

**DESCRIÇÃO DE COMPOSTOS TOPONÍMICOS E
FORMALIZAÇÃO PARA PROCESSAMENTO
AUTOMÁTICO DE LINGUAGEM NATURAL**

Rosana de Vilhena Lima

Dissertação de Mestrado em Estudos Linguísticos

Universidade Federal do Espírito Santo

Vitória, março de 2010

Rosana de Vilhena Lima

DESCRIÇÃO DE COMPOSTOS
TOPONÍMICOS E FORMALIZAÇÃO PARA
PROCESSAMENTO DE LINGUAGEM
NATURAL

Dissertação apresentada ao Programa de Pós-Graduação 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.

Aprovada em 12 de março de 2010.

Comissão Examinadora:

Profª. Drª. Aucione das Dores Smarsaro – UFES
Orientadora, Presidente da Sessão e da Comissão Examinadora

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

Prof. Dr. Orivaldo de Lira Tavares – UFES
Membro Titular Interno da Comissão Examinadora

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

L732d Lima, Rosana de Vilhena, 1967-
Descrição de compostos toponímicos e formalização para processamento automático de linguagem natural / Rosana de Vilhena Lima. – 2010.
131 f. : il.

Orientadora: Aucione das Dores Smarsaro.
Dissertação (mestrado) – Universidade Federal do Espírito Santo, Centro de Ciências Humanas e Naturais.

1. Lingüística - Processamento de dados. 2. Processamento de linguagem natural (Computação). 3. Dicionários eletrônicos. 4. Toponímia. 5. Língua portuguesa - Espírito Santo (Estado) - Palavras compostas. I. Smarsaro, Aucione das Dores. II. Universidade Federal do Espírito Santo. Centro de Ciências Humanas e Naturais. III. Título.

CDU: 80

Dedico esta dissertação à Djacira Gonçalves de Vilhena (*in memoriam*), mulher virtuosa, cujo valor excede ao das jóias mais preciosas.

AGRADECIMENTOS

Antes e acima de tudo minha gratidão, toda honra e todo louvor a Deus. “Porque dEle, e por Ele, e para Ele são todas as coisas; glória, pois, a Ele eternamente” (Rm 11. 36).

Quero expressar minha sincera gratidão

À profa. Dra. Aucione Smarsaro, por ter me apresentado (na graduação, através do seu trabalho de divulgação dos estudos em PLN no Departamento de Línguas e Letras da UFES) os fundamentos da Linguística Computacional e ter-me franqueado a oportunidade de participar das pesquisas do NUPLIC (Núcleo de Pesquisas em Linguística Computacional), pela confiança em mim depositada, pela orientação da pesquisa e por sua amizade.

Ao Prof. Dr. Eric Laporte, pelos preciosos esclarecimentos acerca de conceitos concernentes ao PLN, pelas leituras e sugestões na condução da pesquisa, pela presteza e atenção com que sempre respondeu aos meus questionamentos;

À profa. Dra. Margarida Basílio e à profa. Dra. Marta Sherre pelas contribuições e observações feitas durante o exame de qualificação;

À colega Alzira Davel, pelas conversas que até em momentos difíceis nos renderam boas risadas.

Aos professores, amigos e colegas que de alguma maneira cooperaram para que a pesquisa de mestrado chegasse a bom termo.

Agradeço também à Fundação de Apoio à Ciência e Tecnologia do Espírito Santo (FAPES), pelo suporte financeiro concedido através da bolsa de mestrado.

Agradeço especialmente:

Ao meu pai, Guaracy de Vilhena, por todo amor e pelo ensino de valores eternos que nortearam e norteiam meus passos;

A Carlos Lima, meu maior incentivador, ajudador, companheiro de todas as horas, pelo apoio sempre presente, pela paciência em ouvir-me;

A Juliana e Olavo, presentes de Deus para a minha vida, pela compreensão e pelas muitas palavras de carinho e incentivo.

RESUMO

Um desafio oriundo do desenvolvimento das ciências da computação é a comunicação do ser humano com a máquina através de linguagem natural. Os avanços tecnológicos observados nos últimos tempos impulsionaram o desenvolvimento de pesquisas que exigem o conhecimento de áreas diversas, entre elas a engenharia e a linguística, no campo da inteligência artificial, especificamente, para processamento automático das línguas naturais. Os pesquisadores dessa área têm como meta comum a otimização do acesso do humano às máquinas. Nesse processo a compreensão das estruturas que compõem as línguas naturais e o complexo conjunto de regras que as governam é imprescindível, pois, essas informações formalizadas e codificadas podem ser inseridas nos sistemas informáticos a fim de facilitar o acesso do ser humano ao computador. O objetivo desta pesquisa é a elaboração de um dicionário eletrônico para o português do Brasil, especificamente a descrição de palavras compostas toponímicas do estado do Espírito Santo (Grande Vitória) com estrutura: substantivo - de - substantivo, substantivo - substantivo e substantivo - adjetivo, a partir dos pressupostos teórico-metodológicos do Léxico-Gramática desenvolvidos por Maurice Gross (1975). Um dicionário eletrônico de palavras toponímicas constitui-se numa ferramenta que viabilizará melhorias em *softwares* que lidam com processamento de linguagem natural, como, por exemplo, programas de recuperação de informação, em que os dicionários contribuem na função de identificar os itens lexicais selecionados, levando em conta as propriedades morfossintáticas e semânticas dos compostos toponímicos.

Palavras-chave: *Composto toponímico; processamento automático do português do Brasil; dicionário eletrônico.*

ABSTRACT

One challenge comes from the development of computer science is the communication of man with the machine through natural language. The developments seen in recent years has boosted the development of research that requires knowledge from several areas, including engineering and linguistics in the field of artificial intelligence, specifically in the automatic processing of natural languages. The researchers in this area have the common goal to optimize the human access to machines. In the process of understanding the structures that make natural languages and complex set of rules that govern them is essential, therefore, these formalized and encoded information can be entered in informat systems to facilitate the access of human beings to the computer. The objective of this research is the development of an electronic dictionary for the Portuguese of Brazil, specifically the description of toponymic compounds the state of Espírito Santo (Grande Vitória) frame: noun - of - noun, noun - noun and noun - adjective, from the theoretical and methodological principles of the Lexicon-Grammar developed by Maurice Gross (1975). An electronic dictionary of words to place names is a tool that will enable improvements in software that deals with natural language processing, such as programs, information retrieval, where the dictionaries help identify the function of the lexical items selected, taking into to the morphosyntactic and semantic properties of the toponymic compounds.

Keywords: Toponymic compound; automatic processing of Portuguese in Brazil; electronic dictionary.

NOTAÇÕES

Adj – Adjetivo.

AdjN – Composto formado de um adjetivo seguido de nome.

AN – Acidente natural.

AA – Acidente artificial.

C – Complemento.

Det – Determinante.

+D1 – Indica a ocorrência do determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição.

+D2 – Indica a ocorrência do determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição e a elisão de N1.

+DZ2 – Indica a ausência de determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição e a elisão de N1.

+DZ1 – Indica a ausência de determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição.

E – Indica um elemento neutro ou a ausência de um elemento em uma sequência. Por exemplo, na sequência: (*artigo definido* + **E**) *Ana viajou para São Paulo* significa *A Ana viajou para São Paulo* ou *Ana viajou para São Paulo*.

f – Gênero feminino.

+g1 – Indica que a sequência $NI(Prep)N2$ não apresenta variação de gênero do determinante em decorrência do apagamento de N1.

+g2 – Indica que a sequência $NI(Prep)N2$ apresenta variação de gênero do determinante em decorrência do apagamento de N1.

m – Gênero masculino.

N – Substantivo. Esta notação é usada de forma um tanto ambígua.

Designa um grupo nominal ou os elementos de uma frase. Neste caso os índices numéricos permitem que sejam identificados diferentes grupos nominais, por exemplo, em uma sentença de estrutura SVC (sujeito, verbo e complementos – *Ana entregou os livros para Luíza*) o N0 refere-se ao sujeito (*Ana*), o N1 ao primeiro complemento (*os livros*) o N2 ao segundo complemento (*para Luíza*). A classe dos nomes pode ser especificada, como, por exemplo, Nhum (nome humano), Npr (nome próprio) etc. A letra *N* acrescida de índices numéricos também é utilizada para indicar os elementos

constituintes do nome composto, por exemplo, em *código Morse* o N1 é comum (N1 =: Nc) e o N2 próprio (N2 =: Npr).

NAdj – Composto formado de um nome seguido de um adjetivo.

Nc – Nome comum.

Ncomp – Nome composto.

NdeN – Nome mais preposição *de* mais nome.

NN – Nome mais nome.

Npr – Nome próprio.

N1 – primeiro substantivo do composto.

+N1Z – Indica a elisão de N1 na sequência N1(Prep)N2.

-N1Z – Indica a impossibilidade de elisão de N1 na sequência N1(Prep)N2.

N2 – segundo substantivo do composto, que pode ser simples ou composto.

+N2Z – Indica que a sequência N1(Prep)N2 apresenta elisão de N2.

-N2Z – Indica que a sequência N1(Prep)N2 não apresenta elisão de N2.

+N2p – Indica que o N2 da sequência N1(Prep)N2 é plural.

-N2p – Indica que quando o N2 da sequência N1(Prep)N2 é singular.

p – Plural.

PLN – Processamento automático de linguagem natural.

Prep – Preposição.

Pron – Pronome.

Pron-r – Pronome relativo.

s – Singular.

V – Verbo.

Vc – Verbo copulativo.

W – Complemento.

Z – Elemento neutro ou a ausência do elemento à esquerda.

* – Inaceitabilidade do termo correspondente.

= – É analisado(a).

=: – Para explicar uma classe ou uma estrutura. Por exemplo: N0 =: Npr (o nome que funciona como sujeito é um nome próprio).

+ – Funciona como “ou”. Por exemplo, *Ana escalou o (monte + morro) Mestre Álvaro* = *Ana escalou o monte Mestre Álvaro* **ou** *Ana escalou o morro Mestre Álvaro*.

/ – Funciona como “ou”. *Ana escalou o monte / morro Mestre Álvaro* significa: *Ana escalou o monte Mestre Álvaro* **ou** *Ana escalou o morro Mestre Álvaro*.

- Serve para intruzir um comentário opcional nas entradas dos dicionários do sistema DELA.
- : – Funciona como “ou”, como separador dos códigos flexionais nas entradas dos dicionários do sistema DELA.
 - () – Indicam fatores do produto, o produto é a operação de concatenação não comutativa. Em outras palavras, os parênteses indicam os elementos que concorrem para a obtenção de um resultado (produto). Por exemplo, *Ana comeu um pão (bolo + doce) mole* significa: *Ana comeu um pão mole*, ou *Ana comeu um bolo mole*, ou *Ana comeu um doce mole*.
 - [] – Indicam as transformações. Por exemplo: *Ana beijou Moisés [passiva] = Moisés foi beijado por Ana*.
 - # – Indica descontinuidade na entonação.

LISTA DE FIGURAS

Figura 1 – Parte da lista das entradas do DELAF aplicado ao texto <i>Senhora</i> de José de Alencar	68
Figura 2 – Parte da lista das entradas do DELACF aplicado ao texto <i>Senhora</i> de José de Alencar	68
Figura 3 – Fragmento do texto <i>Senhora</i> de José de Alencar, pré-processado pelo sistema Unitex 2.0	70
Figura 4 – Grafo de um fragmento do texto <i>Senhora</i> de José de Alencar	71
Figura 5 – Grafo do composto toponímico <i>Rio de Janeiro</i>	72
Figura 6 – Grafo das frases (1a) <i>Ana mora na Praia da Costa</i> e (1b) <i>Ana mora no Praia da Costa</i>	72

SUMÁRIO

1 – INTRODUÇÃO	14
1.1 – Apresentação do tema	14
1.2 – Objetivos	19
1.3 – Justificativa	20
2 – PRESSUPOSTOS TEÓRICO-METODOLÓGICOS	24
2.1 – A Gramática Transformacional de Zellig Harris	24
2.2 – O Léxico-Gramática de Maurice Gross	28
2.3 – Metodologia	31
3 – O NOME COMPOSTO	34
3.1 – Na Gramática Tradicional	34
3.2 – Na Linguística	37
3.3 – No Léxico-Gramática	41
4 – O NOME PRÓPRIO	43
4.1 – Abordagens acerca do Nome Próprio	43
4.1.1 – O nome próprio na Gramática Tradicional	43
4.1.2 – O nome próprio na Linguística	46
4.1.3 – O nome próprio na Lexicografia - O uso de inicial maiúscula	49
4.2 – A descrição do nome próprio para o PLN.....	50
4.3 – Do nome comum ao nome próprio – <i>um continuum</i>	52
4.4 – Os topônimos	59
4.5 – Os compostos toponímicos	61
5 – DICIONÁRIO ELETRÔNICO DE NOMES COMPOSTOS TOPONÍMICOS DO PORTUGUES DO BRASIL	64
5.1 – Dicionário digital	65

5.2 – Dicionário eletrônico	66
6 – CRITÉRIOS DE IDENTIFICAÇÃO DOS COMPOSTOS TOPONÍMICOS	73
6.1 – Topônimos com a estrutura NN	77
6.1.1 - Elisão de um dos substantivos	78
6.1.2 – Inserção de um item lexical no grupo nominal	82
6.1.3 – Ruptura paradigmática	83
6.1.4 – Variação em número	85
6.2 – Topônimos com a estrutura NAdj	87
6.2.1 – Perda da predicatividade do adjetivo	87
6.2.2 – Variação do adjetivo em grau	91
6.2.3 – Coordenação do adjetivo com outro adjetivo	93
6.2.4 – Ruptura paradigmática	94
6.3 – Topônimos com a estrutura NdeN	94
6.3.1 – Inserção de um item lexical no grupo nominal	96
6.3.2 – Variação do determinante de N2	97
6.3.3 – Ruptura paradigmática	99
6.3.4 – Variação em número	100
7 – FORMALIZAÇÃO DA DESCRIÇÃO DOS COMPOSTOS TOPONÍMICOS	102
8 – ALGUMAS CONSIDERAÇÕES	115
8.1 – Propriedades distribucionais dos compostos toponímicos e sua formalização	115
8.2 – A descrição linguística dos nomes próprios compostos para o PLN	116
8.3 – Trabalhos futuros	118
REFERÊNCIAS	120
ANEXO A – Referências das sentenças retiradas da <i>internet</i> e outras fontes	127

1 – INTRODUÇÃO

1.1 – Apresentação do tema

Os avanços tecnológicos observados nos últimos tempos acontecem em uma velocidade vertiginosa. Diariamente são vistas, nos meios de comunicação, inovações em equipamentos informatizados que têm como objetivo proporcionar ao indivíduo novos conhecimentos, conforto e praticidade no desempenho de suas atividades diárias, entre tantas outras coisas.

Um dos desafios dessa época de desenvolvimento da ciência da computação é a comunicação do ser humano com a máquina através da linguagem natural.

Nesse contexto surgem pesquisas que exigem o conhecimento de áreas diversas, entre elas a engenharia e a linguística, no processamento automático das línguas naturais. Segundo Dias (1994: 2),

Uma das tarefas da Linguística dentro do Processamento da Linguagem Natural é possibilitar, com os recursos de que dispõem, um equilíbrio entre dois pólos aparentemente inconciliáveis: utilizar ao máximo o que é previsível e determinado dentro da língua e explorar o que ela oferece em termos de interpretação e expressividade, nem sempre previsíveis, pois dependem também de fatores extralingüísticos, como a situação e o conhecimento compartilhado.

Observa-se que a análise linguística para o processamento automático de linguagem natural (doravante PLN) considera os aspectos morfossintáticos, semânticos e pragmáticos do conjunto de elementos que constituem os textos a serem processados. Os aspectos morfossintáticos podem ser considerados como previsíveis enquanto os aspectos semânticos e pragmáticos situam-se em um campo menos previsível e mais dependente dos fatores extralingüísticos. Essas informações, que devem ser descritas da forma mais precisa possível, são reunidas, formalizadas, codificadas e inseridas em sistemas informáticos. Os sistemas informáticos são utilizados em máquinas, com a finalidade de maximizar o seu potencial de comunicação em língua natural com o ser humano.

De acordo com Whitby (2004: 24) cada pesquisador pode enfatizar determinada meta no PLN, ou ainda, um mesmo pesquisador pode enfatizar metas diferentes, conforme seu interesse ou clientela, mas eles “têm as metas simultâneas [...] de tornar os computadores mais fáceis de usar, compreendendo as complexas regras que governam a estrutura das línguas naturais e a descoberta de como os humanos podem aprender a aplicar essas regras”. Nota-se a importância da integração de linguistas e engenheiros no desenvolvimento das pesquisas em inteligência artificial e nos trabalhos de PLN. Essa integração linguística-engenharia proporciona um exame diferenciado dos profissionais de uma e de outra área quanto ao seu objeto de estudo, que redundam em benefício para o avanço das pesquisas através da colaboração mútua.

Nesse cenário insere-se de forma altamente significativa os estudos linguísticos e a descrição exaustiva das estruturas lexicais para o PLN, especialmente a análise de nomes compostos para elaboração de dicionários eletrônicos. As descrições de todos os tipos de estruturas compostas configuram-se um desafio contínuo para os linguistas. Cabe ressaltar que no contexto do PLN são consideradas compostos os itens lexicais constituídos por mais de um elemento e que funcionam como uma unidade, ou seja, nomes compostos por justaposição. No universo dos elementos que formam o conjunto dos compostos encontram-se nomes comuns e nomes próprios. Há compostos comuns que apresentam um nome próprio como elemento formador, por exemplo: código *Morse*, efeito *Doppler*, prêmio *Nobel*. Há também compostos toponímicos que apresentam um nome comum como elemento formador, por exemplo: *córrego Carapina*, *bairro Aribiri*, *canal Praia da Costa*, sem, contudo, deixar de ser um nome próprio.

Os nomes próprios são objeto de estudo da Onomástica e podem ser divididos em dois grandes grupos: *antropônimos* (nomes de pessoas) e *topônimos* (nomes de lugares), ou, como preferem alguns, *corônimos*. Os topônimos, muitas vezes, surgem a partir de nomes comuns que apresentam uma descrição ou indicam determinado aspecto do lugar a que se referem, assumindo posteriormente, pelo uso, o *status* de nome próprio. Dick (1997: 45) afirma que os nomes explicativos, “com o uso contínuo pelos falantes, acabam por se transformar em topônimos”. Citamos como exemplo desse fato o topônimo *Prainha* (acidente¹ artificial [AA] da cidade de Vila Velha, Espírito Santo), formado pelo diminutivo do substantivo comum

¹ Utilizamos o substantivo *acidente* para designar os acidentes geográficos, ou seja, as variações no relevo terrestre. As designações *acidente natural* e *acidente artificial* referem-se respectivamente aos acidentes geográficos encontrados na natureza, como as ilhas, os arquipélagos, as baías, os lagos, os rios etc.; e acidentes geográficos criados pelo ser humano, como as ruas, os edifícios, os bairros, as cidades etc.

praia (designativo de acidente natural [AN]), que com o uso toponimizou-se; ou *Ilha Bela* (AH da cidade de Vitória, Espírito Santo) formado pela justaposição do substantivo comum *ilha* (designativo de acidente natural) e do adjetivo *bela*, que expressa uma impressão do nomeador acerca do lugar nomeado. Esses elementos comuns, com o uso, converteram-se em um composto toponímico.

Na elaboração de um dicionário eletrônico para o português do Brasil é imprescindível uma descrição apropriada dos fenômenos linguísticos, principalmente para a resolução de ambiguidades. Segundo Laporte (2001: 23) “quando o número de acepções descritas cresce, as ambiguidades lexicais também crescem”. Para minimizar a dificuldade gerada pela ocorrência de palavras polissêmicas, que possibilitam diferentes interpretações do enunciado em que se inserem, parte-se da hipótese de que essas palavras devem ser registradas no dicionário eletrônico homonimicamente. Cada acepção da palavra deve representar uma entrada distinta, ser tratada como um item lexical distinto para maior eficácia da utilização do dicionário eletrônico em sistemas informáticos. Nesse processo a homonímia é considerada apenas como forma de entrada dos itens lexicais no dicionário eletrônico e não na análise diacrônica desses itens.

A ocorrência de ambiguidade abarca tanto palavras simples como as palavras compostas. Segundo Smarsaro (2004: 39) “uma palavra composta deve ser analisada computacionalmente como um bloco”. As palavras compostas devem ser observadas como um item lexical que apresenta um sentido global distinto do sentido atribuído às suas partes. A existência de um grupo nominal livre que apresenta a mesma estrutura do composto favorece a interpretação errônea do nome composto. A análise das propriedades linguísticas dos elementos formadores dos nomes compostos considerados independentemente acarreta uma descrição inadequada, como se observa, por exemplo, nas frases²:

(1) “Um israelense foi condenado a nove meses de prisão por atirar uma **cabeça de porco** em uma mesquita de Tel Aviv em agosto de 2005”.

² Dentre os exemplos apresentados encontram-se sentenças construídas e sentenças retiradas da *internet* e de outros suportes. Destas, algumas são transcritas na íntegra (e são grafadas entre aspas) e outras adaptadas. À semelhança dos exemplos construídos, as frases adaptadas também são grafadas sem o uso de aspas. Os exemplos construídos são elaborados a partir da observação da realização da língua falada, em elocuições formais ou informais. As referências das sentenças citadas, modificados ou não, encontram-se na listagem do anexo A.

- (2) “Moravam [...] numa **cabeça de porco**: quarto e banheiro [...], onde se amontoavam mais de vinte famílias – bairro da Glória”.

Os itens destacados em (1) constituem um grupo nominal livre, *cabeça de porco* significa a *parte superior do corpo de um animal, de um suíno*, enquanto na sentença (2) *cabeça de porco* significa *cortiço, moradia coletiva, de pessoas pobres*.

A identificação prévia da ocorrência de palavras compostas nos textos a serem processados automaticamente coopera para que a máquina interprete as propriedades linguísticas desses itens lexicais de forma global.

Smarsaro (2004) propõe a descrição de palavras compostas do português do Brasil e a formalização para o PLN. A autora apresenta uma lista de 1520 palavras compostas com a estrutura nome mais preposição *de* mais nome (NdeN) com a codificação das propriedades linguísticas analisadas. Embora Smarsaro (2004) não tenha restringido o seu *corpus* a palavras comuns observa-se que das 1520 estruturas codificadas 30 apresentam a propriedade *+pr*, ou seja, são tratadas como nomes próprios, em sua maioria topônimos. A partir da observação desse fato constata-se que não foi priorizado um estudo detalhado dos compostos toponímicos. Ao tratar da transformação de itens lexicais comuns em topônimos, Dick (1997: 45) afirma que “é a mudança da categoria gramatical dos substantivos comuns a próprios que lhes confere esse *status*”. Consideramos, entre outros motivos, que os compostos toponímicos devem ser tratados de forma distinta dos nomes compostos comuns por pertencerem a categorias distintas e apresentarem comportamento linguístico distinto em alguns aspectos.

Carvalhinhos (2007a: 179) cita Nicolaisen (1980) e afirma que “além de observar a distribuição e frequência dos genéricos (que, uma vez parte de um sintagma toponímico, são mais estáveis que as lexicais comuns), o autor analisa as áreas de ocorrência das variantes”. Nicolaisen (1980) refere-se às variantes do substantivo que funciona como elemento genérico – que indica o tipo de entidade geográfica nomeada. O autor observa que uma palavra ao integrar um sintagma toponímico torna-se mais estável que sua realização como palavra comum do léxico. Consideramos que os grupos nominais livres (formado por palavras comuns), na conversão a compostos toponímicos (nomes próprios), apresentam fixidez sintática.

O elemento genérico que integra a estrutura dos compostos toponímicos coopera para a desambiguação dos compostos que apresentam elemento específico homônimo. O topônimo *Praia da Costa*, por exemplo, apresenta um elemento genérico (*praia*) que passou a integrar o elemento específico. Essa conversão acarreta a necessidade de inclusão de um novo elemento genérico, especificador do acidente geográfico nomeado, como se observa nas frases:

(3) “Aqui no **Bairro Praia da Costa** acontecem os principais eventos da região de Vila Velha como a apresentação de bandas e shows”.

(4) “A criançada também vai poder dar seus pulos na piscina do Happy Ball, no **Shopping Praia da Costa**”.

O composto toponímico pode ser registrado de forma reduzida, como em

(5) “Artesãos Associados - Exposição ‘O artesanato na decoração’ no **Praia da Costa**”.

Na frase (5) há o apagamento do elemento genérico. *Praia da Costa* é a forma reduzida de *shopping Praia da Costa*.

O substantivo *praia* pode também funcionar como elemento genérico na estrutura *Praia da Costa*, como ocorre em

(6) “Rian e Richard realizam show gratuito nas areias da **Praia da Costa**”.

O elemento genérico além de auxiliar na desambiguação de elementos específicos homônimos coopera na classificação do elemento específico como composto ou simples.

A elaboração de um dicionário eletrônico exige a colaboração entre linguistas e informatas. Tratando-se de um dicionário de topônimos há de se levar em conta o fato de que esses itens lexicais, que pertencem à língua geral, também fazem parte da terminologia geográfica. Dessa forma o dicionário eletrônico de compostos toponímicos do português do Brasil serve igualmente aos estudos da terminologia geográfica quanto aos estudos de língua portuguesa de maneira geral.

Estão envolvidos na construção de um léxico computacional de compostos toponímicos os conhecimentos e as vicissitudes de diversas áreas, como as Ciências da Informação, a

Lexicografia, a Toponímia, a Geografia entre outras. A falta de padronização dos nomes geográficos, por exemplo, constitui-se uma dificuldade na averiguação dos topônimos brasileiros. De acordo com Santos (2008: 167)

Os benefícios dos nomes geográficos padronizados são evidentes nos processos da comunicação humana: eles podem impedir a ambigüidade e fornecer esclarecimentos e economia para o comércio [...].

O Brasil, [*sic*] ainda encontra-se carente, no alvorecer do século XXI, de uma fonte segura de desambigüização, uma Comissão Nacional de Nomes Geográficos que norteie as soluções, para as complexas questões da padronização geonímica.

A padronização dos topônimos acarreta benefícios tanto às pesquisas de geógrafos quanto de lexicógrafos, de toponimistas e demais linguistas. A falta de uma comissão responsável pela padronização dos nomes geográficos e de uma instituição que mantenha, de forma centralizada, um banco de dados com essas informações facilita a ocorrência de nomes grafados de formas diversas, assim como a atribuição de diferentes nomes genéricos a um mesmo termo específico. Esses fatores aumentam, desnecessariamente, os esforços investidos na elaboração de léxicos computacionais de topônimos.

1.2 – Objetivos

O objetivo geral da pesquisa consiste na elaboração de um dicionário eletrônico para o português do Brasil de compostos toponímicos do estado do Espírito Santo – especificamente, dos municípios da região da Grande Vitória (Cariacica, Serra, Viana, Vila Velha e Vitória) – a partir da descrição de compostos com as seguintes estruturas: nome + preposição *de* + nome (NdeN), nome + nome (NN) e nome + adjetivo (NAdj) e da formalização da descrição. Para a descrição desses itens lexicais são consideradas as propriedades estruturais, os aspectos morfossintáticos-semânticos e pragmáticos.

Os objetivos específicos são:

- identificar entre as combinações idênticas de palavras as que representam nomes de lugares (topônimos), distinguindo-as dos nomes comuns e

- identificar os grupos nominais livres e os que funcionam como nomes compostos.

Por exemplo: a distinção entre *córrego fundo* e *Córrego Fundo*, nas sentenças (7) e (8).

(7) “O ponto alto da prova foi à [sic] passagem por um **córrego fundo** e estreito [...]”.

(7a) O ponto alto da prova foi a passagem por um **córrego que era/estava fundo** e estreito.

(8) “Os imigrantes italianos se fixaram em **Córrego Fundo** a partir de 1877 [...]”.

(8a) * Os imigrantes italianos se fixaram em **Córrego que era/estava Fundo** a partir de 1877.

Na sentença (7) *córrego fundo* é um grupo nominal livre, formado por nomes comuns. Pode-se constatar a função predicativa do adjetivo em (7) através da inserção do pronome relativo *que* e de um verbo de ligação *ser / estar* (7a), que gera uma sentença aceitável na língua portuguesa. Em (8) a sequência *Córrego Fundo* é um substantivo, especificamente, um topônimo composto. O conjunto *substantivo (Córrego) + adjetivo (Fundo)* designa uma localidade do estado do Espírito Santo. A inserção do pronome relativo *que* e do verbo de ligação *ser / estar* em (8a) gera uma sentença inaceitável (marcada com um asterisco).

1.3 – Justificativa

Um dos grandes desafios do PLN é dar um tratamento formal às línguas naturais, isto é, tratar as línguas naturais de forma que seja possível a “leitura e compreensão” pela máquina. Russel (2004: 767) explica que:

Uma **linguagem formal** é definida como um conjunto (possivelmente infinito) de **cadeias**. [...]. Linguagens formais [...] têm definições matemáticas rígidas. Isso contrasta com as **linguagens naturais**, [...] que não têm nenhuma definição rígida, mas são utilizadas por uma comunidade de falantes (grifo do autor).

Os padrões observados nas línguas naturais são convertidos em linguagem formal, legível para a máquina, a fim de que a interação humano-máquina ocorra de forma mais eficaz.

Segundo Ranchhod (2001: 30) “há combinações de palavras que tanto podem constituir um composto como corresponder a uma construção livre.” Essas construções podem representar nomes comuns ou nomes próprios. Apesar da distinção entre nomes comuns e nomes próprios ser tarefa aparentemente simples, observam-se muitas dificuldades nesse processo, por motivos diversos que, na língua escrita, vão desde a falta de critério no uso de inicial maiúscula como elemento identificador do nome próprio até a possibilidade de uma palavra comum converter-se em nome próprio e vice-versa.

A ambiguidade é um problema a ser resolvido, no que tange à descrição das línguas naturais. De acordo com Laporte (2001: 51), “as ambigüidades lexicais abrangem uma realidade bem mais extensa do que algumas curiosidades e coincidências”. Pode-se afirmar que os topônimos são elementos constituintes dessa realidade.

A necessidade de se determinar se uma sequência de palavras é composicional ou não composicional torna-se indispensável no PLN. Para Ranchhod (2001: 37) “é desejável que o processamento dos compostos não ambíguos seja feito tão cedo quanto possível, a fim de evitar que os seus constituintes sejam analisados como palavras simples”. Os nomes *mal-estar* e *mão-de-obra* são apresentados como exemplos de compostos não ambíguos. Esses itens lexicais são etiquetados como uma unidade lexical, isto é, são associadas a eles informações linguísticas que podem ser lidas pela máquina. “Quanto aos compostos ambíguos, eles aparecem na lista de compostos, mas os seus constituintes fazem igualmente parte da lista de palavras simples do texto (RANCHHOD, 2001: 41)”. O nome composto *dente de alho*, por exemplo, deve ser etiquetado como um único item lexical, porém os itens *dente* e *alho* fazem parte também do dicionário de palavras simples.

Pode-se observar a ocorrência de ambigüidades tanto em nomes comuns quanto em nomes próprios, conforme os exemplos apresentados anteriormente.

(7) “O ponto alto da prova foi à [*sic*] passagem por um **córrego fundo** e estreito [...]”

(8) “Os imigrantes italianos se fixaram em **Córrego Fundo** a partir de 1877.”

Em (7) as palavras destacadas correspondem a um grupo nominal livre, o significado desse conjunto é obtido de suas partes. O substantivo *córrego* refere-se ao acidente geográfico (que corresponde a um curso de águas tênue), enquanto *fundo* realiza-se como adjetivo (que indica um nível abaixo da superfície).

Na frase (8) as palavras destacadas correspondem a um nome composto, que representa o nome de um lugar. Neste caso o significado não pode ser deduzido das partes. *Córrego Fundo* não é um *córrego* nem é *fundo*. *Córrego Fundo* é o nome de uma localidade do município de Serra, no estado do Espírito Santo.

O grupo nominal *córrego Fundo* apresenta ainda, na frase (9), outra realização toponímica, que designa um *córrego capixaba*.

(9) O **córrego Fundo** é um dos afluentes do rio Santa Maria.

Nesta frase, embora o substantivo *córrego* seja semanticamente transparente (indica o tipo de acidente nomeado), o significado deduz-se, como em (8), do conjunto. Nota-se que indiferentemente do nível de água do *córrego*, o topônimo *Fundo* permanece, ou seja, essa denominação fixou-se independentemente da existência ou não do referente. Em (10) o substantivo *córrego* passa a funcionar como elemento específico, fazendo-se necessária a inclusão de um novo termo genérico para designação do topônimo. No exemplo (8), o composto *Córrego Fundo* pode ser considerado uma forma reduzida da estrutura *localidade de Córrego Fundo*, conforme se observa em:

(10) A Fazenda das Palmas fica na **localidade de Córrego Fundo**, a 4 km do Centro de Aracruz.

O mesmo não ocorre na frase (9), em que *córrego* representa o termo genérico, isso é, indica o tipo de acidente geográfico nomeado pelo topônimo *Fundo*.

Embora haja uma coincidência estrutural entre (7), (8) e (9) o que se tem em (8) e (9) são nomes compostos por justaposição, com propriedades semânticas distintas.

O objeto de análise proposto nesta pesquisa são os compostos toponímicos que apresentam elemento genérico (referente a um acidente natural) como um dos elementos formadores dos nomes candidatos a composto. Limitar-nos-emos a análise dos acidentes naturais. Dentre os

acidentes artificiais são incluídos somente os nomes de bairros, localidades e municípios com as características anteriormente descritas, ou seja, os nomes de acidentes artificiais que apresentam como elemento específico um substantivo que designa um acidente natural (morro, córrego, rio etc.).

Em sua abordagem acerca da toponímia, Dick (2007: 146) destaca os seguintes pontos:

6 - conceitualmente, a onomástica toponímica representa a forma estável da língua, os valores culturais, memorialísticos, históricos, geográficos; sobretudo, valores lingüísticos.

[...]

8- em sua constituição gramatical, o *onoma* compreende lexias simples (sb (+) adj (+); vb; lexias compostas (sb-adj +); (vb + vb) (-) e pode apresentar status de lexia complexa (estável).

A autora reforça a importância das pesquisas linguísticas relacionadas à onomástica toponímica para a manutenção da cultura, a preservação da história e os registros geográficos. Dick (2007) ressalta a existência de topônimos simples e topônimos compostos para a nomeação de lugares e a maior ocorrência de estruturas que apresentam um nome como primeiro elemento do composto. Essa recorrência do nome como primeiro item dos compostos toponímicos justifica a escolha do *corpus* (estruturas NN, NdeN e NAdj).

O desenvolvimento da pesquisa proposta torna-se relevante pelo fato de que os estudos em PLN, sobretudo para a elaboração de um dicionário eletrônico de palavras toponímicas do português do Brasil, são ainda incipientes e exigem o empenho de pesquisadores que possam trazer algum tipo de contribuição para esse grande empreendimento.

2 – PRESSUPOSTOS TEÓRICO-METODOLÓGICOS

A análise e a descrição das estruturas toponímicas compostas para o PLN apresentadas baseiam-se na Gramática Transformacional de Zellig Harris, bem como nos pressupostos teórico-metodológicos do Léxico-Gramática desenvolvidos por Maurice Gross (1975).

2.1 – A Gramática Transformacional De Zellig Harris

Para a descrição das palavras compostas toponímicas parte-se dos pressupostos teóricos da Gramática Transformacional de Zellig Harris. Segundo Gross (1970: 185), na análise transformacional de Harris (1952, 1954, 1964) a ideia fundamental é que todas as frases³ de uma língua natural podem ser descritas a partir da frase simples e das operações/transformações com essas construções.

Harris (1964) considera um primeiro conjunto de geradores e operações de natureza finita que permitem gerar frases núcleo. As frases núcleo, por sua vez, formam um segundo grupo de geradores que geram frases complexas mediante transformações unitárias (que operam sobre uma frase) e binárias (que operam sobre uma dupla de frases).

Harris (1968) propõe a formação de classes de substituição de morfemas únicos, ou seja, um conjunto de morfemas de determinada língua que podem ser substituídos mutuamente dentro de certa ambiência linguística, que apresentam a mesma seleção. Os morfemas que apresentam distribuição ligeiramente distinta são inseridos em uma mesma classe, desde que as diferenças distribucionais entre os seus ambientes correspondam às diferenças distribucionais entre os morfemas.

³ O termo *frase* é utilizado como tradução do texto original, onde se lê: *phrase*, e não como índice de filiação teórica. Os termos *frase*, *sentença* e *cláusula* são usados indistintamente.

O autor considera a realização dos substantivos ingleses *house* e *poem* em ambiências distintas, nas frases: “*That’s a beautiful _____*” e “*I’m writing a whole _____ this time*” (HARRIS 1968: 26, grifo do autor). Na primeira frase não há problema em permutar *poem* por *house*,

*That’s a beautiful **poem**.*

*That’s a beautiful **house**.*

mas o mesmo não ocorre na segunda sentença.

*I’m writing a whole **poem** this time.*

I’m writing a whole **house this time.*

Dessa forma,

Si *poem* et *house* ne diffèrent distributionnellement que par le fait que *poem* apparaît avec *write* et *house* avec *wire*, et par de différences du meme ordre, et si *write* et *wire* à leur tour ne different que parce que *write* apparaît avec *poem* et *wire* avec *house* et par des différences du meme ordre, alors nous mettrons *poem* dans une classe em compagnie de *house*, et, simultanément, *write* dans une classe em compagnie de e *wite*.

(HARRIS, 1968: 27, grifo do autor).

Se a diferença distribucional entre os nomes *poem* e *house*, na sentença *I’m writing a whole _____ this time* é da mesma ordem que a diferença distribucional entre os verbos *write* e *wire*, esses nomes podem, então, ser incluídos na mesma classe, assim como os verbos *write* e *wire* podem ser considerados participantes da mesma classe.

Para Harris (1951: 368-369)

Once we have a number of comparisons available as bases for settings up segmentations or classifications, we select those comparisons which apply to large numbers of elements. [...] This selection, of course, derives not from nature of the comparisons but from our purposes [...].

As a result of these operations, we not only obtain initial elements, but are also able to define new sets of elements as classes or combinations [...] of old ones.

O método distribucional consiste, inicialmente, em identificar e agrupar/classificar os morfemas de determinada língua de acordo com suas similaridades. É possível, a partir dessas classes, se estabelecerem novas classes. Pode-se dizer dessa hierarquia, quanto às entradas, que as classes superiores apresentam elementos mais livres quanto à ocorrência.

Acerca da hierarquia dos elementos linguísticos, Harris afirma que:

La pertinence de la hiérarchie des expansions de mots, qui était organisée en un chaîne ascendante d'équivalances [...] le fait qu'elle ouvrait la voie à une méthode générale de décomposition des phases quelconques em phases élémentaires et donc à um système de décomposition transformationnel.

(HARRIS, 1990: 11).

Nesse sistema de decomposição transformacional uma frase menor é considerada a base das expansões, o elemento mínimo a partir do qual as expansões se dão. Gross (1990b) considera a dificuldade de delimitação do que seja uma *sentença elementar*, geralmente definida pelos elementos que a constituem (sujeito + verbo + complementos essenciais) e representada pelo autor através da seguinte fórmula: $P =: N_o V W$, (onde P = frase, N_o = sujeito, V = verbo e W = complementos). Essa dificuldade consiste em determinar quais complementos devem ser tidos como essenciais a cada verbo a que se associa. Esse fator é de extrema relevância para a classificação dos verbos. Por exemplo, as seguintes acepções do verbo *deseambarcar*:

- (1) “Por fim, os ingleses desembarcaram um carregamento variado de foguetes ‘Rapier’, sistemas de mísseis, barris de combustível, caixas de munições e água potável.” (desembarcar = tirar algo de embarcação)
- (2) “Delegação gremista desembarcou em Porto Alegre.” (desembarcar = chegar; aportar)

O verbo *deseambarcar*, em (1), seleciona como complemento um objeto direto (*a carga*), já em (2) o complemento selecionado não apresenta a mesma natureza.

Acerca da definição de *sentença elementar* e da classificação dos complementos *essenciais*, Gross (1983: 108-109) aponta os critérios adotados pela tradição gramatical e pela gramática gerativa e propõe sua aplicação adaptada aos verbos franceses. As etapas observadas são:

Step 1) all obligatory complements are retained.

Step 2) only three forms of (optional) complements are being considered: direct complements, and complements with the preposition *à* or *de*. The rationale behind this choice is that these complements are linked to the pre-verbal or so called clitic pronouns.

Step 3) the notion of object complement is defined in terms of the shape of the associated interrogative pronouns. Thus, direct objects correspond to the pronouns *qui* (*who*) and *que* (*what*).

O autor aplica os critérios mencionados a algumas sentenças e observa que, em determinadas ambiências com a preposição locativa, a forma do pronome interrogativo utilizada na identificação do complemento não é *quem* ou *que*, mas *onde*. Quanto à classificação dos complementos locativos em *essencial* ou *não essencial*, Gross observa que nas estruturas N_oVN_I há de admitir-se que o segundo complemento é *não essencial*. Porém, a descrição dos verbos franceses *ranger* (*ordenar/classificar*) e *changer* (*mudar/transferir*) é mais apropriada com dois complementos essenciais. Logo, a aplicação sistemática dos critérios acima a casos concretos mostra a necessidade de adaptá-los.

Thus, by step 1 above, we included obligatory complements as essential complements. For example in

Max compte sur Bob
(Max counts on Bob)

The obligatory complement *sur Bob* is incorporated in our description (e.g. $N_oV\ sur\ N_I$). But, when independently we describe the sentence

Cette remarque a agi sur Bob
(This remark had an effect on Bob)

We have to classify *agir* (*to have an effect*) as intransitive (e.g. with structure N_oV), since the optional prepositional complement is excluded by step 2.

(GROSS, 1983: 110, grifo do autor).

A distinção entre o caráter *essencial* e *não essencial* do complemento com preposição fica patente nos exemplos apresentados por Gross.

A *transformação*, como é tratada na gramática transformacional de Harris, de acordo com Dubois (2001: 597-598), define-se da seguinte forma:

se duas ou mais construções (ou seqüências de construções) que contém as mesmas *n* classes aparecessem com as mesmas *n-duplas* de membros dessas classes, dir-se-ia que as construções são transformações uma da outra e que cada qual pode derivar da outra por uma transformação especial.

A distribuição dos elementos na construção é um fator fundamental nesse processo, para exemplificar as transformações o autor apresenta as sentenças: “*O carro atropelou o transeunte* e *O transeunte foi atropelado pelo carro*”. Essa mudança constitui um exemplo de apassivação, que consiste numa transformação unitária em que “as construções $SN_1 + V_1 SN_2$ (sintagma nominal seguido de verbo transitivo e de sintagma nominal) e $SN_2 + ser + V_{part} + por SN_1$, contém os mesmo grupos tríplexes, SN, V, SN [...]”.

O sinal de igualdade (=) é utilizado para especificar a estrutura, ou seja, para marcar a relação de transformação entre duas sentenças. Em uma transformação unitária, como a apassivação, por exemplo, teríamos:

(3) As bailarinas executaram a coreografia.

[passiva] = (4) A coreografia foi executada pelas bailarinas.

A relativização é um exemplo de transformação binária, em que se pode verificar uma dupla de sentenças e a inserção de um pronome relativo:

(3) As bailarinas executaram a coreografia.

(5) A coreografia foi aplaudida pela plateia.

[relativização] = (6) As bailarinas executaram a coreografia, que foi aplaudida pela plateia.

Para Gross (1991), as transformações são as relações de equivalência entre sentenças que não alteram o significado básico. Operações como a *voz passiva*, a *introdução de modais*, a *introdução da negação* e a *relativização* são apresentadas como exemplos dessas transformações. O autor avalia que o método da construção de classes – de equivalência para sentenças elementares – poderia ser aplicado a uma língua natural, o que resultaria em uma cobertura das estruturas complexas possibilitando assim a análise sintática de textos por computador. Os textos seriam reduzidos então aos jogos de unidades elementares de significados (Harris, 1982), permitindo que a grande variedade de informação fosse reduzida a um número mais tratável de formas.

A Gramática Transformacional de Harris é adequada à necessidade de descrição e formalização mais rígida da língua no PLN, associada a outros pressupostos teóricos que tiveram ou não suas origens nos seus estudos.

2.2 – O Léxico-Gramática de Maurice Gross

A partir das pesquisas de Harris, outros trabalhos dedicados à formalização das línguas naturais foram elaborados, dentre eles o Léxico-Gramática desenvolvido por Maurice Gross

(1975) e os pesquisadores do *Laboratoire d'Automatique Documentaire et Linguistique* (LADL) da Universidade Paris 7.

O Léxico-Gramática é constituído das sentenças de uma língua (GROSS, 1984). A unidade básica de informação considerada é a sentença simples (sujeito-verbo-complementos), ao invés da palavra. As entradas no dicionário são a partir dessas sentenças simples, com a descrição distribucional e as propriedades transformacionais. A construção do léxico-gramática do francês proporcionou acumulação de informação linguística a ser utilizada no PLN. Percebe-se a importância dos estudos em PLN no desenvolvimento da descrição de uma língua natural, bem como para o desenvolvimento dos sistemas informáticos em que são utilizados esses dados.

De acordo com Gross (1991) o axioma básico da teoria do Léxico-Gramática é: “a sentença é a unidade linguística de significado”, logo, as palavras não constituem as unidades de significado. Esse fato fica patente no que tange aos nomes compostos, em que o significado se deduz do todo e não da soma do significado das partes. Por exemplo, em:

(7) “O Maracanã tem um **campo grande**, é difícil qualquer time ficar fechado ali 90 minutos”,

os itens em negrito têm significado literal, ou seja, o substantivo *campo* designa um terreno plano e o adjetivo *grande* indica um tamanho avantajado. O mesmo não se dá em:

(8) “Somente neste ano já foram retirados de circulação mais de 100 quilos de entorpecentes das rodovias federais capixabas [...]. Os detidos e a droga foram encaminhados ao DPJ de **Campo Grande** [...]”.

O significado dos itens destacados não é deduzido das partes, não se trata de um *terreno plano de tamanho avantajado*, mas de uma palavra composta. O topônimo *Campo Grande* designa um acidente artificial (bairro) do município capixaba de Cariacica.

Gross (1991) explica que o número de palavras compostas no léxico de uma língua é superior ao número de palavras simples, de acordo com o levantamento feito em diversas línguas, entre elas o português. O fato do léxico de uma língua comportar maior número de palavras compostas em relação às simples e as palavras compostas serem formadas por mais de uma

palavra simples, o que possibilita a ocorrência de ambiguidade, aponta para relevância das pesquisas acerca de palavras compostas para o PLN.

Cabe ressaltar, conforme explica Vale (2001: 69), que embora o Léxico-Gramática tenha sua origem na visão de transformação proposta por Harris e o uso do termo “transformacional” seja muitas vezes associado à gramática gerativa proposta por Chomsky, a abordagem de Maurice Gross se distingue da abordagem de Chomsky. De acordo com o autor, Gross (1975) critica o fato de a gramática gerativa analisar números limitados de exemplos, não levar em conta toda produtividade dos exemplos analisados e escolher fatos ao acaso dentro de um conjunto de fenômenos sem preocupar-se com a sua representatividade. Desta forma, as ocorrências linguísticas não são observadas conforme sua realização, sua recorrência ou não no uso.

De acordo com Laporte (2008: 37), entre os anos de 1970 e 1980 as tabelas do léxico-gramática “não se utilizam em nenhum momento de um corpus, limitando-se assim a uma lingüística exclusivamente introspectiva e manipuladora”, fato justificado pela falta de um número considerável de textos disponíveis em meios eletrônicos que pudessem ser utilizados na elaboração dessas tabelas. O autor explica que nesse período foram tomadas algumas precauções quanto à elaboração de exemplos, como: encontros frequentes dos grupos de linguistas a fim de averiguarem os julgamentos de aceitabilidade e as análises efetuadas e realizarem o questionamento da pertinência dos critérios utilizados para atestar as propriedades sintático-semânticas estudadas.

Gross (1984: 378-379) afirma que os julgamentos de aceitabilidade devem ser de caráter binário, “sequences are accepted as sequences or not. In the grammar we do not use more than the two values: acceptable and ‘*’ (i.e. unacceptable)”. O autor explica que nas discussões pode-se introduzir o sinal de interrogação, o que aponta para a existência de quatro valores (aceitável, inaceitável, aceitabilidade duvidosa e inaceitabilidade duvidosa) ao invés de dois. Gross opta pelo uso de apenas dois valores e explica que o julgamento de aceitabilidade deve ser binário e que o uso da interrogação é característica dos comentários informais acerca das descrições formais.

A avaliação da aceitabilidade de uma sequência de palavras pode ser comparada à execução de uma experiência sintática. Laporte (2008: 34) afirma que “uma experiência, uma medida, só tem um interesse científico se são reprodutíveis, quer dizer, se um experimentador que as

praticar de novo obtiver os mesmos resultados.” Assim, a exigência da reprodutibilidade é crucial como em toda a experiência científica, o que inclui as descrições linguísticas.

Segundo Laporte, a partir da década de 1990 a situação muda, com a larga utilização de exemplos verificados em *corpus* devido aos avanços tecnológicos, o que não implicou abandono das precauções tomadas no período anterior.

Laporte (2008) explica que a descrição sintático-semântica do Léxico-Gramática utiliza-se tanto da abordagem probabilística como da abordagem simbolista. De acordo com autor (2008: 49), “as tradições aparentemente opostas da linguística introspectiva e da linguística de corpus são, pois, complementares e de natureza a se combinar para favorecer o sucesso de tal empresa, sendo contra-produtivo excluir um ou outro”. Pode-se incluir no âmbito da linguística introspectiva os julgamentos de aceitabilidade emitidos quanto aos exemplos construídos e no âmbito da linguística de *corpus* a verificação da ocorrência dos itens a ser descritos, sua maior ou menor ocorrência em textos da língua analisada. Essa abordagem híbrida (simbólica/ probabilista) propicia a obtenção de resultados mais eficazes.

2.3 – Metodologia

A análise proposta, das palavras toponímicas compostas por justaposição, segue a metodologia do Léxico-Gramática proposta por Maurice Gross (1975). A unidade básica de informação considerada é a sentença simples (sujeito-verbo-complementos), logo, a descrição de uma palavra é feita a partir da realização desta em uma frase simples.

O *corpus* selecionado é formado de palavras compostas toponímicas colhidas, a princípio, em mapas do IBGE (Instituto Brasileiro de Geografia e Estatística) e outras cartas. Em seguida verificou-se a ocorrência desses itens lexicais em textos jornalísticos (Folha do Espírito Santo, Gazeta, Jornal Vitória, Século diário – versões *on line*), decretos e leis municipais e outros documentos. Essas sentenças servem de base para a aplicação das transformações. São utilizadas também frases construídas.

Vale ressaltar a afirmação de Laporte (2009: 72) acerca dos exemplos construídos, de que a observação de *corpus* não atesta inaceitabilidade “for instance, the absence of the phrase *aborder de* in a French corpus of 820 000 words does not prove that this expression is not in use, and in fact, it is” (grifo do autor). A não ocorrência de uma palavra em *corpus* não é indício, pois, de inaceitabilidade. O autor afirma que “um dos méritos do léxico-gramática é o de ter-se dotado desde o início de um arsenal de precauções metodológicas contra os riscos próprios a atividade de construção de exemplos” (LAPORTE, 2008: 35). Esses cuidados (como o controle dos julgamentos de aceitabilidade por outros linguistas falantes nativos, verificação da reprodutibilidade dos critérios adotados etc.) são observados no desenvolvimento desta pesquisa.

A identificação dos topônimos capixabas selecionados, quanto ao seu caráter não-composicional, insere-se numa visão sincrônica da língua portuguesa do Brasil. Desse modo, optamos pela não indicação da origem desses nomes ou do fato que motivou a nomeação. De acordo com Dick (2000), “nomes orográficos ou geomorfológicos, empregados na microtoponímia urbana, mostram-se, geralmente, desvinculados do motivo gerador da denominação [...]”. Percebe-se que a motivação da nomeação de montes e montanhas, assim como de outros tipos de acidentes geográficos naturais, pode com o passar do tempo não ser mais identificada no nome atribuído. Esse fato contribui para que esses itens lexicais sejam semanticamente opacos, o que torna o resgate da origem dos topônimos de importância relevante para o resgate da história da comunidade inserida em determinado local em uma época precisa, porém, essas informações de cunho etimológico são de pouca utilidade para descrição linguística dos topônimos para a elaboração de um dicionário eletrônico, nos moldes propostos nesta pesquisa.

Utilizamos também ferramentas computacionais como o Unitex - sistema com tecnologias criadas pelo *Laboratoire d'Automatique Documentaire et Linguistique* (LADL) da Universidade Marne-la-Vallée.

Cabe ressaltar a diferença entre os sistemas Unitex e Intex. Segundo Muniz (2004) o Intex foi criado em 1992 no LADL. O Unitex foi criado no mesmo laboratório, dez anos depois, em 2002.

O Unitex é um sistema baseado no uso de grandes dicionários lexicais. Ele pode ser usado para analisar textos de muitos milhões de palavras. Inclui vários dicionários e gramáticas embutidos representados como autômatos de estados finitos, porém, o

usuário pode adicionar seus próprios dicionários e gramáticas. Estas ferramentas são aplicadas ao texto para localizar padrões léxicos e sintáticos, gerar dicionários lexicais, remover ambigüidades e etiquetar palavras simples como também expressões complexas. Ele pode ser utilizado por lingüistas para análise de corpus, mas também pode ser visto como um sistema de recuperação de informação. (MUNIZ, 2004: 47)

O Intex possui as mesmas funcionalidades do Unitex, contudo, não pode ser modificado pelo usuário.

3 – O NOME COMPOSTO

3.1 – Na Gramática Tradicional

Tradicionalmente a conceituação de composição faz-se através da associação de critério fonológico, morfológico e semântico. Said Ali (1964) e Melo (1978) apresentam o conceito de palavra composta como a combinação de duas ou mais palavras (semanticamente autônomas) com a qual se nomeia algum conceito novo. Essa combinação apresenta um sentido diferenciado do sentido atribuído às partes que a compõe. A concepção de *nome composto* desses autores prioriza o aspecto semântico da composição.

Melo (1978) propõe a divisão dos compostos em *perfeitos* e *imperfeitos*, divisão que se baseia na conservação ou não da independência fonética dos elementos constituintes da palavra composta. A composição *perfeita* seria aquela em que o primeiro elemento perde a tonicidade e o composto apresenta um único acento tônico, ao passo que na composição *imperfeita* os elementos constituintes conservam a sua independência fonética. Pouca importância é atribuída à classificação dos compostos em *aglutinados* e *justapostos*.

Said Ali (1964) observa que muitas palavras compostas eram, a princípio, grupos sintáticos que passaram por um processo de modificação que culminou na fusão semântica dos elementos desses grupos sintáticos, de forma que o novo significado é atribuído ao todo e não aos elementos formadores do composto. Esse processo pode ocorrer em um longo espaço de tempo ou acontecer de forma rápida, assim, há grupos de palavras que se encontram em transição entre um *grupo de itens lexicais simples* e uma *palavra composta*. O fato de algumas construções estarem em uma fase de transição gera dificuldades na identificação desses elementos como grupos nominais livres ou palavras compostas.

Para Bechara (2009: 355), “COMPOSTA é a palavras que possui mais de um radical [...]”, autor afirma que, “a COMPOSIÇÃO consiste na criação de uma palavra nova de significado único e constante, sempre e somente por meio de dois radicais relacionados entre si” (grifo do

autor). Somente as palavras formadas a partir da relação entre dois elementos são consideradas como composição. As estruturas que apresentam mais de dois elementos formadores são consideradas *lexias complexas* ou *sinapsias*. Quanto à distinção entre composição e a sinapsia o autor admite “a dificuldade de se traçarem limites rígidos entre os dois processos” (BECHARA, 2009: 352).

Cunha e Cintra (2007: 119) afirmam que a composição “consiste em formar uma nova palavra pela união de dois ou mais radicais. A palavra composta representa sempre uma ideia única e autônoma, muitas vezes dissociada das noções expressas pelos seus componentes”. Os autores apresentam o conceito de composição nos mesmos moldes que Bechara (2009), contudo, consideram que o processo de formação de palavras por composição pode relacionar mais de dois radicais.

Observa-se a utilização do critério morfológico, na exigência da presença mais de um radical, especificamente dois radicais, na constituição do composto. O critério semântico também é observado na atribuição de “único e constante” significado da composição.

De acordo com Rocha Lima (2005: 225),

Dá-se a *composição* [...] quando se juntam dois ou mais elementos vocabulares de significação própria, para darem ideia de um novo ser ou objeto. O que caracteriza, em última análise, a *composição* é, além da *unidade de significação*, a existência de *mais de um radical* (grifo do autor).

Os compostos são divididos em dois tipos: os justapostos e os aglutinados. Essa classificação baseia-se no critério fonológico, na manutenção da integridade ou não da forma e da acentuação dos itens lexicais que formam esses compostos. Ressalta-se o fato de que a presença do hífen como critério de identificação do composto é ineficaz. Tanto Rocha Lima (2005: 227) como Said Ali (1964: 259) citam Bréal ao afirmar que o sentimento da composição “tem seu critério no espírito”. O critério semântico é enfatizado na identificação dos compostos, atividade considerada como um fazer subjetivo do analista.

De acordo com Cunha e Cintra (2007: 119) a distinção entre os compostos aglutinados e os compostos justapostos, na escrita, pode ser percebida através do uso de hífen entre os elementos formadores dos compostos aglutinados e o registro gráfico em um só vocábulo no caso dos compostos aglutinados. Todos os exemplos de compostos justapostos apresentados no capítulo dedicado à *derivação e composição* apresentam hífen, porém, os autores fazem a

seguinte observação: “Reitere-se que o emprego do hífen é uma simples convenção ortográfica. Nem sempre os elementos justapostos vêm ligados por ele”. De fato, muitos nomes compostos apresentam a estrutura idêntica aos grupos nominais livres sem que a fixidez sintática entre os elementos do composto seja marcada por sinal gráfico.

Para Melo (1978: 55), os elementos que formam os compostos imperfeitos (também chamados de justapostos) devem ser unidos por hífen. A identificação desse tipo de palavra seria uma tarefa de execução menos laboriosa se, de fato, a utilização do hífen nas palavras compostas por justaposição fosse condição imprescindível para o funcionamento desses itens lexicais como um composto. Porém, o uso de hífen entre os elementos formadores de um nome composto não consiste numa regra geral e coerente. A presença do hífen não serve como critério de distinção entre os grupos nominais livres e as palavras compostas já que a ortografia destas não é uniforme. Algumas palavras compostas são grafadas reunindo os elementos em um só vocábulo, outras se escrevem com hífen entre os elementos constituintes e outras têm seus termos grafados separadamente, à semelhança dos grupos nominais livres.

A Gramática Tradicional propõe a utilização dos critérios morfológico, semântico e fonológico na identificação dos compostos ao considerar que a composição é uma combinação que apresenta mais de um radical e um sentido diferenciado do sentido das partes que a compõe podendo ser *imperfeita* ou *perfeita*, *justaposta* ou *aglutinada*, dependendo da manutenção ou não do acento tônico.

Gross (1971) explica que os conceitos da gramática tradicional geralmente associam forma e sentido. A associação entre forma e sentido não é vista como um problema. O que se questiona é a natureza dessa relação. A base semântica dessas noções é vista como um fator que dificulta a sua aplicação. As colocações da gramática tradicional não são desconsideradas. Admite-se a importância de seus conceitos semânticos, contudo, a utilização dos conceitos e critérios adotados carece de explicações. O tratamento dispensado ao nome composto está inserido nessas considerações acerca da abordagem da gramática tradicional.

3.2 – Na Linguística

Sandmann (1988) apresenta os modelos de composição mais produtivos e procura fazer a distinção entre os compostos e os grupos sintáticos. Essa distinção é considerada mais difícil no que diz respeito aos nomes compostos (determinativos) de estrutura adjetivo mais nome (AjdN), nome mais adjetivo (NAdj) e nome mais preposição *de* mais nome (NdeN). Na diferenciação dos nomes compostos dos grupos nominais livres são utilizados os critérios: fonológico, morfológico, sintático e semântico. O autor (1991b: 60-61) considera palavra complexa o item lexical formado por mais de uma raiz e destaca a relação entre os elementos formadores dos substantivos e os adjetivos compostos, que pode ser de coordenação ou de subordinação. A relação de coordenação ou de subordinação aponta para a relevância da identificação do elemento que funciona como núcleo da palavra (aquele que determina o gênero da palavra) e do seu adjunto. A discussão apresentada pelo autor baseia-se em aspectos morfossintáticos, ou seja, ela está amparada mais em critérios formais que na subjetividade do analista.

A ordem dos compostos formados por dois substantivos também é observada. Os substantivos de estrutura NN, como *carro-bomba*, em que N1 – *carro* – é o núcleo e o N2 – *bomba* – o adjunto, são chamados de *vernáculos*. Nesses casos N2 está subordinado a N1. Pode-se dizer que um *carro-bomba* é um *carro no qual se instala uma bomba*, sendo assim, não é possível a inversão dos elementos constituintes. Os compostos NN, em que o adjunto antecede ao núcleo, são chamados *não-vernáculos* e considerados menos produtivos na língua portuguesa que os compostos *vernáculos*.

Quanto aos compostos de estrutura NdeN, Sandmann (1991b) afirma que tanto os nomes compostos quanto os grupos nominais livres apresentam a mesma estrutura, logo, as considerações sintáticas – a identificação do nome que funciona como núcleo e do adjunto – são de pouca relevância. Por exemplo, na sentença:

(1) “Um israelense foi condenado a nove meses de prisão por atirar uma **cabeça de porco** em uma mesquita de Tel Aviv em agosto de 2005”,

os itens destacados constituem um grupo nominal livre, em que N1 (*cabeça*) é o núcleo, o substantivo determinado e o N2 (*porco*) é o adjunto, o determinante. Na sentença:

- (2) “Moravam [...] numa **cabeça de porco**: quarto e banheiro [...], onde se amontoavam mais de vinte famílias – bairro da Glória.”;

o mesmo ocorre, N1 (*cabeça*) é o núcleo e N2 (*porco*) o adjunto. A distinção dá-se quanto ao aspecto semântico. Na sentença (1), *cabeça de porco* significa *a parte superior do corpo de um animal, de um suíno*, enquanto na sentença (2) *cabeça de porco* significa *cortiço, moradia coletiva, de pessoas pobres*.

Sandmann (1991a: 33) admite que a língua portuguesa não possui “um molde, com características fonológicas, morfológicas e sintáticas fixas, de acordo com a qual se formam compostos que se distingam [...] facilmente dos grupos sintáticos correspondentes”. Considerando que a formação de palavras compostas carece de um padrão determinado, a distinção desses nomes dos grupos nominais que apresentam a mesma estrutura torna-se uma tarefa mais laboriosa.

O autor explica que novas palavras que são formadas e as que se estabelecem no léxico sofrem um “processo de lexicalização, ou idiomatização, processo que também pode ser chamado de desmotivação, isto é, o todo não é mais a simples soma das partes” (SANDMANN, 1991a: 29-30). A lexicalização é apresentada como um processo gradual. À medida que o conjunto de palavras que formam o composto distancia-se do comportamento semântico, fonológico e morfológico que individualmente apresentam na língua, maior será a sua fixidez e caracterização como nova unidade lexical.

Dessa forma, o nome *caixa* na sentença (3) aceita, por exemplo, a variação de grau:

- (3) Ana ficou emocionada quando Luiz tirou de sua pasta uma **caixa** preta de veludo com uma pulseira de diamantes.

- (3a) Ana ficou emocionada quando Luiz tirou de sua pasta uma **caixinha** preta de veludo com uma pulseira de diamantes.

O mesmo não ocorre na sentença (4).

- (4) “Autoridades encontram **caixa preta** do Airbus acidentado nas Comores.”

- (4a) *Autoridades encontram **caixinha preta** do Airbus acidentado nas Comores.

A inaceitabilidade da variação em grau do substantivo *caixa* em (4a) aponta para um comportamento morfológico que permite estabelecer a distinção entre *caixa preta* como grupo nominal livre e a forma lexicalizada *caixa preta* – nome composto, através de um critério menos dependente da subjetividade do analista.

Cabe ressaltar que, na realidade, formas perfeitamente lexicalizadas ainda podem sofrer variações morfológicas regulares, por exemplo: *colher de sopa/colherzinha de sopa/colherzona de sopa*.

O conceito tradicional de composição, para Basílio (2000: 14), deixa “duas áreas de indefinição: por um lado, a questão das construções existentes, em oposição aos processos de formação; por outro, a caracterização do composto enquanto palavra”. Dessa maneira, a definição tradicional de composição leva à classificação, dentre as palavras compostas, de construções que representam um processo de composição pouco produtivo na língua portuguesa. O conceito tradicional de composição também não apresenta as propriedades estruturais que permitem o estabelecimento da diferença entre as palavras compostas e os grupos nominais livres. Outra questão apresentada pela autora é dificuldade da identificação das sequências NAdj como compostos morfológicos ou compostos lexicais.

Segundo Basílio (2004: 14) a palavra, do ponto de vista morfológico, estrutura-se de forma específica: “seus elementos componentes, ou formativos, apresentam ordem fixa e são rigidamente ligados uns aos outros, não permitindo qualquer mudança de posição ou interferência de outros elementos” são apresentados como exemplo de palavras: *guarda-chuva, encaixotado, narração*. A autora afirma que *guarda-muita-chuva, encaixonado, çãonarra* não podem ser consideradas palavras da língua portuguesa. O fato do conjunto de itens lexicais *guarda-chuva* não admitir a inserção do intensificador *muito* atesta a fixidez sintática desse composto.

Dentre os problemas remanescentes em torno da conceituação de palavra, insere-se: “a questão das palavras compostas, [...] a situação de nomes próprios de cidade e instituições, e assim por diante” (BASÍLIO, 2004: 14), o que inclui os compostos toponímicos.

Basílio (2007) define a composição como um processo de formação de palavra caracterizado pela estrutura. As palavras oriundas desse processo são formadas por duas bases, presas ou não. A função dos compostos é associada à necessidade de designações mais particularizadas,

logo, a composição apresenta função prioritariamente lexical. Cada elemento, no processo de formação de palavras por composição, tem uma função que é determinada pela estrutura. A autora observa que nos compostos de estrutura NN, N1 é o núcleo do substantivo, responsável pela classificação geral e N2 funciona como modificador (especificador ou elemento particularizante de N1). Outras formas de composição (NAdj e VN) também são analisadas. O aspecto semântico, do distanciamento do significado das partes que formam o nome composto e da aquisição de um novo significado atribuído ao conjunto (ao composto), também é observado por Basílio (2007) e relacionado à função desempenhada pela composição no léxico.

Monteiro (1987: 164) conceitua composto como “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)”. O autor apresenta dois compostos toponímicos (*Porto Alegre* e *Mato Grosso*) como exemplos de compostos que apresentam seus componentes ‘soltos’ (não ligados por hífen) e aponta a dificuldade de estabelecer a distinção entre *vocábulo composto* e uma *locução*.

As características morfossintáticas dos compostos são resumidas em cinco itens, dos quais o autor considera os três últimos não muito confiáveis. São eles:

1. Flexão exclusiva do último componente.
2. Sufixação relacionada a o 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: 164).

Os aspectos morfológicos dos compostos são considerados mais confiáveis que a distribuição sintática dos elementos formadores dos compostos.

O autor explica que palavras como *marechal*, *pedestal* e *vendaval*, classificadas por alguns estudiosos como palavras compostas não passam de vocábulos simples. Para PLN os chamados compostos aglutinados são considerados palavras simples.

3.3 – No Léxico-Gramática

O Léxico-Gramática é o suporte teórico metodológico para diversas pesquisas no processamento de linguagem natural entre elas os trabalhos relativos à elaboração de dicionários eletrônicos do sistema Unitex, que exige do linguista uma descrição detalhada dos itens lexicais que dão entrada nesses dicionários de maneira codificada.

Os nomes compostos são constituídos de palavras simples e obedecem a determinadas regras gramaticais, por exemplo, nos nomes compostos formados por dois substantivos um dos substantivos funcionam como núcleo da palavra e o outro como adjunto. O aumento do número de nomes compostos é associado ao crescimento dos vocabulários técnicos. Segundo Gross (1986a), os nomes compostos apresentam, por natureza, significado não composicional e podem ser descritos em termos da sequência de suas categorias. O autor cita como exemplo de nome composto inglês *board of governors* (grupo que gerencia uma entidade pública ou organização sem fins lucrativos / conselho administrativo) e admite a proximidade entre o nome composto *board of governors* e uma forma livre com a mesma estrutura. O fato do N2 (*governors*) poder ser substituído por limitado número de nomes (*directors* ou *regents*) indica que a sequência *board of governors* constitui um nome composto. A substituição do N1 do topônimo **Monte Mestre Álvaro** (município de Serra – ES) ocorre de forma restrita, admitindo-se a permuta de N1 (*monte*) somente com o substantivo *morro*. Gross (1986a: 4) explica que o significado dos elementos que formam as palavras compostas é não composicional, isso é, há um significado legal ou institucional que os componentes não apresentam claramente se analisados isoladamente. Por exemplo, o significado do substantivo *cachorro-quente* não pode ser deduzido da soma do significado dos elementos que formam o composto, *cachorro-quente* não é um *animal*, *um cão que conserva o calor, que está aquecido*, mas um *pão com salsicha e molho*. Do mesmo modo o adjetivo *solteira* no topônimo *Ilha Solteira* não indica o estado civil da *ilha*, não é próprio de uma porção de terra (ilha) casar-se ou manter-se solteira.

Segundo Constant (2003), as unidades complexas que contêm pelo menos um separador são chamadas palavras compostas. Certos separadores são bastante produtivos em termos de palavras compostas como o hífen “-”, mas existem milhares de outras palavras compostas que contêm outros separadores como, por exemplo, o espaço. Na quase totalidade dos casos, essas

palavras compostas são formadas de palavras simples e apresentam a mesma estrutura que os grupos nominais livres. A distinção entre grupos nominais livres e palavras compostas se faz pela observação das restrições sintático–semânticas que favorecem a fixidez da sequência. A utilização concomitante dos diferentes tipos de critérios coopera para uma descrição mais abrangente e adequada dos fatos linguísticos.

A identificação de palavras compostas do português do Brasil para a elaboração de um dicionário eletrônico é imprescindível. Para Monteleone (2003: 78), as diferenças formais e semânticas que existem entre palavras simples e palavras compostas justificam a separação entre os dois tipos de entradas e, por conseguinte a criação de duas bases de dados distintas. Ou seja, há a necessidade de um dicionário eletrônico de palavras simples e outro dicionário, específico, para a inserção das palavras compostas. O dicionário eletrônico de palavras compostas apresenta a formalização e codificação da descrição, não apenas dos elementos (simples) que formam o composto, mas o comportamento destes como uma unidade lexical. Essas bases de dados (os dicionários eletrônicos) são de crucial importância no processamento automático de textos.

De acordo com Savary (2000) as palavras compostas são, para fins computacionais, identificadas pelo critério da não-composicionalidade, ou seja, as suas propriedades sintáticas, semânticas e/ou distribucionais não podem ser deduzidas das partes que formam o nome composto, mas do todo. Quanto ao aspecto ortográfico, a existência de um separador (hífen, espaço em branco e outros) no interior da palavra é imprescindível para o reconhecimento de um grupo de elementos como palavra composta. Dessa forma as palavras formadas por mais de um radical aglutinados (que não apresentam um separador no seu interior, por exemplo, *planalto*, *aguardente*) são consideradas na elaboração de dicionários eletrônicos como palavras simples.

4 – O NOME PRÓPRIO

O avanço das pesquisas em linguística para o desenvolvimento de um dicionário eletrônico do português do Brasil é ainda incipiente e esbarra-se em alguns entraves, próprios da estrutura da língua em uso. As propriedades que delimitam o nome próprio é um dos problemas encontrados no estudo das línguas naturais e conseqüentemente no PLN.

4.1 – As abordagens acerca do nome próprio

Os estudos acerca do nome próprio apresentam diversas abordagens, das quais destacamos as gramáticas normativas de Said Ali (1964), Rocha Lima (2005) e Cunha e Cintra (2007); a abordagem de Ullmann (1964), Lyons (1980) e Brito (2003); a perspectiva da Lexicografia – quanto ao uso de inicial maiúscula – de acordo com os dicionários de Ferreira (1988), Houaiss (2001) e Luft (2001).

4.1.1 – O nome próprio na Gramática Tradicional

As informações contidas nas gramáticas tradicionais não são, na maioria das vezes, suficientes para que se estabeleça com precisão a diferença entre nomes comuns e nomes próprios. No capítulo dedicado à morfologia, Rocha Lima (2005: 68) faz a distinção entre nomes próprios e comuns da seguinte forma:

Os substantivos podem ser de *extensão* diferente; ora expressam a *espécie* (homem, menina, cidade, rio, etc.), ora um *indivíduo* da espécie (Bruno, Mimi, Belém, São

Francisco, etc.). Os primeiros chamam-se COMUNS; os outros, PRÓPRIOS (grifo do autor).

São elencados juntamente nomes simples (*Bruno, Mimi, Belém*) e compostos (*São Francisco*), antropônimos (nomes de pessoas – *Bruno*) e topônimos (nomes de lugar - *Belém, São Francisco*). O autor explica que os nomes próprios de pessoa admitem a flexão de número, “desde que a sua forma se adapte a pluralização” (ROCHA LIMA, 2005: 82). São apresentados exemplos de pluralização de nomes próprios de pessoas, porém, não se explica em que contextos essa variação em número se dá. O autor afirma que os substantivos próprios referem-se a “um *indivíduo* da espécie” e que esses substantivos podem variar em número, sem maiores explicações acerca desse fato. O consulente pode entender que os sintagmas: *o homem, a mulher, o menino e a criança* são nomes próprios, já que podem se referir a “um indivíduo da espécie”.

Cunha e Cintra (2007) conferem aos nomes próprios tratamento semelhante ao de Rocha Lima (2005) e mencionam a classificação dos nomes em *comum* ou *próprio*, de acordo com sua maior ou menor abrangência, isso é, afirmam que os substantivos do primeiro grupo nomeiam a totalidade dos indivíduos de uma espécie enquanto os do segundo grupo nomeiam um indivíduo de uma espécie. De acordo com os autores:

Os substantivos podem designar a totalidade dos seres de uma espécie (DESIGNAÇÃO GENÉRICA) ou um indivíduo de determinada espécie (DESIGNAÇÃO ESPECÍFICA).

Quando se aplica a todos os seres de uma espécie ou quando designa uma abstração, o substantivo é chamado comum.

Quando se aplica a determinado indivíduo da espécie, o substantivo é PRÓPRIO.

Assim, os substantivos homem, pais e cidade são comuns, porque se empregam para nomear todos os seres e todas as coisas das respectivas classes. Pedro, Brasil e Lisboa, ao contrário, são substantivos próprios, porque se aplicam a um determinado homem, a um dado país e a uma certa cidade.

(CUNHA e CINTRA, 2007: 192, grifo do autor).

Os antropônimos (*Pedro*) e os topônimos (*Brasil e Lisboa*) são tratados indistintamente, questões como a variação de gênero e número são abordadas somente em relação aos nomes comuns. Os casos dos topônimos que são constituídos de mais de um elemento, como *Espírito Santo*, em que nomes comuns funcionam como nomes próprios não são analisados.

Said Ali (1964) trata da abrangência dos nomes e discorre acerca da flexão de gênero dos nomes próprios e de seus determinantes. O autor afirma que,

O gênero dos nomes próprios de pessoas, dos sobrenomes, apelidos ou alcunhas decide-se pelo sexo das pessoas portadoras de tais nomes.

Alguns nomes de batismo têm forma masculina e feminina: *Francisco, Francisca; Luís, Luísa* etc.

(SAID ALI, 1964: 43, grifo do autor).

A afirmação de que o sobrenome é atribuído a uma pessoa de acordo com o sexo do portador parece incoerente, considerando-se que um sobrenome passa de pai para filha e de mãe para filho e, por ser uma herança de família, é atribuído a todos os elementos pertencentes aquele grupo/família independentemente do sexo dos descendentes ou ascendentes. Fica estabelecida, neste item, uma relação estreita entre o sexo do portador do nome e gênero do substantivo. Quanto ao gênero dos determinantes dos topônimos, é enfatizada a concordância do determinante com o termo genérico nos casos em que há elisão deste, por exemplo, *o (rio) Amazonas*, em que *rio* é o termo genérico – que designa o tipo de acidente natural e *Amazonas* o elemento específico. Outros casos de concordância do determinante com o topônimo também são apresentados pelo autor.

A abordagem das gramáticas normativas não contempla determinadas discussões acerca do nome próprio. Serafim da Silva Neto, no prefácio da *Gramática normativa da língua portuguesa* de Rocha Lima, afirma que este autor “estava em condições privilegiadas para dotar a língua portuguesa com uma gramática que, sem erro ou fastio, lhe desse as normas cultas do bem escrever e do bem falar” (ROCHA LIMA, 2005: xviii). Vê-se, pois, que o propósito da gramática normativa, segundo Serafim da Silva Neto é elencar normas inerrantes “do bem escrever e do falar”, sem abordar questões que não satisfazem esse propósito.

A distinção entre gramática normativa e gramática descritiva é apresentada por Bechara (2009: 52), nos seguintes termos:

A gramática descritiva é uma disciplina científica que registra e descreve [...] um sistema lingüístico em todos os aspectos (fonético-fonológico, morfossintático e léxico).

[...]

Cabe à *gramática normativa*, que não é uma disciplina com finalidade científica e sim pedagógica, elencar os fatos recomendados como modelares de exemplaridade idiomática para serem utilizados em circunstancias especiais do convívio social.

A gramática normativa recomenda como se deve falar e escrever segundo o uso e a autoridade dos escritores corretos e dos gramáticos e dicionaristas esclarecidos.

A finalidade das gramáticas normativas exclui, portanto, alguns aspectos considerados por outros estudiosos como relevantes no que concerne aos nomes próprios.

4.1.2 – O nome próprio na Linguística

Ao discorrer acerca dos nomes próprios, Lyons (1980: 183) admite que “a questão de saber se os nomes pertencem ou não a uma língua e têm ou não um significado não admite uma resposta simples e universalmente válida”. Crystal (1977: 99) afirma que

palavras como *João* têm muito mais semelhanças do que diferenças com as palavras como *rapaz* e que [...] podem legitimamente ser consideradas um tipo (mais estritamente, uma ‘subclasse’) do substantivo [...]; e, à medida que a análise fosse progredindo, seria determinada a extensão precisa da diferença estrutural [...] (grifo do autor).

O autor reconhece os nomes próprios como pertencem a uma língua ao classificá-los como substantivos e observa que nos casos em que as semelhanças são parciais existe a possibilidade de desacordos. Optamos por atribuir aos compostos toponímicos tratamento semelhante ao proposto por Crystal. Isto é, consideramos os compostos toponímicos como itens lexicais pertencentes à classe dos nomes compostos e que esses nomes apresentam semelhanças e distinções estruturais com relação aos nomes compostos comuns, essas semelhanças e diferenças são consideradas na descrição dos compostos toponímicos.

Lyons (1980) explica que o nome próprio pode, em certas culturas, assumir associações, esse fato permite que nomes próprios passem a funcionar, com o passar do tempo, como nomes comuns. Pode-se observar na língua em uso que o percurso inverso também ocorre, ou seja, nomes comuns adquirem o *status* de nome próprio com o decorrer do tempo em determinadas circunstâncias. Embora o autor não se dedique à discussão da transposição de um nome comum (ou de um grupo de nomes comuns) a nome próprio de lugar, especificamente, esse fato se dá de forma recorrente na língua.

Em grande parte das abordagens acerca do nome próprio encontra-se explícita ou pressuposta a ideia de que o ser humano faz a distinção entre esses nomes e os comuns sem maiores dificuldades. Acerca da constância no uso de determinados antropônimos e topônimos de uma língua, Lyons (1980: 183) admite que: “alguns nomes de lugar e de pessoa, institucionalizados são tão vulgares em certos países que é de se esperar que todos os locutores da língua aí utilizada reconheçam o seu estatuto de nome próprio”, contudo, esse esperado reconhecimento que o falante nativo faz de um nome próprio de sua língua não é

possível a uma máquina, o que torna imprescindível a averiguação de propriedades destes nomes que possam ser formalizadas e codificadas a fim de serem utilizadas pela máquina.

Nos estudos dedicados à semântica uma questão discutida diz respeito ao sentido do nome próprio. Ullmann (1964: 152) afirma que “muitos filósofos linguistas estão de acordo na consideração dos nomes próprios como marcas de identificação [...], um nome próprio serve apenas para identificar uma pessoa ou objeto”. O autor ressalva a fragilidade da fronteira que separa nomes próprios e nomes comuns, bem como as dificuldades geradas pela transparência semântica⁴ de determinados nomes próprios derivados de substantivos comuns e a possibilidade de um nome próprio converter-se em palavra comum.

De acordo com Ullmann (1964: 160) “embora seja bastante fácil distinguir os nomes próprios dos substantivos comuns, a fronteira entre as duas categorias não é de modo algum decisiva”. O autor explica que muitos nomes próprios são semanticamente transparentes, entre esses nomes próprios são citados os nomes de lugar que têm origem em nomes comuns. A transparência semântica dos nomes de lugar oriundos de nomes comuns pode se verificar nos topônimos capixabas: *Ribeirão Brejo Grande* (Serra), *Rio Claro* (Viana), *Pedra da Pescaria* (Vila Velha) e *Ilha da Fumaça* (Vitória), por exemplo. Há nomes próprios, porém, que apresentam opacidade semântica.

A ambiguidade gerada pela ocorrência de nomes próprios derivados de nomes comuns (por exemplo, os antropônimos *Cândido*, *Clóvis*, *Magno* e os topônimos *Mato Grosso*, *Rio de Janeiro*, *Minas Gerais*), bem como a mudança de nomes próprios a palavras comuns (por exemplo, *gilete* – oriundo do antropônimo King Camp Gillette, *belchior* – comerciante, *donjuán* – mulhêrengo) é assunto relevante nos estudos de PLN.

No que diz respeito ao sentido do nome próprio (sua existência ou não) constata-se a falta de consenso entre os estudiosos. De acordo com Brito (2003: 39-40),

o debate sobre nomes próprios tem estado bastante dividido, de sorte que se poderia, *grosso modo*, dividir as concepções que se confrontam em dois grupos. De um lado estariam aqueles que como Kripke e Mill, afirmam que os nomes próprios, para cumprirem sua função referencial, não carecem de nenhuma vinculação com descrições ou com qualquer coisa como significado. [...]. De outro lado estariam aqueles outros, os quais, na linha de Frege, vinculam aos nomes próprios o sentido

⁴ Adotamos um conceito de *transparência semântica* semelhante ao proposto por Vale (2001:72): “Pode-se entender transparência como a maior proximidade do cálculo do significado total da expressão por seus componentes, enquanto que a opacidade seria a total impossibilidade desse cálculo”.

(*Sinn – meaning* ou *sense*) para esclarecer o modo com chegaram a estar por objetos particulares na sentença (grifo do autor).

As concepções de John Stuart Mill, Saul A. Kripke e Glottob Frege, no que tange aos conceitos de *referência*, *sentido* e *significado* não serão abordadas nesta pesquisa. A citação desses autores restringir-se-á à indicação da divergência quanto ao sentido dos nomes próprios.

Brito (2003: 26-27) afirma que,

Sentenças como:

(7a) “Tigre é meu melhor amigo, mas também o mais briguento dos garotos da rua”
representam um exemplo de que “tigre” ocorre como nome próprio [...] Por outro lado, construções como:

(8a) “Basílio é um João-ninguém”

são plenamente significativas e mostram o uso de um típico nome próprio como “João” na função de um nome de espécie.

[...] O mesmo nome próprio pode nomear diferentes objetos, mas isso abre espaço para instauração de ambigüidades na comunicação [...]. Essa diferença fica clara quando temos de interpretar uma sentença como:

(9a) “João é Tigre”.

Os nomes comuns podem referir-se a diversos objetos sem com isso gerar ambigüidade, porém, a ocorrência de substantivo próprio para designação de objetos diversos (que assume o estatuto de substantivo comum), de acordo com Brito (2003), é um fator gerador de ambigüidade na comunicação. O autor certamente está considerando as ambigüidades geradas na comunicação entre humanos, que têm a capacidade de avaliar diversos elementos que fazem parte de seu conhecimento de mundo e interferir de forma positiva para a interpretação dos enunciados. Esse problema toma maiores dimensões quando se trata do PLN, da comunicação entre humano e máquina, considerando que a máquina necessita que sejam inseridas todas as informações para desfazer ambigüidades, como as citadas por Brito. Um dos critérios de identificação dos nomes próprios, comumente proposto, é o uso de inicial maiúscula, porém, na prática, ele também se mostra ineficaz, como veremos a seguir.

4.1.3 – O nome próprio na Lexicografia - o uso de inicial maiúscula

A ineficácia do critério do uso de iniciais maiúsculas para identificação dos nomes próprios fica evidente em alguns dicionários de língua portuguesa. A falta de precisão no uso de maiúscula não ocorre apenas nos dicionários de língua portuguesa do Brasil, nos dicionários de língua francesa ocorre fato semelhante. Segundo Vaxelaire (2005), em estudo comparativo entre o *Petit Larousse Illustré* (2000) – PLI – o *Nouveau Petit Robert* (2000) – PR1 – e o *Trésor de la langue française informatisé* – TLFi – consultado no mês de março de 2005, foram observadas algumas divergências quanto ao uso de inicial maiúscula. O autor analisa a entrada do item *Sagitário*: de acordo com o primeiro sentido do PR1, *sagitário* é um termo de astronomia cuja definição é: “Constellation zodiacale de l’hémisphère austral” (Constelação zodiacal do hemisfério sul – tradução nossa). O TLFi indica, na acepção B1 deste item, que *Sagitário* escreve-se frequentemente com letra inicial maiúscula. Vaxelaire (2005) sugere que, se os nomes de planetas (como Mercúrio ou Saturno) são considerados nomes próprios, os nomes de constelações deveriam receber o mesmo tratamento.

O dicionário da língua portuguesa Houaiss (2001) apresenta na primeira acepção da entrada de *Saturno* a seguinte definição: “em relação ao Sol e em ordem crescente, o segundo e maior planeta do sistema solar”, sem recomendação do uso de inicial maiúscula, ao passo que na terceira acepção de *Netuno* lê-se: “o oitavo planeta a partir do Sol” e a recomendação do uso de inicial maiúscula. O mesmo se dá na segunda acepção de *Sagitário*: “nona constelação zodiacal, situada entre Escorpião e Capricórnio”. Embora os três itens mencionados sejam nomes de planetas, no primeiro (*Saturno*) não se recomenda o uso de inicial maiúscula como nos dois últimos (*Netuno* e *Sagitário*).

Para o consulente não fica claro o critério adotado para a inclusão de alguns nomes próprios em um dicionário de nomes comuns. O lugar do nome próprio na Lexicografia não parece ser assunto bem definido nos dicionários franceses, ainda que os franceses possuam obras dedicadas exclusivamente a esses nomes. Pode-se constatar, através de uma breve observação, que as escolhas e critérios para a inclusão dos nomes próprios nos dicionários de língua portuguesa são passíveis de revisão.

4.2 – A descrição do nome próprio para o PLN

Os nomes próprios são considerados em algumas abordagens linguísticas como elementos que trazem pouca contribuição para a descrição e o funcionamento da língua. Borba (1996: 125), por exemplo, ao tratar da valência dos nomes concretos afirma que para fins operacionais não serão considerados os nomes próprios, “primeiro porque eles não nomeiam, apenas se referem a objetos físicos [...], e, segundo, porque têm, parece, sintaxe (muito) simples e semântica praticamente nula”. Afirmarões como essas, podem conduzir à compreensão equivocada de que o tratamento dos nomes próprios para o PLN é tarefa simples e talvez desnecessária. De acordo com Correia (2002: 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”. A autora explica que a definição do nome próprio como o designador de um objeto singular gera dúvidas, visto que dois ou mais objetos podem ser designados pelo mesmo nome. Considerando-se que os nomes são elementos linguísticos largamente presentes nos textos e necessários à coerência textual, o seu tratamento é essencial para o PLN sejam eles nomes comuns ou nomes próprios.

Ao tratar dos dicionários eletrônicos Ranchhod (2001: 24) explica que além dos dicionários de palavras simples, palavras flexionadas – geradas a partir das palavras simples – e palavras compostas, “estão a ser elaborados léxicos mais específicos, nomeadamente dicionários terminológicos [...], dicionários de siglas [...], dicionários de topônimos, etc.” Essa afirmação pressupõe o fato de que, até aquele momento (2001), tais léxicos para a língua portuguesa ainda estavam sendo desenvolvidos.

A delimitação entre o nome próprio e o nome comum não é tarefa simples, como pode parecer a princípio. A identificação do nome próprio pelo uso de inicial maiúscula, apresentada como critério distintivo, nem sempre se mostra eficaz. Para Friburguer (2002: 15) identificação dos nomes próprios através da observação do uso de inicial maiúscula é insuficiente porque esses nomes podem ser compostos de vários itens lexicais e alguns deles não são grafados com inicial maiúscula. Além disso, a maiúscula que se encontra na primeira palavra de uma frase é ambígua, podendo tratar-se de um nome próprio ou de uma palavra comum grafada com letra inicial maiúscula por estar no início de uma frase.

Vale ressaltar que as informações contidas nos dicionários eletrônicos diferem das informações observadas comumente nos dicionários de uso. Ranchhod (2001: 16) destaca algumas diferenças entre os dicionários de uso e os dicionários eletrônicos e afirma que “estes dicionários têm que conter informações lingüísticas codificadas e formatadas [...]. Não podem conter lacunas”.

Ao tratar do formalismo e das restrições gramaticais, Laporte (2001: 78) afirma que,

o carácter exacto ou inexacto das restrições formalizadas permanece inalterado no caso de introdução de novas etiquetas lexicais que descrevam novas acepções das mesmas palavras no dicionário. Consideremos por exemplo a seguinte frase:

Ele está feliz.

[...]

Imaginemos agora o aparecimento de uma revista com o título *Ele* [...].

(32) *Ele (= a revista Ele) está cheia de reportagens esta semana.*

O autor considera a realização de *ele* como pronome e a realização de *Ele* como nome próprio acepções da mesma palavra. Nesse caso, as restrições gramaticais e informações das etiquetas lexicais auxiliam na eliminação das ambiguidades. Para resolver a questão da ambiguidade quanto ao gênero do adjetivo nos exemplos propostos, Laporte (2001: 78) apresenta as etiquetas lexicais e regras de restrições.

No exemplo *Ele está feliz*, o item lexical *ele* realiza-se como pronome de terceira pessoa do masculino singular (etiqueta <ele.PRO:3ms>) e exige adjetivo no gênero masculino.

O autor explica que ocorrendo apenas a etiqueta relativa à construção *Ele está feliz* a sentença *Ele está cheia de reportagens esta semana* seria considerada imprópria, por não se enquadrar na restrição de gênero (masculino) do adjetivo. Contudo, com a introdução da informação acerca deste exemplo, ou seja, da ocorrência do nome próprio feminino – gênero do substantivo elíptico *revista* – (etiqueta <Ele.Npr:fs>), que exige adjetivo no feminino, não há necessidade de modificações na etiqueta do pronome *ele* e haverá a possibilidade de análise das duas estruturas.

No tocante aos aspectos teóricos lingüísticos envolvidos nas pesquisas de PLN, Dias da Silva (2006: 131) afirma que,

A análise conceitual desenvolvida [...] sob a denominação de “análise componencial”, acabou por ser aplicada aos estudos de decomposição do significado dos itens lexicais. Com a denominação de “o estudo do significado em uso”, sob a

influência de Austin e Wittgenstein, transformou-se em um método de análise que procura investigar como uma expressão lingüística pode ter significados diferentes em diferentes contextos de uso, enfatizando, portanto, a forte dependência contextual do significado. De outro, com o desenvolvimento da lógica moderna, por lógicos como Frege e Russell, e com a aplicação dos seus métodos e resultados aos estudos lingüísticos, temas como [...] “contextos transparentes e opacos” [...] passaram a fazer parte das discussões sobre a semântica das línguas naturais.

A abordagem do aspecto semântico dos estudos lingüísticos em PLN abriu espaço para que fossem pensadas questões como: o tratamento dos nomes próprios na elaboração de um dicionário eletrônico de topônimos – nomes de lugar e de antropônimos – nomes de pessoas, a questão da homonímia e/ou polissemia na elaboração dos dicionários eletrônicos, a possibilidade de ocorrência de polissemia em nomes próprios etc. Entretanto, nesta pesquisa o enfoque é para os nomes de lugares.

Na elaboração de dicionários eletrônicos cabe ao profissional em língua portuguesa do Brasil realizar as descrições lingüísticas adequadas. Embora o falante nativo tenha uma boa intuição da língua materna, essa intuição não é suficiente para analisar dados complexos.

4.3 – Do nome comum ao nome próprio - um *continuum*

Neste subcapítulo analisamos o percurso que o nome comum faz até assumir o *status* de nome próprio. A trajetória do geônimo – nome de acidente geográfico natural – capixaba *Morro Engenharia* é apresentada como exemplo desse fato lingüístico.

Melo (1978), ao tratar da mudança de classe que os itens lexicais podem sofrer, aponta a possibilidade de alteração do significado e inclui nesse contexto a mudança do nome comum em nome próprio.

Assim é que substantivos comuns se tornam nomes próprios: *Leite*, [...] *Matos*, *Pinheiro*, etc., etc.

Nomes geográficos vindos das classes de substantivos comuns temos em *Campos*, *Campinas*, [...] *Rodeio*, etc., etc.,

Pelo contrário, alguns substantivos próprios passam a comuns, do que são exemplos *carrasco*, *belchior*, *damasco*, *ferrabrás*, *figaro*, *macadame*, [...] *angora*, etc., etc. (MELO, 1968: 91-92, grifo do autor)

Os nomes são elencados sem que se estabeleça uma relação entre um grupo e outro, nem é analisada a maneira como esse processo de mudança ocorre.

O autor observa as alterações morfológicas e semânticas dos itens em questão e admite a possibilidade de mudança de classe ocorrer em dois sentidos: de nome comum a nome próprio e de nome próprio a nome comum. Os substantivos comuns: *leite, matos, pinheiro* podem ser utilizados como antropônimos (sobrenomes principalmente) ou como topônimos. Ocorre que os nomes comuns passam a designar um elemento dentro de um determinado grupo ou um dado conjunto de elementos dentro de um universo maior, como é o caso dos sobrenomes (nomes de família). Nessa mudança os nomes comuns utilizados como antropônimos – seja como prenome (*Rosa, Cláudio, Clara* etc.), seja como sobrenome (*Lima, Pereira, Lobo* etc.) – sofrem um processo que alguns autores nomeiam de *erosão semântica, esvaziamento semântico* ou *desbotamento semântico*, de forma que a associação do nome ao objeto nomeado não se pode perceber nitidamente.

Na mudança de *status* de nome próprio a nome comum, o item lexical usado para distinguir um indivíduo ou objeto de determinada espécie tem seu significado⁵ primitivo embotado e passa por um processo de ressignificação. São associadas ao nome as características mais marcantes atribuídas ao seu portador. O nome da cidade italiana *Veneza*, por exemplo, pode ser relacionado com o povo conhecido como *vênetos*, porém, essa relação foi paulatinamente enfraquecendo, passando o nome *Veneza* a funcionar como elemento particularizador de um determinado espaço físico. Na sentença:

(1) “Troca-troca na ponta da corrida eleitoral da **Veneza** brasileira”,

a expressão *Veneza brasileira* não se refere ao povo conhecido como *vênetos*, nem à cidade localizada da Itália, mas refere-se às características desta cidade – seus rios, suas pontes e a beleza que lhe é atribuída. São ressaltadas essas e outras características na utilização do nome *Veneza* como elemento indicativo da capital de Pernambuco.

Bechara aponta a distinção entre nomes comuns e nomes próprios e apresenta a subclassificação destes em *antropônimo* – que inclui os prenomes, sobrenomes e apelidos – e *topônimos*. O autor aborda a questão da singularidade do referente de alguns nomes comuns e

⁵ Não é objetivo desta pesquisa discutir a denotação e conotação dos nomes próprios. Partimos do pressuposto de que os nomes próprios (especialmente os topônimos) têm uma motivação inicial, que pode perder-se e tornar o nome opaco, não sendo possível, em alguns casos, o resgate da referência primeira.

a flexão de número dos nomes próprios, e explica a passagem dos nomes próprios a comuns nos seguintes termos:

Não nos prendemos apenas à pessoa ou coisa nomeada; observamo-lhes qualidades e defeitos que se podem transferir a um grupo mais numeroso de seres. Os personagens históricos, artísticos e literários pagam o tributo de sua fama como o desgaste do valor com o desgaste individualizante do seu nome próprio que, por isso, passa a comum. [...].

Passam a substantivos comuns os nomes próprios de fabricantes e de certos lugares onde se fazem ou se fabricam certos produtos [...].

(BECHARA, 2009: 114)

Bechara nomeia de “desgaste individualizante do nome próprio”, referindo-se à particularização e abrangência dos nomes (próprios e comuns, respectivamente). Essa perda que o nome próprio sofre na sua qualidade de elemento individualizante pode ser associada a um processo de ressignificação do nome em que há um acréscimo na complexidade semântica desse nome.

A motivação da nomeação dos lugares e o momento de transposição de nomes comuns a topônimos são assuntos de interesse da toponímia. O uso constante dos nomes explicativos (*rio, morro, ilha, córrego* etc) acaba por transformá-los em topônimos.

É a mudança da categoria gramatical dos substantivos comuns em próprios que lhes confere esse *status*. Sua finalidade é indicar, com precisão, o acidente que nomeiam. Assim aconteceu com a série de rios franceses [...]; o mesmo fenômeno ocorreu, entre nós, com os *pará*, "rio ou mar", no guarani; *paraná*, "rio"; *iguaçu*, "rio grande", tornados, pela cristalização semântica do significado, *Pará*, *Paraná*, *Iguaçu*, por exemplo.

(DICK 1997:45).

Observa-se a perda de algumas características do nome comum e a sua particularização, no que concerne à abrangência, nos exemplos apresentados por Dick (1997) os nomes comuns, de origem guarani, passam a funcionar como nomes próprios. Como afirma a autora, o topônimo “[...] não é um signo lingüístico especial, mas, ao contrário, um designativo vocabular comum, acrescido, porém, da função específica de identificação dos lugares [...]” (DICK, 1982: 76). Desse modo, o nome comum *paraná* passa a identificar um acidente artificial (um estado da região Sul do Brasil) e um acidente natural (um rio que demarca a fronteira entre Brasil e Paraguai). A ocorrência da cristalização semântica de *paraná* (rio ou mar) e sua mudança de *status* para nome próprio acarreta a exigência do uso de um elemento genérico (*rio* ou *estado*). A inserção do elemento ou termo genérico, no caso dos topônimos *Paraná* e *Iguaçu*, cooperara para a desambiguação, visto que o nome *Paraná* pode referir-se a

um estado ou a um rio. O topônimo *Iguaçu* pode referir-se a um rio do estado do Paraná, às quedas de água deste rio ou a um rio do estado do Rio de Janeiro. No topônimo *Rio de Janeiro* observa-se também a cristalização do primeiro item (*Rio*), a sua incorporação ao elemento específico e a necessidade de inclusão do termo genérico *estado*, imprescindível na distinção entre o *estado* e o *município* homônimo.

Ao tratar da toponímia do estado de São Paulo, Dick afirma que: “ao toponimista [...] importará conhecer o momento de transposição, ou seja, quando Piratininga deixou de ser assim chamada para denominar-se São Paulo de Piratininga, ou [...] só o *nome português* (São Paulo), simplesmente” (DICK, 1997: 67-68, grifo da autora). A mudança do nome comum para nome próprio é um fato linguístico de grande importância para a elaboração de um dicionário eletrônico de nomes próprios e para estudos toponímicos, por marcar o momento em que o nome comum assume o *status* de nome próprio tornando-se objeto de estudo da Toponímia e elemento passível de ser inserido nos dicionários eletrônicos de nomes próprios.

Os estudos toponímicos lidam com a questão da mudança linguística. Na modificação do nome *Piratininga* para *São Paulo de Piratininga*, e, posteriormente, de *São Paulo de Piratininga* para *São Paulo*, observa-se o fator social do acréscimo do nome português (São Paulo) ao nome indígena (Piratininga) no contato inicial entre esses dois povos – indígena e europeu – e, num momento seguinte, a redução do nome com o apagamento do nome indígena, indicativo provável das pressões de uso e do estabelecimento e domínio do elemento europeu em terras brasileiras.

Neves (2001: 15) nomeia de *gramática funcional* “uma teoria da organização gramatical das línguas naturais que procura integrar-se em uma teoria global da interação social”. Essa visão coaduna-se com a afirmação de Dick (2007: 144) de que “a Onomástica [...] é indício de rumos tomados pelos falares ao longo dos períodos históricos, de comportamentos presentes no cotidiano e de atitudes morais ou operosas valorizadas pela população”, ou seja, o estudo dos topônimos considera as interações sociais motivadoras do ato de nomear, aspecto que se assemelha à abordagem funcionalista da língua.

Ao tratar dos princípios e resultados do Léxico-Gramática, Laporte (2008: 33) afirma que:

As propriedades sintático-semânticas são identificadas por meio de fórmulas [...]. Enfim, somente são estudadas as propriedades para as quais se encontra um procedimento que permita determinar de maneira suficientemente confiável se uma entrada a possui ou não: as propriedades são pois modalizadas como binárias e não como um *continuum* (grifo do autor).

A observação do *continuum* que vai do nome comum ao nome próprio, ou do grupo nominal livre à palavra composta, se faz com o fim de constatar o momento em que se dá essa transposição, visto que o interesse da descrição proposta fundamenta-se numa visão binária, conforme explica Laporte. Os pontos que formam o *continuum* nome comum-nome próprio não são analisados detalhadamente. Ressaltamos o ponto em que a mudança de um nome comum a nome próprio, ou de um grupo nominal livre à palavra composta faz com que esse elemento passe a integrar nosso objeto de estudo. Essa mudança pode ser percebida, entre outros aspectos, pela perda da composicionalidade do grupo nominal e sua maior fixidez sintática.

Observamos o *continuum* entre a expressão descritiva que identifica um dos acidentes naturais da cidade de Vitória e o nome próprio *Morro Engenharia* em um espaço de tempo de cinquenta anos (entre 1958 e 2008), considerado suficiente para a identificação das mudanças ocorridas.

A Prefeitura Municipal de Vitória nomeia uma rua da capital do Espírito Santo através da lei 2.523, de 1977, onde se lê no artigo primeiro:

(2) “Fica denominada “Joaquim José da Vitória”, a rua que partindo a Av. Maruípe, ao lado da padaria Eliana, dá acesso ao **morro conhecido por Engenharia**, nesta capital” (grifo nosso).

O acidente geográfico (*morro*) é identificado a partir do nome do bairro onde está localizado (Engenharia). Nota-se que não se faz referência ao nome oficial do bairro ou do morro onde está localizada a rua nomeada, mas admite-se o nome em uso pela comunidade daquela localidade. Os itens lexicais destacados compõem um grupo nominal livre. Neste caso é aceitável a inserção de um item lexical no conjunto:

(2a) Fica denominada “Joaquim José da Vitória”, a rua que partindo a Av. Maruípe, ao lado da padaria Eliana, dá acesso ao **morro geralmente conhecido por Engenharia**, nesta capital.

O artigo primeiro da lei 3.424, de 1987, apresenta a seguinte redação:

- (3) “Fica denominada MARIA JOSÉ BALDAN a rua que partindo do início da Rua Joaquim José da Vitória [...] vai dar acesso ao **morro do Bairro Engenharia Maruípe**” (grifo nosso).

A lei 3.424 mantém a identificação do acidente geográfico (*morro*) através do nome do bairro em que está situado. Neste caso é mencionada a nomeação sistemática do bairro, mas ainda não atribuído uma denominação específica para o morro.

A designação de lugares através da utilização de referências ao espaço físico em que se encontra indica um alto grau de vinculação do nomeador com o meio no qual vive. Essa forma de referenciar fica notória na localização da rua Joaquim José da Vitória, registrada no artigo primeiro da lei 2.523:

- (2) “Fica denominada ‘Joaquim José da Vitória’, a rua que partindo a Av. Maruípe, **ao lado da padaria Eliana**, dá acesso ao morro conhecido por Engenharia, nesta capital” (grifo nosso).

A numeração do imóvel localizado na avenida Maruípe não é mencionada, a referência utilizada é um estabelecimento comercial que funcionava próximo àquele lugar. Essa forma de localização de um determinado espaço através de um referente pertencente ao conhecimento partilhado daquela comunidade deixa entrever o modo de se situar dos capixabas, com ênfase para as referências locais ao invés da utilização da nomeação e numeração sistemática.

Na redação da lei 3.438, de 1987, que oficializa a nomeação de uma travessa do bairro Engenharia, lê-se:

- (4) “Fica denominada travessa Coronel Manoel da Silva a servidão pública que dá acesso ao **morro da Engenharia** [...]” (grifo nosso).

A utilização da expressão descritiva *morro do Bairro Engenharia Maruípe* ocorre simultaneamente ao uso do sintagma toponímico *morro da Engenharia* (registrados em 09 de janeiro de 1987). Nota-se a redução da quantidade de itens lexicais, a perda pragmática e a maior fixidez sintática na transposição de *morro do Bairro Engenharia Maruípe* para *morro da Engenharia*.

A lei 3.438 registra a forma *morro da Engenharia*, o elemento genérico é ligado a um elemento específico através da preposição *de*. Observa-se a institucionalização de *morro da Engenharia*, que acarreta, por exemplo, a impossibilidade de inserção de itens lexicais no composto:

(4a) * Fica denominada travessa Coronel Manoel da Silva a servidão pública que dá acesso ao **morro alto da Engenharia**.

Percebe-se, na redação da lei 5.515, de 2006, que o poder público municipal dá preferência ao designativo *bairro* em detrimento do uso do designativo *morro*, contudo, reconhece a nomeação espontânea que é registrada no artigo primeiro:

(5) “Fica denominado ‘Bairro Engenharia’, o conglomerado residencial e comercial hoje denominado ‘**Morro Engenharia**’” (grifo nosso).

Segundo Dick (1990a: 98), a denominação espontânea e natural,

aos poucos, por força do costume e do hábito, acaba se arraigando no espírito da população e transformando-se na única forma denominativa conhecida e aceita. Seu característico é a praticidade de uma fácil assimilação e ganha aspectos de uma autêntica forma de nomear [...] porque oriunda do nativo da região. Não lhe foi imposta nem visa homenagens a terceiros estranhos ao meio [...].

A transposição do nome comum a nome próprio – no caso, topônimo – dá-se pelo uso, ficando as nomeações (sistemáticas) impostas por força de lei relegadas a um segundo plano, tornando-se muitas vezes concorrentes daquelas.

Comparando-se os sintagmas toponímicos *morro da Engenharia* e *morro Engenharia* fica patente a perda do elemento de ligação (a preposição *de*) entre os substantivos o que aponta para uma maior vinculação sintática entre os mesmos.

Pode-se constatar, pela trajetória de designações atribuídas a um acidente geográfico da Grande Vitória convertida em nome próprio, que esse percurso é marcado pela perda de composicionalidade, ou seja, o significado não é mais considerado a partir do significado das partes que formam o grupo nominal, o significado se deduz do conjunto como unidade. O sentido de descrição do acidente natural (*morro*) fica enfraquecido e há maior fixidez da ordem.

A análise deu-se a partir do topônimo já estabelecido como tal, mas o processo não foi considerado em sua totalidade. Pode-se observar também embotamento semântico do elemento específico do sintagma toponímico – *Engenharia* – referente à Escola de Engenharia estabelecida naquela localidade.

Pode-se afirmar que na mudança: *morro do Bairro Engenharia Maruípe* > *morro conhecido por Engenharia* > *morro da Engenharia* > *morro Engenharia*, ocorre o surgimento de uma designação (nome próprio) para um determinado acidente natural.

4.4 – Os topônimos

Os estudos toponímicos situam-se no âmbito da ciência chamada *Onomástica*, estudo dos nomes em geral. A *Onomástica* apresenta duas grandes divisões: a *Antroponímia* – que se dedica aos estudos dos nomes próprios de pessoa e a *Toponímia* – área em que se concentram os estudos dos nomes próprios de lugares. Os antropônimos podem ser utilizados para designar lugares funcionando como topônimos, ou melhor, como antropotopônimos (nomes de lugar relativos aos nomes próprios de pessoas). Podemos citar como exemplo desse fato a denominação do bairro *Maria Ortiz*, no município de Vila Velha – Espírito Santo. O nome próprio de pessoa – *Maria Ortiz* – é utilizado como especificador de um acidente artificial (bairro) capixaba.

Embora a distinção entre antropônimos e topônimos e entre o comportamento desses e daqueles nomes pareça uma tarefa simples, alguns equívocos quanto a essa distinção são observados inclusive em obras atinentes ao estudo da língua portuguesa. Borba (1996: 125), por exemplo, em uma nota de rodapé em que diferencia antropônimos e topônimos - quanto ao uso de artigo definido - e trata da topicalização de antropônimos, utiliza topônimos (*Roma*, *Paris* etc.) nos exemplos dados.

Qualquer antropônimo pode receber artigo quando vem topicalizado com especificador – Visitei a Roma dos Césares. Estive na Vila Isabel do Noel Rosa, na Ipanema de Vinícius. Qualquer deles também pode ser topicalizado por demonstrativo – O que eu me diverti naquela Paris!/Está infestada esta Moema.

Borba (1996) apresenta como exemplo de topicalização de antropônimos pela anteposição do artigo definido: “Estive na **Vila Isabel** do Noel Rosa, na **Ipanema** de Vinícius” (grifo nosso). O que ocorre no exemplo proposto pelo autor é a topicalização de nomes de lugar (topônimos). No exemplo dado, os antropônimos são *Noel Rosa* e *Vinícius*. Nas frases: “O que eu me diverti naquela **Paris!**/Está infestada esta **Moema**” (grifo nosso), os nomes topicalizados por especificadores são topônimos, acidentes artificiais e não antropônimos como propõe o autor. Embora *Moema* possa funcionar como um antropônimo (por exemplo, nomeia a heroína do poema *Caramuru* de Santa Rita Durão), na frase proposta *Moema* parece funcionar como nome de lugar (antropotopônimo).

Segundo Dick (1982: 76), “o topônimo não é um signo lingüístico especial, mas, ao contrário, um designativo vocabular comum, acrescido, porém, da função específica de identificação dos lugares”. Desse fato decorre, por vezes, a dificuldade de se estabelecer a diferença entre grupos nominais livres, formados por nomes comuns e um topônimo composto. A autora explica (1990a: 110) que “acidentes geográficos ou nomes comuns ou simples designativos assim tratados no VLB acabaram por se firmar como topônimos, recebendo, eles próprios, em alguns casos, um complemento circunstancial.” Esses nomes que figuram no *Vocabulário da Língua Brasileira*, citados por Dick, enquadram-se na categoria *geomorfotopônimos*, ou seja, são topônimos que fazem referência às conformações topográficas da região nomeada.

De acordo com Dick (2006: 98-99)

A Toponímia desenvolveu-se acompanhando um vocabulário geográfico de ocorrências, pois geográfico era o seu objeto de estudo (topos=lugar) [...].

Na medida que consideramos a Toponímia – neste caso em foco – como uma linguagem de especialidade, mas diferenciada daquelas outras reconhecidamente técnicas [...].

Os topônimos são elementos lingüísticos pertinentes à terminologia geográfica, portanto, empregados em um contexto específico, da Geografia. O uso dos topônimos, porém, não se restringe a essa especialidade, observa-se a ocorrência de topônimos como elementos do discurso na língua geral.

4.5 – Os compostos toponímicos

Na pesquisa linguístico-descritiva da toponímia os nomes de rios e córregos, entre outras formações, por vezes, referem-se aos lugares em que se encontram esses acidentes. Observa-se essa prática, por exemplo, nas seguintes frases:

- (6) A Fazenda das Palmas fica na **Localidade de Córrego Fundo**, a 4 km do Centro de Aracruz.
- (7) A localidade rural de Cidade Nova da Serra, mais conhecida por **Chapada Grande**, distante 10 quilômetros do centro da Serra, terá sua rede de abastecimento de água.

Cabe ressaltar que, Dick (1990: 13-14) considera como *topônimo composto* ou *elemento específico composto* o topônimo “que se apresenta com mais de um elemento formador, de origens diversas entre si, do ponto de vista do conteúdo”, porém, o conceito de *composto toponímico* adotado nesta pesquisa diz respeito à estrutura do item lexical, desconsiderando-se a origem diversa ou comum dos elementos constituintes.

Quanto à estrutura do topônimo, Dick (1990: 10) observa que há uma relação binômica entre o acidente geográfico identificado e o nome próprio. A autora explica que o sintagma toponímico é formado por um *termo* ou *elemento genérico* e um *termo* ou *elemento específico* e que ambos atuam “no sintagma toponímico, de forma justaposta (rio das Amazonas) ou aglutinada (Parauna, ‘rio negro’), conforme, portanto, a natureza da língua que os inscreve”. Assim sendo, pode-se considerar que a estrutura do topônimo enquadra-se nos conceitos de nome composto (justaposto ou aglutinado) apresentados no capítulo anterior. De acordo com Dick apenas o segundo elemento do composto enquadra-se na categoria de nome próprio.

Menezes e Santos (2008: 78) afirmam, acerca dos elementos que constituem o topônimo (termo genérico e termo específico), que a

Cartografia considera impositivamente, esses dois elementos, porquanto a parte genérica de um topônimo indica a que tipo de acidente se refere este nome, se a um curso d’água ou se a uma forma orográfica, por exemplo, enquanto a específica particulariza, identifica e qualifica com precisão o acidente, ao mesmo tempo em que, no aspecto geral, exprime um atributo característico do lugar.

Embora entendamos que nem sempre o termo específico exprime um atributo característico do lugar, consideramos a impositividade deste termo assim como do termo genérico um fato incontestável, não apenas para a cartografia, mas também para a análise linguística desses nomes.

A Gramática Tradicional aborda algumas questões acerca dos compostos toponímicos, principalmente no que concerne à ortografia. Bechara (2009) recomenda o uso de letra inicial maiúscula em qualquer espécie de nome próprio. Entre os exemplos elencados pelo autor encontram-se: *Guanabara*, *Tietê*, *Atlântico* nos quais os termos genéricos (**Baía/Estado da Guanabara**, **Rio Tietê**, **Oceano Atlântico**) são omitidos. De acordo com o autor, devem ser grafados com inicial maiúscula “os nomes de vias e lugares públicos: *Avenida de Rio Branco*, *Beco do Carmo*, *Largo da Carioca*, *Praia do Flamengo*, *Praça da Bandeira*, *Travessa do Comércio*, *Túnel Noel Rosa*, etc.” (BECHARA, 2009: 104, grifo do autor).

Nota-se que não apenas o segundo elemento do composto é tido como nome próprio, mas os termos genéricos *Avenida*, *Beco*, *Largo*, *Praia*, *Praça*, *Travessa* e *Túnel* são também grafados com inicial maiúscula, o que nos leva a inferir que são considerados pelo autor como parte integrante do composto toponímico.

Ao tratar da aposição com a preposição *de*, Bechara (2009: 458) analisa a distinção entre *Ilha de Marajó* e *casa de Pedro* e afirma que “em *ilha* e *Marajó* se trata de uma só realidade, já que ambos querem referir-se a um só conteúdo de pensamento designado”. O autor observa o caráter não-composicional do geomorfotopônimo *Ilha de Marajó* em contraposição ao grupo nominal livre *casa de Pedro*.

De acordo com Rocha Lima (2005: 60) o uso de letra inicial maiúscula é recomendado em qualquer nome próprio. Consideram-se os nomes de pessoas e os nomes de lugar: “[...] *Avenida Presidente Vargas*, *Rua do Ouvidor*” e incluem-se nesse item os nomes de acidentes geográficos:

“*Amazonas*, *Guanabara*, *Vesúvio*, *Morro da Viúva*;

Paquetá, *Danúbio*, *Estreito de Magalhães*” (grifo do autor).

Rocha Lima omite o termo genérico ou entidade geográfica em alguns dos exemplos de topônimos apresentados: (**Rio**) *Amazonas*, (**Baía/Estado da**) *Guanabara*, (**Vulcão**) *Vesúvio*,

(**Ilha de**) *Paquetá*, (**Rio**) *Danúbio*. Contudo, nos exemplos em que há o registro do elemento genérico, ele vem grafado com letra inicial maiúscula: **Morro da Viúva** e **Estreito de Magalhães**. Pode-se deduzir, desses fatos, que o termo geográfico é considerado por esses autores como parte integrante do topônimo, ou seja, apresenta o *status* de nome próprio, ainda que ocorra a elipse do termo genérico.

Segundo Sandmann (1988: 127) nos nomes próprios a relação determinante (DT) - determinado (DM) e vice-versa é mais semântica do que sintática, o autor afirma que o “DM é em geral um substantivo comum (designativo de espécie ou gênero), com uma extensão mais ampla do que a do DT, o qual é, em geral, como o produto da combinação, um nome próprio”. Essa consideração acerca da “origem do nome próprio”, da relação DT-DM/DM-DT, leva-nos a compreender o fato do autor grafar os termos genéricos com inicial maiúscula (*Brasil Turismo, Camisaria Novo Mundo* etc.).

Carvalhinhos e Antunes (2007: 117) explicam que, às vezes, ocorre a elisão do termo genérico, principalmente nos nomes dos aglomerados humanos, e que “em caso de homonímia, o genérico entra como elemento para evitar ambigüidades: assim, temos o *estado do Rio de Janeiro* em oposição à *cidade*” (grifo das autoras). Nesse exemplo a ambigüidade ocorre por causa da homonímia entre *cidade* e *estado*. A homonímia entre *cidade* e *estado* acontece também com o topônimo *São Paulo*. O topônimo *Rio de Janeiro* não gera dúvidas quanto à condição do substantivo *Rio* como parte integrante do termo específico, destituído do sentido de *curso de água*. O mesmo não acontece com alguns topônimos capixabas, como *Morro Engenharia, Praia da Costa e Córrego Fundo*, por exemplo. Esses nomes mantêm ainda estreita relação com o acidente geográfico (*morro, praia e córrego*) podendo ainda referir-se ao acidente artificial a que o acidente natural dá nome. Nesses casos, a elisão do termo genérico *bairro/localidade* colabora com a ocorrência de ambigüidade. A possibilidade de ocorrência de ambigüidade gerada pela elisão do termo genérico aponta para a necessidade do registro do composto toponímico de forma integral no dicionário eletrônico de compostos toponímicos do português do Brasil. As ocorrências de elipse de um dos elementos constituintes do composto são consideradas variações/reduções do nome.

5 – DICIONÁRIO ELETRÔNICO DE NOMES COMPOSTOS TOPONÍMICOS DO PORTUGUÊS DO BRASIL

Os sistemas de informação contam com uma série de ferramentas para o seu bom desempenho. No PLN os dicionários eletrônicos estão entre as ferramentas imprescindíveis para a tradução automática, para o reconhecimento de entidades mencionadas, para a geração de textos e outras aplicações.

Segundo Biderman (2001c: 90) em virtude das pesquisas em “PLN, TA, BDL, Lexicografia Computacional, Terminologia e *Tecnologias da Fala*, o léxico veio ocupar um lugar central nas pesquisas e teorias sobre a linguagem” (grifo da autora)⁶. O desenvolvimento de pesquisas em PLN evidencia a necessidade da construção de léxicos computacionais, de *dicionários eletrônicos* do português do Brasil.

O termo *dicionário eletrônico* é comumente utilizado de forma ambígua. Muitos dicionários de uso, em suporte informático, são comercializados sob a designação de *dicionário eletrônico*. De acordo com Ranchhod (2001: 15),

O que aproxima e justifica a designação de dicionário para os dois tipos de objectos é que, em qualquer dos casos, se trata de repositórios lexicais mais ou menos completos, organizados em *artigos*. Cada artigo é constituído por uma *entrada* (em geral, uma palavra) e um *conteúdo* (a descrição linguística da entrada) (grifo da autora).

A autora observa o emprego do substantivo *dicionário* para designar tanto os dicionários usuais como as ferramentas computacionais utilizadas no PLN (dicionários eletrônicos) e justifica o emprego dessa designação em relação aos léxicos computacionais. Os dicionários podem ser de diversos tipos: bilíngues ou monolíngues, gerais ou especializados. Neste capítulo consideramos apenas os dicionários monolíngues.

Para uma melhor compreensão do uso e do significado da expressão *dicionário eletrônico* no âmbito do PLN é necessário fazer a distinção entre *dicionário eletrônico* e *dicionário digital*.

⁶ TA – Tradução Automática.
BDL – Bases de Dados Lexicais.

5.1 – Dicionário digital

Biderman (2001b: 131-133) afirma que os dicionários “constituem uma organização sistemática do léxico, uma espécie de tentativa de descrição do léxico de uma língua” e que “nenhum dicionário por mais volumoso que seja, dará conta integral de uma língua de civilização”. Essas observações são válidas tanto para os dicionários usuais quanto para os dicionários eletrônicos e apontam para a necessidade de atualização constante dessas obras.

O dicionário digital é uma lista de palavras de uma língua concebida para ser utilizada por humano. Essas listas apresentam entradas com a palavra grafada em sua forma canônica (os substantivos e adjetivos no masculino, singular; os verbos no infinitivo etc.) e informações como o(s) sentido(s) da palavra, sua classe gramatical, sinônimo, antônimo, ortografia, pronúncia, etimologia etc. O conteúdo desses dicionários é semelhante ao conteúdo das edições impressas dessas obras. O dicionário digital é, pois, a versão informatizada do dicionário de papel, sendo que o formato digital proporciona ao usuário agilidade no acesso às informações e economia de espaço.

De acordo com Biderman (2001a: 17-18):

O dicionário de língua faz uma descrição do vocabulário da língua em questão, buscando registrar e definir os signos lexicais que referem os conceitos elaborados e cristalizados na cultura. [...] esse produto cultural deveria ser de uso obrigatório para todos os usuários da língua.

Esse dicionário, disponível em linguagem natural, destina-se exclusivamente ao uso humano. A natureza do usuário a que se destina o dicionário indica a natureza das informações contidas nele. Segundo Biderman (2001c: 91) “com relação aos dicionários tradicionais, eles foram elaborados para os humanos que possuem um grande volume de conhecimentos implícitos sobre a estruturação do léxico”. Os dicionários destinados a humanos não apresentam, explicitamente, informações que podem ser deduzidas pelos consulentes, por exemplo: o dicionário Houaiss (2001) não apresenta entradas de advérbios terminados com o sufixo –*mente*. Há um verbete para o sufixo –*mente* com as seguintes informações:

-**mente** *suf.* formador de advérbios, do lat. *mens, mentis* 'espírito, alma, razão, sabedoria', us. no lat. vulg. no abl., quase sempre com um adj., donde locuções modais do tipo *fortemente* 'de alma decidida', *bona mente* 'com boa vontade' etc., torneios que passaram para várias línguas român.; no port. arcaico, ocorrem ainda f. graficamente separadas dos dois el., mas a partir do sXVI sua juntada se faz

progressivamente regular; por sua orig. fem., a formação até o presente é fiel à adunção das f. femininas ao suf., com a ressalva, porém, de que, nos adj. uniformes presentes e passados, perdura a uniformidade; [...].

(HOUAISS, 2001: 1895, grifo do autor).

A informação morfológica, de que o sufixo *-mente* forma advérbios a partir de adjetivos, agregada ao significado do adjetivo que origina o advérbio permite ao consulente deduzir o significado de palavras como: *calmamente*, *sossegadamente* etc., tornado-se portanto dispensável a inclusão de entradas para esses itens lexicais.

O conteúdo dos dicionários digitais é da mesma natureza que as informações contidas na versão impressa dessas obras. Os avanços tecnológicos permitiram a elaboração de versões de dicionários de uso em suporte digital o que proporcionou um aumento significativo na quantidade de informações disponíveis. A versão *on line* do dicionário Houaiss, por exemplo, apresenta na janela das entradas de verbos um caminho para uma tabela de conjugação do referido verbo. Porém, a busca deve ser feita a partir da forma canônica do item lexical, no caso dos verbos, no infinitivo. O resultado da busca por um verbo conjugado, por exemplo, será negativo. Esse dicionário é, pois, elaborado para um tipo de usuário capaz de fazer deduções a partir de regras e identificar determinados fatos linguísticos.

5.2 – Dicionário eletrônico

O dicionário eletrônico é concebido para ser utilizado pela máquina, sem a intervenção humana, em processamento de linguagem natural. As informações contidas nesse tipo de dicionário são, obrigatoriamente, formalizadas e codificadas, isto é, descritas de forma tal que possam ser lidas e ‘compreendidas’ pela máquina na análise automática de textos. As palavras são tratadas de forma que abarque o maior número de informações possíveis, visto que a máquina não é capaz de fazer deduções a partir de regras a não ser que estas sejam previamente inseridas no sistema, bem como as orientações para o acesso a elas. São consideradas as informações acerca das variáveis dos itens lexicais, como: gênero, número,

grau, pessoa, tempo, modo, regência verbal ou nominal, descrições semânticas e outras que se fizerem necessárias ao bom funcionamento do sistema.

Gross (1989: 4) afirma que “les dictionnaires électroniques entrent dans la plupart des programmes qui mettent en jeu une procédure de reconnaissance d’unités linguistiques significantes”. Os programas que envolvem o reconhecimento das unidades linguísticas utilizam dicionários eletrônicos. Entre as muitas aplicações desse tipo de dicionário está a recuperação de informação. Por exemplo, ao inserirmos em um programa de busca na *internet* o nome *Espírito Santo*, o resultado obtido incluirá todas as ocorrências que contêm as palavras *espírito* e *santo*, independentemente da ordem em que elas aparecem, e sem levar em conta se se trata de um nome próprio ou de um item lexical composto. Há a possibilidade de restrição da busca, através da inserção de símbolos apropriados que levam o sistema a buscar a expressão solicitada como um bloco indissolúvel. Esse recurso resolve em parte o problema. No exemplo dado o programa de busca poderia eliminar resultados com os itens *espírito* e *santo* isoladamente, mas seriam mantidas todas as ocorrências de *espírito santo* tanto em referência à *parte imaterial do ser humano e aquilo/aquela que é considerado essencialmente puro, soberanamente perfeito* como em relação ao estado brasileiro da região Sudeste, quando a busca refere-se a apenas ao nome próprio. O uso do dicionário eletrônico minimiza o problema aqui apresentado, por conter descrições mais precisas acerca das palavras e contemplar além das palavras simples, as palavras compostas.

Destacamos entre as ferramentas computacionais disponíveis atualmente o Unitex. Segundo Paumier (2006: 9) “Unitex est un ensemble de logiciels permettant de traiter des textes en langues naturelles en utilisant des ressources linguistiques.” Esse conjunto de *softwares* destinados a lidar com textos em língua natural utiliza os recursos linguísticos, que vêm em forma de dicionário eletrônicos, gramáticas e tabelas léxico-gramática. Os dicionários são representados segundo o formalismo DELA (Dicionários Eletrônicos do *Laboratoire d’automatique documentaire et linguistique* – LADL).

Os dicionários DELA podem ser de palavras simples (DELAS) ou de palavras compostas (DELAC). Esses dicionários podem ser de formas flexionadas (DELAF – palavras simples flexionadas [figura 1] e DELACF – palavras compostas flexionadas [figura 2]). As entradas lexicais do DELAS são de formas canônicas à semelhança dos dicionários tradicionais, ou seja, verbo no infinitivo, substantivo e adjetivo no masculino singular etc.

```

rindo, rir.V:G
ringir, .V:W1s:W3s:U1s:U3s
rio, .N:ms
rio, rir.V:P1s
riqueza, .N:fs
rir, .V:W1s:W3s:U1s:U3s
risada, .N:fs
risadas, risada.N:fp
risadinha, risada.N:Dfs
riscar, .V:W1s:W3s:U1s:U3s
risco, .N:ms
risco, riscar.V:P1s
riscos, risco.N:mp

```

Figura 1 – Parte da lista das entradas do DELAF aplicado ao texto *Senhora* de José de Alencar.

```

guarda-roupa, .N+VN:ms
guarda-roupas, guarda-roupa.N+VN:mp
lua-de-mel, .N+NDN:fs
lusco-fusco, .N+NA:ms
mata-borrão, .N+VN:ms
meia-noite, .N+AN:fs
meio-dia, .N+AN:ms
mil-réis, .A+DETN:ms:mp:fs:fp
pobre-diabo, .N+AN:ms
prima-dona, .N+AN:fs
quarta-feira, .N+DETN:fs
queima-roupa, .N+MN:fs
sem-cerimônia, .N+PM:fs

```

Figura 2 – Parte da lista das entradas do DELACF aplicado ao texto *Senhora* de José de Alencar.

As figuras 1 e 2 apresentam, respectivamente, entradas dos dicionários de palavras simples e do dicionário de palavras compostas do sistema DELA. As entradas dos itens lexicais, como aparecem no texto, são registradas à esquerda da janela (em letra azul), seguidas de vírgula (,) e ponto (.), quando essas ocorrências correspondem à forma canônica da palavra. Quando o item lexical ocorre no texto de forma flexionada, após a vírgula é registrada (em fonte vermelha) a forma canônica da palavra. Os códigos em fonte verde, após o ponto, indicam a classe a que a palavra pertence e outras informações gramaticais. O sinal de mais (+) é usado como separador entre as informações como se observa em

guarda-roupa,.N+VN:ms

o *N* indica a classe a que pertence a palavra composta (*nome*) e as letras *V* e *N* indicam a classe a que pertencem os elementos constituintes desse composto (*verbo* e *nome*). Já em

este,.PRO+Dem:ms

a informação (*Dem*) após o sinal de mais (+) apresenta a sub-categorização do item lexical.

Após os dois pontos (:) são inseridas as informações flexionais (*masculino e singular* nos exemplos apresentados).

A organização de um dicionário eletrônico aponta para a necessidade de constante atualização, como a inserção de novas palavras (simples e compostas) que passam a integrar o léxico. A expansão do sistema de dicionário DELA inclui também a inserção de dados linguísticos. A cada grupo de palavra é associado um tipo de informação (por exemplo, informações morfológicas acerca de *verbos* – formas conjugadas ou não; de *nomes e adjetivos* – singular e plural). As palavras homônimas recebem entradas distintas por apresentarem informações gramaticais diferentes, o que facilita a inserção de novos dados. Um dicionário eletrônico pode conter apenas informações morfológicas, neste caso não haverá, por exemplo, distinção das informações atribuídas ao item lexical *pregador* com o sentido de *aquele que faz pregações* e *pregador* com o sentido de *aquilo que serve para pregar*. A etiqueta atribuída aos dois sentidos da palavra será:

pregador,.N:ms

pois, os dois itens lexicais apresentam as mesmas características morfológicas, contudo, existe a possibilidade de expansão do dicionário, do acréscimo, por exemplo, de informações semânticas. As entradas distintas das diferentes acepções do item lexical *pregador* facilita o acréscimo de informações, como se observa no exemplo:

pregador,.N:ms/aquele que faz pregações⁷

pregador,N:ms/aquilo que serve para pregar

O fato das entradas nos dicionários do sistema DELA serem independentes facilita a inserção de novas informações linguísticas nas entradas e a inclusão de novos itens lexicais.

Gross (1989), ao tratar da organização dos dados lexicográficos e sintáticos no dicionário eletrônico do LADL, explica que as formas linguísticas são divididas em dois grupos: vocabulários correntes e léxicos satélites. Os léxicos satélites correspondem aos vocabulários

⁷ Nas entradas dos dicionários do sistema DELA os comentários opcionais são antecidos pelo símbolo /.

científicos e técnicos (nomes simples e nomes compostos), aos vocabulários e formas sintáticas dialetais. O autor afirma que

On notera bien qu'il existe une imbrication importante du noyau et des satellites : un mot donné peut avoir une ou plusieurs entrées (sens et construction) dans le noyau, et d'autres dans des vocabulaires satellites. En particulier, on a rencontré cette situation avec les noms propres et les sigles: d'une part, il est souhaitable de disposer de dictionnaires électroniques des noms de pays par exemple, mais alors il faudra y associer (dans le DELAS ?) des mots dérivés [...].

(Gross, 1989: 18).

A observação de que uma determinada palavra pode ter uma ou mais entradas (significado e construção) no núcleo e em outros vocabulários satélites é exemplificada com topônimos (nomes próprios de país). Para Gross (1989) é desejável a construção de dicionários eletrônicos que abarquem os nomes de países, por exemplo, mas ele deverá ser associado (ao DELAS?) às palavras derivadas, assim tem-se *França* no dicionário de nomes próprios e, por outro lado, *francês, francófono* etc., no dicionário de palavras comuns. Pode-se estender essa observação aos compostos toponímicos e seus derivados como (estado do) *Mato Grosso* e *mato-grossense* ou (município de) *Vila Velha* e *vila-velhense*, por exemplo.

A busca nos dicionários usuais é feita através da inserção de um item lexical descontextualizado, nos dicionários eletrônicos a análise dá-se a partir de um texto a ser processado. Nos programas de reconhecimento de unidades lexicais, por exemplo, o texto é inicialmente segmentado em sentenças (figura 3) e o dicionário eletrônico é consultado a fim de definir a natureza do item lexical. Se uma palavra não consta no dicionário eletrônico o seu reconhecimento torna-se inviável nesse processo e conseqüentemente a análise da frase em que está inserido o item lexical é bloqueada ou quando realizada apresenta um resultado impreciso.

```

Senhora
{S}José de Alencar
{S}Projeto Lácio-WEB
{S}Disponível em: http://www.nilc.icmc.usp.br/lacioweb/
{S}Ao Leitor
{S}Este livro, como os dois que o precederam, não são da própria lavra do escritor, a quem geralmente os atribuem.
{S}A história é verdadeira:{S} e a narração vem de pessoa que recebeu diretamente, e em circunstâncias que ignoro, a confiança dos principais atores deste drama curioso.
{S}O suposto autor não passa rigorosamente de editor.{S} É certo que tomando a si o encargo de corrigir a forma e dar-lhe um lavor literário, de algum modo apropria-se não a obra mas o livro.
{S}Em todo o caso, encontram-se muitas vezes nestas páginas, exuberâncias de linguagem e afoutezas de imaginação, a que já não se lança a pena sóbria e refletida do escritor sem ilusões e sem entusiasmos.

```

Figura 3 – Fragmento do texto *Senhora* de José de Alencar, pré-processado pelo sistema Unitex 2.0.

O reconhecimento dos compostos toponímicos e a distinção entre estes e os compostos comuns é desejável para que o processamento de textos em língua natural seja mais preciso. O topônimo *Rio de Janeiro*, por exemplo, não figura na lista das entradas do DELACF aplicado ao texto *Senhora* de José de Alencar (figura 2). O substantivo *rio* consta da lista das entradas do DELAF aplicado ao texto (figura 1) assim como o substantivo *janeiro*. Esses nomes, porém, escritos com letra inicial maiúscula (*Rio* e *Janeiro*), estão na listagem do arquivo *err*, arquivo de formas desconhecidas, formas que não constam da lista de palavras simples nem da lista de palavras compostas.

Observa-se (grafo da figura 4) que o topônimo *Rio de Janeiro* não foi etiquetado como um nome próprio composto, suas partes foram etiquetadas isoladamente. O sistema Unitex viabiliza a construção de grafos como os das figuras 4 e 5.

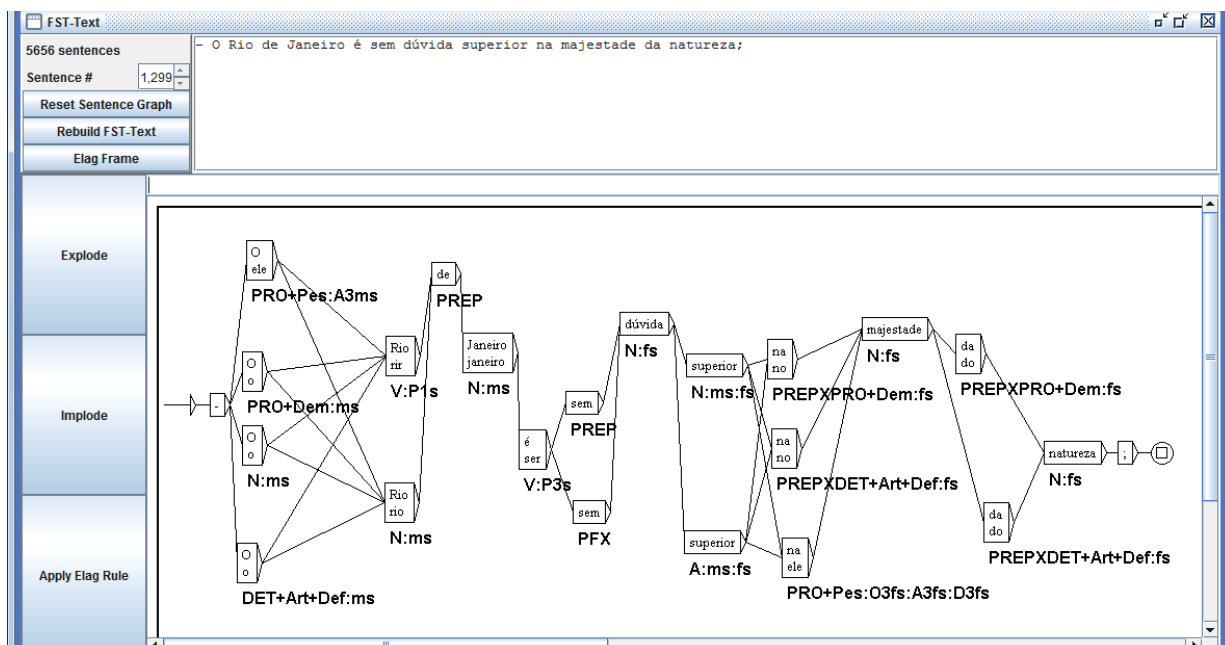
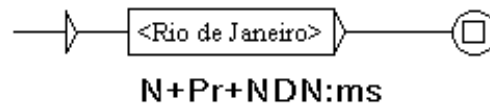


Figura 4 – Grafo de um fragmento do texto *Senhora* de José de Alencar.

De acordo com Paumier (2006: 67) “Le symbole en forme de flèche est l'état initial du graphe. Le symbole rond contenant un carré est l'état final du graphe. La grammaire ne reconnaît que les expressions décrites par des chemins reliant l'état initial à l'état final”. A gramática reconhece, portanto, as expressões que constam da descrição do caminho entre o estado inicial (simbolizado pela seta na extremidade esquerda do grafo – figuras 4 e 5) e o estado final (simbolizado pela circunferência com um quadrado circunscrito na extremidade direita do grafo – figuras 4 e 5).



Rio de Janeiro.grf

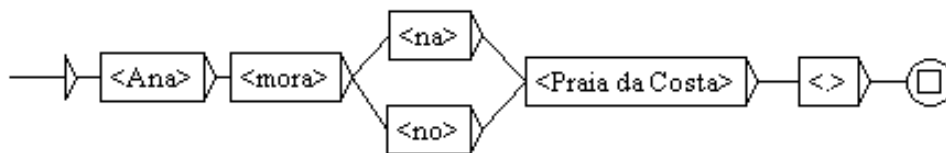
Figura 5 – Grafo do composto toponímico *Rio de Janeiro*.

Através dos grafos é possível registrar algumas variações como em

(1) Ana mora no **bairro Praia da Costa**.

(1a) Ana mora na **Praia da Costa**.

(1b) Ana mora no **Praia da Costa**.



Ana mora na+no Praia da Costa.grf

Figura 6 – Grafo das frases (1a) *Ana mora na Praia da Costa* e (1b) *Ana mora no Praia da Costa*.

A inclusão dos compostos toponímicos no DELACF, a etiquetagem desses itens lexicais como nomes próprios é um fator de otimização dos sistemas informáticos que se utilizam dos dicionários do sistema DELA. A atribuição da etiqueta $\langle N+PR+NDN:ms \rangle$ ao composto toponímico *Rio de Janeiro* permite que ele seja identificado adequadamente (como um composto e não como um grupo nominal livre). O dicionário eletrônico de nomes compostos toponímicos do português do Brasil colabora, pois, de maneira significativa para o bom desempenho dos sistemas informáticos.

O essencial na construção de léxicos e gramáticas formalizadas é a pesquisa do conhecimento dos mecanismos linguísticos que são os mecanismos gerais e indispensáveis para a descrição de uma língua natural.

6 – CRITÉRIOS DE IDENTIFICAÇÃO DOS NOMES TOPONÍMICOS

Os itens lexicais selecionados apresentam transparência semântica parcial, visto que um dos elementos formadores do composto descreve o acidente geográfico a que se refere o topônimo. Esse fato pode gerar dificuldade quanto à identificação da não-composicionalidade do significado dos topônimos. Afirmar que o significado de uma palavra é não-composicional significa dizer que ele é deduzido do todo, do conjunto de elementos constituintes, e não da soma dos significados desses elementos. Embora o primeiro substantivo dos compostos toponímicos selecionados apresente transparência semântica, o significado não é deduzido desse elemento isoladamente, mas do conjunto. O topônimo *Praia da Costa*, por exemplo, nomeia um acidente artificial (bairro) que tem como referência uma *praia*, logo, fica na *costa*, no litoral do Espírito Santo. O nome *Praia da Costa* não se refere a todo e qualquer bairro situado próximo a uma praia, que se desenvolveu na costa capixaba, mas designa um acidente artificial específico que tem como limites: a baía de Vitória, o bairro Jardim Itapoã, o oceano Atlântico e o centro de Vila Velha. Apesar da transparência semântica parcial do nome *Praia da Costa*, o seu significado não é deduzido das partes que o constitui, mas é deduzido do todo, trata-se de um nome não-composicional.

A identificação dos grupos nominais é feita através da aplicação de testes que comprovam as propriedades morfológicas, sintáticas e semânticas dos topônimos candidatos a compostos a partir de frases elementares, ou seja, sentenças constituídas de *sujeito*, *verbo* e *complementos essenciais*.

Dick (1990: 10b) explica que 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). O termo específico por sua vez pode ser constituído por itens lexicais de categorias gramaticais diversas que funcionam como elemento distintivo do termo genérico. O termo específico pode ser, por exemplo, um substantivo (córrego *Batata* – Viana – ES), um adjetivo (praia *Grande* – Vila Velha – ES) e até mesmo um verbo (cidade de *Varre e Sai* – RJ) e pode ser constituído de um ou mais elementos, unidos ou não por preposição, como: morro *São Benedito* – Vitória – ES (Npr =: Adj+N), córrego *Doutor Robson* – Serra – (Npr =: N+N), canal *Praia da Costa* – Vila Velha – ES (Npr =: N+prep+N).

De acordo com Bechara (2009: 355),

A COMPOSIÇÃO consiste na criação de uma palavra nova de significado único e constante, sempre e somente por meio de dois radicais relacionados entre si. Isso não impede que um dos elementos do composto seja ele mesmo já um composto, contando como um termo único, pelo princípio dos constituintes imediatos (grifo do autor).

Dessa forma, consideramos como elementos integrantes do conjunto de compostos NN os nomes compostos em que o N2 é, também, constituído por mais de um substantivo. Os topônimos de estrutura NNN (*córrego Doutor Robson*) são contados como parte da classe de nomes compostos NN, assim como os nomes NdeNcomp (*Morro do Pão de Açúcar – Vila Velha – ES*) são parte da classe NdeN. Sandmann (1988: 124), por exemplo, inclui entre os compostos predicativos do tipo NN o nome *Ministro-Chefe do Gabinete Militar*, em que o N2 é constituído de N+prep+art+N+Adj.

Os nomes analisados neste capítulo apresentam a estrutura geral dos topônimos, isso é: *nome comum* (Nc) + *nome próprio* (Npr), em que Npr pode ser constituído por um ou mais elementos lexicais (Npr =: N [córrego *Carapebus* – Serra – ES]; Adj [praia *Grande* – Vila Velha – ES]; NN [córrego *Doutor Robson* – Serra – ES]; AdjN [morro *Santa Clara* – Vitória – ES]; NAdj [ribeirão *Ponto Doce* – Vila Velha – ES]; NdeN [serra *Pé de Urubu* – Cariacica – ES]).

Para Menezes e Santos (2008) nomes geográficos e topônimos são sinônimos. Os autores afirmam que

A forma de escrever, como escrever, ou seja, a sua grafia, pode alterar substancialmente o seu significado. Uma vez que computacionalmente qualquer alteração por mínima que seja, pode gerar um nome diferente em uma base de dados, justifica que haja uma padronização, seja na grafia dos nomes existentes, como também em uma orientação para a criação de novos nomes geográficos, notadamente na esfera executiva, através de nomes administrativos.

(MENEZES e SANTOS, 2008: 77).

A pesar da recomendação dos geógrafos, o que se percebe quanto à ortografia desses nomes é a falta de consenso entre os estudiosos. Se o primeiro item dos sintagmas toponímicos apresenta o estatuto de nome comum, deduz-se que ele deveria ser grafado com letra inicial minúscula, entretanto, Rocha Lima (2005) e Bechara (2009) grafam o primeiro termo do composto toponímico com inicial maiúscula. Diante dessa divergência optamos por grafar os compostos toponímicos de acordo com a forma registrada no *corpus*. Nos exemplos construídos registramos o primeiro item do composto toponímico com inicial minúscula, por

entendermos tratar-se de um nome comum. Essa opção baseia-se no *status* de nome comum que o termo genérico apresenta e não em juízos de valor. Ullmann (1964: 151) afirma que: “Numa novela de F. Molnár, um membro de uma quadrilha de rapazes, acusado de um comportamento indigno, sofre a suprema humilhação de o seu ‘pobre o honesto nome’ escrito com uma inicial minúscula”. A utilização de inicial maiúscula torna-se, neste caso, um critério subjetivo e pouco eficiente na identificação do nome próprio.

A falta de consenso quanto à ortografia dos topônimos não se restringe ao uso ou não de inicial maiúscula. Ocorre, em alguns casos, que o poder público atribui a um espaço físico, determinado nome através de lei ou decreto (nomeação sistemática), porém, por motivos diversos, o topônimo nem sempre é acatado pela população, que utiliza outro nome para designar o mesmo lugar. Esta nomeação espontânea, conferida a um lugar por sujeitos anônimos, passa a coexistir com a nomeação sistemática, de forma que o mesmo referente passa a ser portador de nomes distintos. Constatamos esse fato na seguinte sentença:

(1) “O corpo de Edilma Maria Ferreira, 26, foi localizado próximo ao **viaduto Oscar Brito**, também conhecido como **viaduto dos Cabritos**”.

O viaduto localizado na cidade do Rio de Janeiro, bairro de Campo Grande, embora tenha, sistematicamente, recebido o nome de *viaduto Engenheiro Oscar Brito*, é popularmente conhecido como *viaduto dos Cabritos*, possivelmente pela semelhança fônica que se observa entre os grupos nominais *os Cabritos* e *Oscar Brito*. Na elaboração de um dicionário eletrônico de compostos toponímicos do português do Brasil, os nomes que fazem referência à mesma realidade espacial, mas apresentam composição e comportamento linguístico distintos são registrados em entradas diferentes. Os topônimos *viaduto dos Cabritos* e *viaduto Engenheiro Oscar Brito* são exemplo dessa situação. Vale ressaltar a elipse do substantivo *Engenheiro* no topônimo *viaduto Engenheiro Oscar Brito*. Nesses casos – de elipse de um dos itens que constitui os compostos toponímicos – registramos a forma longa do nome e consideramos as ocorrências de elisão como variações do nome.

Acerca dos nomes de lugar Constant (2003: 126) considera “ces objets linguistiques comme des formes composées comprenant un classifieur de lieu [...]. Par exemple, *Méditerranée* est la forme réduite de *mer Méditerranée* et a un comportement syntaxique différent de *mer du Nord*” (grifo do autor). Adotamos a mesma postura nesta pesquisa. A diferença de comportamento

sintático dos diversos compostos toponímicos leva a aplicação de critérios diferenciados para a constatação das propriedades das estruturas.

Quanto ao determinante dos nomes de ilhas, cabe ressaltar a observação de Garrigues (1993: 293-294) de que “la répartition des noms d’îles est un [...] cas beaucoup plus complete” e que “le système des classifieurs d’îles ne peut donc être construit que par l’examen de chaque cas”. Essas afirmações são válidas também para o português do Brasil. A distribuição dos nomes de ilhas é bastante complexa e o sistema de classificação dos determinantes deve considerar caso a caso. Os nomes de ilhas podem apresentar ausência do determinante, determinante no singular e determinante no plural, como nas frases a seguir:

(2) “**Ilha das Caieiras** terá museu com histórias contadas pelos moradores”.

(3) “Lula passará réveillon na **Ilha de Fernando de Noronha**”.

(4) “Caso das **Ilhas Cayman** reforça sentença contra Dantas”.

O plural do elemento genérico *ilha* geralmente deve-se à ocorrência de homonímia entre as diversas ilhas que constituem determinado arquipélago. A variação em número associa-se, portanto, à natureza decomponível do termo genérico. A ausência ou a presença do artigo definido por sua vez relaciona-se ao contexto linguístico em que o nome se realiza. O nome *ilha das Caieiras* figura tanto na frase (2) como na frase (5):

(2) “**Ilha das Caieiras** terá museu com histórias contadas pelos moradores”.

(5) “Obra de drenagem vai desviar trânsito na **Ilha das Caieiras**”.

Contudo, em (5) a anteposição da preposição *em* ao topônimo (na função de complemento) propicia a inserção do artigo definido.

6.1 - Topônimos com a estrutura NN

Segundo Batista (1994: 68), “A identificação dos *NN* não implica uma distinção entre grupos nominais livres e nomes compostos, já que não há, na língua, grupos nominais livres com esta estrutura interna”, o que não significa afirmar a total inexistência de grupos nominais livres formados por dois substantivos.

De acordo com Vivès (1990: 98):

On ne peut pas confronter le status des deux derniers noms dans chacune de ces deux phrases:

On a élu ce concierge président

On a élu cet ingénieur chimiste.

Dans la première phrase *président* est attribut de *concierge*, dans la seconde, *ingénieur chimiste* ne forme qu’une seule unité syntaxique, commutable d’ailleurs avec un mot simple. Il y a clairement ellipse d’un argument du verbe dans la seconde phrase [...].

O autor observa que na primeira frase (*Elegeu-se este zelador presidente*) o nome *presidente* é atributo de *zelador*. Os substantivos *zelador* e *presidente* são argumentos do verbo *eleger*, assim como o sujeito. Na segunda frase (*Elegeu-se este engenheiro químico*) o grupo NN funciona como uma unidade e pode ser substituído por uma palavra simples, ou seja, nesta frase o verbo *eleger* conta com apenas dois argumentos (*ingénieur chimiste* e o sujeito). A identificação do número de argumentos do verbo *eleger* nas frases apresentadas confirma o fato de que *engenheiro químico* funciona como uma unidade, um nome composto.

Destacamos entre os topônimos candidatos a compostos que apresentam a estrutura NN: *córrego Montanha* (município de Serra - ES), *rio Aribiri* (município de Vila Velha - ES), *ilha Itatiaia* (município de Vila Velha - ES) e *morro São Benedito* (município de Vitória - ES). Os critérios formais utilizados na identificação das propriedades morfossintáticas e semânticas dos topônimos NN candidatos a compostos são:

- ✓ Elisão de um dos substantivos.
- ✓ Inserção de um item lexical no grupo nominal.
- ✓ Ruptura paradigmática.
- ✓ Variação em número.

6.1.1 - Elisão de um dos substantivos

Sandmann (1988) apresenta entre os nomes compostos de estrutura NN alguns nomes próprios em que N1 é o determinado e N2 o determinante: *Rede Manchete*, *Óticas Brasil*, *Camisaria Novo Mundo*, *Colégio Andrews*, *Bar Amarelinho* e outros em que N1 é o determinante e N2 o determinado, entre eles: *Chico's Bar*, *Pão de Açúcar Well's*, *Antonio's* e *Alvaro's*. Nestes exemplos observa-se a presença do genitivo anglo-saxão ('s) ao final do N1 e nos dois a elipse do termo genérico. Carvalhinhos e Antunes (2007) explicam que pode ocorrer a elisão do termo genérico, principalmente nos nomes dos aglomerados humanos.

Os nomes compostos, em geral, não admitem a elisão de um de seus elementos constituintes. Nos topônimos selecionados, porém, há possibilidade de elisão do N1, pois, sendo N2 o elemento identificador de N1 pode-se inferir, em alguns casos, o N1 a partir do N2, como nos exemplos a seguir:

(6) Os madeireiros acamparam às margens do **córrego Montanha**.

(6a) Os madeireiros acamparam às margens do **Montanha**.

(6b) Os madeireiros acamparam às margens do **córrego**.

Inferre-se em (6a) a elipse do termo genérico (*córrego* - substantivo masculino) pela presença do artigo masculino *o* anteposto a um substantivo do gênero feminino (*Montanha*). A presença substantivo *margem* e do artigo masculino anteposto ao nome *Montanha* possibilita a desambiguação do próprio nome, isso é, colabora para que o nome *Montanha* possa seja identificado como elemento que nomeia o *córrego* e não como o substantivo comum que designa uma *elevação significativamente alta*. Em (6a) o N2 mantém o significado geral do composto (*córrego Montanha*).

Pode-se considerar (6b) como uma variante de (6) pela ocorrência de um apagamento bastante regular na língua, (6b) pode ser também variante de outras frases. Como frase isolada (6b) é ambígua, pois não há elementos linguísticos na frase que identifiquem o *córrego*. Contudo, a frase é aceitável e gramaticalmente adequada.

De acordo com Baptista (1994: 71):

Nos casos em que o segundo substantivo funciona como um classificador do primeiro, a elisão de *N2* é por vezes possível, ficando *N1* com um significado genérico:

[...]

Podemos, pois, considerar que, de um modo geral, *N1* mantém na combinação *NN* o seu significado, como nos indica a aceitabilidade das frases definidoras, onde é possível inserir o determinante *um tipo de* (grifo do autor).

Nos topônimos de estrutura *NN* seleccionados, observa-se a transparência semântica de *N1*, (que funciona como elemento genérico) através da elaboração de frases definidoras como nos exemplos a seguir:

O **córrego Montanha** é um córrego.

O **rio Aribiri** é um rio.

A **ilha Itatiaia** é uma ilha.

O **morro São Benedito** é um morro.

Em alguns casos há a possibilidade de elisão de um dos substantivos que formam o composto, sem detrimento da compreensão da sentença.

(7) “O **rio Aribiri** desaguava na baía de Vitória”.

(7a) O **Aribiri** desaguava na baía de Vitória.

(7b) O **rio** desaguava na baía de Vitória.

Embora haja, no município de Vila Velha, o bairro Aribiri, o verbo *desaguar* posposto ao topônimo *Aribiri* – em (7a) – sugere que este se refere a um curso de água, não a um acidente artificial.

A elisão de *N2*, em (7b), assim como em (6b) impossibilita a identificação do elemento genérico ao qual o enunciador refere-se. Pode ser que apenas o *rio Aribiri* deságue na baía de Vitória, porém, esse conhecimento é pertinente a geografia, portanto, externo aos propósitos desta pesquisa. A frase (7b) é perfeitamente aceitável, embora seja ambígua.

Nos compostos das sentenças (6) e (7), o *N2* é o elemento determinante de *N1*. O mesmo acontece com o exemplo (8):

(8) “Outra alternativa é [...] visitar a **Ilha Itatiaia**”.

(8a) Outra alternativa é visitar a **Itatiaia**.

(8b) Outra alternativa é visitar a **Ilha**.

A sentença (8a) é gramatical e aceitável, embora haja em (8a) falta de informação importante para o leitor não nativo da região. Apesar de ser ambígua (8b) pode ser considerada uma variação de (8) em que há o apagamento de N2.

Said Ali (1964: 43, 68) afirma que: “Aos nomes de ilhas usados uns com artigo, outros sem ele, correspondem qualificativos e pronomes do gênero feminino referido ao vocábulo ‘ilha’”. Isso é, “os adjetivos e pronomes irão para o feminino de acordo com o termo geográfico não expresso. Diz-se contudo *os Açôres, os Abrolhos*” (grifo do autor). A elipse do N1 em (8a) pode acarretar ausência do artigo definido:

(8a') Outra alternativa é visitar **Itatiaia**.

Em (8a') a ausência do artigo definido é admissível, visto que na língua portuguesa do Brasil não há obrigatoriedade de uso de determinante diante de nomes próprios. Quando ocorre a elisão do N2, observa-se a necessidade do uso de determinante que individualize o acidente geográfico em questão, como em (8b). Assim sendo, é inaceitável a frase:

(8b') *Outra alternativa é visitar **Ilha**.

Nas frases seguintes, embora a estrutura geral do topônimo (NcNpr) seja a mesma das anteriores (6), (7) e (8), o Npr ou N2 é um nome composto.

(9) “**Morro São Benedito** tem feira comunitária ambiental”.

(9a) **São Benedito** tem feira comunitária ambiental

(9b) **Morro** tem feira comunitária ambiental.

A sentença (9a) é considerada aceitável. Embora não haja na grande Vitória outra região nomeada de *São Benedito*, senão o *morro*, a aceitabilidade da frase (9a) não se resume a esse fato. Ainda que existisse outra região da Grande Vitória com o mesmo nome, o apagamento de *morro* poderia ser considerado aceitável, produzindo uma ambiguidade que não impede o

funcionamento da língua. Mesmo uma sequência que não pode ser inferida do contexto pode sofrer apagamento e gerar uma sentença aceitável.

Em (9b) a elisão do N2 faz com que a informação contida na sentença seja incompleta, pela falta de elementos que possam identificar de forma mais precisa de qual *morro* da cidade de Vitória trata o texto, esse fato, contudo não acarreta a inaceitabilidade da frase.

Os exemplos (9a) e (9b) apresentam a elipse de um dos elementos que constituem a formação NcNpr, isso é, em (9a) ocorre o apagamento do Nc e em (9b) do Npr. Porém, o Npr é também um nome composto constituído por um adjetivo e um substantivo. Nas sentenças (9c) e (9d), efetuamos a elisão de um dos elementos que constitui o composto Npr =: AdjN.

Cabe ressaltar que o composto *São Benedito* é formado por um substantivo precedido de adjetivo (AdjN). Segundo Dick (1997: 182): “Como o adjetivo, lingüisticamente, carece de uma autonomia sintagmática, liga-se, via de consequência, ao substantivo que qualifica, acarretando uma incorporação de formas que acabam por se transmitir ao topônimo, não necessária mas essencialmente”. A explicação da autora refere-se à anteposição do adjetivo ao substantivo e a sua incorporação ao topônimo, conforme se observa no topônimo *São Benedito* – frase (9). Quanto aos compostos que apresentam a estrutura AdjN, Baptista (1994: 59) informa: “verifica-se que constituem unidades lexicais com uma elevada fixidez sintáctica”, as sentenças (9c) e (9d) atestam esse fato.

(9) **Morro São Benedito** tem feira comunitária ambiental.

(9c) ***Morro São** tem feira comunitária ambiental.

(9d) ***Morro Benedito** tem feira comunitária ambiental.

A elisão do adjetivo *São* em (9c) e do substantivo *Benedito* em (9d) gera sentenças inaceitáveis, pois, os dois últimos itens lexicais da estrutura NcNpr são fixos, representam uma composição, assim, a elisão de um dos elementos constituintes do Npr acarreta na formação de compostos toponímicos inaceitáveis.

6.1.2 – Inserção de um item lexical no grupo nominal

Em um grupo nominal livre é possível a inserção de um item lexical entre os elementos do grupo nominal. A não aceitabilidade da inserção de um adjetivo entre os itens que constituem o composto NN, ou que se refira especificamente a cada um dos nomes, caracteriza a fixidez dos compostos *córrego Montanha*, *rio Aribiri*, *Ilha Itatiaia* e *Morro São Benedito*.

(6) Os madeireiros acamparam às margens do **córrego Montanha**.

(6c) *Os madeireiros acamparam às margens do **córrego belo Montanha**.

(6d) *Os madeireiros acamparam às margens do **córrego alta Montanha**.

(7) “O **rio Aribiri** desaguava na baía de Vitória”.

(7c) *O **rio caudaloso Aribiri** desaguava na baía de Vitória

(8) Outra alternativa é visitar a **Ilha Itatiaia**.

(8c) *Outra alternativa é visitar a **Ilha grande Itatiaia**.

(9) **Morro São Benedito** tem feira comunitária ambiental

(9e) ***Morro alto São Benedito** tem feira comunitária ambiental

(9f) ***Morro São simpático Benedito** tem feira comunitária ambiental.

Na sentença (9f) fica caracterizada também a fixidez da estrutura do Npr (o hagiônimo⁸ *São Benedito*).

Os adjetivos *belo* e *alta* em (6c) e (6d) referem-se aos substantivos *córrego* e *Montanha*, respectivamente. Nota-se que, embora o topônimo *Montanha* tenha origem em um nome comum ele é destituído de algumas de suas características morfossintáticas e semânticas ao assumir o estatuto de nome próprio. Pode-se afirmar que *córrego Montanha* é um córrego, porém, *córrego Montanha* não é uma montanha. Em (6d) o adjetivo *alta* é inserido entre os

⁸ Hagiônimo - nome próprio constituído pelo nome de um santo.

elementos que constituem o composto, porém, mesmo a sua posposição ao N2 constitui uma sentença inaceitável.

(6e) *Os madeireiros acamparam às margens do **córrego Montanha alta**.

A inaceitabilidade da posposição do adjetivo ao Npr decorre do estatuto de nome próprio que *Montanha* apresenta e, conseqüentemente, seu funcionamento linguístico diferenciado do nome comum. Outro fato a ser observado é a não composicionalidade do significado de *córrego Montanha*, isso é, o significado deve ser deduzido do todo, não das partes constituintes. Os nomes *córrego* e *Montanha* formam um composto, não sendo possível atribuir qualidades a esses itens de forma isolada. Em alguns casos é possível a anteposição do adjetivo, quando este se refere ao composto de forma global:

(6f) Os madeireiros acamparam às margens do **belo córrego Montanha**.

(7d) O **caudaloso rio Aribiri** desaguava na baía de Vitória.

(8d) Outra alternativa é visitar a **grande/paradisíaca Ilha Itatiaia**.

(9g) O **conhecido Morro São Benedito** tem feira comunitária ambiental.

A anteposição do item lexical *conhecido* ao nome (*Morro São Benedito*) em (9g) torna necessária a presença do artigo definido (*o*) no início da sentença por uma questão de adequação gramatical.

6.1.3 – Ruptura paradigmática

Batista (1994) afirma que, “na maior parte dos casos, os nomes compostos *NN* não admitem qualquer substituição dos substantivos da combinação por outros elementos [...]”, o mesmo observa-se nas frases a seguir:

(6) Os madeireiros acamparam às margens do **córrego Montanha**.

(6g) *Os madeireiros acamparam às margens do **rio Montanha**.

(6h) *Os madeireiros acamparam às margens do **córrego Monte**.

(7) O **rio Aribiri** desaguava na baía de Vitória.

(7e) * O **córrego Aribiri** desaguava na baía de Vitória.

(7f) * O **rio Itaparica** desaguava na baía de Vitória.

Em (6g), (6h), (7e) e (7f) foram utilizados elementos de natureza semântica semelhante na substituição de N1, contudo, as sentenças geradas são inaceitáveis. Os topônimos formados não são elementos linguísticos institucionalizados, reconhecidos legalmente como nomes de lugar, sequer são fruto da nomeação espontânea, reconhecida pelo uso. O mesmo acontece nos exemplos a seguir:

(8) Outra alternativa é visitar a **Ilha Itatiaia**.

(8e) *Outra alternativa é visitar a **Montanha Itatiaia**.

(8f) *Outra alternativa é visitar a **Ilha Itaparica**.

(9) **Morro São Benedito** tem feira comunitária ambiental.

(9h) ***Montanha São Benedito** tem feira comunitária ambiental.

(9i) ***Morro beato Benedito** tem feira comunitária ambiental.

(9j) ***Morro São Lourenço** tem feira comunitária ambiental.

Em (9i) e (9j) observa-se a impossibilidade de substituição dos elementos que compõem o Npr, ainda que a substituição seja por outro hagiônimo (nome de santo), como em (9j).

As frases (10) e (11) apresentam exemplo de composto toponímico que admite a variação do N1.

(10) O famoso **Morro Mestre Álvaro** encontra-se no município de Serra.

(11) “A Serra natural era dotada de diversidade ambiental, ou seja, Mata Atlântica, [...], praias, o **monte Mestre Álvaro** (833m) [...], etc”.

Embora os nomes compostos apresentem elevada fixidez e sejam praticamente unidades singulares, que em geral não admitem a substituição dos elementos constituintes, sobretudo os topônimos, que por seu estatuto de nome próprio apresenta uma natureza particularizante, há alguns compostos topônimos que admitem a variação do elemento genérico (N1) dentro de um paradigma distribucional bastante reduzido.

6.1.4 – Variação em número

Os nomes próprios geralmente são caracterizados pela sua função de especificar, isso é de identificar determinado objeto. Ao tratar da variação em número dos compostos de estrutura NN, Batista (1994) afirma que: “Alguns não admitem plural por designarem objectos únicos”, esse fato se dá com os topônimos, ainda que haja a ocorrência de semelhança fônica e gráfica (homografia) entre dois ou mais topônimos, eles designam objetos únicos, lugares únicos. Razão pela qual sentenças como (6i) e (6j) são inaceitáveis.

(6) Os madeireiros acamparam às margens do **córrego Montanha**.

(6i) *Os madeireiros acamparam às margens dos **córregos Montanha**.

(6j) *Os madeireiros acamparam às margens do **córrego Montanhas**.

O substantivo *Montanha* não admite flexão em número por fazer referência à um acidente geográfico específico que nomeia também o município homônimo.

Difícilmente ocorrerá o plural de N1 ou N2 em contextos como o das sentenças (7g), (7h), (8g), (9h) e (9i).

(7) O **rio Aribiri** desaguavam na baía de Vitória.

(7g) * Os **rios Aribiri** desaguavam na baía de Vitória.

(7h) * O **rio Aribiris** desaguava na baía de Vitória.

(8) Outra alternativa é visitar a **Ilha Itatiaia**.

(8g) Outra alternativa é visitar as **Ilhas Itatiaia**.

(8h) * Outra alternativa é visitar a **Ilha Itatiaias**.

(9) **Morro São Benedito** têm feira comunitária ambiental.

(9k) * **Morros São Benedito** têm feira comunitária ambiental.

(9l) * **Morro Sãos Beneditos** tem feira comunitária ambiental.

O composto *ilha Itatiaia* admite flexão de número do N1, como se observa em (8g), por tratar-se de um conjunto de ilhas chamadas *Itatiaia*. *Ilha Itatiaia* é, portanto, uma maneira de referir-se a um dos acidentes naturais que formam o arquipélago das *ilhas Itatiaia* ou ao seu conjunto. O plural de N1 também é possível em contextos como o da sentença a seguir:

(12) Existem dois **córregos Carapina**, um no estado do Espírito Santo e outro no estado de Goiás,

considerando que o N1 funciona como um termo que apresenta a designação genérica do composto.

No caso dos topônimos homônimos o mais usual é a indicação da localização desses para que seja desfeita a ambiguidade, como se observa no exemplo que segue:

(13) “Compramos um terreno para a congregação de Gaivotas, que passa a funcionar em Vila Nova. Vila nova de Vila Velha, porque tem Vila Nova na Serra e Vila Nova em Cariacica, mas a nossa congregação fica em Vila Nova de Vila Velha”.

O enunciador preocupa-se em apontar a localização do município onde se encontra o bairro a que se refere, a fim de desfazer as possíveis ambiguidades geradas pela homônimia dos bairros capixabas dos municípios de Cariacica, Serra e Vila Velha.

Nas frases (8g) e (12) ocorre a faturação do elemento genérico (*ilha* e *córrego*), os compostos toponímicos – em (8g) e (12) – referem-se a objetos distintos.

6.2 - Topônimos com a estrutura NAdj

As construções analisadas neste item apresentam um adjetivo posposto ao substantivo. O adjetivo está sintaticamente associado ao substantivo, “é essencialmente um modificador do substantivo” (CUNHA e CINTRA 2008: 259). Destacamos entre os topônimos candidatos a compostos que apresentam a estrutura NAdj os acidentes naturais: *córrego Fundo* (município de Serra - ES), *praia Grande* (município de Vila Velha - ES), *praia Mole* (município de Vitória - ES) e *rio Marinho* (município de Cariacica - ES). As propriedades observadas nessas estruturas são:

- ✓ Perda da predicatividade do adjetivo.
- ✓ Variação do adjetivo em grau.
- ✓ Coordenação do adjetivo com outro adjetivo.
- ✓ Ruptura paradigmática.

6.2.1 – Perda da predicatividade do adjetivo

O adjetivo exerce função predicativa, nesse caso, ocorre como núcleo do predicado em frases com a presença de verbo copulativo como, por exemplo, em

(14) O pão ficou mole.

O adjetivo *mole*, posposto ao verbo de ligação *ficar*, funciona como predicativo do sujeito. O adjetivo pode aparecer também em construções em que o verbo de ligação não está explícito como, por exemplo, em

(15) Ana comeu um pão mole.

A frase (15) admite a paráfrase (15a), resultado da inserção do pronome relativo *que* e do verbo de ligação *estar*:

(15a) Ana comeu um pão **que estava** mole.

Borba (1996: 155) afirma que “o adjetivo em posição pedicativa é matriz para o adjetivo em posição adnominal, por redução de oração relativa (ex.: sua cabeleira é branca > cabeleira que é branca > cabeleira branca)”. O adjetivo *mole* em (15a) é matriz para o adjetivo em posição adnominal em (15).

A sentença (15) apresenta reduções que podem ser consideradas como resultado das seguintes transformações:

Ana comeu um pão # O pão estava mole.

[Relativização] = (15a) Ana comeu um pão *que estava* mole.

[Pron-r Z], [Vc Z] = (15) Ana comeu um pão mole.⁹

As transformações atestam a equivalência entre as frases: *Ana comeu um pão # O pão estava mole, Ana comeu um pão que estava mole e Ana comeu um pão mole*. Caso a transformação gere uma frase não equivalente à construção inicial, a frase gerada é marcada com asterístico, como iniceitável.

De acordo com Rio-Torto (2006: 105)

O adjectivo desempenha essencialmente funções predicativas, entendendo-se por predicação a capacidade que ele activa de atribuir e/ou de modificar uma determinada propriedade ao denotado pelo nome a que se encontra associado.

⁹ [Pron-r Z] lê-se: elisão do pronome relativo (*que*).
[Vc Z] lê-se: elisão do verbo copulativo (*estar*)

Em alguns nomes compostos formados por um substantivo e um adjetivo a função predicativa do adjetivo não fica evidenciada. Nesses casos não é aceitável a elaboração de frases com a inserção do pronome relativo *que* e de um verbo copulativo entre o substantivo e o adjetivo. Considera-se que há perda da predicatividade do adjetivo como, por exemplo, nas frases:

(16) O **córrego Fundo** deságua no rio Santa Maria.

(16a) *O **córrego que é/está Fundo** deságua no rio Santa Maria.

(17) A **praia Grande** possui lagoas costeiras.

(17a) *A **praia que é Grande** possui lagoas costeiras.

Os adjetivos *fundo* e *grande* nas sentenças (16) e (17) funcionam como elemento identificador dos acidentes geográficos naturais *córrego* e *praia* respectivamente. Os nomes *córrego Fundo* e *praia Grande* não podem ser parafraseados pelas expressões *córrego que é/está Fundo* e *praia que é Grande*. Esse fato aponta para fixidez sintática dos compostos toponímicos destacados (*córrego Fundo* e *praia Grande*). O uso dos elementos especificadores *fundo* e *grande*, nas frases (16) e (17) independe do nível ou profundidade do *córrego* especificado ou da extensão da *praia* nomeada, ou seja, os adjetivos *fundo* e *grande* não cumprem, nesses exemplos, o seu papel predicativo, de atribuir ou modificar uma característica ou propriedade atribuída ao nome a que se refere.

Na frase (15) *Ana comeu um pão mole* o adjetivo *mole* atribui ao substantivo a que se associa (*pão*) a propriedade de *ser macio*, *ser pouco consistente*. Já em

(18) A **praia Mole** tem um porto de uso privado,

o adjetivo funciona como elemento identificador de um acidente geográfico natural do município de Vitória. Assim sendo, torna-se inaceitável a frase:

(18a) *A **praia que é Mole** tem um porto de uso privado.

O adjetivo *mole* em (18) não atribui ao substantivo *praia* a propriedade de *ser macia*, *ser pouco consistente* como ocorre na frase (15).

(19) O **rio Marinho** separa os municípios de Cariacica e Vila Velha.

(19a) *O rio **que é Marinho** separa os municípios de Cariacica e Vila Velha.

Dick (1997: 19) afirma que “certos nomes registrados acabam por perder a sua função identificadora, com o desaparecimento do referencial a que se ligavam, fosse ele de natureza pessoal ou real.” Nas pesquisas em PLN a descrição do comportamento linguístico dos compostos toponímicos é pautada em uma abordagem sincrônica da língua. Dessa forma, a permanência ou não da motivação que levou o nomeador a eleger os adjetivos *fundo*, *grande*, *mole* e *marinho* como elemento especificador de acidentes geográficos capixabas não é objeto de estudo desta pesquisa.

Bechara (2009: 143-144) ao tratar dos instrumentos gramaticais de determinação nominal inclui nesse grupo os adjetivos e explica que esses itens lexicais podem ser elementos de determinação identificadora e que

Os identificadores podem ser ocasionais (Caxias, *Maranhão*), usuais (São Lourenço *do Sul*), ou constantes (*Nova York*, *Porto Alegre*). Os indentificadores usuais e constantes, como partes integrantes de um signo, constituem, com seus determinados, verdadeiros *nomes compostos* [...] (grido do autor).

O autor considera ocasional a indicação da unidade federativa a que pertence determinado município. Por exemplo, ao referir-se ao município de Caxias o falante pode fazer a indicação do estado a que pertence esse município (Maranhão), a indicação torna-se dispensável se o ato comunicativo se der no mesmo espaço físico em que se localiza o referido município. Da mesma forma dificilmente um morador da Baixada Fuminense (Rio de Janeiro) dirá que vai à (Duque de) Caxias, Rio de Janeiro. Nesses casos, o elemento indentificador não é parte integrante de um composto toponímico. Os indentificadores usuais por sua vez são elementos formadores de compostos toponímicos que são passíveis de apagamento. Para exemplificar os indentificadores usuais Bechara (2009: 144) apresenta o topônimo *São Lourenço do Sul* (município do Rio Grande do Sul) e afirma que “o habitante de *São Lourenço do Sul* usará apenas *São Lourenço*, dentro do mesmo entorno” (grifo do autor). Poder-se-ia dizer igualmente que o habitante de *São Lourenço* (município de Minas Gerais) usará *São Lourenço do Sul* para referir-se ao município gaúcho. Bechara (2009) considera como “verdadeiros nomes compostos” os indentificadores constantes, isso é, os adjetivos que integram o elemento específico do composto toponímico.

Consideramos *Porto Alegre* a forma reduzida de **Cidade de Porto Alegre** e *Nova York* a forma reduzida de **Cidade de Nova York**, ou seja, formas em que há a elisão do elemento genérico

cidade. Os topônimos *córrego Fundo*, *praia Grande*, *praia Mole* e *rio Marinho* anotados nas frases (16), (17), (18) e (19) apresentam um adjetivo associado ao elemento genérico, nesses exemplos o substantivo não faz parte do elemento específico, contudo, observa-se que esses nomes poderiam ser inseridos, na classificação proposta por Bechara (2009), entre os identificadores constantes.

A classificação dos instrumentos gramaticais de determinação nominal proposta por Bechara (2009) não é adotada nesta pesquisa. É observado se esses instrumentos gramaticais, apresentados pelo autor nos exemplos mencionados, integram ou não o composto toponímico. Os grupos nominais que se referem a determinado lugar são analisados quanto à sua condição de compostos ou de grupos nominais livres. Essa abordagem binária descarta a identificação proposta por Bechara (2006) dos “verdadeiros nomes compostos” (dos graus de fixidez dos compostos), considerando-se que uma sequência é ou não é composta.

6.2.2 – Variação do adjetivo em grau

Os adjetivos em geral admitem variação em grau, em (15) *Ana comeu um pão mole*, por exemplo, o adjetivo *mole*, pode variar em grau. Segundo Bechara (2009: 147) o grau superlativo sintético do adjetivo “é obtido por meio do sufixo derivacional *-íssimo* (ou outro de valor intensivo) acrescido ao adjetivo”. Assim, com a variação em grau do adjetivo em (15) obtém-se (15b) *Ana comeu um pão molíssimo*. O mesmo não ocorre com os adjetivos que constituem os compostos toponímicos, como se observa em

(16) O **córrego Fundo** deságua no rio Santa Maria.

(16b) *O **córrego Fundíssimo** deságua no rio Santa Maria.

(17) A **praia Grande** possui lagoas costeiras.

(17b) *A **praia Grandíssima** possui lagoas costeiras.

(18b) A **praia Mole** tem um porto de uso privado.

(18b) *A **praia Molíssima** tem um porto de uso privado.

(19) O **rio Marinho** separa os municípios de Cariacica e Vila Velha.

(19b) *O **rio Marinhíssimo** separa os municípios de Cariacica e Vila Velha.

A inaceitabilidade dos exemplos (16b) a (19b) deve-se à fixidez sintática dos compostos NcNpr =: Nadj em decorrência de sua institucionalização.

Cunha e Cintra (2008: 259-260) apontam duas funções dos adjetivos: “caracterizar os seres, os objetos ou as noções nomeadas pelo substantivo” e “estabelecer com o substantivo uma relação [...]”, estes são chamados *adjetivos relacionais*. Os autores afirmam que os adjetivos relacionais “são de natureza classificatória” e “não admitem graus de intensidade”. O adjetivo *marinho* (*relativo ao mar/do mar*) pertence ao grupo dos chamados *adjetivos relacionais*, assim sendo, não é possível a variação em grau desse item lexical. De acordo com Borba (1996: 177) os adjetivos chamados de classificadores (CL) “ampliam nosso conhecimento do nome [...]. Um discurso *corajoso* é uma apreciação subjetiva [...]; um discurso político é uma especificação que restringe o alcance do nome” (grifo do autor). Os adjetivos qualificadores (QL) são passíveis de gradação por representar a atribuição de determinada propriedade a um nome ao passo que a inclusão de um nome em determinada classe (através de um adjetivo relacional ou classificador) é definitiva, portanto, não é passível de gradação. O autor (1996: 178) explica que “a gradação parece ser traço sintático típico do adjetivo, razão por que há quem considere só os QL como verdadeiros adjetivos, colocando CL na categoria de pseudo-adjetivos”. O reconhecimento dos adjetivos chamados de qualificadores como verdadeiros relaciona-se com sua maior vinculação com o nome a que se associa.

A inaceitabilidade da frase

(19b) *O **rio Marinhíssimo** separa os municípios de Cariacica e Vila Velha.

é corroborada pela identificação do adjetivo *marinho* como classificador ou relacional.

6.2.3 – Coordenação do adjetivo com outro adjetivo

De acordo com Bechara (2009: 48) “orações que apresentam uma mesma função textual, palavras e grupos de palavras de mesma função” podem ser coordenados. O adjetivo *mole* desempenha a função de predicativo do objeto, na frase:

(15) Ana comeu um pão **mole**.

É possível estabelecer uma relação de coordenação do adjetivo *mole* com outro adjetivo que desempenhe a mesma função, como se observa em

(15b) Ana comeu um pão **mole e insosso**.

Os adjetivos *mole e insosso* desempenham a função de predicativo do objeto (*pão*).

Com os nomes compostos de estrutura NAdj a coordenação com outro adjetivo não é possível, como se observa nas sentenças:

(16) O **córrego Fundo** deságua no rio Santa Maria.

(16c) *O **córrego Fundo e caudaloso** deságua no rio Santa Maria.

(17) A **praia Grande** possui lagoas costeiras.

(17c) *A **praia Grande e paradisíaca** possui lagoas costeiras.

(18) A **praia Mole** tem um porto de uso privado.

(18c) *A **praia Mole e calma** tem um porto de uso privado.

(19) O **rio Marinho** separa os municípios de Cariacica e Vila Velha.

(19c) *O **rio Marinho e navegável** separa os municípios de Cariacica e Vila Velha.

O adjetivo que forma o nome composto constitui com o substantivo que o antecede uma unidade lexical fixa. Em (16c) o adjetivo *fundo* forma com o substantivo *córrego* uma unidade lexical fixa impossibilitando a coordenação com o adjetivo *caudaloso*, por este

adjetivo referir-se a uma parte do composto (o substantivo *córrego*) de forma isolada. O fato de não ser possível decompor os nomes *praia Grande*, *praia Mole* e *rio Marinho* e analisar morfosintática e semanticamente as suas partes (isoladamente) torna inadequada a coordenação do adjetivo formador desses nomes com um outro adjetivo.

6.2.4 – Ruptura paradigmática

Em grupos nominais livres formados por um substantivo seguido de adjetivo é possível comutar tanto o substantivo como o adjetivo, respeitando-se as restrições impostas pelo verbo e pelo substantivo, por exemplo:

(20) Ana passeava por uma praia (ilha + cidade + praça + região) **belíssima** (calma + tranquila + linda + deslumbrante).

Os compostos toponímicos que apresentam a mesma estrutura (substantivo + adjetivo) não admitem a permuta dos seus elementos, formadores ainda que esses elementos sejam substituídos por outros da mesma classe gramatical e campo semântico, como se observa nas frases a seguir:

(16) O **córrego Fundo** deságua no rio Santa Maria.

(16d) *O **córrego Profundo** deságua no rio Santa Maria.

(17) A **praia Grande** possui lagoas costeiras.

(17d) *A **praia Enorme** possui lagoas costeiras.

(18) A **praia Mole** tem um porto de uso privado.

(18d) *A **praia Macia** tem um porto de uso privado.

(19) O **rio Marinho** separa os municípios de Cariacica e Vila Velha.

(19d) *O rio *Pluvial* separa os municípios de Cariacica e Vila Velha.

De acordo com Gaston Gross (1996: 86) nos grupos nominais livres cada posição nominal constitui um paradigma e “la figement dans un groupe peut être calculé par les restrictions imposées sur ces paradigmes”. Considerando que a fixidez de um grupo nominal pode ser calculada pelas restrições impostas sobre estes paradigmas pode-se dizer que os itens lexicais grifados nas frases (16) a (19) são fixos, ou seja, são compostos toponímicos.

6.3 – Topônimos com a estrutura NdeN

Para Vivès (1990: 98) “des formes composées N Prép N (une baouchée à la reine, um homme de paille), souvent plus délicates à distinguer de formes libres (une encoste à la cheville, um verre de vin)”. A distinção entre os compostos de estrutura NdeN e os grupos nominais livres é considerada, portanto, mais delicada que a distinção dos compostos NN e os grupos nominais livres.

Destacamos entre os topônimos candidatos a compostos que apresentam a estrutura NdeN: *Ilha de Vitória* (município de Vitória - ES), *Praia de Itapoã* (município de Vila Velha - ES), *Morro do Céu* (município de Serra - ES), *Rio da Passagem* (município de Vitória - ES). Os critérios formais utilizados na identificação das propriedades morfossintáticas e semânticas dos topônimos são:

- ✓ Inserção de um item lexical no grupo nominal.
- ✓ Variação do determinante de N2.
- ✓ Ruptura paradigmática.
- ✓ Variação em número.

6.3.1 – Inserção de um item lexical no grupo nominal

A inserção de um item lexical (um adjetivo, por exemplo) em um grupo nominal livre com a estrutura NdeN é gramaticalmente aceitável. O adjetivo pode referir-se ao N1 ou ao N2, de forma isolada, como se observa em

(21) Ana bebeu um **copo de leite** antes de dormir.

(21a) Ana bebeu um **copo grande de leite** antes de dormir.

(21b) Ana bebeu um **copo de leite quente** antes de dormir.

Os nomes compostos, por sua vez, não admitem a inserção de um item lexical entre os elementos formantes como se observa nas frases a seguir:

(22) “Os índios chamavam a **Ilha de Vitória** de Guanaaní ou ‘Ilha do Mel’ [...]”.

(22a) *Os índios chamavam a **Ilha bela de Vitória** de Guanaaní ou ‘Ilha do Mel’.

(23) “A **Praia de Itapoã** também vai receber as novas lixeiras e contentores [...]”.

(23a) *A **Praia linda de Itapoã** também vai receber as novas lixeiras e contentores.

(24) O **Morro do Céu** possui 414 metros de altura.

(24a) *O **Morro grande do Céu** possui 414 metros de altura.

(25) O **Rio da Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

(25a) *O **Rio caudaloso da Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

Observa-se a impossibilidade de se associar um adjetivo ao N2 nos nomes compostos.

(22b) *Os índios chamavam a **Ilha de Vitória acirrada** de Guanaaní ou ‘Ilha do Mel’.

(23b) *A **Praia de Itapoã** *nova* também vai receber as novas lixeiras e contentores.

(24b) *O **Morro do Céu** *estrelado* possui 414 metros de altura.

(25b) *O **Rio da Passagem** *estreita* recebe grande parte do esgoto doméstico produzido na Grande Vitória.

Nas frases (22a) a (25b) o adjetivo refere-se a apenas um dos nomes da estrutura NdeN. Essas frases são consideradas inaceitáveis, não constituem paráfrases das sentenças (22) a (25), o que comprova a fixidez dessas construções. O adjetivo deve referir-se ao conjunto (NdeN) como um todo. A associação de um adjetivo ao composto toponímico deve ser feita considerando-se que os elementos que constituem o composto representam um conjunto único.

6.3.2 – Variação do determinante de N2

Segundo Bechara (2009: 411) “os determinantes estão, em geral, representados pelas seguintes classes de palavras: *adjetivo, artigo e pronome demonstrativo* ou equivalentes de adjetivos” (grifo do autor). Nas frases a seguir procede-se à substituição do artigo definido por um pronome demonstrativo ou a inserção do pronome nos casos de ausência do artigo a fim de verificar se os grupos nominais apresentam fixidez.

(22c) *Os índios chamavam a **Ilha** *desta* **Vitória** de Guanaaní ou ‘Ilha do Mel’.

(23c) *A **Praia** *deste* **Itapoã** também vai receber as novas lixeiras e contentores.

(24c) *O **Morro** *deste* **Céu** possui 414 metros de altura.

(25c) *O **Rio** *daquela* **Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

Bechara (2009: 155) afirma que o artigo definido “costuma aparecer ao lado de certos nomes próprios geográficos, principalmente os que denotam países, oceanos, rios, montanhas, ilhas”.

Nos exemplos apresentados pelo autor não são registrados os termos genéricos: “*a Suécia, o Atlântico, o Amazonas, os Andes, a Groelândia*” (grifo do autor). Os acidentes naturais são apresentados em sua forma reduzida: *o oceano Atlântico, o rio Amazonas, a cordilheira dos Andes, a ilha da Groelândia*.

Nas frases (22*d*) e (23*d*) os compostos toponímicos são registrado em sua forma longa, sem a elisão de N1 e o artigo definido é anteposto ao N2.

(22*d*) *Os índios chamavam a **Ilha da Vitória** de Guanaaní ou ‘Ilha do Mel’,

A inserção do determinante em (22*d*) é considerada inaceitável, enquanto paráfrase de *ilha de Vitória*, assim como é inaceitável a inserção de um pronome demonstrativo anteposto ao N2 em (22*c*). Em

(23*d*) A **Praia do Itapoã** também vai receber as novas lixeiras e contentores,

consideramos aceitável a formação *Praia do Itapoã* por ser esse um nome institucionalizado, registrado nas cartas do IBGE, órgão responsável, no Brasil, pelo registro dos nomes próprios de lugar. No *corpus* da pesquisa registramos também o nome *Praia de Itapoã* por ser essa uma forma em uso, reconhecida e aceita pelos falantes nativos do português do Brasil.

Observa-se a presença de um determinante (artigo definido) anteposto ao N2 nas frases:

(24) O **Morro do Céu** possui 414 metros de altura.

(25) O **Rio da Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

Porém, não é possível a substituição por outro determinante (pronome demonstrativo), o que atesta a fixidez do conjunto, como se observa em

(24*c*) *O **Morro deste Céu** possui 414 metros de altura.

(25*c*) *O **Rio daquela Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

6.3.3 – Ruptura paradigmática

Nos grupos nominais livres de estrutura NdeN é aceitável a substituição de cada um dos substantivos que formam o grupo por outro que tenha a mesma distribuição, visto que cada posição nominal representa um paradigma, como por exemplo:

(26) Ana comprou um cesto (uma cesta + um balaio) de bambu (taquara + cipó).

Os compostos toponímicos não apresentam essa propriedade, é inaceitável a substituição de um dos nomes da estrutura NdeN, ainda que o substantivo seja substituído por outro que apresente a mesma distribuição, conforme se observa a seguir:

(22) Os índios chamavam a **Ilha de Vitória** de Guanaaní ou ‘Ilha do Mel’.

(22e) *Os índios chamavam a **Ilha da Conquista** de Guanaaní ou ‘Ilha do Mel’.

(23) A **Praia de Itapoã** também vai receber as novas lixeiras e contentores.

(23e) *A **Praia de Itaparica** também vai receber as novas lixeiras e contentores.

(24) O **Morro do Céu** possui 414 metros de altura.

(24d) *O **Morro do Firmamento** possui 414 metros de altura.

(25) O **Rio da Passagem** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

(25d) *O **Rio do Caminho** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

A sequência apresentada em (23e) pode ser considerada como possível de se realizar na língua portuguesa, contudo é considerada como inaceitável em relação à frase

(23) “A **Praia de Itapoã** também vai receber as novas lixeiras e contentores [...]”.

Em (23) o nome *praia de Itapoã* pode ser parafraseado por *praia* ou *Itapoã*, mas não pode ser parafraseado por *praia de Itaparica*.

6.3.4 – Variação em número

De acordo com Bechara (2009: 130) “nos compostos onde haja preposição, clara ou oculta”, somente o primeiro elemento varia em número. Essa orientação, porém, não se aplica aos compostos toponímicos de estrutura NdeN. O autor observa, quanto à variação em número, que “um nome como *João*, *Isabel* ou *Açores* só acidentalmente se aplicará a várias pessoas ou ilhas [...]. Por isso cada *João*, cada *Isabel* e cada *Açores* é uma pessoa ou ilha [...]” (BECHARA, 2009: 113, grifo do autor). Os compostos toponímicos não apresentam variação de número do N1 devido ao seu caráter individualizante como se pode notar, por exemplo, pela inaceitabilidade das frases em que o N1 do composto toponímico é registrado no plural:

(22f) *Os índios chamavam as **Ilhas de Vitória** de Guanaaní ou ‘Ilhas do Mel’.

(23f) *As **Praias de Itapoã** também vão receber as novas lixeiras e contentores.

(24e) *Os **Morros do Céu** possuem 414 metros de altura.

(25e) *Os **Rios da Passagem** recebem grande parte do esgoto doméstico produzido na Grande Vitória.

O plural de N1, nas frases (22f), (23f), (23e) e (25e) faz com que se perca o sentido da composição, isto é, as sequências NdeN apresentam o comportamento linguístico de grupos nominais livres. A frase (22f), por exemplo, não apresenta o mesmo sentido da frase

(22) “Os índios chamavam a Ilha de Vitória de Guanaaní ou ‘Ilha do Mel’ [...]”.

A frase (22) afirma que os índios chamavam de Guanaaní ou ‘Ilha do Mel’ o acidente geográfico natural nomeado de Ilha de Vitória. A interpretação da frase (22f) aponta para o fato de que os índios chamavam de Guanaaní ou ‘Ilha do Mel’ os acidentes geográficos naturais (as ilhas) localizados no município de Vitória, na faixa marítima que pertence a esse município.

Smarsaro (2004: 74) afirma que “a categoria dos nomes é susceptível de variação em número. Quando um grupo nominal apresenta obrigatoriamente N1 ou N2 no plural trata-se de um

indício de fixidez”. Pode-se dizer também que o fato de um grupo nominal apresentar N1 ou N2 obrigatoriamente no singular representa um indício de fixidez, como se observa em

(22g) *Os índios chamavam a **Ilha de *Vitórias*** de Guanaaní ou ‘Ilha do Mel’.

(23g) *A **Praia de *Itapoãs*** também vai receber as novas lixeiras e contentores.

(24f) *O **Morro dos *Céus*** possui 414 metros de altura

(25f) *O **Rio das *Passagens*** recebe grande parte do esgoto doméstico produzido na Grande Vitória.

A inaceitabilidade das frases em que se efetua a flexão de número de N2 apontam para a fixidez dos topônimos que elas contêm.

7 – FORMALIZAÇÃO DA DESCRIÇÃO DOS COMPOSTOS TOPONÍMICOS

A formalização das descrições linguísticas é de extrema importância para o desempenho dos sistemas informáticos que dela se utilizam. Formalizar as descrições linguísticas consiste em elaborar e anotar tais descrições de maneira que os resultados sejam utilizáveis por máquinas. Por exemplo, o resultado da descrição dos compostos toponímicos apresentada nesta pesquisa assume a forma de frases como "expressões do tipo 'baía de Vitória' são topônimos". Essa informação não está formalizada. As máquinas não são dotadas da habilidade de avaliar se uma dada expressão enquadra-se ou não no tipo "baía de Vitória", nem 'compreendem' as frases escritas em língua natural. Além disso, a frase "expressões do tipo 'baía de Vitória' são topônimos" não informa, por exemplo, se "baía de Vitória" apresenta determinante obrigatório ou se a presença do determinante é dispensável. Formalizar a descrição, nesse caso, significa elaborar uma lista de compostos toponímicos (do tipo "baía de Vitória") e codificar as informações sobre a obrigatoriedade e o gênero do determinante de cada expressão. Desse modo, as informações podem tornar-se mais completas, menos ambíguas e utilizáveis pela máquina.

A codificação representa regularidades e generalidades das descrições linguísticas, por meio de símbolos que obedecem convenções. Por exemplo, para indicar que "baía de Vitória" não apresenta determinante obrigatório. Pode ser estabelecida uma convenção, o uso do código +DZ, em vez de uma frase como "não é obrigatório o uso de determinante do composto".

Consideramos as seguintes propriedades de $N1(Prep)N2$ na formalização apresentada:

não ser obrigatório o uso de determinante do composto,

apagamento de N1,

apagamento de N2 e

variação de gênero do determinante do composto em decorrência do apagamento de N1.

Onde:

N1 – primeiro substantivo do composto,

N2 – segundo substantivo do composto, que pode ser simples ou composto,

Prep – preposição.

O uso dos parênteses na expressão $NI(Prep)N2$ indica que a presença da preposição entre o NI e o $N2$ é facultativa, isto é, consideramos tanto as sequências constituídas de NI e $N2$ ligadas por preposição (*Praia de Itapoã*) como as que não apresentam preposição (*Ilha de Itatiaia*).

A letra N indica a classe a que o composto pertence (nome).

Anota-se $+DZI$ quando há ausência de determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição, por exemplo:

(1) Ana e Moisés moram em **Alto do Suá**.

Anota-se $+DI$ quando há ocorrência do determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição, por exemplo:

(2) “Em Ponto Alto, já são 36 afetadas pela enchente do **Rio Jucu Braço Sul**”.

Anota-se $+DI+DZI$ nos casos em que a presença do determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição é facultativa.

Anota-se $+NIZ$ quando há elisão de $N1$ na sequência $NI(Prep)N2$, por exemplo:

(3) **Bairro Praia da Costa** conta com patrulhamento 24 horas.

(3a) **Praia da Costa** conta com patrulhamento 24 horas.

Anota-se $-NIZ$ quando não há elisão de $N1$ na sequência $NI(Prep)N2$, por exemplo:

(4) **Praia da Costa** recebe o projeto 'capoeira nas praias'.

(4a)***Da Costa/Costa** recebe o projeto 'capoeira nas praias'.

Anota-se $+DZ2$ quando há ausência de determinante da sequência $NI(Prep)N2$ em função de complemento antecedido por preposição e a elisão de $N1$, por exemplo:

(5) Ana mudou-se para o **Bairro São Pedro**.

(5a) Ana mudou-se para **São Pedro**.

Anota-se +D2 quando há ocorrência do determinante da sequência *NI(Prep)N2* em função de complemento antecedido por preposição e a elisão de N1, por exemplo:

(6) A lagoa Vermelha fica próxima ao **Bairro Morro da Lagoa**.

(6a) A lagoa Vermelha fica próxima ao **Morro da Lagoa**.

Anota-se +D2+DZ2 nos casos em que a presença do determinante da sequência *NI(Prep)N2*, em função de complemento antecedido por preposição e a elisão de N1 é facultativa.

Anota-se +g1 quando a sequência *NI(Prep)N2* não apresenta variação de gênero do determinante em decorrência do apagamento de N1, por exemplo:

(7) O **rio Aribiri** deságua na baía de Vitória.

(7a) O **Aribiri** deságua na baía de Vitória.

Anota-se +g2 quando a sequência *NI(Prep)N2* apresenta variação de gênero do determinante em decorrência do apagamento de N1, por exemplo:

(8) Ana mudou-se para **o bairro Ilha dos Aires**.

(8a) Ana mudou-se para **a Ilha dos Aires**.

Anota-se +g1+g2 nos casos em que variação de gênero do determinante em decorrência do apagamento de N1 é facultativa.

Anota-se +N2Z quando a sequência *NI(Prep)N2* apresenta elisão de N2, por exemplo:

(4) **Praia da Costa** recebe o projeto 'capoeira nas praias'.

(4b) **Praia** recebe o projeto 'capoeira nas praias'.

Anota-se -N2Z quando a sequência *NI(Prep)N2* não apresenta elisão de N2, por exemplo:

(1) Ana e Moisés moram em **Alto do Suá**.

(1a) *Ana e Moisés moram em **Alto**.

Anota-se +N2p quando o N2 da sequência *N1(Prep)N2* é plural, por exemplo:

(9) “**Praia das Gaivotas** recebe projeto Arte e Cultura nos bairros”.

Anota-se -N2p quando o N2 da sequência *N1(Prep)N2* é singular, por exemplo:

(10) “O projeto ‘Rua Segura’ também vem sendo desenvolvido no **Bairro Praia do Canto**, na cidade de Vitória”.

As notações :fs, :fp, :ms e :mp referem-se, respectivamente, aos seguintes códigos flexionais dos compostos: *feminino singular*, *feminino plural*, *masculino singular* e *masculino plural*.

Na entrada

Bairro Praia da Costa N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms

a letra *N* indica a classe a que pertence o composto (nome), o código +*Pr* assinala que *Bairro Praia da Costa* é um *nome próprio*. Trata-se de um nome próprio de lugar, um *topônimo*, fato indicado através do código +*Top*. O código +*D1* aponta para a presença de determinante diante da forma longa desse composto toponímico quando em função de complemento antecedido por preposição. Por exemplo:

(11) O ônibus passa pelo **bairro Praia da Costa**.

É possível o apagamento do N1 (*bairro*), fato registrado através da anotação +*N1Z*, como se observa nas frases:

(3) **Bairro Praia da Costa** conta com patrulhamento 24 horas.

(3a) **Praia da Costa** conta com patrulhamento 24 horas.

A presença de determinante diante da forma reduzida do composto toponímico (com o apagamento de N1) quando em função de complemento antecedido por preposição é marcada pelo código +*D2*. O código +*g1* indica manutenção do gênero do determinante do composto toponímico (o mesmo que o registrado diante da forma longa) quando ocorre a elisão de N1, mas a forma reduzida do composto toponímico (*Praia da Costa*) admite também a variação do gênero do determinante, daí anotar-se +*g2*.

(12) “Apartamento para locação **no praia da costa** em Vila Velha-ES, 120m², R\$ 1.800,00”.

(13) A professora mora *na Praia da Costa*.

Há possibilidade de elisão do N2 do composto toponímico *Bairro Praia da Costa*, marcada através do código +N2Z.

(3) **Bairro Praia da Costa** conta com patrulhamento 24 horas.

(3b) **Bairro** conta com patrulhamento 24 horas

O código -N2p indica que o N2 do composto toponímico é singular. As letras *ms*, após os dois pontos indicam que se trata de um composto masculino singular.

Lista das palavras compostas toponímicas por estrutura e a formalização de sua descrição.

Compostos toponímicos com estrutura NN:

1. Arquipélago Martin Vaz	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
2. Bairro Alto Lage	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
3. Bairro Barra do Jucu	N+Pr+Top+D1+N1Z+D2+DZ2+g1+g2+N2Z-N2p:ms
4. Bairro Campinho da Serra	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
5. Bairro Campo Grande	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
6. Bairro Ilha da Conceição	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:ms
7. Bairro Ilha das Flores	N+Pr+Top+D1+N1Z +D2+DZ2+g2+N2Z-N2p:ms
8. Bairro Ilha dos Aires	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:ms
9. Bairro Ilha dos Bentos	N+Pr+Top+D1+N1Z+DZ2+g2+N2Z-N2p:ms
10. Bairro Morro da Lagoa	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
11. Bairro Pedra dos Búzios	N+Pr+Top+D1+N1Z+D2+DZ2+g2+g2+N2Z-N2p:ms
12. Bairro Ponta da Fruta	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:ms
13. Bairro Praia da Costa	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
14. Bairro Praia das Gaivotas	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
15. Bairro Praia de Carapebus	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
16. Bairro Praia de Itaparica	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
17. Bairro Praia do Canto	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
18. Bairro Praia dos Recifes	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
19. Bairro Prainha da Glória	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
20. Bairro Rio Marinho	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
21. Bairro Vale do Sol	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
22. Bairro Vale dos Reis	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
23. Bairro Vale Encantado	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
24. Bairro Vale Esperança	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
25. Canal Bigossi	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
26. Canal Camboapina	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
27. Canal Dique	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
28. Canal Jacarandá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
29. Canal Praia da Costa	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
30. Córrego Água Funda	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
31. Córrego Arataca	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
32. Córrego Areinha	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms

33. Córrego Aruaba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
34. Córrego Barata	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
35. Córrego Barro Branco	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
36. Córrego Bico do Grilo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
37. Córrego Biriricas	N+Pr+Top+D1+N1Z+D2+g1+N2Z+N2p:ms
38. Córrego Boca da Mata	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
39. Córrego Boca do Mato	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
40. Córrego Boqueirão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
41. Córrego Caburé	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
42. Córrego Caçaroca	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
43. Córrego Cachoeira do Putiri	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
44. Córrego Cachoeira	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
45. Córrego Cachoeirinha	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
46. Córrego Calogi	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
47. Córrego Campo Grande	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
48. Córrego Canto Grande	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
49. Córrego Capivari	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
50. Córrego Capuba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
51. Córrego Carapebus	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
52. Córrego Carapina	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
53. Córrego Castelo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
54. Córrego Cavada	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
55. Córrego Chapada Grande	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
56. Córrego Doutor Robson	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
57. Córrego Garanhões	N+Pr+Top+D1+N1Z+D2+g1+N2Z+N2p:ms
58. Córrego Garanhuns	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
59. Córrego Independência	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
60. Córrego Irema	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
61. Córrego Itaiobaia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
62. Córrego Itapebuna	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
63. Córrego Itatiaia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
64. Córrego Jacaraípe	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
65. Córrego Jacarandá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
66. Córrego Jacaré	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
67. Córrego Jaconé	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
68. Córrego Jacuném	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
69. Córrego Jaquitá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
70. Córrego João Dias	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
71. Córrego Joãozinho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
72. Córrego Juara	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
73. Córrego Laje	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
74. Córrego Laranjeiras	N+Pr+Top+D1+N1Z+D2+g1+N2Z+N2p:ms
75. Córrego Limão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
76. Córrego Manguinhos	N+Pr+Top+D1+N1Z+D2+g1+N2Z+N2p:ms
77. Córrego Maringá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
78. Córrego Moitavão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
79. Córrego Montanha	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
80. Córrego Morro Baixo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
81. Córrego Nova Manta	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
82. Córrego Pau-amarelo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms

83. Córrego Pedra Mulata	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
84. Córrego Piaíba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
85. Córrego Piranem	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
86. Córrego Piranema	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
87. Córrego Porto do Engenho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
88. Córrego Quatizeiro	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
89. Córrego Quibebe	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
90. Córrego Relógio	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
91. Córrego Roda-d'água	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
92. Córrego Roncador	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
93. Córrego Santa Luzia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
94. Córrego Santo Agostinho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
95. Córrego São Bento	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
96. Córrego São Domingos	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
97. Córrego São Paulo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
98. Córrego Sapucaia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
99. Córrego Tabatinga	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
100. Córrego Tamanco	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
101. Córrego Tanque	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
102. Córrego Trincheira	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
103. Córrego Vargem Grande	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
104. Córrego Vasco Coutinho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
105. Córrego Vêner	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
106. Córrego Vimeiro	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
107. Ilha Crisógono	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
108. Ilha Doutor Américo de Oliveira	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
109. Ilha Gonçalves Martins	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
110. Ilha Guruçá	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
111. Ilha Margarida	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
112. Ilha Maria Catoré	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
113. Ilha Rabello	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
114. Ilha Tipoá	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
115. Ilha Wetzel	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
116. Ilha Itatiaia	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
117. Ilhas Itatiaia	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fp
118. Lagoa Doutor Robson	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
119. Lagoa Largo do Juara	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
120. Localidade Chapada Grande	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
121. Localidade Córrego Fundo	N+Pr+Top+D1+N1Z+DZ2+g2+N2Z-N2p:fs
122. Maciço Garrafão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
123. Monte Horebe	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z-N2p:ms
124. Monte Mestre Álvaro	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z-N2p:ms
125. Morro Alto Caratoíra	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
126. Morro Alto Itararé	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
127. Morro Andorinhas	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
128. Morro Araçatiba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
129. Morro Bananal	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
130. Morro Barro Vermelho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
131. Morro Bela Vista	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms

132. Morro Bento Ferreira	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
133. Morro Boa Vista	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
134. Morro Conquista	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
135. Morro Consolação	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
136. Morro Constantino	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
137. Morro Engenharia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
138. Morro Fonte Grande	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:ms
139. Morro Forte São João	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
140. Morro Guajuru	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
141. Morro Gurigica	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
142. Morro Ilha das Caieiras	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
143. Morro Itaúnas	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z+N2p:ms
144. Morro Jaburu	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z -2p:ms
145. Morro Jaburuna	N+Pr+Top+D1+N1Z+D2+D2+g1+N2Z-N2p:ms
146. Morro Jesus de Nazareth	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
147. Morro Kokó	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
148. Morro Maria Ortiz	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
149. Morro Monte Belo	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
150. Morro Mutrapina	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:MS
151. Morro Pedra dos Olhos	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:ms
152. Morro Santa Clara	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
153. Morro Santa Maria	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
154. Morro Santiago	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
155. Morro Santos Dumont	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
156. Morro São Benedito	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
157. Morro São João	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
158. Morro São José	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
159. Morro Senhor do Bonfim	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
160. Morro Tabuazeiro	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
161. Morro Una	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
162. Morro Universitário	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
163. Morro Xavier	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
164. Ponta Carapebus	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
165. Praia Camburi	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
166. Praia Carapebus	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
167. Praia Itaparica	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
168. Praia Peitoril	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
169. Praia Ponta dos Fachos	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
170. Praia Santa Helena	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
171. Ribeirão Brejo Grande	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
172. Ribeirão Juara	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
173. Ribeirão Peixe Verde	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
174. Ribeirão Ponto Doce	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
175. Ribeirão Santo Agostinho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
176. Ribeirão São Vicente	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
177. Ribeirão Sauanha	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
178. Rio Araçatiba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
179. Rio Aribiri	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
180. Rio Braço do Mangaraí	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
181. Rio Bubu	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms

182. Rio Chury	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
183. Rio Formate	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
184. Rio Jacaraípe	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
185. Rio Jacarandá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
186. Rio Jequitibá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
187. Rio Jucu Braço Norte	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
188. Rio Jucu	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
189. Rio Mangaral	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
190. Rio Reis Magos	N+Pr+Top+D1+N1Z+D2+g1+N2Z+N2p:ms
191. Rio Santa Maria	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
192. Serra Bragança	N+Pr+Top+D1+DZ1+N1Z+g1+N2Z-N2p:fs
193. Serra Mês Alvo	N+Pr+Top+D1+DZ1+N1Z+D2+g2+N2Z-N2p:fs
194. Serra Mororom	N+Pr+Top+D1+DZ1+N1Z+D2+g2+N2Z-N2p:fs
195. Serra Munchuara	N+Pr+Top+D1+DZ1+N1Z+D2+g2+N2Z-N2p:fs
196. Serra Pé de Urubu	N+Pr+Top+D1+DZ1+N1Z+D2+g2+N2Z-N2p:fs

Compostos toponímicos com estrutura NAdj:

1. Córrego Carioca	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
2. Córrego Fundo	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
3. Córrego Pelado	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
4. Córrego Queimado	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
5. Ilha Bela	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
6. Ilha Cinzenta	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
7. Ilha Grande	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
8. Ilha Rasa	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
9. Ilha Solteira	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
10. Lagoa Grande	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
11. Lagoa Vermelha	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
12. Maciço Central	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
13. Morro Agudo	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:ms
14. Morro Alagoano	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
15. Morro Grande	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:ms
16. Morro Seco	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:ms
17. Pedra Preta	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
18. Ponta Formosa	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
19. Praia Comprida	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
20. Praia Grande	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
21. Praia Mole	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
22. Rio Calçado	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
23. Rio Claro	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
24. Rio Fundão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
25. Rio Indiviso	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
26. Rio Marinho	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
27. Rio Melgaço	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
28. Rio Preto	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
29. Serra Escalvada	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z-N2p:fs

Compostos toponímicos com estrutura NdeN:

1. Alto de Santa Helena	N+Pr+Top+DZ1+N1Z+DZ2+g1-N2Z-N2p:ms
2. Alto do Suá	N+Pr+Top+DZ1+N1Z+D2+g1-N2Z-N2p:ms
3. Baía de Vitória	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
4. Barra do Jucu	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
5. Canal dos Escravos	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:ms
6. Córrego da Forquilha	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
7. Córrego da Poça	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
8. Córrego da Ribeira	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
9. Córrego de Boa Vista	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
10. Córrego do Bonito	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
11. Córrego do Congo	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
12. Córrego do Moinho	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
13. Córrego do Sabão	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
14. Córrego do Tanque	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
15. Enseada de Jacaraípe	N+Pr+Top+D1+N1Z+DZ2+g2+N2Z-N2p:fs
16. Enseada do Suá	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
17. Ilha da Baleia	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
18. Ilha da Forca	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
19. Ilha da Fumaça	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
20. Ilha da Galeta de Dentro	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
21. Ilha da Galeta de Fora	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
22. Ilha da Madeira	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
23. Ilha da Palha	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
24. Ilha da Pesca	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
25. Ilha da Pólvora	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
26. Ilha da Rainha	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
27. Ilha das Andorinhas	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
28. Ilha das Caieiras	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z+N2p:fs
29. Ilha das Cobras	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
30. Ilha das Garças	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
31. Ilha das Pombas	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
32. Ilha das Tendas	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
33. Ilha de Monte Belo	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g2+N2Z-N2p:fs
34. Ilha de Santa Maria	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
35. Ilha de Trindade	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
36. Ilha de Vitória	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
37. Ilha do Apicum	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
38. Ilha do Araçá	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
39. Ilha do Bode	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
40. Ilha do Boi	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
41. Ilha do Caju	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
42. Ilha do Cal	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
43. Ilha do Campinho	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
44. Ilha do Cavalo	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
45. Ilha do Cercado	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
46. Ilha do Fato	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
47. Ilha do Frade	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
48. Ilha do Gaspar	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs

49. Ilha do Grajuru	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
50. Ilha do Igarapé	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
51. Ilha do Lameirão	N+Pr+Top+D1+DZ1+N1Z+D2+g2+N2Z-N2p:fs
52. Ilha do Meio	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
53. Ilha do Papagaio	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
54. Ilha do Paraíso	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
55. Ilha do Pato	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
56. Ilha do Príncipe	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
57. Ilha do Sapo	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
58. Ilha do Socó	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
59. Ilha do Sururu	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
60. Ilha do Urubu	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
61. Ilha dos Índios	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
62. Ilha dos Itaitis	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z+N2p:fs
63. Ilha dos Pacotes	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
64. Ilha dos Práticos	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z+N2p:fs
65. Lagoa de Jabaeté	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
66. Lagoa de Jacarenema	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
67. Lagoa de Jacuném	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
68. Lagoa do Cocal	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
69. Lagoa do Queimado	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
70. Morro da Aparecida	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
71. Morro da Areia Preta	N+Pr+Top+D1+N1Z+D2+DZ2+g1+g2+N2Z-N2p:ms
72. Morro da Argola	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
73. Morro da Barrinha	N+Pr+Top+D1+N1Z+D2+DZ2+g1+g2+N2Z-N2p:ms
74. Morro da Bela Vista	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
75. Morro da Camela	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
76. Morro da Cangaiba	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
77. Morro da Capixaba	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
78. Morro da Cavada	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
79. Morro da Chapada	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
80. Morro da Concha	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
81. Morro da Coréa	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
82. Morro da Coréia	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
83. Morro da Gameia	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
84. Morro da Gamela	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
85. Morro da Ilha do Príncipe	N+Pr+Top+D1+N1Z+D2+DZ2+g1+g2+N2Z-N2p:ms
86. Morro da Lagoa	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
87. Morro da Mantegueira	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
88. Morro da Palha	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
89. Morro da Pedra do Bode	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:ms
90. Morro da Penha	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
91. Morro da Piedade	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
92. Morro da Torre	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
93. Morro da Vamela	N+Pr+Top+D1+N1Z+D2+g1+g2+N2Z-N2p:ms
94. Morro da Vigia	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
95. Morro das Araras	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z+N2p:ms
96. Morro de Itapebuçu	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
97. Morro de Itararé	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
98. Morro de Jucutuquara	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms

99. Morro de Santa Helena	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
100. Morro de São Benedito	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
101. Morro de São Francisco	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
102. Morro de São José	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
103. Morro do Alto de Caratoína	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:ms
104. Morro do Bananal	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
105. Morro do Barro Vermelho	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
106. Morro do Bode	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
107. Morro do Bonfim	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
108. Morro do Cabral	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
109. Morro do Céu	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
110. Morro do Convento	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
111. Morro do Cruzamento	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
112. Morro do Exército	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
113. Morro do Grajuru	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
114. Morro do Itapenambi	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
115. Morro do Macaco	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
116. Morro do Martelo	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
117. Morro do Meio	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
118. Morro do Moreno	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
119. Morro do Moscoso	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
120. Morro do Pão de Açúcar	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
121. Morro do Penedo	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
122. Morro do Pico	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
123. Morro do Pinto	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
124. Morro do Quadro	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
125. Morro do Quartel	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
126. Morro do Reservatório	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
127. Morro do Rio Branco	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:ms
128. Morro do Romão	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
129. Morro do Suá	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
130. Morro do Vilante	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:ms
131. Morro dos Bastos	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
132. Pedra da Baleia	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
133. Pedra da Pescaria	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
134. Pedra do Frade	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
135. Pedra do Martha	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
136. Pedra do Tirol	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
137. Pedra dos Búzios	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:fs
138. Pedra dos Olhos	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:fs
139. Pedra dos Ovos	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:fs
140. Pico da Pedra Azul	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:ms
141. Ponta do Fato	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
142. Praia da Baleia	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
143. Praia da Barrinha	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:fs
144. Praia da Castanheira	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
145. Praia da Chaleirinha	N+Pr+Top+D1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
146. Praia da Costa	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
147. Praia da Enseada	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:fs
148. Praia da Fruta	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs

149. Praia da Sereia	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
150. Praia de Bicanga	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
151. Praia de Coaral do Meio	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:fs
152. Praia de Coqueiral	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
153. Praia de Interlagos	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z+N2p:fs
154. Praia de Itapoã	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
155. Praia de Jacaraípe	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
156. Praia de Manguinhos	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
157. Praia de Santa Monica	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
158. Praia do Barrão	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
159. Praia do Barrote	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
160. Praia do Canto	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
161. Praia do Coqueiral	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
162. Praia do Governador	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:fs
163. Praia do Itapoã	N+Pr+Top+D1+N1Z+DZ2+g1+N2Z-N2p:fs
164. Praia do Nenei	N+Pr+Top+D1+N1Z+D2+g1+N2Z-N2p:fs
165. Praia do Ribeiro	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
166. Praia do Solemar	N+Pr+Top+D1+N1Z+D2+DZ2+g2+N2Z-N2p:fs
167. Praia do Suá	N+Pr+Top+D1+N1Z+D2+g2+N2Z-N2p:fs
168. Praia dos Pacotes	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:fs
169. Praia dos Recifes	N+Pr+Top+D1-N1Z+g1+N2Z+N2p:fs
170. Rio da Draga	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
171. Rio da Passagem	N+Pr+Top+D1-N1Z+g1+N2Z-N2p:ms
172. Serra da Andorinha	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
173. Serra da Pedra Azul	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
174. Serra da Samambaia	N+Pr+Top+D1+DZ1+N1Z+D2+DZ2+g1+N2Z-N2p:fs
175. Serra da Santa Lúcia	N+Pr+Top+D1+DZ1+N1Z+DZ2+g1+N2Z-N2p:fs
176. Serra do Adriano	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
177. Serra do Anil	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
178. Serra do Batatal	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs
179. Serra do Mestre Álvaro	N+Pr+Top+D1+DZ1+N1Z+D2+g1+N2Z-N2p:fs
180. Serra do Redentor	N+Pr+Top+D1+DZ1-N1Z+g1+N2Z-N2p:fs

Os nomes *Alto de Santa Helena* e *Alto do Suá* foram inseridos na listagem dos compostos toponímicos com estrutura NdeN por considerarmos que nesses casos *alto* funciona como nome, em detrimento da utilização do substantivo *morro*. A utilização de *alto* em lugar do substantivo *morro* fica evidenciada na redação dos artigos primeiro e segundo da lei 5.543:

“**Art. 1º.** O Morro de Santa Helena, na Praia do Suá, passa a denominar-se Alto de Santa Helena.

Art. 2º. Ficam mantidos os limites e as confrontações atuais do Morro de Santa Helena.”

8 – ALGUMAS CONSIDERAÇÕES

Apresentamos e formalizamos a descrição de compostos toponímicos dos municípios da Grande Vitória (Cariacica, Serra, Viana, Vila Velha e Vitória). Foram selecionados os nomes formados por mais de um elemento em que um substantivo indicasse o tipo de acidente natural (morro, pedra, praia, rio, serra etc.).

8.1 – Propriedades distribucionais dos compostos toponímicos e sua formalização

A utilização de critérios formais e a observação das propriedades distribucionais das sequências constituídas de um elemento genérico e um elemento específico (simples ou composto) possibilita a diferenciação entre os compostos toponímicos e os grupos nominais livres que apresentam a mesma estrutura. Procedeu-se a descrição dos compostos toponímicos e a codificação dos resultados de acordo com os pressupostos teórico-metodológicos do léxico-gramática, assim sendo, as sequências candidatas a compostos toponímicos foram analisadas a partir da sua inserção em uma frase elementar (constituída de sujeito, verbo e complementos essenciais) e não pela comparação entre grupos nominais descontextualizados.

Os compostos comuns podem apresentar transparência ou opacidade semântica de N1, por exemplo, N1 é transparente no composto *chinelo de dedo* (um *chinelo de dedo* é um *chinelo*) e opaco em *lua-de-mel* (uma *lua-de-mel* não é uma *lua*). No primeiro caso há possibilidade do apagamento de N2 (Ana calçou o *chinelo de dedo*. / Ana calçou o *chinelo*.) o mesmo geralmente não ocorre com os compostos em que o N1 é semanticamente opaco (Ana e Moisés estão em *lua-de-mel*. / *Ana e Moisés estão em+na *lua*). A estrutura geral dos compostos toponímicos (Nc+Npr ou elemento genérico + elemento específico) confere a esses nomes características próprias, o N1 é semanticamente transparente e contém o significado geral do composto, por exemplo, a *baía de Vitória* (Espírito Santo) é uma baía, é uma porção de mar (no caso, do oceano Atlântico) rodeada por terra. Há, portanto, a

possibilidade de elisão de N2 nos compostos toponímicos (sua forma longa) de estrutura NN e NAdj e NdeN. Considerando que o N2 é o elemento identificador de N1 pode-se inferir o N1 a partir do N2, desde que haja o conhecimento prévio da combinação, da forma longa do composto. Logo, é possível a elisão do N1 na forma longa compostos toponímicos de estrutura NN e NAdj em que o Adj seja relacional, ou classificador.

Para a averiguação das propriedades morfossintáticas e semânticas dos compostos toponímicos constatou-se a necessidade de utilização de critérios diferenciados, visto que essas propriedades não são idênticas em todos os tipos de compostos.

8.2 – A descrição linguística dos nomes próprios compostos para o PLN

O fato de se atribuir, por vezes, pouca relevância ao estudo dos nomes próprios para a teoria linguística pode conduzir à falsa compreensão de que a descrição e formalização desses itens lexicais para o PLN é tarefa desnecessária. Segundo Oliveira e Freitas (2006: 186)

a compreensão de nome próprio é crucial na análise de textos [...].

O fato de tais nomes constituírem uma classe ainda mais ‘aberta’ do que a dos substantivos comuns, uma vez que novos nomes são criados a todo momento, deixa ainda mais evidente a necessidade de atualização constante e, conseqüentemente, de metodologias capazes [*sic*] acrescentar nomes – e suas classes semânticas – automaticamente.

A ocorrência de compostos toponímicos em textos jornalísticos, em guias turísticos, textos de História e de Geografia, entre outros, atesta a necessidade da identificação e etiquetagem adequada desses nomes para o processamento automático de textos em língua natural.

A inserção de novas informações nos dicionários do sistema DELA é facilitada por causa da independência das entradas lexicais e de seus conteúdos. O sistema Unitex utiliza tanto dicionários como grafos dicionários no processamento de textos em língua natural. Os grafos dicionários possibilitam a realização de buscas levando-se em conta a diferenciação entre letras maiúsculas e letras minúsculas. Os grafos dicionários “peuvent exploiter les résultats fournis par les dictionnaires appliqués précédemment. Ainsi, on peut appliquer le dictionnaire

général, puis étiqueter comme noms propres les mots inconnus commençant par une majuscule à l'aide du graphe NPR+” (PAUMIER, 2006: 47).

Após a aplicação dos dicionários o sistema Unitex exibe o resultado da busca através de listas (de palavras simples, de palavras compostas e de palavras desconhecidas) em ordem alfabética. A identificação das palavras desconhecidas começadas por uma maiúscula facilita a etiquetagem desses nomes através de um grafo apropriado, este expediente é eficaz, principalmente quanto aos compostos toponímicos que apresentam um antropônimo na sua formação. O Unitex oferece ainda recursos que podem ser utilizados na expansão de seus dicionários.

A descrição linguística dos nomes compostos e dos nomes próprios é muitas vezes classificada como inviável devido ao surgimento de novas combinações de palavras (não-composicionais) e à utilização de substantivos comuns como nomes próprios, o que torna a descrição laboriosa e carente de atualizações constantes. Os numerosos estudos dedicados ao fenômeno da composição demonstram a possibilidade de pesquisas quanto a essas formações, o mesmo pode-se dizer acerca dos nomes próprios. Os recursos advindos do desenvolvimento tecnológico, de sistemas informáticos (a disponibilização de um número maior de *corpora* e léxicos em suporte eletrônico) auxiliam na tarefa de descrição linguística para o aperfeiçoamento desses sistemas. O que se observa é a necessidade do engajamento de um maior número de linguistas na descrição do português do Brasil para o processamento automático. A descrição linguística para o PLN é por vezes classificada como uma atividade laboriosa e entediante, esse tipo de observação é desprovida de rigor científico, conforme afirma Laporte (2009: 68-69):

On of the most commonplace phrases in scientific articles about language processing is *tedious and time-consuming*. Authors use it to explain their systems do not make use manually constructed resources: their construction would be tedious and time-consuming. [...].

This argument shows a striking lack of scientific rigour.

Os pressupostos teórico-metodológicos do léxico-gramática incluem o uso concomitante de construção de frases e verificação em *corpus*. A suposta falta de entusiasmo do pesquisador em relação as descrição linguística para o PLN por vezes está associada a sua falta de habilidade para o desenvolvimento de tal tarefa ou a outros fatores de ordem psíquica que nada têm a ver com o método empregado.

8.3 – Trabalhos futuros

O desenvolvimento da pesquisa aponta a possibilidade de outros estudos quanto aos compostos toponímicos, por exemplo, a substituição do elemento genérico *morro*, por *bairro*¹⁰; a ocorrência de topônimos transplantados e seu significado cultural¹¹; o levantamento de áreas de ocorrência de determinado termo específico. Para o PLN, considerando-se o *corpus* desta pesquisa, pode-se observar as formar variantes dos compostos toponímicos, como o apagamento da preposição *de*, presença/ausência do artigo definido anteposto ao N2 nos compostos NdeN. Há ainda a necessidade de formalização e codificação dos demais compostos toponímicos do Espírito Santo (tanto acidentes geográficos naturais como artificiais) e dos demais estados brasileiros. Observou-se a possibilidade de estudo dos chamados verbos de movimento em relação aos compostos toponímicos e a obrigatoriedade da presença/ausência das preposições locativas.

O estudo descritivo dos compostos toponímicos capixabas para fins computacionais, permitiu a observação de possibilidades de pesquisas linguísticas (na Toponímia, na Análise do Discurso, na Linguística Computacional etc.) e a constatação da importância da presença do profissional em Língua Portuguesa do Brasil em equipes multidisciplinares como, por exemplo, a sua cooperação para pesquisas que permitam o resgate histórico e construção da identidade capixaba através dos nomes de lugares (atribuídos de forma sistemática ou espontânea), ou ainda sua participação na formação de uma comissão nacional de padronização de nomes de lugares considerando-se as normas gramaticais e a língua em uso. A proposta de padronização dos topônimos não consiste numa tentativa de refrear as mudanças que ocorrem na língua, mas de registrá-las para que o processamento automático de textos em linguagem natural ocorra de forma mais eficiente.

Considerando-se a observação de Biderman (2001b: 131-133), que os dicionários “constituem uma organização sistemática do léxico, uma espécie de tentativa de descrição do léxico de uma língua” e que “nenhum dicionário por mais volumoso que seja, dará conta integral de uma língua de civilização”, pode-se afirmar que a construção de dicionários eletrônicos para o

¹⁰ LIMA, Rosana de Vilhena. Geomorfotopônimos da cidade de Vitória - a substituição da denominação MORRO por BAIRRO. **ABRALIN em Cena Espírito Santo**, João Pessoa, p. 340-347, 2009.

¹¹ LIMA, Rosana de Vilhena. Topônimos transplantados: a Europa em Vitória. **Anais do II Simpósio Nacional de Estudos Filológicos e Linguísticos**. Rio de Janeiro, v. 02, p. 01-09, 2009.

português do Brasil e a sua expansão (inserção de novos itens lexicais bem como a atualização e acréscimo de informações linguísticas) franqueia aos profissionais da Língua Portuguesa no Brasil espaço para o desenvolvimento de novas pesquisas para o PLN.

REFERÊNCIAS

BAPTISTA, Jorge. **Estabelecimento e formalização de classes de nomes compostos**. 1994. Dissertação (mestrado) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 1994.

BASÍLIO, Margarida Maria de Paula. Em torno da palavra como unidade lexical: palavras e composição. **Veredas**, Juiz de Fora, v. 4, n. 2, p. 9-18, 2000.

_____. **Formação e classes de palavras no Português do Brasil**. 1. ed. São Paulo: Contexto, 2004.

_____. **Teoria Lexical**. 8. ed. São Paulo: Ática, 2007.

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

BELLEIL, Claude; MAUREL, Denis. Un dictionnaire relationnel des noms propres liés à la géographie consulté par transducteurs, **Meta**, v. 42, n. 2, p. 273-282, 1997.

BIDERMAN, Maria Tereza Camargo. As ciências do léxico. In: OLIVEIRA, Ana Maria P. P. de; ISQUERDO, Aparecida Negri (Org.). **As ciências do léxico**. Lexicologia, lexicografia, terminologia. 2. ed., Campo Grande: Ed. UFMS, p. 11-20, 2001a.

_____. Os dicionários na contemporaneidade: arquitetura, métodos e técnicas. In: OLIVEIRA, Ana Maria P. P. de; ISQUERDO, Aparecida Negri (Org.). **As ciências do léxico**. Lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: Ed. UFMS, p. 131-144, 2001b.

_____. **Teoria lingüística: teoria lexical e lingüística computacional**. 2. ed. São Paulo: Martins Fontes, 2001c.

BORBA, Francisco S.. **Uma gramática de valências para o português**. São Paulo: Ática, 1996.

BRITO, Adriano Naves de. **Nomes próprios: semântica e ontologia**. Brasília: UnB, 2003.

CARVALHINHOS, Patrícia de Jesus. Variantes lexicais na toponímia portuguesa: os elementos genéricos (entidades geográficas) denominados. Estudo de caso: diferenças terminológicas entre português do Brasil e português europeu. **Cadernos do CNLF**, Rio de Janeiro, v. 11, n. 11, p. 177-194, 2007a.

_____; ANTUNES, Alessandra Martins. Princípios teóricos de Onomástica. Toponímia e Antroponímia - O nome próprio. **Cadernos do CNLF**, Rio de Janeiro, v. 11, n. 02, p. 108-121, 2007b.

CONSTANT, Matthieu. On the analysis of locative prepositional phrases: the classifier/proper noun pairing, **Lecture Notes in Artificial Intelligence (LNAI) 2389, International conference Portugal for Natural Language Processing (PorTAL)**, Faro, Berlin, p. 33-42, 2002.

CONSTANT, Matthieu **Grammaires locales pour l'analyse automatique de textes: Méthodes de construction et outils de gestion**. 2003. Thèse (doctorat) – Université de Marne-la-Vallée, Paris, 2003.

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

CRYSTAL, David. **A lingüística**. Lisboa: Dom Quixote, 1977.

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

DIAS, Maria Carmelita Pádua. **O léxico em sistemas de análise e geração automática e textos em língua portuguesa**. 1994. Tese (doutorado) – Programa de Pós-Graduação em Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1994.

DIAS DA SILVA, Bento Carlos. O estudo Lingüístico-Computacional da Linguagem. **Letras de Hoje**. Porto Alegre. v. 41, n. 2, p. 103-138, 2006.

DICK, Maria Vicentina de Paula do Amaral. Origens Históricas da Toponímia Brasileira. Os Nomes Transplantados. **Revista do Instituto de Estudos Brasileiros**, São Paulo, v. 24, p. 75-96, 1982.

_____. Os Vocábulos Toponímicos Básicos no “Vocabulário da Língua Brasília” e sua Relação Geográfica. **Revista do Instituto de Estudos Brasileiros**, São Paulo, v. 31, p. 95-112, 1990a.

_____. **Toponímia e Antroponímia do Brasil**. Coletânea de Estudos. 2. ed. São Paulo: Serviço de Artes Gráficas da FFLCH/USP, 1990b.

_____. A Projeção Léxico-Cultural na Onomástica. **Revista do Instituto de Estudos Brasileiros**, São Paulo, v. 40, p. 161-173, 1996.

_____. **A dinâmica dos nomes na cidade de São Paulo: 1554-1897**. 2. ed. São Paulo: Annablume, 1997.

_____. Fundamentos teóricos da Toponímia. Estudo de caso: o Projeto ATEMIG - Atlas Toponímico do Estado de Minas Gerais - Variante Regional do Atlas Toponímico do Brasil. In: SEABRA, Maria Cândida T. C. de (Org.). **O Léxico em Estudo**. Belo Horizonte: FALE UFMG, p. 91-119, 2006.

_____. Atlas toponímico do Brasil: Teoria e Prática II. **Trama**, v. 3, n. 5, p. 141-155, 2007.

DUBOIS, Jean. **Dicionário de lingüística**. 8. ed. São Paulo: Cultrix, 2001.

FERREIRA, Aurélio Buarque de Holanda. **Minidicionário da língua portuguesa**. 2. ed. Rio de Janeiro: Nova Fronteira, 1988.

FREITAS, Maria Cláudia de. **Elaboração automática de ontologias de domínio**: discussão e resultados. 2007. Tese (doutorado) – Programa de Pós-Graduação em Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

FRIBURGER, Nathalie. **Reconnaissance automatique des noms propres** – Application à la classification automatique de textes journalistiques. 2002. Thèse (doctorat) – Université François Rabelais, Tours, 2002.

GARRIGUES, Mylène. Prépositionns et noms de pays et d'iles: une grammaire locale pour l'analyse automatique des textes. **Lingvistic Investigations**, Amsterdam/Filadélfia, v. 11, n. 1, p. 281-305, 1993.

GROSS, Gaston. Définition des mots composés dans un lexique-grammaire. **Langue française**, v. 87, n. 1, p. 84-90, 1990.

GROSS, Maurice. **Grammaire transformationnelle du français - syntaxe du verbe**. v. 1, Paris: Larousse, 1968.

_____; LENTIN André. **Notions sur les grmmaires formelles**. 2. ed. rev. e cor. Paris: Gauthier-Villars, 1970.

_____. Grammaire transformationnelle et enseignement du français. **Langue française**, v. 11, n. 1, p. 4-14, 1971.

_____. **Méthodes en syntaxe**. Paris: Hermann, 1975.

_____. **Grammaire transformationnelle du français - syntaxe du nom**. v. 2, Paris: Larousse, 1977.

_____. Les bases empiriques de la notion de prédicat sémantique, **Langages**, v. 63, p. 07-52, Paris: Larousse, 1981.

_____. On structuring the lexicon. **Quaderni di Semantica**, v. 4, n. 1, p.107-120, 1983.

_____. A linguistic environment for comparative Romance syntax. In: **Papers from the XIIth Linguistic Symposium on Romance Languages**, Ph. Baldi (ed.), Amsterdam studies en the theory and history of linguistic scince IV (26), Amsterdam/Filadélfia: Benjamins, p. 373-446, 1984.

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

_____. Methods and tactics in the construction of a Lexicon-Grammar. In: **Linguistics in the Morning Calm 2**, Selected papers from SICOL, Seoul: Hanshin, p. 177-197, 1986b.

_____. La constuction de dictionnaires életroniques. **Annales des Télécommunications**, Issy-les-Moulineaux/Lannion, v. 44, n. 1-2, p. 4-19, 1989.

_____. Le programme d'extension des lexiques életroniques. **Langue Française**, Paris, n.

87, p. 123-127, 1990a.

_____. Sur la notion harrissienne de transformation et son application au français. **Langages**, v. 25, n. 99, p. 39-56, 1990b.

_____. Linguistics representations and text analysis. In: **Linguistic Unity and Linguistic Diversity in Europe**. London: Academia Europaea, p. 31-61, 1991.

HARRIS, Zellig Sabbetai. **Methods in Structural Linguistics**. Chicago: University of Chicago Press, 1951.

_____. Discourse Analysis. **Langages**, v. 28, n. 1, p. 1-30, 1952.

_____. Elementary transformations, Philadelphie: University of Pennsylvania, **Transformations and Discourse Analysis Papers** 54. 1964. Reimpresso em **Papers on Syntax. Structural and Transformational Linguistics**. Henry Hiz, Reidel: Dordrecht, p. 211-235, 1970.

_____. Du morphème à l'expression. **Langages**, v. 3, n. 9, p. 23-50, 1968.

_____. **A grammar of English on mathematical principles**. New York: Wiley, 1982.

_____. La genèse de l'analyse des transformations et de la métalangue. **Langages**, v. 25, n. 99, p. 9-20, 1990.

HOUAISS, Antônio; VILLAR, Mauro de Salles. **Grande dicionário Houaiss da Língua Portuguesa**, Instituto Antônio Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa, Rio de Janeiro: Objetiva, 2001.

LAPORTE, Eric. Resolução de ambigüidades. In: RANCHHOD, Elisabete Marques. **Tratamento das línguas por computador**. Uma introdução à lingüística computacional e suas aplicações. Lisboa: Caminho, 2001.

_____. Exemplos atestados e exemplos construídos na prática do léxico-gramática. **(Con)Textos Lingüísticos**, Vitória, v. 2, p. 26-51, 2008.

_____. Lexicons and grammars for language processing: industrial or handcrafted products? In: REZENDE, Letícia Marcondes (org.). **Léxico e gramática: dos sentidos à construção de significação**. São Paulo: Cultura Acadêmica, 2009.

LUFT, Celso Pedro. **Minidicionário Luft**. 20. ed. São Paulo: Ática, 2001.

LYONS, John. **Semântica**. Lisboa: Presença/Martins Fontes, 1980.

MAUREL, Denis; PITON, Odile. Un dictionnaire de noms propres pour Intex: Les noms propres géographiques. **Lingvisticæ Investigationes**, Faison, n. 22, p. 279-289, 1999.

MELO, Gladstone Chaves de. **Gramática fundamental da língua portuguesa**. 3. ed. Rio de Janeiro: Ao Livro Técnico, 1978.

MENEZES, Paulo M.L.; SANTOS, Cláudio J.B. **Portal da Cartografia**. Londrina, v.1, n.1, p. 75 - 92, 2008. Disponível em: <www.uel.br/revistas/uel/index.php/portalcartografia> Acesso em: 30 de outubro de 2007.

MONTEIRO, José Lemos. **Morfologia portuguesa**. Fortaleza: EDUFC, 1987.

MONTELEONE, Mario. **Lexicographie et dictionnaires électroniques. Des usages linguistiques aux bases de données lexicales**. 2003. Thèse (doctorat). Université de Marne-la-Vallée, 2003.

MUNIZ, Marcelo Caetano Martins. **A construção de recursos linguísticos-computacionais para o português do Brasil: o projeto Unitex-PB**. 2004. Dissertação (mestrado) – Instituto de Ciências Matemáticas e da Computação, Universidade de São Paulo, São Carlos, 2004.

NEVES, Maria Helena de Moura. **A gramática funcional**. São Paulo: Martins Fontes, 2001.

NICOLAISEN, W. F. H. Onomastic Dialects. **American Speech**, v. 55, n. 1, p. 36-45, 1980.

OLIVEIRA, Cláudia; FREITAS, Maria Cláudia de. Classes de palavras e etiquetagem na Lingüística Computacional. **Calidoscópico**, Porto Alegre, v. 4, n. 3, p. 179-188, 2006.

PAUMIER, Sébastien. **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 de dezembro de 2009.

RANCHHOD, Elisabete Marques. O uso de dicionários e de autômatos finitos na representação lexical das línguas naturais. In: RANCHHOD, Elisabete Marques. (Org.). **Tratamento das línguas por Computador: uma introdução à Lingüística Computacional e suas aplicações**. Lisboa: Caminho. 2001.

RIO-TORTO, Graça. Para uma gramática do Adjectivo. **Alfa**, v. 50, n. 2, p. 103-129, 2006.

ROCHA LIMA, Carlos Henrique da. **Gramática da língua portuguesa**. 44. ed. Rio de Janeiro: José Olympio, 2005 [31. ed. 1991].

RUSSEL, Stuart Jonathan. **Inteligência artificial**. Tradução da 2. ed. Rio de Janeiro: Elsevier, 2004.

SAID ALI, Manoel. **Gramática secundaria e gramática histórica da língua portuguesa**. 3. ed. Brasília: UnB, 1964.

SANDMANN, Antônio José. **Formação de palavras no português brasileiro contemporâneo**. Scientia et Labor/Ícone, 1988.

_____. **Competência lexical**. Curitiba: UFPR, 1991a.

_____. **Morfologia geral**. São Paulo: Contexto, 1991b.

_____. **Morfologia lexical**. São Paulo: Contexto, 1992.

SANTOS, Cláudio João Barreto dos. **Geonímia do Brasil: a padronização dos nomes geográficos num estudo de casos dos municípios fluminenses**. 2008. Tese (doutorado) – Programa de Pós-Graduação em Geografia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

SAVARY, Agata. **Recensement et description des mots composés – méthodes et applications**. 2000. Thèse (doctorat) – Université de Marne-la-Vallée, 2000.

SMARSARO, Aucione das Dores. **Descrição e formalização de palavras compostas do português do Brasil para a elaboração de um dicionário eletrônico**. 2004. Tese (doutorado) – Programa de Pós-Graduação em Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

TRAN, Mickäel; GRASS, T.; MAUREL, Denis. **An ontology for multilingual treatment of proper names**, Ontologies and Lexical Resources in Distributed Environments (OntoLex 2004), in Association with LREC2004 (Actes p. 75-78), Lisboa, Portugal, 2004.

TRAN, Mickäel; MAUREL, Denis. Prolexbase - Un dictionnaire relationnel multilingue de noms propres, **TAL**, v. 47, 2006.

ULLMANN, Stephen. **Semântica: uma introdução a ciência do significado**. 2. ed. - Lisboa: Fundação Calouste Gulbenkian, 1964.

VALE, Oto Araújo. **Expressões cristalizadas do português do Brasil: uma proposta de tipologia**. Araraquara. 2001. Tese (Doutorado) – Universidade Estadual Paulista, Araraquara, 2001.

VAXELAIRE Jean Louis. Nom propre et lexicographie française. **Corela**, Le traitement lexicographique des noms propres. 2005. Disponível em: <<http://edel.univ-poitiers.fr/corela/document.php?id=444>>. Acesso em: 01 de agosto de 2008.

VIVÈS, Robert. Les composés nominaux par juxtaposition. **Langue Française**, Paris, n. 87, pp. 98-103, 1990.

VITÓRIA (ES). **Lei nº 2.523, de 1977**. Vitória, 1977. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

VITÓRIA (ES). **Lei nº 3.424, de 1987**. Vitória, 1987. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

VITÓRIA (ES). **Lei nº 3.438, de 1987**. Vitória, 1977. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

VITÓRIA (ES). **Lei nº 5.515, de 2006**. Vitória, 2006. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

VITÓRIA (ES). **Lei nº 5.543, de 2002**. Vitória, 2002. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

WHITBY, Blay (1953). **Inteligência artificial**: um guia para iniciantes. Tradução de Cláudio Blanc. São Paulo: Medras, 2004.

ANEXO A

Referências das sentenças retiradas da *internet* e outras fontes.

1 – INTRODUÇÃO

(1) “Um israelense foi condenado a nove meses de prisão por atirar uma cabeça de porco em uma mesquita de Tel Aviv em agosto de 2005”.

Disponível em: <<http://g1.globo.com>>. Acesso em: 12 de junho de 2009.

(2) “Moravam até então, numa cabeça de porco: quarto e banheiro num imenso casarão do século dezoito, onde se amontoavam mais de vinte famílias – bairro da Glória”.

ALMEIDA, Ronaldo José de Almeida. **Os mineiros**. Disponível em: <<http://recantodasletras.uol.com.br>>. Acesso em: 15 de junho de 2009.

(3) “Aqui no Bairro Praia da Costa acontecem os principais eventos da região de Vila Velha como a apresentação de bandas e shows”.

Disponível em: <www.praiadacosta.imb.br>. Acesso em: 09 de outubro de 2009.

(4) “Até lá, a criançada também vai poder dar seus pulos na piscina do Happy Ball, no Shopping Praia da Costa, que terá recreação especial amanhã e domingo”.

Disponível em: <<http://gazetaonline.globo.com>>. Acesso em: 09 de outubro de 2009.

(5) “**Artesãos Associados - Exposição ‘O artesanato na decoração’ no Praia da Costa**”.

Disponível em: <<http://gazetaonline.globo.com>>. Acesso em: 09 de outubro de 2009.

(6) “Rian e Richard realizam show gratuito nas areias da Praia da Costa”.

Disponível em: <<http://gazetaonline.globo.com>>. Acesso em: 09 de outubro de 2009.

(7) “O ponto alto da prova foi à passagem por um córrego fundo e estreito [...]”.

AKIYAMA, Adriano. **2º Etapa Copa Vale Raid** - São José dos Campos / SP. Disponível em: <www.crossnaveia.com.br>. Acesso em: 16 de abril de 2009.

(8) “Os imigrantes italianos se fixaram em Córrego Fundo a partir de 1877 [...]”.

INSTITUTO JONES DOS SANTOS NEVES. Disponível em: <www.ijsn.es.gov.br>. Acesso em: 18 de junho de 2008.

(9) “Temos também dezenas de córregos. Citamos os mais conhecidos: [...] CÓRREGO FUNDO, é afluente do rio Santa Maria”.

Disponível em: <<http://vitoria-es-brasil.com>>. Acesso em: 18 de junho de 2008.

(10) “Fazenda das Palmas:

Localização: Localidade de Córrego Fundo, 4 km do Centro”.

Disponível em: <www.brasilchannel.com.br>. Acesso em: 30 de abril de 2009.

2 – PRESSUPOSTOS TEÓRICO-METODOLÓGICOS

(1) “Por fim, os ingleses desembarcaram um carregamento variado de foguetes ‘Rapier’, sistemas de mísseis, barris de combustível, caixas de munições e água potável”.

PEPORTAGENS. revista **Veja**. 26 de maio de 1982. Disponível em: <<http://veja.abril.com.br>>. Acesso em: 16 de maio de 2009.

(2) “Delegação gremista desembarcou em Porto Alegre pela manhã”.

Jornal O PIONEIRO Caxias do Sul. Disponível em: <www.clicrbs.com.br>. Acesso em: 16 de maio de 2009.

(3) “[...] as bailarinas da Apress-Arte executaram uma coreografia [...]”.

Posted By: Apress-Arte **Last Edit:** 02 de abril de 2009. Disponível em: <www.alexpress-arte.pt>. Acesso em: 16 de maio de 2009.

(6) “O Maracanã tem um campo grande, é difícil qualquer time ficar fechado ali 90 minutos”.

RIZZO, Marcel; LOUZADA, Bruno. O Estado de São Paulo. Disponível em: <www.estadao.com.br>. Acesso em: 16 de maio de 2009.

(7) “Somente neste ano já foram retirados de circulação mais de 100 quilos de entorpecentes das rodovias federais capixabas. [...] Os detidos e a droga foram encaminhados ao DPJ de Campo Grande [...]”.

Disponível em: <www.folhaes.com.br>. Acesso em: 03 de janeiro de 2009.

3 – O NOME COMPOSTO

(1) “Um israelense foi condenado a nove meses de prisão por atirar uma cabeça de porco em uma mesquita de Tel Aviv em agosto de 2005”.

Disponível em: <<http://g1.globo.com>>. Acesso em: 12 de junho de 2009.

- (2) “Moravam até então, numa cabeça de porco: quarto e banheiro num imenso casarão do século dezoito, onde se amontoavam mais de vinte famílias – bairro da Glória”.

ALMEIDA. Ronaldo José de. **Os mineiros**. Disponível em: <<http://recantodasletras.uol.com.br>>. Acesso em: 15 de junho de 2009.

- (4) “Autoridades encontram caixa preta do Airbus acidentado nas Comores”.

Disponível em: <www.correiobraziliense.com.br>. Acesso em: 14 de outubro de 2009.

4 – O NOME PRÓPRIO

- (1) “Troca-troca na ponta da corrida eleitoral da Veneza brasileira”.

Disponível em: <<http://www.terra.com.br/istoe>>. Acesso em: 09 de junho de 2009.

- (2) “Fica denominada ‘Joaquim José da Vitória’, a rua que partindo a Av. Maruípe, ao lado da padaria Eliana, dá acesso ao morro conhecido por Engenharia, nesta capital”.

VITÓRIA (ES). **Lei nº 2.523, de 1977**. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 02 de dezembro de 2008.

- (3) “Fica denominada MARIA JOSÉ BALDAN a rua que partindo do início da Rua Joaquim José da Vitória [...] vai dar acesso ao morro do Bairro Engenharia Maruípe”.

VITÓRIA (ES). **Lei nº 3.424, de 1987**. Disponível em: <<http://sistemas.vitoria.es.gov.br>>. Acesso em: 02 de dezembro de 2008.

- (4) “Fica denominada travessa Coronel Manoel da Silva a servidão pública que dá acesso ao morro da Engenharia [...]”.

VITÓRIA (ES). **Lei nº 3.438, de 1987**. Vitória, 1987. Disponível em: <<http://siste.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

- (5) “Fica denominado ‘Bairro Engenharia’, o conglomerado residencial e comercial hoje denominado ‘Morro Engenharia’”.

VITÓRIA (ES). **Lei nº 5.515, de 2006**. Vitória, 2006. Disponível em: <<http://siste.vitoria.es.gov.br>>. Acesso em: 24 de setembro de 2008.

- (6) A Fazenda das Palmas fica na **Localidade de Córrego Fundo**, a 4 km do Centro de Aracruz.

“Fazenda das Palmas:

Localização: Localidade de Córrego Fundo, 4 km do Centro”.

Disponível em: <www.brasilchannel.com.br>. Acesso em: 30 de abril de 2009.

- (7) “A localidade rural de Cidade Nova da Serra, mais conhecida por Chapada Grande, distante 10 quilômetros do centro da Serra, terá sua rede de abastecimento de água”.

Disponível em: <www.serra.es.gov.br>. Acesso em: 10 de outubro de 2009.

6 – CRITÉRIOS DE IDENTIFICAÇÃO DOS COMPOSTOS TOPONÍMICOS

- (1) “O corpo de Edilma Maria Ferreira, 26, foi localizado próximo ao viaduto Oscar Brito, também conhecido como viaduto dos Cabritos”.

BRITO, Diana. Folha Online. Disponível em: <www1.folha.uol.com.br/>. Acesso em: 16 de junho de 2009.

- (2) “Ilha das Caieiras terá museu com histórias contadas pelos moradores”.

Disponível em: <www.folhavoria.com.br>. Acesso em: 05 de novembro de 2009.

- (3) “Lula passará réveillon na Ilha de Fernando de Noronha”.

Disponível em: <www.folhavoria.com.br>. Acesso em: 04 de novembro de 2009.

- (4) “Caso das Ilhas Cayman reforça sentença contra Dantas”.

Disponível em: <www.folhavoria.com.br>. Acesso em: 04 de novembro de 2009.

- (5) “Obra de drenagem vai desviar trânsito na Ilha das Caieiras”.

Disponível em: <www.folhavoria.com.br>. Acesso em: 05 de novembro de 2009.

- (7) “O rio Aribiri desaguava na baía de Vitória, onde ficava o porto de mesmo nome”.

Disponível em: <www.ceats.org.br>. Acesso em: 25 de abril de 2009.

- (8) “Outra alternativa é alugar um barco através de uma licença e visitar a Ilha Itatiaia [...]”.

Disponível em: <www.vvinflash.hpg.com.br>. Acesso em: 22 de junho de 2009.

- (9) “Morro São Benedito tem feira comunitária ambiental”.

Disponível em: <www.vitoria.es.gov.br>. Acesso em: 22 de junho de 2009.

- (11) “A Serra natural era dotada de diversidade ambiental, ou seja, Mata Atlântica, [...], praias, o monte Mestre Álvaro (833m) [...], etc.”.

Disponível em: <<http://www.serra.es.gov.br>>. Acesso em: 15 de dezembro de 2008.

(13) “Compramos um terreno para a congregação de Gaivotas, que passa a funcionar em Vila Nova. Vila nova de Vilha Velha, porque tem Vila Nova na Serra e Vila Nova em Cariacica, mas a nossa congregação fica em Vila Nova de Vila Velha”.

Informação verbal - Declaração do Pastor Eduardo Vieira Gomes em culto realizado no dia 05 de julho de 2008, na Igreja Evangélica Assembleia de Deus Nova Vida em Itapoã – ES.

(17) “Nas bacias do rios Jucu e Santa Maria, esta unidade ocorre ao longo do litoral. Alcançam maiores expressões à retaguarda das lagoas costeiras da Praia Grande”.

Disponível em: <http://www.seculodiario.com/cesan/bacia_do_rio_jucu.doc>. Acesso em: 15 de agosto de 2008.

(22) “Os índios chamavam a Ilha de Vitória de Guanaaní ou ‘Ilha do Mel’ pela beleza de sua geografia e amenidade do clima com a baía de águas viscosas e manguezal repleto de moluscos, peixes, pássaros e muita vida”.

Disponível em: <<http://www.vitoria.es.gov.br/secretarias/sedec/histvitoria.htm>>. Acesso em: 19 de dezembro de 2009.

(23) “A Praia de Itapoã também vai receber as novas lixeiras e contentores, já instalados na Praia da Costa”.

Disponível em: <<http://www.vilavelha.es.gov.br>>. Acesso em: 19 de dezembro de 2009.

7 – FORMALIZAÇÃO DA DESCRIÇÃO DOS COMPOSTOS TOPONÍMICOS

(2) “Em Ponto Alto, já são 36 afetadas pela enchente do Rio Jucu Braço Sul”.

Disponível em: <<http://www.folhavoria.com.br>>. Acesso em: 19 de dezembro de 2009.

(9) “Praia das Gaivotas recebe projeto Arte e Cultura nos bairros”.

Disponível em: <<http://www.vilavelha.es.gov.br>>. Acesso em: 19 de dezembro de 2009.

(10) “O projeto ‘Rua Segura’ também vem sendo desenvolvido no Bairro Praia do Canto, na cidade de Vitória”.

Disponível em: <<http://www.sindinoticias.com>>. Acesso em: 19 de dezembro de 2009.

(12) “Apartamento para locação no praia da costa [*sic*] em Vila Velha-ES, 120m², R\$ 1.800,00”.

Disponível em: <<http://imoveis.mitula.com.br/>>. Acesso em: 8 de março de 2010.