

## COMPOSTOS TOPONÍMICOS - UMA ABORDAGEM PARA O PROCESSAMENTO AUTOMÁTICO DO PORTUGUÊS DO BRASIL

Rosana de Vilhena Lima\*

**Resumo:** Os estudos linguísticos com vistas ao processamento automático das línguas naturais ganhou impulso diante da necessidade de fornecer à máquina uma descrição exaustiva e formalizada dos fenômenos linguísticos a fim de que essas informações estejam disponíveis para a utilização por programas de geração automática de textos (orais e escritos), tradução de textos, recuperação de informação entre outros. Para que esses programas funcionem de maneira eficiente faz-se necessária a inserção de dados acerca do maior número de estruturas linguísticas possível, descritas detalhadamente. Apresenta-se neste artigo uma abordagem acerca dos nomes próprios de lugar (compostos), os compostos toponímicos, para o processamento automático do português do Brasil.

**Palavras-chave:** Compostos toponímicos. Processamento automático do português do Brasil. Descrição linguística. Formalização.

**Abstract:** The language studies with a view to automatic processing of natural language has gained momentum with the necessity to provide the right machine a thorough description of linguistic phenomena and formalized so that information is available for use by programs for automatic generation of texts (oral and written), text translation, information retrieval and others. For these programs to work efficiently it is necessary to enter data on the greatest number of possible linguistic structures, described in detail. This paper presents an approach on the place of proper names (compound), toponymic compounds for automatic processing of Portuguese in Brazil.

**Keywords:** Toponymic compounds. Automatic processing of Portuguese in Brazil. Description language. Formalization.

### Introdução

Com o advento da inteligência artificial avultaram-se os esforços na tentativa de facilitar o acesso do ser humano à máquina, nesse contexto evidencia-se a necessidade de se estabelecer uma comunicação eficaz entre o ser humano e a máquina. Essa necessidade de comunicação esbarra-se no fato de que o ser humano utiliza-se de uma linguagem natural, complexa, que conta com uma série de recursos não perceptíveis à máquina, como as expressões faciais, os gestos, o conhecimento de mundo dos interlocutores, entre outros. A máquina não é, portanto, capaz de “compreender” as nuances da linguagem natural. Uma das

---

\* Mestre em Estudos Linguísticos pelo Departamento de Línguas e Letras do Centro de Ciências Humanas e Naturais da Universidade Federal do Espírito Santo – Brasil. Endereço eletrônico - rvlima5@hotmail.com. Orientação: Prof.<sup>a</sup> Dr.<sup>a</sup> Aucione Das Dores Smarsaro.

soluções para a minimização dessa dificuldade é a inserção do máximo de informações acerca da estrutura e do funcionamento das línguas naturais na máquina. A tarefa exige a participação de linguistas, profissionais habilitados a realizar descrições linguísticas apropriadas e de cientistas da computação, habilitados a implementar nas máquinas as descrições linguísticas formalizadas.

A descrição das estruturas de uma língua natural é fundamental para o bom funcionamento dos sistemas de informática destinados ao tratamento automático de textos em língua natural. Os compostos toponímicos figuram entre as estruturas linguísticas cuja formalização da descrição coopera para a otimização dos sistemas de tratamento automático de textos em língua natural. O artigo apresenta uma abordagem dos compostos toponímicos, quanto aos aspectos linguísticos, para o processamento automático do português do Brasil .

### **Os nomes compostos comuns e os compostos toponímicos – critérios de identificação**

Os nomes compostos são objeto de estudo de gramáticos, linguistas e lexicógrafos. As pesquisas acerca dos compostos abordam diferentes aspectos desses nomes, especialmente dos nomes compostos comuns.

A presença do hífen como separador dos elementos de uma sequência de palavras, embora seja facilmente encontrada nos exemplos de palavras compostas elencados nas gramáticas normativas, não se constitui um critério eficaz de identificação de um grupo nominal como composto. Ranchhod e Carvalho (2003, p. 212-213) afirmam que

Os nomes compostos que se grafam com hífen são normalmente identificados pelos lexicógrafos e registrados nos dicionários. Contudo, muitos deles não têm tal grafia, nem esta seria um critério lingüístico adequado para caracterizar uma sequência de palavras como unidade lexical. De facto, na identificação de compostos há que utilizar um conjunto de critérios linguísticos [...] (RANCHHOD; CARVALHO, 2003, p. 212-213).

A presença do hífen como separador dos elementos de uma sequência de palavras pode ser considerada como um aspecto gráfico que facilita a identificação de alguns grupos nominais como compostos, porém, a presença do hífen nos compostos não é condição imprescindível.

Através da observação das propriedades linguísticas dos grupos nominais é possível estabelecer quais sequências devem ser consideradas nomes compostos, essas propriedades

nem sempre são comuns aos diversos tipos de compostos. Os critérios adotados na identificação de uma sequência como unidade lexical, por evidenciarem determinadas propriedades linguísticas, não asseguram a identificação de todo e qualquer composto. Há de se considerar o comportamento linguístico de cada sequência de palavras candidata a composto e eleger os critérios que melhor salientem as propriedades linguísticas que permitem o reconhecimento de tal sequência como um composto.

De acordo com Monteiro (1987, p. 164) constitui um composto “o vocábulo formado pela união de dois ou mais semantemas. Os componentes graficamente podem estar ligados (aguardente, passatempo), hifenizados (vira-lata, franco-suíço) ou soltos (Porto Alegre, Mato Grosso)”. Os elementos que formam os compostos podem aparecer “soltos”, não ligados por hífen. Os nomes que apresentam os elementos formantes unidos através de hífen ou que apresentam um espaço em branco entre os elementos formantes são considerados, no processamento automático de linguagem natural (PLN), nomes compostos. Os compostos que apresentam os elementos formantes “graficamente ligados”, sem a presença de um separador, são tratados como palavras simples.

O autor apresenta algumas características morfossintáticas que auxiliam na identificação dos compostos, são elas:

1. Flexão exclusiva do último componente.
  2. Sufixação relacionada ao composto como um todo.
  3. Impossibilidade de intercalação de novos determinantes.
  4. Impossibilidade de disjunção ou alteração da ordem dos constituintes.
  5. Impossibilidade de supressão de um dos elementos.
- (MONTEIRO, 1987, p. 164).

Monteiro considera a *impossibilidade de intercalação de novos determinantes*, a *impossibilidade de disjunção ou alteração da ordem dos constituintes* e *impossibilidade de supressão de um dos elementos*, aspectos que dizem respeito à distribuição sintática dos elementos formadores dos compostos, menos confiáveis que os aspectos morfológicos.

O critério da *flexão exclusiva do último componente* é considerado por Monteiro um critério mais confiável que os critérios sintáticos na identificação dos de nomes compostos. Observa-se que os exemplos de compostos comuns apresentados pelo autor admitem a flexão apenas do último elemento:

aguardente → aguardentes

passatempo → passatempos

vira-lata → vira-latas

franco-suíço → franco-suíços

Há, contudo, compostos em que a variação de número ocorre apenas no primeiro elemento e outros em que a variação ocorre nos dois elementos, é o caso, respectivamente, dos “compostos onde haja preposição, clara ou oculta” ou dos “compostos de dois substantivos, onde o segundo exprime ideia de *fim*, *semelhança*, ou limita a significação do primeiro” e dos “compostos de dois *substantivos*, de um *substantivo* e um adjetivo ou de um *adjetivo* e um *substantivo*” (BECHARA, 2010, p. 130, grifo do autor). Observa-se que a flexão de número dá-se de forma distinta nos compostos de formações diversas e que o comportamento dos elementos que formam os compostos é distinto da sua realização como item lexical (individualmente). O substantivo *moleque*, por exemplo, admite plural (*moleques*), porém, ao integrar um composto, como em *pé de moleque*, a variação em número do substantivo *moleque* não é admissível (plural de *pé de moleque* – *pés de moleque*).

Monteiro apresenta como exemplos de compostos “soltos” os nomes próprios *Porto Alegre* e *Mato Grosso* (nomes de cidade e estado, respectivamente) fato que inviabiliza a flexão de número de um de seus elementos, visto que os compostos toponímicos são institucionalizados. A aplicação do critério da *flexão exclusiva do último componente* a esses nomes não evidencia as propriedades linguísticas que permitem identificá-los como compostos. A aplicação de critérios de identificação de compostos deve levar em consideração os diferentes tipos de compostos (quanto à estrutura e estatuto de nome comum ou próprio).

Os topônimos constituem-se, em geral, de um termo genérico (que indica o tipo de acidente geográfico nomeado) e de um elemento específico. Há casos em que a variação de número do elemento genérico é possível devido à ocorrência de homonímia entre diversos acidentes geográficos que constituem determinado conjunto. Essa variação de número do primeiro elemento do composto toponímico relaciona-se, portanto, à natureza decomponível do termo genérico, como por exemplo, em *Ilhas Canárias* (arquipélago formado de sete ilhas) e *Montes Pirineus* (cordilheira – cadeia de montanhas).

### Variação do grau do adjetivo nos nomes compostos de estrutura nome+adjetivo

De acordo com Gross (1986) um nome composto é, por natureza, não composicional, seu significado deve ser deduzido do todo e não das partes que o formam. As propriedades morfossintáticas dos compostos relacionam-se ao conjunto e não aos elementos formadores de maneira isolada. Por exemplo, o adjetivo *teimoso* é passível de gradação quando ocorre como um item lexical, como se observa em

(1) Moisés é uma pessoa muito *teimosa*.

(1b) Moisés é uma pessoa *teimosíssima*.

Contudo, não se admite a variação do grau do adjetivo *teimoso* quando ele realiza-se como parte integrante de um composto, conforme observa-se nas frases:

(2) O filho de Moisés ganhou um *joão-teimoso*.

(2a) \*O filho de Moisés ganhou um *joão-teimosíssimo*.<sup>1</sup>

O adjetivo *teimoso*, como parte integrante de um nome composto, apresenta um comportamento distinto do observado na sua realização como item lexical independente.

Há adjetivos, contudo, que não admitem gradação. Segundo Cunha e Cintra (2008, p. 260) uma das funções dos adjetivos é “estabelecer com o substantivo uma relação”, os adjetivos que desempenham prioritariamente essa função são chamados pelos autores de *adjetivos relacionais*. Os adjetivos relacionais, por seu caráter classificatório, não admitem gradação. Assim sendo, a utilização do critério de variação do grau do adjetivo em compostos de estrutura nome-adjetivo (NAdj) apresentada resultados mais satisfatórios quando aplicado aos compostos NAdj em que o segundo constituinte é um adjetivo não relacional, ou seja, um adjetivo qualificador.

---

<sup>1</sup> O asterisco marca a inaceitabilidade da frase, ou seja, indica que a frase marcada não pode ser considerada uma paráfrase da sequência apresentada como modelo.

### A sufixação relacionada ao composto como um todo

A atribuição de sufixos aos compostos toponímicos dá-se da mesma forma que nos nomes compostos comuns. A sufixação relacionada ao composto como um todo é uma das características morfossintáticas que, segundo Monteiro (1987, p. 164), auxiliam na identificação dos compostos. O caráter não composicional dos nomes compostos acarreta a inviabilidade da atribuição de sufixo a apenas um dos elementos formadores desses nomes, ou seja, o sufixo deve referir-se ao composto como um conjunto indissociável, como se observa nos exemplos:

(3) “No Fantástico, nós já mostramos o *pão-duro* mor do Brasil: José Adelço, do Rio Grande do Norte” (grifo nosso).<sup>2</sup>

(4) “Outro que entende tudo de *pão-durismo* é o carioca e jornalista Gustavo Nagib” (grifo nosso).<sup>3</sup>

O sufixo *-ismo* (4) refere-se ao nome *pão-duro* como um conjunto. Nota-se que a forma *durismo* sequer realiza-se isoladamente na língua portuguesa, como um item lexical, ocorre apenas como elemento constituinte do composto. Nas frases (3) e (4) as palavras *pão-duro* e *pão-durismo* mantêm o estatuto de palavra comum.

Na frase (6) o sufixo *-ense* refere-se ao nome *Mato Grosso* como um todo, por tratar-se de um nome composto. A sufixação do composto toponímico (*estado do Mato Grosso*)<sup>4</sup> mantêm a condição de composto do nome do qual se origina, à semelhança do que ocorre em (4), porém, em (6) ocorre a mudança de estatuto da palavra, de nome próprio (topônimo - *Mato Grosso*) a nome comum (*mato-grossense*).

(5) “O *Mato Grosso* é uma das 27 unidades federativas do Brasil” (grifo do autor).<sup>5</sup>

(6) “Kart: *Mato-grossense* [sic] Gabriel Sereia se isola na liderança na Copa São Paulo de Kart Granja Viana” (grifo nosso).<sup>6</sup>

<sup>2</sup> Disponível em: <<http://g1.globo.com/Noticias/Brasil>>. Acesso em: 11 mai. 2010.

<sup>3</sup> Disponível em: <<http://g1.globo.com/Noticias/Brasil>>. Acesso em: 11 mai. 2010.

<sup>4</sup> Consideramos *Mato Grosso* a forma reduzida do composto toponímico *estado do Mato Grosso*.

<sup>5</sup> Disponível em: <<http://pt.wikipedia.org/wiki>>. Acesso em: 11 mai. 2010.

<sup>6</sup> Disponível em: <<http://www.velocidadeonline.com.br>>. Acesso em: 11 mai. 2010.

Outra mudança decorrente do acréscimo do sufixo *-ense* ao topônimo *Mato Grosso*, em (6), é a presença do hífen como separador dos elementos formantes do composto sufixado enquanto o separador dos elementos formantes do composto toponímico (estado do) *Mato Grosso* é o espaço em branco.

### Apagamento de um dos elementos do composto

De acordo com Dick (1990, p. 10) o sintagma toponímico apresenta na sua estrutura um elemento relativo à entidade geográfica que recebe a nomeação (termo ou elemento genérico), o outro elemento é o topônimo propriamente dito (ou termo específico). Considerando-se o elemento genérico como parte integrante do composto toponímico e o fato de que ele contém o significado geral do composto, pode-se afirmar que é possível o apagamento do segundo nome (N2) que constitui os compostos toponímicos de estrutura nome-nome (NN) e nome- preposição *de*-nome (NdeN), com se observa em

(7) “A *praia de Copacabana* esteve lotada no local do torneio e a comunidade do Futevôlei estava em festa, como um reencontro dos maiores do esporte e seus amigos” (grifo nosso).<sup>7</sup>

(7a) A *praia* esteve lotada no local do torneio e a comunidade do Futevôlei estava em festa, como um reencontro dos maiores do esporte e seus amigos.

(8) “A *Lagoa Rodrigues de Freitas* faz parte de um complexo lagunar em meio urbano, apresentando em seu entorno aspectos peculiares quanto à questão de circulação das águas [...] (grifo nosso)”.<sup>8</sup>

(8a) A *Lagoa* faz parte de um complexo lagunar em meio urbano, apresentando em seu entorno aspectos peculiares quanto à questão de circulação das águas.

O apagamento do N2 nos compostos toponímicos *praia de Copacabana* e *lagoa Rodrigues de Freitas* não acarreta a inaceitabilidade das frases (7a) e (8a).

As frases (7a) e (8a) podem ser consideradas paráfrases das frases (7) e (8) respectivamente, contudo, a compreensão das sequências (7a) e (8a) requer do leitor maior

<sup>7</sup> Disponível em: <<http://www.futevolei.com.br/>>. Acesso em: 10 mai. 2010.

<sup>8</sup> Disponível em: <<http://www.inea.rj.gov.br/rfreitas/index.asp>>. Acesso em: 10 mai. 2010.

conhecimento do contexto devido a falta de elementos que permitam a identificação da *praia* e da *lagoa* a que se referem essas frases.

### Os compostos toponímicos e o processamento automático da linguagem natural

Em uma etapa inicial do processamento automático de textos escritos em língua natural procede-se a segmentação do texto em parágrafos, frases e palavras, estas são então etiquetadas.

A identificação e etiquetagem adequada dos compostos toponímicos tornam mais eficaz a geração e tradução automática de textos em língua natural, bem como a recuperação de informação. *Etiquetar* uma palavra significa atribuir a ela determinada informação linguística, como, por exemplo, a classe gramatical a que pertence. Pode-se atribuir mais de uma etiqueta a uma mesma forma, por exemplo, o item *como* admite etiquetas diversas (*verbo* [comer], *conjunção* e *advérbio*), porém, cada etiqueta corresponde a uma classe gramatical somente. As etiquetas podem conter informações morfológicas e sintáticas, como a classe da palavra, a variação de número, variação de grau, variação de gênero, etc.

“A maioria das operações de processamento automático das línguas naturais tem como objectivo último extrair dos textos unidades de significado” (RANCHHOD; CARVALHO, 2003, p. 212). Nesse processo a identificação prévia dos nomes compostos é imprescindível, a fim de evitar o equívoco de tratar cada um dos elementos formantes do composto isoladamente, como uma palavra simples.

Os compostos toponímicos que apresentam o hífen ou o apóstrofo como separador dos elementos formantes são identificados facilmente como uma unidade lexical. Savary (2000, p. 42) explica que

Un mot composé est non ambigu s'il est composé quel que soit son contexte d'occurrence. Le nombre de composés non ambigu est petit par rapport au nombre de tous les composés. On y compte surtout les séquences qui contiennent des constituants n'ayant pas de statut indépendant<sup>9</sup> (SAVARY, 2000, p. 42).

---

<sup>9</sup>“Uma palavra composta é não ambígua se ela for composta independentemente do seu contexto de ocorrência. O número de compostos não ambíguos é pequeno em relação ao número de todos os compostos. Contam-se sobretudo as sequências que contêm constituintes que não têm estatuto independente” (tradução nossa).



As palavras compostas não ambíguas são, portanto, aquelas que, em sua sequência apresentam um componente que não se realiza de forma independente na língua, como, por exemplo, *durismo* e *grossense* nas frases (4) e (6).

(4) “Outro que entende tudo de pão-*durismo* é o carioca e jornalista Gustavo Nagib” (grifo nosso).

6) “Kart: Mato-*grossense* Gabriel Sereia se isola na liderança na Copa São Paulo de Kart Granja Viana” (grifo nosso).

O número de compostos não ambíguos, porém, é pequeno em relação à totalidade dos nomes compostos. Os compostos que apresentam o espaço em branco como separador podem ser considerados compostos ambíguos.

Os nomes próprios “servem para conferir identidade exclusiva a um membro de uma classe ou espécie” (AZEREDO, 2008, p. 156) não é de se esperar, portanto, que sejam elementos geradores de ambiguidade. Há, contudo, a possibilidade de conversão de um nome próprio em nome comum ou de um nome comum em nome próprio. De acordo com Azeredo (2008, p. 156) “as palavras podem mudar de classe ou de subclasse no âmbito da respectiva classe”. O autor apresenta como exemplos desse processo antropônimos (nomes próprios de pessoas – prenomes e sobrenomes) oriundos de nomes comuns (*Violeta, Coelho, Rosa, Leitão, Madeira, Machado e Lírio*) e nomes comuns oriundos de nomes próprios (*xerox, gilete, gari e quixote*). Os exemplos apresentados são palavras simples. Processo semelhante ocorre com os compostos (os nomes compostos comuns e os nomes compostos próprios), sejam eles antropônimos ou topônimos (nomes de lugar), como pode-se verificar no verbete: “*maria-isabel* s.f. CUL B C.-O. refogado de carne-seca com arroz ” (HOUAISS 2001, p. 1853, grifo do autor). O substantivo comum *maria-isabel* apresenta como elementos formantes os antropônimos *Maria* e *Isabel*.

Quanto aos nomes próprios de lugar, Dick (1997, p. 45) afirma que os nomes explicativos “com o uso contínuo pelos falantes, acabam por se transformar em topônimos”. Expressões que denotam características do local indicado podem, com o uso, passar a funcionar como compostos toponímicos. A mudança de *status* de expressão descritiva a nome próprio acarreta algumas alterações no comportamento linguístico dessas palavras. O registro das alterações decorrentes da mudança de *status* dos nomes (de comum à próprio e de próprio à comum) é de suma importância para que esses dados possam ser

utilizados no processamento automático de textos do português do Brasil. A mudança observada na passagem da expressão descritiva de *um morro em que se encontra uma ucharia* (um armazém) para o composto toponímico *morro da Ucharia* (município de Vila Velha - ES)<sup>10</sup> diz respeito não apenas à conversão de uma sequência de palavras comuns em nome próprio, mas indica também a conversão de um grupo nominal livre em nome composto. A percepção e registro dessas mudanças é imprescindível ao PLN, por contribuir para a descrição linguística. O registro das motivações do processo não é essencial para o PLN, mas os resultados dele advindos. Na frase:

(9) “A Ilha de Páscoa [...] é uma das ilhas mais isoladas do mundo”<sup>11</sup>

o substantivo *ilha* realiza-se como nome próprio, um elemento constituinte do composto toponímico *Ilha de Páscoa* e pode realizar-se como nome comum, ao indicar um tipo de acidente geográfico natural. A identificação do topônimo *Ilha de Páscoa* como composto permite que lhe sejam atribuídas etiquetas adequadas, referentes ao conjunto e contribui para que os substantivos *Ilha* e *Páscoa* não sejam analisados isoladamente.

Para que os nomes *Ilha de Páscoa*, *ilha* e *páscoa* sejam reconhecidos adequadamente pela máquina é preciso que cada um deles seja etiquetado, por exemplo:

<*Ilha de Páscoa*.Npr:fs>

<*ilha*.N:fs>

<*páscoa*.N:fs>

A letra *N* indica a classe da palavra (*nome*), as letras *pr* indicam que se trata de um nome *próprio* e as letras após os dois pontos representam as informações acerca da flexão dos nomes (*fs* = feminino singular). Outras informações linguísticas podem ser adicionadas às etiquetas.

A disponibilização apenas da etiqueta <*Ilha de Páscoa*.Npr:fs> em uma ferramenta ou software usado no tratamento automático de textos em língua natural afeta o processamento automático de frases, como

<sup>10</sup> Informação toponímica acerca de *morro da Ucharia* disponível em:

<<http://www.morrodomoreno.com.br/roberto2.htm>>. Acesso em: 10 mai. 2010.

<sup>11</sup> Disponível em: <[www.revistaturismo.com.br](http://www.revistaturismo.com.br)>. Acesso em: 10 mai. 2010.

(10) Ana e Moisés passaram o fim de semana da *páscoa* em uma *ilha* paradisíaca.

O reconhecimento das palavras *ilha* e *páscoa* como itens lexicais que se realizam de forma independente na língua fica prejudicado.

Uma ferramenta ou software usado no tratamento automático de textos em língua natural que disponha apenas das etiquetas <*ilha*.N:fs> e <*páscoa*.N:fs> também é prejudicial para o reconhecimento do composto toponímico *ilha de Páscoa* e o processamento automático da frase

(9) “A Ilha de Páscoa [...] é uma das ilhas mais isoladas do mundo”.

A atribuição das etiquetas <*ilha de Páscoa*.Npr:fs>, <*ilha*.N:fs>, <*páscoa*.N:fs> é igualmente importante .

### **Algumas considerações**

O bom desempenho dos sistemas de informática que lidam com o PLN está intimamente relacionado à qualidade das descrições neles inseridas. Os recursos linguísticos desses sistemas podem ser, entre outros, em forma de dicionários, tábuas e gramáticas. A construção e implementação desses recursos dependem dos trabalhos desenvolvidos principalmente por linguistas e engenheiros da computação. Cabe aos linguistas a realização de descrições linguísticas precisas e exaustivas e aos engenheiros da computação a implementação dessas informações nos sistemas de informática.

O processamento automático de textos em língua natural exige a descrição de todos os elementos que o compõe e formalização da descrição. O tratamento dos nomes próprios na teoria linguística é por vezes relegado a um segundo plano e segundo Correia (2002, p. 118) as discussões acerca dos nomes próprios (Npr) “correm o risco de se reduzirem, apenas, à discussão do valor referencial dos nomes, afastando-se da distinção pertinente entre o que é, sob o ponto de vista da lingüística um Npr e um Nc”. Para o PLN essa distinção é essencial devido a ambigüidade causada, por exemplo, pela semelhança entre os compostos toponímicos e os grupos nominais livres e pelo fato do comportamento dos elementos que constituem um composto toponímico ser diferenciado da sua realização de forma

independente, como palavras simples da língua. Os estudos acerca dos compostos toponímicos vão, portanto, além das discussões acerca do seu valor referencial. As pesquisas de descrição linguística dos compostos toponímicos podem contribuir de forma significativa para a melhoria dos sistemas de informática destinados ao PLN bem como para a melhoria do ensino da língua portuguesa.

## Referências

AZEREDO, J. C. de. *Gramática Houaiss da língua portuguesa*. 2. ed. São Paulo: Publifolha, 2008.

BECHARA, E. *Moderna gramática portuguesa*. 37. ed. rev. e ampl. e atual. conforme o novo Acordo Ortográfico. Rio de Janeiro: Nova Fronteira, 2009.

CORREIA, C. N. *Estudos de determinação: a operação de quantificação-qualificação em sintagmas nominais*. Lisboa: Fundação Calouste Gulbenkian, 2002.

CUNHA, C.; CINTRA, L. F. L. *Nova gramática do português contemporâneo*. 4. ed. rev. e ampl. Rio de Janeiro: Lexikon, 2007.

DICK, M. V. de P. do A. *A dinâmica dos nomes na cidade de São Paulo: 1554-1897*. 2. ed. São Paulo: Annablume, 1997.

GROSS, M. Lexicon-Grammar. The Representation of Compound Words. In: *COLING-1986 Proceedings*. Bonn, 1986. p. 1-6.

HOUAISS, A.; VILLAR, M. de S. *Grande dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2001.

MONTEIRO, J. L. *Morfologia portuguesa*. Fortaleza: EDUFC, 1987.

PAUMIER, S. *Unitex 1.2*. Manuel de l'utilisation. Université de Paris-Est, Marne-la-Vallée, 2006. Disponível em: <<http://www-igm.univ-mlv.fr/~unitex>>. Acesso em: 14 dez. 2009.

RANCHHOD, Elisabete Marques; CARVALHO, Paula. Unidades lexicais complexas. Problemas de análise e etiquetagem. In: *Actas del VIII Simpósio Internacional de Comunicación Social*. Santiago de Cuba, 2003. p. 212-217.