

DESCRIÇÃO DE EXPRESSÕES FIXAS COM *NOMES DE PARTES DO CORPO* DO PORTUGUÊS BRASILEIRO PARA FINS COMPUTACIONAIS.

Vangevaldo Cardoso dos Santos*

Resumo: Este trabalho enfoca um estudo descritivo e formalizado de expressões fixas do português brasileiro, especialmente aquelas constituídas de *nomes de partes do corpo*, com **estrutura argumental verbal** (*abrir mão, dobrar a língua, bater boca*), com intuito de inseri-las em um ambiente computacional para processamento automático da linguagem natural, com vistas à correção de falhas, que possam prejudicar a qualidade de um texto, especialmente quando processado automaticamente.

Palavras-chave: Linguística Computacional, Processamento de Linguagem Natural, expressões fixas, descrição e formalização.

Abstract: This work focuses on a descriptive and formalized study of fixed expressions in Brazilian Portuguese, especially those consisting by names of parts of the body, with a **verbal argument structure** (*abrir mão, dobrar a língua, bater boca*) with a view to incorporating them into a computing environment for automatic processing of natural language, in order to fix flaws that could undermine the quality of a text, especially when automatically processed.

Keywords: Computational Linguistics, Natural Language Processing, fixed expressions, description and formalization.

INTRODUÇÃO

Na Linguística Computacional, o homem procura fazer uma interação entre a linguística e a informática¹. 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. Com a 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. Mas devido aos 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 transposta para uma língua artificial, capaz de ser executável em um ambiente computacional.

* Mestre em Estudos Linguísticos – UFES / ES.

¹ Segundo Youssef e Fernandez (1988, 25-26), a palavra informática surgiu na década de 1960 como uma nova ciência, e esse termo foi criado pela junção dos vocábulos informação e automática.

APRESENTAÇÃO DO TEMA

Este trabalho apresenta um estudo descritivo e formalizado das expressões fixas com *nomes de parte do corpo*, com **estrutura argumental verbal**, para a elaboração de dicionário eletrônico.

Segundo Makkai (1978), as expressões fixas são numerosas em qualquer língua, e conforme Gross (1988), a fixidez é uma característica universal das línguas naturais. 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 a expressão Eder **bateu braço**, pois essa não adquire outro sentido além do literal.

Ainda, uma expressão fixa, como **esfriar a cabeça**, 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).

De acordo com Vale (2001, p. 3-4), as expressões são um problema de peso quando olhadas com mais vagar. No aspecto sintático, as expressões cristalizadas² são frases aparentemente normais, semelhantes às frases "comuns". Já 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.

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 fixas. 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.

Portanto, para inserir as expressões fixas *com nomes de partes do corpo* em um programa de PLN, faz-se necessário realizar um estudo ancorado em critérios linguísticos, passando por um quadro descritivo, representativo e formalizado. Com isso, tem-se a possibilidade de viabilizar a elaboração de um dicionário eletrônico, com o intuito de melhorar a qualidade de programas computacionais em PLN.

Assim, para o PLN a descrição de expressão fixa deve seguir uma:

² O autor caracteriza as expressões fixas como expressões cristalizadas.

- a) Análise;
- b) Descrição;
- c) Formalização.

A análise das estruturas lexicais das 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.

Selecionou-se um *corpus* com 175 expressões fixas a partir de dicionários de língua (impressos e digitais), de jornal impresso (A Tribuna) e por informantes da comunidade linguística local.

As expressões fixas selecionadas foram analisadas através de uma descrição mais pormenorizada das suas propriedades, obedecendo alguns critérios sintático-semânticos para identificá-las como expressões fixas. A partir do estudo descritivo, ancorado em critérios formais, fez-se, então, a formalização. Essa formalização viabiliza a possibilidade de implementações e aplicações computacionais. Após essa formalização é possível inserir as expressões fixas em um software como o Unitex³ para o tratamento de textos em língua natural.

Este trabalho está ancorado na teoria do Léxico-gramática (GROSS, 1975), por ser uma teoria experimental, que tem como propósito designar uma metodologia e uma prática efetiva de descrição manual sintático-semântica do léxico de uma língua. Assim, o Léxico-gramática exige que o *corpus* analisado e descrito seja formalizado/codificado.

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 línguas.

O Léxico-gramática 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

³ "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, 2006, p. 2). Diferente dos outros dicionários usuais, o dicionário proposto no Unitex se define como um programa de processamento de texto com dados linguísticos.

lexicais com as propriedades morfossintático-semânticas, e os resultados obtidos constituem uma base de informações sintático-semânticas (GROSS, 1975).

A DESCRIÇÃO

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. Portanto, deve-se analisar morfologicamente 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, observando se os significados das palavras que compõem o item lexical relacionam-se entre si ou se apontam para outro significado.

Sem as descrições coerentes dos dados linguísticos, o processamento automático fica bastante comprometido e sujeito a falhas. Por isso, a qualidade dos softwares (PLN) depende da qualidade das descrições linguísticas de um léxico.

OS CRITÉRIOS DE IDENTIFICAÇÃO DAS EXPRESSÕES FIXAS VERBAIS

A análise das propriedades internas das estruturas linguísticas das expressões fixas, ancorada no Léxico-gramática, seguirá os seguintes critérios:

- a) critérios morfológicos:
- b) critérios sintáticos:
- c) critérios semânticos.

Seguindo o estudo de Smarsaro (2004), a aplicação dos critérios também 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).

APLICAÇÃO DOS CRITÉRIOS

Critério morfológico:

- a) **Variação em número de N**

Segundo as gramáticas tradicionais, os substantivos podem sofrer variação em número, gênero e grau. Entretanto, as expressões fixas, por serem fixas, costumam ter restrições quanto às regras gramaticais. Tais restrições são indícios de fixidez. Para comprovar se há realmente fixidez, o elemento substantival (N) da estrutura argumental verbal VartN da expressão será submetido à flexão em número, como no exemplo abaixo:

- (1) Eder e Marcelo *torceram o nariz* para Mônica no trabalho.
- (1a) *⁴Eder e Marcelo *torceram os narizes* para Mônica no trabalho.
- (1b) Eder *torceu o nariz* para Mônica no trabalho.
- (2) Eder *queimou a língua* quando acusou Marcelo injustamente.
- (2a) *Eder e Kátia *queimaram as línguas* quando acusaram Marcelo injustamente.
- (2b) Eder e Kátia *queimaram a língua* quando acusaram Marcelo injustamente.

Pode-se perceber que, contrariando algumas normas gramaticais, tanto em (1a) quanto em (2a), os substantivos *nariz* e *língua* permaneceram invariáveis, apesar dos sujeitos das frases estarem no plural.

A grande maioria das expressões fixas, de estrutura argumental VartN, o substantivo (N) sempre será apresentado no singular, como ocorreu em (1a) e (2a), independentemente do sujeito no enunciado estar ou não no singular.

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

- (3) Eder *deu as caras* por aqui, depois foi embora.
- (3a) *Eder *deu a cara* por aqui, depois foi embora.
- (3b) Eder e Marcelo *deram as caras* por aqui, depois foi embora.

Critérios sintáticos:

a) Distribuição sintática dos itens lexicais.

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, como abaixo:

- (4) Eder *esquentou a cabeça* durante a avaliação num concurso.
- (4a) **a cabeça* de Eder esquentou muito e ferveu.
- (4b) **a cabeça* de Eder foi esquentada em fogo baixo.
- (4c) *Eder queimou a mão quando a encostou na *cabeça quente*.

⁴ O código (*) representa uma sequência inaceitável em consonância ao sentido da expressão fixa em relação às restrições impostas por essa expressão.

O verbo *esquentar*, como parte da expressão, não se enquadra como uma ação praticada pelo sujeito na frase, mas se refere a um estado e que esse sujeito se encontra, referindo-se a uma **preocupação** ou até mesmo uma **aflição** por causa do concurso. A distribuição sintática dos itens acima comprova a restrição do item lexical *esquentar a cabeça*, caracterizando-a como uma expressão fixa. Assim, a partir da distribuição sintática, pode-se 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.

b) Negação obrigatória

Apesar de não ser uma constância, algumas expressões fixas aparecem, obrigatoriamente, precedidas de uma forma negativa, como abaixo:

(5) Todos se foram, 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:

(5a) Todos se foram, mas Eder (*não + nem + sequer + nunca + jamais*) *arredou o pé* da festa.

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

(5b) O dia amanheceu, mas *ninguém arredou o pé* da festa.

A negação também pode ser a preposição *sem*:

(5c) Todos continuaram *sem arredar o pé* da festa.

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

(5d) Por que Eder ia *arredar o pé* daqui?

(5e) Eder vai ter que *arredar o pé* daqui.

(5f) Quando Eder *arredar o pé* daqui...

c) Inserção de um elemento lexical

Algumas 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:

(6) Eder *salvou o pescoço* de Marcelo.

- (6a) Eder *salvou* (**bastante* + **logo* + **muito* + **de imediato*) o pescoço de Marcelo.

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

- (6b) Até Eder *deu as caras* na festa.
(6c) Até Eder *deu (logo) as caras* na festa.

Portanto, dependendo do item lexical, essa inserção é totalmente possível sem causar nenhuma perda para o significado da expressão fixa, como em (6c) e abaixo:

- (6d) Eder *livrou a cara* para não ser preso pela polícia por fraude.
(6e) Eder *livrou a (própria + sua) cara* para não ser preso pela polícia por fraude.

Em (6e), 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*.

d) Determinante obrigatório

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

- **Determinante indefinido:**

- (7) 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* pelo determinante *a*, tem-se, então, a quebra do seu sentido, e, por conseguinte, a perda de sua fixidez:

- (7a) A irmã de Eder mal casou e já *carrega (uma + *a) 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:

- (8) Será que Eder *conseguirá uma boca* na empresa do amigo?
(8a) Será que Eder *conseguirá (uma + aquela + outra + alguma) boca* na empresa do amigo?

Em (8a), 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.

- **Determinante definido:**

Em algumas expressões fixas não é possível substituir o determinante definido por certos determinantes, como abaixo:

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

(9a) Eder fez Marcelo *dobrar (a + *uma + *esta + *aquela + *a maioria das + *ambas + *alguma + *diversas) língua(s)* diante de todo o mundo.

A expressão fixa *dobrar a língua* tende aceitar certos determinantes possessivos como *minha, sua e nossa* sem a perda de sentido:

(9b) Eder fez Marcelo *dobrar (a minha + a sua + a nossa) língua* diante de todo o mundo.

Essa possibilidade só acontece, como em (9b), porque os determinantes possessivos não substituem o determinante definido feminino *a*, mas apenas são acrescentados à expressão. Assim, o sentido figurado da expressão fixa em (9) e (9b), que significa fazer alguém se reconsiderar, ou até mesmo falar com respeito, é mantido.

Critérios semânticos:

a) Ruptura paradigmática

Em uma sequência fixa, de estrutura argumental VartN, a possibilidade de se fazer um paradigma distribucional com N é praticamente nula:

(10) Eder *passou a perna* em Marcelo e ficou com todo o lucro da empresa.

(10) Eder *passou a (perna + *mão + *unha)* em Marcelo e ficou com todo o lucro da empresa.

A tentativa de substituir paradigmaticamente o substantivo (N) *perna* por outros na expressão fixa *passar a perna* desfaz o sentido da expressão, ocasionando uma ruptura paradigmática.

Ainda sim, há certos casos em que é possível se fazer um paradigma do N por outros sem a perda de sentido, não havendo, portanto, ruptura paradigmática:

(11) Eder *torceu o nariz* para Marcelo.

(11a) Eder *torceu o (nariz + rosto)* para Marcelo.

(12) Eder *mostrou as unhas* para Marcelo.

(12a) Eder *mostrou as (unhas + garras)* para Marcelo.

Em (11a), a substituição do substantivo *nariz* por *rosto* na expressão fixa *torcer o nariz* não ocasiona perda de sentido da expressão. Isso acontece também com a expressão fixa *mostrar as unhas*, como em (12a).

b) Substituição do verbo por outro com sentido próximo.

Em determinadas expressões fixas, a fixidez impede que se faça uma substituição do verbo por outro verbo sem desfazer o sentido da expressão, como abaixo:

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

(13a) Eder (*cortou + *amputou + *deceitou*) *as asas* do funcionário que só chegava atrasado.

(14) Eder *livrou a cara* do amigo de uma encrenca.

(14a) Eder (*livrou + *libertou + *tirou + *soltou*) *a cara* do amigo de uma encrenca.

Tanto em *cortar as asas* como em *livrar a cara* não aceitam substituir seus respectivos verbos nas frases acima. Isso acontece porque em (13a), 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 (14a), tem-se a perda de seu sentido, que é sair de uma situação embaraçosa. A impossibilidade de substituição dos verbos, em (13a) e (14a), confirmam que essas sequências são realmente expressões fixas.

No entanto, 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:

(15) Eder *amarrrou a cara* para Marcelo.

(15^a) Eder (*amarrrou + fechou*) *a cara* para Marcelo.

Em (15a), 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.

Algumas expressões fixas que admitem a troca do verbo por outro sinónimo:

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

(16a) Eder (*fechou + tapou*) *os ouvidos* para as fofocas do trabalho.

A troca do verbo *fechar* pelo verbo sinónimo *tapar*, em (16a), não prejudica o sentido metafórico da expressão fixa *fechar os ouvidos*.

A FORMALIZAÇÃO

Após a descrição das expressões fixas, de estrutura argumental VartN, tornou-se possível criar três classes representativas dessa estrutura, como na tabela abaixo:

Classe	Estrutura	Exemplo	Efetivo
PB-CA1	N ₀ V Art C _{1pc}	Eder <i>afiou a língua</i>	57
PB-CADN	N ₀ V Art (C de N) _{1pc}	Eder <i>calou a boca</i> de Marcelo	54
PB-CA1PN	N ₀ V Art C _{1pc} Prep N	Eder <i>fechou a cara</i> para Marcelo	64

Para cada classe tem-se uma tábua. As tábuas foram construídas de acordo com as propriedades observadas durante a descrição de cada expressão fixa. Na construção dessas tábuas, fez-se necessário também observar o tipo de complemento, a posição do elemento lexical, tanto livre quanto fixo, através das diversas transformações sintáticas.

As estruturas dessas tábuas seguem os mesmos códigos convencionais da teoria do Léxico-gramática. Todos propostos inicialmente por Gross (1982). Tem-se abaixo o fragmento de uma das tábuas do Léxico-gramática prontas:

Tábua da classe PB-CA1, com estrutura N₀ V Art C_{1pc};

N ₀ =: Nhum N ₀ =: N-hum	Negação obrigatória	Verbo	Inserção lexical	Det. obrigatório	Varição do determinante	Subst. do det. por pron. oblíquo	C _{1pc}	C _{1pc} =: Nsingular C _{1pc} =: Nplural	Forma passiva	Subst. de V por outro V	Varição em nº do C _{1pc}
+	-	<abastecer>	bastante / bem	o		+	estômago	+	-	+	-
+	-	<Abrir>		a		+	cabeça	+	-	-	-
+	-	<Abrir>		o		+	pulso	+	-	-	-
+	-	<Abrir>		os		-	bofes	-	+	+	-
+	-	<Amarrar>		a		-	testa	+	-	-	-

Fragmento da Tábua PB-CA1, representando a estrutura N₀ V Art C_{1pc}

Nessa tábua, as propriedades observadas nas expressões fixas são de maneira simples e objetiva, que podem ser da seguinte forma:

- a) nas linhas imaginárias das tábuas tem-se as expressões;

- b) nas colunas tem-se suas propriedades;
- c) nas colunas à esquerda encontram-se os sujeitos das frases, que podem ser humanos, representados pelo código $N_0=Nhum$, ou não humanos, representados pelo código $N_0=N-hum$;
- d) ainda à esquerda tem-se uma coluna que indica se a expressão apresenta obrigatoriedade de negação ou não.
- e) o código $< >$ indica que a palavra, representada pelo verbo é possível de flexão.
- f) a quarta coluna indica as expressões fixas que admitem uma inserção lexical.
- g) a quinta coluna diz respeito a obrigatoriedade e o tipo de determinante.
- h) a sexta coluna indica se há possibilidade de variação desse determinante obrigatório.
- i) já a sétima coluna indica se a expressão fixa admite substituir o determinante por um pronome possessivo.
- j) na oitava coluna, o código C_{1pc} indica um complemento fixo direto da expressão, com nome de parte do corpo.
- k) a nona coluna indica se o C_{1pc} vem obrigatoriamente no singular ou plural.
- l) a décima coluna indica se há possibilidade de passivação da expressão.
- m) a penúltima coluna à direita indica se é possível trocar o verbo pertencente a expressão por outro verbo de igual valor semântico.
- n) a última coluna à direita indica se é possível ou não variar em número o N (substantivo) da expressão fixa.
- o) o sinal (“+”) representa a ocorrência de determinada propriedade nas tábuas. Já o sinal (“-”) indica a não ocorrência de determinada propriedade nas tábuas.

CONSIDERAÇÕES FINAIS

Neste trabalho, foi apresentado uma descrição e uma formalização das expressões fixas de estrutura argumental verbal VartN, com *nomes de partes do corpo*, para processamento de linguagem natural, com objetivo de inseri-las em um ambiente computacional, considerando que os recursos linguísticos utilizados foram construídos manualmente, o que possibilitou uma garantia de maior fidelidade às informações sintático-semânticas.

Há, ainda, muitas estruturas argumentais que merecem uma investigação mais detalhada, visto que a criatividade do falante é um fator que interfere sobremaneira e continuamente no uso das expressões fixas. Portanto, cabe nesse contexto considerar outras formas de análise, levando-se em conta o tempo e o espaço em que as expressões fixas são analisadas. Nesse sentido, registra-se que há uma inesgotável fonte de pesquisa no tratamento e descrição das expressões fixas.

A descrição ainda pode ser mais detalhada, mas a formalização apresentada já possibilita sua implementação em dicionário eletrônico por meio de um software, viabilizando o processamento automático da linguagem natural do *corpus* analisado.

REFERÊNCIAS

BASILIO, Margarida. **Teoria lexical**. 7. ed. São Paulo: Ática, 2004.

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, 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.

_____. **Méthodes en syntaxe**. Le régime des constructions complétives. Paris: Hermann, 1975.

_____. **Une classification des phrases ‘figges’ em français**. Revue Québécoise de linguistique, Vol. 11, n. 2, 1982, p. 151-185.

LAPORTE, Éric. 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.

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.

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>.

SMARSARO, Aucione Das Dores. **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, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

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.

YOUSSEF, A. Nicolau; FERNANDEZ, V. Paz. **Informática e sociedade**. 2. ed. São Paulo: Ática, 1988.