marcelo ***dobrar (a minha + a sua + a nossa) língua*** diante de todo o mundo.

Essa possibilidade só acontece, como em (18b), porque os determinantes possessivos não substituem o determinante feminino *a*, mas são entendidos como uma inserção lexical na expressão. Assim, o sentido figurado da expressão fixa em (18) e (18b), que significa fazer alguém se reconsiderar, ou até mesmo falar com respeito, é mantido.

5.1.2.4 Negação obrigatória

Várias expressões, de estrutura argumental VartN, admitem passar para a forma negativa:

(19) Eder ***deu as caras*** por aqui logo cedo.

(20) Hoje, Eder ***não deu as caras*** por aqui.

No entanto, apesar de não ser uma constância, algumas expressões fixas aparecem, obrigatoriamente, na forma negativa:

(21) Todos foram embora, mas Eder **não arredou o pé** da festa.

(21a) *Todos foram embora, mas Eder **arredou o pé** da festa.

Ou, às vezes, dependendo do contexto, não há necessidade da forma negativa para que a expressão tenha um sentido negativo:

(21b) Todos ficaram, mas Eder **não arredou o pé** da festa.

Muitas vezes, a forma negativa de uma sequência, como em **não arredar o pé**, pode-se perfeitamente substituir o **não** por outras formas negativas sem o comprometimento de seu sentido:

(21c) Todos foram embora, mas Eder (**não + nunca + jamais**) **arredará o pé** da festa.

Dependendo de como está conjugado o verbo da expressão, também se aceita outras formas negativas:

(21d) Todos foram embora, mas Eder (**não + nem + sequer**) **arredou o pé** da festa.

Em (21c), por o verbo **arredar** estar no futuro, as formas negativas **nunca** e **jamais** são mais apropriadas para substituir a forma negativa **não**. Mas se o verbo **arredar** estiver no passado, as formas mais adequadas para substituir o **não** são **nem** e **sequer**, como em (21d).

A forma negativa **não** pode também ser substituída pela forma **ninguém**:

(21e) O dia amanheceu, mas **ninguém arredou o pé** da festa.

Ou até mesmo pelas formas **nunca** e **nenhum(a)**:

(21f) Pode o dia amanhecer, mas Eder **nunca arredará o pé** da festa.

(21g) A bebida e a comida acabou, mas **nenhum arredou o pé** da festa.

Ou por **tampouco**:

(21h) Quase todos foram embora, mas Eder não desanimou, **tampouco arredou o pé** da festa.

A negação também pode ser a preposição **sem**:

(21i) Todos continuaram **sem arredar o pé** da festa.

A negação também pode estar um pouco mais longe:

(21j) Eder **não pretende arredar o pé** daqui.

A negação também pode estar ausente, mas sem perder a noção de negação:

(21k) Por que Eder ia **arredar o pé** daqui?

(21l) Eder vai ter que **arredar o pé** daqui.

(21m) Quando Eder **arredar o pé** daqui...

5.1.2.5 Inserção de um elemento lexical

As expressões fixas tendem a não aceitar inserção de nenhum elemento lexical em seu grupo constituinte. Se isso acontecer, tem-se o total comprometimento do sentido da expressão:

(22) Até a mãe de Luciana, Vera Gimenez – musa do cinema nos anos 70 e 80 – **deu as caras** no “Superpop” (AT - 05/08/05).

(22a) Até a mãe de Luciana, Vera Gimenez – musa do cinema nos anos 70 e 80 – **deu** (*bastante + *muito + *de imediato) **as caras** no “Superpop”.

Entretanto, é perfeitamente possível inserir o item lexical **logo**:

(22b) Até a mãe de Luciana, Vera Gimenez – musa do cinema nos anos 70 e 80 – **deu (logo) as caras** no “Superpop”.

Por isso, dizer que somente sequências livres aceitam inserir itens lexicais e as expressões fixas não, por causa da fixidez é, de certa forma, arriscado. Dependendo do item lexical, essa inserção é totalmente possível sem causar nenhuma perda para o significado da expressão fixa:

- (23) Eder disse que Marcelo **deu as costas** quando precisou de um empréstimo.
- (23a) Eder disse que Marcelo **deu (lhe) as costas** quando precisou de um empréstimo.
- (24) Eder **livrou a cara** para não ser preso pela polícia por fraude.
- (24a) Eder **livrou a (própria + sua) cara** para não ser preso pela polícia por fraude.

Nos exemplos acima, como em (23a), a inserção do pronome **lhe** na expressão fixa **dar as costas** ocorreu entre o verbo **dar** e o artigo feminino **a** sem que houvesse problema no sentido da expressão. Também em (24a), a ocorrência do adjetivo **próprio** e do pronome **sua** entre o artigo feminino **a** e o substantivo **cara**, fazendo referência ao sujeito do enunciado, não desfez a fixidez da expressão **livrar a cara**.

5.1.2.6 Flexão em número de N

De acordo com as gramáticas tradicionais, os substantivos podem sofrer variação em número, gênero e grau. De modo geral, os substantivos das expressões livres, de estrutura verbal, são submetidos às regras gramaticais quanto à flexão. Entretanto, as expressões fixas, por serem opacas, costumam ter restrições quanto às regras gramaticais. Tais restrições são indícios de fixidez ou opacidade. Para comprovar se há realmente fixidez, o elemento substantival de VartN da expressão verbal será submetido à flexão em número, como nos exemplos abaixo:

- (25) Eder e Marcelo **torceram o nariz** para Mônica no trabalho.
- (25a) Eder e Marcelo **torceram (*os narizes)** para Mônica no trabalho.
- (25b) Eder **torceu o nariz** para Mônica no trabalho.
- (26) Ninguém do PT tem coragem de declarar que jamais fez negócios com Toninho da Barcelona, porque todos **queimaram a língua** quando negaram as acusações de Roberto Jefferson (AT-21/08/05).

(26a) Ninguém do PT tem coragem de declarar que jamais fez negócios com Toninho da Barcelona, porque todos **queimaram** (***as línguas**) quando negaram as acusações de Roberto Jefferson.

(27) Eder **queimou a língua** quando acusou Marcelo injustamente.

Tanto em (25a) como em (26a), os substantivos **nariz** e **língua** permaneceram invariáveis, apesar dos sujeitos das frases estarem no plural.

Na grande maioria das expressões fixas, de estrutura argumental VartN o substantivo sempre será apresentado no singular, como em (25) e (26), independentemente do sujeito no enunciado estar ou não no singular.

Também existem aquelas expressões fixas em que o N sempre aparecerá no plural:

(28) Eder **deu as caras** por aqui, depois foi embora.

(28a) Eder **deu** (***a cara**) por aqui, depois foi embora.

(29) Engenheiros do Hawaii **dão as caras** em Vitória (AT 24/06/05).

(29a) Engenheiros do Hawaii **dão** (***a cara**) em Vitória.

5.1.3 VprepN

As expressões fixas de estrutura argumental VprepN, ou seja, aquelas formadas por um verbo e um nome, unidas por uma preposição podem ser avaliadas pelos seguintes critérios:

- a) distribuição sintática dos itens lexicais;
- b) negação obrigatória;
- c) inserção lexical;
- d) substituição do verbo com outro verbo;
- e) preposição obrigatória;

f) variação em número de N.

5.1.3.1 Distribuição sintática dos itens lexicais

Através da distribuição dos itens, como no exemplo (30), pode-se perceber que a expressão **botar na cabeça** se trata de uma expressão fixa. Isso se confirma porque o verbo **botar** aceita apenas complementos abstratos como **idéia** ou **decisão** após o substantivo **cabeça**:

- (30) Eder **botou na cabeça** a **idéia** de abrir uma lanchonete.
- (30a) Eder **botou na cabeça** a **decisão** de abrir uma lanchonete.
- (30b) *Eder botou na cabeça a idéia e a mão.
- (30c) *ainda há bastante espaço na cabeça.
- (30d) *a cabeça de Eder está bastante pesada por causa do excesso de coisas que tem lá.
- (30e) *Eder pôs dentro da cabeça o celular e o bloco de anotação.

Após a distribuição sintática, tem-se a confirmação de que a junção do verbo **botar** mais o substantivo **cabeça** formam uma unidade lexical que denota uma **atitude** ou **decisão** (35a). Em (35b), por exemplo, a coordenação do substantivo abstrato **ideia** com o substantivo concreto **mão** prejudica o sentido metafórico da expressão fixa **botar na cabeça**.

Observando a distribuição sintática da expressão (36) **falar pelos cotovelos**:

- (31) Eder **falou pelos cotovelos** na reunião dos pais.
- (31a) *Eder usou os cotovelos e os microfones para falar.
- (31b) *Os cotovelos estão cansados de tanto falar.
- (31c) *Os cotovelos não pararam de falar.
- (31d) *para Eder gritar, ele abre bastante os cotovelos.
- (31e) *Os cotovelos podem falar.

(31f) *Sem os cotovelos, Eder ficaria mudo.

Anatomicamente, o ato de falar é restrito unicamente a uma parte do nosso corpo chamada de **boca**, situada na cabeça e delimitada externamente por lábios inferior e superior. A junção do verbo **falar** com o substantivo **cotovelo** só faz sentido quando se tem a intenção de formar um novo significado a partir dessa junção, que é falar excessivamente ou falar o que não deve.

O substantivo **cotovelo**, na anatomia geral, denota a parte posterior da articulação entre o braço e o antebraço, por isso, não tem a propriedade da fala, como se pode ver pela distribuição sintática acima, o que caracteriza como uma expressão fixa.

Em (31a), é possível falar pelos microfones, mas estes são usados apenas como instrumentos auxiliares da boca, para que a voz saia um pouco mais amplificada. Por si só, assim como o cotovelo, o microfone não tem a capacidade inata de produzir som.

Em **estar sem cabeça**:

(32) Hoje Eder **está sem cabeça** para desenhar.

(32a) *Eder está sem cabeça e sem lápis para desenhar

(32b) *Eder perdeu a cabeça e não sabe onde pôs

(32c) *Eder tirou a cabeça e pôs no conserto

(32d) * Eder não desenhou porque não tem cabeça.

O verbo **estar** indica sempre uma transitoriedade ou um estado acidental do sujeito em relação ao enunciado. Quando esse verbo vem acompanhado da preposição **sem** também pode denotar uma condição emocional apresentada pelo sujeito, que pode ser: estar sem condições de **raciocinar** ou **pensar**, que é o sentido conotativo da expressão fixa **estar sem cabeça**. Essa interpretação é possível, porque o significado da expressão não é deduzido pela soma dos seus constituintes.

O substantivo **cabeça** não é um objeto que Eder pode condicionalmente ficar sem, mesmo que seja temporariamente, como no exemplo a seguir:

(33) Eder está sem caderno / giz / pincel / caneta para desenhar.

Em (33), o ato de desenhar depende diretamente de objetos, como caderno, lápis, caneta, giz ou outros. Logo, **estar sem cabeça** expressa um sentido metafórico, que é estar sem condições, vontade ou paciência para desenhar, sendo, portanto, uma expressão fixa.

5.1.3.2 negação obrigatória

Assim como as expressões fixas, de estrutura argumental VartN, as expressões fixas, de estrutura argumental VprepN também apresentam algumas expressões que só admitem a forma negativa:

(34) Eder **não chega aos calcanhares** de Marcelo.

(34a) *Eder **chega aos calcanhares** de Marcelo.

De acordo com Vale (2001), a forma negativa **não** pode ser substituída por outras formas negativas. Assim, seguindo a intuição de Vale, nessa expressão acima, a forma negativa **não** pode ser substituída por outras formas negativas sem causar prejuízos ao seu sentido, como as formas **sequer, jamais, nunca, nem, ninguém, nenhum(a)**:

(35b) Eder (**sequer + jamais + nunca + nem**) **chega aos calcanhares** de Marcelo.

(35c) **Ninguém chega aos calcanhares** de Marcelo.

(35d) Eles podem até tentar, mas **nenhum chegará aos calcanhares** de Marcelo.

(35e) **Nada chega aos calcanhares** de Marcelo.

Mas nem todas as expressões fixas aceitam substituir sua forma negativa por qualquer outra forma negativa:

(36) Esse bolo **não será para o bico** de Marcelo.

(36b) Esse bolo (**sequer + jamais + nunca + *nem*) **será para o bico** de Marcelo.

Apesar de no *corpus* desta pesquisa, a ocorrência de expressões fixas, com formas negativas, serem poucas, não é possível inserir uma única observação para todas, como se pode ver entre (35b) e (36b). Cada caso é um caso, nem todas as expressões com a mesma estrutura apresentam de maneira igual, por isso é que devem ser analisadas uma a uma.

5.1.3.3 Substituição do verbo por outro verbo

Em grupos nominais livres, como *pegar no braço*, é perfeitamente possível fazer a substituição do verbo por outros verbos sem comprometer o sentido da sentença, como no exemplo abaixo:

(37) Eder *pegou no braço* de Marcelo com violência.

(37a) Eder (*pegou + agarrou + segurou*) no braço de Marcelo com violência.

Mas, quando se trata de expressões fixas, como em *pegar no pé* e *cair de boca*, essa substituição é bloqueada:

(38) Eder ***pegou no pé*** de Marcelo por causa do vestibular.

(38a) Eder (***pegou + *agarrou + *segurou***) ***no pé*** de Marcelo por causa do vestibular.

(39) O doce estava tão bom que Eder ***caiu de boca*** nele.

(39a) O doce estava tão bom que Eder (***caiu + *desequilibrrou + *escorregou***) ***de boca*** nele.

Diferentemente de (37a), em (38a) e (39a), por serem expressões fixas, não foi possível fazer uma substituição dos verbos *pegar* e *cair* por outros verbos. Quando essa coordenação é feita, desfaz o sentido figurado da expressão. A substituição do verbo *cair* por *desequilibrar* ou *escorregar* só é possível em uma sequência livre, como abaixo:

(40) Eder *caiu no doce que estava no chão*.

(40a) Eder (*caiu + desequilibrou + escorregou*) no doce que estava no chão.

Em (40a), essa substituição do verbo só é possível porque não se trata de uma expressão fixa, caso contrário, aconteceria um bloqueio, como aconteceu em (39a).

Apesar de não ser muito comum, algumas expressões fixas admitem substituir o verbo por outro verbo, como em **acabar em sangue**:

(41) A briga entre Eder e Marcelo **acabou em sangue**.

(41a) A briga entre Eder e Marcelo (**acabou + terminou**) **em sangue**.

Percebe-se que em (41a), a substituição do verbo **acabar** por **terminar** é perfeitamente possível. Isso também acontece com a expressão **botar na cabeça**:

(42) Eder **botou na cabeça** de Marcelo idéias horríveis.

(42a) Eder (**botou + pôs + meteu + enfiou + colocou**) **na cabeça** de Marcelo idéias horríveis.

Isso acontece também com a expressão fixa **carregar nas costas**:

(43) Eder **carrega nas costas** todo o trabalho da empresa.

(43a) Eder (**carrega + leva**) **nas costas** todo o trabalho da empresa.

É possível também encontrar expressão fixa que admite substituir o verbo por outro que não seja, necessariamente, sinónimo:

(44) Eder **matou no peito** a bola e chutou para o gol.

(44a) Eder (**matou + ajeitou**) **no peito** a bola e chutou para o gol.

Em (44a), na expressão fixa **matar no peito**, percebe-se que não relação sinonímica entre **matar** e **ajeitar**, entretanto, isso não impediu que o verbo **matar** pudesse ser substituído pelo verbo **ajeitar** sem prejudicar o sentido da expressão.

5.1.3.4 Inserção lexical

Diversas expressões fixas, de estrutura argumental VprepN, aceitam inserir algum tipo de item lexical em sua estrutura sem que ocorra a perda de seu sentido:

(45) A notícia **chegou pelas mãos** de Marcelo.

(45a) A notícia **chegou** (* **muito** + ***imediatamente** + ***bastante** + **novamente**) **pelas mãos** de Marcelo.

Em (45a), na expressão fixa **chegar pelas mãos**, apesar de não aceitar os itens lexicais **muito**, **imediatamente**, **bastante**, aceita o item lexical **novamente** sem comprometer o seu sentido.

Já a expressão fixa **sofrer nas mãos** admite os itens lexicais **novamente muito** e **bastante**:

(46) Eder **sofreu nas mãos** de Marcelo.

(46a) Eder sofreu (**novamente** + **muito** + ***imediatamente** + **bastante**) **nas mãos** de Marcelo.

Nessa expressão, o item lexical imediatamente causa estranhamento à expressão. No entanto, na expressão fixa chegar pelas mãos, a inserção do item lexical imediatamente é bem aceita:

(47) A notícia chegou pelas mãos de Eder.

(47a) A notícia chegou (**novamente** + ***muito** + **imediatamente** + ***bastante**) pelas mãos de Eder.

Tanto em (45a) como em (46a) e (47a), a inserção é feita entre o V e a prep da expressão fixa, mas existem expressões fixas, de estruturas argumental VprepN, que admitem inserir o item lexical entre a prep e o N:

(48) Hoje, Eder **sofreu na carne**.

(48a) Hoje, Eder **sofreu na (própria) carne**.

Em (48a), a inserção do item lexical *própria*, na expressão fica *sofrer na carne*, realça a idéia de alguém que passou por algum tipo de sofrimento.

5.1.3.5 Preposição obrigatória

Diversas expressões fixas, de estrutura argumental VprepN, tem, obrigatoriamente, em suas estruturas algum tipo de preposição.

Em algumas expressões fixas, a preposição *a* vem acompanhada do artigo definido *o* ou *a*:

(49) A notícia *chegou aos ouvidos* de Eder.

(49a) Eder *deu às costas* para Marcelo.

Ou somente pela preposição *de*:

(50) Eder *caiu de cabeça* no trabalho.

Ou a preposição *de* acompanhada do artigo definido *o*:

(50a) Depois que sua irmã casou, Eder *largou do pé* dela.

Ainda, pela preposição *em*:

(51) A briga *acabou em sangue*.

Tem-se também a preposição *em*, seguida do artigo definido *a* ou *o*:

(52) Eder *deu na cara* de Marcelo.

(52a) Eder *ficou no pé* de Marcelo.

Pode-se também ter a preposição *pelo*:

(53) Eder *trouxe* Marcelo *pelo beijo*.

A preposição *pelo* também pode ser acompanhada do artigo definido *a*:

(53a) A carta *chegou pelas mãos* de Eder.

Somente pela preposição *com*:

(54) Esta reunião *está com cara* de festa.

Ou pela preposição **para**:

(55) Esta vaga na empresa **não é para o bico** de Eder.

Ou ainda, pela preposição **sem**:

(56) Eder **está sem cabeça** para o concurso.

5.1.3.6 Variação em número de N

De modo geral, o substantivo em uma sequência livre, de estrutura verbal, de estrutura argumental VprepN, como a sequência *estar sem carro*, o substantivo *carros* aceita variar em número:

(57) Eder *está sem carro* para vender.

(57a) Eder *está sem carros* para vender .

Em (57) e (57a), o substantivo *carro* tanto pode ficar no singular como no plural sem nenhum problema para a sentença.

Entretanto, as expressões fixas, também com estrutura argumental VprepN, costumam ter restrições quanto à variação do N em número:

(57) Eder **está sem cabeça** para viajar hoje.

(57a) *Eder **está sem cabeças** para viajar hoje.

(57b) Eles **estão sem cabeça** para viajar hoje.

(57c) *Eles **estão sem cabeças** para viajar hoje.

A sequência em (57) **está sem cabeça** não aceita pluralizar o substantivo **cabeça**, mesmo que o sujeito da sentença esteja no plural, como em (57b).

Do mesmo modo, na expressão fixa **está de olho**, o substantivo **olho** também não é suscetível à variação em número:

(58) Os fiscais **estão de olho** na qualidade do peixe e do palmito vendidos aos consumidores (AT – 23/03/05).

- (58a) *Os fiscais **estão de olhos** na qualidade do peixe e do palmito vendidos aos consumidores.

No exemplo em (58a), a flexão do substantivo **olho** de singular para o plural desfaz a unidade de sentido da expressão, tornado-se uma sequência inaceitável para a sentença.

5.1.3.6.1 Plural obrigatório de N

Algumas expressões fixas apresentam obrigatoriamente o N no plural. Quando se tem uma expressão fixa, de estrutura argumental VprepN, com o N no plural, mesmo que o sujeito na sentença estiver ou não no plural, esse N fica invariável, como na expressão fixa **falar pelos cotovelos** abaixo:

- (59) Quando Eder **fala pelos cotovelos** ninguém aguênta.

- (59a) *Quando Eder **fala pelo cotovelo** ninguém aguênta.

- (59) *Quando eles **falam pelo cotovelo** ninguém aguênta.

Outros substantivos, como **mão** e **coxa** pertencentes às expressões fixas como, por exemplo, **ter em mãos** e **fazer nas coxas**, também são apresentadas somente no plural:

- (60) Eder **tem em mãos** as idéias para salvar a firma da falência.

- (60a) *Eder **tem em mão** as idéias para salvar a firma da falência.

- (61) Eder **fez nas coxas** a limpeza do carro.

- (61a) *Eder **fez na coxa** a limpeza do carro.

O emprego de N obrigatoriamente no plural nas expressões **ter em mãos** e **fazer nas coxas**, como em (60) e (61), revela certo grau de fixidez dessas expressões.

5.1.4 Estrutura argumental VN

As expressões fixas, de estrutura argumental VN, ou seja, aquelas constituídas apenas por uma forma verbal e um substantivo, também se mostrou bem produtiva no *corpus* desta pesquisa. Para descrever essas expressões e,

assim, incluí-las como status de expressões fixas, fez-se necessário submetê-las aos seguintes critérios:

- a) distribuição sintática dos itens lexicais;
- b) inserção lexical;
- c) negação obrigatória;
- d) ausência de determinante ou preposição;
- e) substituição do verbo com outro verbo;
- f) variação em número do N.

Antes de iniciar a descrição a partir dos critérios seleccionados acima, pode-se abordar, de modo rápido¹⁵, o comportamento da forma verbal nessa estrutura argumental.

A estrutura interna das expressões verbais VN não é tão regular, já que a forma verbal pode ser flexionada tanto na 1ª, 2ª e na 3ª pessoa do singular como na 1ª, 2ª ou 3ª pessoa do plural do modo indicativo:

- (62) Eu **abro mão** da herança
- (62a) Tu **abres mão** da herança
- (62b) Ele **abre mão** da herança
- (62c) Nós **abrimos mão** da herança
- (62d) Vós **abris mão** da herança
- (62e) Eles **abrem mão** da herança

¹⁵ É preciso salientar que a abordagem de maneira rápida da forma verbal, não só dessa estrutura argumental, mas também das outras estruturas argumentais verbais seleccionadas para esta pesquisa, se dá pelo simples motivo de essas formas verbais serem sempre submetidas à conjugação verbal em consonância com as regras gramaticais. O que varia muito é a forma nominal e adjetival dessas estruturas. Por isso que “nossos olhos” estão mais atentos às formais nominais e adjetivais, já que essas duas nem sempre seguem “à risca” as normas, como manda os preceitos das gramáticas tradicionais.

É também totalmente possível flexionar o verbo para todos os tempos e todas as pessoas do modo indicativo sem que ocorra nenhum dano ao sentido da expressão:

Presente: Eder **abre mão** da herança;

Pretérito imperfeito: Eder **abria** mão da herança;

Pretérito perfeito: Eder **abriu** mão da herança;

Pretérito mais-que-perfeito: Eder **abrira** mão da herança;

Futuro do presente: Eder **abrirá** mão da herança;

Futuro do pretérito: Eder **abriria** mão da herança.

Da mesma forma, não há nenhum problema em flexionar o verbo nos tempos do subjuntivo:

Presente do subjuntivo: espero que ele **abra** mão da herança;

Pretérito imperfeito do subjuntivo: se ele **abrisse** mão da herança, as coisas ficariam mais fáceis;

Futuro do subjuntivo: Quando ele **abrir** mão da herança, podes viajar tranquilo.

Assim como não também há problemas no modo imperativo:

Imperativo afirmativo: Eder, **abra** mão da herança!

Imperativo negativo: **não abra** mão da herança, Eder!

Ainda com relação ao verbo, há três formas verbais que por não ter precisão do tempo nem do modo, e são classificadas como formas nominais, que são:

Infinitivo: é necessário **bater perna** para comprar bons presentes;

Gerúndio: sem solução, o problema foi **tomando corpo**;

Particípio: (depois de) **tomado corpo** o problema, o pai procurou uma solução.

No infinitivo, o verbo **bater** não indica tempo, modo nem aspecto. No gerúndio, o verbo **tomar** indica um fato incompleto e há também certa duração, o que indica um aspecto imperfeito e durativo. Já no particípio, o verbo **tomar** indica um fato completo com relação a outro fato anterior, caracterizando um aspecto perfeito. Questões referentes ao aspecto verbal são muito interessantes, entretanto, não é objetivo desta pesquisa investir reflexões, neste momento, nessas questões.

Nas formas nominais citadas acima, os verbos podem vir precedidos de outros verbos, principalmente por verbos auxiliares, formando uma locução verbal:

- (63) Eder **vai bater perna** no shopping para comprar bons presentes;
- (63a) O problema **está tomando corpo** e a situação pode piorar;
- (63b) O problema **tem tomado corpo** de forma irreparável.

Esse conjunto de verbo é formado por:

Verbo auxiliar + forma nominal

No português brasileiro é muito comum usar um verbo auxiliar junto a outro verbo, isso também pode ocorrer até em expressões fixas sem nenhuma perda de sentido. Mas é na linguagem informal que a locução verbal acontece com tanta frequência. Podemos, por exemplo, dizer:

- (64) Eder **está abrindo mão** da herança = Eder **abre mão** da herança (presente).
- (65) Nós **vamos bater perna** no shopping = **bateremos perna** no shopping (futuro).
- (66) O problema já **tinha tomado corpo** = o problema **tomara corpo** (mais-que-perfeito).

5.1.4.1 Distribuição sintática dos itens lexicais

Através da distribuição sintática dos itens de uma sequência em uma sentença, como no exemplo abaixo, pode avaliar se essa sequência mantém o mesmo sentido ou se há alteração:

- (67) **Faltou perna** para Eder ganhar a corrida.
- (67a) *tem pouca perna no mercado para se comprar.
- (67b) *Eder não conseguiu uma perna a tempo para a corrida.
- (67c) *a escassez de perna atrapalhou Eder ganhar a corrida.
- (67d) *Eder se esqueceu de levar a perna para a corrida.
- (67e) *faltou perna e braço para Eder ganhar a corrida.

Não nos parece que foi a falta do objeto físico **perna** que fez com que Eder perdesse a corrida. Na sequência **faltar perna**, a forma verbal **Faltar** pode ser considerado um verbo transitivo indireto, pois pode-se perfeitamente dizer:

- (67f) **Faltaram lhe pernas** para ganhar a corrida

com o sentido de falha. Entretanto, a sequência perde seu sentido de fixidez se substituirmos o verbo **faltar** por **falhar**:

- (67g) ***falhou perna** para Eder ganhar a corrida.

O mau êxito de Eder na corrida nos remete não a uma **falha** ou **falta**, mas a uma **perda**, ou seja, o verbo **faltar** quando se junta ao substantivo **perna** para formar a expressão fixa **faltar perna**, passa a ter a interpretação de que Eder **perdeu** velocidade ou capacidade de correr e, a partir daí, passou a andar sem grande esforço, tornando impossível sua vitória.

Na expressão fixa **ter estômago**:

- (68) Eder **teve estômago** para resolver a situação da empresa.
- (68a) *O Estômago de Eder o ajudou na hora certa
- (68b) *Eder teve estômago e barriga para resolver a situação da empresa.
- (68c) *O estômago era de Eder
- (68d) *Sem seu estômago, Eder não conseguiria resolver a situação da empresa.

A distribuição sintática reforça a idéia de que a soma do verbo mais o substantivo em **ter estômago**, o verbo **ter** não indica uma posse com relação ao substantivo **barriga**, mas uma disposição, uma capacidade, um ânimo ou energia para enfrentar situações difíceis.

Em uma sequência livre, do tipo *ter dinheiro*, o verbo *ter* pode ser substituído pelo verbo *possuir* como também o substantivo *dinheiro* por outros substantivos, sem que haja problemas para o sentido da sentença:

- (69) Eder *teve* dinheiro para resolver a situação da empresa.
- (69a) Eder *possuía* dinheiro para resolver a situação da empresa.
- (69b) Eder *teve carro* para resolver a situação da empresa.
- (69c) Eder *teve* material didático para resolver a situação da empresa.

Já na expressão **ter estômago**, a fixidez impede essa mesma substituição do verbo **ter**, como em (70).

- (70) Eder **teve** estômago para resolver a situação da empresa.
- (70a) *Eder **possuía** estômago para resolver a situação da empresa.
- (70b) *Eder **conseguiu** estômago para resolver a situação da empresa.
- (70c) *Eder **arranjou** estômago para resolver a situação da empresa.

Essa quebra de sentido se dá, porque o substantivo **estômago** não é entendido como uma parte física, concreta, como acontece em (70), com relação ao sujeito da sentença, mas sim, a algo abstrato, metafórico.

5.1.4.2 Substituição do verbo com outros verbos

Em uma sequência livre, como *pegar bala*, é possível coordenar o verbo com outro:

- (71) A irmã de Eder *pegou bala* na festa.
- (71a) A irmã de Eder *pegou / deu e chupou bala* na festa.

Em (88a), os verbos *pegar*, *dar* e *chupar* são da mesma natureza sintática em relação ao substantivo *bala*, não havendo, portanto, nenhum problema com relação à coordenação entre os verbos. Mas quando se trata de expressões fixas, a coordenação entre esses mesmos verbos é bem restrita:

(71) A irmã de Eder **pegou barriga** na festa.

(71a) *A irmã de Eder **pegou / deu e chupou barriga** na festa.

A sequência **pegar barriga**, em (88), não aceita a coordenação entre os verbos **pegar**, **chupar** e **dar** porque se trata de uma expressão fixa. Não há uma relação sintática entre os verbos **dar** e **chupar** e o substantivo **barriga** com o sentido metafórico da expressão fixa **pegar barriga**.

É totalmente aceitável a sequência **dar barriga**, mas aí é outra expressão, com outro sentido. Já a sequência **chupar barriga** não é considerada como uma expressão, como também é pouco aceitável em uma sequência livre, num determinado enunciado, como em (71a), porque não há compatibilidade semântica entre o verbo **chupar** e o substantivo **barriga** nessa sentença.

Da mesma forma, não é possível coordenar os verbos de cada expressões fixas com outros verbos, como abaixo:

Bater perna = *bater e colidir perna

Abrir mão = *abrir e descerrar mão

Dar ouvidos = *dar e ganhar ouvidos

Nessas três expressões fixas, assim como em outras, os verbos coordenados não possuem compatibilidade semântica com relação aos sentidos figurados delas.

5.1.4.3 Inserção lexical

Nas expressões fixas, de estrutura argumental VN, algumas admitem inserir itens lexicais que haja alteração seus sentidos:

(72) Eder **bateu boca** com Marcelo.

(72a) Eder **bateu** (*muita* + *bastante*) **boca** com Marcelo.

Assim como na expressão fixa **bater boca**, em **quebrar cabeça** também admite inserir item lexical na sua estrutura:

(73) Eder **quebrou cabeça** na prova de matemática.

(73a) Eder **quebrou** (*bastante* + **muita* + **novamente*) **cabeça** na prova de matemática.

Em (73a), a inserção do item lexical **bastante** não causa nenhum problema para o sentido da expressão fixa **bater cabeça**, mas quanto ao uso do item lexical **muita**, a princípio, pode também não ter nenhum problema, entretanto, a palavra **muita**, diferentemente de **bastante**, não representa um advérbio de intensidade, mas a um adjetivo. Por isso, o correto seria usar o advérbio de intensidade **muito**, que também causaria estranhamento para o sentido da expressão, como se pode ver abaixo:

(73b) Eder **quebrou** (*bastante* + **muito* + **novamente*) **cabeça** na prova de matemática.

Esse estranhamento acontece porque para a aceitabilidade de **muito** na expressão, deve-se acrescentar o artigo definido **a** após o advérbio **muito**, o que também mudaria totalmente o sentido da expressão, passando a ter um sentido não mais figurado, mas a um sentido real:

(73c) Eder **quebrou** (**muito*) **a cabeça** na prova de matemática.

Diferentemente de **quebrar cabeça**, a sequência **quebrar a cabeça** não se caracteriza como uma expressão fixa, mas um grupo nominal livre.

Existem também aquelas que não aceitam nenhum tipo de inserção lexical em suas estruturas:

(74) Eder **abriu mão** da herança que iria receber.

(74a) Eder **abriu** (**bastante* + **muito* + **novamente*) **mão** da herança que iria receber.

O item lexical **novamente** só é aceito se for inserido antes ou após a expressão fixa, mas não no interior de sua estrutura:

(74b) Eder *novamente* **abriu mão** da herança.

(74c) Eder **abriu mão** da herança *novamente*.

Da mesma forma, a expressão fixa **tomar boca** não aceita item lexical na sua estrutura:

(75) A notícia logo **tomou boca**.

(75a) A notícia logo **tomou** (**bastante* + **muita* + **novamente* + **imediatamente*) **boca**.

Qualquer tentativa de inserção de algum dos itens lexicais presentes em (75a) desfaz a fixidez e, por conseguinte, quebra o sentido da expressão fixa **tomar boca**.

5.1.4.4 Negação obrigatória

Na estrutura argumental VN também se pode encontrar algumas expressões fixas que em sua estrutura tem-se a forma negativa, como na expressão fixa **não ter coração**:

(76) Eder **não tem coração**, pois vive maltratando os animais.

(76a) *Eder **tem coração**, pois vive maltratando os animais.

Essa mesma expressão pode ter um sentido afirmativo que se muda a segunda parte da sentença:

(76b) Eder **tem coração**, porque vive tratando bem os animais.

Na expressão fixa **não ter cabeça**, se retirar a forma negativa **não**, perde-se o seu sentido:

(77) Eder **não tem cabeça** para receber Marcelo.

(77a) *Eder **tem cabeça** para receber Marcelo.

Assim como nas estruturas argumentais VartN e VprepN, na estrutura argumental VN o forma negativa também pode ser substituída por outras formas negativas sem prejuízos para a expressão fixa:

- (78) Eder (**não + sequer + nunca + nem**) **tem coração**, pois vive maltratando os animais.

Nessa expressão também se aceita substituir o sujeito do enunciado pela forma negativa ninguém:

- (78a) **Ninguém tem coração**, pois vivem maltratando os animais.

5.1.4.5 Ausência de determinante ou preposição

Nas expressões fixas, de estrutura argumental VN, são formadas sem a presença de um determinante, seja ele definido ou indefinido:

- (79) Eder **comeu barriga** na entrevista.

- (79a) Eder **comeu (*a + *uma) barriga** na entrevista.

Tanto a presença do determinante definido feminino quanto a do indefinido feminino na estrutura da expressão fixa comer barriga desfaz o sentido de fixidez. Isso também acontece com a expressão fixa dar ouvidos:

- (80) Eder **não deu ouvidos** a Marcelo.

- (80a) Eder **não deu (*os + *uns) ouvidos** a Marcelo.

Mesmo não sendo comum, existem algumas expressões que, apesar de não ser obrigatório, aceita a presença facultativa de um determinante definido em sua estrutura:

- (81) Eder **bateu pé** e não aceitou a decisão do diretoria.

- (81a) Eder **bateu (o) pé** e não aceitou a decisão do diretoria.

Em (81a), o uso do artigo definido masculino **o** não alterou o sentido da expressão fixa **bater pé**.

Essa estrutura também não aceita a presença de nenhuma preposição:

(82) Eder **bateu boca** com Marcelo.

(82a) Eder **bateu** (**de* + **pela* + **na*) **boca** com Marcelo.

A introdução de preposição na sua estrutura expressão fixa **bater boca**, como em (82a), passa a ter um sentido inaceitável. Ainda em (82a), a presença da preposição **de** é até aceitável, o que caracteriza a expressão fixa **bater de boca**, mas com o sentido totalmente diferente da expressão fixa **bater boca** com alguém, que se remete uma discussão com alguém.

A expressão fixa **tomar pé** também não aceita nenhuma preposição:

(83) Eder **tomou pé** da situação na empresa.

(83a) Eder **tomou** (**de* + **com* + **em* + **ao* + **pelo*) **pé** da situação na empresa.

Em (83a), não foi possível introduzir nenhuma dessas preposições na estrutura da expressão fixa. Quando isso acontece tem-se a sua perda de sentido.

5.1.4.6 Variação em número de N

Em uma sequência livre com a estrutura argumental VN, o N é suscetível à variação em número:

(84) Foi necessário *abrir buraco* no asfalto para consertar a rede de esgoto.

(84a) Foi necessário *abrir buracos* no asfalto para consertar a rede de esgoto.

(84b) Foi necessário que *abrissem buracos* no asfalto para consertar a rede de esgoto.

Na sequência livre *abrir buraco*, o N, representado pelo substantivo *buraco* admitiu variação em número sem comprometer o sentido da sentença.

Já a maioria das expressões fixas com a mesma estrutura argumental VN, o N tende a não aceitar a variação em número:

(85) Eder **abriu mão** de tudo para viver um grande amor.

(85a) *Eder **abriu mãos** de tudo para viver um grande amor.

Na sequência **abrir mão**, que é uma expressão fixa, a variação em número do substantivo **mão** para **mãos**, como em (80a), perde totalmente o seu sentido. Essa perda de sentido acontece independentemente de o sujeito na sentença estar ou não no plural:

(86) Foi necessário que ele **abrissem mão** de tudo para viver um grande amor.

(86a) Foi necessário que eles **abrissem mão** de tudo para viver um grande amor.

(86b) *Foi necessário que eles **abrissem mãos** de tudo para viver um grande amor.

Há também algumas expressões fixas, de estrutura argumental verbal VN, que o N é apresentado obrigatoriamente no plural, como as expressões fixas abaixo:

Dar ouvidos, faltar braços, ter nervos, ter pernas

A fixidez sintática dessas expressões impede qualquer tentativa de variar seus substantivos de plural para singular:

(87) Eder **deu ouvidos** para tudo que ela disse.

(87a) *Eder **deu ouvido** para tudo que Lea disse.

(88) Faltou braços para Eder chegar até o outro lado do rio.

(88a) *Faltou braço para Eder chegar até o outro lado do rio.

Entretanto, com a expressão fixa **bater boca**, foi possível identificar algumas ocorrências curiosas. Ora ela é grafada com hífen:

(89) CPI vira motivo de **bate-boca** no Japão. O clima entre congressistas convidados por Lula esquentou após instalação de comissão (AT - 27/05/05).

Ora o N variando em número, mas mantendo o V no singular:

- (90) A sessão da Assembléia Legislativa foi marcada ontem por diversos **bate-bocas** entre os deputados (AT - 22/08/05).

Desta vez, o V e o N permaneceram no singular:

- (91) Jefferson fez questão de inocentar Lula (...), chegou a chamar José Dirceu de Rasputim e **bateu boca** com o presidente do PL, Valdemar Costa Neto (AT - 15/06/05).

Tem-se também o V no plural, que concorda com os sujeitos da sentença, entretanto, o N permanece invariável:

- (92) Tucano e petista **batem boca** (AT - 22/08/05).

Já abaixo, essa mesma expressão fixa, também grafada por hífen, está na posição de sujeito do enunciado de forma invariável, tanto do V quanto do N:

- (93) O **bate-boca** durou mais de uma hora e teve seu ponto alto em uma pesada troca de ofensas com Fernando Gabeira (AT - 31/08/05).

A diversidade de formas encontradas nessa expressão também é identificada em outras, como **bater papo / bate-papo** e **quebrar cabeça / quebra-cabeça**. Mas as expressões fixas, de estrutura argumental VN, na sua ampla maioria, não admitem variação em número do N.

No corpus foi possível também identificar algumas expressões fixas, de estrutura VN, em que o N tanto pode ficar no plural como no singular:

- (94) Eder **pôs ombros** à obra assim que chegou de viagem.

- (94a) Eder **pôs ombro** à obra assim que chegou de viagem.

Em (94a), a expressão fixa **por ombro(s)**, que significa dar início, é formado com o N tanto no plural como no singular.

Da mesma forma, a sequência fixa **cortar gordura** também é formada com o N tanto no singular como no plural:

- (95) Eder **cortou gordura** da sua refeição.

(95a) Eder **cortou gorduras** da sua refeição.

5.2 Expressões fixas de estruturas argumentais substantivais

Nas expressões fixas, de estrutura argumental substantival, foi possível identificar dois tipos, aquelas formadas por substantivos seguidos de adjetivos (Nadj) e aquelas formadas por dois substantivos ligados por preposições (NprepN), como na tabela abaixo:

SUBSTANTIVAL	EXEMPLOS	EFETIVOS
Nadj	Barba-azul	90
NprepN	Cabeça de vento	110
TOTAL		200

Tabela 6. Representação da estrutura argumental substantival.

5.2.1 Critérios de identificação das expressões fixas substantivais

Assim como as expressões fixas verbais, as substantivais também são analisadas a partir de critérios morfossintáticos e semânticos para uma possível comprovação da sua fixidez. Os critérios selecionados para as expressões fixas substantivais são aplicados de acordo com as estruturas das expressões, levando-se em conta as distribuições sintáticas e a interpretação linguística, para que, dessa forma, seja possível julgar se uma sequência é tida como fixa ou não.

5.2.2 Estrutura argumental Nadj

Nessa estrutura argumental, que é formada por um substantivo e um adjetivo pós-nominal, foi possível identificar expressões fixas com duas propriedades: (N0 ter Nadj) e (N0 ser Nadj). O símbolo N0 representa o sujeito livre que antecede a expressão fixa.

A propriedade (N0 ter Nadj) indica que a expressão fixa funciona apenas como uma *parte* do corpo pertencente ao sujeito livre (N0) na sentença como, por exemplo:

- (1) Eder tem um **cabelo bom**.
- (2) Kátia tem uma **língua afiada**.

Em (1), a expressão fixa **cabelo bom** indica que o substantivo **cabelo** e o adjetivo **bom** representam apenas uma parte do corpo que está relacionado ao sujeito livre (N0). Isso também acontece em (2).

Já a propriedade (N0 ser Nadj) indica que a expressão fixa não se relaciona apenas como uma parte do corpo do sujeito livre (N0) na sentença, mas o sujeito como um todo. Também pode ser entendido que um Nadj é um tipo de N0. Por exemplo:

- (3) Eder é um **pé-frio**.
- (4) Eder é um **cabeça-dura**.

Em (3), a expressão fixa **pé-frio** não representa apenas uma parte do corpo de N0, mas N0 como um todo, ou seja, representa um tipo de pessoa sem sorte, azarada. Da mesma forma, em (4), a expressão fixa **cabeça-dura** também representa o próprio sujeito livre (N0) ou uma pessoa que não aceita argumentos, não se deixa convencer facilmente, sendo chamada também de teimosa.

Ainda, como em (3), em (4) não se pode dizer que Eder tem uma **cabeça dura**, pois dessa forma, perde-se a noção de fixidez da sequência, passando a ter um sentido literal, de uma *cabeça* realmente *dura* e não mais como uma teimosia, que é o sentido da expressão fixa **cabeça-dura**. Devido a essas diferenças, não se pode enquadrar todas as expressões fixas substantivais, de estrutura argumental Nadj, em uma única propriedade.

A partir das observações acima, foi possível definir os seguintes critérios utilizados para as propriedades morfossintático-semânticas da estruturas argumentais Nadj:

- a) perda da predicatividade do adjetivo;
- b) coordenação do adjetivo com outro adjetivo;
- c) elisão do adjetivo;
- d) ruptura paradigmática;
- e) variação em número.

5.2.2.1 Perda da predicatividade do adjetivo

Em uma sequência livre, o adjetivo pós-nominal pode perfeitamente exercer a função de adjetivos predicativos com o acompanhamento de verbos copulativos¹⁶, como os verbos *ser* e/ou *estar*.

(5) Eder comeu uma *banana verde*.

(5a) Eder comeu uma *banana* que (era + estava) *verde*.

(5b) A *banana* (era + estava) *verde*.

Em (5a), o adjetivo *verde* após o substantivo *banana* se enquadra na função de adjetivo predicativo por causa dos verbos copulativos *ser* e *estar*.

O adjetivo *azul*, combinado com determinados substantivos como *carro*, também passa a ter uma função predicativa:

(6) Eder tem um *carro azul*.

(6a) Eder tem um *carro* que (é + *está) *azul*.

(6b) O *carro* (é + *está) *azul*.

Entretanto, quando esse mesmo adjetivo não aceitar a condição de predicativo com outros substantivos como ***sangue***, pode-se dizer, então, que houve perda de predicatividade nessa combinação, como abaixo:

(7) Eder tem ***sangue azul***.

(7a) Eder tem ***sangue*** que (*é + *está) ***azul***.

¹⁶ Para Almeida (1980, p. 239), os verbos copulativos são aqueles verbos que tem a função de unir um adjetivo a um substantivo.

(7b) O **sangue** (*é + *está) **azul**.

A perda da predicatividade na combinação do substantivo **sangue** com o adjetivo **azul**, na estrutura argumental Nadj, pode indicar forte sinal de fixidez sintática, o que eleva a sequência **sangue azul** à categoria de uma expressão fixa, com um significado totalmente diferente da soma das palavras que compõem essa sequência. O seu significado aponta exclusivamente para uma pessoa de ascendência nobre, como os fidalgos. Percebe-se, então, que a ocupação do adjetivo **azul** após o substantivo **sangue** só é aceitável nesse contexto, pois o adjetivo **azul** não é uma característica inerente ao substantivo **sangue**.

A perda da predicatividade também acontece com a sequência fixa **braço direito**:

(8) Eder era o **braço direito** do seu pai.

(8a) Eder era o *braço* que (*era + *estava) *direito* do seu pai.

(8b) O **braço** (*era + *estava) **direito** do seu pai.

Através das transformações acima, pode-se comprovar que quando se trata de uma expressão fixa, a perda da predicatividade ocorre com bastante frequência. Ou seja, nessas condições, o adjetivo não pode cumprir seu papel de predicatividade.

5.2.2.2 Coordenação do adjetivo com outro adjetivo

Em uma sequência livre do tipo Nadj, é possível coordenar o adjetivo com outros adjetivos:

(9) Eder tem um *olho ferido*.

(9a) Eder tem um *olho ferido* (*e + inchado + sangrento*).

Essa possibilidade de coordenação do adjetivo na sequência livre *olho ferido*, como em (9a), se dá porque os adjetivos *ferido*, *inchado* e *sangrento* são modificadores do substantivo *olho*.

Porém, em sequência fixa também do tipo Nadj, o adjetivo dificilmente poderá

ser coordenado:

(10) Eder tem um **olho gordo**, não pode ver nada que quer para si.

(10a) Eder tem um **olho gordo** (e + ***inchado** + ***sangrento**), não pode ver nada que quer para si.

Em (10), na sequência **olho gordo**, o substantivo **olho** com o adjetivo **gordo** forma uma nova unidade lexical, com um significado totalmente diferente da soma dos seus constituintes. Por isso, em (10a), não foi possível coordenar o adjetivo **gordo** com outros adjetivos, porque, nessa sequência, os adjetivos não são da mesma natureza sintática, ou seja, eles não são predicativos e nem modificam o substantivo **olho**.

A sequência fixa **cabeça pequena** também não admite coordenar o adjetivo **pequena** com determinados adjetivos, como abaixo:

(11) Eder tem uma **cabeça pequena** para esses assuntos.

(11a) Eder tem uma **cabeça pequena** (e + ***redonda**) para esses assuntos.

Essa coordenação do adjetivo **pequena** com o adjetivo **redonda** só seria aceitável se a sequência **cabeça pequena** fosse um grupo nominal livre:

(12) Eder nasceu com uma **cabeça pequena**.

(12a) Eder nasceu com uma **cabeça pequena** (e + **redonda**).

Na sentença (12a), a coordenação entre os adjetivos **pequena** e **redonda** é aceitável porque o contexto aponta para uma compatibilidade semântica entre eles.

Entretanto, se a sentença estivesse desta forma:

(13) Eder tem uma **cabeça pequena** e **quadrada**, não aceita os tempos modernos.

Em (13) apesar de haver compatibilidade semântica entre o adjetivo **pequena** e o substantivo **cabeça**, entretanto não há compatibilidade semântica entre o

substantivo **cabeça** e o adjetivo **quadrada**. Assim, a coordenação do adjetivo **pequena** e **quadrada** entre as sequências **cabeça pequena** e **cabeça quadrada** só são aceitáveis porque se tratam de expressões fixas. Caso contrário, seria problemático imaginar que um ser humano tenha uma cabeça pequena e ao mesmo tempo quadrada. Pode até ser potencial alguém nascer com uma deformação genética e ter uma *cabeça quadrada*, mas não é uma situação real. Isso seria um caso raríssimo que não pode ser levado em consideração. Logo é forçoso tornar aceitável a sequência livre na sentença abaixo:

(14) Eder nasceu com uma cabeça *pequena*.

(14^a) *Eder nasceu com uma *cabeça (pequena e + quadrada)*.

Existem também algumas expressões fixas que aceitam coordenar o adjetivo fixo com outros adjetivos que não fazem parte da estrutura da expressão fixa:

(15) Eder tem um **cabelo bom**.

(15a) Eder tem um **cabelo (bom e + curto + longo + liso)**.

A expressão fixa **cabelo bom** aceitou sem problemas coordenar o adjetivo **bom** com os adjetivos **curto**, **longo** e **liso**.

5.2.2.3 Elisão do adjetivo

Em uma sequência livre, do tipo Nadj, o adjetivo dessa sequência pode ser reduzido sem perda de sentido na sentença, como abaixo:

(16) Eder comprou uma *casa pequena*.

(16a) Eder comprou uma *casa*.

Em (16a), com a elisão do adjetivo modificador *pequena* na sequência *casa pequena*, reduzindo a sequência para *casa*, não alterou de forma drástica a estrutura da sentença. No entanto, quando se tem a elisão do adjetivo em posição pós-nominal de uma expressão fixa também do tipo Nadj, o seu sentido é essencialmente alterado:

(17) Fazem do sistema financeiro nacional o que bem entendem e anunciam, na maior ***cara dura***, lucros absurdos (AT - 22/08/05).

(17a) *Fazem do sistema financeiro nacional o que bem entendem e anunciam, na maior ***cara***, lucros absurdos.

A elisão do adjetivo ***dura*** da expressão fixa ***cara dura***, em (17a), alterou de forma significativa o sentido do enunciado, fazendo com que haja a total quebra de sentido da expressão fixa, tornando-a inaceitável.

Ainda, em uma sequência livre, como *olho azul*, o adjetivo dessa sequência, que também é um modificador, pode se retirado sem que haja danos para o sentido da sentença:

(18) Eder tem um *olho azul* que todo mundo nota.

(18a) Eder tem um *olho* que todo mundo nota.

A retirada do adjetivo modificador *olho*, como em (18a), foi perfeitamente possível. Também houve uma verificação que em frases definidoras, como em (16) e (18), a estrutura argumental Nadj representará um N. Quando isso acontece, é perfeitamente possível elidir o adjetivo sem alterar o significado do substantivo, pois:

Uma *casa pequena* é uma *casa*.

Um *olho azul* é um *olho*.

Entretanto, em muitas sequências fixas, como ***olho gordo***, a elisão do adjetivo ***gordo*** torna a sentença sem sentido ou altera de forma significativa o sentido no enunciado:

(19) Eder tem um ***olho gordo***, não pode ver nada que quer para si.

(19a) *Eder tem um ***olho***, não pode ver nada que quer para si.

A impossibilidade de elisão do adjetivo ***gordo*** em (13a), revela a fixidez sintática da sequência ***olho gordo***. Da mesma forma, não se trata de frases definidoras, nem o adjetivo *gordo* é tido como um adjetivo modificador, pois:

Um **olho gordo** não é um olho.

Assim como em (18), dificilmente poderá elidir o adjetivo **azul**, da expressão fixa **barba-azul**. Se isso acontecer, ora a sentença perde todo o sentido ora muda de sentido:

(20) Eder é um **barda-azul**, já ficou viúvo sete vezes.

(20a) *Eder é um **barba**, já ficou viúvo sete vezes.

Quando se trata de uma expressão fixa, a elisão do adjetivo, como em (20a), destrói o sentido da sentença.

5.2.2.4 Ruptura paradigmática

Em uma sequência livre, de estrutura argumental Nadj, é perfeitamente possível comutar o adjetivo com outros adjetivos em um mesmo paradigma distribucional:

(21) Eder disse que Kátia só usa *cabelo solto*.

(21a) Eder disse que Kátia só usa *cabelo (solto + preso + trançado + amarrado)*.

Em (21a), o substantivo *cabelo*, da sequência livre *cabelo solto*, foi modificado pelos adjetivos *solto*, *preso*, *trançado* e *amarrado* sem nenhum problema para o sentido da sentença.

Mas quando numa sequência de Nadj ocorrer uma ruptura paradigmática, pode-se, então, verificar fortes indícios de fixidez dessa sequência:

(22) Eder é mesmo um **bunda-mole**, não reage quando deve.

(22a) Eder é mesmo um **bunda (mole + *macia + *tenra)**, não reage quando deve.

Em (22a), a sequência **bunda-mole** não aceita uma distribuição paradigmática para o adjetivo **mole**. Essa ruptura paradigmática indica fortes indícios de fixidez da sequência acima.

Em **língua venenosa** também apresenta ruptura nos paradigmas

distribucionais para o adjetivo **venenosa**:

(23) Por ter uma **língua venenosa**, Eder não foi convidado para o casamento.

(23a) Por ter uma **língua (venenosa + *tóxica + *peçonhenta)**, Eder não foi convidado para o casamento.

A fixidez da expressão **língua venenosa**, como em (23a), bloqueia a distribuição sintática do adjetivo **venenosa** com outros adjetivos pertencentes ao mesmo campo semântico.

Da mesma forma, em **dedo duro**, o Adjetivo **duro** não permite nenhuma variação lexical dentro do paradigma distribucional:

(24) **Dedos duros** da noite. Com câmeras e celulares que tiram fotos, jovens saem à caça de micos, caras bonitas e romances e ganham fama (AT - 03/07/05).

(24a) **Dedos (duros + *rígidos + *sólidos)** da noite. Com câmeras e celulares que tiram fotos, jovens saem à caça de micos, caras bonitas e romances e ganham fama.

Apesar dos adjetivos **rígidos** e **sólidos** pertencerem ao mesmo campo semântico do adjetivo **duro**, a ruptura paradigmática acontece porque a sequência **dedo duro** forma uma expressão fixa.

5.2.2.5 Variação em número

Segundo Bechara (1988, p. 82), nas formações compostas de um substantivo e um adjetivo, ambos vão para o plural:

Cabra-cega = cabras-cegas

Amor-perfeito = amores-perfeitos

No *corpus* desta pesquisa, várias expressões fixas analisadas parecem estar em consonância com as regras gramaticais presentes em Bechara:

(25) o **barba-azul** chega hoje para mais um casamento.

(25a) os **barbas-azuis** chegam hoje para mais um casamento.

(26) Eder é um **bunda-mole**.

(26a) Eles são uns **bundas-moles**.

Essa regra parece funcionar, em sua grande maioria, quando se trata de sequências fixas ligadas por hífen. Entretanto, nem todas as expressões fixas, de estrutura argumental Nadj, são sequências ligadas por hífen. Várias expressões se apresentam sem esse traço de união. Nestas expressões fixas, nem sempre é possível flexionar o adjetivo em numero:

(27) Eder manteve a **cabeça fria** e não brigou.

(27a) Eder e Marcelo mantiveram a **cabeça fria** e não brigaram.

(27b) *Eder e Marcelo mantiveram as **cabeças frias** e não brigaram.

Na expressão fixa **cabeça fria**, quando se varia tanto o substantivo **cabeça** como o adjetivo **fria** em numero, perde-se o seu sentido de fixidez.

A expressão fixa **cara dura** também não admite variação nem do substantivo **cara** e nem do adjetivo **duro**:

(28) fazem do Sistema Financeiros Nacional o que bem entendem e anunciam, na maior **cara dura**, lucros absurdos (AT – 22/08/05).

(28a) *fazem do Sistema Financeiros Nacional o que bem entendem e anunciam, nas maiores **caras duras**, lucros absurdos.

Já na expressão fixa **sangue-frio**, apesar de ter um traço união, dependendo do contexto, ora o adjetivo pode variar ora pode permanecer no singular:

(29) A frase de que ele é do tipo **sangue-frio** veio de Roberto Justus, que mandou e desmandou no programa “O Aprendiz” (AT - 19/010/05).

(29a) A frase de que eles são do tipo **sangues-frios** veio de Roberto Justus, que mandou e desmandou no programa “O Aprendiz”.

(30) Os meninos (...) foram covardemente assassinados com pancadas na cabeça, desferidas a **sangue-frio**, sem que as crianças tivessem chance de defesa (AT - 03/08/05).

(30a) Os meninos (...) foram covardemente assassinados com pancadas na cabeça, desferidas a **sangues-frios**, sem que as crianças tivessem chance de defesa.

A expressão fixa **sangue-frio** foi usado em duas sentenças com contextos diferentes. Em (29), **sangue-frio** dá uma idéia de totalidade, ou seja, o sujeito da sentença é um **sangue-frio** e não do tipo que tem o **sangue-frio**. Já em (30), representa não mais uma totalidade, mas o modo como os meninos foram assassinados; e a preposição **a** antes da expressão fixa **sangue-frio** fez com o substantivo e, automaticamente, o adjetivo **frio** ficasse no singular.

Seja como for, ao fazer a descrição, é possível encontrar diversas observações com relação à variação em número do adjetivo na estrutura argumental Nadj. Essas observações vão muito mais além do que regras formais encontradas nas gramáticas. Por isso, para o tratamento automático das expressões fixas, não é possível encontrar nas gramáticas regras de flexões em número para que se possa aplicar em cada expressão fixa. Portanto, a flexão em número do adjetivo, deve ser feita caso a caso, pois, dessa forma, tem-se a certeza de que as informações observadas são essenciais e relevantes para serem inseridas em um ambiente computacional.

5.2.3 Estrutura argumental NprepN

A estrutura argumental NprepN, formada por dois substantivos, ligados exclusivamente pela preposição **de**, mostrou-se bastante produtiva. Seguindo o modelo da pesquisa de Smarsaro (2004), foi possível identificar três tipos de propriedades para essa estrutura:

- a) N1deN2 ser N1 – representa uma relação entre N1deN2 e N1. Por exemplo, **dente de leite** é um dente.
- b) N1deN2 ser N2 – representa uma relação entre N1deN2 e N2. Por exemplo, **cabeça de cebola** é uma cebola.

- c) N1deN2 não ser N1 nem N2 – não representa nenhuma relação entre N1deN2 em N1eN2. Por exemplo, **língua-de-sogra** não é uma língua nem uma sogra.

O N1 está relacionado ao primeiro substantivo e N2 ao segundo substantivo da sequência NdeN.

A estrutura argumental NdeN é bastante diversificada, pois pode-se encontrar grupos nominais livres, com estruturas formalmente idênticas aos grupos que caracterizam uma expressão fixa, como abaixo:

Cobertor de lã – caracteriza um grupo nominal livre.

Cobertor de orelha – caracteriza uma expressão fixa.

Qualquer falante nativo de uma língua saberia que *cobertor de lã* não é uma expressão fixa, mas ***cobertor de orelha*** sim. Isso se dá por causa da intuição e do conhecimento implícito que se tem da língua, mas o computador não tem essa intuição nem esse conhecimento. O processamento de uma língua consiste em compreender, gerar e interpretar uma língua natural por meio de representação dessa língua em língua artificial. Por isso, para sua investigação e maior clareza, com intuito de diferenciar os grupos nominais livres das expressões fixas, as propriedades dessa estrutura argumental serão analisadas a partir dos seguintes critérios:

- a) coordenação de grupos nominais;
- b) elisão de um dos elementos;
- c) inserção de um elemento lexical;
- d) ruptura paradigmática;
- e) variação em número.

5.2.3.1 Coordenação de grupos nominais

Em grupos nominais livres é possível coordenar as sequências, desde que elas tenham um mesmo referente:

(31) Eder comprou uma *camisa de algodão* e uma *camisa de seda*.

Também é possível pronominalizar o N1 da segunda sequência:

(31a) Eder comprou uma *camisa de algodão* e uma *de seda*.

Se uma das sequências coordenadas constituir uma expressão fixa, não é possível coordená-las porque os referentes são diferentes:

(32) *Eder conseguiu um *cobertor de lã* e um ***cobertor de orelha***.

Em (32), apesar da sequência livre *cobertor de lã* ter em N1 o mesmo item lexical de N1 da expressão fixa ***cobertor de orelha***, a inaceitabilidade ocorre porque os referentes são bem distintos.

A pronominalização também não é aceitável:

(32a) *?Eder conseguiu um *cobertor de lã* e um ***de orelha***.

Mesmo fazendo a inversão, não é possível pronominalizar:

(32b) *?Eder conseguiu um ***cobertor de orelha*** e um *de lã*.

O resultado também é o mesmo para a sequência livre *língua de vaca* e a sequência fixa ***língua de sogra***:

(33) *Eder comprou uma *língua de vaca* e uma ***língua de sogra***.

A pronominalização do N1 da segunda sequência também é bloqueada:

(33a) *Eder comprou uma *língua de vaca* e uma ***de sogra***.

Assim como em (32), em (33), os referentes são completamente diferentes. pois, diferentemente da sequência livre ***língua de vaca***, que denota literalmente uma língua, na sequência fixa ***língua de sogra***, o N1 *língua* forma com o N2 ***sogra*** uma nova unidade lexical, e aponta para um sentido bem diferente das partes que compõem essa unidade. ***Língua de sogra*** se refere a um tipo de apito, feito de papel enroscado, que, quando se sopra, desenrola fazendo um som e formando uma espécie de língua comprida.

Quando duas sequências fixas apresentam o mesmo referente, é possível coordená-las:

- (34) Eder conseguiu um **cabeça de área** e um **cabeça-de-chave** para compor seu time.

Também é possível pronominalizar o N1 da segunda sequência:

- (34a) Eder conseguiu um **cabeça de área** e um **de chave** para compor seu time.

A pronominalização é possível porque tanto em **cabeça de área** como em **cabeça de chave** os referentes apontam para uma mesma direção, ou seja, apontam para um tipo específico de jogador de futebol em uma equipe. Essa compatibilidade semântica evidencia a aceitabilidade tanto para a coordenação como para a pronominalização.

Existem também expressões fixas que o primeiro N de cada uma é o mesmo, mas não é possível coordená-los:

- (35) *Eder é um **cabeça de bagre** e um **cabeça de vento**.

Da mesma forma, não é possível pronominalizar o N1 da segunda sequência:

- (35a) *Eder é um **cabeça de bagre** e um de **vento**.

Não há compatibilidade semântica entre os referentes das expressões fixas **cabeça de bagre** e **cabeça de vento**, por isso não há possibilidade de coordenação ou pronominalização.

5.2.3.2 Elisão de um dos elementos

Nas expressões fixas NdeN, o apagamento do N2 desfaz totalmente a fixidez:

- (36) Eder ganhou uma **língua-de-sogra** na festa.

- (36a) * Eder ganhou uma **língua** na festa.

- (37) Eder comprou um **olho de boi** no leilão.

- (37a) *Eder comprou um **olho** no leilão.

Da mesma forma, dificilmente também se poderá elidir o N1 de uma expressão fixa:

(38) Eder teve **jogo de cintura** diante da situação.

(38a) *Eder teve **cintura** diante da situação.

(39) Eder comprou **olho de sogra** para o aniversário do filho.

(39a) *Eder comprou **sogra** para o aniversário do filho.

Em sua grande maioria não é possível o apagamento do N1 e nem do N2 porque ambos formam uma unidade lexical fixa.

Entretanto, foi possível encontrar dois exemplos de expressões fixas em que é possível elidir o N1 sem que houvesse nenhuma alteração no seu sentido:

(40) Eder usou uma **cabeça de cebola** para preparar o jantar.

(40a) Eder usou uma **cebola** para preparar o jantar.

(41) Eder usou um **dente de alho** no molho.

(41a) Eder usou um **alho** no molho.

Pode-se perceber que tanto na expressão fixa **cabeça de cebola** como em **dente de alho** ocorreu o apagamento do N1 sem ocasionar estranhamento ou mudança de sentido na sentença. Essa aceitabilidade se dá porque os substantivos **cabeça** e **dente** representam uma totalidade com relação ao produto, que é **cebola** e **alho**. Uma **cabeça de cebola** é uma **cebola**, assim como um **dente de alho** é um **alho**. Isso não acontece com as expressões fixas **língua-de-sogra**, **olho de boi**, **jogo de cintura** e **olho de sogra** e tantas outras. Não temos uma **língua-de-sogra** como uma **sogra**, mas um objeto que se enrola e desenrola imitando uma língua. Também **olho de boi** não é um **boi**, mas sim, um selo; **jogo de cintura** não é uma **cintura**, mas uma flexibilidade para solucionar problemas ou situações difíceis. Da mesma forma, **olho de sogra** não se caracteriza nem um **olho** e nem uma **sogra**, mas um tipo de doce feito com ameixa, recheada com uma massa de ovos e coco.

Pode-se perceber também essa totalidade em expressões como **cabeça de boi** ou **cabeça de gado**. Para Borba (1996, p. 130), a palavra **cabeça** tem uma relação *parte-todo*, pois tanto a designação da parte como o enfoque da posição da parte (parte mais alta / parte da frente) tornam preciso o valor do todo pelo esquema *parte (cabeça) + de + todo (N)*: *cabeça de boi / de alho / de alfinete*. No Houaiss (2009), a expressão **cabeça de gado** é considerada como uma unidade de um grupo. O exemplo abaixo esclarece o que Houaiss considera como unidade de um grupo:

- (42) O comerciante Ronaldo Luz, 31 anos, está desaparecido desde o dia 1, última segunda-feira, quando saiu de casa, em Itacibá, Cariacica, para receber cinco mil reais referentes à venda de sete **cabeças de gados** (AT – 17/07/05).

Borba (1996, p. 134) ainda menciona que partes de um todo passam a unidade de medida quando associados a nomes massivos, tendo, portanto, a projeção dirigida para a esquerda. Logo, o nome passa a unidade de medida por causa do nome massivo, como dez *cabeças* de gado, duas *colheres* (das de sopa) de manteiga.

Tem-se também a expressão fixa **carne de sol** que, por o N2 especificar um determinado tipo de N1, não permite o apagamento de N2:

- (43) Eder comprou uma **carne de sol** no supermercado.
 (43a) *Eder comprou uma **carne** no supermercado.

Com a elisão do N2, a frase é bem aceitável, mas não mais como uma expressão fixa, que representa uma distinção para N1.

5.2.3.3 Inserção de um elemento lexical

No *corpus* desta pesquisa, a estrutura argumental NprepN apresentou-se uma elevada fixidez sintática. Isso impossibilitou que se inserisse qualquer tipo de elemento lexical dentro da estrutura argumental NprepN:

- (44) Ganhei **jogo de cintura** com o “BandTempo” (TA – 22/08/05).
 (44a) *Ganhei **jogo de muita cintura** com o “BandTempo”.

(44b) *Ganhei **jogo muito de cintura** com o “BandTempo”.

A tentativa de inserção do item lexical modificador *muito* tanto para o N1 como para o N2 desfez totalmente a fixidez da expressão **Jogo de cintura**.

Esse mesmo modificador só é aceito se for posicionado antes da sequência fixa:

(44c) Ganhei *muito* **jogo de cintura** com o “BandTempo”.

(44d) *Ganhei **jogo de cintura muito** com o “BandTempo”.

Também é possível alternar o modificar *muito* com outros modificadores, desde que seja inserido antes da sequência fixa:

(44e) Ganhei (*bastante + pouco + elevado + enorme + o maior + significativo*) **jogo de cintura** com o “BandTempo”.

Algumas sequências fixas não aceitam inserir itens lexicais nem antes nem após a sequência:

(45) Eder elogiou Kátia, dizendo que ela tem uma **mão de fada**.

(45a) Eder elogiou Kátia, dizendo que ela tem uma (**grande + *nova + *boa + *interessante*) **mão de fada**.

(45b) Eder d elogiou Kátia, dizendo que ela tem uma **mão de fada** (**grande + *nova + *boa + *interessante*).

Já algumas admitem inserir um item lexical antes ou após a sequência fixa:

(46) Eder informou à policia que existe uma **boca-de-fumo** perto de sua casa.

(46a) Eder informou à policia que existe uma *enorme* **boca-de-fumo** perto de sua casa.

(46b) Eder informou à policia que existe uma **boca-de-fumo enorme** perto de sua casa.

A inserção do item lexical modificador *enorme* antes ou após a sequência fixa **boca-de-fumo** não alterou o seu sentido.

A constatação de que só é possível inserir alguns tipos de itens lexicais antes ou após a determinadas sequências fixas, mas nunca depois de N1 e antes de N2 nas estruturas argumentais NprepN, comprova a total fixidez dessas expressões.

5.2.3.4 Ruptura paradigmática

Em uma sequência livre NprepN, a comutação do substantivo por outros do mesmo paradigma é totalmente possível:

(47) Eder mandou fazer um **dente de ouro** para por na dentadura nova.

(47a) Eder mandou fazer um **dente de (ouro + prata + resina + diamante)** para por na dentadura nova.

Pode-se observar que os substantivos **prata**, **resina** e **diamante** podem substituir o substantivo **ouro** sem nenhum problema para a sentença. Isso acontece porque esses substantivos pertencem a um mesmo campo semântico em relação à sequência livre **dente de ouro**.

Mas, quando se tem uma sequência fixa, como **dente de leite**, não é possível constituir um paradigma distribucional:

(48) Eder retirou seu **dente de leite** quando tinha 7 anos.

(48a) Eder retirou seu **dente de (leite + *café + *chá)** quando tinha 7 anos.

Na sequência fixa **dente de leite** não foi possível estabelecer um paradigma como ocorreu com **dente de ouro**. Essa impossibilidade demonstra que o sentido para **dente de leite** não corresponde ao sentido da soma dos seus constituintes, como em **dente de ouro**, mas aponta para um dente que surge nas crianças entre os seis e os trinta meses de idade, e que por volta dos seis anos de idade, começa a ser substituído pelo dente permanente.

Outra sequência que não admite distribuição paradigmática é a sequência **rabo de saia**:

(49) Mal casou e Eder já arranhou um **rabo de saia**.

(49a) Mal casou e Eder já arranhou um **rabo de (saia + *vestido + *calça + *camisa)**.

A ruptura paradigmática em uma dada combinação, como nas combinações sintáticas **dente de leite** e **rabo de saia**, é um forte indício de que tal combinação seja realmente uma expressão fixa.

5.2.3.5 Variação em número

Grande parte das sequências fixas com estruturas argumentais NprepN são suscetíveis à variação em número. De modo geral, o N2 não admite variação em número:

(50) Nosso herói encontrou um malandro que lhe deu o endereço de uma **boca-de-fumo** localizada em um bairro da periferia (AT 13/05/05).

(51) As **bocas-de-fumo** do bairro da Penha têm grande movimentação de entorpecentes e o trabalho da repressão é um pouco mais complexo, em virtude da localização do bairro (AT 14/01/05).

(51a) *As **bocas-de-fumos** do bairro da Penha têm grande movimentação de entorpecentes e o trabalho da repressão é um pouco mais complexo, em virtude da localização do bairro.

(52) A polícia comandou com **mão-de-ferro** o combate ao tráfico na Grande Vitória (AT 10/06/05).

(52a) A polícia comandou com **mãos-de-ferro** o combate ao tráfico na Grande Vitória.

(52b) *A polícia comandou com **mãos-de-ferros** o combate ao tráfico na Grande Vitória.

Qualquer tentativa de pluralizar o N2 das sequências fixas **boca-de-fumo** e **mão-de-ferro** desfaz essa fixidez.

Algumas expressões fixas não admitem variar nem o N1 e nem o N2:

(53) Diante do espanto geral, o velho “*cana-dura*” explicou com a maior ***cara de pau*** do mundo (AT 14/08/05).

(53a) A ***cara de pau*** dos banqueiros é digna de comédia, pela maneira debochada com que vem apresentar números recordes (AT 22/08/05).

(53b) *As ***caras de pau*** dos banqueiros são dignas de comédia

(53c) *As ***cara de paus*** dos banqueiros são dignas de comédia

(53d) *As ***caras de paus*** dos banqueiros são dignas de comédia

Algumas já admitem variar o N1 e, conseqüentemente, o N2:

(54) Eder comprou uma ***cabeça de gado*** do Marcelo.

(54a) Eder comprou cinco ***cabeças de gados*** do Marcelo.

Algumas expressões fixas já apresentam o N2 obrigatoriamente no plural, como abaixo:

Calcanhar de Aquiles, Corpo de bombeiros, Língua de trapos, Menina dos olhos.

A partir dos exemplos acima, pode-se perceber que existem algumas formas de se variar em número as expressões fixas. Por isso, quando se havia a possibilidade de variar uma sequência NdeN, fazia-se, então, essa variação.

A ambiguidade

6.1 Conceituando

Em quase todas as línguas naturais existem milhares de palavras que possuem duplo sentido em uma construção sintática, permitindo mais de uma interpretação, constituindo assim, a ambiguidade. A ambiguidade é a possibilidade de uma construção ou sentença ter mais de um sentido. Isso muitas vezes pode distorcer o raciocínio lógico ou torná-lo obscuro, incerto ou equivocado. Por isso, a ambiguidade é, de certa forma, considerada como um vício de linguagem. Entretanto, na poesia, os autores se valem da licença poética para “lançarem mão” da polissemia¹⁷ das palavras para, de maneira proposital, utilizar a ambiguidade com intuito de fazer insinuações, trocadilhos ou até mesmo brincar com o locutor através de jogos de palavras. A ambiguidade também é muito utilizada como artifícios estilísticos em contextos literários, publicitários, provérbios, em frases ou sequências fixas e em tantos outros casos.

Segundo Zavaglia (2003b, *apud* SILVA, 2006, p. 23), os muitos fenômenos causadores da ambiguidade, como a polissemia, homonímia, metáfora, ambivalência, duplo sentido de palavras, revelam-se uma riqueza da expressão linguística.

Já Leffa (1996, p. 872) aponta que, do ponto de vista das gramáticas, há duas posições diferentes em relação à ambiguidade. Para a gramática prescritiva, a ambiguidade é simplesmente um fenômeno que deve ser evitado, mesmo que

¹⁷ A ambiguidade é considerada como um caso especial de polissemia, que, Para Boniatti e Bidarra (2005, *apud* Silva, 2006. P. 39), está presente nas palavras que têm a capacidade de assumir significados diferentes, mas mantendo uma relação de sentido entre elas.

possa ser resolvido por informações do contexto maior. Já na gramática descritiva, a ambiguidade é um fenômeno natural da língua, e cabe à gramática normativa analisá-lo do modo adequado, mostrando, fora de contexto, a estrutura profunda de cada interpretação.

Leffa também afirma que a ambiguidade pode envolver casos de polissemia, homonímia e metáfora. Para a autora, há polissemia quando os diferentes sentidos de uma palavra estão relacionados entre si; e cita como exemplo os sentidos da palavra *bico*, em sequências como *o bico da ave*, *o bico do sapato*, *o bico de gás*, em que todas compartilham do mesmo traço semântico, que é *extremidade aguçada*. Na homonímia, alguns casos podem até envolver palavras de diferentes etimologias; e geralmente não apresentam qualquer relação entre os significados, como por exemplo, a palavra *manga*, como parte do vestuário, que nada tem em comum com *manga*, o fruto da mangueira. Já a metáfora é vista como a aquisição de um novo sentido para uma mesma palavra. A palavra *braço*, por exemplo, pode adquirir o sentido de *afluente* na expressão *o braço do rio*. Segundo Leffa, quando isso acontece, e principalmente se a expressão ficar de uso corrente na língua, o que era inicialmente uma metáfora, uma transposição quase insciente de uma palavra para outro âmbito semântico, passa a ser, na realidade, mais um caso de polissemia.

Leffa considera, ainda, a ambiguidade como um fato acidental, intencional ou interlinguístico. Ela é acidental quando produzida não intencionalmente; neste caso ela não existe para o produtor do texto, mas apenas para o receptor. É intencional quando existe para ambos, logo, tem-se o duplo sentido. E finalmente, ela pode ser interlinguístico, isto é, quando existe apenas na passagem de uma língua para outra.

Silva (2006, p. 19) argumenta que a ambiguidade na língua está associada aos fenômenos da conotação e da polissemia. Para o autor, a conotação é entendida como um significado secundário ou subjacente que uma palavra possui, para além da acepção que é empregada. Já a polissemia está associada à qualidade de uma palavra ter muitos significados. Em consonância

com a argumentação do autor, convém citar alguns exemplos de sequências fixas que possuem ambiguidade conotativa, como abaixo:

- (1) Eder **enfeitou a testa** da esposa ao traí-la com a vizinha.
- (2) Eder **forrou o estômago** pela manhã, por isso está sem fome.
- (3) Eder sempre **pega no pé** da irmã em época de provas.
- (4) Eder **bateu perna** na rua o dia inteiro.

Seguindo a argumentação de Silva, em todos os exemplos supracitados, os verbos **enfeitar**, **forrar**, **pegar** e **bater** são usados com significados secundários, constituindo uma ambiguidade conotativa. Pode-se observar tal diferença, a partir dos exemplos abaixo, quando se usa os mesmos verbos em contextos sem ambiguidade conotativa:

- (5) Eder *enfeitou a mesa* para o aniversário.
- (6) Eder *forrou o botão* para seu novo paletó.
- (7) Eder *pegou no pé* da irmã e segurou firme.
- (8) Eder *bateu a perna* na árvore sem querer.

Já como ambiguidade polissêmica, têm-se os exemplos:

- (9) Eder ganhou um **pé de coelho**.
- (10) Eder comprou um **pé de cabra**.
- (11) Eder fotografou um **pé de moleque**.

Nas sentenças (9), (10) e (11), é possível perceber que se trata tanto do **pé** de um **animal** e de um **garoto** quanto de um **amuleto**, uma **ferramenta** e um **doce**.

Apesar da ambiguidade ser um recurso muito utilizado na linguagem informal, sobretudo no cotidiano pela modalidade oral, quando há presença do interlocutor desfaz uma dúvida eventual. No entanto, na modalidade escrita, a ambiguidade torna-se um recurso indesejável. Isso porque, uma vez inserida

determinada palavra com mais de um sentido, a interpretação correta da mensagem fica comprometida, já que nem sempre é possível a presença do emissor da mensagem para desfazer a dúvida sobre a verdadeira intenção comunicativa.

6.2 O problema da ambiguidade para o PLN

Para Ranchhod (2003, p. 213), na análise automática de textos, um dos grandes problemas que se coloca é o da sua eventual ambiguidade. De fato, apesar de as palavras simples serem bastante ambíguas, existem também muitas sequências semanticamente ambíguas, que tanto podem ser interpretadas como uma expressão fixa, constituindo um único item lexical, como também podem ser consideradas como sequências livres de palavras simples, como demonstrado em (5), (6), (7) e (8).

No exemplo abaixo:

(12) Eder tem a **boca suja**.

A sequência **boca suja**, na sentença (12), não deixa claro de que se trata de uma pessoa que costuma dizer palavrões ou obscenidades ou se está referindo a uma pessoa, cuja boca está realmente suja de alguma coisa. Para desfazer tal ambiguidade é preciso observar os complementos que indicam o sentido da sequência, deixando clara a real intenção da sentença:

(13) Eder tem a **boca suja**, pois vive xingando as pessoas.

(14) Eder tem a **boca suja**, pois acabou de almoçar.

Já nas sentenças abaixo:

(15) Eder é um **boca suja**.

(16) Eder está um **boca suja**.

Em (15) e (16), a escolha dos verbos **ser** e **estar** impede que se tenha qualquer dúvida com relação ao substantivo **boca** e ao adjetivo **sujo**. Percebe-se, então, que a expressão fixa **boca suja**, acompanhado do verbo **ter**, como

em (13) e (14), traz uma ambiguidade que só se desfaz com complementos que revelam o significado na sentença.

Os exemplos acima demonstram porque a ambiguidade tem se apresentado como uma das grandes preocupações àqueles que se propõem a descrever e formalizar uma língua natural para o PLN. Quando um sistema de processamento automático encontra diferentes significados para uma mesma palavra, é necessário que esse sistema esteja preparado para distinguir e determinar, de acordo com o contexto, qual o significado mais correto, para que, assim, o usuário tenha uma resposta mais precisa daquilo que busca. Por isso, quando se tem uma palavra ambígua, há sempre um risco para uma má interpretação ou incompreensão da informação solicitada; e no PLN, a ambiguidade pode causar falhas irrecuperáveis em uma análise informacional. Deve-se, então reforçar os cuidados quanto à ambiguidade das palavras para que elas não interfiram na precisão dessa informação quando processada.

A ambiguidade, que é causada pelos fenômenos linguísticos, como o morfológico, lexical, semântico e pelo fenômeno extralinguístico, como o pragmático, por também levar em consideração o uso, torna-se um problema para o sistema computacional se as bases de conhecimentos disponíveis nesse sistema não estiverem devidamente descritas. Desse modo, em um enunciado com informação ambígua, toda a informação que até então se pensava ser relevante, passa a ser irrelevante ou confusa, causando, inevitavelmente, a perda parcial ou total da dada informação. Assim, como foi afirmado anteriormente, a ambiguidade ocasiona falhas irrecuperáveis na informação durante o processamento de um texto em uma língua natural.

Laporte (2001, p. 53), menciona que numa ambiguidade lexical, uma palavra pode perfeitamente tornar ambígua a frase onde se encontra; e cita como exemplo a palavra *decorar*.

(17) Os alunos **decoraram** o soneto.

(18) as vendedoras **decoraram** a vitrina.

Para o autor, em (17), pode também perfeitamente ser possível interpretar que os alunos **decoraram** a página do soneto com desenho nas margens. Por isso

é que os problemas encontrados no tratamento automático de um texto, causados pela ambiguidade, podem gerar resultados incoerentes ou inadequados por programas de PLN. Para evitar grandes problemas, um estudo investigativo e adequado da ambiguidade passa a ser uma tarefa básica para o PLN, e a interpretação precisa desse fenómeno linguístico, em particular para a Tradução Automática, faz toda a diferença, especialmente porque, oposto do ser humano, a máquina não pode contar com a experiência e nem com o conhecimento de mundo em uma língua natural.

6.3 Resolução de ambiguidade

Na linguística, utiliza-se o termo desambiguação quando se refere ao processo de explicação a um determinado enunciado que possui mais de um sentido. Foi visto que um termo torna-se ambíguo, quando esse termo traz uma mensagem ou instrução confusa, sendo interpretada por mais de uma maneira. Por isso, como afirma Ide e Véronis (1998, p. 1), a desambiguação é essencial para aplicações de compreensão da linguagem como, por exemplo, a compreensão da mensagem, da comunicação homem-máquina etc. Portanto, a desambiguação é pelo menos útil e aplicável em casos onde não se tem o entendimento da linguagem.

Martins (1999, p.1) determina que,

Por mais que tenha fracassado a tentativa de aprisionamento da língua natural na aritmética, e por mais que a pragmática linguística tenha operado a recusa do reducionismo lógico, este *Aufbau* persiste ainda como fundamento principal de número considerável de sistemas de processamento automático das línguas naturais, quer pela absoluta falta de outros meios digitalizáveis de representação do significado, quer pela ignorância dos problemas que a abordagem oferece, quer pela crença de que as limitações identificadas são antes periféricas ao desempenho das ferramentas. A representação unívoca (lógica) do conteúdo semântico das sentenças pontua, sob a forma de interlíngua, nos sistemas de tradução automática que, sob o argumento da portabilidade, preferem prever, entre duas línguas naturais, uma língua intermediária artificial, comum não apenas às duas línguas comparadas, mas a todas as línguas comparáveis. Esta representação semântica universal conservaria todo o poder expressivo (ou pelo menos todo o poder representativo) das línguas naturais, sem envolver, porém, os acidentes (como a homonímia e a ambigüidade sintática) que as caracterizam e tanto ruído produzem durante o seu processamento.

Segundo Martins, os acidentes, nas línguas naturais, não são exatamente acidentais, no sentido aristotélico do termo, pois a ambiguidade e a imprecisão

não são elementos contingenciais, mas constitutivos das línguas naturais, pelo que a formalização de uma representação semântica deve necessariamente contemplar a indeterminação inerente às línguas naturais, sob o risco de verificar-se uma considerável dialeção da linguagem lógica, mecanismo compensatório de sua falta de plasticidade.

Ainda em Martins (1999, p. 5), admitir a existência de uma ambiguidade fundacional na semântica da linguagem, ambiguidade que não pode ser expurgada por nenhum movimento metalinguístico (a própria metalíngua, sendo linguagem, é ambígua), e que deve ser incorporada como elemento fundamental no processo de codificação parece menos ingênuo. Para Martins, não haverá jamais metalíngua completamente unívoca, menos por defeitos intrínsecos à especificação do que pela heterogeneidade de seus usuários. Segundo o autor, trata-se, neste caso, de investigar como a ambiguidade pode ser inserida na definição da própria especificação, de forma a não impedir a heterogeneidade do uso, mas a controlar suas implicações.

Os homens sempre usaram a língua natural para a comunicação entre si. Entretanto, quando se trata de PLN, é necessário que o homem dê à máquina recursos suficientes para que ela também utilize essa mesma língua natural emprestada por uma língua artificial na interação entre homem-máquina. Por isso, a ambiguidade, que é algo inerente à língua natural, está constantemente presente no PLN. Isso quer dizer que, para elaborar tal tecnologia, é necessário explorar todos os níveis linguísticos, tais como, o léxico, a morfologia, a sintaxe, a semântica, e a pragmática.

Considerando, então, os níveis linguísticos, Silva (2006, p. 23) diz que é possível afirmar que a ambiguidade se apresenta de diversas maneiras. A multiplicidade conceitual ou categorial que uma palavra pode assumir, caracterizando a ambiguidade léxica, ou multiplicidade relacionada à ambiguidade estrutural, na representação sintática de uma frase, pode gerar erros no processo do texto executado pela máquina e, conseqüentemente, alterações no sentido real trazido pelo texto. Ainda segundo o autor, a resolução de ambiguidade por meio de PLN enfrenta vários problemas de difíceis resoluções. Isso porque o conhecimento de contexto que não está

explícito no texto analisado, e o conhecimento de mundo ou costumes de uma determinada região, um objeto, um ser ou uma realidade que possui um significado específico pode, simplesmente, não ter o mesmo significado em outra região.

Segundo Ranchhod (2003, p. 216), a resolução da ambiguidade passa pela elaboração de gramáticas específicas. Para Laporte (2001, 51), a resolução da ambiguidade é uma operação básica cujos resultados são úteis para todos os processamentos de textos escritos, inclusive para alguns dos processamentos mais simples. Entretanto, é possível perceber que a resolução da ambiguidade não é uma tarefa tão fácil. E de fato não é. Para o autor, o processamento de textos escritos não pode ser feito sem informações linguísticas relativas às palavras; e para dispor de tais informações rápida e convenientemente, os programas informáticos costumam associá-las às próprias palavras dos textos sob a forma de etiquetas lexicais. A etiqueta lexical de uma palavra reúne, portanto, todas as informações linguísticas disponíveis sobre ela e úteis ao processamento, desde a forma que consta no texto até aos dados gramaticais, morfológicos, sintáticos ou semânticos, dependendo da natureza do processamento.

Laporte (2001, 51) ainda afirma que uma etapa primordial consiste em segmentar o texto, identificar as unidades mínimas e representá-las por etiquetas; e essa etapa é designada como análise lexical ou etiquetagem lexical. Por meio da etiquetagem lexical, o programa procura as palavras do texto num dicionário que associa etiquetas a todas as palavras da língua. Para Laporte, este método dá os resultados mais fiáveis, mas é preciso que o conteúdo do dicionário esteja conforme à realidade da língua e suficientemente perto da exaustividade descritiva da palavra.

Segundo Laporte, nos exemplos abaixo:

- (19) Um especialista não se **forma** facilmente.
- (20) O contrato foi redigido de **forma** a satisfazer ambas as partes.

A necessidade de associar duas etiquetas diferentes ao verbo (**forma**), conforme (19) e ao nome (**forma**), conforme (20), como etapa prévia ao reconhecimento da estrutura sintática das frases é evidente.

Ainda para Laporte, nos exemplos abaixo:

(21) O gato **foge**.

(22) Por favor, **foge** agora.

Nos dois elementos (**foge**) de um mesmo paradigma dos exemplos (21) e (22), as etiquetas respectivas diferem por traços flexionais, no caso, o presente do indicativo, em (21) ou o imperativo, em (22).

Laporte segue afirmando que, em geral, as propriedades sintáticas das várias acepções são diferentes: no caso dos verbos, o número de complementos essenciais, as preposições associadas, as transformações sintáticas aplicáveis, a distribuição dos nomes que podem ser colocados na frase na posição do sujeito ou do complemento podem diferir. E tais informações sintáticas, necessárias para o reconhecimento da estrutura da frase, devem ser incluídas nas etiquetas lexicais.

6.3.1 Resolução de ambiguidade entre as sequências livres e fixas

Segundo Laporte (2001, p. 53), a tradição subestima a importância das palavras compostas nos textos. Para o autor, o conteúdo técnico de um texto, inclusive a maioria dos termos técnicos, encontra-se mais nessa parte do que na outra, formada por palavras simples. Laporte acredita que, em comparação com a tecnologia padrão, geralmente aceita pela comunidade de linguística computacional, a operação de etiquetagem lexical aplica-se a unidades muito mais complexas do que as palavras simples. Logo, por as sequências fixas representarem uma parte significativa do léxico em uma língua, etiquetar só as palavras simples seria limitar-se a uma parte superficial da língua.

Como foi mencionado, as sequências formadas por combinações de palavras simples constituem-se expressões fixas em quantidades surpreendentes, como por exemplo, **dar as caras**, **fazer nas coxas**, **ter estômago**, **braço-de-ferro**, e **cabeça oca**. Por serem fixas, essas sequências possuem restrições em suas

propriedades estruturais. Entretanto, há várias sequências que são ambíguas. Isso faz com que sejam analisadas ora como sequências livres de palavras simples, ora como expressões fixas.

Observa-se em:

(23) O **braço direito** do chefe pediu um café.

(24) O *braço direito* do chefe está fraturado.

Apesar de as duas sentenças conterem as mesmas sequências (**braço direito**), o sentido em ambos é o mesmo. Em (23), diferentemente de (24), tem-se uma unidade mínima de significado, constituindo uma expressão fixa. Por isso, a etiqueta deverá conter informações sintáticas e semânticas relacionadas com a forma e o tipo de complementos das sequências, como em (23) **braço direito**, expressão fixa que representa alguém eficaz e principal auxiliar, e em (24), *braço direito*, sequência livre que, linguisticamente, expressa uma parte do corpo.

Ranchhod (2001, p. 21) orienta que, se um sistema de análise lexical contemplar a existência de compostos e der prioridade à identificação destas sequências, como **andar modelo** e **a curto prazo**, serão analisadas como um bloco e, respectivamente, funcionando como um **nome** e um **advérbio** e receberão apenas essas etiquetas.

Logo, a sequência fixa **a curto prazo** deve ser analisada da seguinte forma:

A curto prazo = advérbio

E não:

a = determinante

a = preposição

a = pronome

curto = adjetivo

prazo = nome

como se fosse uma sequência livre, com a necessidade de analisar individualmente as palavras que compõem essa sequência.

Por isso, a sequência **andar modelo**, em um sistema que não leva em conta a existência de sequências fixas, seja qual for a situação linguística em que se encontre, será analisada individualmente como palavras simples e terão, pelo menos, dois valores (duas etiquetas) cada uma, como **verbo** e **nome**. Portanto, para PLN, a sequência fixa **braço direito**, como em (23), deve ser analisada somente como um único item lexical e não como dois itens lexicais, no caso, um **nome** e um **adjetivo**. Assim, terá a entrada de apenas uma etiqueta e não de duas, como a sequência livre *braço direito*, em (24).

Entretanto, Ranchhod (1998, p. 215) argumenta que é totalmente possível identificar sequências fixas que permitem igualmente uma análise como sequências livres. A autora cita como exemplo a sequência **mesa redonda**:

(25) Eles detestam **mesas redondas**.

(26) Eles compraram *mesas redondas*.

(27) Eles organizaram **mesas redondas**.

Em (25), **Mesa redonda**, na posição de complemento do verbo **detestar**, confere à frase uma ambiguidade irreduzível; em (26), *mesa redonda* constitui uma sequência livre; e em (27), forma uma sequência fixa.

Para a autora, tanto no caso das sequências livres como das fixas, os resultados da análise lexical só serão adequados, ou aproximar-se-ão dos desejados se o analisador utilizar informações sintáticas integradas em gramáticas de resolução de ambiguidades.

Segundo Laporte (2001, p. 57-58), existem exemplos de frases ambíguas em relação às quais a resolução total das ambiguidades lexicais necessita da análise completa das estruturas sintáticas envolvidas. Para o autor, a ambiguidade da sentença:

(28) Encheu a **colher de chá**

provém da ambiguidade lexical do nome **colher de chá**. Pois, numa das interpretações, o nome composto **colher de chá** é o complemento direto do verbo **encher**, e o segundo complemento essencial do verbo está ausente. Em outras interpretações, o nome **colher** é o complemento direto, e **de chá** é um segundo complemento. Então, como esse complemento é facultativo, as duas interpretações são possíveis. Para o autor, é preciso fazer uma análise completa da frase, a partir das propriedades sintáticas do verbo, para chegar a essa conclusão. Laporte ainda complementa que uma análise análoga da seguinte sentença leva à resolução da mesma ambiguidade:

(29) Encheu a **colher de chá** de sal

Nesse exemplo, a posição pré-nominal ou pós-nominal dos adjetivos fornece outros exemplos em que a resolução total das ambiguidades lexicais requer uma análise sintática mais aprofundada do que a que foi antes referida.

Laporte salienta que a determinação correta das etiquetas lexicais associáveis às palavras pode depender do reconhecimento da estrutura sintática global da frase. Pois, trata-se de uma dependência circular, já que o reconhecimento das estruturas sintáticas se baseia nas informações lexicais incluídas nas etiquetas. Assim, uma das dificuldades inerentes à análise sintática é precisamente essa dependência circular.

Para Ranchhod (2003, p. 213), com intuito de analisar corretamente as sequências ambíguas, que descontextualizadas pertencem mais de uma categoria gramatical, constrói-se gramáticas que restringem as possibilidades de co-ocorrência das diversas categorias gramaticais. Ainda em Ranchhod, essas gramáticas estão formalizadas em transdutores de estados finitos e são aplicadas ao texto, em combinação com os dicionários [eletrônicos] de palavras simples e compostas, pelo sistema INTEX¹⁸.

¹⁸ O sistema INTEX é uma ferramenta computacional que utiliza dados linguísticos especificamente elaborados por especialistas em PLN. Nesse sistema, os textos, os dicionários e as gramáticas são representados por transdutores de estado finitos, que tem a forma de grafos (SILBERSTEIN, 1997 *apud* RANCHHOD, 2001, p. 8). Esta característica do sistema permite o desenvolvimento e a aplicação ao processamento de grandes *corpora*, de dicionários e de gramáticas de ampla cobertura. Em operações de análise automática de texto, os dicionários e gramáticas são aplicados em combinação, a fim de, entre outras coisas: (i) indexar o texto de vários modos possíveis; (ii) reconhecer unidades lexicais, simples e compostas; (iii) identificar estruturas sintáticas ou léxico-sintáticas; (iv) resolver ambiguidades; (v)

De acordo com Ranchhod, a Fig. 1 abaixo representa parcialmente uma dessas gramáticas; e a Gramática da Fig. 1 é bastante simples, pois se aplica a sequências que contenham um nome, especificado por um ou mais determinantes, representados na subgramática *DET* (nó sombreado), e modificado por um adjetivo predicativo de cor, *A+Pco*. A autora ainda acrescenta que, em português, os constituintes da sequência têm de concordar em gênero e número, e duas estruturas paralelas estão ainda representadas na gramática, uma em que os constituintes se encontram no masculino singular e outra no feminino singular. Na versão integral da gramática estão, naturalmente, especificadas as estruturas cujos constituintes são formas do masculino e feminino plural.

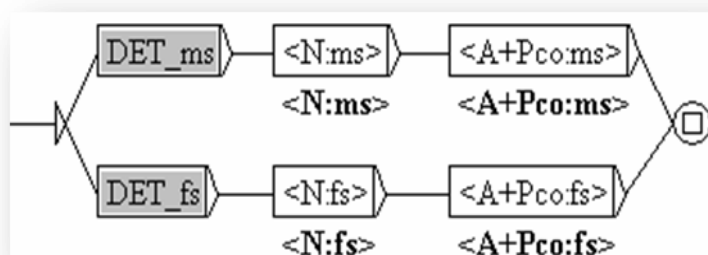


Figura 1 – Grafo usado para ajudar na resolução de ambiguidade.

Para ilustrar o desempenho da gramática representada na Fig. 1, a autora cita como exemplo o caso da sequência: *uma colher de açúcar amarela*, cuja ambiguidade está explícita no transdutor da Fig. 2

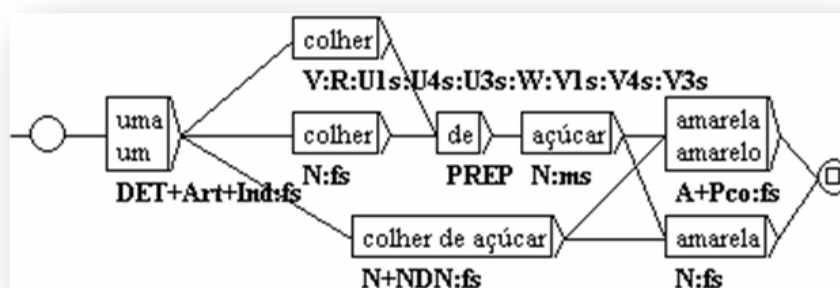


Figura 2 – Grafo representando uma ambiguidade.

etiquetar palavras ou expressões; (vi) solicitar concordâncias parametrizáveis de estruturas linguísticas variadas; (vii) obter dados estatísticos sobre o texto. O sistema permite ainda uma manutenção fácil (automática e semi-automática) dos dicionários e gramáticas (RANCHHOD, 2001, P. 8-9). No endereço: <http://www.ladl.jussieu.fr/INTEX/index.html> há uma apresentação geral do sistema e uma demonstração de algumas das suas potencialidades.

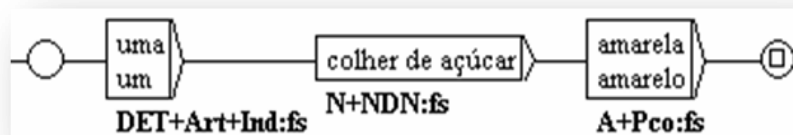


Figura 3 – Grafo representativo dos traços morfológicos de uma sequência.

A partir da análise da Fig. 2, a sequência fixa *colher de açúcar* é ambígua com uma estrutura sintática livre do tipo: *Nome Preposição Nome*. Contudo, a gramática utilizada obriga à verificação dos traços morfológicos das várias categorias gramaticais, e, por isso, analisa corretamente *colher de açúcar* como sequência fixa, a única análise que satisfaz as condições de concordância impostas pela gramática.

Ranchhod (2003, p. 216), reitera que as sequências formadas por sequências de palavras simples formam inúmeras sequências fixas; e elas constituem uma parte significativa do léxico de qualquer língua, por isso, são muito frequentes em qualquer tipo de texto, em particular nos textos de natureza técnica e científica. Contudo, várias sequências são ambíguas, já que também podem ser interpretadas como uma combinação livre de palavras. Por isso, a resolução dessa ambiguidade passa, necessariamente, pela elaboração de gramáticas específicas para que possam ser analisadas como sequências livres ou fixas.

Assim como Ranchhod, Laporte também acredita que a resolução automática de ambiguidades lexicais envolve análise e reconhecimento do contexto gramatical, a fim de verificar restrições locais. Trata-se, em geral, de restrições distribucionais, gramaticais e combinatórias sobre as sequências de palavras ou de etiquetas lexicais. Para Laporte, estas restrições, devidamente formalizadas, constituem os dados linguísticos do sistema, e podem ser designadas por gramáticas de resolução de ambiguidades.

Por isso, Laporte advoga que a construção de gramáticas de resolução de ambiguidades lexicais é uma atividade importante na medida em que contribui para um trabalho persistente de elaboração progressiva da gramática das

línguas, de formalização e aplicação à análise automática dos textos. Para tanto, conforme afirma Laporte (2001, p. 70), a resolução das ambiguidades lexicais pressupõe uma descrição de restrições gramaticais durante a construção do sistema e uma filtragem automática das análises conformes às restrições, durante a aplicação do sistema ao processamento de texto. Também acrescenta que a descrição e a automatização da filtragem não podem ser levadas à prática sem um formalismo de descrição que defina as convenções de formalização das restrições gramaticais e a interpretação da descrição. Laporte considera que a falta de formalização pode tornar a descrição aproximativa, incompleta, incoerente ou inaplicável; e com a ausência de um formalismo definido, é impossível definir uma noção de restrição gramatical correta, pois a interpretação das restrições será sempre aproximativa. Portanto, Laporte evidencia que o formalismo garante que uma correta acumulação de restrições seja correta; e a construção de uma gramática de resolução de ambiguidades lexicais por acumulação de restrições gramaticais torna-se mais simples e mais segura.

Laporte (2001, p. 80) conclui que uma gramática de resolução de ambiguidades lexicais, expressa com um formalismo de filtragem independente, possui também duas outras aplicações importantes, que são:

- Pode ser usada para detectar erros não lexicais: quando a gramática é correta e elimina todas as análises de uma frase, isso significa que a frase não tem nenhuma análise correta e que é, portanto, incorreta.
- Pode também ser usada para resolver as homofonias do tipo *paço/passo*, que perturbam o reconhecimento da fala por ausência de razões fonéticas para escolher a ortografia certa. Pois, se uma palavra for transcrita *paço/passo*, e se uma gramática de resolução de ambiguidades lexicais estabelecer, por observação do contexto, que a palavra é um verbo, a hipótese *paço* pode ser eliminada.

Laporte ressalta que o progresso da resolução da ambiguidade não poderá ser alcançado se não levar em consideração análises linguísticas mais elaboradas, e que o conteúdo informativo dos dados linguísticos é tão importante no

domínio da linguística computacional quanto é nas outras áreas da linguística. Só assim será possível a elaboração de boas gramáticas de resolução de ambiguidades.

Formalização das expressões fixas

Esta pesquisa apresenta um estudo sobre dois tipos de expressões fixas: a) aquelas de estrutura argumental verbal e b) aquelas de estrutura argumental substantival. Por isso, resulta em dois tipos de formalização. Nas expressões fixas de estrutura argumental verbal, a formalização é representada por tábuas, seguindo o modelo das tábuas do Léxico-gramática. Já na formalização das estruturas argumentais substantivais, segue-se o modelo das codificações das estruturas substantivais apresentado no trabalho de Smarsaro (2004).

7.1 Formalização das propriedades das expressões fixa verbais

Na construção das tábuas, que representam a formalização das expressões fixas verbais, foi de suma importância o trabalho de Vale (2001), pois assim como esta pesquisa, Vale também faz um estudo das expressões fixas. Mas, diferentemente de Vale, em que seu estudo direciona para uma proposta tipológica, com intuito de investigar a estrutura valencial das expressões fixas verbais, esta pesquisa não investiga a estrutura valencial do verbo das expressões fixas selecionadas no *corpus*. Ainda sim, “volta e meia” fez-se necessário recorrer a Vale (2001) para a realização do trabalho de descrição das expressões, para a construção das tábuas. Por isso, os modelos das tábuas nesta pesquisa se aproximam muito das tábuas criadas por Vale, que por sua vez, também se aproximou dos modelos das tábuas do Léxico-gramática.

Segundo Fernandes, Correia e Baptista (2004, p. 2), um dicionário eletrônico de frases fixas é composto de várias matrizes, uma para cada classe formal. Por isso, nessas matrizes, cada classe formal é uma sentença de expressões

fixas e as suas colunas contêm os elementos lexicais da sentença e suas propriedades sintáticas (distribucional e transformacional). Logo, o conjunto de matrizes constitui o Léxico-gramática das expressões fixas. E esse conjunto de matrizes, representado por tábuas, é de imensa importância para o PLN, especialmente para a constituição de um ambiente computacional.

A seguir, apresenta-se a tabela representativa das classes das expressões fixas formadas por nomes de *partes de corpo*, bem como as tábuas dessas classes.

7.1.1 Tábuas – alguns comentários

Após a descrição de um *corpus* com 351 expressões fixas, de estrutura argumental verbal (cf. cap. 5), foi possível criar nove tábuas diferentes. Cada tábua é representada por uma classe, e a estrutura dessa classe segue os mesmos códigos convencionais propostos inicialmente por Gross (1982), como na tabela abaixo:

Classe	Estrutura	Exemplo	Efetivo
PB-C1	$N_0 V C_{1pc}$	Kátia pegou barriga	35
PB-C1PN	$N_0 V C_{1pc} Prep N$	Eder abriu mão da herança	29
PB-CA1	$N_0 V Art C_{1pc}$	Eder afiou a língua	57
PB-CADN	$N_0 V Art (C de N)_{1pc}$	Eder calou a boca de Marcelo	54
PB-CA1PN	$N_0 V Art C_{1pc} Prep N$	Eder fechou a cara para Marcelo	64
PB-CP1	$N_0 V Prep C_{1pc}$	Eder fala pelos cotovelos	36
PB-CPN	$N_0 V Prep (C Prep N)_{1pc}$	Eder caiu nas unhas de Marcelo	35
PB-CP1PN	$N_0 V Pre C_{1pc} Prep N_1$	Eder caiu de boca no bolo	16
PB-CNP2	$N_0 V N_1 Prep C_{1pc}$	Eder apunhalou Marcelo pelos costas	25

Tabela 7 - Classificação das expressões fixas verbais

Cada tábua foi construída de acordo com as propriedades observadas durante a descrição de cada expressão fixa. Na construção dessas tábuas, fez-se necessário observar o tipo de complemento, a posição do elemento lexical, tanto livre como fixo, através das diversas transformações sintáticas.

Dessa forma, as expressões fixas verbais, de estrutura argumental VN, resultaram em duas classes distintas:

- a) classe PB-C1, com estrutura $N_0 V C_{1pc}$;
- b) classe PB-C1PN, com estrutura $N_0 V C_{1pc} Prep N$.

Abaixo, seguem-se breves comentários dessas classes:

Classe PB-C1, com estrutura $N_0 V C_{1pc}$

Nessa classe, tem-se uma sequência fixa, sem a necessidade de um complemento livre:

- (1) O problema **ganhou corpo**.

O sujeito da sentença pode ser humano:

- (1a) Eder **bateu cabeça** para fazer o teste de motorista.

Ou por um sujeito não humano:

- (1b) O acordo de ontem **deu pé**.

O substantivo **perna** da expressão fixa **ter perna** pode vir no singular:

- (1c) Eder não **teve perna** para ganhar a corrida.

Ou no plural:

- (1d) Eder não **teve pernas** para ganhar a corrida.

Também pode ser passivada:

- (1e) Eder **cortou gordura**.
- (1f) **Gordura foi cortada** por Eder.

Essa classe, que geralmente não necessita de completo livre, é representada na tábua abaixo:

Tábua PB-C1

N₀ V C_{1pc}

N ₀ =: Nhum N ₀ =: N-hum	Negação obrigatória	Verbo	Inserção lexical	Det. opcional	C _{1pc}	C _{1pc} =: Nsingular C _{1pc} =: Nplural	Forma passiva	Subst. de V por outro V	Varição em n° do C _{1pc}
+	-	<Bater>			cabeça	-	+	-	-
+	-	<Bater>			orelha	+	-	-	-
+	-	<Bater>		o	pé	+	-	-	-

Figura 4 – Fragmento da Tábua PB-C1, representando a estrutura N₀ V C_{1pc}

Classe PB-C1PN, com estrutura N₀ V C_{1pc} Prep N

Essa classe exige um complemento preposicional livre, que pode ser um sujeito humano ou um sujeito não humano, como abaixo:

(2) Eder **pediu bexiga** para Marcelo.

(2a) Eder **abriu mão** dos bens.

A expressão fixa **pedir bexiga** exige de um complemento livre humano, seguido da preposição **para**. Já a expressão fixa **abrir mão** exige um complemento livre não humano. Essa expressão também pode apresentar um complemento preposicionado livre humano, como em

(2b) Eder **abriu mão** de Kátia.

O complemento livre pode ser afetado pelo sujeito livre da frase:

(2c) Eder **botou chifre** em Kátia.

Ou o sujeito livre da frase pode ser elemento afetado:

(2d) Eder **comeu barriga** no sorteio da loteria.

Nesse caso, não há a possibilidade de se aceitar a forma passiva:

(2e) Eder **bateu boca** com Marcelo.

(2f) *A **Boca foi batida** por Eder e Marcelo.

Essa classe está representada através do fragmento da tábua abaixo:

Tábua PB-C1PN

$N_0 = \text{Nhum}$ $N_0 = \text{N-hum}$	$N_0 = \text{afetado}$	Negeção obrigatória	Verbo	Inserção lexical	C_{1pc}	$C_1 = \text{Nsingular}$ $C_1 = \text{Nplural}$	Preposição	Determinante	$N_{1pc} = \text{Nhum}$ $N_{1pc} = \text{N-hum}$	$N_1 = \text{afetado}$	Subst. de V por outro V	Forma passiva	Varição em nº do C_{1pc}
+	-	-	<Abrir>		mão	+	de	a / o	+	+		-	-
+	-	+	<Bater>	muita / bastante	boca	+	com		+	-		-	-
+	-	+	<Bater>	muita / bastante	perna	+	em / com	a / o	+	+		-	+
+	-	-	<Botar>	muito / bastante	chifre	+	em		+	-	<Colocar / Por>	+	-
+	-	+	<Comer>		barriga	+	em	a / o	-	+		-	-
+	-	+	<Dar>	muita / bastante	cabeçada	+	em	a / o	-	+		-	-

Figura 5 – Fragmento da Tábua PB-C1PN, representando a estrutura $N_0 V C_{1pc}$ Prep N

Para as expressões fixas verbais, de estrutura argumental VartN, verificou-se três classes distintas:

- classe PB-CA1, com estrutura $N_0 V \text{ Art } C_{1pc}$;
- classe PB-CADN, com estrutura $N_0 V \text{ Art } (C \text{ de } N)_{1pc}$;
- classe PB-CA1PN, de estrutura $N_0 V \text{ Art } C_{1pc} \text{ Prep } N$

Têm-se abaixo, comentários dessas três classes:

Classe PB-CA1, com estrutura N₀ V Art C_{1pc}

Essa classe, geralmente também não necessita de complemento livre após a expressão fixa:

(3) Eder **assentou a cabeça**.

O sujeito da frase é um sujeito humano:

(3a) Eder **abasteceu o estômago**.

O determinante obrigatório pode ser substituído por um pronome oblíquo:

(3b) Eder **abasteceu seu estômago**.

Em (3b), o artigo definido **o**, que é parte obrigatória na estrutura da expressão fixa **abastecer o estômago**, foi substituído pelo pronome oblíquo **seu** sem causar problema no sentido da expressão.

Também pode ser apassivada:

(3c) **O estômago foi abastecido** por Eder.

Essa classe está representada na tábua abaixo:

Tábua PB-CA1

N ₀ =:Nhum N ₀ =:N-hum	Negação obrigatória	Verbo	Inserção lexical	Det. obrigatório	Variação do determinante	Subst. do det. por pron. oblíquo	C _{1pc}	C _{1pc} =:Nsingular C _{1pc} =:Nplural	Forma passiva	Subst. de V por outro V	Variação em n° do C _{1pc}
+	-	<abastecer>	bastante / bem	o	+	estômago	+	-	+		-
+	-	<Abrir>		a	+	cabeça	+	-	-		-
+	-	<Abrir>		o	+	pulso	+	-	-		-
+	-	<Abrir>		os	-	bofes	-	+	+		-
+	-	<Amarrar>		a	-	testa	+	-	-		-

Figura 6 – Fragmento da Tábua PB-CA1, representando a estrutura N₀ V Art C_{1pc}

Classe PB-CADN, com estrutura N₀ V Art (C de N)_{1pc}

Nessa classe, o complemento livre após a expressão fixa pode ser representado por um complemento livre humano ou algo que represente coletividade humana, como povo, nação entre outros:

- (4) Eder **calou a boca** de Marcelo.
- (4a) Eder **calou a boca** (do povo + da nação).

O sujeito livre pode ser um sujeito humano ou um sujeito não humano:

- (4b) Eder **babou o ovo** de Marcelo.
- (4c) a derrota **custou os bofes** de Marcelo.

Em sua grande maioria, ocorre aceitabilidade do determinante possessivo:

- (4d) Eder **calou a boca** de Vander.
- (4e) Eder **calou a boca** de mim.
- (4f) Eder **calou a minha boca**.

O complemento livre pode ser afetado pelo sujeito livre da frase:

- (4g) Eder **cortou as asas** de Marcelo.

A expressão fixa só admite complemento livre humano:

- (4h) Eder **encheu a cabeça** de Marcelo.
- (4i) Eder **encheu a cabeça** do (*cachorro + *macaco + *gato).

A expressão fixa **encher a cabeça** não aceita nenhum tipo de complementos livre não humano, como em (4i).

Algumas aceitam a construção de passiva, como abaixo:

- (4j) Eder **confortou o coração** de Kátia.
- (4k) **O coração** de Kátia **foi confortado** por Eder.

Outras, como **babar o ovo**, não admitem esse tipo de construção:

(4l) Eder **babou o ovo** do chefe.

(4m) ***O ovo** do chefe **foi babado** por Eder.

Essa classe está representada na tábua abaixo:

Tábua PB-CADN

N ₀ =:Nhum N ₀ =:N-hum	N ₀ =:afetado	Negação obrigatória	Verbo	Inserção lexical	Det. obrigatório	Det. possessivo	C _{1pc}	C _{1pc} =:Nsingular C _{1pc} =:Nplural	Preposição	Apassivação	Subst. de V por outro V	N ₁ =:Nhum N ₁ =:N-hum	N ₁ =:afetado	Varição em nº do C _{1pc}
+	+	-	<Abrir>	- a	+	cabeça		+ -	de -			+ -	+	-
+	+	-	<Abrir>	+ os	+	olhos		- +	de -			+ -	+	-
+	-	-	<Amarrar>	- a	+	testa		+ -	de -			+ -	+	-
+	-	-	<Apalpar>	- as	-	costelas		- +	de +			+ -	+	-
+	-	-	<Assentar>	- o	-	cabelo		+ -	de -			+ -	+	-
+	-	-	<Babar>	+ o	+	ovo		+ -	de -			+ -	+	-

Figura 7 – Fragmento da Tábua PB-CADN, representando a estrutura N₀ V Art (C de N)_{1pc}

Classe PB-CA1PN, de estrutura N₀ V Art C_{1pc} Prep N

Essa classe é formada pelas expressões fixas que apresentam um complemento fixo direto e um complemento preposicional livre:

(5) Eder **atrelou os bigodes** com Kátia.

Essa classe é muito parecida com a classe PB-CADN, que também é formada por um complemento fixo direto e um complemento preposicional livre. O que diferencia uma da outra é que, na classe PB-CA1PN, ora o sujeito livre da frase é afetado:

(5a) Eder **curvou a espinha** para Marcelo.

Ora o complemento preposicional livre é afetado pelo sujeito livre da frase:

(5b) Eder **deitou as unhas** em Marcelo.

Ora ambos são afetados ao mesmo tempo:

(5c) Eder **emendou os bigodes** com Marcelo.

O sujeito livre pode ser tanto humano como não humano:

(5d) Eder **não arreda o pé** do quintal.

(5e) O cachorro **não arreda o pé** do quintal.

Ainda nessa classe, o determinante possessivo dificilmente é aceito:

(5f) Eder não **arredou o pé** de mim.

(5g) *Eder não **arredou o meu pé**.

(5h) Eder **atrelou os bigodes** comigo.

(5i) *Eder **atrelou os meus bigodes**.

Tem-se também um complemento preposicional livre humano:

(5j) Eder **fechou a cara** para Marcelo.

Ou não humano:

(5k) Eder **deu o sangue** no trabalho.

A expressão fixa só admite complemento livre humano:

(5l) Eder **encheu a cabeça** de Marcelo.

(5m) Eder **encheu a cabeça** do (*cachorro + *macaco + *gato).

Ou um complemento livre que ora pode ser humano, ora pode ser não humano:

(5n) Eder **abriu os olhos** com Marcelo

(5o) Eder **abriu os olhos** com a falência da empresa.

Essa classe está representada na tábua abaixo:

Tábua PB-CA1PN

N ₀ =:Nhum N ₀ =:N-hum	N ₀ =:afetado	Negação obrigatória	Verbo	Inserção lexical	Determinante obrigatório	Determinante possessivo	C _{1pc}	C _{1pc} =:Nsingular C _{1pc} =: Nplural	Preposição	Determinante	N ₁ =:Nhum N ₁ =:N-hum	Apassivação	Subst. de V por outro V	N ₁ =:afetado	Varição em n° do C _{1pc}
+	-	+	<Abaixar>	+	a	+	cabeça	+	-	para	+	-	-	-	-
+	-	+	<Abrir>	+	as	-	pernas	-	+	para	+	-	+	-	-
+	-	-	<Abrir>	-	o	-	bico	+	-	para / contra	+	-	-	+	-
+	-	+	<Abrir>	+	o	+	coração	+	-	para	+	-	-	-	-
+	-	+	<Abrir>	+	o	-	peito	+	-	para / em	+	-	-	-	-

Figura 8 – Fragmento da Tábua PB-CA1PN, representando a estrutura N₀ V Art C_{1pc} Prep N

Já as expressões fixas, de estrutura argumental V_{prep}N, são representadas por quatro classes distintas:

- a) classe PB-CP, com estrutura N₀ V Prep C_{1pc};
- b) classe PB-CPN, com estrutura N₀ V Prep (C Prep N)_{1pc};
- c) classe PB-CP1PN, com estrutura N₀ V Prep C_{1pc} Prep N;
- d) classe PB-CNP2, com estrutura N₀ V N₁ Prep C_{1pc}.

Abaixo, Têm-se também, breves comentários sobre essas quatro classes:

Classe PB-CP, com estrutura N₀ V Prep C_{1pc}

Nessa classe, as expressões fixas não exigem complementos livres determinativos:

- (6) o debate *acabou em sangue*.

O sujeito livre da frase por ser humano:

- (6a) Eder *sofreu na carne*.

Ou por um sujeito livre não humano:

(6b) o cavalo **deu no pé**.

Todos os exemplos (6, 6a e 6b) estão representados na tábua abaixo:

Tábua PB-CP1

$N_0 = N_{hum}$ $N_0 = N_{-hum}$	Negação obrigatória	Verbo	Inserção lexical	Prep. obrigatória	Determinante	Inserção lexical	C_{1pc}	$C_{1pc} = N_{singular}$ $C_{1pc} = N_{plural}$	Subst. de V por outro V	Varição em nº do C_{1pc}
- +	-	<Acabar>	novamente	em			sangue	+ -	<Terminar>	-
+ +	-	<Andar>		em	O		colo	+ -		-
+ -	-	<cair>		de			pé	+ -		-
+ -	-	<Dar>		à			espinha	+ -		-
+ -	-	<dar>		à			língua	+ -		-
+ -	-	<dar>		às			canelas	- +		-

Figura 9 – Fragmento da Tábua PB-CP1, representando a estrutura $N_0 V Prep C_{1pc}$

Classe PB-CPN, com estrutura $N_0 V Prep (C Prep N)_{1pc}$

Essa classe é formada por expressões fixas que pedem um complemento determinativo livre humano e um sujeito livre não humano:

(7) A carta **chegou pelas mãos** de Eder.

Ou um sujeito livre humano, que pode ser afetado pelo complemento livre:

(7a) Eder **caiu nas unhas** de Marcelo.

O complemento livre humano também pode ser afetado pelo sujeito livre na frase:

(7b) Eder **deu na cara** de Marcelo.

O complemento livre pode ser não humano:

(7c) Eder **está com cara** de cachorro Bulldog.

Há, ainda, a possibilidade de aceitação do pronome possessivo:

(7d) A carta **chegou pelas mãos** de mim = a carta **chegou pelas minhas mãos**.

(7e) Eder **caiu nas unhas** de mim = Eder **caiu nas minhas unhas**.

Essa classe é representada pela tábua abaixo:

Tábua PB-CPN

N ₀ =:Nhum N ₀ =:N-hum	N ₀ =:afetado	Negação obrigatória	Verbo	Inserção lexical	Prep. obrigatória	Determinante	Pron. possessivo	C _{1pc}	C _{1pc} =:Nsingular C _{1pc} =:Nplural	Preposição	N _i =:Nhum N _i =:N-hum	N _i =:afetado	Subst. do V por outro V	Varição em nº do C _{1pc}
+ -	+	-	<Atirar>		a	os	+	pés	- +	de	+ -	-		-
+ -	-	-	<Botar>		em	a	+	cabeça	+ -	de	+ -	+	<Por> / <Meter>	-
+ -	+	-	<Cair>		em	as	+	unhas	- +	de	+ -	-		-
+ +	-	+	<Chegar>		a	os	+	calcanhares	- +	de	+ -	+		-
- +	+	-	<Chegar>		a	os	+	ouvidos	- +	de	+ -	-		-
+ +	+	-	<Chegar>		por	as	+	mãos	- +	de	+ -	-		-

Figura 10 – Fragmento da Tábua PB-CPN, representando a estrutura N₀ V Prep (C Prep N)_{1pc}

Classe PB-CP1PN, com estrutura N₀ V Prep C_{1pc} Prep N

Nessa classe, a frase é formada por um sujeito livre à direita do complemento fixo e um complemento preposicional livre após o complemento fixo:

(8) Eder **caiu de boca** no bolo.

Assim como o sujeito livre, o complemento livre também pode ser humano ou não humano:

(8a) Eder **deu às costa** para Marcelo.

(8b) o cachorro **caiu de boca** no osso.

Diferentemente da classe **PB-CPN**, essa classe dificilmente aceitará um pronome possessivo:

(8c) Eder **deu às costa** comigo = *Eder **deu às minhas costa**.

Também não é possível apassivar:

(8d) Eder **deu às costa** para Marcelo = * **às minhas costa foi dada** a Marcelo por Eder.

(8e) Eder **deu de rosto** com Marcelo = *Marcelo **foi dado de rosto** por Eder.

Em algumas expressões o verbo pode ser perfeitamente substituído por outro verbo de igual valor semântico:

(8f) Eder (**entrou + mergulhou**) **de cabeça** no novo projeto.

Já outras não aceitam substituir o verbo por outro verbo:

(8g) Eder (**deu + *deparou + *encontrou**) **de cara** com seu pai.

Essa classe está representada na tábua abaixo:

Tábua PB-CP1PN

N ₀ =:N _{hum} N ₀ =:N- _{hum}	Negação obrigatória	Verbo	Prep. obrigatória	Determinante	Pron. possessivo	C _{1pc}	C _{1pc} =:N _{singular} C _{1pc} =:N _{plural}	Preposição	Determinante	N ₁ =:N _{hum} N ₁ =:N- _{hum}	Forma passiva	Subst. do V por outro V	Varição em n° do C _{1pc}
+ -	-	<Andar>	de		-	olho	+ -	em	a / o	+ +	-		-
+ +	-	<Cair>	de		-	boca	+ -	em	a / o	- +	-		-
+ -	-	<Cair	de		-	cabeça	+ -	em	a / o	- +	-		-
+ -	-	<Dar>	às		-	costas	- +	para		+ -	-		-
+ -	-	<Dar>	com	os	-	olhos	- +	em	a / o	+ +	-		-
+ -	-	<Dar>	com	os	-	ossos	- +	em	a / o	- +	-		-

Figura 11 – Fragmento da Tábua PB-CP1PN, representando a estrutura N₀ V Prep C_{1pc} Prep N₁

Classe PB-CNP2, com estrutura N₀ V N₁ Prep C_{1pc}

Nessa classe, o complemento livre é posicionado entre o verbo e o complemento preposicionado fixo:

(9) Eder **apunhalou** Marcelo **pelas costas**.

O sujeito livre da frase é sempre humano, mas o complemento livre pode ser humano ou não humano:

(9a) Eder **tem** Marcelo **pela barba**.

(9b) Eder **carrega** (Marcelo + o serviço da empresa) **nas costas**.

O complemento livre pode ser afetado pelo sujeito livre na frase:

(9c) Eder **passou** Kátia **na cara**.

Também pode se realizar com a construção passiva da expressão fixa:

(9d) Eder **apunhalou** Marcelo **pelas costas**.

(9e) Marcelo **foi apunhalado pelas costas** por Eder.

Essa classe está representada na tábua abaixo:

Tábua PB-CNP2

N ₀ ≠: N _{hum} N ₀ =: N _{-hum}	N ₀ ≠: N _{-hum} N ₀ =: N _{-hum}	N ₀ ≠: afetado	Negação obrigatória	Verbo	N ₁	N ₁ ≠: N _{hum} N ₁ =: N _{-hum}	N ₁ ≠: N _{-hum} N ₁ =: N _{-hum}	N ₁ ≠: afetado	Prep. obrigatória	Determinante	C _{1pc}	C _{1pc} ≠: N _{singular} C _{1pc} =: N _{plural}	Forma passiva	Subst. do V por outro V	Varição em n° do C _{1pc}
+	-	-	-	<Apunhalar>	N	+	-	+	por	as	costas	-	+		-
+	-	+	-	<Botar>	N	-	+	-	em	a	cabeça	+	-		-
+	-	+	-	<Carregar>	N	+	+	-	em	as	costas	-	+	<Levar>	-
+	-	-	-	<Deixar>	N	+	-	+	em	a	mão	+	+		-
+	-	+	-	<Destacar>	N	-	+	-	em	a	pele	+	-	<Brilhar>	-
+	-	-	-	<Fazer>	N	-	+	+	em	as	coxas	-	+		-

Figura 12 – Fragmento da Tábua PB-CNP2, representando a estrutura N₀ V N₁ Prep C_{1pc}

7.2 Codificação das descrições das propriedades das expressões fixas substantivais de estruturas argumentais Adj e NdeN

A codificação é uma representação por meio de símbolos (cf p. 9-12) que se relacionam com as propriedades estruturais, tais como morfológica, sintática, semântica e pragmática, uma vez que esta leva em consideração o consenso e o uso entre falantes nativos. Essas estruturas são examinadas por meio de critérios formais que permitem a codificação das propriedades para a formalização.

A formalização é de suma importância para o PLN, pois é a partir dela que se tem resultado final de um estudo analisado e descrito sobre determinadas estruturas linguísticas para implementação em um ambiente computacional. Sem a formalização das descrições fica quase impossível utilizar as descrições em sistemas informáticos.

A importância da formalização se dá porque as máquinas não dispõem de habilidades avaliativas para julgar, como os humanos, em uma língua natural, se uma dada sequência é um grupo nominal livre ou uma expressão fixa. Essas informações devem ser dadas à máquina através de códigos¹⁹ representativos das descrições linguísticas estudadas

A codificação das expressões fixas, de estrutura argumental substantival, se baseia nos resultados da aplicação dos critérios das estruturas argumentais substantivais NAdj e NdeN.

7.2.1 Propriedades das estruturas argumentais NAdj

Na estrutura argumental substantival NAdj, a codificação se dá a partir das seguintes propriedades: (N0 ter NAdj) e (N0 ser NAdj).

"N0" representa o sujeito livre que antecede a expressão fixa no enunciado.

"N" representa uma palavra com categoria de substantivo da expressão fixa. Indica também que a expressão em foco é uma expressão de

¹⁹ Segundo Youssef e Fernandez (1988, p. 52), códigos são sistemas de regras e convenções de acordo com os quais é possível formar, transmitir, receber e processar sinais que representam dados.

estrutura argumental substantival. Por isso, todas as expressões codificadas, que forem de estrutura argumental substantival, serão iniciadas com a letra N.

"Adj" representa o adjetivo da expressão fixa.

"+pc" indica que a expressão fixa é formada por um substantivo com nome de uma parte qualquer do corpo, como cabeça, perna, rosto, pé, unha e tantas outras.

Por meio de códigos, pode-se estabelecer uma convenção para determinadas descrições linguísticas, como abaixo:

- 1 Quando uma sequência fixa representar a propriedade (N0 ter NAdj), anota-se "+ter".
- 2 Quando uma sequência fixa representar a propriedade (N0 ser NAdj), anota-se "+ser".

Por exemplo, *língua afiada* (N+pc+ter), *barba-azul* (N+pc+ser). Portanto, não há necessidade de uma frase explicando que a expressão fixa *barba-azul* é formada pelo substantivo *barba*, e que esse substantivo é uma parte do corpo, o código "+pc" é uma representação dessa informação.

- 3 Quando uma expressão fixa apresentar Predicatividade do adjetivo em relação ao sujeito do enunciado, anota-se "+Pred". Quando NÃO apresentar Predicatividade, "-Pred".

Ex.:

- (1) Ede é um *barba-azul*.

Barba-azul: (N+pc+ser-Pred)

- (2) Eder tem uma *língua afiada*.

Língua afiada (N+pc+ter+Pred)

- 4 Quando o NAdj ser N, tem-se, assim, "+1". Quando o NAdj NÃO ser N, tem-se "-1".

Ex.:

(3) Eder é um **barba-azul**.

Barba-azul é uma barba? Não: (N+pc+ser-Pred-1)

(4) Eder tem um **cabelo ruim**.

Cabelo ruim é um cabelo? Sim: (N+pc+ser+Pred+1)

5 Quando o NAdj ser Adj, anota-se "+2". Quando o NAdj NÃO ser Adj, anota-se "-2".

Ex.:

(5) Eder é um **barba-azul**.

Em **barba-azul** é uma barba? Não, e é azul? Não: (N+pc+ser-Pred-1-2)

(6) Eder tem um **cabelo ruim**.

Cabelo ruim é um cabelo? Sim, e é ruim? Não: (N+pc+ser+Pred+1-2)

6 Quando for possível Coordenar o Adj com outros adjetivos, anota-se, "+AdjC". Quando NÃO for possível Coordenar o Adj com outros adjetivos, anota-se, "-AdjC".

(7) Eder tem um **cabelo ruim**.

(7a) Eder tem um **cabelo (ruim e + curto + longo)**.

Cabelo ruim: (N+pc+ter+Pred+1-2+AdjC)

(8) Eder é um **barba-azul**.

(8a) Eder é um **barba (azul e + *grande + *vermelho)**.

Barba-azul: (N+pc+ser-Pred-1-2-AdjC)

- 7 Quando for possível Elidir o Adj da sequência fixa NAdj, anota-se "+EAdj". Quando NÃO for possível Elidir o Adj da sequência fixa NAdj, anota-se "-EAdj".

Ex.:

(9) Eder é um **barba-azul**.

(9a) *Eder é um **barba**.

Barba-azul: (N+pc+ser-Pred-1-2-AdjC-EAdj)

(10) Eder tem uma **língua afiada**.

(10a) *Eder tem uma **língua**.

Língua afiada (N+pc+ter+Pred-1-2+AdjC+EAdj)

- 8 Quando em uma sequência fixa NAdj, tanto o N como o Adj permanecerem invariáveis em número, tem-se "-n". Quando em uma sequência fixa NAdj, tanto o N como o Adj FOREM variáveis em número, tem-se "+n".

Ex.:

(11) Eder tem uma **língua afiada**.

(11a) *?²⁰Eder e Marcelo têm umas **línguas afiadas**.

Língua afiada (N+pc+ter+Pred-1-2+AdjC-EAdj-n)

(12) Eder é um **barba-azul**.

(12a) Eder e Marcelo são uns **barbas-azuis**.

Barba-azul: (N+pc+ser-Pred-1-2-AdjC-EAdj+n)

- 9 Quando o N da estrutura Nadj for feminino singular, anota-se :fs, se for plural, :fp. Da mesma forma, quando o N for masculino singular, tem-se :ms, se for plural, :mp.

²⁰ O sinal interrogativo (?) indica uma dúvida se esta frase está ou não inaceitável.

Ex.:

Barba-azul: (N+pc+ser-Pred-1-2-AdjC-EAdj+n:ms)

Cabelo ruim: (N+pc+ter+Pred+1-2+AdjC-EAdj+n:ms)

Em um enunciado como abaixo:

Eder é um **Barba-azul**

A entrada final da expressão fixa **barba-azul** após a codificação será da seguinte forma:

Barba-azul = N+pc+ser-Pred-1-2-AdjC-EAdj+n:ms

Assim, o "N" indica que se trata de uma expressão fixa substantival, o código "+pc" aponta para uma expressão formada por um substantivo com nome de uma parte do corpo, o "+ser" demonstra que a expressão fixa representa o sujeito livre em um enunciado, o "-Pred" indica que não há predicatividade do adjetivo. Os códigos "-1" e "-2" indicam que não se trata de um substantivo barba nem de um adjetivo azul. O código "-AdjC" indica que não é possível coordenar o Adj da expressão com outros adjetivos. O código "-EAdj" indica que não é possível elidir o adjetivo da expressão. "+n" indica que é possível variar a expressão em número. ":ms" indica que a expressão funciona como um substantivo masculino com relação ao sujeito no enunciado e este sujeito está no singular.

Através da descrição, foi possível identificar e classificar as propriedades das expressões fixas em dois grupos:

Grupo 1 – São expressões fixas constituídas por um substantivo e um adjetivo pós-nominal, cujo sentido da expressão está relacionado apenas com uma parte do corpo do sujeito livre no enunciado.

Ex.:

Eder tem uma **boca nervosa**.

Eder tem um **coração duro**.

Eder tem as **costas-largas**.

Eder tem um **olho clínico**.

Todas representadas pela propriedade N0 ter Nadj

Grupo 2 – São as expressões fixas, também constituídas por um substantivo e um adjetivo pós-nominal, mas o sentido da expressão representa o sujeito livre do enunciado.

Ex.:

Eder é um **cabo eleitoral**.

Eder é um **canela-verde**.

Eder é um **dedo-duro**.

Eder é um **mão leve**.

Todas representadas pela propriedade N0 ser Nadj

Abaixo, tem-se uma amostra das expressões fixas, de estrutura substantival argumental Nadj, após a formalização de suas descrições linguística:

1. língua solta	N+pc+ter+Pred-1-2+AdjC-EAdj-n:fs
2. barba-azul	N+pc+ser-Pred-1-2-AdjC-EAdj+n:ms
3. boca-mole	N+pc+ter+Pred-1-2-AdjC-EAdj+n:fs
4. cara suja	N+pc+ter+Pred-1-2-AdjC-EAdj-n:fs
5. braço direito	N+pc+ser-Pred-1-2+AdjC-EAdj+n:ms
6. dente canino	N+pc+ter+Pred+1-2-AdjC-EAdj+n:ms
7. braço forte	N+pc+ser-Pred-1-2+AdjC-EAdj+n:ms
8. bunda-suja	N+pc+ser-Pred-1-2+AdjC-EAdj+n:ms
9. bunda-mole	N+pc+ser-Pred-1-2+AdjC-EAdj+n:ms

7.2.2 Estrutura argumental NdeN

Seguindo a pesquisa de Smarsaro (2004), as expressões fixas, de estrutura argumental NprepN, compostas pela preposição **de**, serão codificadas a partir das seguintes propriedades:

N1deN2 ser N1

N1deN2 ser N2

N1deN2 não ser nem N1 nem N2

Assim como nas estruturas argumentais substantivais Nadj, em NdeN o “N” estará sempre no início de cada expressão codificada, para que, dessa forma, seja possível identificar que tal expressão pertence a uma estrutura substantival. Ou seja, o N representa uma categoria de substantivo.

O código “+pc” indica que as sequências fixas são formadas por substantivos com nomes de partes do corpo. Ex.: **cintura de pilão** (N+pc).

A partir dessas observações, pode-se estabelecer uma convenção para descrições linguísticas das propriedades da estrutura argumental NdeN, como abaixo:

- 1 Quando a propriedade das estruturas argumentais NdeN for (N1deN2 ser N1), anota-se “+1”. Se NÃO for, anota-se “-1”.
- 2 Quando a propriedade das estruturas argumentais NdeN for (N1deN2 ser N2), anota-se “+2”. Se NÃO for, anota-se “-2”.
- 3 Quando a propriedade das estruturas argumentais NdeN for (N1deN2 não ser nem N1 nem N2), anota-se “-1-2”.

Assim, para as expressões fixas abaixo, tem-se os seguintes códigos iniciais:

Relógio de pulso é um **relógio**.

Relógio de pulso = N+pc+1-2

Pé de mato é um **mato**.

Pé de mato = N+pc-1+2

Pé-de-meia não é um **pé** nem uma **meia**.

Pé-de-meia = N+pc-1-2

- 4 Quando for possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas, tem-se “+N1Pron”. Quando NÃO for possível Pronominalizar o “N1” da segunda sequência em duas sequências fixas, tem-se “-N1Pron”.

Ex.:

Eder contratou um **cabeça de área** e um **cabeça de chave** para seu novo time de futebol.

Eder contratou um **cabeça de área** e um **de chave** para seu novo time de futebol.

Cabeça de área e de chave = N+pc-1-2+N1Pron

Eder contratou um **cabeça de área** e um **cabeça de vento** para seu novo time de futebol.

Eder contratou um **cabeça de área** e um ***de vento** para seu novo time de futebol.

Cabeça de área e *de vento = N+pc-1-2-N1Pron

- 5 Quando for possível Elidir O “N1” da expressão fixa, anota-se “+N1E”. Quando NÃO for possível Elidir o “N1”, anota-se “-N1E”.

Ex.:

Kátia tem uma **cintura de vespa**. = *Kátia tem uma **vespa**.

Cintura de vespa = N+pc+1-2-N1Pron-N1E

Eder comprou uma **cabeça de cebola**. = Eder comprou uma **cebola**.

Cabeça de cebola = N+pc-1+2+N1Pron+N1E

- 6 Quando for possível Elidir o "N2", anota-se "+N2E". Quando NÃO for possível Elidir o "N2", anota-se "-N2E".

Ex.:

Eder ganhou um **relógio de pulso** = Eder ganhou um **relógio**.

Relógio de pulso = N+pc+1-2+N1Pron-N1E+N2E

Eder comprou uma **cabeça de cebola** = *Eder comprou uma **cabeça**

Cabeça de cebola = N+pc-1+2+N1Pron+N1E-N2E

- 7 Quando for possível variar em número a sequência fixa "NdeN", tem-se "+n". Quando NÃO for possível variar em número a sequência fixa "NdeN", tem-se "-n".

Ex.:

Eder tem **mão de fada**.

Eles têm **mãos de fada**.

Mãos de fada = N+pc-1-2-N1Pron-N1E-N2E+n

- 8 Quando a sequência fixa "NdeN" apresentar na forma masculina singular, em um enunciado, tem-se ":ms". Mas se apresentar no plural, tem-se ":mp".

Eder ganhou um **relógio de pulso**.

Relógio de pulso = N+pc+1-2+N1Pron-N1E+N2E+n:ms

Eder ganhou dois **relógios de pulso**.

Relógios de pulso = N+pc+1-2+N1Pron-N1E+N2E+n:mp

- 9 Quando a sequência fixa "NdeN" apresentar na forma feminina singular, em um enunciado, tem-se ":fs". Mas se for no plural, tem-se ":fp".

Ex.:

Eder comprou uma **cabeça de cebola**.

Cabeça de cebola = N+pc-1+2+N1Pron+N1E-N2E+n:fs

Eder comprou duas **cabeças de cebola**.

Cabeças de cebola = N+pc-1+2+N1Pron+N1E-N2E+n:fp

Segue abaixo uma amostra representativa de codificação das expressões fixas substantivais, com estrutura argumental NdeN:

- | | |
|---------------------|------------------------------|
| 1. Barba de bode | N+pc+1-2+N1Pron-N1E-N2E+n:fs |
| 2. dente de alho | N+pc-1+2-N1Pron+N1E-N2E+n:ms |
| 3. dente de siso | N+pc+1-2+N1Pron+N1E+N2E+n:ms |
| 4. língua de víbora | N+pc-1-2-N1Pron-N1E-N2E+n:fs |
| 5. mão de tinta | N+pc-1+2+N1Pron+N1E-N2E+n:fs |
| 6. olho de boi | N+pc-1-2+N1Pron-N1E-N2E+n:ms |
| 7. orelha de abano | N+pc+1-2-N1Pron-N1E-N2E+n:fs |
| 8. pé de vento | N+pc-1+2-N1Pron+N1E-N2E+n:ms |
| 9. rabo de foguete | N+pc-1-2-N1Pron-N1E-N2E+n:ms |

Conclusões e contribuições

Segundo Vieira, Strube de Lima (2001, p. 38), a área de linguística computacional envolve um grande conjunto de atividades voltadas para o objetivo de tornar possível a comunicação com as máquinas utilizando as habilidades naturais de comunicação humana. Por isso, o objetivo desta pesquisa foi estabelecer a descrição linguística de um *corpus* com 558 expressões fixas do português brasileiro, sendo 351 de estruturas argumentais verbais e 207 de estruturas substantivais, para uma possível inserção delas em um ambiente computacional.

Todas as expressões selecionadas têm como palavras-chave nomes de qualquer *parte do corpo*, como **boca**, **mão** e **língua**. Essa escolha mostrou-se acertada porque as expressões fixas formadas por esses nomes são bastante recorrentes no uso cotidiano, tanto na modalidade falada quanto na modalidade escrita.

No decorrer da pesquisa, ficou demonstrado que essas expressões são de grande importância para a ampliação do léxico em qualquer língua natural. Demonstrou também que o léxico de uma língua não é formado por apenas palavras simples, mas também de sequências fixas constituídas por duas ou mais palavras, como por exemplo, as expressões fixas em questão.

Por as expressões fixas serem bastante numerosas nas línguas naturais, esta pesquisa se respalda em Gross (1982), uma vez que esse autor defende a idéia de que as expressões fixas são bastante relevantes, por isso, devem ser descritas e formalizadas para serem inseridas em um léxico computacional de programas que lidem com PLN.

A partir de critérios formais, foi possível analisar as sequências. Dessa forma, segundo Smarsaro (2004, p. 88), permite de maneira sustentável, estabelecer critérios de identificação para as sequências livres e para as sequências fixas com a mesma estrutura. Para tanto, Smarsaro acrescenta que as sequências fixas não podem ser estudadas se comparadas umas com as outras. É preciso que essas sequências sejam inseridas em enunciados para constatar o comportamento sintático-semântico das palavras que as compõem.

Muitos dos exemplos utilizados foram construídos manualmente para verificar, com maior clareza, o comportamento sintático-semântico dessas sequências fixas dentro dos enunciados. Portanto, a Teoria do Léxico-gramática se mostrou como um grande suporte nesta pesquisa por ser considerada como uma teoria experimental, que tem como propósito designar uma metodologia e uma prática efetiva da descrição manual sintático-semântica do léxico numa determinada língua. Assim, o Léxico-gramática não trabalha com itens lexicais isolados, mas em contexto a partir de enunciados, com sujeitos, verbos e complementos necessários. Isso ajuda a observar o comportamento sintático-semântico das propriedades estruturais dessas expressões fixas em situação de uso pelos falantes nativos.

Durante a aplicação de um conjunto de critérios formais, tais como:

- coordenação do verbo com outro verbo;
- inserção lexical;
- perda da predicatividade do adjetivo;
- ruptura paradigmática;
- elisão de um dos elementos;
- coordenação de grupos nominais;
- e variação em número.

Verificou-se que as expressões fixas, tanto de estrutura argumental verbal como de substantival, apesar de apresentarem como um único item lexical, de homogeneidade singular, quando inseridas em enunciados, por meio das

distribuições sintáticas dos seus componentes, podem sofrer variações diferentes. Isso levou a uma descrição detalhada em cada expressão para verificar suas propriedades sintáticas.

Como resultado, foi possível observar que, por haver restrições morfossintático-semânticas nas estruturas internas dessas expressões fixas, não se podia incluí-las em uma única categoria. Pois cada caso é um caso. O que necessitou de extrema atenção na hora da descrição para se conseguir uma descrição satisfatória. Assim, a descrição se mostra como uma saída confiável e com menos riscos de se ter problemas na hora da formalização.

Duas observações foram determinantes para o resultado final da descrição e da formalização das expressões fixas:

- 1 Nas expressões fixas, de estrutura argumental verbal, levou-se em consideração a PRESENÇA de um verbo sempre atrelado à sua estrutura, como em ***perder a cabeça, crescer o olho, falar pelos cotovelos, pegar no pé, abrir mão, bater perna.***
- 2 Nas expressões fixas, de estrutura argumental substantival, há total AUSÊNCIA de um verbo atrelado à estrutura, como em ***sangue-azul, braço direito, cabeça quadrada, língua de sogra, cabeça de bagre, boca-de-fumo.***

A partir dessas observações, criou-se dois tipos distintos de formalizações, um para as expressões fixas, de estrutura argumental verbal (cf. p. 156, do capítulo 7.1) e um para as expressões fixas, de estrutura argumental substantival (cf. p. 167, do capítulo 7.2).

Na estrutura argumental verbal, após a descrição, foi possível criar nove classes diferentes. Cada classe é representada por uma tábua, e a estrutura de cada classe segue os mesmos códigos convencionais propostos inicialmente por Gross (1982), dentro da Teoria do Léxico-gramática, como na tabela abaixo:

Classe	Estrutura	Nº efetivo da classe
PB-C1	N ₀ V C _{1pc}	35
PB-C1PN	N ₀ V C _{1pc} Prep N	29
PB-CA1	N ₀ V Art C _{1pc}	57
PB-CADN	N ₀ V Art (C de N) _{1pc}	54
PB-CA1PN	N ₀ V Art C _{1pc} Prep N	64
PB-CP1	N ₀ V Prep C _{1pc}	36
PB-CPN	N ₀ V Prep (C Prep N) _{1pc}	35
PB-CP1PN	N ₀ V Prep C _{1pc} Prep N ₁	16
PB-CNP2	N ₀ V N ₁ Prep C _{1pc}	25

Tabela 8- Representação das classes verbais

Assim, foi possível fazer a formalização das expressões verbais, em forma de tabuas, construídas de acordo com as propriedades observadas durante a descrição de cada expressão fixa. Também se fez necessário observar o tipo de complemento, a posição do elemento lexical, tanto livre como fixo, através das diversas transformações sintáticas.

Na estrutura argumental substantival, a formalização também ocorreu por meio de critérios formais, mas seguiu uma codificação sem a necessidade de construção de tábuas, como ocorreu na estrutura argumental verbal. Para as expressões substantivais que possuem a estrutura argumental NAdj, após os testes com base nas propriedades sintáticas, foram codificadas, levando-se em consideração as seguintes formas: (N0 ter NAdj) e (N0 ser NAdj). Já as expressões substantivais, de estrutura argumental NdeN, foram codificadas a partir das seguintes formas: (N1deN2 ser N1), (N1deN2 ser N2) e (N1deN2 não ser nem N1 nem N2).

A codificação das descrições das expressões fixas verbais e substantivais foi baseada em um conjunto de regras e critérios para defini-las como um único item lexical. Dessa maneira, foi possível dar a essas expressões fixas uma

representação formal, com intuito de serem inseridas em um ambiente computacional, em forma de dicionário eletrônico, para ser usado por sistemas de PLN.

Vale ressaltar que um dicionário eletrônico feito com entradas de expressões fixas é de imensa importância para o PLN, pois além de tornar mais abrangente o léxico computacional, "pode ser utilizável na engenharia da linguagem, em aplicações tais como a busca de informações, a correção gramatical e a sumarização automática (SMARSARO, 2004, p. 90)". Sua aplicação também é direcionada tanto para análise e geração automática de textos quanto para a indexação e a geração de sumários e, em especial, para a tradução automática²¹.

No português brasileiro, a área computacional, apesar de já existir uma crescente e promissora pesquisa em PLN, com a finalidade de descrever os fenômenos linguísticos em uso, percebe-se que ainda é uma área bastante tímida e desafiadora. Há muito por avançar em estudos descritivos. Por isso, os resultados da descrição e da formalização das expressões fixas, com estruturas verbais e substantivais, propostos nesta pesquisa, se mostram como uma grande contribuição para o crescimento e desenvolvimento dos recursos linguísticos na linguística brasileira para programas computacionais. Esses recursos podem ser utilizados tanto por pesquisadores atuais como por novos pesquisadores que lidam com o PLN.

Esta pesquisa também é de grande relevância para possíveis soluções satisfatórias de problemas linguísticos, como a ambiguidade, típica em qualquer ferramenta computacional. Assim, pode-se evitar comprometimento de sentido das informações, eliminando possíveis dúvidas sobre a capacidade dos programas especializados em processamento das línguas naturais.

Espera-se que a partir deste estudo com as expressões fixas, abra-se caminho para futuros estudos com outros tipos de estruturas argumentais, como as estruturas argumentais adverbiais (**a olho nu**) ou adjetivais (**baixo-ventre**),

²¹ Na tradução automática, a descrição e formalização evitam que uma expressão fixa seja convertida apenas como um fragmento sintático, sem sentido, dentro de determinado texto quando traduzido de uma língua para outra.

para que se amplie cada vez mais trabalhos descritos, formalizados e codificados das expressões fixas do português brasileiro, com nomes de *partes do corpo*. Para Conteratto e Chrishman (2005, p. 2196), um dos fatores que implica diretamente na eficiência de um sistema computacional é de se ter uma descrição linguística suficiente informativa e organizada. Assim, quanto maior for a cobertura por meio da descrição linguística das expressões fixas, maior será o seu acesso a sistemas computacionais, para processar uma língua natural com maior eficiência.

Portanto, há ainda muito por fazer nesta área, e isso não é uma tarefa simples, pois a linguística computacional é um campo bastante ambicioso, que requer dedicação e tempo por aqueles que se propõem formalizar e codificar o léxico de uma língua natural, mas o resultado tem se mostrado sempre satisfatório.

Referências

ALMEIDA, Napoleão M. **Gramática metódica da língua portuguesa**. 29. ed. São Paulo: Saraiva, 1963.

ALVES, Maria Alves. **Neologismo: criação lexical**. 2. ed. São Paulo: Ática, 2004.

BAPTISTA, Jorge. **Estabelecimento e formalização de classes de nomes compostos**. Dissertação (Mestrado em Letras) - Programa de Pós-Graduação em Educação, Faculdade de Letras da Universidade de Lisboa, Lisboa, 2004.

BAPTISTA, Jorge; CORREIA, A.; FERNANDES, Graça. Léxico-gramática das frases fixas do português europeu. Breve presentación. In: **Cadernos de Fraseoloxía Galega 7**, Santiago de Compostela: Xunta de Galicia, 2005, p. 41-53.

BASÍLIO, Margarida. **Formação e classes de palavras no português do Brasil**. 1. ed. São Paulo: Contexto, 2004.

_____. Questões clássicas e recentes na delimitação de unidades lexicais. In: Margarida Basílio (Org.), **Revista palavra, nº 5**. Rio de Janeiro: Departamento de Letras da PUC, 1999, p. 9-18.

_____. **Teoria lexical**. 7. ed. São Paulo: Ática, 2004.

BECHARA, Evanildo. **Moderna gramática portuguesa**. 32 ed. Rio de Janeiro: Companhia Editora Nacional, 1988.

BIDERMAN, Maria Tereza Camargo. **Teoria linguística: teoria lexical e linguística computacional**. 2. ed. São Paulo: Martins Fontes, 2001.

_____. Conceito linguístico de palavra. In: Margarida Basílio (Org.), **Revista palavra, nº 5**. Rio de Janeiro: Departamento de Letras da PUC, 1999, p. 81-97.

BLOOMFIELD, Leonard. A set of postulates for the science of language. *Language* 2, 1926. In: M. Joos (Org), **Readings in linguistics I**. The University of Chicago Press, 1957, 153-64.

BOONS, Jean-Paul; GUILLET, Alain; LECLÈRE, Christian. **La structure des phrases simples en français: constructions intransitives**. Genebra: Droz, 1976.

BORBA, Francisco da Silva. **Pequeno vocabulário de linguística moderna**. Vol. 31, São Paulo: Nacional e USP, 1971.

CÂMARA JR, Joaquim Mattoso. **Dicionário de linguística e gramática**. 23. ed. Petrópolis: Vozes, 2002.

CARONE, Flavia de Barros. **Morfossintaxe**. São Paulo: Ática, 1986.

CARVALHO, Paula. **Gramáticas de resolução de ambiguidades resultantes da homografia de nomes e adjetivos**. Dissertação (Mestrado em Letras) - Programa de Pós-Graduação em Educação, Faculdade de Letras da Universidade de Lisboa, Lisboa, 2001.

CONTERATTO, G. B. Hinrichs; CHISHMAN, R. L. de Oliveira. Semântica: uma tentativa de tornar as estruturas com predicado secundário acessíveis aos sistemas de PLN. In: **III Workshop em Tecnologia da Informação e da Linguagem Humana**, São Leopoldo, 2005, p. 2188-2197.

COWIE, A. P. Stable and creative aspects of the use of vocabulary. In: R. Carter; M. McCarthy (eds.), **Vocabulary and language teaching**. Harlow: Longman, 1988, p. 126-137.

D'AGOSTINO, Emilio; ANNIBALE, Elia; VIETRI, Simonetta. **Lexicon-Grammar, Electronic Dictionaries and Local Grammars of Italian**. *Lexique, Syntaxe et Lexique-grammaire. Papers in honour of Maurice Gross*, Coll.

Linguisticae Investigationes Supplementa, 24. Amsterdam/Philadelphia: Benjamins, 2004, p. 125-136.

DANLOS, Laurence. **La morphosyntaxe des expressions figées**. Langages 63, Paris : Larousse, 198, p. 53-74.

DUARTE, Sérgio Nogueira. **Língua viva III: uma análise simples e bem-humorada da linguagem do brasileiro**. Rio de Janeiro: Rocco, 2000.

ELSON, Benjamin; PICKETT, Velma. **Introdução à morfologia e à sintaxe**. Tradução de Aryon D. Rodrigues *et al*; coordenação e revisão de Marta Coelho. 2. ed. Petrópolis: Vozes, 1978.

FERNANDES, Graça. **Léxico-Gramática das Frases Fixas do Português Europeu. Construções Intransitivas**. 2007. Dissertação (Mestrado em Letras), Faro, Universidade do Algarve/FCHS, 2007.

FERNANDES, Graça; BAPTISTA, Jorge. **Reconhecimento automático de expressões idiomáticas em corpus – algumas experiências**. Proceedings of TIL, 2007 – V Workshop em tecnologia da informação e linguagem humana, Rio de Janeiro, Brasil, June, 2007, p. 1725-1728.

FERREIRA, Aurélio Buarque de Holanda. **Enriqueça o seu vocabulário**. 4. ed. Rio de Janeiro: Nova Fronteira, 1984.

FRANCIS, N.; KUCERA, H. **Frequency Analysis of English Usage**. Lexicon and Grammar, New York: Houghton Mifflin. 1982.

GARRÃO, Milena de Uzeda. **Um estudo de expressões cristalizadas e sua inclusão em um tradutor automático bilíngue (português/inglês): o caso de “bater + SN”**. 2001. Dissertação de Mestrado.

GROSS, Gaston. **Degré de figement des noms composés**. Langages, 1988, n. 90, p. 56-70.

GROSS, Maurice. **Methods and Tactics in the Construction of a Lexicon-Grammar**. Linguistics in the Morning Calm 2, Selected Papers from SICOL, Seoul: Hanshin Publishing Company 1986, p.177-197.

_____. **Les bases empiriques de la notion de prédicat sémantique.** Langages, Paris: Larousse, 1981, n. 63, p.7-52.

_____. **Lexicon-Grammar. The Representation of Compound Words.** In Proceedings of Coling, 1986, p. 1- 6.

_____. **Grammaire transformationnelle du français: 2 - Syntaxe du nom,** 1977. Paris: Cantilène, 1986.

_____. Apresentação. In: BOONS, Jean-Paul; GUILLET, Alain; LECLÈRE, Christian. **La structure des phrases simples en français: constructions intransitives.** Genebra: Droz, 1976, p.7-52.

_____. **Constructing Lexicon-grammars.** Computational Approaches to the Lexicon, Atkins and Zampolli (Eds.), Oxford University Press, 1994, p. 213-263.

_____. État du lexique-grammaire du français et perspectives d'extension. In: Sylvain Aurox et al (Eds). **History of the Language Sciences, Tome 3,** Berlin-New york: Walter de Gruyter, 2006, p. 2122-2129.

_____. **Grammaire transformationnelle du français 1 : syntaxe Du verbe.** Paris: Cantilène, 1986a.

_____. **Grammaire transformationnelle du français 2 : syntaxe Du nom.** Paris: Cantilène, 1986b.

_____. **Grammaire transformationnelle du français 3: syntaxe de l'adverbe.** Paris: ASSTRIL, 1986c.

_____. **Les bases empiriques de la notion de prédicat sémantique.** Langages, n. 63, 1981, p. 7-52.

_____. **Les expressions figées. Une description des expressions françaises et ses conséquences théoriques.** Programme de Recherches Coordonnées Informatique Linguistique, Rapport technique n. 8 Paris: LADL, Université Paris 7, CERIL, 1989. (mimeogr.).

_____. **Les limites de la phrase figée.** Langages, n. 90, 1988, p. 7-22.

_____. **Lexicon-Grammar**. En BROWN, K. e MILLER, J. (Eds.): *Concise Encyclopaedia of Syntactic Theories*. Pergamon, Cambridge, 1996, p. 244-258.

_____. **Lexicon-Grammar: the representation of compound words**. In **Proceedings of the 11th International Conference on Computational Linguistics, COLING'86**, Bonn, West Germany, 1986, p. 1-6.

_____. **Méthodes en syntaxe**. Le régime des constructions complétives. Paris: Hermann, 1975.

_____. **Sur les déterminants dans les expressions figées**. *Languages*, n. 79, 1985, p. 89-117.

_____. **The L exicon-grammar of a language: application to French**. *Encyclopedia of language and linguistics*, London, Pergamon Press, 1994.

_____. **Une classification des phrases 'figg es' em fran ais**. *Revue Qu ebecoise de linguistique*, Vol. 11, n. 2, 1982, p. 151-185.

HARRIS, Z. S. **Mathematical Structures of Language**. New York: Wiley-Interscience, 1968.

_____. **Notes du Cours de Syntaxe**. Maurice Gross (Trad.). Paris:  dition du Seuil, 1976.

_____. **Papers on Syntax**. Henry Hiz (Ed.). Dordrecht: D. Reidel Publishing Company, 1981.

_____. **The elementary transformations**. In Harris, 1981, p. 211-235.

HOUAISS, Antonio. **Dicion rio digital Houaiss da l ngua portuguesa**. Rio de Janeiro: Objetiva, [2009]. 1 CD-ROM.

_____. **Dicion rio Houaiss da l ngua portuguesa**. Rio de Janeiro: Objetiva, 2003.

IGN CIO, Sebasti o Expedito. **An lise sint tica em tr s dimens es: uma proposta pedag gica**. 2. ed. S o Paulo: Ribeir o gr fica, 2003.

JACKENDOFF, R. **The architecture of the language faculty**. Cambridge,

Massachusetts: MIT Press, 1997.

KYRIACOPOULOU, Tita. **Analyse automatique des textes écrits: le cas du grec moderne.** Thessaloniki: University Press, 2004.

LACAVALLA, Claudia. **Lexique-grammaire des proverbes en Quand/Quando: Comparaison français-italien et représentation par grammaires locales.** Thèse de doctorat. Università degli Studi di Bari. 2008.

LAMIROY, Béatrice. **In memoriam Maurice Gross.** Travaux de Linguistique, 2003, p. 145-158.

LAPORTE, Éric. Resolução de ambiguidade. In Elisabete Ranchhod (Ed), **Tratamento das línguas por computador. Uma introdução à linguística computacional e suas aplicações.** Lisboa: Caminho, 2001, p. 49-89.

_____. Lexicons and grammars for language processing: industrial or handcrafted products?. In REZENDE, M. Letícia, SILVA, Bento Carlos Dias da, BARBOSA, Juliana Bertucci (Org.), **Léxico e gramática: dos sentidos à construção da significação.** Séries Trilhas Linguísticas nº 16, Faculdade de Ciências e Letras, Araraquara, SP: Cultura acadêmica Editora, 2009, p. 51-82.

_____. Exemplos atestados e exemplos construídos na prática do Léxico-gramática. Tradução do francês: Francisco Antônio P. Léllis. In **Revista (Com)textos Linguísticos 2.** 2008, p. 26-51.

LAPORTE, Eric; NAKAMURA, Takuya; VOYATZI, Stavroula. A French Corpus Annotated for Multiword Expressions with dverbial Function. Author manuscript, published. In "**Language Resources and Evaluation Conference**" (LREC). Linguistic Annotation Workshop, Marrakech: Morocco (2008).

LEFFA, Vilson J. **A resolução da ambiguidade lexical sem apoio do conhecimento de mundo.** Intercâmbio. São Paulo, PUC: v. 6, Parte I, 1996, p. 869-889.

LEWIS, Michael. Pedagogical implications of the lexical approach. In: J. Coady, T. Huckin (Eds.), **Acquisition of second language vocabulary: A rationale for pedagogy.** Cambridge: Cambridge University Press, 1997a, p. 255-270.

_____. **Implementing the Lexical Approach.** Language Teaching Publications, Hove, England, 1997b.

MAKKAI, A. **Idiomacity as a language universal.** In: Greenberg, J.H. Universals of human language. Vol. 3, Stanford: Stanford University Press, 1978, p. 401-448.

MARCUS, M.; SANTORINI, B.; MARCINKIEWICZ, M.A. **Building a large annotated corpus of English.** The Penn Treebank. *Computational Linguistics* 19, 1993, p. 313-330.

MARQUES, Cristina P. C.; MATTOS, M. Isabel L. de; TAYLLE, Yves de La. **Computador e ensino: uma aplicação à língua portuguesa.** São Paulo: Ática, 1986.

MARTINS, Ronaldo Teixeira (NILC / Universidade São Francisco). **Dos problemas da ambiguidade semântica em um modelo de tradução automática baseado em interlíngua: Apontamentos do projeto UNL-Brasil.** 1999. Disponível em <http://orbita.starmedia.com/~lcorpus/simp99/ronaldo.pdf>. Acessado: 02-03-10

MUNIZ, Marcelo Caetano Martins. **A construção de recursos lingüístico-computacionais para o português do Brasil: o projeto Unitex-PB.** 2004. Dissertação (Mestrado em Ciências Matemáticas e de Computação). Instituto de Ciências Matemáticas de São Carlos, USP, 2004.

NAM, Jee-sun. **Lexique grammairre des adjectifs coréens et analyse syntaxique automatique.** Langages 126, paris: Larousse, 1997.

NANCY IDE, Jean Véronis. **Word disambiguation: the state of the art.** *Computational Linguistics*, 24 (1), 1998, p. 1-40.

NUNES, M. Volpe; SILVA, BENTO C. Dias da; MONTILHA, Gisele; RINO, Lúcia H. Machado; SPECIA, Lucia; OLIVEIRA Jr., Osvaldo Novais de; MARTINS, Ronaldo Teixeira; PARDO, Thiago A. Salgueiro. **Introdução ao processamento das línguas naturais e algumas aplicações.** São Carlos: ICMC – USP, 1999.

OTHERO, Gabriel de Ávila. Linguística computacional: uma breve introdução. **Introdução a estudos de fonologia do português brasileiro**. Leda Bisol (Org.). 4. ed. Porto Alegre: EDIPUCS, 2006.

PAUMIER, Sébastien. Unitex Manual, 2006. <http://univ-mlv.fr/~unitex>.

P.V.O.L.P. **Pequeno vocabulário ortográfico da língua portuguesa**. Academia Brasileira de Letras, 1943.

PERINI, Mario A. **Gramática descritiva do português**. 2. ed. São Paulo: Ática, 1996.

QUENTAL, Violeta; DIAS, Maria Carmelita. **Tendências do processamento computacional do português**. Palavras – Departamento de Letras da PUC – Rio. 2004, p. 7-12.

RABUSKE, Renato Antônio. **Inteligência artificial**. Florianópolis: Ed. da UFSC, 1995.

RANCHHOD, Elisabete Marques. **Expressões ‘fixas’. Questões conceituais e terminológicas**. Materiais bibliográficos elaborados no âmbito do mestrado em Linguística “O lugar das expressões fixas na gramática do português, FLUL (não publicados), 1993.

_____. **Frozen adverbs. Comparative forms como C in portuguese**. *Linguisticae investigationes*, XV: 1, Amsterdam/Philadelphia: JOHN Benjamins, 1991, p. 141-170.

_____. O uso de Dicionários e de autômatos finitos na representação lexical das línguas naturais. In: Elisabete M. Ranchhod (Org.), **Tratamento das línguas por computador. Uma introdução à lingüística computacional e suas aplicações**, Lisboa: Caminho, 2001, p. 13-47.

_____. Dicionários eletrônicos e análise lexical automática. In: Paumira Marrafa, Maria Antônia Mota (Orgs), **Linguística computacional. Investigação fundamental e aplicações**. Lisboa: Colibri, 1998, p. 207-220.

_____. **Sintaxe dos Predicados Nominais com Estar**. *Linguística* 1-2,

Lisboa: Instituto Nacional de Investigação Científica, 1990.

_____. O Lugar das Expressões 'Fixas' na Gramática do Português. In: CASTRO, Ivo; I. Duarte (Orgs.), **Razões e Emoção**. Miscelânea de estudos oferecida a Maria Helena Mira Mateus, Lisboa: Imprensa Nacional - Casa da Moeda, 2003, p. 239-254.

RANCHHOD, Elisabete; CARVALHO, P; MOTA, Cristina; BARREIRO, A. Portuguese Large-scale Language Resources for NLP Applications. In: **Proceedings of the 4th LREC**, Lisbon, 2004, p. 1755-1758.

RANCHHOD, Elisabete; CARVALHO, Paula. Unidades lexicais complexas: problemas de análise e etiquetagem. In: **Actas Del VIII Simposio Internacional de Comunicación Social**. Santiago de Cuba, 2003, p. 212-221.

ROCHA LIMA, Carlos Henrique da. **Gramática normativa da língua portuguesa**. 34. ed. Rio de Janeiro: José Olympio, 1997.

SANDMANN, Antônio José. **Competência lexical**: produtividade, restrições e bloqueios. Curitiba: Ed. da UFRR, 1991.

_____. **Formação Palavras no Português Contemporâneo**. Curitiba: Ícone Editora Ltda, 1989.

SEHELLART, Jean. Reconnaissance automatique des entrées du lexiquegrammaire des phrases figées. En: LAMIROY, B. (Ed.), **Le Lexique-Grammaire**. Travaux de Linguistique 37, Ducolot, Bruxelles, 1998, p. 109-125.

SILVA, Bento Carlos da. O estudo linguístico-computacional na linguagem. **Introdução a estudos de fonologia do português brasileiro**. Leda Bisol (Org.). 4. ed. Porto Alegre: EDIPUCS, 2005, p. 103-133.

SILVA, Lúcio Buzon da. **Ambiguidades da língua portuguesa: recorte classificatório para elaboração de um modelo ontológico**. Dissertação (Mestrado em Ciências da Computação e Documentação) – Universidade de Brasília, 2006.

SMARSARO, Aucione. **Descrição e formalização de palavras compostas do português do Brasil para elaboração de um dicionário eletrônico**. 2004.

Tese (Doutorado em Letras) – Programa de Pós-graduação em Educação, Pontífica Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

_____. O processamento automático de Linguagem Natural: uma introdução à linguística computacional. In: SILVA, Alacir de Araújo; LINS, Maria da Penha Pereira (Orgs), **Recortes linguísticos**. Inst. de Ensino Saberes, Vitória: Lisboa, 2000.

TAGNIN, Stella Ortweiler. **Expressões idiomáticas e convencionais**. São Paulo: Ática, 1989.

TRASK, R. L. **Dicionário de linguagem e lingüística**. Tradução de Rodolfo Ilari; revisão técnica de Ingedore Villaça Koch, Thaís Cristófaró Silva. São Paulo: Contexto, 2004.

TÔRRES, Artur de Almeida. **Moderna gramática expositiva da língua portuguesa**. 16. ed. São Paulo / Rio de Janeiro, 1964, 214.

VALE, Oto Araújo. **Expressões cristalizadas do português do Brasil: uma proposta de tipologia**. 2001. Tese (Doutorado em Letras) – Universidade Estadual Julio Mesquita Filho - Campus de Araraquara, São Paulo, 2001.

_____. Transparência e opacidade de expressões cristalizadas. In: Hirata-Vale, Flávia B. M. (Orgs), **Anais do IV Seminário Nacional de Literatura e Crítica e do II Seminário Nacional de Lingüística e Língua Portuguesa**, Goiânia, Gráfica e Editora Vieira, 2001, p. 240-246.

VIEIRA, Renata; LIMA, Vera Lúcia Strube de. Linguística computacional: princípios e aplicações. In: Ana Teresa Martins; DÍbio Leandro Borges (Orgs.), **As tecnologias da informação e a questão social: anais**. 1 ed. Fortaleza: SBC, 2001, v.3.

YOUSSEF, A. Nicolau; FERNANDEZ, V. Paz. **Informática e sociedade**. 2. ed. São Paulo: Ática, 1988.