

Uma proposta de representação das relações de polissemia em *wordnets*

Isa Mara da Rosa Alves¹, Rove Luiza de Oliveira Chishman²

^{1,2}PPG-Linguística Aplicada – Universidade do Vale do Rio dos Sinos (UNISINOS)
Caixa Postal 15.064 – 91.501-970 – São Leopoldo – RS – Brasil
{ialves, rove}@unisin.br

Resumo. *O objetivo deste trabalho é apresentar uma proposta de representação de relações de polissemia para as wordnets, particularmente, para a WordNet.Br.*

Abstract. *This research aims at representing noun polysemy so that it can be useful to wordnet bases, particularly in WordNet.Br.*

1. Introdução

Esta pesquisa desenvolve-se seguindo uma metodologia em três domínios mutuamente complementares [DIAS-DA-SILVA, 1996]: linguístico, linguístico-computacional e computacional. As investigações empreendidas no domínio linguístico-computacional revelaram a importância de introduzir níveis distintos de generalidade entre os sentidos em uma *wordnet*, de modo a reduzir (ou não aumentar) a complexidade do processamento lexical. O domínio linguístico ficou responsável pela reflexão sobre os subsídios teórico-metodológicos que embasam a representação proposta. Por fim, o domínio computacional envolve as atividades de codificação das representações propostas no domínio linguístico-computacional em programas de computador.

2. A polissemia como uma rede de sentidos

A Semântica Lexical Cognitiva (SLC) mostrou ser a teoria mais adequada para os propósitos deste trabalho. Para a SLC, os sentidos polissêmicos devem ser descritos como entidades cognitivas complexas, isto é, como uma categoria de sentidos [LAKOFF, 1987], descrevendo o potencial de sentidos convencionalmente associado a uma estrutura fonológica, representada em forma de rede, uma *rede de polissemia*.

O estudo da polissemia está intimamente ligado ao estudo da homonímia e da monossemia; não há uma fronteira bem delimitada entre esses três fenômenos. Entender a polissemia como um fenômeno categorial implica entender que o fenômeno é caracterizado quando for possível delimitar certas porções de conteúdo conceitual (i) como um agrupamento de sentidos (potencial de sentidos) relacionados (ii) que fazem parte da estrutura semântica (categoria de sentidos) subjacente a uma unidade linguística, que tem *status* simbólico. A negação da parte (i) dessa definição torna os múltiplos sentidos como instância de homonímia por não serem percebidos como um conjunto (uma categoria), fato que pode ter como consequência considerar que estão sendo percebidos dois ou mais itens sem relação (e não um) instanciados na superfície da língua a partir da mesma estrutura fonológica. A negação da parte (ii) leva a entendermos a variação semântica como uma simples especificação contextual da estrutura semântica (categoria de sentidos), caracterizando a monossemia.

3. O modelo *polyset*

As redes de polissemia sincrônicas, em sua configuração livre e multidimensional, tal como Blank (2003) e Geeraerts (2006) propõem, são referência para o modelo *polyset* por possibilitarem um modelo de representação flexível em termos de relações polissêmicas. Há flexibilidade nesse modelo porque “na medida em que o foco desce para sentidos específicos, o item vai sendo visto como mais polissêmico e à medida que o foco passa para os níveis mais altos, mais esquemáticos, o item é visto como menos polissêmico, ou até monossêmico” [TAYLOR, 1995, p. 288].

Cada componente da estrutura do *polyset* contribui de algum modo para a definição dos *nós* (sentidos) e dos *arcos* (relações) que constituem a rede, são eles: (a) o *item lexical polissêmico*, aquele que ativa um determinado tipo de rotina cognitiva, de modo que uma mesma estrutura fonológica provê acesso a diferentes porções relacionadas de conteúdo no espaço conceitual; (b) a *constante*, que representa a *estrutura fonológica/grafêmica* de um item lexical, por esta ser a porção formal comum aos sentidos que compõem o *polyset*; (c) os *sentidos polissêmicos*, caracterizados com base em Croft e Cruse (2004), são os nós da rede e podem ser tanto sentidos plenos quanto os subsentidos (facetas e micros sentidos); (d) os *conjunto de sentidos* que apresentam algum fator de coerência, ou seja, alguma relação semântica; (e) a relação entre os sentidos (polissemia regular ou polissemia irregular, do tipo automeronímia ou autohiponímia); (f) os *efeitos de saliência*. A seguir, serão apresentados detalhes sobre os elementos (e) e (f).

As relações entre os sentidos polissêmicos podem ser de dois tipos: regulares ou irregulares. Entre as relações regulares [CRUSE, 2000], podemos ter a *automeronímia/autoholonímia* (relação de PARTE-TODO) ou a *autohiponímia/autosuperordenação* (relação É-UM). A primeira ocorre quando os subsentidos (do tipo facetas) associam-se por relação de *parte-todo* (ex.: porção/substância, objeto físico/contéudo). Como exemplo, destacamos *livro* ‘objeto físico’ e *livro* ‘contéudo’. O segundo tipo de polissemia ocorre quando um dos subsentidos (do tipo micros sentidos) é construído de modo mais específico em relação aos outros. Como exemplo, destacamos *faca* como ‘utensílio doméstico’, ‘arma branca’ e ‘instrumento cirúrgico’.

As relações de polissemia irregular caracterizam similaridade entre sentidos com alto grau de independência entre si, mas, ainda assim, com alguma similaridade entre eles. Ao contrário das relações que se detectam na polissemia regular, trata-se de uma relação horizontal, não havendo herança entre os sentidos e que apresenta um grau de subjetividade bem maior. O que ocorre é uma comparação entre os sentidos de modo que podemos dizer que há algum grau de similaridade entre eles. Nessas situações, há um grau médio de antagonismo e significativo grau de autonomia entre os sentidos polissêmicos, que se associam de algum modo a uma mesma “noção geral” (objetiva ou subjetiva). Destacamos como exemplo *fonte* como ‘origem, causa, procedência’, ‘nascente de água’ e ‘instalação para aproveitar a água nascente ou mesmo a água encanada’, sentidos que podem ser associados à noção geral *ponto de origem*. O rótulo a ser utilizado para codificar essa é *é-similar-a*.

Sobre os efeitos de saliência de sentido, é preciso dizer que [cf. GEERAERTS, 2006] podemos ter os seguintes cenários típicos: (a) sentido único – quando há um

único sentido saliente que se sobrepõe aos outros, caracterizando a polissemia irregular (ex.: *fonte*, com o sentido ‘origem, causa, procedência’, ‘nascente de água’); (b) sentidos múltiplos e independentes – quando a independência é parcial, caracteriza-se a polissemia irregular ou a polissemia regular do tipo autohiponímia (ex.: *faca*, *cão*); (c) sentidos múltiplos e sobrepostos – quando há a ocorrência de um modo independente, caracterizando a polissemia regular do tipo automeronímia (ex.: *livro*, *universidade*). Neste trabalho, para a identificação do núcleo de uma categoria de sentidos representada em termos de um *polyset*, consideraremos dois critérios fundamentais: a predominância na rede de polissemia e a frequência de ocorrência em *corpus*, conforme descreveremos. O sentido considerado predominante é o sentido que possui maior número de conexões na rede. Ele é, portanto, o sentido com maior *saliência funcional*. O critério é considerado válido por duas razões. Entende-se que ele é o sentido cognitivamente mais saliente e, portanto, o que tem maior vantagem psicológica, por facilitar o acesso aos demais. O sentido mais frequente no *corpus web* é aquele que possui maior *saliência estrutural*.

4. A construção do *polyset*

As etapas linguísticas de construção de um *polyset* são realizadas em dois momentos. O primeiro deles, envolve a definição dos *nós* da rede de polissemia e a indicação dos conjuntos de sentidos. O segundo envolve as etapas de análise e de representação linguística propriamente dita. O item lexical *banco*, que aqui exemplifica o modelo, é frequentemente citado na literatura como exemplo de homonímia por não poder ser percebida qualquer similaridade entre *banco* ‘instituição financeira’ e ‘assento’. Esses não são, no entanto, os únicos sentidos convencionalmente associados ao item, de maneira que diferentes agrupamentos podem ser percebidos em seu conteúdo semântico. O item *banco* é associado a dois *polysets* independentes. Para ilustrar agrupamento polissêmico 1 associado a *banco*, observemos a Figura 1.

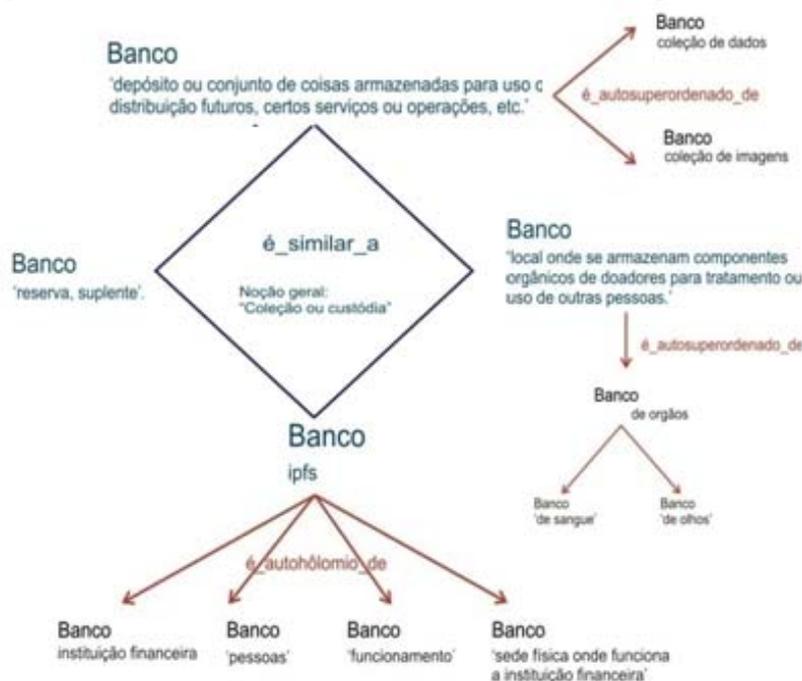


Figura 1 - *Polyset*_{banco}

A Figura 1 representa o primeiro *polyset* associado a *banco*. A rede é estruturada em termos de polissemia regular e irregular. Optamos por organizar o conjunto de sentidos identificados para *banco* associados à noção geral ‘coleção ou custódia’ de algo de valor em diferentes níveis. O primeiro nível relaciona três sentidos por similaridade subjetiva, que tem como referência a função de armazenamento de algo de valor. Um segundo nível de relações, e está aqui o nosso acréscimo, associa os sentidos por polissemia regular, em que o sentido pleno ‘local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas’, serve como conteúdo esquemático para as leituras codificadas como micros sentidos ‘banco de sangue’ e ‘banco de córnea’, relação de *autohiponímia*. Entre os sentidos subordinados à *ipfs* (‘instituição financeira’, ‘pessoas’, ‘funcionamento’ e ‘sede física onde funciona a instituição financeira’), pode ser percebida uma relação de polissemia regular do tipo *automeronímia*. A regularidade é percebida por podermos encontrar outros itens lexicais que apresentam o mesmo tipo de variação entre ‘instituição’/ ‘pessoas’/ ‘funcionamento’/ ‘estrutura física’, entre eles, *escola, creche, universidade*.

A Figura 2 ilustra o *Polyset2* associado ao rótulo *banco*, relacionando os sentidos ‘móvel usado como assento’ e ‘aglomeração de areia, conchas, fósseis e detritos de rochas’ por similaridade.

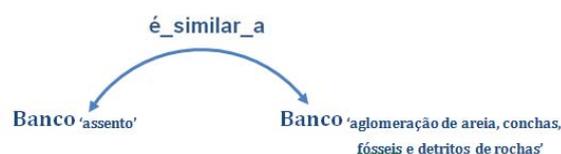


Figura 2 - *Polyset2* banco

A construção do *Polyset2* justifica-se pela identificação de uma similaridade subjetiva relacionada à noção geral *função de apoio* ou da similaridade objetiva baseada na forma de um banco do tipo móvel e um banco de areia ou fósseis e detritos. Há alto grau de autonomia e antagonismo entre esses sentidos, configurando-se como sentidos plenos relacionados por similaridade.

A etapa computacional de construção de um *polyset* prevê sua representação no Protégé-OWL. Para tal, foram utilizadas as recomendações da World Wide Web Consortium (W3C) para a codificação da base WordNet de Princeton em Ontology Web Language (OWL), com o objetivo de criar um esquema representacional comum aos diferentes projetos, facilitando o intercâmbio de dados.

7. Conclusão

Procuramos mostrar aqui que uma abordagem flexível como a de redes de polissemia é uma alternativa à estruturação de bases de dados lexicais por permitir a representação dos sentidos polissêmicos em seus diferentes níveis de generalidade. Para entender o desafio envolvido, basta olharmos para a estrutura global de tal base: os sentidos são agrupados em conjuntos de sinônimos cognitivos (*synsets*) e relações são estabelecidas entre os conceitos expressos linguisticamente pelos itens que compõem o *synset*. O que se observa é que a grande dificuldade na representação da polissemia em tais ontologias, ainda sem envolver a sua definição, é o fato de o fenômeno exigir um olhar para o léxico sob a perspectiva semasiológica e as *wordnets*, por definição, serem representações onomasiológicas.

Referências

- Blank, A. (2003) “Polysemy in the lexicon and discourse. In: NERLICH, B. et al (Ed.). Polysemy: flexible patterns of meaning in mind and language. Hawthorne, N.Y.: Mouton de Gruyter, p. 267-293.
- Croft, W. e Cruse, D. A. (2004) “Cognitive linguistics”. Cambridge: Cambridge University Press.
- Cruse, D.A. (2000) “Meaning in language: an introduction to semantics and pragmatics”. Cambridge: Cambridge University Press.
- Dias-da-Silva, B. C. (1996). “A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais”. 272 f. Tese (Doutorado em Linguística e Língua Portuguesa)—Faculdade de Ciências e Letras, UNESP/Araraquara.
- Fellbaum, C. (ed.). (1998) “WordNet: an electronic lexical database”. Cambridge: The MIT Press.
- Geeraerts, D. (2006) “Words and other wonders: papers on lexical and semantic topics”. Berlin: Mouton de Gruyter.
- Lakoff, G. (1987) “Women, fire, and dangerous things: what categories reveal about the mind”. Chicago: University of Chicago Press.
- Taylor, J. R. (1995) “Linguistic categorization: prototypes in linguistic theory”. London: Clarendon Press; New York: Oxford University Press.