

## LÉXICO E SEMÂNTICA

### NOME PRÓPRIO: UMA ABORDAGEM PARA O PROCESSAMENTO DA LINGUAGEM NATURAL

Rosana de Vilhena Lima (UFES)  
[rvillema5@hotmail.com](mailto:rvillema5@hotmail.com)

#### INTRODUÇÃO

O processamento automático da linguagem natural (PLN) é um campo de pesquisa que exige conhecimentos de áreas diversas, sobretudo da linguística e da informática. Para que o homem faça um melhor uso da máquina, torna-se urgente a descrição das línguas naturais e sua formalização. Isso possibilitará uma “leitura e compreensão” das línguas naturais pela máquina, favorecendo o homem em grande medida, nos diversos aspectos da comunicação e da significação lexical, viabilizando a interação humano-máquina.

A elaboração de um dicionário eletrônico para o português do Brasil, que envolva uma ampla cobertura lexical, é uma das aspirações dos pesquisadores em linguística e informática envolvidos com o PLN. Vale ressaltar que a expressão *dicionário eletrônico* não diz respeito aos dicionários digitais – ou seja, às versões informatizadas de dicionários de uso – mas refere-se ao léxico computacional, que abarca o maior número de informações sobre os itens lexicais, codificadas para o uso em máquina no processamento automático de textos, dispensando assim a intervenção humana.

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.

A ocorrência de palavras que possuem a mesma forma e vários significados é uma das dificuldades a ser trabalhada na elaboração do dicionário eletrônico do português do Brasil. Esse fenômeno (polissemia) será abordado no artigo no que diz respeito aos nomes próprios.

A POLISSEMIA

A polissemia, ocorrência de um item lexical com vários significados, é um fato incontestável nas línguas naturais. De acordo com Lyons (1987, p. 143):

Os vários significados de um lexema polissêmico único [...] são normalmente tidos por relacionados entre si [...].

É fácil ver que, enquanto a identidade entre as formas é uma questão de sim ou não, a relação entre os significados é um problema de mais e ou menos.

Nesta perspectiva, pode-se dizer que todas as palavras são potencialmente polissêmicas. Considerando que a relação entre significados é expressa em termos de *mais* ou *menos*, pode-se pensar na existência de palavras *mais* ou *menos* polissêmicas.

Perini (1995, p. 251, 252) considera que a maior parte das palavras apresenta algum grau de polissemia e afirma que, “a polissemia é uma propriedade fundamental das línguas, que sem ela não poderiam funcionar eficientemente”. O potencial polissêmico das palavras é um dos fatores que atribui às línguas flexibilidade e possibilita a expressão dos diversos aspectos da realidade, segundo o linguista.

O autor considera problemático o critério da distinção semântica e apresenta o seguinte questionamento: “Como distinguir uma com segurança uma diferença semântica ‘grande’ de uma ‘pequena’?” (Perini, 1995, p. 251). São apresentados alguns casos em que se pode identificar a diferença de significado entre os itens lexicais (*fio* de linha – *fio* elétrico, *roda* de carroça – *roda* de amigos e outros) sem que se possa, contudo, determinar a dimensão dessa diferença.

Outro fenômeno abordado por gramáticos e linguistas é a *homonímia* – ocorrência de itens lexicais com a mesma grafia e pronúncia e significados completamente distintos – frequentemente confundida com a polissemia. Embora não seja objetivo do artigo tratar da diferença entre homonímia e polissemia, pode-se dizer que esse aspecto é extremamente relevante na construção de dicionários.

No caso de ocorrência de itens lexicais que apresentem a mesma forma (gráfica e fônica – homógrafos e homófonos), a defi-

## LÉXICO E SEMÂNTICA

nição da ocorrência de polissemia ou de homonímia determinará o número de entradas no dicionário. Ao considerar-se que há ocorrência de polissemia, o dicionário apresentará entrada única para o item com as acepções descritas nessa entrada, caso contrário – ao considerar-se que os itens são homônimos – o dicionário apresentará entradas distintas para cada item lexical. Um exemplo dessa dificuldade observa-se no tratamento do substantivo *manga* em dois minidicionários da língua portuguesa:

- *Minidicionário Aurélio* (Ferreira, 1988, p. 325)

“**manga**<sup>1</sup> *sf.* 1. Parte do vestuário onde se enfia o braço. 2. Qualquer peça em forma de tubo que reveste ou protege outra peça.

**manga**<sup>2</sup> *sf.* O fruto da mangueira.”

- *Minidicionário Luft* (Luft, 2001, p. 441)

“**man.ga** *s. f.* 1. Parte do vestuário na qual se enfia o braço. 2. Fruto da mangueira. 3. Chaminé de candeeiro. 4. Tromba-d’água.”

Embora a palavra *manga* seja um exemplo clássico de homonímia (Cf. Borba, 1976, p. 62 e Perini, 1995, p. 251) constata-se a falta de consenso na construção de dicionários. Ferreira (1988) considera a existência de dois itens lexicais ao passo que Luft (2001) aponta a ocorrência de apenas um item. Um fator que leva o lexicógrafo a optar por entradas polissêmicas é a economia de espaço, sobretudo em dicionários em suporte de papel.

Os dicionários de uso em suporte digital contam com um espaço maior, o que possibilita o registro homonímico das entradas bem como o acréscimo de informações acerca dos itens lexicais. Por exemplo, o dicionário Houaiss – formato digital<sup>22</sup> – no verbete *manga*, registra três entradas, com diversas acepções em cada uma delas, além de informações quanto à etimologia, gramática e ocorrência de homonímia.

Cabe ressaltar que, embora os dicionários em formato digital possibilitem uma quantidade maior de informações sobre as entradas, a natureza dessas informações difere das que se encontram nos dicionários eletrônicos.

---

<sup>22</sup> Disponível em: <http://houaiss.uol.com.br>.

Nos dicionários de uso as palavras são inseridas na sua forma canônica (por exemplo, os nomes no masculino singular e os verbos no infinitivo) enquanto nos dicionários eletrônicos as formas flexionadas também são consideradas. A inserção dessas formas favorece a ocorrência da ambiguidade. No caso do item lexical *manga*, além das realizações apontadas anteriormente, no dicionário eletrônico encontra-se, por exemplo, *manga* como a terceira pessoa do presente do indicativo do verbo *mangar*.

Na elaboração do dicionário eletrônico não há a preocupação com relação ao espaço a ser ocupado pelas informações a respeito dos itens lexicais. Sendo assim opta-se, na maioria das vezes, pelas entradas homônimas, que ajudam a evitar a ocorrência de ambiguidade.

## O NOME PRÓPRIO

O estudo do nome próprio assume espaço significativo na Linguística nas duas grandes divisões da ciência Onomástica: a antropônimo – estudo do nome de pessoa – e a toponímia – estudo do nome de lugar.

Ao tratar dos dicionários eletrônicos Ranchhod (2001, p. 24) afirma 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.” A elaboração de dicionários de topônimos atesta a relevância do estudo dos nomes próprios na elaboração de dicionários eletrônicos.

A delimitação entre nome próprio e nome comum não é tarefa simples, como pode parecer à primeira vista. A marcação do nome próprio com inicial maiúscula, apresentada como critério distintivo, nem sempre se mostra eficaz. Vaxelaire (2005) observa esse fato nos dicionários franceses: *Nouveau Petit Robert* (2000) – doravante PR1 – e *Petit Robert des noms propres* (2000) – doravante PR2 – e aponta alguns exemplos, dos quais destacamos: a entrada *magos*, na segunda acepção do PR1 com destaque para *Magos* – personagens bíblicos que (de acordo com a narrativa dos Evangelhos) foram home-

## LÉXICO E SEMÂNTICA

nagear Jesus, e a entrada dessa palavra no PR2, apenas referindo-se a Baltazar, a Gaspar e Melchior. O autor menciona ainda a segunda acepção da entrada de *hexágono* no PR2, que se refere à forma do mapa da França (que pode ser inscrita num hexágono)<sup>23</sup>.

Situação semelhante ocorre com dicionários de língua portuguesa. Pode-se constatar no *Minidicionário Luft* (M.L.) algumas palavras que dão entrada no dicionário com inicial maiúscula, ou que se recomenda o uso de inicial maiúscula, conforme os exemplos a seguir:

**Pás.coa** *s.f.* Comemoração festiva da ressurreição de Cristo.

**Pen.te.cos.tes** *s.m.2n.* (Cat.) Festa em que a Igreja comemora a descida do Espírito Santo sobre os apóstolos; festa do divino.

**na.tal** [...] *S.m.2.* Dia (25 de dezembro) em que se comemora o nascimento de Cristo (inic. maiúsc.).

No *Minidicionário Aurélio* (M.A.) encontra-se:

**páscoa** *s.f.* **1.** Festa anual dos hebreus, que comemora a sua saída do Egito. **2.** Festa anual dos cristãos, que comemora a ressurreição de Cristo.

**pentecostes** *s.m.2n.* Festa católica celebrada 50 dias depois da Páscoa, em comemoração da descida do Espírito Santo sobre os apóstolos.

**na.tal** [...] Dia em que se comemora o nascimento de Cristo (25 de dezembro). [Com *n* maiúsculo, nesta acepç.]

Embora não haja recomendação de se grafar *Páscoa* com inicial maiúscula o M.A. apresenta a palavra registrada dessa forma ao definir *Pentecostes*.

Não fica claro o critério adotado para o uso de inicial maiúscula na definição do nome próprio. Os verbetes apontados anteriormente são de cunho religioso, porém, não parece ter sido esse o parâmetro utilizado por Luft (2001) na utilização de iniciais maiúsculas, visto que no M.L. *iemanjá*, *xangô* e *tupã* não são grafados com inicial minúscula nem há recomendação do uso de maiúsculas. Caso

---

<sup>23</sup> Vexalair, 2005: "Le sens 2 de mage dans le PR1 évoque des personnes particulières "SPÉCIALT. Les Mages: les personnages qui, selon l'Évangile, vinrent rendre hommage à l'Enfant Jésus". À nouveau, le PR2 possède une entrée pour ce mot, même si elle ne fait que renvoyer à Balthazar, Gaspard et Melchior. À l'inverse, si le sens 2 de hexagone renvoie sans hésitation à un nom propre ("L'HEXAGONE: la France métropolitaine2 [à cause de la forme de la carte de France, qu'on peut inscrire dans un hexagone]."), il n'y a pas d'article Hexagone dans le PR2."

fosse considerado o aspecto religioso dos itens lexicais na utilização de iniciais maiúsculas para caracterizar os nomes próprios, as palavras mencionadas seriam grafadas da seguinte forma: *Iemanjá, Xangô e Tupã*, considerando que fazem referência a elementos pertinentes à religião. Os nomes mencionados não apresentam registro no M.A.

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

Observa-se a falta de consenso na questão do sentido do nome próprio (sua existência ou não). De acordo com Brito (2003, p. 39-40),

o debate sobre nomes próprios tem estado bastante dividido [...]. De um lado estariam aqueles que como Kripke e Mill, afirmam que os nomes próprios [...] 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 [...].

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

A distinção de nome próprio e nome comum a partir da abrangência desses – ou seja, afirmar que alcance dos nomes próprios é restrito (marca de identificação de um objeto) e o alcance dos nomes comuns é abrangente – carece de maiores explicações.

Segundo Ullmann, há nomes próprios derivados de substantivos comuns que apresentam transparência semântica e que se observa também o processo em direção inversa, há nomes próprios que se transformam em palavras comuns, neste caso a transparência semân-

## LÉXICO E SEMÂNTICA

tica dependerá do conhecimento que se tem do nome próprio. O autor explica que no primeiro caso “a mudança pode ser acompanhada por uma restrição no seu alcance [...]. No entanto, não há restrição em alguns dos apelidos e nomes de batismo que derivam de substantivos comuns [...]” (Ullmann, 1964, p. 160-164).

### A AMBIGUIDADE DOS NOMES PRÓPRIOS NA ELABORAÇÃO DE UM DICIONÁRIO ELETRÔNICO PARA O PORTUGUÊS DO BRASIL

Embora Ullmann (1964, p. 164) considere “de importância secundária” a ampliação ou restrição quanto aos substantivos próprios e às palavras comuns, respectivamente, a ambiguidade gerada pela ocorrência de nomes próprios derivados de nomes comuns, bem como a mudança de nomes próprios a palavras comuns é assunto de considerada relevância nos estudos de PLN.

De acordo com Brito (2003, p. 26-27),

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 ambiguidades na comunicação [...]. Essa diferença fica clara quando temos de interpretar uma sentença como:

(9a) “João é Tigre”.

Outro fator gerador de ambiguidade é a ocorrência de substantivo próprio para designação de objetos diversos, conforme exemplo apresentado por Brito.

Esses fatores, aliados à possibilidade de ocorrência da polissemia dificultam a diferenciação entre nomes próprios e palavras comuns.

Como já foi dito, as informações contidas nos dicionários eletrônicos diferem das informações observadas comumente nos dicionários de uso. Ranchhod (2001, p. 16) destaca algumas diferenças entre os dicionários de uso e os dicionários eletrônicos e afirma que, “estes dicionários têm que conter informações linguísticas codificadas e formatadas [...]. Não podem conter lacunas”. A autora explica que os dicionários eletrônicos apresentam informações sobre os itens lexicais e as propriedades sintático-semânticas das combinações entre eles. A associação das informações linguísticas concernentes aos itens lexicais de um texto é chamada de *etiquetagem*.

Ao tratar do formalismo e das restrições gramaticais, Laporte (2001, p. 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 - palavra comum - e a realização de *Ele* como nome próprio, acepções da mesma palavra, ou seja, uma ocorrência de polissemia. Nesse caso, as restrições gramaticais e informações das etiquetas lexicais auxiliam na eliminação das ambiguidades. Para resolver a ambiguidade quanto ao gênero do adjetivo nos exemplos propostos, Laporte (2001, p. 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 <e-le.PRO:3ms>) e exige adjetivo no 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 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 <E-

## LÉXICO E SEMÂNTICA

*le.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 linguísticos envolvidos nas pesquisas de PLN, Dias da Silva (2006, p. 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 linguística 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 linguísticos, 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 linguísticos em PLN abriu espaço para que fossem pensadas questões como: a elaboração de um dicionário eletrônico de topônimos – nomes próprios de lugar (Ranchhod, 2001, p. 24) e de antopônimos – nomes próprios de pessoas, o tratamento da homonímia e/ou polissemia na elaboração dos dicionários eletrônicos, a possibilidade de ocorrência de polissemia em nomes próprios etc.

### CONSIDERAÇÕES FINAIS

O desenvolvimento das pesquisas em PLN requer um trabalho conjunto de profissionais de diferentes áreas de conhecimento. Nessa empreitada faz-se necessário o reconhecimento da importância de cada elemento desse conjunto. Para que os dicionários eletrônicos do português do Brasil tenham um bom desempenho em sua aplicação é imprescindível que as descrições e formalizações neles inseridas contenham o menor número de erros possível. Uma descrição equivocada redundará em uma análise textual distorcida, em sistema de tradução automática ou na recuperação de informação não adequada, ou seja, um dicionário eletrônico com descrições inexatas não coopera para o bom funcionamento dos programas informáticos que dele se utilizam.

## *Círculo Fluminense de Estudos Filológicos e Linguísticos*

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

Dias da Silva aponta algumas dificuldades no desenvolvimento das pesquisas em PLN, das quais destacamos

[...] a tímida interação entre os engenheiros e os cientistas da linguagem, alimentada por desencontros e descrédito mútuo [...]

[...] linguistas que chegam a criticar severamente os grupos de pesquisa envolvidos com o PLN. (Dias da Silva, 2006, p. 117, 118)

Observa-se que os entraves das pesquisas em PLN não se limitam aos complexos sistemas teóricos pertencentes aos diversos tipos de conhecimento. A interdisciplinaridade, própria da pesquisa solicita a interação de sujeitos que possuem saberes distintos e de igual relevância. A dificuldade entre tais profissionais muitas vezes é justificada pelo fato de se tratar de pessoas que pertencem a duas áreas de conhecimento: as ciências exatas (engenheiros) e as ciências humanas (linguistas). Contudo, mesmo havendo uma interação adequada entre esses profissionais, os linguistas, por vezes encontram tenaz resistência entre seus pares. Esses fatos apontam para a necessidade de maior colaboração de profissionais das diversas áreas de conhecimento, bem como dos linguistas engajadas em diferentes linhas de pesquisa.

As reflexões apresentadas apontam para o desenvolvimento das pesquisas brasileiras em PLN como um campo vasto, que está além de uma visão simplista e meramente mercadológica, e constitui um grande campo de conhecimentos com solo fértil, pronto para receber sementes de variadas espécies.

As pesquisas em processamento automático do português do Brasil, que inclui a construção de dicionários eletrônicos, evocam questões concernentes à estrutura e ao uso da língua que colaboram para a consolidação das teorias linguísticas em âmbito nacional.

## LÉXICO E SEMÂNTICA

### REFERÊNCIAS

- BORBA, Francisco da Silva. *Pequeno vocabulário de linguística moderna*. 2ª ed. São Paulo: Cia. Ed. Nacional, 1976.
- BRITO, Adriano Naves de. *Nomes próprios: semântica e ontologia*. Brasília: Universidade de Brasília, 2003.
- DIAS DA SILVA, Bento Carlos. O estudo Linguístico-Computacional da Linguagem. *Letras de Hoje*. Porto Alegre, v. 41, nº 2, p. 103-138, junho, 2006.
- FERREIRA, Aurélio Buarque de Holanda. *Minidicionário da língua portuguesa*. 2ª ed. Rio de Janeiro: Nova Fronteira, 1988.
- LAPORTE, Eric. Resolução de ambiguidades. **In:** RANCHHOD, Elisabete Marques. *Tratamento das línguas por computador*. Uma introdução à linguística computacional e suas aplicações. Lisboa: Caminho, 2001.
- LUFT, Celso Pedro. *Minidicionário Luft*. 20ª ed. São Paulo: Ática, 2001.
- LYONS, John. *Linguagem e Linguística*. Uma introdução. Rio de Janeiro: Guanabara Koogan. 1987.
- PERINI, Mario A. *Gramática descritiva do português*. São Paulo: Ática, 1995.
- RANCHHOD, Elisabete Marques. O uso de dicionários e de autômatos finitos na representação lexical das línguas naturais. **In:** —. (Org.). *Tratamento das línguas por computador: uma introdução à Linguística Computacional e suas aplicações*. Lisboa: Caminho. 2001.
- ULLMANN, Stephen. *Semântica: uma introdução a ciência do significado*. 2ª ed. Lisboa: Fundação Calouste Gulbenkian, 1964.
- VAXELAIRE, Jean Louis. *Nom propre et lexicographie française*. *Corela*, número especial, 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.