

ISA MARA DA ROSA ALVES

***Polyset: Modelo Linguístico-Computacional para a Estruturação de
Redes de Polissemia de Nominais***

Araraquara

2009

ISA MARA DA ROSA ALVES

***Polyset: Modelo Linguístico-Computacional para a Estruturação de
Redes de Polissemia de Nominais***

Tese apresentada à Faculdade de Ciências e Letras, Universidade Estadual Paulista – Campus de Araraquara, como parte dos requisitos para a obtenção do título de Doutor em Letras (Área de Concentração: Linguística e Língua Portuguesa).

Linhas de pesquisa: Estudos do léxico; Análise fonológica, morfológica, semântica e pragmática.

Orientador: *Prof. Dr. Bento Carlos Dias da Silva*

Co-orientadora: *Profa. Dra. Rove Luiza de Oliveira Chishman*

Araraquara

2009

A474p

Alves, Isa Mara da Rosa.

Polyset: modelo linguístico-computacional para a estruturação de redes de polissemia de nominais / Isa Mara da Rosa Alves. – 2009.

212 f. : il. ; 30 cm.

Tese (doutorado) – Universidade Estadual Paulista, Faculdade de Ciências e Letras de Araraquara, 2009.

“Orientador: Prof. Dr. Bento Carlos Dias da Silva ; co-orientadora: Profa. Dra. Rove Luiza de Oliveria Chishman”.

1. Polissemia. 2. Linguística. 3. Semântica. 4. WordNet. 5. Linguística – Processamento de dados. 6. Linguística cognitiva. 7. Processamento de linguagem natural (Computação). I. Título. II. As redes de polissemia e as bases relacionais de dados lexicais : proposta de representação linguístico-computacional de nominais com sentidos múltiplos.

CDD-410.285
CDU-81'822

ISA MARA DA ROSA ALVES

***Polyset: Modelo Linguístico-Computacional para a Estruturação de
Redes de Polissemia de Nominais***

COMISSÃO JULGADORA

TESE PARA OBTENÇÃO DO GRAU DE DOUTOR

Presidente e Orientador: **Prof. Dr. Bento Carlos Dias da Silva**

1º Examinador: Profa. Dra. Beatriz Nunes de Oliveira Longo

2º Examinador: Profa. Dra. Gladis Maria de Barcellos Almeida

3º Examinador: Prof. Dr. Thiago A. S. Pardo

4º Examinador: Prof. Dr. Heronides M. M. Moura

Araraquara, 01 de Outubro de 2009.

Ao Daniel
e a meus pais, Cícero e Isabel.

AGRADECIMENTOS

Mais do que nunca, tenho muito a agradecer pelos vários anjos que me acompanharam ao longo dos últimos 4 anos. Para mencionar apenas algumas dessas pessoas especiais, agradeço:

Aos meus avós e a minhas queridas irmãs, por estarem de diferentes maneiras ao meu lado sempre;

Aos meus amigos e colegas do NILC, da UNESP e da UNISINOS;

As minhas queridas amigas Ana Carolina Sperança, Cristina Damim e Silvana Kissmann;

Ao meu orientador, Prof. Dr. Bento Dias da Silva, pelos ensinamentos preciosos e pela compreensão;

A minha co-orientadora, Profa. Dra. Rove Luiza de Oliveira Chishman, por seguir ao meu lado em mais essa caminhada sempre com uma postura exigente, sincera e amiga.

“That a great deal to make one word mean, Alice Said in a thoughtful tone.

When I make a word do a lot of work like that, said Humpty Dumpty, I always pay it extra.”

(Carroll, Lewis. *Alice in Wonderland*, p. 225)

RESUMO

Esta pesquisa visa a propor uma representação da polissemia de nominais compatível com sistemas computacionais; mais especificamente, o objetivo deste trabalho é incluir a especificação de relações de polissemia em bases *wordnets*, em especial, na WordNet.Br (DIAS-DA-SILVA, 1996, 1998, 2003). A metodologia do trabalho está baseada em Dias-da-Silva (1996, 1998, 2003, 2006), abrangendo três domínios mutuamente complementares: o linguístico, o linguístico-computacional e o computacional. O domínio linguístico-computacional forneceu o tema para esta pesquisa e articulou a relação entre os domínios linguístico e computacional. Das investigações realizadas no cenário linguístico-computacional, destacamos a relevância da introdução de níveis distintos de generalidade entre os sentidos em uma base de dados de modo a otimizar o processamento lexical a ser realizada pelo sistema. Percebe-se que esse tipo de tarefa é ainda um desafio para as *wordnets*. Do cenário linguístico, destacamos que a Semântica Lexical Cognitiva foi considerada a teoria mais adequada aos propósitos desta tese. Olhar para o fenômeno do significado múltiplo sob o viés cognitivo possibilitou descrever os sentidos como uma entidade complexa, estruturada em termos de redes. As redes de polissemia sincrônicas, em sua configuração livre e multidimensional, conforme propõem Blank (2003) e Geeraerts (2006), demonstraram ser a estratégia descritiva mais adequada à representação da flexibilidade do sentido para os propósitos desta tese. Respondendo à fase aplicada dos domínios linguístico e linguístico-computacional, propomos um modelo de representação denominado *polyset*. Os *polysets* são constructos estruturados em termos de redes de polissemia, de maneira que possibilitam representar diferentes níveis de generalidade entre os sentidos, diferentes graus de saliência e diferentes tipos de relações de polissemia. Para dar conta do domínio computacional, propomos, seguindo a metodologia proposta por Ganguemi (2004), uma estratégia para a codificação dos *polysets* em OWL com auxílio do editor de ontologias Protégé-OWL. O modelo *polyset* representa uma tentativa de contribuir para a representação da polissemia de nominais no paradigma *wordnet*, em especial, para a WN.Br.

Palavras-chave: polissemia; redes de polissemia; relação de similaridade; automeronímia; autohiponímia; relações de sentido.

ABSTRACT

This research aims at representing noun polysemy so that it can be useful to computational systems; more specifically, the subject of this work is the inclusion of specification of polysemy relations in *wordnet* bases, particularly in WordNet.Br (DIAS-DA-SILVA, 1996, 1998, 2003). The methodology has been based on Dias-da-Silva (1996, 1998, 2003, 2006), comprehending three mutually complementary domains: linguistic, computational-linguistic, and computational ones. The computational-linguistic domain has both provided the subject for this research and articulated the relationship between the linguistic domain and the computational domain. From the investigations carried out in the computational-linguistic scene, we have highlighted the relevance of the introduction of distinct levels of generality among meanings in a database, so as to reduce the amount of lexical processing to be carried out by the system. At the same time, that multiple representation provides the necessary information for a system that needs a higher degree of meaning detailing. This kind of task is still a challenge to *wordnets*. From the linguistic scene, we have highlighted that Cognitive Lexical Semantics has shown to be the most suitable theory for the purposes of this thesis. Regarding the phenomenon of the multiple meaning from the cognitive perspective has allowed for describing meanings as a complex entity, structured in terms of nets. The nets of synchronic polysemy, in their free, multidimensional configuration, as Blank (2003) and Geeraerts (2006) have proposed, have shown to be the most suitable descriptive strategy for the representation of the meaning flexibility for the purposes of this thesis. Answering to the applied phase of both the linguistic and computational-linguistic domains, we have proposed a representation model called *polyset*. *Polysets* are constructs structured in terms of polysemy nets, allowing for representing different levels of generality among meanings, different degrees of salience, and different types of polysemy relationships. Following the methodology pointed out by Ganguemi (2004), in order to consider the computational domain, we have proposed a strategy for the codification of *polysets* in OWL with the help of the Protégé-OWL ontology editor. The *polyset* model represents an attempt at contributing towards the representation of noun polysemy in the *wordnet* paradigm, especially WN.Br.

Keywords: polysemy; polysemy network; similarity relation; autonymy; autohyponymy; sense relation.

LISTA DE FIGURAS

Figura 1 – O <i>synset book</i> e informações associadas.....	19
Figura 2 - Os 25 unique beginners da hierarquia dos synsets para os nomes da WN.Pr	20
Figura 3 - <i>Synsets</i> que contêm o item lexical <i>bank</i> na WN.Pr.2.0	23
Figura 4 – Exemplo de itens lexicais primos.....	25
Figura 5 - Exemplo sentidos irmãos.....	26
Figura 6 - Consulta WN.2.0: <i>sole</i>	27
Figura 7 - WN.Pt: teste de relação léxico-semântica independente.....	29
Figura 8 - Polissemia Regular na GN – <i>bank</i>	31
Figura 9 - Relação <i>forma-conteúdo</i>	58
Figura 10 - Configurações de redes	83
Figura 11 - Ilustração do modelo de rede radial	86
Figura 12 – Modelo de rede esquemática	88
Figura 13 – Rede esquemática complexa	89
Figura 14 - Rede esquemática <i>uncle</i> e <i>ear</i>	90
Figura 15 – Rede sincrônica de polissemia	93
Figura 16 - Figura e fundo.....	102
Figura 17 - Autohiponímia <i>cachorro</i>	132
Figura 18 – Autossuperordenação <i>homem</i>	133
Figura 19 - Representação gráfica dos automerônimos de <i>livro</i>	156
Figura 20 - <i>Polyset</i> fonte.....	163
Figura 21 - <i>Polyset1</i> banco.....	167
Figura 22 - <i>Polyset2</i> banco.....	168
Figura 23 - <i>Polyset</i> Café	171
Figura 24 - Os componentes do <i>polyset</i> em OWL.....	174
Figura 25 - Classes genéricas de uma <i>wordnet</i> em OWL.....	175
Figura 26 - Classes genéricas do modelo <i>polyset</i>	175
Figura 27 - Os nós como classes em estrutura hierárquica	176
Figura 28 - A relação entre estrutura fonológica e semântica como uma <i>object property</i> do tipo <i>hasSense</i>	177

Figura 29 - Conjunto de sentidos polissêmicos como <i>object property</i> do tipo <i>inSynset</i>	178
Figura 30 - Relações como <i>object properties</i>	179
Figura 31 - Relação de autohiponímia/autosuperordenação	181
Figura 32 - Relação de automeronímia/autoholonímia.....	182
Figura 33 - Graus de saliência.....	183
Figura 34 - Visão geral dos <i>polysets</i> codificados no Protégé-OWL.....	184
Figura 35 - <i>wordnet</i> e <i>polysets</i> em OLW no Protégé-OWL	185

LISTA DE QUADROS

Quadro 1- Relações do tipo <i>nome-nome</i> da WN-Pr	21
Quadro 2 - CoreLexII - Classes de polissemia regular.....	34
Quadro 3 - A polissemia e a descrição da semântica lexical	67
Quadro 4 – Inventário de tipos de <i>automeronímia</i>	131
Quadro 5 – Síntese da seção 4.	137
Quadro 6 - Evidências de coerência e de autonomia	150
Quadro 7 – Propriedades centrais da relação de polissemia.	154
Quadro 8 – Níveis gerais de estruturação de um <i>polyset</i>	155
Quadro 9 - Resultado das etapas preliminares de construção do <i>polyset</i>	159
Quadro 10 - Resultado das etapas de construção do <i>polyset</i>	160
Quadro 11 – Resultado da etapa preliminar de construção do <i>polyset</i> _{fonte}	162
Quadro 12 – Resultados das etapas preliminares de construção do <i>polyset</i> _{banco}	166
Quadro 13 – Resultado das etapas preliminares de construção do <i>polyset</i> _{café} ...	170
Quadro 14 – Domínio e alvo nas relações de sentido.....	180

SUMÁRIO

Seção 1 - Introdução	1
1.1 Delimitação do tema e contextualização	1
1.2 Objetivos e estratégias metodológicas globais	6
1.3 Organização do trabalho.....	8
Seção 2 - A polissemia como um desafio à estruturação de bases relacionais de dados lexicais do tipo <i>wordnet</i>	10
2.1 A interação entre polissemia, processamento automático de língua natural e wordnets	10
2.2 Caracterização de uma base relacional de dados lexicais do tipo <i>wordnet</i>	18
2.3 A proposta da WordNet.Pr.....	22
2.4 A proposta da WordNet.Pt.....	28
2.5 A proposta da GermaNet	30
2.6 A proposta do CoreLex.....	32
2.7 Síntese da seção	35
Seção 3 - A representação da polissemia sob a ótica da Semântica Lexical Cognitiva	37
3.1 Que desafios a polissemia apresenta a uma teoria linguística?.....	38
3.2 Por que escolher o olhar da Semântica Lexical Cognitiva?	42
3.2.1 Visão Panorâmica	43
3.2.2 Contribuições da Semântica Lexical Cognitiva para o Estudo dos Sentidos.....	52
3.3 A caracterização da polissemia como um fenômeno categorial.....	56
3.3.1 A relação entre forma e conteúdo	57
3.3.2 A polissemia como categorização.....	59
3.4 Aspectos estruturais e funcionais da dimensão semasiológica do léxico.....	65
3.5 As abordagens homonímica, polissêmica e monossêmica.....	68
3.5.1 Caracterização e problemas implicados	69
3.5.2 Uma estratégia alternativa	76
3.6 As redes de polissemia	83
3.7 Síntese da seção	95
Seção 4 - Os sentidos polissêmicos e suas relações.....	97
4.1 A delimitação dos sentidos.....	97
4.2 Efeitos de saliência entre os sentidos.....	101
4.2.1 Saliência de Perspectiva (<i>Realce</i>)	103
4.2.2 Saliência Semasiológica Paradigmática (<i>Prototypicalidade</i>)	104
4.2.3 Saliência estrutural (<i>distintividade</i>).....	104
4.2.4 Saliência Psicológica (<i>Funcional</i>).....	104
4.3 A similaridade e a polissemia	106
4.3.1 Caracterização da Similaridade.....	107
4.3.2 A Existência ou Negação de Similaridade entre os Sentidos.....	110
4.3.2.1 Grau insignificante de similaridade entre sentidos plenos.....	112
4.3.2.2 Grau relevante de similaridade entre sentidos plenos.....	116
4.3.2.3 A similaridade entre facetas de sentidos	120
4.3.2.4 A similaridade entre os microssentidos.....	125

4.3.3 Tipos de similaridade: relações de polissemia.....	128
4.3.3.1 Relação de Polissemia Regular.....	129
4.3.3.2 Relação de Polissemia Irregular.....	134
4.4 Quadro síntese da seção	136

Seção 5 - delineamento linguístico do modelo de representação: o constructo

<i>polyset</i>.....	138
5.1 O <i>polyset</i> como um constructo de <i>status</i> múltiplo	140
5.2 Os componentes e a estrutura do <i>polyset</i>	142
5.2.1 Os <i>nós</i> e a caracterização dos sentidos polissêmicos.....	143
5.2.1.1 O item lexical.....	143
5.2.1.2 A constante	144
5.2.1.3 Os sentidos.....	144
5.2.1.4 Um conjunto de sentidos.....	148
5.2.1.5 A construção dos <i>polysets</i> : a delimitação dos tipos sentidos	149
5.2.1.6 Efeitos de saliência entre os sentidos.....	150
5.2.2 Os arcos da rede de sentidos polissêmicos.....	151
5.2.2.1 A relação de polissemia regular.....	152
5.2.2.2 A relação de polissemia irregular	153
5.2.3 Síntese da seção.....	154

Seção 6 - Construção e implementação do *polyset*.....

6.1 a face linguística de construção do <i>polyset</i>.....	158
6.1.1 Análise do item lexical fonte.....	161
6.1.2 Análise do item lexical <i>banco</i>	164
6.1.3 Café	169
6.2 A face computacional de construção do <i>polyset</i>	172
6.2.1 A codificação de um <i>polyset</i> em OWL através do Protégé-OWL	173
6.3 Síntese das contribuições do modelo <i>polyset</i> para a wordnet.br.....	184
6.4 Síntese da seção	187

Seção 7 - Conclusão

Referências bibliográficas.....

193

SEÇÃO 1 - INTRODUÇÃO

1.1 DELIMITAÇÃO DO TEMA E CONTEXTUALIZAÇÃO

Precisaremos esperar por uma *posição* dos diretores, mas, enquanto esperamos, podemos ficar em posição confortável.

Acabo de saber que mudarei de posição na empresa, pois fui promovido!

Felipe Massa chegou em primeira posição.

Ainda não escolhi em que banco vou abrir uma conta.

O banco está precisando de sangue do tipo AB.

Não importa se você tem um *cão* Shipdog ou uma cadela Rotweiller, todo *cão* precisa ser vacinado periodicamente.

Vamos combinar o seguinte: o José vai escolhendo a melhor fonte para o panfleto, enquanto a Maria descobre a fonte daquelas informações.

Criamos os enunciados acima para exemplificar o fenômeno linguístico em foco nesta tese: a *polissemia* de nominais. Não há como definir polissemia¹ sem um comprometimento teórico, mas, em sua forma mais simples, ela pode ser entendida, sincronicamente, como a propriedade que certos itens lexicais têm de poder ser empregados em mais de um contexto instanciando sentidos² distintos, mas relacionados. É o que pretendemos ilustrar com os usos de *posição*, *banco*, *cão* e *fonte* na abertura desta seção. Embora não tenha sido fornecido mais do que um contexto sentencial para cada um desses itens polissêmicos, por representarem usos comuns em português brasileiro (PB), é altamente provável que qualquer falante do PB compreenda facilmente cada uma das situações, delimitando adequadamente os sentidos em cada contexto. Além disso, é provável que os falantes percebam a existência de alguma similaridade entre os sentidos. Nós, humanos, executamos tarefas como essas

¹ Quando usarmos o termo *polissemia* pretendemos denotar o fenômeno como um todo, tanto polissemia regular quanto irregular.

² Os termos sentido e significado são empregados nesta tese de maneira intercambiável, como variações estilísticas, assim como fazem, por exemplo, Taylor (1995; 2003), Tyler e Evans (2003) e Soares Da Silva (2006).

intuitivamente e com muito sucesso. Entretanto, assim como ocorre com a maior parte das atividades, a facilidade que humanos têm para armazenar, acessar e processar múltiplos sentidos associados a um mesmo item é inversamente proporcional à dificuldade que os sistemas computacionais enfrentam. Armazenar as informações semânticas de itens lexicais polissêmicos de maneira suficientemente precisa, estruturada e manipulável para que sistemas computacionais realizem inferências sobre tais conteúdos é um grande desafio.

A situação descrita acima ilustra o contexto desta pesquisa, que se propõe a representar a polissemia de nominais de maneira compatível com sistemas computacionais; mais especificamente, o tema deste trabalho é a inclusão da especificação de relações de polissemia em bases *wordnets*, em especial, na WordNet.Br (DIAS-DA-SILVA 1996, 1998, 2003). Em linhas gerais, as *wordnets* são repositórios de conteúdo semântico estruturados em termos de relações paradigmáticas. A *wordnet* mãe, a WordNet de Princeton (WN.Pr) (MILLER e FELLBAUM, 2007), foi concebida originalmente como um modelo cognitivo geral da organização semântica humana, mas logo se tornou recurso lexical de referência para a comunidade de Processamento Automático de língua Natural, doravante PLN³ (BOYD-GRABER *et. al*, 2006), de maneira que os esforços que são atualmente empreendidos na ampliação das *wordnets* têm o fim de melhor servir às necessidades computacionais. Nas palavras de seus idealizadores, a WN.Pr “objetiva servir à comunidade e continuar a ser de interesse de colegas que servem tanto a perspectivas simbólicas quanto estatísticas de língua” (MILLER e FELLBAUM, 2007, p. 213).

Incluir em uma *wordnet* relações entre sentidos polissêmicos exige, de início, desafiar os princípios globais de sua estrutura, e o motivo é evidente: as *wordnets*, tal como originalmente concebidas, são organizadas seguindo a lógica onomasiológica de estruturação do conteúdo lexical, e a representação da polissemia exige um olhar sob a perspectiva semasiológica. Esse fato fica evidente ao percebermos que o paradigma *wordnet* prevê uma estrutura em termos de *synset*, constructo que representa o sentido de um conjunto de itens lexicais que codificam um mesmo conceito. As relações típicas de uma *wordnet*, portanto, não são relações entre itens lexicais, mas entre os conceitos expressos linguisticamente pelos itens que compõem o *synset*. Conforme observaremos

³ A área conhecida como Processamento Automático de Língua Natural, em linhas gerais, tem como objetivo “capturar” em estruturas formais o conteúdo linguístico.

ao longo do trabalho, no entanto, essas diferentes visões não são incompatíveis, pelo contrário, são complementares.

Estamos cientes de que esse tipo de estudo é altamente custoso e que exige julgamentos semânticos muito delicados, sutis, e de que há uma porção de subjetividade na tarefa, fato que, na medida do possível, precisa ser minimizado para fins aplicados ao PLN. Todavia, é preciso notar que todo esse esforço é justificado quando pensamos nos benefícios da inclusão desse tipo de informação em uma base de dados lexicais que se propõe a ser útil também a sistemas que realizam tarefas igualmente minuciosas, como tradução automática e a interação homem-máquina por meio de sistemas de pergunta e resposta, por exemplo. Pesquisas que realizamos em 2002 (ALVES, 2002) e 2005 (ALVES, 2005) demonstram em mais detalhes o tipo de necessidade de tais sistemas e deixam-nos certos dessa posição.

O fato de termos como objetivo resolver uma tarefa específica, a inclusão da representação da polissemia na WordNet.Br, revela a nossa ideologia de pesquisa, que é do tipo “demand-side” (NIRENBURG e RASKIN, 1996, p. 26) ou “task-driven” (NIRENBURG e RASKIN, 2004, p. 14)⁴. De acordo com Nirenburg & Raskin (1996), quem trabalha nessa perspectiva opta por, antes de adotar qualquer teoria, certificar-se de que há uma função específica para sua aplicação na descrição dos fatos linguísticos a serem implementados no sistema de PLN em foco. Enfatizam-se a adequação e a eficiência descritiva da teoria escolhida, mais que sua potencial capacidade de prever generalizações. Essa é a justificativa para situarmos nossas investigações no contexto teórico geral da Semântica Lexical. Tendo em vista o fim aplicado deste trabalho, fazemos interface com as áreas da Semântica Lexical Computacional e do PLN. A Semântica Lexical, em linhas gerais, concentra-se em delimitar e relacionar sentidos associados a itens lexicais. A Semântica Lexical Computacional, por sua vez, descreve os sentidos individualmente, bem como seus múltiplos relacionamentos, porém o seu foco é a utilidade de suas descrições para os sistemas computacionais que se dedicam a processar automaticamente a língua natural. A Semântica Lexical Computacional liga-se diretamente ao PLN, que é uma subárea da Inteligência Artificial voltada à construção de sistemas diversos que realizam interação homem-máquina através da língua humana

⁴ Essa postura é oposta a uma perspectiva “supply side” (NIRENBURG e RASKIN, 1996, p. 26) ou “method-driven” (NIRENBURG e RASKIN, 2004, p. 14), que tem como foco avaliar o poder descritivo de uma teoria específica, com a intenção de captar o máximo de generalizações a partir da aplicação das regras propostas por essa teoria.

(ex.: tradutores automáticos, corretores de textos, sumarizadores automáticos, sistemas de busca, sistemas de extração de informações, sistemas de pergunta e resposta, etc.), conforme discute Dias-da-Silva (1996) em sua tese de doutorado.

Dizer que o contexto teórico deste trabalho é a Semântica Lexical significa dizer que estamos nos filiando a uma comunidade altamente heterogênea. Conforme reportam Nirenburg e Raskin (1996), trata-se de um grupo que inclui funcionalistas, formalistas, cognitivistas e estruturalistas, além de uma grande comunidade de PLN, que inclui tanto linguistas quanto cientistas da computação, com diferentes objetivos e foco: alguns têm uma forte queda para as probabilidades estatísticas; outros dedicam-se ao uso de grandes *córpus* e de dicionários legíveis por máquina, bases de dados lexicais, ontologias; uns participam ativamente no desenvolvimento de sistemas PLN; há, ainda, outros que trabalham em projetos para os sistemas futuros. Como as comunidades tendem a fazer, continuam os autores, semanticistas debatem sobre os mesmos exemplos e temas, ficam empolgados com promessas de novos métodos e seguem as eternas mudanças de tendência científicas e sucessões de descobertas. Todavia, os objetivos e as premissas são bastante distintos; há diferentes visões de teoria, metodologia e aplicações; são variados também os limites, as restrições, assim como as concepções sobre o que é e o que não é viável; são seguidos diferentes métodos de raciocínio; e, ainda, chega-se a resultados muito distintos. É também justo dizer, acrescentam Nirenburg e Raskin, que todos os semanticistas são capazes de reconhecer uma entrada lexical adequadamente especificada em diferentes formas e paradigmas, mas, ao mesmo tempo, todos são unânimes ao afirmar que há relativamente poucas entradas realmente adequadas. A terminologia e as diferenças conceituais percebidas entre os semanticistas podem ser fruto dessas diferenças, que nem sempre são percebidas, afirmam.

De maneira leve e ilustrativa, Nirenburg e Raskin (1996) enfatizam a natureza múltipla da Semântica Lexical, e, a partir deles, podemos entender a origem de parte das discrepâncias constatadas no estudo deste fenômeno que é o tema central da teoria: a multiplicidade de sentidos, que abrange a polissemia e fenômenos correlatos, nomeadamente, a *homonímia* (quando, entre os múltiplos sentidos associados a um item, não é possível percebermos qualquer relação; por exemplo, os sentidos 'instituição financeira' e 'assento' associados a *banco*) e a *monosseμία* (quando as distinções entre as possibilidades de uso de um item são consideradas especificações contextuais de um

único sentido; por exemplo, o único sentido que é inespecífico em relação ao sexo e que é associado a *criança*). Em razão da dimensão das diferenças de princípios e enfoques que compõem a Semântica Lexical, do ponto de vista linguístico, não basta dizer que essa é a perspectiva teórica de base de nossa investigação – em seguida, seremos mais específicos.

A compreensão da polissemia e a proposição de uma representação computacionalmente tratável para esse fenômeno ocupam posição de destaque na Semântica Lexical Computacional. Ninguém discute a importância de uma base de dados lexicais representar as diferentes nuances de sentido decorrentes da polissemia, mas essa é uma das poucas unanimidades relacionadas ao tema. Quando passamos a refletir sobre a melhor estratégia para a descrição desse tipo de informação, seja no âmbito da Linguística, seja no âmbito do PLN, começamos a perceber as discordâncias entre os pesquisadores.

A *polissemia*, termo de origem grega que deriva da combinação de *πολυξ* ('muito') e *σημειον* ('sentido'/'significado'/'significação') (cf. BRÉAL, 1992, p. 102 [originalmente, 1897]), recebeu seus primeiros tratamentos na linguística em período pré-estruturalista. Ao longo do desenvolvimento dos estudos semânticos, em alguns momentos, esteve sob olhares atentos (ex. Bréal), em outros, foi considerada intratável (ex. Chomsky), até que o tema voltou ao primeiro plano sob dois enfoques: linguístico e computacional. Em Soares da Silva (2006), encontramos o termo "(re)polissemização" para denotar esse movimento em duas vias que se deu no estudo da polissemia. No âmbito da linguística teórica, a redescoberta da polissemia ficou por conta da Semântica Lexical Cognitiva (ex. Lakoff, Langacker, Taylor, Geeraerts, Cruse), que deu ao fenômeno a posição de destaque a partir do final dos anos 80. Em abordagens propostas sob essa perspectiva, embora a maior parte dedique-se às irregularidades da polissemia, é possível encontrar reflexões sobre o fenômeno da multiplicidade de sentidos como um todo (*homonímia*, *polissemia irregular*, *polissemia regular* e *monossemia*). Nessa perspectiva de estudo da semântica lexical a partir de uma visão não-modular de cognição, segundo a qual o conteúdo linguístico está diretamente ligado à experiência corpórea, não reconhecendo distinção estrita entre conhecimento linguístico e extralinguístico, a multiplicidade de sentidos é vista como uma riqueza da língua que revela uma das propriedades fundamentais dos sentidos: a flexibilidade e a sua sensibilidade ao contexto. No âmbito do PLN, o interesse pela multiplicidade de

sentidos se deve às necessidades, conforme observamos anteriormente, apresentadas por sistemas de PLN, que, de maneira mais precisa ou mais geral, necessitam realizar atividades de desambiguação. A partir desse cenário, passaram a ser frequentes os estudos linguísticos que se dedicam a aplicações computacionais e estudos computacionais que aprofundam pesquisas linguísticas (ex. RAVIN & LEACOCK, 2000).

Nesse contexto interdisciplinar, como não poderia deixar de ser, a proposição de uma representação da polissemia em uma base de dados relacionais do tipo *wordnet* exige a adoção de uma metodologia, como a proposta por Dias-da-Silva (1996, 1998; 2003; 2006), que abrange uma investigação em três domínios mutuamente complementares: o linguístico, o linguístico-computacional e o computacional, conforme descreveremos a seguir.

1.2 OBJETIVOS E ESTRATÉGIAS METODOLÓGICAS GLOBAIS

Seguindo a ideologia de pesquisa *task-driven*, o **objetivo geral** desta tese é propor uma estratégia de representação linguístico-computacional da semântica de nominais polissêmicos que complemente as informações léxico-conceituais codificadas em *wordnets*, com vistas a ampliar a utilidade dessas bases à ampliação da capacidade de os sistemas de PLN realizarem inferências sobre eles. As bases de dados lexicais do tipo *wordnet* funcionam, neste trabalho, como elo entre o estudo de questões linguísticas e linguístico-computacionais. Sendo assim, as reflexões responsáveis pelo equacionamento do **domínio linguístico-computacional**, previsto em Dias-da-Silva (1996, 1998; 2003; 2006), ocupam aqui a posição de “pivô”, por serem responsáveis por duas funções: (i) a de motivar a escolha de um fenômeno linguístico relevante, delimitando o tipo de necessidade representacional (entendendo-se aqui *representação* como descrição formal ou implementável de informações linguísticas) e (ii) a de estabelecer o diálogo com o domínio computacional através de uma representação linguístico-computacional. Os **objetivos específicos do trabalho relacionados a essa etapa** dividem-se, portanto, em duas fases. Na primeira, precisa-se (a) entender em que medida a polissemia constitui um problema para PLN e (b) investigar como a questão tem sido enfrentada pelas bases *wordnets*, a fim de avaliar que desafios fazem parte da

codificação da multiplicidade de sentidos. Na segunda, levando-se em conta as necessidades e exigências delimitadas na fase anterior, objetiva-se propor uma representação linguístico-computacional aplicável ao contexto da *wordnet* brasileira (WordNet.Br). Esse objetivo linguístico-computacional será somente alcançado quando forem finalizadas as tarefas do domínio linguístico, revelando a mútua complementaridade entre os domínios prevista por Dias-da-Silva (1996, 1998; 2003; 2006).

Delineada a demanda de bases do tipo *wordnet* em relação à inclusão da polissemia de nominais, dedicaremos-nos aos estudos do **domínio linguístico**. Será neste domínio que refletiremos sobre os subsídios teórico-metodológicos que embasarão a representação proposta, que deve ter completude descritiva e representacional, sem perder de vista o rigor teórico necessário a uma representação linguístico-computacional cientificamente válida. Sabe-se que o modo como é entendido cada um dos diferentes graus de multiplicidade de sentidos é determinante para sua caracterização precisa e para a proposição de estratégias representacionais. Os **objetivos específicos do trabalho relacionados a esses estudos** desenvolvem-se em duas fases. Na primeira, o foco é demonstrar em que medida a Semântica Lexical Cognitiva pode ser considerada uma teoria (a) suficientemente abrangente e com (b) poder descritivo adequado à representação da polissemia de nominais de maneira útil aos fins computacionais. Na segunda, o foco é descrever, de acordo com o ferramental teórico fornecido pela teoria delimitada, ocorrências de polissemia em nominais do português. A descrição proposta será apresentada já nos moldes mencionados no **domínio linguístico-computacional**, de maneira que ela terá *status* múltiplo: o de representação linguística, pois caracteriza um fenômeno que se manifesta na superfície da língua; o de representação linguístico-computacional, pois configura-se como uma descrição formalizável das informações linguísticas; e, ainda, o de representação linguístico-cognitiva, pois tem o *status* de “isomorfismo funcional”, entendendo-se que a realidade psicológica de uma representação não é imediata, mas pode ser representativa do que temos na mente (SANDRA e RICE, 1995, p. 102).

Por fim, o **domínio computacional** envolverá as atividades de codificação das representações propostas no domínio linguístico-computacional em programas de computador. O **objetivo específico do trabalho relacionado a essas atividades** é a codificação das representações no editor Protégé-OWL, de maneira a possibilitar que os

dados sejam convertidos para as linguagens de programação diretamente aproveitadas por sistemas de PLN. Para a inclusão dos dados no sistema, adaptaremos as recomendações da World Wide Web Consórcio (W3C)⁵ para a codificação da base *wordnet* às necessidades do fenômeno descrito. O *status* que pretendemos dar à representação proposta no domínio computacional é o que Nirenburg e Raskin (2004, p. 14) denominam “tese fraca de IA”. Uma representação de conhecimento entendida como tese fraca de Inteligência Artificial nega que um sistema computacional modele diretamente o conhecimento humano sobre semântica. Ao invés disso, esta tese sugere haver equivalência funcional, isto é, sugere que os programas computacionais podem alcançar resultados com qualidade análoga à alcançada por humanos, apesar de não usarem exatamente os mesmos métodos.

1.3 ORGANIZAÇÃO DO TRABALHO

Levando em consideração os objetivos e os procedimentos metodológicos apresentados acima, o trabalho foi organizado do seguinte modo:

A Seção 2 desenvolve a fase inicial do domínio linguístico-computacional, dedicada a (a) entender em que medida a polissemia constitui um problema para o PLN e (b) avaliar como a questão tem sido enfrentada pelas bases *wordnets*, percebendo possíveis avanços e limitações na codificação da multiplicidade de sentidos. Para equacionarmos as questões, apresentaremos, em 2.1, *a interação entre polissemia, processamento automático de língua natural e wordnets*; em 2.2, *a caracterização de uma base relacional de dados lexicais do tipo wordnet*; em 2.3, *a proposta da WordNet.Pr*; em 2.4, *a proposta da WordNet.Pt*; em 2.5, *a proposta da GermaNet*; e, por fim, em 2.6, *a proposta do CoreLex*.

As Seções 3 e 4 desenvolvem a fase inicial do domínio linguístico, dedicada a demonstrar em que medida a Semântica Lexical Cognitiva pode ser considerada uma teoria (a) suficientemente abrangente e (b) com poder descritivo adequado à representação da polissemia de nominais para os fins propostos. Para tanto, na Seção 3, investigaremos os seguintes tópicos: em 3.1, *que desafios a polissemia apresenta a uma*

⁵ Endereço: <http://www.w3c.br/>

teoria linguística; em 3.2, *por que escolher o olhar da semântica lexical cognitiva*; em 3.3, *a caracterização da polissemia como um fenômeno categorial*; em 3.4, *aspectos estruturais e funcionais da dimensão semasiológica do léxico*; em 3.5, as abordagens homonímica, polissêmica e monossêmica; por fim, em 3.6, *as redes de polissemia*. Na Seção 4, abordaremos os seguintes tópicos: em 4.1, *a delimitação dos sentidos*; em 4.2, *efeitos de saliência entre os sentidos*; por fim, em 4.3, *a similaridade e a polissemia*.

A Seção 5 desenvolve a proposta de representação da polissemia em termos do constructo *polyset*, respondendo às questões representacionais dos domínios linguístico e linguístico-computacional. Os seguintes tópicos são abordados: em 5.1, *o polyset como um constructo de status múltiplo* e, em 5.2, *os componentes e a estrutura do polyset*.

A Seção 6, por fim, descreve a construção e implementação computacional do *polyset* através do editor Protégé-OLW. A seção é organizada em duas grandes subseções: a face linguística de construção do *polyset*, em 6.1 e a face computacional de construção do *polyset*, em 6.2.

SEÇÃO 2 - A POLISSEMIA COMO UM DESAFIO À ESTRUTURAÇÃO DE BASES RELACIONAIS DE DADOS LEXICAIS DO TIPO *WORDNET*

Oferecer uma representação linguístico-computacional da polissemia de nominais compatível com sistemas de PLN via *wordnet*, especialmente, WN.Br, é relevante para contribuir para a solução de uma das limitações identificadas nesse tipo de recurso, que é a “relativa baixa densidade de relações” (BOYD-GRABER *et. al*, 2006, p.2). Para reconhecermos as dimensões dessa tarefa, demanda que nos propusemos a abordar nesta seção, refletiremos inicialmente sobre a interação entre o PLN, a polissemia e as *wordnets* e, posteriormente, investigaremos o modo como as diferentes *wordnets* e o CoreLex, recurso relacionado, têm enfrentado os desafios elencados. A partir dessas questões e da avaliação do estágio de desenvolvimento atual da WN.Br, na seção final da tese, teremos mais elementos aptos a identificar os desafios a serem solucionados por nossa proposta de representação. Esta seção dedica-se, portanto, à abordagem da primeira fase do domínio linguístico-computacional e está organizada do seguinte modo: 2.1, *a interação entre polissemia, processamento automático de língua natural e wordnets*; 2.2, *a caracterização de uma base relacional de dados lexicais do tipo wordnet*; 2.3, *a proposta da WordNet.Pr*; 2.4, *a proposta da WordNet.Pt*; 2.5, *a proposta da GermaNet*; e, por fim, em 2.6, *a proposta do CoreLex*.

2.1 A INTERAÇÃO ENTRE POLISSEMIA, PROCESSAMENTO AUTOMÁTICO DE LÍNGUA NATURAL E WORDNETS

O que entendemos por polissemia? Qual é a relação entre polissemia e PLN? Por que tratar computacionalmente a polissemia? Como tratar da polissemia de forma útil ao PLN? Em que medida as *wordnets* inscrevem-se nesse contexto? O que já se sabe sobre a representação da polissemia e sua utilidade para sistemas de PLN? Em que medida a polissemia é um desafio ao PLN e às bases *wordnets*? Essas são algumas das questões que nos vêm à mente quando pensamos sobre a relação entre polissemia, PLN e *wordnets*. Ao longo desta seção, investigaremos cada uma delas, a fim de compreendermos em que medida a investigação da polissemia é relevante para o PLN e,

por consequência, por que é relevante incluir a representação da polissemia em uma base de dados lexicais do tipo *wordnet*.

Iniciaremos esta seção fazendo uma primeira delimitação do fenômeno linguístico em foco. Nosso objetivo, neste momento, é apresentar diferentes graus de polissemia e fenômenos correlatos (homonímia e monossemia), conhecimentos que estão pressupostos ao discutirmos sua representação. Para nos auxiliar nesta tarefa, tomaremos como base as definições apresentadas por Apresjan (1973), por ser autor de referência para grande parte dos trabalhos sobre o tema em PLN. Não faremos ainda aqui nenhuma problematização das noções trazidas pelo autor por entendermos que elas são adequadas e suficientes para a delimitação do fenômeno nesta fase do trabalho. O objetivo agora é apenas apresentar de um modo geral a abrangência do fenômeno na superfície linguística para que possamos olhar para as *wordnets* e refletir sobre o tratamento dado a ele por cada base. Para a problematização necessária à definição de polissemia, conforme observamos na introdução, reservamos as Seções 3 e 4.

Tradicionalmente, os estudos sobre multiplicidade de sentidos dividem-se nas perspectivas diacrônica e sincrônica. Em ambas as perspectivas, o que se procura identificar é a *relação*, de étimo, na primeira, e de sentido, na segunda. Para que um falante compreenda um enunciado, não há dúvidas de que o conhecimento etimológico não é relevante. Do mesmo modo, não é a descrição da língua com base na etimologia, mas na semântica sincrônica, a descrição útil à construção de uma base de dados que vise a ampliar a precisão analítica de um sistema de PLN (PUSTEJOVSKY, 1995). Não há distinção estrita entre os fenômenos de multiplicidade de sentidos. Assim, homonímia, polissemia regular e polissemia irregular são “conceitos relativos” (APRESJAN, 1973), no sentido de não serem absolutos. A delimitação desses conceitos não se dá por oposição dicotômica, mas em uma escala contínua e graduada, conforme demonstraremos oportunamente. Sem perder de vista esse contexto e, seguindo esse autor, para a definição breve do fenômeno, apresentaremos uma fórmula lógica tomada como definição de situações prototípicas para cada um dos tipos de multiplicidade de sentidos:

- (a) A **polissemia regular** - “a polissemia de um item lexical A com os sentidos a_i e a_j é regular em uma dada língua se existir pelo menos outro item lexical B com os sentidos b_i e b_j que sejam semanticamente distintos um do outro

exatamente da mesma maneira” (APRESJAN, 1973, p. 16). Exemplo: *livro* como ‘conteúdo’, em *eu li o livro e gostei*, e como ‘objeto físico’, em *o livro molhou com a chuvarada*. Esse mesmo tipo de variação de sentido pode ser observado em *apostila*, *cartilha*, *polígrafo*, entre outros tipos de publicação.

- (b) A **polissemia irregular** - “a polissemia irregular caracteriza-se quando a distinção entre a_i e a_j não for exemplificada em qualquer outra palavra de uma dada língua” (APRESJAN, 1973, p.15). A seguinte fórmula lógica pode ser tomada como critério geral para a identificação da polissemia irregular: um item lexical (A) com os sentidos a_1 e a_2 apresenta polissemia em uma dada língua se **NÃO** existir outra unidade lexical B com sentidos b_1 e b_2 que sejam semanticamente distintos um do outro exatamente da mesma maneira da distinção entre a_1 e a_2 . Exemplo: *posição* como ‘disposição’, em *gostei da posição dos móveis*, e como ‘situação’, em *minha posição na empresa melhorou*.
- (c) A **homonímia** – a homonímia caracteriza-se por uma coincidência puramente externa de dois ou mais itens lexicais cujos sentidos não têm nada em comum (APRESJAN, 1973). A seguinte fórmula lógica, construída a partir dos modelos anteriores, que pode ser tomada como critério geral para a identificação da homonímia: os sentidos a_1 e a_2 associados à forma lexical A são homônimos se entre eles **NÃO** existir nenhum fator de coerência semântica. Exemplo: *manga* como ‘fruta’, em *comprei uma manga bem doce*, e como ‘parte uma vestimenta’, em *minha blusa tem uma manga azul e uma preta*.
- (d) A **monossemia** – a monossemia é o extremo oposto da homonímia, pois não há duplicidade de sentidos, apenas uma “disjunção inclusiva” de componentes de um mesmo sentido: “se $A = 'B \text{ ou } C'$, então $A = 'tanto B, quanto C, \text{ ou } B \text{ e } C \text{ ao mesmo tempo}'$ (APRESJAN, 1973, p. 14). Exemplos: *criança*, item que pode ser empregado tanto para meninos quanto para meninas; *meia-idade*, item que pode ser empregado para expressar que não sabemos exatamente quantos anos a pessoa tem, mas isso não justifica dizer que há sentidos distintos.

Apresentados os diferentes graus de multiplicidade de sentidos, passaremos a tratar das questões que elencamos acima sobre o PLN, sua relação com a polissemia e com as bases *wordnets*. Olhando especificamente para a relação PLN e semântica, podemos dizer que o PLN procura *representar o significado*, entendendo-se, nesse contexto, *representação* como termo equivalente à descrição formal ou formalizável da língua. A representação do significado torna possível a construção de sistemas computacionais capazes de reconhecer e produzir informação apresentada em língua natural. Conforme registrado na abertura do *site* da Comissão de PLN na SBC, fundada em 2007, os sistemas construídos a partir de PLN são

Tradução Automática, Sumarização Automática de Textos, Ferramentas de Auxílio à Escrita, Perguntas e Respostas, Categorização Textual, Recuperação e Extração de Informação, entre muitas outras, além das tarefas relacionadas de criação e disponibilização de dicionários/léxicos e *córpus* eletrônicos, desenvolvimento de taxonomias e ontologias, investigações em linguística de *córpus*, desenvolvimento de esquemas de marcação e anotação de conhecimento linguístico-computacional, resolução anafórica, análise morfossintática automática, análise semântico-discursiva automática, *etc.*" (www.sbc.org.br/cepln)

Vemos aí que a tarefa com a qual colaboraremos nesta tese - a construção de uma porção de uma *base relacional de dados lexicais* do tipo *wordnet* - está prevista, pois pode ser considerada, para esse fim, equivalente à tarefa de criação de dicionários/léxicos/taxonomias/ontologias. A construção de recursos léxicos no contexto do PLN pode ser feita de maneira automática ou semi-automática. A porção "semi" de um processo automático nesse contexto remete a qualquer tipo de intervenção humana durante o processo. Cientistas da computação que defendem a importância dessa intervenção manual consideram que um processo semiautomático pode ser muito útil para encontrar as melhores alternativas à, por exemplo, realização de uma tradução, auxiliando na indicação de escolhas a serem feitas pelo sistema ou em sua correção, contribuindo para o processo como um todo. É consensual, também, em PLN, entender que essa contribuição é altamente custosa em termos de tempo, mas muito importante do ponto de vista do processamento semântico porque auxilia especialmente na realização da tarefa que é um dos principais desafios à máquina: a desambiguação (atividade conhecida como Word Sense Disambiguation – WSD).

Escolher adequadamente o sentido de um item lexical, ou seja, desambiguar esse item lexical, é tarefa central para uma série de aplicações, entre elas, a de recuperação de informações. Conforme explica Voorhees (1998), encontrar o(s) documento(s) desejado(s) em um conjunto heterogêneo de textos requer a compatibilização adequada entre a palavra de busca e o conteúdo do documento. Sabe-se que o grau de detalhamento do conteúdo exigido pelos diferentes sistemas de PLN pode variar: em alguns momentos, a intenção pode ser apenas identificar o tópico do texto (sistema de classificação de textos); em outros, a intenção pode ser identificar o texto que contém a chave de busca (sistema de busca de informação); pode haver, ainda, a necessidade de o sistema construir uma resposta a partir da pergunta do usuário (sistema de pergunta e resposta). Portanto, a realização inadequada de inferências sobre o sentido do item lexical a ser delimitado pode levar à recuperação de documentos e informações irrelevantes pelo sistema, tornando-o ineficiente. A problematização nesse contexto, em termos gerais, envolve a determinação de quais sentidos representar e como representá-los de modo a servir a qualquer tipo de sistema de PLN, possibilitando a seleção do sentido adequado ao contexto em questão.

Uma primeira estratégia de representação é a denominada *homonímica* (também conhecida como a estratégia do *léxico enumerativo*), que consiste na descrição em termos de uma lista exaustiva de sentidos, de modo que uma mesma forma é listada o número de vezes que ela for associada a um novo sentido. A abordagem é criticada por diferentes razões, entre elas: por somente listar os sentidos sem estabelecer qualquer relação entre eles; pela arbitrariedade (ou pelo menos pela falta de critério consistente) para selecionar e cobrir os sentidos necessários; por falhar frequentemente ao registrar o escopo completo de uso de um item lexical; e pela inabilidade de prever novos usos/sentidos (NIRENBURG & RASKIN, 2003). Entretanto, a descrição do conteúdo lexical de modo enumerativo, desde que cuidadosamente planejada e com procedimentos baseados em teoria e metodologia eficientes, conforme afirmam Nirenburg e Raskin (2004), pode vir a ser adequada a certas tarefas.

Situaremos em posição contrária a da estratégia homonímica a estratégia denominada polissêmicas. Um léxico estruturado segundo essa estratégia lista os diferentes sentidos, porém indica as conexões entre eles, o que possibilita perceber similaridades e, inclusive, se olharmos para a hierarquia de um léxico como um todo, inferir novos sentidos. Representar explicitamente as diferentes conexões que se

estabelecem entre os sentidos polissêmicos tem se mostrado estratégia que contribui para melhorar o desempenho de sistemas computacionais de diferentes tipos (ex. KROVETZ, 1997; BUITELAAR, 2000; STOKOE, 2005). Krovetz (2005), por exemplo, testou a capacidade de desambiguação de sistemas de busca de textos e de informações em relação à sua sensibilidade à polissemia e à homonímia. Para a realização do experimento, foram incluídos itens lexicais fictícios na base de dados do sistema. Em um primeiro momento, os itens incluídos foram marcados como homônimos e, em um segundo momento, como polissêmicos. Os resultados revelaram que o sistema é mais sensível à polissemia do que à homonímia. Essa conclusão levou os autores a perceber que a representação de sentidos em granularidade fina e com diferentes níveis de conexão (a referida estratégia conhecida como polissêmica, em oposição à homonímica) oferece benefícios adicionais para a desambiguação em comparação à representação com pouca granularidade (estratégia conhecida como monossêmica).

A representação de graus de polissemia mais sutis, como a polissemia regular, também é considerada relevante (ex. BUITELAAR, 2000). Representar explicitamente relações regulares entre sentidos polissêmicos pode otimizar o processamento da semântica lexical por um sistema de PLN, tendo em vista que as decisões a tomar durante a desambiguação podem ser restritas àqueles casos que envolvem maior contraste e, portanto, ambiguidade real (BUITELAAR, 2000). Diferentemente do que ocorre com a homonímia, itens lexicais regularmente polissêmicos nem sempre precisam ser desambiguados porque possuem sentidos distintos muito próximos. Em muitas aplicações, por exemplo, na categorização de documentos e em outros tipos de sistemas de extração de informação, pode ser suficiente saber se dado item pertence a determinado grupo polissêmico mais do que saber qual dos sentidos (relacionados) está sendo empregado no texto (BUITELAAR, 2000). Em outras palavras, pode ser suficiente reconhecer níveis mais gerais de sentido que mantêm diversas informações subespecificadas, deixando que, em um momento futuro, se houver necessidade, elas sejam especificadas.

As situações mencionadas acima levam à conclusão de que a granularidade ideal para a representação dos sentidos varia de aplicação para aplicação (KROVETZ, 1997; BUITELAAR, 2000; STOKOE, 2005). Para termos um exemplo mais específico, é interessante observar duas situações distintas mencionadas por Buitelaar (2000) relacionadas a sistemas de busca. Um primeiro caso é aquele em que não é necessário

distinguir entre duas interpretações do item polissêmico, por exemplo, *hazelnut*, como ‘fruto avelã’ e como ‘árvore aveleira’. Essa situação pode ocorrer quando o usuário está procurando tanto pelo *fruto* quanto pela *árvore*. Outro caso pode ser caracterizado quando um usuário está interessado especificamente em conhecer as propriedades do *fruto* e não deseja saber sobre a *árvore*. Nessa situação, a seleção de um dos sentidos *árvore/fruto* se torna necessária. A extração de classes de nominais regularmente polissêmicos, como *hazelnut*, pode ser usada como filtros em tarefas de WSD em várias aplicações que necessitam de granularidade baixa, mas, conforme vimos, tendo em vista que o grau alto de granularidade pode ser o nível ideal para certas tarefas automáticas, uma base de dados lexicais construída com fim genérico precisa fornecer informações úteis e suficientes para que diferentes sistemas de PLN possam dela extrair o grau de detalhamento necessário à tarefa em questão.

Conforme observamos ao longo da seção, para diferentes tarefas de PLN, é necessário que o sistema acesse o significado lexical que pode estar armazenado em diferentes níveis de generalidade ou granularidade. A partir daí, uma questão que precisa ser resolvida é como extrair tais informações dos léxicos e que recursos são necessários para a realização de WSD. As tarefas de WSD baseiam-se fundamentalmente em técnicas automáticas e semiautomáticas de identificação de *similaridade de sentidos*. A noção de similaridade é empregada para sugerir uma série de agrupamentos de sentidos. Buscando uma definição para a similaridade que não seja baseada em uma técnica específica, Lin (1998) elenca três intuições que são responsáveis pela atribuição ou não do rótulo de similar a dois ou mais sentidos: (a) a similaridade entre A e B está relacionada ao que há de comum entre eles, de modo que, quanto mais fatores comuns eles compartilharem, mais similares eles serão; (b) a similaridade entre A e B está relacionada às diferenças entre eles, de modo que, quanto mais diferenças eles expressarem, menos similares eles serão; (c) o máximo de similaridade entre A e B é alcançado quando A e B são idênticos, não importando se compartilham elementos comuns ou não. A partir dessa definição tripartida, o autor cria modelos matemáticos para extrair similaridade de textos e hierarquias. A literatura sobre técnicas de identificação de similaridades é vastíssima (entre os nomes citados, estão MCGILL et al., 1979; LEE et al., 1989;; LEACOCK-CHODOROW, 1998; HIRST-ST.ONGE, 1998; HINDLE, 1990; FRAKES e BAEZA-YATES, 1992; RESNIK, 1995; PATWARDHAN, 2003).

Agirre e Edmonds (2007), na introdução do livro *Word Sense Disambiguation Algorithms and Applications*, relatam que a WN.Pr (juntamente com as *wordnets* para outras línguas) tem sido o inventário de sentidos gerais mais usados em pesquisa para WSD, visto que ela armazena o conteúdo lexical de maneira computacionalmente acessível, organizando os sentidos hierarquicamente a partir dos chamados *synsets* (conjunto de sinônimos) (AGIRRE e EDMONDS, 2007). Os autores também salientam que aplicações que fazem uso das *wordnets* têm obtido sucesso tanto em procedimentos semiautomáticos quanto estatísticos e de aprendizagem de máquina. Métodos que fazem anotação manual de cópús (isto é, métodos de aprendizagem de máquina supervisionada) têm sido predominantes em WSD e têm obtido os melhores resultados na competição conhecida como Senseval⁶, que se dedica a avaliar sistemas de WSD (AGIRRE e EDMONDS, 2007).

Entre as estratégias de utilização das *wordnets* para a desambiguação, está a identificação de similaridades entre sentidos através das relações de hiponímia e de sinonímia ou, ainda, através da extração automática de informações a partir das glosas. Nesse contexto, há um projeto, denominado WordNet::Similarity, dedicado a propor medidas automáticas para extrair similaridades exclusivamente de *wordnets*. Em linhas gerais, o que ocorre é que a desambiguação a partir das *wordnets* é feita com base na comparação entre itens lexicais distintos. A utilização das *wordnets* para tarefas de WSD, no entanto, não revela somente vantagens. Kerner (2008) chama a atenção para o fato de que a granularidade fina e os sentidos enumerados em um único nível – características das *wordnets* – exigem maior esforço do sistema. Isso porque exigem que o sistema decida entre múltiplos sentidos, que, muitas vezes, apresentam diferenças sutis, independentemente da necessidade do sistema, que pode ser a classificação de textos (que exige menor precisão) ou a tradução automática (que exige maior precisão).

Conforme mencionamos a partir da definição de Apresjan (1973), a polissemia é também um tipo de similaridade de sentidos que se manifesta na superfície linguística associada a um único item lexical (simples ou composto). Assim, tendo em vista as informações apresentadas nesta seção que revelaram a importância do agrupamento de sentidos para o aperfeiçoamento de sistemas de PLN, nas seções seguintes, estudaremos as *wordnets* a fim investigar a ocorrência e o tipo de técnica para a codificação de

⁶ <http://www.senseval.org/>

similaridade empregada por *wordnets* selecionadas e pelo CoreLex, recurso que faz uso da WN.Pr em sua construção.

2.2 CARACTERIZAÇÃO DE UMA BASE RELACIONAL DE DADOS LEXICAIS DO TIPO *WORDNET*

Tendo em vista que o objetivo geral desta seção é identificar os desafios a serem solucionados por nossa proposta de representação aplicada à WN.Br, faz-se necessário, antes de iniciarmos a reflexão sobre as estratégias apresentadas por cada base de dados, conhecermos com mais detalhes as características de sua organização. A caracterização do paradigma *wordnet* tem aqui duas funções. A primeira delas é a caracterização da hierarquia geral de uma base relacional de dados lexicais, a fim de percebermos as possibilidades oferecidas e as limitações da estrutura para o tratamento das relações de polissemia a serem propostas nesta tese. Somente a partir de um detalhado estudo sobre os recursos descritivos oferecidos por uma *wordnet* é que poderemos propor uma estratégia descritiva para as relações de polissemia. A segunda é identificarmos o de informação que pode ser representada em uma *wordnet* para podermos formular os critérios de delimitação dos sentidos de um item lexical polissêmico.

A primeira *wordnet* – a WordNet.Pr – foi construída com o objetivo de compreender como as crianças aprendem novas palavras. Essa foi a motivação do Professor de Psicologia George A. Miller e de sua equipe para iniciar a construção, na Universidade de Princeton/EUA, nos anos 80, da WN.Pr para a língua inglesa. Atualmente, embora haja ainda preocupação com a motivação psicolinguística, é a preocupação com a utilidade dessa rede para pesquisas de PLN, que é a principal razão para as reelaborações propostas por seus desenvolvedores (FELLBAUM & MILLER, 2006). A WN.Pr foi concebida como um repositório altamente enriquecido de conteúdo léxico-semântico estruturado de maneira a simular o que se acredita ser a organização do léxico mental humano (MILLER & FELLBAUM, 2007). A partir da WN.Pr, diversos projetos surgiram e redes, seguindo proposta similar, têm sido construídas individualmente para diferentes línguas (ex.: português, alemão, italiano, japonês, latim,

espanhol, hebreu, romeno, japonês, etc.), além de recursos que se propõem a integrar essas diferentes bases (EuroWordNet⁷ e MultiWordNet⁸).

A arquitetura de uma *wordnet* traz de um dicionário tradicional a inserção de glosas para representar o sentido de nominais, verbos, adjetivos e advérbios. Entretanto, dele se diferencia por não seguir o critério alfabético em sua macroestrutura, mas o de estruturação semântica. Em uma *wordnet*, o léxico de uma língua é representado como uma rede de formas lexicais que mapeiam para um ou mais sentidos representados por meio de *synsets* (BOAS e FELLBAUM, 2009). Desse modo, um item lexical é representado por uma sequência de caracteres (*string*), e um sentido é representado por um conjunto formado por um ou mais itens lexicais sinônimos (o *synset*). Os *synsets* são formados por itens lexicais de mesma classe gramatical que podem ser substituídos em pelo menos um contexto de uso (VOSSEN, 1999), são os *sinônimos contextuais* ou *cognitivos*. Por exemplo, o *synset* {*carro, automóvel*} tem seus elementos intersubstituíveis no contexto *Ele precisa de um ___ para ir ao trabalho* e aponta para o sentido glosado por “veículo de quatro rodas impulsionado por um motor de combustão interna”⁹. Cada *synset* define os sentidos associados a uma forma lexical de maneira implícita a partir dos seus membros (os outros itens lexicais do *synset*) e possibilita que o sentido seja inferido automaticamente por sistemas computacionais a partir das relações léxico-conceituais explicitamente codificadas entre os diferentes *synsets*. A figura 1 ilustra o *synset* da WN.Pr (versão 2.0) do qual *book* é parte e as diferentes informações a ele associadas.

{02870092} <noun.artifact>S: (n) book, volume (physical objects consisting of a number of pages bound together) "he used a large book as a doorstep"

Figura 1 – O *synset* *book* e informações associadas

⁷ <http://www.illc.uva.nl/EuroWordNet/>

⁸ Disponível em <http://multiwordnet.itc.it/english/home.php>

⁹ Não se pode confundir a noção de sinonímia adotada pelas *wordnets*, a sinonímia contextual, com a rara ou inexistente noção de sinonímia absoluta, fenômeno que, sob uma perspectiva contextual, corresponde à completa identidade de sentidos em todos os contextos; ou com a noção de sinonímia proposicional, quando dois itens lexicais podem ser substituídos em qualquer proposição sem alteração na verdade condicional, havendo apenas diferença na expressividade da significação, no nível estilístico (coloquial/ formal) ou no campo pressuposto de discurso (CRUSE, 2000). Uma *wordnet* adota, portanto, uma noção bastante ampla de sinonímia.

O *synset* em que se insere *book* é identificado pelo número {02870092}, é formado pelos itens lexicais *book*, *volume* e recebe de *tipo semântico* (*unique beginner*) <noun.artifact>. Os *synsets* organizam-se hierarquicamente. No topo da hierarquia, estão os *unique beginners*, conceitos genéricos que iniciam diferentes hierarquias que englobam o vocabulário pertencente a campos semânticos distintos, sendo que cada campo semântico é representado por um *unique beginner*. A WN.Pr organiza a hierarquia de substantivos a partir de 25 *unique beginners*, indicados na figura 2.

{act, action, activity}	{natural object}
{animal, fauna}	{natural phenomenon}
{artifact}	{person, human being}
{attribute, property}	{plant, flora}
{body, corpus}	{possession}
{cognition, knowledge}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation}
{feeling, emotion}	{shape}
{food}	{state, condition}
{group, collection}	{substance}
{location, place}	{time}
{motive}	

Figura 2 - Os 25 *unique beginners* da hierarquia dos *synsets* para os nomes da WN.Pr
Fonte: MILLER, 1990, p. 17

A partir de cada *unique beginner*, os demais *synsets* são, hierarquicamente, apresentados através de relações. As relações entre itens lexicais são denominadas *relações lexicais*. As relações entre *sentidos* ou *conceitos* associados aos itens lexicais distintos, isto é, entre *synsets*, são denominadas *relações semânticas* ou *conceituais*. Tendo em vista que itens lexicais e sentidos estão intimamente interligados, as expressões *relações léxico-semânticas* ou *léxico-conceituais* são frequentemente empregadas quando não é necessário ou possível estabelecer distinção (EVENS, 1988). Entre as relações lexicais codificadas nas *wordnets*, estão a *sinonímia* e a *antonímia*. De acordo com os critérios das *wordnets*, *cão* e *cachorro* devem ser considerados sinônimos, visto que podemos dizer que há acarretamento bilateral entre esses elementos em um contexto como *Todos devem levar seus cachorros/cães para vacinar*. Os nomes *homem* e *mulher* devem ser codificados nas *wordnets* como antônimos, uma vez que cada um

deles caracteriza propriedades disjuntas, caracterizando-se um tipo de antonímia binária (CANÇADO, 2005).

Entre as relações semânticas mais relevantes codificadas nas *wordnets* para a descrição do conteúdo associado aos nominais, destacam-se a *hiponímia-hiperonímia* e a *meronímia-holonímia*. A relação de *hiponímia-hiperonímia* (ou relação de *é-um*) é uma relação de acarretamento unilateral. Como exemplo, podemos ter *apartamento*, que é um tipo de *imóvel*, uma vez que *apartamento* (hipônimo) é uma entidade mais específica do que *imóvel* (hiperônimo). A relação semântica de *meronímia-holonímia* (ou de *parte-todo*) pode ser percebida entre *braço* e *corpo*, visto que *braço* é parte do *corpo*. Algumas *wordnets* (ex. EuroWordNet, WordNet.Pt) incluem detalhamentos da relação de meronímia como *mero-membro* (ex. *professor* é mero-membro do *corpo docente*), *mero-porção* (ex. *fatia* é mero-porção de *pão*), *mero-matéria* (ex. *bambu* é mero-matéria de *colher de bambu*), *mero-localização* (ex. *centro* é mero-localização de *cidade*). Visto que as relações semânticas são relações entre sentidos associados a formas lexicais distintas e que os sentidos em uma *wordnet* são representados pelos *synsets*, é natural que as relações semânticas nas *wordnets* sejam representadas como rótulos que indicam relações entre os *synsets*. O quadro 1 mostra relações da WN-Pr que se estabelecem entre nomes.

<i>Relação</i>	<i>Classe gramatical relacionada</i>	<i>Exemplo</i>
SINONÍMIA	nome/nome	posição/opinião
ANTONÍMIA	nome/nome	homem/mulher
HIPONÍMIA- HIPERONÍMIA	nome/nome	animal/cachorro
MERONÍMIA- HOLONÍMIA	nome/nome	cachorro/patas

Quadro 1- Relações do tipo *nome-nome* da WN-Pr
(adaptado de VOSSEN, 1999)

O quadro 1 ilustra as relações semânticas básicas que estruturam as *wordnets* em uma rede maior com **nós** linguisticamente rotulados (MILLER & FELLBAUM, 2007). Cada um desses nós é formado por um ou mais itens lexicais equivalentes em sua denotação (*synsets*). Sob o enfoque topológico, entende-se que cada *synset* constitui um nó da rede e que as relações léxico-semânticas são os arcos que ligam esses nós. A

existência de um **arco** entre os *synsets* representa, por exemplo, a proposição *um cachorro é um tipo de animal*. Outro tipo de arco, expressando parte, pode relacionar *cachorro* e *patas*, expressando o fato de que *patas são parte de cachorro* e, através de herança, uma *pata* é parte de todos os tipos de *cachorro*, tais como *rottweiler*, *schnauzer*, *poodle*.

Apresentamos nesta seção as principais características de uma *wordnet* que são mais gerais e comuns aos exemplos de bases que selecionamos para tratar. Além disso, inventariamos aqui informações-chave que serão importantes mesmo que secundariamente à proposição da representação de nominais polissêmicos; são elas: (a) a estrutura base da *wordnet* a partir de *synsets*, constructo que deve ser entendido como a representação de um sentido, em certa medida, comum a todos os seus membros; (c) a organização geral da base a partir de relações semânticas que conectam *synsets* e não itens lexicais individuais; (d) os tipos de relações semânticas; e, por fim, (e) a indicação de um tipo semântico mais geral particular a um grupo de sentidos que são os *unique beginners*. Essas informações serão retomadas explicitamente ao longo da apresentação dos exemplos de bases de dados relacionais, e, posteriormente, algumas delas são necessárias para a apresentação da estratégia de representação a ser proposta na Seção 5. As bases escolhidas como exemplo de estratégias comuns às *wordnets* foram a WN.Pr, a WN.Pt e a GermaNet. O CoreLex será também apresentado por representar uma robusta base de dados de polissemia regular extraída da WN.Pr e da GN.

2.3 A PROPOSTA DA WORDNET.PR

Na WordNet.Pr¹⁰, a multiplicidade de sentidos é codificada sob o rótulo de *polissemia*. Um item lexical é definido como “polissêmico” quando participa de mais de um *synset* associado à mesma categoria sintática, independentemente de quão similares são os sentidos a ele associados (BOAS e FELLBAUM, 2009). O rótulo *polissemia* é empregado, portanto, de uma maneira mais geral do que a que estamos adotando neste trabalho. Trata-se de um parâmetro para demonstrar o grau de familiaridade (*familiarity*) dos itens lexicais, apresentando uma contagem de *polissemia* (*polysemy*

¹⁰ Disponível em <http://wordnetweb.princeton.edu/perl/webwn>.

count) que indica o número de *synsets* em que a unidade ocorre. Nenhum desdobramento maior dessa noção é levado em conta, fato que resulta na não-discriminação sistemática da relação existente entre os múltiplos sentidos associados a uma mesma forma lexical, ficando esta difusamente representada em termos de índices e ocorrências de uma mesma forma em *synsets* distintos. O item lexical *bank*, por exemplo, está associado ao número de *polissemia* 10, indicando que essa unidade ocorre em 10 *synsets* distintos. A figura 3 ilustra esses *synsets* e suas respectivas informações associadas.

10 senses of bank

Sense 1

<noun.group> depository financial institution, bank, banking concern, banking company -- (a financial institution that accepts deposits and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the mortgage on my home")

Sense 2

<noun.object> bank1 -- (sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents")

Sense 3

<noun.possession> bank -- (a supply or stock held in reserve for future use (especially in emergencies))

Sense 4

<noun.artifact> bank, bank building -- (a building in which commercial banking is transacted; "the bank is on the corner of Nassau and Witherspoon")

Sense 5

<noun.group> bank1 -- (an arrangement of similar objects in a row or in tiers; "he operated a bank of switches")

Sense 6

<noun.artifact> savings bank, coin bank, money box 1, bank1 -- (a container (usually with a slot in the top) for keeping money at home; "the coin bank was empty")

Sense 7

<noun.object> bank -- (a long ridge or pile; "a huge bank of earth")

Sense 8

<noun.possession> bank1 -- (the funds held by a gambling house or the dealer in some gambling games; "he tried to break the bank at Monte Carlo")

Sense 9

<noun.object> bank2, cant, camber -- (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)

Sense 10

<noun.act> bank -- (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning); "the plane went into a steep bank")

Figura 3 - Synsets que contêm o item lexical *bank* na WN.Pr.2.0

Na figura 3, observamos as características que mencionamos na seção anterior como limitações da rede para tarefas de WSD: uma delas é a não-distinção entre polissemia e monossemia, e a outra é em relação à não-diferenciação entre níveis de generalidades entre os sentidos polissêmicos. Observa-se que os *synsets* 1 e 2 remetem a sentidos sem qualquer relação: (1) *instituição financeira* e (2) *margem, costa* (de rio, por exemplo) – caracterizando a *homonímia*. Os *synsets* 1 e 4 expressam sentidos que não são excludentes, mas complementares, remetendo tanto a (1) *instituição* quanto a (4) *prédio* – caracterizando a *polissemia regular*. Entre os *synsets* (1, 3, 4, 6, 8), há um traço semântico compartilhado: *local para armazenamento de algo de valor*. O mesmo ocorre entre os *synsets* (2, 7, 9, 10), sendo *elevação* o traço comum. Esses dois grupos de *synsets*, {1,3,4,6,8} e {2,7,9,10}, ilustram, cada um, a *polissemia irregular*.

Miller (1998) justifica a falta de indicação de graus de similaridade e níveis distintos de sentidos polissêmicos com base no imenso custo que essa tarefa exigiria, uma vez que depende de delicados julgamentos de humanos sobre propriedades semânticas altamente sutis. Acrescenta, ainda, que discordâncias entre os julgamentos, em razão da sutileza das nuances de sentidos, seriam frequentes. Embora o agrupamento de sentidos similares tenha sido implementado apenas para certos *synsets* – exemplo, Miller (1998) sugere uma alternativa para extração e codificação de conexões entre sentidos similares, criando um segundo nível de refinamento de sentido. Entre as estratégias para a utilização das relações da WN.Pr para a identificação de tipos de similaridade entre sentidos, destacamos as relações de: *cousins* (primos) e *sisters* (irmãs), de relevância comprovada também por outros estudos (ex.: VOSSSEN et. al, 1998).

Se dois sentidos associados a um mesmo item lexical são similares, então os sentidos dos seus hipônimos deveriam ser também similares da mesma maneira (Miller, 1988). Esse é o primeiro critério de similaridade proposto por Miller, denominado *primo*. A partir desse critério, é possível localizar classes de nomes regularmente polissêmicos. A tarefa de localizar itens lexicais primos exige que sejam encontrados nós da rede em posição mais alta possível, de modo que estabeleçam um tipo de similaridade comum a um conjunto de itens, como ocorre entre *animal* e *alimento* (exemplo adaptado de Miller, 1998, p. 42). Para *peixe*, por exemplo, a relação se dá do seguinte modo: se dois sentidos de *peixe* estão relacionados tanto a *animal* quanto a *alimento*, então também todos os seus hipônimos (ex.: *lambari, sardinha, garoupa,*

bacalhau, etc.) terão a mesma relação. Se o sentido de *sardinha*₁ é hipônimo de *peixe*₁ (animal) e *sardinha*₂ é hipônimo do alimento *peixe*₂ (alimento), essas relações caracterizam o fato de que os dois sentidos de *sardinha* estão também relacionados tanto a *peixe* ('animal') quanto a *peixe* ('alimento'), conforme mostra a figura 4.

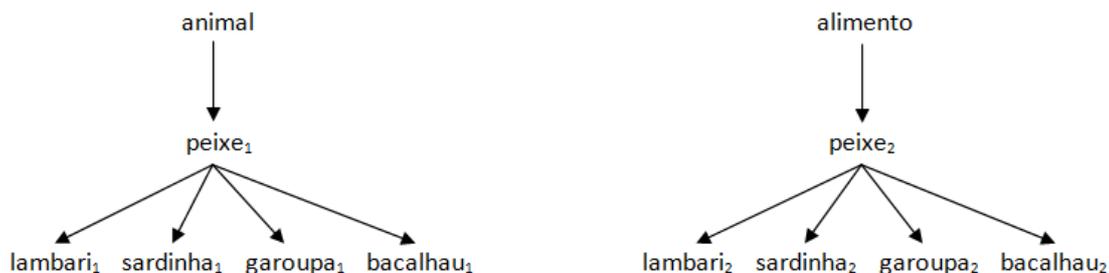


Figura 4 – Exemplo de itens lexicais primos

Conforme mostra a figura 4, os pares de sentidos indicados pelos índices 1 e 2 são primos: *peixe*, *lambari*, *sardinha*, *garoupa*, *bacalhau*. Essa metodologia foi automaticamente testada, e Miller chama a atenção para a necessidade de criação manual de uma lista de exceções, o que dificulta a geração automática de itens lexicais primos. Adaptando o exemplo do autor, citamos *coral*; embora seja hipônimo tanto de *animal* quanto de *alimento*, um coral não é alimento do mesmo modo que um peixe, visto que não é seu corpo que é comestível, mas parte dele (referindo-se a *lobster roe* – ovas de crustáceos comestíveis).

O segundo critério de agrupamento de sentidos similares proposto por Miller é a relação de itens lexicais irmãos (*sister*), frequentemente encontrada em estruturas arbóreas e caracterizada quando dois sentidos associados a uma mesma forma lexical são hipônimos imediatos de um mesmo nó. Essa relação é também conhecida como *co-hiponímia* em uma perspectiva de relação léxico-semântica. O pressuposto é que são irmãos dois sentidos associados a uma mesma forma lexical que compartilham o mesmo hiperônimo. O exemplo originalmente citado pelo autor é *flounder*₁ e *flounder*₂, que têm como hiperônimo *flatfish* já que *flounder* refere-se a diferentes tipos de *flatfish*. Em PB, um exemplo semelhante é o de *faca*, que pode ser vista, por exemplo, como ‘arma branca’, como ‘utensílio doméstico’ ou, em uso informal, como ‘instrumento cirúrgico’, conforme mostra a figura 5.

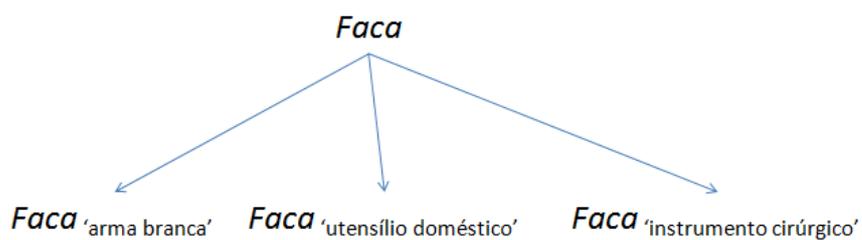


Figura 5 - Exemplo sentidos irmãos

Para estruturar a WN.Pr, Miller sugere que os critérios de agrupamentos apresentados sejam considerados como uma alternativa à visualização por frequência de ocorrência dos *synsets*. Na base da WN2.0, encontramos a seguinte representação a partir da palavra de busca *sole*.

Search Word: Redisplay Overview

Searches for sole: Senses:

sole -- (right-eyed flatfish; many are valued as food; most common in warm seas especially European)

- => flatfish -- (any of several families of fishes having flattened bodies that swim along the sea floor on one side of the body with both eyes on the upper side)
 - => spiny-finned fish, acanthopterygian -- (a teleost fish with fins that are supported by sharp inflexible rays)
 - => teleost fish, teleost, teleostan -- (a bony fish of the subclass Teleostei)
 - => bony fish -- (any fish of the class Osteichthyes)
 - => fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills; "the shark is a large fish"; "in the livingroom there was a tank of colorful fish")
 - => aquatic vertebrate -- (animal living wholly or chiefly in or on water)
 - => vertebrate, craniate -- (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
 - => animal, animate being, beast, brute, creature, fauna -- (a living organism characterized by voluntary movement)
 - => organism, being -- (a living thing that has (or can develop) the ability to act or function independently)
 - => living thing, animate thing -- (a living (or once living) entity)
 - => object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
 - => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> food fish -- (any fish used for food by human beings)

- => fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills; "the shark is a large fish"; "in the livingroom there was a tank of colorful fish")
 - => aquatic vertebrate -- (animal living wholly or chiefly in or on water)
 - => vertebrate, craniate -- (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
 - => chordate -- (any animal of the phylum Chordata having a notochord or spinal column)
 - => animal, animate being, beast, brute, creature, fauna -- (a living organism characterized by voluntary movement)
 - => organism, being -- (a living thing that has (or can develop) the ability to act or function independently)
 - => living thing, animate thing -- (a living (or once living) entity)
 - => object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
 - => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"Hypernyms (this is a kind of...)" search for noun "sole"

Figura 6 - Consulta WN.2.0: sole

A figura 6 mostra que os sentidos de *sole* como 'animal' e como 'alimento' (marcados em vermelho) estão agrupados como subentradas de um mesmo *synset* unitário (*sole*). Miller (1998) observa que essa estratégia, implementada na WN.Pr, apenas para fins de exemplificação, não permite tantos agrupamentos quanto o fenômeno da polissemia possibilitaria. À observação de Miller, acrescentamos o fato de que esse tipo de agrupamento funcionou bem para *sole* porque o item faz parte de um

synset unitário. Para o item lexical *bank*, por exemplo, que faz parte de *synsets* não unitários, a estratégia não se mostra satisfatória, isso porque não podemos dizer que tanto *{depository financial institution, bank, banking concern, banking company}* quanto *{bank, bank building}* podem ser, ao mesmo tempo, hipônimos de *{institution, establishment}* e de *{facility, installation}*. Para representar esse tipo de relação, é preciso, em algum momento da representação, abandonar o nível de *synsets* (conceitual/semântico) e olhar para os itens que compõem *synsets* (o nível lexical). Regras dedicadas à extração de itens lexicais primos, ou seja, de polissemia regular, têm sido amplamente exploradas na literatura.

Destaca-se que esta seção foi importante para reconhecermos propostas dos próprios criadores da *wordnet* para a exploração da base com a finalidade de extrair similaridades entre sentidos. A relação *primos* é equivalente ao que definimos como polissemia regular. Ficou evidente, na figura 6, que a solução de visualização proposta não se aplica a itens que participam de *synsets* não unitários, questão cuja solução fica em aberto. A relação de *irmãos* é equivalente ao que descreveremos na Seção 3 como *autohiponímia*, visto que entre os sentidos associados a um mesmo item pode ser percebida uma relação que segue a mesma lógica da relação de hiponímia convencional. O CoreLex, como veremos, explorará a relação de *primos* para construir sua base a partir da WN.Pr e da GN.

2.4 A PROPOSTA DA WORDNET.PT

A WordNet.Pt (WN.Pt)¹¹ (MARRAFA, 2001), *wordnet* do português europeu, seguindo o paradigma EuroWordNet, base que segue os princípios gerais das *wordnets*, mas que se propõe a integrar em um só recurso informações das *wordnets* particulares construídas para as línguas europeias, inclui idéias de representação de múltiplos sentidos associados a uma única forma lexical (ex. MENDES e CHAVES, 2001; MARRAFA, 2001), porém, não se tem notícias de implementações.

Embora as soluções não tenham sido implementadas, conforme afirma Marrafa, a inclusão de propriedades marcantes da polissemia é reconhecida como relevante. A

¹¹ Disponível em <http://www.clul.ul.pt/wn2/>

primeira delas é a representação do conteúdo que leva em consideração a noção de *gradiência*. Embora não seja possível que o tratamento dado ao fenômeno seja uniforme, dada a própria falta de unicidade do fenômeno, afirma Marrafa, é negada a dicotomia entre *polissemia* e *monossemia*, seguindo-se uma abordagem balanceada, na linha de estudos como os de Cruse (2000a). Os testes propostos pelo autor para avaliar o grau de *discretude* (quão discretos ou independentes são dois sentidos) e de *antagonismo* (quão similares ou relacionados são dois sentidos) são previstos para utilização na definição de quantos sentidos novos serão inseridos na base. Os testes utilizados são o de identidade, o de condição independente de verdade, o de relação léxico-conceitual independente e o definicional. As noções de gradiência e os testes de antagonismo e de discretude serão discutidos nas Seções 3 e 4. Traremos aqui apenas um exemplo de teste empregado na construção da WN.Pt como ilustração da metodologia adotada: o teste de relação léxico-conceitual independente. Esse teste parte do princípio de que os sentidos de um item lexical com certo grau de discretude não pertencem ao mesmo domínio semântico; por essa razão, cada sentido possui um hiperônimo distinto, conforme exemplificação na figura 7.

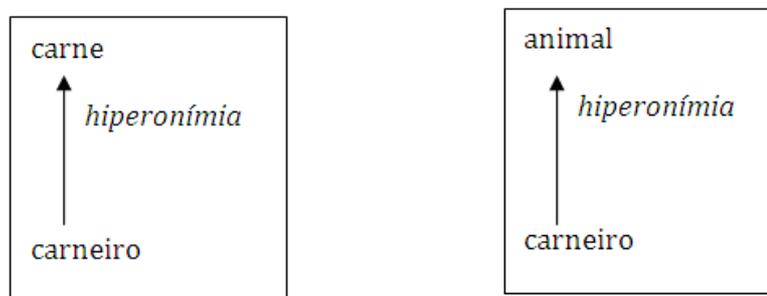


Figura 7 - WN.Pt: teste de relação léxico-semântica independente

Conforme demonstraremos, a interpretação dos testes precisa ser cautelosa. Por concordar com tal posição, Marrafa explica que os testes são aplicados de maneira conjunta e que são utilizados combinatoriamente como parâmetros para determinar um grau relevante de distinção. Ela salienta que os resultados nem sempre são conclusivos, de modo que os testes não representam critérios completos e definitivos para a identificação de antagonismo e discretude entre dois sentidos.

A apresentação que Marrafa faz da organização da WN.Pr em relação à polissemia mostra avanços por prever, pelo menos em teoria, uma estratégia sintonizada com a noção de gradiência, marcante na caracterização da polissemia. Vimos destacada a importância de reflexão sobre quando uma nuance de sentido deve ser considerada ou não um novo sentido a partir da aplicação dos “testes de polissemia”. Os avanços da WN.Pr, no entanto, não vão adiante. A codificação da relação entre os diferentes sentidos associados a uma mesma forma lexical que participa de *synsets* distintos é enfatizada por Marrafa como uma tarefa importante, porém difícil de ser realizada, de modo que a autora não propõe solução para tal tipo de informação.

2.5 A PROPOSTA DA GERMANET

A GermaNet (GN)¹² (FELWEG, 1997) revela a preocupação com a codificação de nomes regularmente polissêmicos. A GN está ligada ao projeto EuroWordNet e segue os princípios de base da WN.Pr. A proposta a ser apresentada está disponível no *site* da base e foi publicada por Hamp e Feldweg (1997). A estratégia não foi implementada, e os exemplos disponibilizados utilizam a base do inglês.

Hamp e Feldweg (1997) chamam atenção para o fato de que um léxico puramente enumerativo não é suficiente para a representação do fenômeno; pelo contrário, destacam que um tratamento baseado em princípios mais gerais precisa ser incluído, seguindo os princípios propostos em Pustejovsky (1995). Na tentativa de superar as limitações para a representação do fenômeno em uma base de dados organizada a partir de *synsets* que deve obedecer a regras de herança, os pesquisadores propõem a inclusão de *pointers* (apontadores) para indicar que há uma relação de polissemia regular envolvendo aqueles *synsets*. O *apontador*, indicando a existência de polissemia regular entre sentidos que fazem parte de entradas separadas do léxico, é inserido no ponto mais alto da hierarquia que manifesta a polissemia regular. É o que mostra a figura 8 para a relação regular que se estabelece entre *instituição* (institution) e *prédio* (facility) via *bank*.

¹² Disponível em <http://www.sfs.uni-tuebingen.de/GermaNet/>

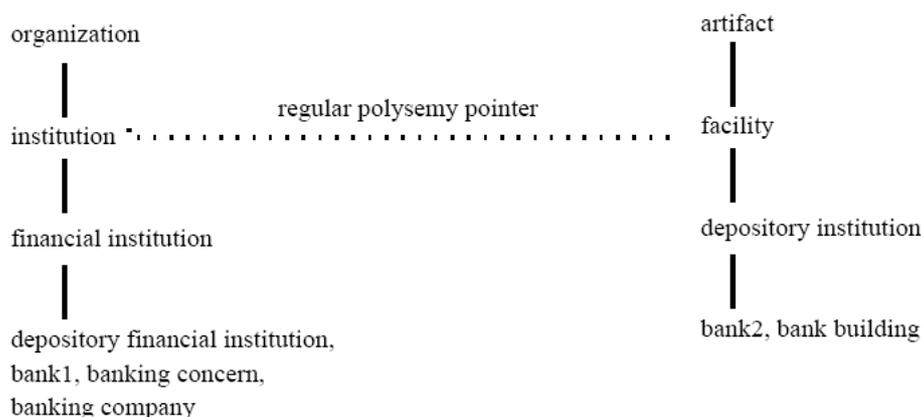


Figura 8 - Polissemia Regular na GN – *bank*
(HAMP e FELDWEG, 1997)

Tendo em vista que o apontador de polissemia regular é reversível e bidirecional, é suficiente, afirmam os autores, marcar manualmente somente um elemento para conectar o par. A relação reversa pode ser gerada automaticamente. O apontador pode ser visto, conforme argumentam Hamp e Feldweg (1997), como uma implementação de um padrão simples através do qual a *wordnet* pode dar conta da produtividade e da regularidade da língua de maneira efetiva.

A marcação de classes polissêmicas estatisticamente frequentes é feita manualmente, declaram os autores. A partir desse processo, componentes menos comuns dessas classes vão sendo reconhecidos de maneira automática a partir de regras de heranças. O apontador é, portanto, visto como auxiliar no reconhecimento de itens regularmente polissêmicos pouco comuns, como *crocodilo*, instância das classes *animal* e *alimento*, por exemplo, em *Comi crocodilo no almoço*. Isso mostra que é possível recuperar, através do apontador de polissemia regular, o sentido de *alimento* para animais, mesmo que esse sentido não esteja entre os representantes clássicos e frequentes das classes polissêmicas. Essa é uma das vantagens destacadas pelos autores. Mas, mas, conforme já vimos em Miller (1999) com o exemplo de *coral*, ela pode se tornar uma limitação.

A partir do exposto, observa-se que a proposta da GN trata a polissemia regular como uma relação entre *synsets*, tal como é a hiponímia, meronímia, etc., que é herdada por todos os hipônimos. Para o exemplo sugerido pelos autores, a técnica funciona perfeitamente. Entretanto, duas possíveis limitações são percebidas: (a) conforme vimos

em Miller (1998), uma lista de exceções precisa ser construída para certas ocorrências de polissemia regular, e (b) o apontador funciona para os sentidos e ignora a sua lexicalização, de modo que os autores não mencionam como fariam para restringir o apontador a apenas um dos elementos do *synset* ou como bloquear a herança para um ou mais componentes do *synset*, como pode ser necessário.

2.6 A PROPOSTA DO CORELEX

O CoreLex (BUITELAAR, 1998) representa o resultado de um esforço para a extração semiautomática de similaridades a partir da WN.Pr com o intuito de construir uma base mais útil à WSD (BOAS & FELLBAUM, 2009). A primeira versão do léxico foi proposta em Buitelaar (1998), e a segunda será aqui apresentada a partir de um artigo de Buitelaar (2000). Nesse projeto, a proposta é extrair classes de nominais regularmente polissêmicos a partir da WN.Pr e da GN. O CoreLex é um léxico estruturado de modo a refletir a sistematicidade e a previsibilidade de propriedades semânticas dos itens de uma língua, centrado nos princípios do Léxico Gerativo (PUSTEJOVSKY, 1995). Explorando a relação de primos (ex. *animal/alimento*), Buitelaar (1998) reconheceu na WN.Pr 105 classes de nominais regularmente polissêmicos. Esse número foi ampliado na versão mais recente. Os critérios para a identificação de classes de itens regularmente polissêmicos serão descritos a seguir, tomando-se como exemplo o item lexical *book*.

No CoreLexI (BUITELAAR, 1998), as classes polissêmicas são identificadas a partir dos tipos básicos (*unique beginners*) da WN.Pr. O primeiro passo para a identificação das classes é extrair da WN.Pr os sentidos associados à forma lexical em análise. Para *book*, a base codifica sete sentidos: (1) 'publication'; (2) 'product, production'; (3) 'fact'; (4) 'dramatic_composition', 'dramatic_work'; (5) 'record'; (6) 'section', 'subdivision'; (7) 'journal'. O segundo passo é a identificação dos tipos semânticos correspondentes. No topo da hierarquia da WN, esses sete sentidos podem ser reduzidos a dois tipos básicos: o conteúdo que está sendo comunicado (*communication*) ou o objeto físico (*artifact*). Os autores explicam que a constatação da existência de uma classe polissêmica se dá através da constatação de que outras

palavras compartilham esses mesmos sentidos e que são, então, polissêmicas da mesma maneira. Por exemplo, os sete sentidos diferentes que a WN.Pr discrimina para *book* podem ser reduzidos a dois tipos básicos: *communication* e *artifact*. A proposta é fazer isso para cada nome e então agrupá-los em classes de acordo com a combinação dos tipos básicos correspondentes. Finalmente, a partir de julgamento humano, diversas classes são agrupadas, porque seus membros são suficientemente similares, explica Buitelaar. Entre as classes resultantes, estão aquelas esperadas e já muito analisadas na literatura, como *animal/alimento*, *planta/ produto natural*. Além dessas, outras classes menos esperadas também foram identificadas, tal como *artefato/atributo/substância*, envolvendo os nomes *chalk*, *charcoal*, *daub*, *fiber*, *fibre*, *tincture*, que se referem a um *objeto* que é ao mesmo tempo um *artefato* feito de alguma *substância* que tem também a função de *atributo*.

No CoreLexII (BUIELAAR, 2000), as regras descritas acima são detalhadas, considerando os avanços de pesquisas posteriores (ex. KRYMOLOWSKI e ROTH, 1998; PETERS, PETERS e VOSSEN, 1998; TOMURO 1998). São propostos critérios mais livres para a identificação das classes polissêmicas. Essa nova proposta surgiu com o intuito de resolver críticas que o primeiro modelo sofreu, tais como: a escolha dos tipos básicos ser arbitrária e em nível muito alto na hierarquia, de modo que somente as classes polissêmicas que se manifestam em níveis altos podem ser identificadas; a arbitrariedade (e ineficiência) da intervenção humana para o agrupamento das classes resultantes a partir de critérios intuitivos de similaridade entre os membros (BUIELAAR, 2000).

Na nova abordagem, são comparados os *synsets* em todos os níveis, não somente no nível dos *unique beginners*. No CoreLexII, conforme descreve Buitelaar (2000), são executadas as seguintes etapas: (i) análise dos sentidos de cada nome na WN.Pr e na GN (nível1 *synsets*); (ii) se um nome tiver mais que um sentido, ele será incluído em uma lista separada, que será usada para próximos processamentos – nomes com somente um sentido são descartados; (iii) com o intuito de comparar os nomes em todos os níveis da hierarquia, também todos os hiperônimos de cada *synset* serão armazenados; (iv) a distribuição de cada sentido selecionado nas etapas anteriores é comparada automaticamente a partir da métrica de JACCARD, que compara objetos de acordo com os seus atributos individuais e os que ele compartilha; (v) se a similaridade for maior do que certo limiar, os pares de nomes são armazenados em uma matriz, que será

posteriormente utilizada; (vi) o agrupamento final se dá a partir de um algoritmo de relação simples e único que agrupa objetos em *clusters* discretos, ou seja, quebra-se a organização por *synset* (BUITELAAR, 2000).

O quadro 2 apresenta as classes de nomes regularmente polissêmicos geradas a partir do algoritmo proposto por Buitelaar (2000). No quadro, há o nome original da classe em inglês, seu respectivo equivalente no português e exemplos originais e possíveis ocorrências em PB, não necessariamente equivalentes às originais.

| |
|---|
| <p>1. <i>comunicação/barulho (communication/noise)</i>
 Exemplo original: <i>clamor, hiss, roaring, screeching, whisper.</i>
 Exemplo PB: ronco, berro, grito, murmúrio, sussurro</p> |
| <p>2. <i>gosto/sensação (taste_property/sensation)</i>
 Exemplo original: <i>acuity, aroma, odor, pungency</i>
 Exemplo PB: acidez, acridão, azedo, acre
 aroma, odor, fragrância, cheiro, perfume, sabor</p> |
| <p>3. <i>fio/tecido (cord/fabric)</i>
 Exemplo original: <i>chenille, lace, lanyard, ripcord</i>
 Exemplo em PB: chenile, renda, lã</p> |
| <p>4. <i>continente/unidade de medida (vessel/measure)</i>
 Exemplo original: <i>bottle, bucket, cask, tub, flask</i>
 Exemplo em PB: garrafa, copo, prato, xícara, colher</p> |
| <p>5. <i>planta/tipo de fruto (plant/berry)</i>
 Exemplo original: <i>blueberry, checkerberry, cranberry</i>
 *classe não-produtiva em PB</p> |
| <p>6. <i>planta/ nós (plant/nut)</i>
 Exemplo original: <i>hazelnut, pistachio, butternut</i>
 Exemplo em PB: pistache, avelã</p> |
| <p>7. <i>peixe/alimento (fish/food)</i>
 Exemplo original: <i>cappie, dolphinfish, flatfish</i>
 Exemplo em PB: carpa, tainha, sardinha</p> |

Quadro 2 - CoreLexII - Classes de polissemia regular

Conclui-se que o estudo das duas versões do CoreLex foi importante para investigarmos estratégias para identificação de classes polissêmicas automáticas e

semiautomáticas. A partir de Buitelaar (2000), tem-se a expansão da noção de polissemia regular para classes formadas por elementos situados em níveis distintos de uma hierarquia. Uma importante característica do modelo é que, ao contrário do que ocorre na GN, os *synsets* são quebrados, e passa-se a dar foco para a forma lexical regularmente polissêmica. As classes detectadas, conforme o quadro 2, revelaram-se úteis para descrever o português do Brasil, comprovando o que a literatura afirma (CROFT e CRUSE, 2004) sobre certa universalidade desse tipo de polissemia entre línguas.

2.7 SÍNTESE DA SEÇÃO

As investigações apresentadas nesta seção possibilitaram, de um lado, delinear o cenário linguístico-computacional que justifica a proposição de uma representação linguístico-computacional da polissemia de nominais e, de outro, identificar, a partir das propostas, parcialmente implementadas nas redes, lacunas na representação do fenômeno de forma abrangente e a dificuldade de inclusão da polissemia em uma base de dados organizada a partir de *synsets*, ou seja, seguindo um critério onomasiológico, que estrutura o léxico a partir de conceitos ou sentidos.

Do cenário linguístico-computacional, é importante destacarmos a importância da representação da polissemia para a realização de tarefas de PLN. A introdução de níveis distintos de generalidade entre os sentidos em uma base de dados pode reduzir a quantidade de processamento lexical, tendo em vista que o número de decisões pode restringir-se a situações com grau maior de antagonismo e, ao mesmo tempo, possibilita que a base forneça as informações necessárias a um sistema que necessite grau maior de detalhamento do sentido.

O estudo das *wordnets* possibilitou entender como se configura esse tipo de base relacional de dados lexicais, identificando possibilidades e limitações. Entre as possibilidades ou vantagens para a representação da polissemia, destacamos a organização hierárquica dos conceitos e a codificação de relações léxico-conceituais que podem ser úteis à delimitação de um sentido polissêmico. Como limitação evidente na estrutura padrão de uma *wordnet*, destacamos a organização de base onomasiológica,

que organiza os itens em termos de conjuntos de sinônimos (*synsets*/sentidos) e codifica relações entre esses conjuntos. Sabe-se, de antemão, que olhar para a polissemia pressupõe olhar para cada item lexical isoladamente e analisar os múltiplos sentidos associados a ele. As propostas já apresentadas para a representação da polissemia nas *wordnets* e no CoreLex são parciais, pois focam fundamentalmente na representação da polissemia sistemática. Entre as *wordnets*, fica evidente o não enfrentamento do problema, pois embora ele seja reconhecido e propostas tenham sido apresentadas, elas não foram implementadas por completo, pelo menos até onde tivemos notícia. Ficou evidente nesta seção que buscar uma estratégia representacional para a polissemia adequada ao contexto das *wordnets* e que habilite a base a fornecer informações úteis a sistemas computacionais com diferentes objetivos é tarefa necessária e altamente complexa.

SEÇÃO 3 - A REPRESENTAÇÃO DA POLISSEMIA SOB A ÓTICA DA SEMÂNTICA LEXICAL COGNITIVA

A partir do levantamento de propostas práticas realizado na Seção 2, ficou evidente que a proposição de uma descrição formalizável para a polissemia é uma demanda das bases relacionais de dados lexicais que pretendem servir a sistemas de PLN. Sem perder de vista essa necessidade, nesta seção, trataremos da polissemia sob o viés linguístico¹, tendo como objetivo responder à primeira fase do domínio linguístico da tese.

As questões levantadas na Seção 2 sobre a *definição* e a *representação* da polissemia têm sua contraparte na Linguística, conforme veremos aqui. Muito já se falou sobre polissemia nesse contexto, e unanimidade não faz parte das investigações sobre o tema. Não temos de forma alguma a pretensão de resolver as questões que a literatura apresenta, mas sim de propor uma metodologia para a descrição da polissemia dos nominais do PB que seja útil à explicitação das principais relações que se estabelecem entre os diferentes sentidos de nominais polissêmicos.

Não há como definir polissemia sem um comprometimento com uma perspectiva teórica. Apresentamos, na Seção 2, uma primeira caracterização do fenômeno. Nessa definição, ficou expressa a posição clássica adotada na literatura: o olhar para os múltiplos sentidos como um fenômeno exclusivamente da superfície da língua. Nesta seção, essa noção será revista e ampliada a partir do paradigma proposto pela Semântica Lexical Cognitiva (doravante, SLC).

Para a caracterização da polissemia de acordo com o paradigma cognitivo da semântica, refletiremos sobre os seguintes tópicos: em 3.1, *que desafios a polissemia apresenta a uma teoria linguística*; em 3.2, *por que escolher o olhar da semântica lexical cognitiva*; em 3.3, *a caracterização da polissemia como fenômeno categorial*; em 3.4, *aspectos estruturais e funcionais da dimensão semasiológica do léxico*; em 3.5, *as abordagens homonímica polissêmica e monossêmica*; em 3.6, *as redes de polissemia*.

¹ Não estamos ainda aqui estabelecendo distinção entre os domínios linguístico e cognitivo. Quando essa distinção for necessária, chamaremos a atenção.

3.1 QUE DESAFIOS A POLISSEMIA APRESENTA A UMA TEORIA LINGUÍSTICA?

Há tempos a polissemia deixou de ser vista como possível origem de desentendimento entre falantes de uma língua. Não há dúvidas de que nós, humanos, ao contrário do que acontece com as máquinas, sabemos lidar bem com os múltiplos sentidos de uma palavra em situações do cotidiano. Os problemas surgem quando se pretende delimitar teoricamente o fenômeno.

Dizer o que é polissemia exige dizer o que ela não é, pois há fenômenos correlatos que estão intimamente ligados a ela, de modo que não se podem, muitas vezes, desenhar limites precisos entre eles. Por essa razão, temos apresentado nesta tese considerações também sobre homonímia e monossemia, olhando, portanto, para a multiplicidade de sentidos como um todo. Na Seção 2, apresentamos uma caracterização dos tipos de multiplicidade de sentidos sob a visão de Apresjan (1973). Tomamos essa caracterização como uma primeira definição, por ser ela representativa do que comumente se entende por polissemia (regular e irregular) e fenômenos correlatos (homonímia e monossemia). Embora aparentemente simples e clara, ela suscita uma série de questões conceituais e metodológicas, de modo que as dificuldades de definição e descrições enfrentadas pelo PLN na construção de bases relacionais de dados lexicais têm sua contraparte na linguística.

Atrelada à tarefa de definição de um fenômeno linguístico, está a sua nomeação. Também no domínio linguístico desta tese, seguiremos as opções de nomeação para os tipos de multiplicidade de sentidos apresentados em Apresjan. A escolha se deve ao fato de que, embora o *status* do fenômeno seja limitado nessa visão, as denominações empregadas pelo autor são altamente frequentes tanto em PLN quanto em Linguística. Vale mencionar, no entanto, que encontramos na literatura, de maneira imprecisa e sobreposta, uma variedade de outras denominações. Com o objetivo de organizar esse emaranhado de termos e noções, apresentaremos, a partir de Alves (2002), uma sistematização das principais nomeações encontradas. Para esse inventário de tipos, levaremos em consideração a relação entre a ocorrência do fenômeno na língua (exemplo) e o termo tipicamente utilizado para nomeá-lo, sem considerar as diferenças teóricas implicadas.

- (a) Para a **polissemia irregular** (cf.: APRESJAN, 1973; KILLGARRIFF, 1998), tal como ocorre em *fonte* (1. mecanismo hidráulico que fornece água e 2. fonte de alimentação, dispositivo eletrônico que fornece energia elétrica a um aparelho), destacamos os seguintes equivalentes: *polissemia assistemática* (ex.: BUITELAAR, 1998; CRUSE, 2000; COPESTAKE e BRISCOE, 1995), *polissemia convencional* ou *polissemia idiossincrática* (ex. COPESTAKE e BRISCOE, 1995), *polissemia metafórica* (BLANK, 2003) ou simplesmente *polissemia* quando não é relevante a distinção entre regular e irregular (ex.: TUGGY, 2003; BLANK, 2003; ZLATEV, 2003; GEERAERTS, 2003; 2006; RICE, 2003; TAYLOR, 2003; 2006; CROFT e CRUSE, 2004; FAUCONIER e TURNER, 2006; LEHRER, 2003; SOARES DA SILVA, 2006).
- (b) Para a **polissemia regular** (cf.: APRESJAN, 1973; KILLGARRIFF, 1998), tal como ocorre em *livro* (1. Objeto físico e 2. conteúdo), destacamos os seguintes equivalentes: *polissemia lógica* ou *polissemia complementar* (ex.: WEINREICH, 1964; PUSTEJOVSKY, 1995), *polissemia semiprodutiva* (ex.: COPESTAKE e BRISCOE, 1995). Como subtipos, destacamos: *polissemia metonímica*, *polissemia taxonômica*, *polissemia por co-hiponímia* (ex.: BLANK, 2003), *autohiponímia*, *autossuperordenação*, *automeronímia* e *autoholonímia* (ex.: CRUSE, 2000).
- (c) Para a **homonímia** (ex.: ULLMAN, 1962; LYONS, 1977), como ocorre em *fonte* (1. mecanismo hidráulico e 3. fonte tipográfica, tipos de letras), há poucas variações, podendo ser encontrado o termo *polissemia contrastiva* (ex.: WEINREICH 1964; PUSTEJOVSKY 1995). Relacionada à noção de homônimos, encontramos *homógrafos* (itens com a mesma forma) e *homófonos* (itens com o mesmo som).
- (d) Para **monossemia** (termo recorrente, ex. JANSSEN, 2003; SOARES DA SILVA, 2006; RICE, 2003; GEERAERTS, 2003, 2006; TAYLOR, 2003, 2006), como ocorre em *meia-idade* (época da vida entre a maturidade e a velhice), destacamos os seguintes equivalentes: *modulação contextual* (CRUSE, 1986; CROFT e CRUSE, 2004), *vagueza* ou *vaguidade*, originalmente, *vagueness* (ex.: KILLGARRIFF, 1992; CRUSE, 2000,

GEERAERTS, 2003; TUGGY, 2003, 2006; SOARES DA SILVA, 2006), *variação contextual* (BLANK, 2003), *generalidade* (ZLAVEV, 2003).

Essa pluralidade de termos e noções ilustra apenas uma pequena parcela das questões a serem resolvidas ao estudarmos polissemia (sem dúvida, a mais simples delas). Ultrapassada essa etapa, para delimitar o fenômeno, é preciso que se tenha ideia clara sobre que tipo de entidade é o *sentido (significado)*² lexical. Além disso, precisamos de procedimentos confiáveis para identificar tais entidades e de critérios para determinar se e de que forma essas entidades, uma vez identificadas, estão relacionadas (cf. TAYLOR, 2003a). A partir de então, uma série de questões teórico-metodológicas se apresentam (cf. TAYLOR, 2002, 2003a, 2003b; GEERAERTS, 2006; SOARES DA SILVA, 2006).

Conforme já mencionamos, o estudo da polissemia está intimamente ligado ao estudo da homonímia e da monossemia; não há uma fronteira bem delimitada entre esses três fenômenos. Assim, os problemas de definição, ou demarcação, remetem às distinções entre polissemia e monossemia, de um lado, e entre polissemia e homonímia, de outro.

Distinguir polissemia de monossemia implica reconhecer a existência de diferentes sentidos ou de simples modulação contextual. Sabe-se que a polissemia, assim como a homonímia, é um fenômeno que pode causar *ambiguidade lexical*, ou seja, mais de uma possibilidade de interpretação de um enunciado em razão da presença de pelo menos uma palavra associada a mais de um sentido. Embora essa distinção seja relativamente clara e possa parecer simples de ser aplicada, diversas questões surgem ao tentarmos demarcar os limites entre um único sentido e mais de um. Do ponto de vista teórico, surgem questões como: *qual é a natureza do sentido dos itens lexicais? Qual é o nível de generalidade ideal de identificação dos sentidos? Os sentidos são entidades fixas e delimitáveis? Se não, poderemos manter a distinção entre polissemia e monossemia? Ou, no caso de uma palavra presumivelmente polissêmica, será alguma vez possível determinar com precisão quantos sentidos diferentes tem essa palavra?* (cf. TAYLOR, 2002, 2003a, 2003b; GEERAERTS, 2006; SOARES DA SILVA, 2006). Do ponto de vista metodológico, destacamos as seguintes questões: *como os sentidos podem ser*

² Também nas seções dedicadas ao domínio linguístico desta tese, os termos *sentido* e *significado* são empregados aqui de maneira intercambiável.

identificados? Como eles podem ser descritos? De que maneira eles podem estar relacionados? Qual é realidade psicológica do constructo proposto para descrição? Que tipo de unidade linguística é candidata à análise polissêmica? (cf. SOARES DA SILVA, 2006). Diferentes estratégias têm sido apresentadas na literatura em busca de soluções e diagnósticos. Mesmo antes de apresentarmos qualquer investigação sobre o tema, é importante anteciparmos que é consensual o fato de que “uma distinção exata entre polissemia e monossema é frequentemente difícil de se estabelecer” (TAYLOR, 2002, p. 470).

A demarcação entre polissemia e homonímia é desenhada em termos da existência de algum tipo de *relação* entre os diferentes sentidos associados a uma mesma forma (polissemia) ou da inexistência de qualquer relação (homonímia). Detectada a homonímia, a posição comum dos teóricos é a de dizer que se trata de dois itens lexicais diferentes que, por acaso, compartilham a mesma forma. Na tentativa de delinear essa suposta dicotomia, assim como no caso anterior, questões teóricas e metodológicas se apresentam. Do ponto de vista teórico, é preciso reflexão sobre: “o que significa dizer que há relação entre sentidos de uma palavra?”; “qual é a natureza dessa relação? Histórica e, portanto, apenas do conhecimento dos filólogos e das pessoas com formação filológica; ou então sincrônica e, assim, do conhecimento de todos os falantes? Se é sincrônica, será psicologicamente real e, principalmente, intersubjetivamente compartilhada?” (SOARES DA SILVA, 2006, p.36). Do ponto de vista metodológico, é preciso definir: “que critérios utilizar para determinar a existência de relação”; “que critérios utilizar para identificar o tipo de relação”, “que estratégia utilizar para representar essa relação” (SOARES DA SILVA, 2006, p.36). Também nesse caso, tem-se uma certeza – conforme Lyons (1977), mas também conforme Geeraerts (2003, 2006), Tuggy (1993, 2006), Croft e Cruse (2004), Soares da Silva (2006) e Taylor (2006), entre outros: a relação entre os sentidos é um fenômeno gradual, e a percepção da relação frequentemente varia de um falante para outro. Portanto, não há uma dicotomia, mas a distinção se dá em um *continuum*. Para o português, um exemplo recorrente que evidencia esse fato é o do item *porto* – ‘local para chegada e abrigo de embarcações’, ‘tipo de vinho’ e, ainda, o ‘nome da cidade’, que se associa ao fato de haver um porto (exemplo de Lyons 1977, reportado por Geeraerts, 2006; Soares da Silva, 2006). Há evidências de que historicamente os sentidos de *porto* estão relacionados; entretanto, essa relação não necessariamente é percebida por todos os falantes do português e, caso

seja, é bastante improvável que seja acessada no momento do uso (GEERAERTS, 2006; SOARES DA SILVA, 2006).

O critério da *relação* é empregado tradicionalmente para reconhecer a polissemia tanto na perspectiva diacrônica quanto na sincrônica: na primeira busca-se pela relação de étimo; no segundo, de sentido. Ao contrário do que ocorre na distinção sincrônica, diacronicamente, a diferenciação é uma questão de *sim/não*, embora na prática nem sempre isso aconteça em razão de falta do histórico sobre muitas palavras (CROFT e CRUSE, 2004). Ainda que a perspectiva sincrônica seja menos firmemente embasada por ser uma questão de gradação, é esse o critério que poderá vir a ser útil a um falante que não conhece o sentido de um item e, de maneira análoga, é esse o critério útil à construção de um léxico que vise a ampliar a habilidade de realização de inferências de um sistema de PLN (*cf.* PUSTEJOVSKY, 1995).

É importante mencionar que a falta de dicotomia entre polissemia/monossemia/homonímia não invalida a importância de reconhecermos particularidades de cada fenômeno, a fim de, para mencionar o nosso interesse, avaliar e propor estratégias distintas para a descrição da semântica associada a itens homônimos, polissêmicos ou monossêmicos. Conforme salienta Taylor (2003, p. 645), não há motivos para acreditar na existência de dicotomia entre os referidos fenômenos, e detalharemos ao longo desta Seção 3 os motivos.

As questões apresentadas acima representam os principais problemas implicados na tarefa de delimitação teórica da polissemia e sintetizam os tópicos que abordaremos nas seções seguintes. Vimos, na seção anterior, que a prática comum de listar os diferentes sentidos associados às formas lexicais não é a maneira mais eficiente de armazenamento para um sistema de PLN. Nesta seção, veremos que, sob o viés da SLC, há evidências de que essa não é a estratégia cognitivamente real e, a partir daí, proporemos uma alternativa ao problema.

3.2 POR QUE ESCOLHER O OLHAR DA SEMÂNTICA LEXICAL COGNITIVA?

Para justificarmos a opção pela Semântica Lexical Cognitiva (doravante, SLC), é importante que voltemos um pouco a momentos pontuais que marcaram a história da

polissemia. Vale, aqui, chamar a atenção para o fato de que essa história reflete a própria trajetória de abandono e encanto pela qual a semântica lexical passou até se tornar o centro dos estudos linguísticos. Lembraremos fatos marcantes do desenvolvimento do estudo da polissemia, com o objetivo de registrarmos estudos importantes. Não nos dedicaremos a apresentar nenhuma das propostas particulares nem princípios teóricos de fundo até chegarmos à SLC. Esta seção foi organizada em dois momentos: em 3.2.1, traremos uma visão panorâmica e, em 3.2.2, focaremos nas contribuições da SLC para o estudo dos sentidos.

3.2.1 Visão Panorâmica

A literatura mostra que as raízes das investigações sobre a polissemia remetem pelo menos aos Estóicos e a Aristóteles; entretanto, o termo moderno *polissemia* (originalmente, *polysémie*) foi popularizado somente em 1897, pelo francês Michael Bréal, com o livro *Essai de Sémantique: Science des significations*. Foi a partir de Bréal que a polissemia deixou de ser estudada somente atrelada à Etimologia e à Lexicografia e passou a ser vista como um fenômeno de uso, da aquisição, de mudança da língua e como consequência de processos cognitivos. Bréal entendeu que a polissemia representa o fenômeno que melhor responde às necessidades cognitivas e sociais dos falantes, e é a partir do uso que eles fazem da língua, da relação com a cognição e com a cultura que, de acordo com o autor, ela deve ser estudada. Nesse cenário, a polissemia é tida como um efeito sincrônico da mudança da semântica lexical. Nas palavras de Blank (2003), ela é entendida como o seu “lado sincrônico” (*cf.* BLANK, 2003).

Os estudos de Bréal, além de terem sido responsáveis pela popularização do estudo da polissemia (e da semântica) no contexto geral da linguística, merecem destaque aqui por representarem o embrião da SLC, visto que seguem uma concepção mentalista do significado lexical, acreditando que o significado das unidades linguísticas se associa por correspondência a itens do conteúdo mental (ideias e conceitos).

Tomando como critério a natureza da relação semântica que se estabelece entre um sentido antigo e um sentido recente, Bréal reconhece tipos distintos de polissemia, estudo que tem sido revisitado em trabalhos recentes no contexto da SLC (*ex.* BLANK,

2003). Para fins de ilustração desse tipo de olhar, embora a apresentação de propostas particulares não seja nosso objetivo aqui, abro parênteses para trazer os “novos” tipos de polissemia inventariados por Blank (2003) em razão de sua riqueza descritiva. Nesse inventário, há aleatoriamente exemplos para o inglês, o italiano, o latim, o espanhol, o grego e o português. Apresentaremos os tipos selecionados pelo autor com os exemplos originais: *polissemia metafórica* (ing. - *mouse* ‘pequeno roedor’ ou ‘acessório de computador’); *polissemia por co-hiponímia* (port. - *aborrecer* ‘causar tédio em alguém’ ou ‘perturbar alguém’), *polissemia taxonômica* (esp. - *hombre* ‘ser humano’ e ‘ser humano do sexo masculino’), *polissemia metonímia* (lat. - *língua* ‘tongue’, ‘language’), *polissemia por autoconversão* (ital. - *noleggiare* ‘emprestar algo para alguém’ e ‘tomar algo emprestado de alguém’), *polissemia antifrástica* (ital. - *brava donna* ‘senhora admirável’ e ‘prostituta’, em uso informal), *polissemia autoantonímica* (ing. - *bad* ‘o que não é bom’ e ‘excelente’, em uso informal). A tipologia de Blank é proposta por analogia a uma tipologia da mudança lexical. Porém, fica claro no artigo que a polissemia, embora possa ser estudada em comparação com a sua motivação (diacrônica), é um fenômeno sincrônico da língua.

Voltando à nossa retrospectiva para justificar a escolha da SLC, vemos que o estudo da polissemia, reconhecida pela tradição histórico-filológica, passou por um longo “período de latência” na Linguística sincrônica do século XX. Durante esse período, o que ficou marcante na maior parte dos estudos – tanto entre estruturalistas quanto entre gerativistas – foi a minimização do fenômeno e até seu abandono devido ao fato de que a complexa e criativa relação entre os itens lexicais e seus múltiplos sentidos desafiava princípios considerados científicos. Nas palavras de Soares da Silva (2006, p. 3):

A polissemia é menosprezada e negada em nome do ideal semiótico ‘uma forma, um significado’, da pretensa existência de significados ‘invariantes’ ou ‘significados fundamentais’ abstractos e unívocos, da proclamadíssima tese da autonomia da linguagem (*sistema autônomo*, que se basta a si mesmo, ou *faculdade autônoma* inata e independente de outras faculdades mentais), da idéia de língua e de gramática como sistema de regras formais, dos requisitos científicos de formalização e economia, ou ainda de famosas dicotomias como ‘conhecimento linguístico’ vs. ‘conhecimento enciclopédico, gramática (regras) vs. ‘parole’(uso) ou ‘competência’ (conhecimento individual do sistema) vs. ‘performance’ (uso) – dicotomias com as quais, invariavelmente, se quis secundarizar ou mesmo desprezar o seu segundo elemento.

Observa-se, no entanto, que esse esquecimento não foi total, conforme podemos ver, por exemplo, nos clássicos estudos de Ullman (1962), Weinreich (1964) e Apresjan (1973). Em Ullman (1962), assim como em Bréal, princípios mentalistas do significado estão presentes. Encontramos no clássico *Semantics: an introduction to the science of meaning*, uma reflexão sobre o estado da arte da semântica até o início do século XX, e a polissemia tem seu lugar reconhecido, sendo apresentados diferentes tipos de extensões semânticas como suas causas (ex. especialização, generalização, analogia, sinédoque, metáfora, metonímia, etc.). Weinreich (1964), por discordar do tratamento que lexicógrafos estavam dando aos sentidos, descrevendo-os como unidades discretas agrupadas arbitrariamente, sugere uma nova organização para as entradas lexicais a partir das noções de “ambiguidade contrastiva” e “ambiguidade complementar”³ (cf. PUSTEJOVSKY, 1995). A *ambiguidade contrastiva* de Weinreich é aquele tipo de multiplicidade de sentidos a que nos referimos como *homonímia* (ex. *manga* como ‘fruta’ ou ‘parte de uma roupa’); e a *ambiguidade complementar* é a que denominamos *polissemia regular* (ex. *livro*, como ‘conteúdo’ ou ‘objeto físico’), que envolve sentidos lexicais que são manifestações de um mesmo sentido básico.

O resultado do trabalho de Apresjan (1973)⁴ já foi mencionado na Seção 2, visto que, dada a precisão e objetividade de suas definições, tornou-se obra de referência para o PLN. A posição do autor é semelhante à de Weinreich, discordando da postura adotada pelos lexicógrafos da época, que estabelecem fronteiras consideradas excessivamente precisas entre os sentidos. É com o intuito de mostrar a similaridade interna da *polissemia lexical*, ou seja, de mostrar que as relações semânticas entre os sentidos de um item lexical polissêmico podem ser similares e tão recorrentes quanto as relações semânticas entre itens lexicais distintos, que o autor realiza o estudo apresentado em “Regular Polysemy” (APRESJAN, 1973).

Em fins dos anos 80 e início dos 90, os olhares desinteressados ou céticos em relação à descrição da polissemia perdem força, e o tema volta a ser o centro em razão de avanços de estudos psicológicos e computacionais. É iniciado um movimento de “(re)polissemização”⁵, que se dá em duas vias: PLN e Linguística Cognitiva. No PLN, estudos como os de Pustejovsky (1995), Kilgarriff (1992) e Buitelaar (1998), entre

³ O termo *ambiguidade* está sendo usado como equivalente a múltiplos sentidos. De acordo com Weinreich, um item é ambíguo se puder ser associado a mais de um sentido.

⁴ Originalmente publicado em russo em 1971.

⁵ Termo empregado em Soares da Silva (2006).

vários outros reportados, por exemplo, em Ravin e Leacock (2000), marcaram o desejo (e a necessidade) que engenheiros da linguagem⁶ têm de manipular os sentidos associados aos itens lexicais de modo a possibilitar que sistemas computacionais realizem inferências sobre os textos, simulando o que os seres humanos fazem para interpretá-los – posição que destacamos na Seção 2.

Foi seguindo a tendência de estudos aplicados à computação que Pustejovsky (1995) desenvolveu a Teoria do Léxico Gerativo (TLG), que tem o tratamento da polissemia regular como um de seus focos principais de atenção. A TGL busca tratar da semântica lexical de maneira útil à computação, porém sem repetir o erro de trabalhos anteriores, ou seja, sem tratar o léxico como um conjunto de itens lexicais com sentidos estáticos e descritos a partir de traços sintáticos, morfológicos e semânticos fixos. Seu objetivo foi propor uma descrição robusta o suficiente para dar conta do (i) uso criativo dos itens em novos contextos e (ii) da avaliação de modelos de semântica lexical com base na noção de composicionalidade.

De acordo com Geeraerts (2006), Pustejovsky deu um novo ímpeto ao ideal katziano de construção de uma representação semântica formalizada com base na lógica e não na análise componencial. Ao mesmo tempo, ele vai além da abordagem katziana por enfatizar a necessidade de construir um léxico que seja gerativo – no sentido de que ele não só contenha uma enumeração de sentidos, mas que ele incorpore a possibilidade de derivar formalmente novos sentidos a partir daqueles já enumerados. Sendo assim, a abordagem de Pustejovsky é classificada por Geeraerts como “neogerativista”, uma vez que, ainda dentro de uma concepção de cognição modular, desenvolve o ideal katziano de representação formal da semântica a partir da introdução de flexibilidade e formalismo lógico.

A delimitação dos tipos de multiplicidade de sentidos na TLG segue a proposta de Weinreich (1964). Para itens homônimos, o autor sugere que os sentidos sejam representados em entradas distintas no léxico, seguindo o que é regra geral nos estudos. Para itens que apresentam polissemia regular, diferentemente, os sentidos devem ser descritos em uma só entrada subespecificada a partir do seu modelo em quatro níveis

⁶ O termo *engenheiros da linguagem* está sendo usado aqui em sentido informal e neutro para denotar aquele que manipula a linguagem. Não pretendemos fazer referência à dicotomia entre aqueles profissionais dedicados à Engenharia da Linguagem como sendo aqueles que tratam da língua de maneira mais prática e ao Processamento de Língua Natural como sendo aqueles dedicados ao estudo mais teórico, conforme apresentam Nirenburg e Haskin (1996).

que constituem a denominada *estrutura qualia*, capaz de descrever, conforme salienta Chishman (2000, p. 57), “o poder relacional de um item lexical” e de prever estratégias de inferência para a ativação dos diferentes sentidos a partir de mecanismos gerativos.

Devido a sua robusta operacionalidade, a TLG teve grande aceitação da comunidade de PLN. Muitos são os trabalhos que utilizam o modelo de modo que cabe aqui uma justificativa sobre por que não adotá-lo. Enumeraremos três motivos. O primeiro deve-se ao fato de que muito já se falou sobre a teoria, seja em investigações teóricas, seja em investigações aplicadas à computação de modo que são amplamente conhecidas suas características e limitações; sendo assim, o motivo se apresenta: não restam muitas dúvidas sobre seu poder descritivo ou sobre sua validade para fins computacionais. O segundo, como consequência desse primeiro motivo, deve-se ao fato de a validade da descrição proposta pelo modelo ser restrita à polissemia regular, ficando excluídos os tipos de polissemia irregular, fato que impossibilitaria a utilização da TLG como modelo único para darmos conta da demanda que nos propusemos. Por fim, destacamos o terceiro motivo, que se deve ao fato de que, embora a TLG esteja tão interessada na polissemia quanto à SLC, sua abordagem reluta em aderir ao que é um dos princípios de base da SLC: a natureza enciclopédica do significado lexical (*cf.* GEERAERTS, 2006). A ideia de que é altamente útil para a análise automática da língua natural aderir a uma distinção estrita entre os níveis linguístico do significado *stricto sensu* e o nível de conhecimento de mundo é, de acordo com Geeraerts (2006), o motivo para Pustejovsky manter sua separação, aproximando-o, de certo modo, da semântica estrutural clássica.

Chegamos, então, aos estudos relacionados à área da Linguística Cognitiva. Foi com a intenção de aprofundar o entendimento sobre questões como *o que é a faculdade mental? como geramos sentido a partir de nossa experiência? o que é um sistema conceitual e como ele é organizado? todas as pessoas usam o mesmo sistema conceitual? em caso afirmativo, que sistema é esse? em caso negativo, exatamente o que há em comum à forma como os seres humanos pensam? qual é a relação entre língua e cognição? em que medida modelos cognitivos explicam o uso que fazemos da língua e o que vemos expresso nela?* que Lakoff (1987) publica “Women, Fire and Dangerous Thing”, obra que, juntamente com Lakoff e Johnson (1980), Talmy (1985; 2000) e Langacker (1987), se tornou referência para a inauguração de um novo paradigma na Linguística, que trouxe a polissemia de volta ao cerne das investigações: a Linguística Cognitiva (LC).

Opondo-se à concepção de que *língua, cognição e cultura* funcionam de forma independente, surge a LC. Integrando conhecimentos adquiridos em áreas como Psicologia, Linguística, Antropologia, Filosofia e Ciências da Computação, a LC estuda a língua de maneira compatível com o que é conhecido sobre a mente humana (LAKOFF, 1987). A língua é caracterizada, nesse contexto, ao mesmo tempo como uma criação da cognição humana e como um instrumento ao seu serviço (TAYLOR, 1995). Lakoff (1987) chama a atenção para o fato de que as questões que elencamos acima não são novas, mas que algumas das respostas dadas pela LC são. Esse novo paradigma cognitivo da linguística, um *pool* de teorias, tem o estudo do significado como centro e, portanto, a SLC como um dos seus principais ramos de pesquisa; além dela, registram-se subteorias, como a Gramática Cognitiva⁷ e a Gramática de Construções⁸. Destacamos que o termo *gramática*, empregado para denominar esses dois ramos de estudo, não é usado em seu sentido estrito, focando em sintaxe e morfologia, mas em seu sentido amplo, como um inventário estruturado de unidades linguísticas convencionais (LANGACKER, 1987). Isso se deve ao fato de que, para a LC, como um todo, léxico e sintaxe (também fonologia, morfologia, entonação, etc.) organizam-se em um *continuum*, sendo essas porções da língua que podem ser descritas em termos de sua natureza simbólica; “a diferença entre essas unidades é uma questão de graus” (LANGACKER, 2000, p.18). Nesse contexto, o *léxico* é entendido como um “repositório de conhecimento de mundo, uma coleção

⁷ Mantendo o foco no papel do significado, a Gramática Cognitiva, proposta por Langacker (1987), configura-se como uma teoria específica da linguagem, com terminologia e técnicas descritivas particulares, mas que segue os mesmos princípios teóricos da LC para delinear os pressupostos relacionados a aspectos gerais da cognição que servem à estrutura de uma gramática (TAYLOR, 2003; EVANS e GREEN, 2006). Langacker destaca o fato de que o léxico não é um repositório arbitrário de itens sem relação, pelo contrário: o léxico constitui uma rede elaborada de associação entre forma e sentido (LANGACKER, 1987; 1991) em que cada forma é associada a uma rede semântica ou a um *continuum* (BRISARD, 1997).

⁸ A Gramática de Construções foi proposta mais recentemente, por Goldberg (1995), Kay e Fillmore (1999), Croft (2002), entre outros. Cada autor representa uma das vertentes da teoria que apresentam certas particularidades, sem manter a unidade de acordo com os parâmetros fundantes da LC. O que é particular à Gramática de Construções como um todo é o fato de que ela toma como unidade básica de análise o que tem sido denominado *construção*, de acordo com Goldberg (2006), um agrupamento de forma e significado. Dependendo da vertente, uma construção é somente composta por dois ou mais itens lexicais, por exemplo, relação entre verbo e seus complementos; entretanto, um item lexical também pode contar como uma construção, embora não seja a unidade mínima preferencial (cf. GOLDBERG, 1995). Assim, a polissemia lexical é minimizada em detrimento da polissemia construcional: se a polissemia lexical é resolvida quando o item participa de uma construção, diz-se que aquele item não é polissêmico (BROCCIAS, 2006). Embora o foco sejam as construções, é importante destacarmos das considerações de Goldberg (1995) sua reflexão sobre a estrutura de uma rede baseada em relações de polissemia, nos termos da autora.

estruturada de categorias significativas que nos ajuda a lidar com novas experiências e armazenar informações sobre as antigas” (GEERAETS, 1995, p.112-113).

Seguindo o que propõem fundamentalmente Lakoff e Langacker, apresentaremos noções centrais à LC que se relacionam diretamente com o estudo da polissemia no contexto da SLC. Entre suas principais contribuições, destacamos os seguintes postulados: (i) a língua não é uma faculdade cognitiva autônoma; (ii) o conhecimento da língua emerge do uso; (iii) gramática é conceitualização.

O **primeiro postulado** marca a oposição que a LC estabelece com a gramática gerativa em relação ao princípio de que a língua é inata, autônoma e modular, separando habilidades cognitivas linguísticas e não-linguísticas. A partir daí, deduz-se (cf. CROFT e CRUSE, 2004) que conhecimento e uso da língua (sentidos e formas) – em particular, a construção e a comunicação de significado – são estruturas e rotinas conceituais com características próprias, porém regidas por princípios cognitivos gerais que também orientam habilidades como memorização, percepção, atenção e categorização. Sendo assim, organização e recuperação do conhecimento linguístico não são procedimentos significativamente diferentes de organização e recuperação de outros conhecimentos na mente. Podemos dizer que a LC entende que a língua oferece uma “janela” a funções cognitivas gerais, possibilitando descobertas sobre a natureza, a estrutura e a organização de pensamentos e ideias que nascem nas e das nossas vivências e experiências (EVANS e GREEN, 2006).

Modelos psicológicos da memória têm inspirado modelos linguísticos de organização do conhecimento através de, por exemplo, *frames*/domínios (corpo de conhecimento que faz parte do nosso sistema conceitual que armazena e organiza ideias relacionadas e experiências)⁹ e o conhecimento gramatical¹⁰ em redes estruturadas a

⁹ De acordo com a Semântica de *Frames*, de Charles Fillmore, o *frame* pode ser definido em linhas gerais como uma estrutura cognitiva que pode ser definida como a base para uma interação significativa em dada comunidade de fala (SUBIRATS e SATO, 2004; FILLMORE, JOHNSON e PETRUCK, 2003). Em uma descrição do significado a partir de *frames* semânticos, seguindo Fillmore, entende-se que cada item lexical em contexto evoca um *frame* semântico particular e caracteriza algum elemento do *frame*, que podem ser padrões de comportamento, práticas, instituições, imagens ou algum outro aspecto do *frame* em questão. Um exemplo extraído da base eletrônica FrameNet (construída de acordo com essa perspectiva) e adaptado para o português seria a descrição do conteúdo semântico associado a *livro* a partir da identificação dos *frames* por ele evocados, por exemplo, TEXTO e PARTE_ORDENADA_POR_SEGMENTOS. No primeiro exemplo, o *frame* TEXTO pode ser definido como uma entidade que contém informação linguística sobre um TÓPICO, criado por um AUTOR em algum TEMPO. Infere-se, então, que os elementos TEXTO, TÓPICO, AUTOR e TEMPO são participantes de um dos *frames* evocados por livro.

¹⁰ Seguimos usando a concepção ampla do termo *gramática*, tal como definida anteriormente.

partir de relações taxonômicas, entre outras. Modelos psicológicos de atenção e percepção têm explicado muitos processos de conceitualização em semântica, isto é, têm explicado processos cognitivos desencadeados para a construção do significado que se dá a partir da experiência corpórea. Modelos psicológicos de *categorização*, em particular, *protótipos*, *centralidade gradual*¹¹ e *modelos de estrutura de categoria em forma de redes* (ex. radial, esquemática, multidimensional), têm sido os que mais influenciaram os estudos semânticos (ex. LAKOFF, 1987; TAYLOR 1989). Antes de seguirmos, é importante destacar que essa posição representa uma negação da tese da autonomia, e não da tese da capacidade humana inata para o uso da língua (cf. CROFT e CRUSE, 2004).

O **segundo postulado** segue na direção contrária à de tendências reducionistas da gramática gerativa e da semântica de condições de verdade, que acreditam que o conhecimento sobre a língua existe de modo independente de sua experiência. A crença de que o conhecimento da língua emerge do seu uso relaciona-se à ideia de que as propriedades de certas entidades linguísticas são consequência da natureza das capacidades biológicas humanas e da sua experiência de funcionamento em um ambiente físico e social (LAKOFF, 1987). Isto é, entidades e propriedades semânticas, sintáticas, morfológicas, fonológicas, de entonação, entre outras, são construídas a partir da nossa percepção de enunciados específicos proferidos em situações específicas – por isso, organizam-se em termos de um *continuum*. Esse processo indutivo de abstração e esquematização leva em conta as convenções em relação ao significado e às diferenças encontradas entre construções sintáticas que se realizam em situações particulares. Entende-se que o mundo exterior provê o material bruto para o sistema conceitual, que o processa dinamicamente (CROFT e CRUSE, 2004).

O **terceiro postulado** remete ao *slogan* de Langacker – “grammar is conceptualization” –, referindo-se a uma concepção corporificada (do inglês, *embodied*) da língua que se opõe à semântica de condições de verdade em razão da impossibilidade de a estrutura conceitual ser reduzida a uma simples correspondência de verdade ou falsidade em relação ao mundo exterior, embora essa relação também deva ser levada em conta. O estudo da língua nessa perspectiva pode ser caracterizado como o estudo

¹¹ Centralidade: ideia de que alguns membros de uma categoria podem ser “exemplos melhores” dessa categoria do que outros. Centralidade gradual: ideia de que os membros (ou subcategorias) que se encontram claramente dentro dos limites da categoria ainda podem ser mais ou menos centrais (LAKOFF, 1987, p. 12).

dos padrões de *conceitualização*. O foco de atenção do linguista cognitivista é, fundamentalmente, o processo da “categorização linguística dos objetos da língua” (nomes, verbos, fonemas, sentenças) (*cf.* TAYLOR, 1995, p. ix). Nesse contexto, entende-se que as *unidades linguísticas* são porções da língua (itens lexicais/palavras, morfemas, sentenças, orações, expressões fixas) compostas pela associação simbólica entre *forma* (expressão material) e *conteúdo* (sentido/significado) reconhecida pelos falantes a partir de convenção; dito de outro modo, são agrupamentos de forma (faladas, escritas) e conteúdo semântico convencional (EVENS e GREEN, 2006). É a frequência da associação entre uma mesma forma e um mesmo conteúdo conceitual que faz com que tenhamos a sensação de haver uma relação de posse entre um item e determinado sentido – posição negada em SLC. Os itens da língua, portanto, não referem diretamente elementos do “mundo real”, mas simbolizam um inventário de conteúdo conceitual (estrutura semântica) a eles associada. Essas estruturas mentais estão recheadas por elementos de diferentes níveis de generalidade (ex.: conceitos, esquemas, *frames*, etc.) que interpretam o mundo a partir de padrões de experiência e da percepção humana (LANGACKER, 1987; 1991). Não se pode, no entanto, pensar em igualdade entre sistema linguístico e sistema conceitual, visto que a estrutura semântica (sentidos/significados/conceitos lexicais) associada às unidades da língua representa apenas um subgrupo de possíveis estruturas conceituais acessíveis via léxico (EVANS e GREEN, 2006).

Considerando o que apresentamos até este momento, já é possível entendermos parcialmente em que medida é o paradigma cognitivo da semântica lexical o mais útil à descrição da semântica de nominais polissêmicos. Vimos que foi a LC, a partir da década de 80, que, percebendo que ainda havia muito por se descobrir sobre a relação entre significado lexical e mente, retomou os problemas levantados por Bréal (1992), dando novo ânimo aos estudos da polissemia. É a partir de um olhar curioso, atento e flexível que a LC como um todo estuda o significado múltiplo, investigando tanto suas regularidades quanto suas irregularidades. Falta-nos, ainda, entender de que modo os postulados da LC são adotados especificamente pela SLC em seus estudos sobre a relação entre *item lexical*, *sentido (múltiplo)*, *mente* e *uso* – tópico a ser abordado na próxima seção.

3.2.2 Contribuições da Semântica Lexical Cognitiva para o Estudo dos Sentidos

O compromisso cognitivo da LC assumido a partir de uma concepção não-modular de mente, apresentado no primeiro postulado, reflete-se no estudo do significado lexical, que passa a ser entendido como a expressão de processos gerais de categorização do mundo (TAYLOR, 1995). Partindo da concepção de que um item lexical (assim como as demais unidades linguísticas) pode ser estudado em termos de uma categoria conceitual – uma categoria linguística –, refletindo os avanços da Psicologia Cognitiva, destacamos como “núcleo comum” da SLC nas últimas duas décadas os resultados de pesquisas relacionadas aos seguintes tópicos: (a) a estrutura interna das categorias lexicais (ex.: estrutura de protótipo, estrutura por semelhança de família, estrutura de redes); (b) a natureza polissêmica das unidades linguísticas e os princípios cognitivos motivadores das relações entre seus diferentes sentidos (ex.: metáfora, metonímia, transformações por esquemas imagéticos); (c) estruturas conceituais mais amplas (ex.: pesquisas sobre metáforas, semântica de *frames*) (TAYLOR *et. al*, 2003).

O segundo postulado, o de que a língua emerge no uso, gera implicações para entendermos qual é a natureza dos sentidos associados aos itens lexicais e direciona seu estudo no contexto da SLC. Nega-se a concepção de significado como um objeto *reificado*, que está fixamente implicado nos itens lexicais de maneira independente do contexto em que eles são encontrados (TAYLOR *et. al*, 2003). Entre as alternativas propostas a essa noção, está a concepção de “potencial de significado” (*cf.* ALWOOD, 2003; SLATEV, 2003; JANSSEN, 2003; CROFT e CRUSE, 2004), entendido como aquela informação (porção de conteúdo conceitual) que poderá ser acessada por um item lexical a partir de informações contextuais e restrições de convenção. Entende-se, assim, que o potencial de significado de uma unidade linguística resulta do conhecimento armazenado sobre como aquela unidade tem sido utilizada em suas ocorrências anteriores. De acordo com Alwood (2003), essa noção serve como uma alternativa à necessidade de distinguir entre sentidos relacionados armazenados na memória de longa distância (polissemia) e sentidos construídos no momento do uso (monossemia).

Nesse contexto, conhecer uma unidade linguística envolve aprender um conjunto (possivelmente grande e aberto) de usos estabelecidos e padrões de uso (*cf.* TAYLOR, 2008). O significado lexical não igualado à interpretação corresponde a um potencial de

informação a ser construído ao se fazer uso de determinado item; a interpretação, diferentemente, é entendida como o processo que se dá em tempo real e a partir da compatibilização da porção do conteúdo conceitual acessada pelas unidades linguísticas (sentido) com as informações contextuais da situação (contexto linguístico imediato e conhecimento enciclopédico) e o conjunto de restrições convencionais (padrões de uso). O conteúdo conceitual que um falante deseja simbolizar a partir de determinada porção linguística em uma ocasião particular não corresponde, portanto, exatamente aos recursos disponíveis em nossa memória de longo termo. Extensões, reduções, associações são alguns dos processos cognitivos realizados pelo falante e pelo interlocutor no momento de construir a interpretação. Na medida em que essas novas construções de significado vão sendo repetidas, elas se tornam “incrustadas” (originalmente, *entrenched*) na língua e se tornam conteúdos disponíveis para próximas situações de uso. O fato de uma mesma unidade poder designar situações muito diferentes, ou seja, o fato de que o conteúdo linguístico pode ser elaborado de diferentes modos em diferentes contextos (TAYLOR et. al, 2003) é um dos principais desafios enfrentados pela SLC e está diretamente relacionado ao tema deste trabalho. O que está em questão aí, na verdade, é a relação entre uma concepção de sentido fixo e estável *versus* sentido flexível e sensível ao contexto. Aquela porção do significado lexical mais fixa e estável expressa em uma língua é a que é registrada em dicionários tradicionais, mas é importante notar que ela não representa para a SLC o todo do conteúdo acessível por um item.

Ao ser abandonada a visão reificada do significado lexical, é preciso abandonar também a visão tradicional de armazenamento e representação desse conteúdo. A estratégia convencional de armazenamento de sentidos a partir de uma lista de itens lexicais com sentidos estáveis e independentes é conhecida em SLC como a visão de léxico como um “dicionário tradicional” (cf. TAYLOR et. al, 2003; CROFT e CRUSE, 2004). Taylor et. al (2003) propõem que a chamada “visão dicionária” seja substituída pela estratégia do “cópus mental”, perspectiva que é representativa daquela adotada em SLC como um todo. Entende-se que o cópus compreende a memória de registros previamente encontrados na língua especificados tanto em seus aspectos de conteúdo quanto estruturais, bem como em generalizações sobre eles. De maneira análoga à concepção de que o sistema linguístico não é inteiramente comparável ao sistema cognitivo (que é mais amplo), devemos entender que o cópus mental, sob essa ótica,

não é inteiramente comparável a um *cópus* de língua real, tal como aqueles construídos pela Linguística de *Cópus*. Os sentidos associados aos itens lexicais que constituem o *cópus* mental são acessados de acordo com seu potencial para contribuir à compreensão dos contextos reais da língua em que ocorrem, pois representam generalizações de suas ocorrências particulares (suas instâncias), e não as ocorrências propriamente ditas.

Significado lexical é conceitualização – essa é a versão da SLC para *slogan* de Langacker associado à tese simbólica da língua, apresentado no terceiro postulado geral. Essa versão do *slogan* pode ser justificada no contexto da SLC, e no contexto da LC como um todo, pela concepção de que o conhecimento que uma pessoa tem da língua reside precisamente no conhecimento dos itens lexicais e de suas propriedades (cf. TAYLOR, 1995). Por serem unidades simbólicas, os itens lexicais são entendidos como *motivadores* (*prompts*) de operações cognitivas responsáveis pela *construção do significado* (do inglês, *construal of meaning*) a ser registrado e atualizado no sistema conceitual humano a partir da sua experiência com a língua no mundo; por isso, diz-se que as estruturas semânticas são estruturas conceituais corporificadas (cf. FAUCONNIER e TURNER, 2003).

É frequente em SLC a posição de que são raras as situações em que um item é associado a apenas um sentido e, por essa razão, Langacker (cf. apresenta Taylor, 2002, p. 461) propõe que uma unidade linguística – “unidade simbólica”, nos termos do autor - não seja tratada como um composto que envolve uma relação de *um-para-um*, mas como a associação de uma rede de representações do conteúdo semântico a uma rede de representações da forma linguística (materialização do sentido). Taylor (2002) faz tal consideração em um capítulo dedicado ao tratamento de “categorias linguísticas complexas” e questões em torno do tema polissemia.

Essa visão panorâmica que apresentamos aqui sobre os objetos de estudo da SLC e sobre o modo com os princípios gerais da LC são instanciados no estudo do significado lexical foi importante para entendermos agora com mais completude os motivos pelos quais é a SLC o paradigma teórico que mais respostas tem a dar para um moderno estudo da polissemia. Para encerrar esta seção, retomaremos quatro das questões apresentadas em 3.1 por serem centrais ao tratamento da polissemia e trazerem respostas satisfatórias no paradigma da SLC:

(i) *Qual é a natureza da representação dos sentidos?* Tendo em vista que a categorização linguística é um fenômeno cognitivo, assim como outras capacidades cognitivas humanas, é importante estudar a relação entre essas capacidades. A partir daí, entende-se que informação linguística, informação enciclopédica e informação conceitual são estruturas integradas e interdependentes, de maneira que uma unidade linguística simboliza o conteúdo representado em estruturas conceituais a partir da interação do falante com o mundo. Desse modo, a língua não refere diretamente o “mundo real”, mas refere o que está representado no sistema conceitual humano, que é composto por estruturas conceituais que indiretamente refletem e interpretam o mundo por intermédio da experiência e da percepção humana (TYLER e EVANS, 2003).

(ii) *Os sentidos são entidades estáveis e fixas?* Acreditar que o sentido emerge do uso pressupõe negar que os sentidos sejam entidades estáveis e fixas. O conteúdo associado aos itens lexicais é, de acordo com a SLC, construído de maneira corporificada, sendo que é a experiência com a língua que o torna simultaneamente estável (estabelecido) e flexível, e é a frequência de uso que cria as regiões de maior estabilidade no conteúdo semântico acessível a um item. Questões como *de que maneira tratar teoricamente essa relativa instabilidade dos sentidos, qual é o nível de generalidade de identificação e descrição dos sentidos*, entre outras, são objetos de estudo da SLC e serão abordadas nas seções seguintes.

(iii) *Como os sentidos podem ser enumerados?* Vimos que, no contexto da SLC, foi abandonada a visão de léxico como dicionário e apresentada uma visão de léxico estruturado de maneira flexível e psicologicamente coerente. Além disso, vimos sugeridas estruturas de rede para a representação do conteúdo semântico associado a um item lexical.

(iv) *Que tipo de unidade pode apresentar polissemia?* Ao se adotar uma visão não-modular de mente, abandona-se a separação estrita entre léxico, sintaxe, morfologia, texto, etc. Por consequência, todas as unidades linguísticas são entendidas da mesma maneira, associadas a categorias cognitivas. Por essa razão, a polissemia não é um fenômeno restrito aos itens lexicais (palavras), mas que faz parte de unidades fonológicas, morfológicas, sintáticas, de expressões fixas, de unidades de entonação, etc.

Para finalizar, acrescentamos ainda que os sentidos ou os significados lexicais não são, em SLC, unidades atômicas. Podemos falar de *sentidos plenos*, *subsentidos*, *facetas de sentido* e, ainda, de *microsentido* (cf. CRUSE, 1995; 2000a; CROFT e CRUSE, 2004). Esses termos voltarão a ser tratados mais adiante, pois interferem no tratamento da polissemia. Agora, é importante que fique claro que, quando empregamos os termos *sentido*, *significado lexical* ou, de maneira mais neutra, *leitura*, não estamos pretendendo diferenciar *sentidos plenos* de *subsentidos*; quando essa for a intenção, usaremos os próprios termos *sentido pleno*, *subsentido*, *faceta de sentido*, *microsentido* – todos eles seguindo os mesmos princípios de fundo descritos nesta seção.

3.3 A CARACTERIZAÇÃO DA POLISSEMIA COMO UM FENÔMENO CATEGORIAL

A partir do que apresentamos em 3.1, para nós, não restam dúvidas de que, em Linguística, a SLC é o paradigma teórico mais produtivo para o estudo da polissemia, tendo em vista sua visão de língua e o ferramental teórico de que lança mão para explicar a língua; portanto, esse é o paradigma segundo o qual delimitaremos o fenômeno a ser investigado¹². Um dos princípios de base da SLC que precisamos ter em mente para entendermos em que medida a polissemia pode ser caracterizada como uma noção categorial é o fato de que os itens lexicais instanciados na superfície linguística são entendidos como entidades simbólicas associadas a um pareamento entre forma (estrutura fonológica) e sentido (estrutura semântica/conceitual) (LAKOFF, 1987; LANGACKER, 1987; 1999; TAYLOR, 2003). De acordo com Langacker (1999, p.4), “essa relação representa a essência de experiências compartilhadas, de modo que qualquer unidade linguística (sentenças, orações, expressões cristalizadas, morfemas ou mesmo

¹² Vimos que a LC pode ser considerada uma família de teorias, sendo que, embora cada membro tenha um foco e termos específicos, todos seguem os mesmos princípios gerais para analisar itens lexicais ou expressões fixas, por exemplo. Entendemos, portanto, que as semelhanças entre os paradigmas são mais marcantes do que as diferenças, não se fazendo necessário, para os fins desta tese, optarmos por apenas um deles. Ficará marcado em nosso texto quando estivermos apresentando noções que nasceram no contexto da Gramática Cognitiva, da Gramática das Construções ou da Semântica Lexical Cognitiva ao referirmos autores como Langacker, Goldberg, Cruse e Geeraerts, que são tomados aqui, respectivamente, como representantes de cada vertente. Nossa posição segue em certa medida a postura de Taylor, que é simpatizante da Gramática Cognitiva e profundamente interessado em um dos temas mais caros à SLC (se não o mais) que é a polissemia.

um texto), tendo em vista a sua função simbólica, pode ser descrita em termos desse tipo de relação”.

Antes de avançarmos, consideramos relevante caracterizar com mais precisão, no contexto teórico em que nos inserimos, essa famigerada relação entre *forma* e *conteúdo* (cf. LANGACKER, 1987; TAYLOR, 2003), que está na base da compreensão do fenômeno da polissemia – tema da próxima seção, tema da Seção 3.3.1. Feitas essas observações iniciais, passaremos ao delineamento propriamente dito da polissemia como um fenômeno categorial, seção 3.3.2.

3.3.1 A relação entre forma e conteúdo

Olhar para a relação *forma* e *conteúdo* no contexto teórico desta tese pressupõe olhar:

- (a) para a *estrutura fonológica* (forma), que representa aquela porção perceptível de uma unidade linguística. Ela pode ser entendida, em sentido estrito, como o som ou, em sentido amplo, como a forma linguística, sequência de caracteres, ícones. Interessa-nos a noção de estrutura fonológica em sentido amplo; por exemplo, *árvore*, grafada aqui em letras minúsculas e em itálico (TAYLOR, 2003);
- (b) para a *estrutura semântica* (conteúdo), que representa o espaço para o conteúdo semântico simbolizado pela estrutura fonológica. Ela pode ser um sentido específico de um item ou um conteúdo construído (do inglês, *construal*) a partir de uma situação complexa (PALMER, 2006). Essa é a porção do espaço conceitual onde os sentidos ou os significados lexicais residem. A notação adotada aqui para representar a estrutura semântica é a de aspas simples, por exemplo, ‘vegetal lenhoso de porte variável’;
- (c) para as *relações* entre estrutura fonológica e estrutura semântica, consideradas bidirecionais, que podem ser justificadas por se entender que

um som (estrutura fonológica) pode evocar um pensamento (estrutura semântica) ou que o pensamento pode evocar um som (PALMER, 2006). Um exemplo de relação é a que podemos perceber entre *cachorro* como ‘macho da raça canina’ (menor grau de detalhamento) e como ‘membro da raça canina’ (maior grau de detalhamento). Nos termos de Langacker, trata-se de uma relação entre o *esquema* (sentido geral) e sua *instância* (sentido específico). Referindo-se ao mesmo tipo de relação, Cruse (2000a) faz uso dos termos *autohiponímia* e *autossuperordenação*, termos que já vimos nas bases *wordnets* e detalhamos nesta parte linguística do trabalho.

Partindo do pressuposto de que, conforme a SLC, a relação de mapeamento mais frequente entre os polos fonológico e semântico não segue o ideal semiótico *um-para-um*, mas que ela ocorre de *muitos-para-muitos*, considera-se que ambos os polos são formados por estruturas complexas. A polissemia é um fenômeno que envolve, portanto, estruturas fonológicas comuns ligadas pelo mesmo polo semântico, mas não a relação entre uma forma (em sentido estrito) e muitos sentidos. As relações que interessam ao tratamento da polissemia serão, nesta tese, denominadas *relações de sentido*, termo que empregamos para diferenciá-las das clássicas relações semânticas que se estabelecem entre itens lexicais distintos, tal como ocorre entre *cachorro* e *animal* (relação de hiponímia). A figura 9 registra graficamente essa relação entre estruturas complexas em ambos os polos que compõem uma unidade linguística.



Figura 9 - Relação *forma-conteúdo*

Vemos na figura 9 que há mais de um ponto dentro de cada quadro; além disso, os pontos estão ligados formando uma espécie de rede. Através dessa representação,

pretendemos ilustrar a composição complexa de ambas as estruturas, fonológica e semântica. A relação entre os polos é indicada pela seta bidirecional. Nessa figura, fica em destaque o fato de que não podemos pensar a polissemia como uma relação entre uma forma (simples) e múltiplos sentidos, visto que o polo fonológico é por múltiplos elementos representado, o que, na língua, podemos perceber em expressões fixas.

É importante lembramos ainda que os sentidos associados às unidades linguísticas são, em SLC (*cf.* CROFT e CRUSE, 2004), entendidos como perspectivas do nosso conhecimento de mundo, isso porque são decorrentes de processos de categorização do mundo, rotinas cognitivas responsáveis por delimitar conceitos que serão, a partir de padrões de acesso e de regras de convenção, acessados pelos itens lexicais em diferentes situações de uso na forma do que estamos denominando *sentido* ou *significado lexical*. É a partir da frequência de uso que determinada interpretação torna-se *estabelecida* ou *incrustada* e que passará a adquirir o *status* de sentido na memória de longa duração, tornando-se disponível para próximos usos e operações de interpretação. E o fenômeno da multiplicidade de sentidos, para dizer em linhas gerais, é entendido como decorrente do potencial de um item lexical de ser usado em diferentes contextos como porta de acesso para diferentes entidades conceituais (estejam elas agrupadas em uma mesma *categoria de sentidos*, ou façam parte de espaços conceituais disjuntos).

Feitas essas observações iniciais, passamos ao estudo da relação entre polissemia e categorização, principal responsável pela formação dos pontos que estruturam o polo semântico e pela construção de associações com o polo fonológico.

3.3.2 A polissemia como categorização

Meaning is equated with conceptualization. Linguistic semantic must therefore attempt the structural analysis and explicit description of abstract entities like thoughts and concepts. (...) Because conceptualization resides in cognitive processing, our ultimate objective must be to characterize the types of cognitive events whose occurrence constitutes a given mental experience (LANGACKER, 2002, p.2).

A habilidade de categorizar é central aos seres humanos. Ela pode ser definida como a nossa habilidade para identificar entidades como membros de grupos (EVANS e

GREEN, 2006). A compreensão sobre como nós, seres humanos, categorizamos o mundo é central para entendermos como pensamos, funcionamos (LAKOFF, 1987) e, por consequência, como nos expressamos linguisticamente. Essa não é uma posição exclusiva e inovadora de Lakoff ou da LC, porém ela é representativa do compromisso cognitivo assumido pelo paradigma, razão pela qual se espera que a organização linguística se dê em função de processos cognitivos gerais, como é a categorização (EVANS e GREEN, 2006, p. 168). Nesse contexto, a polissemia é, então, entendida como o efeito de um tipo de processo de categorização, aquele responsável pela construção de categorias lexicais semanticamente complexas. Essa posição foi introduzida inicialmente por Brugman (1981), foi popularizada e reelaborada por Lakoff (1987) e recentemente recebe atenção de pesquisadores como Taylor (ex.: 1989; 1995; 2003), Cruse (2000b), Croft e Cruse (2004), Geeraerts (ex.: 1993; 2006), Tyler e Evens (2003a, b), Soares da Silva (2006), entre vários outros no contexto da SLC.

Sabemos que o processo geral de categorização se dá, na maior parte das vezes, de modo inconsciente (LAKOFF, 1987): toda vez que vemos uma coisa de algum *tipo*, por exemplo, uma árvore; sempre que pensamos sobre *tipos* de coisas (cadeiras, nações, doenças, emoções, qualquer coisa), estamos aplicando categorias; quando intencionalmente realizamos algum *tipo* de ação, dizemos alguma coisa, martelamos, cozinhamos ou estudamos, estamos categorizando, isso porque a ação particular que desempenhamos nessas ocasiões é um *tipo* de atividade motora que é uma categoria particular de uma ação motora (ex. falar, escrever, passar roupa). Embora um *tipo* de “coisa” como uma ação nunca seja feita exatamente da mesma maneira, apesar das diferenças em movimentos particulares, são movimentos de um *tipo*, e sabemos como realizar movimentos de determinado *tipo*. E em qualquer momento em que produzimos ou compreendemos qualquer enunciado de qualquer tamanho, aplicamos dúzias, se não centenas, de categorias: categorias de sons da fala, de palavras, de frases e orações, bem como categorias de sentidos (LAKOFF, 1987).

O fato de existir em uma língua particular uma única unidade linguística (simples ou complexa) que possa instanciar múltiplos sentidos é consequência da existência percebida de uma estrutura conceitual ou semântica complexa a ela convencionalmente associada (estrutura ilustrada na figura 9). Essa estrutura conceitual é denominada por Lakoff (1987) *categoria de sentidos*, pois compreende o potencial de sentido convencionalmente associado a uma estrutura fonológica. A atividade cognitiva

responsável pelo isolamento de diferentes porções desse espaço conceitual (delimitação de sentidos) em diferentes ocasiões pode ter como consequência a polissemia, dependendo de quão independentes e similares forem os sentidos delimitados (CROFT e CRUSE, 2004)¹³. Esse processo cognitivo desencadeado para a construção e acesso a determinado sentido é conhecido em SLC como operação de *construção do significado* (originalmente, “construal of meaning”, cf. Croft e Cruse, 2004), e a categorização é um de seus principais instrumentos. Entende-se, a partir daí, que a polissemia é um fenômeno cognitivo instanciado na superfície da língua via itens lexicais, morfemas, fonemas, expressões fixas, unidades de entonação, etc., mas não é exclusivo do nível da superfície (TAYLOR, 1995; CRUSE, 1995; CUYCKENS e ZAWADA, 1997; TAYLOR, 2006).

Uma categoria de sentidos em SLC é um tipo de categoria linguística. Isso faz com que o tratamento dado a uma categoria conceitual seja entendido como aplicável às categorias de sentidos ou categorias semânticas. Embora os estudos tenham evidenciado essa compatibilidade, a noção de *categoria* precisa ser detalhada para entendermos em que medida os sentidos associados a um item lexical podem ser considerados categorias, concepção que precisa ser diferente daquela originalmente proposta por Rosch (TAYLOR, 2008; GEERAERTS, 2006). O termo *categoria* pode ser empregado tanto para denotar itens lexicais (polissêmicos ou não), sentidos individuais de um item ou, ainda, um conjunto de sentidos relacionados associados a um item. No primeiro caso, fala-se de *categoria lexical* (GEERAERTS, 2006); nos outros, de *categoria semântica* (GEERAERTS, 2006) ou *categoria de sentidos* (LAKOFF, 1987).

Os membros de uma categoria de sentidos, na visão clássica de categorização que dominava o ocidente desde Aristóteles, com afinidades com a Filosofia e a Lógica, eram definidos a partir de um conjunto de condições necessárias e suficientes, o que acarretava que a delimitação de membros de uma categoria fosse definida como sim/não – tudo/nada (EVANS e GREEN, 2006). A partir de Wittgenstein (1958), no contexto da Filosofia da Linguagem, com o famoso estudo sobre a palavra *game*, ficou evidente que, ao contrário do que se pensava, as categorias têm fronteiras que podem variar, e a estrutura interna das categorias passou a ser pensada fundamentalmente a partir de efeitos de semelhança de família, contendo elementos que se sobrepõem e estabelecem cruzamentos de modo semelhante ao que ocorre em famílias (cf. RAVIN e

¹³ A noção de delimitação entre sentidos será analisada e detalhada em 4.1.

LEACOCK, 2000). Rosch demonstrou, seguindo essa concepção de semelhança de família, que a categorização não é uma questão binária de tudo ou nada, mas que o julgamento sobre a categorização se dá em função de semelhança a um membro prototípico da categoria, dando origem à noção de “membership gradience” (a percepção de que há graus de pertencimento a uma categoria), princípio da Teoria de Protótipos, que estuda os diferentes critérios para definir o elemento central de uma categoria. O exemplo clássico é a categoria *pássaro*: quando categorizamos *pássaros*, seja em uma perspectiva extensional, seja em uma perspectiva intensional, certos tipos, como sabiás e pardais, são julgados como “melhores” exemplares da categoria (natural ou de sentidos), ou seja, como o mais próximo ao que prototipicamente se espera dessa categoria, diferentemente de outros (como pinguins).

Rosch propôs dois tipos de modelos de protótipos: em um deles, uma categoria é composta por apenas um membro prototípico que possui o maior número de traços característicos da classe; no outro, há diversos protótipos, cada um possuindo um conjunto de traços característicos, que não necessariamente estabelece igualdade com o outro (*cf.* RAVIN e LEACOCK, 2000). Foi a partir desse segundo modelo que linguistas passaram a investigar em que medida os princípios de estrutura de uma categoria natural podem ser aplicados a uma categoria lexical polissêmica – reflexão que culminou na proposição dos modelos de redes radiais (BRUGMAN, 1981; LAKOFF, 1989) e esquemáticas (LANGACKER), conforme veremos em 3.6. Nesse contexto de estruturação de categorias linguísticas em torno de um (ou mais de um) centro prototípico, a polissemia expressa na língua pode ser vista, conforme Taylor (1995), como reflexo de um processo de “categorização policêntrica”.

Esse olhar, que passou a ser recorrente após Brugman e Lakoff, alterou os parâmetros para a delimitação da polissemia: de um fenômeno exclusivamente lexical (tal como apresentamos na versão de Apresjan, 1973), passou a ser entendido como um fenômeno cognitivo e categorial. A compreensão dessa mudança passa pelo entendimento de distinção entre os níveis lexical, conceitual e referencial, entendendo-se que o nível lexical é composto pelas unidades linguísticas como formas – a estrutura fonológica; que o nível conceitual é composto pelas categorias cognitivas – consideradas, ao mesmo tempo, os sentidos dos itens lexicais, a estrutura semântica; e que o nível referencial é composto pelas entidades do mundo real ou de mundos possíveis (GEERAERTS, 2006). Como uma noção não-categorial, a polissemia é, então,

caracterizada como múltiplos sentidos relacionados associados a uma única forma linguística. Como uma noção categorial, a polissemia pode ser definida como uma questão de delineamento de limites entre categorias e subcategorias de sentidos, nas palavras de Croft e Cruse (2004), como uma questão de criação de fronteiras entre unidades de sentido mais ou menos autônomas; e, diferentemente do que tradicionalmente se entende, não é a unidade da forma sua consequência direta, mas a unidade da *categoria lexical* que se expressa na língua via estrutura fonológica, até porque não se pode esquecer que nem sempre uma *categoria de sentidos* ou *semântica* é instanciada na língua por uma única forma, conforme ocorre, por exemplo, com expressões fixas, como *dar uma volta* em *Carol foi dar uma volta*, no sentido de que Carol ‘saiu’.

Olhar para a polissemia sob o viés da cognição geral é posição comumente justificada em SLC a partir da noção de *motivação*, que é percebida como relacionada à polissemia desde Bréal. Tal como é tradicionalmente definida em SLC, a motivação envolve os princípios que explicam (ou que tornam plausíveis) por que uma unidade linguística significa o que ela significa. Mas não podemos confundir: o que a motivação explica não é por que determinada expressão tem determinado significado, mas por que determinados conteúdos conceituais existem para que sejam posteriormente associados às unidades linguísticas a partir de regras de convenção (BRUGMAN e LAKOFF, 2006). Essa noção de motivação aparece como uma alternativa para a noção mais tradicional de *previsibilidade*, entendendo-se que, mesmo que os significados não sejam previsíveis, eles podem ser motivados pela existência de tendências e esquemas cognitivos (ex. metáfora e metonímia¹⁴) que existem no sistema conceitual de maneira independente de determinada unidade linguística (BRUGMAN e LAKOFF, 2006). E são esses padrões cognitivos os responsáveis por criar diferentes *graus de redundância* no conteúdo conceitual; ao serem instanciados na superfície da língua, caracterizam a polissemia (e não a homonímia). Sob essa perspectiva, de acordo com Brugman e Lakoff (2006, p. 131

¹⁴ As definições de metáfora e metonímia não são unânimes. Delimitar em detalhes essas noções exigirá aprofundadas investigações que ultrapassam os objetivos desta tese. Apresentaremos apenas uma definição geral. A extensão metafórica ocorre quando um sentido associado a um domínio conceitual é construído em termos de outro (cf. LAKOFF e JOHNSON, 1980; LAKOFF, 1987). Ex.: *Ela tem um poder estranho sobre mim*, em que o sentido de ‘controle’ é tido como derivado do sentido de ‘acima’ a partir da associação metafórica ‘CONTROL IS UP’ (EVANS e GREEN, 2006, p. 36). A metonímia conceitual ocorre quando um sentido ocorre no lugar de outro, sendo ambos pertencentes ao mesmo domínio. Ex.: O exemplo clássico de metonímia *ham sandwich* representa uma instância de metonímia porque tanto o *customer* quanto o *ham sandwich* pertencem ao mesmo domínio CAFÉ.

[originalmente, 1988]), “é o caráter independentemente motivado das relações [entre os sentidos polissêmicos] que permite explicar por que a polissemia deve existir como um fenômeno geral”. Nessa perspectiva, evidencia-se um olhar para a polissemia que não é inteiramente sincrônico, visto que investiga fundamentalmente as rotinas cognitivas responsáveis pela mudança semântica, mais do que as relações sincrônicas que podem ser percebidas entre os sentidos polissêmicos. Essa é a postura mais frequente em SLC, e foi, conforme demonstraremos em mais detalhes adiante, fundamentalmente a partir de seus resultados que se avançou na compreensão do fenômeno. É possível, no entanto, olhar para a polissemia como um fenômeno categorial, minimizando os fatores diacrônicos e maximizando as relações sincrônicas de polissemia, visão que também está presente em SLC – embora em menor número – e que interessa a este trabalho.

Para concluir esta seção e sintetizar as informações apresentadas aqui, lembramos que 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), mas como um fenômeno acidental (não motivado) que não constitui um agrupamento na forma de uma categoria de sentidos, 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. Embora seja verdadeira e útil, essa caracterização ainda não é suficiente para realizarmos a tarefa de descrição da polissemia.

3.4 ASPECTOS ESTRUTURAIS E FUNCIONAIS DA DIMENSÃO SEMASIOLOGICA DO LÉXICO

As informações apresentadas em 3.1 e 3.2 nos possibilitaram dar mais um passo em direção ao equacionamento de nossos objetivos por prepararem o cenário para a descrição do fenômeno em foco no contexto da SLC. O olhar para a polissemia como um fenômeno categorial possibilita lançarmos mão do instrumental teórico disponível ao tratamento da semântica sob o viés cognitivo. A partir desta seção, analisaremos elementos necessários à representação de nominais polissêmicos sob a perspectiva da SLC. Iniciaremos dando destaque a duas questões implicadas na descrição do conteúdo lexical como um todo que interferem diretamente no tratamento a ser dado à polissemia, são elas: (i) as dimensões *semasiológica* e *onomasiológica* de estruturação semântica e (ii) seus aspectos *qualitativos* e *quantitativos*. As considerações feitas por Geeraerts (2006) e Soares da Silva (2006) guiarão o conteúdo aqui apresentado.

É frequente em SLC explicar a estruturação do conteúdo lexical lançando mão das noções comuns à Lexicografia, que são as dimensões *semasiológica* e *onomasiológica* do léxico. A reflexão sobre essas dimensões ajuda-nos a entender diferentes visões para o estudo do conhecimento lexical. As dimensões *semasiológica* e *onomasiológica* envolvem duas perspectivas inversas. Na dimensão *semasiológica*, o ponto de partida é a forma lexical, vai-se da forma para o conteúdo (nível intensional), ou o referente (nível extensional); por outro lado, na dimensão *onomasiológica*, o ponto de partida é o conteúdo, ou o referente, passando-se, então, para a forma lexical a ele associada. A polissemia é um fenômeno *semasiológico* típico, visto que seu estudo parte da forma para os sentidos (e referentes) que a ela se associam. Na dimensão *onomasiológica*, a estruturação lexical pode ocorrer por meio de relações semânticas (ex. sinonímia, antonímia, hiponímia, meronímia), campos semânticos, *frames* semânticos, etc. A sinonímia é o inverso *onomasiológico* mais direto da polissemia.

A descrição do conteúdo associado a uma categoria polissêmica deve também levar em consideração aspectos *qualitativos* e *quantitativos* das estruturas lexicais tanto do ponto de vista *semasiológico* quanto *onomasiológico*. Numa perspectiva *semasiológica*, estão envolvidos aspectos estruturais ou qualitativos do léxico que

remetem a entidades e a suas *relações*. Olhando especificamente para a polissemia, tratamos dos sentidos associados a um mesmo item e das relações entre eles. As questões fundamentais dessa dimensão são identificar os diferentes sentidos associados a uma mesma forma lexical e as relações que entre eles podem se estabelecer (ex.: especialização, generalização). Além de propriedades estruturais, aspectos funcionais do uso ou *quantitativos*, que remetem a diferenças de *saliência*, precisam ser considerados para a descrição do conteúdo lexical a partir da ótica semasiológica. Para a descrição de um item polissêmico, o que entra em questão nesse aspecto são diferenças de *saliência* entre os sentidos agrupados em uma categoria polissêmica (ex. *saliência psicológica*, *saliência estatística*, *saliência referencial*, etc. – que serão detalhadas em 3.7). A questão principal que se apresenta nessa dimensão é saber se os sentidos associados a um item têm o mesmo “peso estrutural” ou o mesmo poder de ativação em um contexto. As diferenças de *saliência* são entendidas também em termos de *efeitos de prototipicidade* entre os diferentes sentidos de uma categoria polissêmica. A definição do tipo de estrutura adotada para descrever um complexo polissêmico é também questão a ser tratada na dimensão funcional ou quantitativa de estruturação do conteúdo lexical. Tendo em vista os diferentes pesos estruturais dos componentes de uma categoria de sentidos, diferentes organizações podem ser pensadas. As mais populares são: a rede radial (proposta por Brugman e Lakoff), que prevê diferentes graus de centralidade para os sentidos, havendo um (ou mais de um) elemento prototípico que se relaciona radialmente com sua(s) extensão(ões); e a rede esquemática (proposta por Langacker), que prevê a representação de um sentido prototípico diretamente relacionado à sua extensão, estando ambos relacionados a um possível sentido esquemático (abstrato, englobante) – em foco na Seção 3.7.

Numa perspectiva onomasiológica, os aspectos qualitativos tratam dos variados conjuntos que podem ser formados com itens lexicais distintos que se relacionam por participarem de um mesmo campo lexical, de uma mesma taxionomia, de um mesmo *frame*, além de outros tipos de agrupamentos por meio de relações paradigmáticas, por exemplo, como sinonímia, antonímia, meronímia, hiponímia. Os aspectos quantitativos da dimensão onomasiológica envolvem as diferenças de *saliência* ou de “fixação”, o que remete à noção de *entrenchment* (termo apresentado em Langacker (1987) e traduzido

por Soares da Silva (2006) como “incrustamento conceitual”, para denotar o grau de fixação de um sentido em uma língua), entre diferentes categorias conceituais.

O quadro 3 sintetiza as informações reportadas aqui sobre o tipo de conteúdo semântico implicado na estruturação de uma categoria polissêmica. Por consequência, o quadro serve também como uma síntese do tipo de contribuição que uma descrição da polissemia fornece à estruturação do conteúdo lexical como um todo.

| | QUALIDADE:
entidades e relações | QUANTIDADE:
diferenças de saliência |
|---------------|---|---|
| SEMASIOLOGIA | Polissemia
Sentidos e relações (rede)

(generalização,
especialização) | Polissemia

Efeitos de saliência |
| ONOMASIOLOGIA | Itens lexicais e relações
(campos lexicais, taxionomias,
<i>frames</i> , relações) | Incrustamento |

Quadro 3 - A polissemia e a descrição da semântica lexical¹⁵

O quadro 3 põe em destaque questões semânticas que fazem parte da descrição da polissemia lexical e ajuda-nos a avaliar em que medida seu tratamento está relacionado com o tratamento do léxico como um todo. Um primeiro tópico a ser destacado, conforme Soares da Silva (2006), é o aparecimento de informações do tipo *relações* tanto na dimensão semasiológica quanto onomasiológica, numa perspectiva estrutural. O segundo ponto a destacar é a existência de implicações recíprocas entre semasiologia e onomasiologia, desse modo, a investigação das conexões entre sentidos associados a uma mesma forma não pode ocorrer de modo independente da investigação sobre conexões que esses itens lexicais estabelecem com outros semanticamente próximos (SOARES DA SILVA, 2006).

¹⁵ Adaptado de Soares da Silva (2006, p.14).

Tendo em vista que olhar para unidades linguísticas polissêmicas é olhar para a *dimensão semasiológica* do conteúdo lexical (cf. GEERAETS, 2006; SOARES DA SILVA, 2006), as informações apresentadas aqui foram importantes para revelar parte da complexidade da tarefa que nos propusemos: incluir uma representação que envolve a dimensão semasiológica do léxico em um recurso de base onomasiológica, como é uma *wordnet*. Um dos principais desafios está no fato de que a estruturação do léxico como um todo é diretamente afetada pelo tratamento que se dá à polissemia, razão pela qual não há como tratar da polissemia levando em consideração somente informações sobre relações semânticas em uma dimensão onomasiológica. Ao mesmo tempo em que a inclusão de relações semasiológicas em uma base de dados tipicamente onomasiológica pode parecer um empecilho, se levarmos em consideração o que defendem Geeraerts (2006) e Soares da Silva (2006), conforme apresentado ao longo desta seção, concluímos que essa integração de dimensões não só é possível, como é inevitável. Nas palavras de Soares da Silva (2006, p. 14), “condicionamentos recíprocos entre semasiologia e onomasiologia [são inevitáveis], pelo que ao estudo da polissemia de uma palavra ou de outra expressão não pode ser estranho ao estudo das relações entre essa palavra e outras semanticamente próximas”.

3.5 AS ABORDAGENS HOMONÍMICA, POLISSÊMICA E MONOSSÊMICA

Conforme mencionamos na seção anterior, uma das questões centrais para a descrição lexical sob a visão semasiológica é identificar os sentidos e caracterizá-los; para fazermos isso, precisamos saber: onde estão os sentidos? Eles são propriedades herdadas pelos itens lexicais ou fazem parte do conhecimento enciclopédico? O que exatamente essa pergunta significa quando pretendemos descrever (representar) o conhecimento semântico associado às unidades linguísticas? Os sentidos estão ligados a situações particulares ou são mais gerais? Se forem mais gerais, em que nível de generalidade eles podem ser encontrados e devem ser registrados? Não será que, no léxico mental, cada palavra tem apenas um único sentido? Em busca de resposta a questões como essas é que olharemos, nesta seção, para resultados de estudos linguísticos e psicológicos que a elas se dedicam. Previamente, é preciso dizer que a

literatura evidencia (ex. GEERAERTS, 1993; 2006; CRUSE, 1995; 2000; 2004; SOARES DA SILVA, 1999, 2006; TUGGY, 1993; 2006) que uma resposta única e unânime não é possível. Entretanto, é fundamental que opções sejam tomadas sobre quando e como cada novo uso de um item deve ser codificado em uma representação lexical.

Não há dúvidas de que perceber a relação entre os sentidos é um dos critérios de base para identificar a polissemia e diferenciá-la da homonímia; além disso, detectar quando e como os múltiplos sentidos estão registrados na nossa memória de longo termo (léxico mental) e quando eles podem ser gerados *on-line* – ou seja, a cada ocorrência em contexto específico de uso – é considerado o critério de base para diferenciar polissemia de monossemia. Investigar se as pessoas percebem as semelhanças entre os sentidos, como elas armazenam esses conteúdos e como processam itens lexicais com múltiplos sentidos tem desafiado linguistas, psicolinguistas e informatas, que adotam posições muito distintas, conforme veremos ao longo desta seção.

3.5.1 Caracterização e problemas implicados

As estratégias de representação encontradas na literatura são distintas. Em posições extremas, estão abordagens que entendem que cada estrutura fonológica é conectada a um único complexo semântico geral e que os sentidos particulares de um item se caracterizam somente em contexto. Nessa perspectiva, mesmo os sentidos sem relação (homônimos) coexistem em uma única estrutura semântica e são resolvidos pela análise do contexto linguístico em que são instanciados (BROWN, 2008). No outro extremo, estão abordagens que propõem que os sentidos associados a uma mesma forma, relacionados (polissêmicos) ou não (homonímicos), sejam representados em uma estrutura semântica distinta, com nada além da forma lexical em comum (BROWN, 2008).

Estudos psicolinguísticos entendem que, para que haja realidade psicológica em qualquer dessas estratégias, o tempo e a correção na identificação dos múltiplos sentidos relacionados que se associam a determinado item em pares de contextos deveriam ser idênticos ao que ocorre na identificação de pares com sentidos não-

relacionados (BROWN, 2008). Isso não é o que tem sido percebido nos experimentos, fato que evidencia que o acesso ao léxico mental não se dá da mesma maneira em todas as situações (BROWN, 2008). Por exemplo, se alguém é apresentado aos enunciados (a) *Maria escolheu o banco branco de madeira*, (b) *Maria escolheu o banco com taxas menores* e (c) *Maria escolheu o banco de sangue mais próximo de sua casa para fazer a doação*, o tempo de resposta para a identificação do sentido de *banco* e sua correção deveriam ser idênticos para (a-b-c). Os resultados de Brown (2008) revelaram, no entanto, que há diferença significativa no processamento, de maneira que, ao serem apresentados pares relacionados (em nossos exemplos, b-c), a identificação do sentido acontece mais rápida e corretamente do que quando a pessoa tem acesso a pares que não revelam nenhum tipo de relação (em nossos exemplos, a-e, a-c). Esses resultados sugerem que estratégias distintas devam ser adotadas para uma descrição do léxico que reflita nossa organização mental. Conforme vimos na primeira parte deste trabalho, em pesquisas voltadas ao PLN, do mesmo modo, os chamados “léxicos enumerativos” (contendo uma nova “entrada” lexical independente para cada novo sentido) também são preteridos, tendo em vista a relevância do registro de similaridades de sentidos para tarefas WSD, por exemplo.

Em posição intermediária, estão abordagens que sugerem que os sentidos relacionados compartilham uma representação semântica comum. Duas posições são percebidas: (i) o mínimo possível de polissemia e (ii) o máximo possível de polissemia.

Na posição (i), que procura associar os itens lexicais ao mínimo possível de sentidos, propõe-se que somente uma única estrutura nuclear e genérica seja armazenada na memória de longo termo, sendo os sentidos individuais gerados *on-line* através da combinação de informações pragmáticas e de padrões gerais de extensão e inferências com base no contexto de uso (BROWN, 2008) (trata-se da *abordagem monossêmica*). Essa posição pode ser percebida no clássico exemplo de Searle para o verbo *open* (*abrir*, em PB) (cf. TAYLOR, 1995; 2006). Searle chama a atenção para o fato de que uma pessoa encontra um mesmo item lexical em vários contextos e que frequentemente consegue perceber similaridades de diferentes tipos entre os usos, por exemplo: (a) em *open the office* (abrir o escritório) e *open the envelope* (abrir o envelope), pode ser percebida a ideia de obter acesso ao interior de algum contêiner; (b) em *open the door* (abrir a porta) e *open the lid* (abrir a tampa), o foco é dado para a maneira através da qual o acesso é obtido, manipulando-se algum dispositivo para criar

uma abertura; (c) em *open one's shirt* (abrir a camisa de alguém) e *open a penknife* (abrir um canivete), o foco é a movimentação de alguma parte de um dos componentes de um objeto; (d) em *open a nature reserve* (abrir uma reserva natural) e *open a road* (abrir uma rua), pode ser percebida a ideia de tornar algo acessível para o grande público; outro exemplo ainda são os contextos (e) *open a discussion* (abrir uma discussão) e *open a debate* (abrir um debate), que remetem ao início de um processo de interação (cf. TAYLOR, 1995; 2006). Todos esses usos são destacados pelo autor como convencionais e peculiares ao inglês para o verbo *open* e, nos exemplos enumerados, encontram correspondência com *abrir* em PB. Defendendo a posição de que qualquer adulto falante de inglês (e o mesmo ocorre para PB) reconhece esses usos se tiver aprendido os tipos de situação das quais essas expressões tipicamente fazem parte (*background knowledge*), Searle argumenta que *open* tem um único sentido constante (excluindo-se exemplos metafóricos) e que as variações percebidas em (a-e) se devem a componentes contextuais inferidos pelo falante a partir desse conhecimento cultural (TAYLOR, 2006). A posição de Searle, portanto, é representativa da postura de maximização da monosssemia e minimização da polissemia.

Argumentando na direção contrária à de Searle, vários problemas são percebidos na abordagem monossêmica, entre eles, destacamos (i) a delimitação do sentido; (ii) a aplicabilidade; (iii) a realidade cognitiva; e (iv) a maximização excessiva do contexto situacional e cultural (cf. KLEIN e MURPHY, 2002; SOARES DA SILVA, 2003; TAYLOR, 2006; BROWN, 2008). Sobre a dificuldade de *delimitação do sentido de um item lexical*, comentando o exemplo de Searle, Taylor (2006) destaca os problemas para estabelecer um sentido unitário para *open*, abstraindo as particularidades das várias expressões e atividades que elas designam. Uma possibilidade levantada pelo autor é o sentido 'criação de um espaço em uma entidade pela separação de suas partes'. Entretanto, logo esse sentido unitário é invalidado, pois a mesma definição seria aplicável a um verbo como *cut* ('cortar', em PB), em algumas situações. Com o objetivo de investigarmos se as dificuldades são também percebidas na delimitação de um sentido geral (monossêmico) para substantivos, avaliaremos, a seguir, o item *livro*.

Um item lexical como *livro*, em um contexto como (1) *comprei um livro*, pode ser considerado instância do sentido geral 'conjunto encadernado de folhas, cobertas por uma capa resistente [objeto físico] e preenchidas por conteúdo textual de cunho literário, artístico, científico, técnico, documentativo, etc. [conteúdo]'. Outros usos de

livro podem ser observados em (2) *usei aquele livro amarelo como apoio para porta* e (3) *o livro tem uma linguagem muito agradável*. Ao percebermos que o sentido geral, ilustrado em (1), de fato abarca os usos (2) ‘objeto físico’ e (3) ‘conteúdo’, de modo que podem ser considerados instâncias do sentido geral, a tendência seria inferirmos que a abordagem monossêmica é válida para esse tipo de exemplo. Outra evidência que pode levar à mesma conclusão é o fato de que esse mesmo tipo de variação entre sentido geral vs. sentido específico 1 ‘objeto físico’ e sentido específico 2 ‘conteúdo’ é recorrente no léxico (ex.: *jornal, revista, polígrafo, CD, vídeo, filme, etc.*). Sendo assim, parece simples concluirmos que os sentidos específicos são gerados por regras ou padrões lexicais, já que a possibilidade de ativar tanto o sentido ‘objeto físico’ quanto ‘conteúdo’ é recorrente no léxico.

É evidente a importância de se pensar em um conteúdo mais geral que possa ser instanciado a usos específicos quando se pretende descrever o significado lexical, ou seja, é válida a atividade de “puxar o significado para cima” (nos termos de Soares da Silva, 2003, p. 95). Há, no entanto, evidências de que esse não parece ser um recurso aplicável a todo o léxico como critério para a delimitação de um sentido unitário. O próprio sentido geral que selecionamos para *livro* não possui o grau de distintividade onomasiológica necessário para diferenciá-lo de *polígrafo* e *apostila*, por exemplo, visto que eles também podem ser definidos como ‘conjunto encadernado de folhas, cobertas por uma capa resistente [objeto físico] e preenchidas por conteúdo textual de cunho literário, artístico, científico, técnico, documentativo, etc. [conteúdo]’. O que aconteceu ao tentarmos definir *livro* é destacado como um dos problemas recorrentes a esse tipo de abordagem: a delimitação de um sentido nuclear que

[...] não é nem semasiologicamente unitário – porque é muito difícil prever nele e dele ‘derivar’, não apenas todas as como as próprias significações mais importantes, e porque ele não serve com certeza à ‘condição necessária e suficiente’ para o conhecimento, o uso e a interpretação que os falantes fazem –, nem onomasiologicamente distintivo – porque é incapaz de o distinguir de outros [itens lexicais]” (SOARES DA SILVA, 2003, p. 96-97).

Em complemento à crítica da delimitação do sentido, surge a da *aplicabilidade* da noção de sentido unitário. A literatura mostra que são raras (se existirem) as situações em que um único sentido geral é aplicável a todos os usos de um item, o que, nesse caso,

possibilitaria gerar novos usos a partir de informações de contexto (BROWN, 2008; SOARES DA SILVA, 2003; TAYLOR, 2006). Supondo-se a possibilidade de identificação de um sentido unitário, seriam necessários mecanismos geradores de novas interpretações para aquele item nos diferentes contextos em que ele puder aparecer (TAYLOR, 2006). Esses mecanismos precisariam ser de natureza geral, no sentido de que eles não sejam aplicáveis somente a determinado item lexical tal como ele é usado em contextos particulares; caso contrário, enfatiza Taylor (2006), voltaríamos a cair no erro de deixar em segundo plano a polissemia.

A *realidade psicológica* da abordagem monossêmica é também questionada. Estudos como os de Sandra e Rice (1995) e de Klein e Murphy (2002) argumentam que ela não é uma estratégia cognitivamente real. Experimentos realizados por eles demonstraram que os falantes não percebem os diferentes sentidos de itens lexicais associados a uma estrutura semântica unitária. Os autores afirmaram, ainda, que os mecanismos cognitivos para o processamento de itens lexicais como *livro* são mais semelhantes ao que ocorre com itens homônimos do que ao que ocorre com itens monossêmicos.

Por fim, destacamos o problema da *maximização excessiva do contexto situacional e cultural* (o *background knowledge*) e *minimização do potencial semântico do léxico* presente em uma abordagem monossêmica. Conforme vimos, essa foi a posição adotada por Searle para explicar os diferentes usos de *open*. Taylor (2005) chama a atenção para o fato de que, se atribuirmos as condições de verdade de um item unicamente à interpretação contextualizada, presumivelmente, outras propriedades lógicas de uma expressão, tal como seu acarretamento e suas pressuposições, também precisariam ser atributos do contexto. Essas são, no entanto, salienta o autor, propriedades que não podem ser inferidas a partir de significados linguísticos (unitários) e de regras de combinação. Por essa razão, sua validade é questionada pelo autor.

Na posição intermediária (ii), a da *polissemia extrema*, entende-se que os sentidos individuais são armazenados na memória de longo termo e que os sentidos relacionados compartilham uma representação nuclear ou ocupam espaços distintos e sobrepostos de uma estrutura semântica (BROWN, 2008) – *abordagem polissêmica*. Vimos que os resultados de Sandra e Rice (1995), Klein e Murphy (2002) e Brown (2008) evidenciam que os sentidos relacionados não são representados da mesma maneira que os não-relacionados e que, além disso, quanto mais próximos semanticamente estão os sentidos

polissêmicos, mais rapidamente eles são processados. Tendo em vista esses fatos e as críticas às abordagens homonímica e monossêmica, podemos concluir – até então – que a polissemia é a abordagem ideal. Vamos avaliar essa posição tomando como exemplo a postura adotada por Lakoff (1987) e, como referência, a avaliação feita por Taylor (2006).

Conforme Lakoff, *window (janela)* é um item lexical que pode ser usado no inglês para instanciar “diferentes categorias de coisas” (LAKOFF, 1987, p. 416), ou seja, diferentes entidades em variadas situações. Tomando como referência os contextos analisados por Lakoff, para ilustrar, criamos alguns que são aplicáveis tanto ao inglês quanto ao PB; são eles: (a) se alguém joga uma pedra e quebra a janela, o que é quebrado é (provavelmente) o vidro, que é parte da janela; (b) se alguém pinta a janela, o que está sendo pintado é (provavelmente) a estrutura de apoio da janela, que geralmente é de madeira ou de ferro; (c) se alguém senta na janela, ocupa uma porção específica do apoio da janela, aquele mesmo que usualmente é pintado; (d) quando dizemos que alguém irá entregar a janela que foi encomendada, espera-se que a pessoa entregue a estrutura toda e os vidros, ou somente a estrutura, mas dificilmente só os vidros; (e) outra possibilidade, ainda, é uma pessoa passar pela janela, situação que não põe em foco a parte física da janela (estrutura e vidros), mas a abertura de uma janela. Para Lakoff, o fato de *window* poder ser utilizado em contextos distintos como os descritos em (a-e) justifica um tratamento polissêmico ao item, visto como associado a múltiplos sentidos distintos, mas relacionados (cf. TAYLOR, 2006). O que observamos aí é uma posição radicalmente distinta daquela apresentada anteriormente para *open*, na perspectiva de Searle. Postura que dá voz a questionamentos como: podemos entender como razoável que os falantes de uma língua mantenham armazenado na memória de longo termo um grande número de leituras polissêmicas e que a compreensão de distinções banais de sentidos requerem um detalhado processo de desambiguação? (cf. TAYLOR, 2006). Os críticos de posturas como a de Lakoff argumentam ser uma estratégia de descrição demasiadamente tolerante à polissemia, por gerar o que muitos consideram uma proliferação excessiva de sentidos. Seguindo sua perspectiva, o número de sentidos associado a um item seria potencialmente infinito, assim como o número de situações em que ele pode participar e gerar conjuntos diferentes de condições de verdade (cf. TAYLOR, 2006). Defendendo posição contrária à chamada “rampant

polysemy” (CUYCKENS e ZAWADA, 2001, p. xvii) estão Sandra (1998), Tyler e Evans (2001, 2003).

Experimentos psicolinguísticos evidenciam (ex. BROWN, 2008) que os pares de sentido relacionados (mesmo através de uma relação distante) são acessados mais rápida e corretamente do que pares de sentidos totalmente distintos, conforme dissemos acima. Esse fato contribui para que se entenda que há apenas uma “entrada” no léxico mental para sentidos relacionados e que essa relação é percebida pelos falantes (BROWN, 2008). A partir de tais resultados, Brown (2008) sugere a seguinte organização: sentidos não-relacionados (homônimos) são vistos como tendo representações semânticas distintas, podendo ser considerados constituintes de categorias de sentidos distintas, ou “entradas” distintas no léxico mental – mesma estratégia seguida por dicionários tradicionais; diferentemente, sentidos relacionados (polissêmicos) são vistos como partes de uma mesma porção de representação semântica, com nuances de sentido interagindo em uma categoria de sentidos complexa, estrutura que pode ser descrita como uma única “entrada” para o item lexical polissêmico, com “subentradas” para cada sentido diferente e relacionado.

Escolher entre uma estratégia homonímica ou polissêmica – embora gere vários questionamentos sobre o tipo de relação percebida entre os sentidos, conforme veremos – é relativamente simples, tendo em vista as desvantagens de uma abordagem homonímica para sistemas de PLN (destacadas na Seção 2) e a falta de realidade psicológica para uma estruturação dos sentidos unicamente homonímica (cf. KLEIN e MURPHY, 2002; BROWN, 2008). A opção pela polissemia pode ser metodologicamente executada a partir da seguinte estratégia: opta-se pela abordagem polissêmica sempre que algum fator de coerência for identificado entre os sentidos. A escolha entre as abordagens polissêmica e monossêmica, por outro lado, é considerada mais completa tanto teórica quanto metodologicamente, visto que envolve perceber distinção entre aspectos semânticos que geram sentidos polissêmicos e manifestações de um único sentido que é especificado *on-line* (cf. CUYCKENS e ZAWADA, 1997). Não se sabe claramente quando precisamos distinguir entre informações propriamente semânticas e informações contextualmente determinadas (cf. CUYCKENS e ZAWADA, 1997). Sabe-se, no entanto, que, ao pensar como Lakoff para o caso de *window*, corre-se o risco de cair no excesso de proliferação de sentidos para um item (cf. TAYLOR, 2006). Experimentos psicolinguísticos (ex.: KLEIN e MURPHY, 2002; BROWN, 2008) inspiram abordagens da

SLC a defender a estratégia polissêmica, mas nem sempre em versão tão extremada quanto a de Lakoff (1987) (ex.: TYLER e EVANS, 2003, 2004; TAYLOR, 1995, 2003, 2006; CRUSE, 1995; CRUSE, 2000; CROFT e CRUSE, 2004; GEERAERTS, 2006). A estratégia de representação através da indicação explícita de conexões entre os diferentes sentidos associados a um mesmo item é o procedimento considerado como o que mais facilita que as “pessoas passem de um sentido a outro” (BROWN, 2008); pode-se entender, por consequência, que essa é a estratégia que mais facilita a simulação automática desse processo por sistemas de PLN.

Chegamos ao final desta seção com a certeza de que não só as estratégias homonímica e monossêmica apresentam problemas, mas também a polissêmica; ainda assim, alguns posicionamentos já podem ser tomados. O primeiro é que a adoção de uma estratégia exclusivamente homonímica deve ser, de antemão, excluída do escopo deste trabalho por ser adequada somente para situações em que não percebemos fatores de coerência. A questão que permanece em aberto é, ao percebermos algum fator de coerência entre os sentidos, ou seja, grau maior ou menor de *similaridade* (noção que ainda precisa ser detalhada), que estratégia adotar: polissêmica ou monossêmica?

3.5.2 UMA ESTRATÉGIA ALTERNATIVA

Apresentadas vantagens e limitações das diferentes abordagens de estruturação do léxico sob uma perspectiva semasiológica, passaremos a tecer os comentários que levarão ao delineamento da estratégia mais adequada aos propósitos desta tese. As questões levantadas até este ponto sobre a economia de armazenamento do conteúdo lexical seguem em duas direções: (i) as que defendem a representação via sentidos unitários, que economizam armazenamento, mas que necessitam de processamento de informações contextuais; e (ii) as que defendem a polissemia, que economiza processamento das variações contextuais, mas que aumenta a demanda de armazenamento e requer procedimentos de desambiguação poderosos (TAYLOR, 2006). Vimos que essa segunda distinção é a que é tida como mais problemática de ser estabelecida em termos teóricos, isso porque “não sabemos em que grau de profundidade um falante articula [uma rede de] polissemia” (LANGACKER, 1992, p.

267). Além disso, os falantes podem divergir muito a esse respeito (CUYCKENS e ZAWADA, 1997; TUGGY, 2006; ALLWOOD, 2003; GEERAERTS, 2006a; SILVA, 2006; KLEPOUSNIOTOU, 2007; BROWN, 2008). A percepção sobre a existência de maior ou menor influência do contexto na delimitação do sentido, na verdade, conforme Taylor (2006), pode ser considerada, do ponto de vista da descrição do conhecimento, como uma questão de definição sobre o nível de generalidade/especificidade da delimitação dos sentidos associados a um item. Sob essa perspectiva, as questões metodológicas que envolvem a tarefa de delimitação do sentido ficam facilitadas, e veremos em seguida as razões.

No contexto da SLC, a falta de distinção precisa entre polissemia e monosssemia não é necessariamente um problema (cf. ex.: CUYCKENS e ZAWADA, 1997; TYLER e EVANS, 2004; TAYLOR, 1995; 2006). E o motivo é que a necessidade de detalhamento do significado lexical varia dependendo do propósito da descrição a ser realizada. Um dos objetivos pode ser a representação do significado lexical em termos de generalização (esquemas, definições unitárias) que capturam maximamente as informações semânticas entre os muitos usos. Outro objetivo pode ser a representação em termos de informação semântica específica com instâncias (e potenciais sobreposições) de esquemas gerais (TYLER e EVANS, 2004). Taylor (2006) salienta que é sabido que, para uma descrição lexical ser funcional, ela precisa apresentar estratégias representacionais “elegantes”, ou “econômicas”, mas que não há motivos suficientes que exijam que se estabeleça um limite para a quantidade de polissemia que o léxico pode tolerar ou para decidir, *a priori*, qual é o grau “ótimo” de polissemia a ser codificado (TAYLOR, 2006). Seguindo essa linha de raciocínio, procurando evitar os problemas que ambas as estratégias, monossêmica e polissêmicas, suscitam quando tomadas em seu sentido estrito, Taylor (2006, p. 54) sugere uma perspectiva alternativa para a descrição dos sentidos. Trata-se da estratégia do “meio-termo” (TAYLOR, 2006, p. 58), que está baseada em dois compromissos: (i) *ir para o meio do caminho* e (ii) *ter ambas as direções*. Conforme apresenta o autor, ela foi construída considerando conclusões de Langacker (1990), Tyler e Evans (2003), além de Geeraerts (1992).

O primeiro compromisso, o de *ir para o meio do caminho*, procura reduzir a polissemia incluindo representações que generalizem ao máximo as nuances de sentido específicas de cada ocorrência de um item lexical. Trata-se de uma tentativa de “excluir do léxico aquelas nuances de sentido que podem ser derivadas ou previstas a partir de

princípios gerais” (TAYLOR, 2006, p. 55). Ao mesmo tempo, o modelo permite polissemia em casos em que sentidos abstratos unificadores não sejam adequados. Dito de outro modo, o que o autor propõe é seguir em dois caminhos: (a) buscando o mínimo possível de polissemia, mas permitindo (b) o máximo necessário, abandonando tanto a polissemia excessiva quanto a monossemia estrita. O compromisso de seguir em direção ao meio do caminho prevê a redução do excesso de polissemia, deixando que as sutis diferenças de sentido sejam inferidas a partir da combinação dos itens coocorrentes em um contexto sentencial e do conhecimento de mundo geral. Seguindo esse compromisso, nem todas as diferenças de sentido precisam estar consagradas no léxico mental como sentidos polissêmicos distintos, mas podem.

O segundo compromisso, o de *ter ambas as direções*, permite tanto uma descrição semântica fina, com alto grau de granularidade, quanto uma descrição mais geral, incluindo sentidos gerais (esquemáticos) que capturam o que há de comum entre os grupos de sentidos específicos.

A proposta de Taylor (2006) para dar conta desse compromisso bidirecional prevê que os sentidos, descritos em vários níveis de generalidade (esquematicidade), sejam representados em um constructo em formato de *rede* com sentidos mais específicos funcionando como elaborações de sentidos mais esquemáticos. Adotar como estratégia descritiva o modelo de rede, destaca Taylor (2006), “é claramente um avanço para a solução do ‘meio termo’, visto que o analista não é forçado a escolher entre uma série de sentidos distintos e um sentido esquemático global, mas simplesmente incorpora ambos” (p. 58). Mais uma das vantagens mencionadas para modelo de redes é o fato de que ele possibilita lidar facilmente tanto com casos de polissemia menos produtiva (irregular) quanto mais produtiva (regular). Sentidos específicos, acrescenta Taylor (2006), mesmo quando previsíveis a partir de princípios gerais, podem ser incluídos em uma representação desse tipo, sem danos à estrutura do constructo ou a seus princípios teóricos. Além disso, se algum sentido não atestado pelo uso não tiver sido registrado na rede, ele pode ser captado de maneira relativamente simples a partir de inferências baseadas nos sentidos já representados. Para entendermos a liberdade de representação que uma estrutura de redes possibilita, vamos considerar o seguinte exemplo:

1. Isabel foi correndo até aquele *banco*; e Cristina *fez o mesmo*.

Para a interpretação de (1), sentença com estrutura tipicamente utilizada em um dos testes linguísticos de identificação de polissemia (teste de restrição de identidade), o mais provável é selecionarmos o mesmo sentido (seja qual for) para o elemento anaforizante e para o anaforizado. Entendermos, por exemplo, que Isabel correu a um estabelecimento bancário e que Cristina correu até determinado *banco* de areia é improvável, a menos que a intenção seja fazer um jogo de palavras (cf. CROFT e CRUSE, 2004). Se olharmos para (1) considerando os usos de *banco* como ‘instituição financeira’ e como *banco* ‘de areia’, é difícil pensarmos em um elemento unificador entre eles que seja mais específico do que *entidade, coisa, local*. A partir dessa constatação, poderíamos concluir que não há vantagens na inclusão de *banco* em uma única estrutura de representação semântica, ou seja, como uma única estrutura fonológica que mapeia para uma única rede, visto que não percebemos qualquer fator de coerência (cf. CROFT e CRUSE, 2004). A conclusão a que se chega, a partir daí, é a de que há duas estruturas fonológicas distintas e que, portanto, mapeiam para duas redes também distintas – posição tradicional para situações de homonímia, conforme referimos anteriormente. Não há dúvidas de que essa posição é adequada e verdadeira para a maior parte dos contextos em que *banco* ocorre, no entanto, ao nos depararmos com uma sentença como (2), a tendência é, no mínimo, questionarmos essa conclusão:

2. A instituição financeira *banco* é semelhante a um *banco* de areia que encontramos no mar; eles controlam, respectivamente, o fluxo de dinheiro e de água¹⁶.

A sentença (2) convida-nos a focar em propriedades semânticas de semelhança entre os dois tipos de *banco*, ‘instituição financeira’ e ‘*banco* de areia’; força-nos a acessar (ou até mesmo a criar) um sentido esquemático unificando as duas leituras de *banco* (TAYLOR, 1995). Se o objetivo da representação for dar conta também de usos pouco frequentes como (2), será importante que ambos os usos sejam inseridos em uma única rede, e é aí, justamente, que reside a grande vantagem do modelo e a justificativa de sua flexibilidade: é útil à representação do sentido em diferentes níveis de generalização. A definição sobre até onde se vai para cima ou para baixo na rede será determinada pela intenção de representar sentidos mais específicos (e mais

¹⁶ Adaptado de Taylor (1995, p. 289).

polissêmicos) ou sentidos mais genéricos (e mais monossêmicos) em razão de fatores teóricos ou práticos.

Não há, no entanto, apenas vantagens no modelo de rede. Conforme salienta Taylor (2006, p. 59), ele não pode ser considerado como uma “palavra final”. Suas vantagens são várias, como mencionamos, mas é preciso que sejam ponderadas as críticas que o modelo tem recebido tanto em termos metodológicos quanto em sua natureza conceitual (cf. SANDRA e RICE, 1995; TAYLOR, 2006; TYLER e EVENS, 2001; 2003).

Os problemas metodológicos mencionados na literatura para o modelo de redes giram em torno da falta de clareza na apresentação dos critérios, um problema recorrente em vários níveis: (i) a falta de metodologia explícita para distinguir entre os sentidos (cf. SANDRA e RICE, 1995; TYLER e EVENS, 2003); (ii) subjetividade da estrutura (SANDRA e RICE, 1995; TAYLOR, 2006; TYLER e EVENS, 2001; 2003); (iii) falta de consenso sobre as convenções para representação dos resultados das análises. A falta de metodologia clara para distinguir entre os sentidos é percebida como um dos principais problemas de redes do tipo radial tradicionais, tal como proposta por Brugman (1981) e Lakoff (1979). De acordo com Sandra e Rice (1995), parece haver tantos modelos diferentes para construção da rede quantos teóricos responsáveis por sua construção. Tyler e Evans (2003) concordam sobre o alto grau de subjetividade das redes tradicionais e propõem um modelo de rede que eles denominam “principled” que procura apresentar critérios claros para distinguir entre interpretações que são construídas *on-line* e sentidos distintos que são instanciados na memória de longo termo do falante. Notamos, então, que aquela que foi destacada como a principal vantagem de uma abordagem de redes pode ser considerada também sua principal limitação: a flexibilidade para a inclusão de sentidos em diferentes graus de generalidade e de convenção. Sandra e Rice (1995) e Taylor (2006) fazem referência a esse problema como a “vagueza das representações” reticulares (termos de Sandra e Rice, 1995, p. 94), que envolve tanto a falta de consenso sobre as convenções para o registro dos resultados das análises quanto a possibilidade de a estrutura variar muito de pesquisador para pesquisador, não somente em relação ao número de sentidos associados a um item (seu grau de polissemia), mas também ao conteúdo a ser representado para cada sentido e, além disso, ao tipo de relação que estabelece agrupamentos entre os sentidos (TAYLOR, 1995; 2006; TYLER e EVANS, 2001; 2003).

Como exemplo concreto dessas discrepâncias, são mencionadas as redes de polissemia radiais para *over*, inicialmente construídas por Brugman (1981) e que tiveram reelaborações em Lakoff (1987), Brugman e Lakoff (2006 [originalmente, 1988]), Geeraerts (1992; 2006) e Tyler e Evans (2001; 2003), todas elas com versões distintas em função de propósitos particulares.

O segundo problema destacado para o modelo de redes, que é *de ordem conceitual*, remete ao *status* ontológico das redes (cf. TAYLOR, 1995). Como ilustração, Taylor (2006) apresenta as seguintes questões: as redes têm realidade psicológica? Elas representam hipóteses sobre o conhecimento que os falantes têm sobre os itens lexicais? Em caso afirmativo, como é possível testar a realidade psicológica das redes? Em caso negativo, o que se supõe que uma rede represente? Qual é a relevância de uma rede para a aquisição de língua? Em que ponto da rede um item lexical é acessado para sua produção ou interpretação? Foi por preocupar-se com questões como essas que Sandra e Rice (1995) escreveram o artigo “Network analyses of prepositional meaning: Mirroring whose mind – the linguist’s or the language user’s?”.

Encerramos a Seção 3.5 certos de que não há consenso para a escolha de nenhuma das abordagens analisadas e destacamos duas como as principais causas dessas discrepâncias: (i) a flexibilidade inerente ao significado lexical (seja ela em maior ou menor grau) – já que na língua sempre há espaço para a criatividade, tendo em vista que os sentidos são construídos a partir do uso e que os usos de um item são governados pelas necessidades comunicativas (que são as mais variadas); (ii) embora haja evidências sobre diferentes padrões de adequação representacional dependendo das propriedades dos sentidos em questão, parece-nos plausível a posição de que a escolha de uma das estratégias apresentadas (homonímica, polissêmica, monossêmica, meio-termo) é determinada pelo objetivo do analista, que, em algumas vezes, pode ter a intenção de representar o maior número de instâncias de conteúdo conceitual em uma língua e, em outras, a de explorar níveis com grau maior de abstração do conteúdo conceitual acessível a partir de um item. Tendo em vista fundamentalmente essas questões, concordamos com Taylor (1995; 2006) e argumentamos a favor de uma representação de itens lexicais polissêmicos em formato de rede por percebermos vantagens teóricas e metodológicas.

Do *ponto de vista teórico*, destacamos as evidências de que o conceito de rede é – pelo menos em alguma medida – representativo da organização do léxico mental

humano, embora sua estrutura ideal ainda seja uma incógnita. Uma posição recorrente em SLC, que segue Langacker (1987; 2002), é a de que um item lexical típico instancia uma categoria complexa, de modo que ele não se associa somente a um único sentido, mas a uma variedade de sentidos relacionados, com vários graus de incrustamento. Entende-se que esse tipo de estrutura facilita a realização de rotinas cognitivas de acesso a diferentes porções do conteúdo conceitual a partir de diferentes padrões de ativação (CRUSE, 1995). Para Sandra e Rice (2003), o mínimo que se pode dizer é que o modelo de redes (especialmente as redes de preposições polissêmicas) está minimamente comprometido com um processo psicológico de categorização humana. Do *ponto de vista metodológico*, percebe-se, em uma estrutura de redes, alta flexibilidade em sua construção e, ao mesmo tempo, em seu uso/acesso, que pode ocorrer a partir de diferentes níveis de abstração. Há flexibilidade porque, “à medida 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).

Para concluirmos esta seção, daremos destaque aos seguintes tópicos investigados: (a) os experimentos descritos mostram que temos sentidos registrados em nosso léxico mental em diferentes níveis de especificidade, estabilidade e sensibilidade ao contexto, de modo que a representação ideal é determinada pelo objetivo do analista em relação ao que se pretende registrar; (b) há evidências de que a representação mental não é igual para todos os sentidos (nem para todas as pessoas), de que percebemos (em maior ou menor grau) relação entre sentidos e de que essa percepção facilita o processamento mental; (c) a escolha por uma das estratégias - monossêmica, polissêmica ou homonímica - precisa levar em conta, de um lado, a existência ou não de relação entre os sentidos e, de outro, a participação mais ou menos efetiva do contexto situacional na delimitação do sentido. Infere-se, portanto, que, para que sejamos fiéis à variedade de informações semânticas associadas aos itens lexicais, precisamos de um modelo flexível e igualmente robusto e múltiplo. Por fim, antes de detalharmos os modelos de rede propostos no contexto da SLC, é importante garantirmos que encerramos esta seção sabendo *onde estão os sentidos*: acreditando-se que os sentidos não são propriedades herdadas pelos itens lexicais, mas que são construídos a partir do uso da língua e que é a partir de sua frequência de uso que se dá o registro na memória

de longo termo, chega-se à conclusão de que é no léxico mental e no uso que devemos buscar essa porção do conteúdo conceitual acessível a partir das unidades linguísticas.

3.6 AS REDES DE POLISSEMIA

Definida a opção por um modelo representacional baseado no constructo das redes, nesta seção, serão apresentados diferentes tipos de rede com o objetivo de investigar que informações semânticas cabem em uma rede de polissemia. O olhar para a semântica dos itens lexicais polissêmicos a partir desses modelos tem sido posição frequente em SLC. Sabe-se que, em linhas gerais, as *redes de polissemia* buscam modelar conteúdo lexical em termos de uma estrutura que reflita o pressuposto de base da SLC: o estudo do léxico como um “sistema coordenado” (Rice, 1993, p. 206). Em uma rede de polissemia, um item lexical é tratado como uma categoria conceitual, que se subordina a uma variedade de sentidos relacionados, isto é, polissêmicos (TYLER e EVANS, 2003). Ao falarmos de rede, o modelo gráfico talvez seja o primeiro que venha à nossa mente: um constructo composto por *nós*, representando os sentidos, e *arcos*, representando as relações entre eles – elementos que podem ser estruturados visualmente de diferentes formas, por exemplo:

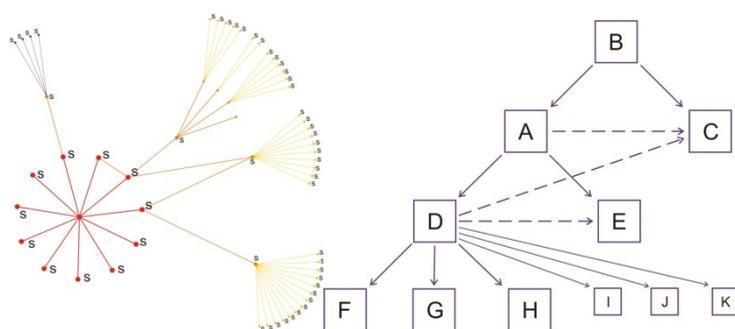


Figura 10 - Configurações de redes

Esse contexto gráfico ou formal não é, entretanto, o contexto que pretendemos enfatizar nesta parte do trabalho; ele será motivo para reflexão na seção dedicada à formalização computacional de nossa proposta. Descrições sobre como as configurações de redes podem ser visualizados adquirem, aqui, *status* secundário, e observações sobre

essas questões servem como um recurso explicativo para facilitar a compreensão dos conceitos apresentados. O foco é, portanto, o constructo de rede enquanto modelo teórico.

São várias as propostas de redes para tratar do léxico em perspectiva onomasiológica ou semasiológicas. No contexto da LC, a posição mais frequente é a que acredita no que Sandra e Rice (1995, p. 101) chamam de “perspectiva isomórfica” entre língua e mente. Entender que as redes linguísticas estabelecem um isomorfismo com sua contraparte cognitiva é pressupor que nelas se vê representada uma imagem da forma como a informação semântica associada aos itens está armazenada no léxico mental (SANDRA e RICE, 1995). A crença no isomorfismo pode ser completa, “isomorfismo radical”, ou parcial, “isomorfismo funcional” (SANDRA e RICE, 1995, p. 102).

A posição de isomorfismo radical olha para as redes fundamentalmente com interesse nos princípios que estão por trás, nos princípios motivadores do que se vê representado nessas estruturas gráficas. Nessa perspectiva, buscam-se estratégias para se estabelecer correspondência exata entre as redes linguísticas e sua contraparte mental (SANDRA e RICE, 1995). Lakoff, Langacker, Miller e Johnson-Laird são importantes representantes desse olhar na Ciência Cognitiva. Miller e Johnson-Laird (1976), tratando das redes com enfoque onomasiológico, inspiraram a construção das *wordnets* e foram marcantes para o fortalecimento de uma interpretação isomórfica da análise linguística por defenderem um conceito de rede que contribuiu para que a conotação psicológica desses modelos passasse a ser algo natural nos diversos contextos (SANDRA e RICE, 1995), não só nas ciências cognitivas, mas também em PLN. Tratando da semântica da polissemia, perspectiva semasiológica do léxico, Lakoff (1987), ao fazer uso de noções como *esquemas imagéticos* e *transformações por esquemas imagéticos*, e Langacker (1987), ao discutir as *relações de categorização*, estão claramente ultrapassando o nível linguístico e tratando apenas das estruturas mentais correspondentes às redes (SANDRA e RICE, 1995) – modelos que apresentaremos a seguir.

Em posição menos radical, no *isomorfismo funcional*, estão autores como Sandra e Rice (1995), que defendem que a correspondência psicológica de uma rede de polissemia não é imediata, visto que ela é um dispositivo estático; assim, sua relevância psicológica inerente reside no nível *representacional* (noção entendida em consonância com a Psicologia). As redes são tratadas como dispositivos representacionais que

captam aspectos da representação mental do uso da língua; isto é, a realidade psicológica de uma rede é justificada pela crença de que a representação linguística (descrições de língua) se associa - em alguma medida - a uma representação mental (estruturas cognitivas). Conforme afirmam Sandra e Rice (1995), os esforços desse tipo de estudo voltam-se para as informações que estão sendo representadas pelas redes, ou seja, para o conteúdo das redes ou dos nós que estão dentro delas. O foco, explicam os autores, está na perspectiva funcional de representação dos sentidos, sendo que sua natureza real permanece desconhecida, embora se acredite na compatibilidade da rede com as estruturas cognitivas correspondentes. Mesmo nessa perspectiva moderada de isomorfismo, sai-se do campo da linguística propriamente dito e ingressa-se no domínio da psicologia cognitiva, salientam Sandra e Rice (1995). Nesse domínio, investiga-se a validade representacional (em sentido psicológico) das redes, e considera-se importante determinar em que medida uma rede de polissemia corresponde a uma representação mental. Com foco na sincronia do fenômeno da polissemia explicado via redes, em Sandra e Rice (1995) vemos argumentos a favor de sua posição a partir de estudo empírico. Seus resultados revelam que muitas das relações comumente propostas pelas teorias são facilmente percebidas pelos falantes. A partir de constatações como essas, a construção de uma rede de polissemia retrata a realidade psicológica da polissemia sob um viés sincrônico.

Estudos realizados sob ambas as perspectivas trazem resultados úteis aos nossos propósitos, visto que são plenamente compatíveis e complementares, conforme temos procurado mostrar; entretanto, a visão de *isomorfismo funcional* é a que melhor representa a que adotaremos. Para fazer aparecer como, ao nosso entender, é o modelo de rede de polissemia mais apropriado aos nossos propósitos, é preciso revisar modelos que nasceram antes e que serviram de inspiração para a proposta. É com esse intuito que apresentamos as redes radiais e esquemáticas, constructos que - seguindo uma perspectiva radical de isomorfismo - seguem o pressuposto de que as extensões semânticas instanciadas na língua não são acidentais, mas são motivadas por princípios psicológicos de categorização (SANDRA e RICE, 1995). Entende-se que são esses os princípios que tornam possível percebermos que determinado uso de um item é parte de uma categoria conceitual. O objetivo, nesse tipo de abordagem, é representar a natureza dos princípios de extensão semântica, e Lakoff (1987) e Langacker (1987) propõem estruturas de redes que representam um estudo diacrônico que mapeia a

evolução do significado lexical. Além dessas propostas clássicas, seguindo perspectiva similar, apresentaremos as redes multidimensionais, propostas por Geeraerts (1993; 2006), conforme descreveremos a seguir.

A primeira proposta de rede de polissemia foi a de Brugman (1981), estendida e popularizada por Lakoff (1987) para tratar da polissemia da preposição *over* seguindo uma *estrutura cognitiva radial*. A proposta, em sua versão clássica, apresenta uma categoria como múltiplos sentidos organizados radialmente em uma estrutura do tipo centro-periferia: no centro, está o *protótipo* (que pode ser mais de um) e, em torno dele, estão as suas *extensões*, que podem se tornar *subcentros* para outras relações centro-periferia (LAKOFF, 1987). Uma rede radial tem como objetivo indicar que há um encadeamento de relações entre os elementos que compõem uma categoria linguística complexa e que essas relações partem de um elemento central e se estendem a elementos menos nucleares (LAKOFF, 1987) que dele derivam por diferentes mecanismos de extensão de sentido. A figura 11 ilustra o modelo.

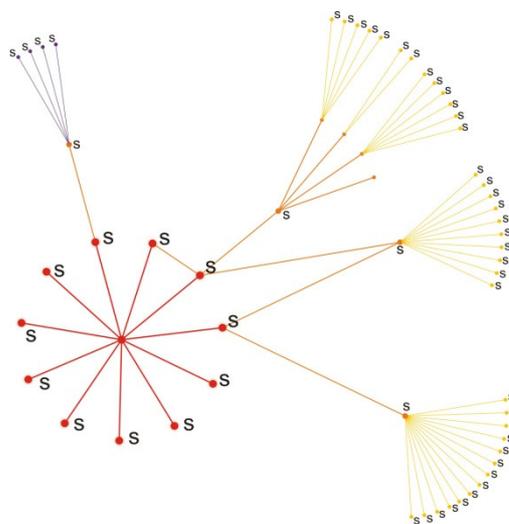


Figura 11 - Ilustração do modelo de rede radial

A figura 11 ilustra uma rede radial altamente ramificada, mas nem sempre ela precisa ser assim. Os pontos representam os sentidos, e as linhas ou arcos, as relações entre eles. O ponto central é o elemento prototípico, e a ele todos os demais elementos estão relacionados direta ou indiretamente. Uma propriedade interessante nessa estrutura, que mostra algum indício de simultaneidade de conexões, é o fato de que nós

terciários podem estar relacionados com múltiplos nós secundários (SANDRA e RICE, 1995).

O modelo de rede radial tem recebido fundamentalmente dois tipos de críticas: (a) o excesso de polissemia, por adotar uma representação que não prevê distinção entre usos gerados *on-line* e usos mais estáveis e abstratos (ex. SANDRA e RICE; TYLER e EVANS, 2003), conforme referimos ao descrever a perspectiva de Lakoff para o verbo *open*, na seção anterior; (b) por estar estruturado em uma arquitetura tipicamente bidimensional, por prever relação entre um centro prototípico (ou mais de um) e usos mais ou menos próximos desse(s) centro(s), dele(s) direta ou indiretamente derivados, sendo a representação de sentidos sobrepostos algo limitado (ex. SANDRA e RICE, 1995; TYLER e EVANS, 2003; GEERAERTS, 2006, SOARES DA SILVA, 2006). Uma forma comumente usada para avaliar as redes é a investigação sobre sua realidade psicológica a partir de experimentos focados em aquisição lexical. O estudo de Rice (2003), sob essa ótica, traz mais uma crítica ao modelo de Brugman e Lakoff: foram percebidas evidências de que a aquisição lexical das crianças não segue a lógica radial, partindo do que se presume ser o protótipo, mas inicia com base nos usos que, em termos do modelo de rede radial, são considerados periféricos.

Observa-se na literatura grande influência das redes radiais; no entanto, a *estratégia do meio-termo*, que apresentamos na seção anterior, defendida por Taylor (1995; 2006), tomou como base o modelo de *rede esquemática*, proposto por Langacker (1987), estrutura que tem grandes méritos por prever maior flexibilidade do conteúdo conceitual. Trata-se de um modelo representacional considerado válido para descrever a construção de conexões entre os elementos da língua como um todo, não somente para itens lexicais polissêmicos. Seguindo princípios de isomorfismo pleno do mesmo modo que Brugman e Lakoff, mas já com a preocupação forte de incluir sobreposições de sentidos em diferentes dimensões da rede, o modelo cognitivo de Langacker é baseado no conceito de *esquema*. A estrutura nuclear de uma rede esquemática pode ser representada graficamente a partir de um modelo em três pontos: (a) *protótipo* (sentido base e instanciação do sentido esquemático), (b) *esquema* (sentido geral e abstrato que *elabora* os demais) e (c) *extensão* (instanciação do esquema e extensão do protótipo), formando o chamado triângulo de categorização de Langacker (cf. TAYLOR, 2003). Os sentidos estão ligados horizontalmente por relações de similaridade e verticalmente por relações de esquema e instância. Um exemplo clássico é *tree*; fazendo a adaptação para o

PB, ilustraremos com *árvore*, que tem como um primeiro sentido, o prototípico, ‘vegetal lenhoso alto’; como segundo sentido, a extensão, ‘árvore genealógica’; e como elemento comum, sentido esquemático, a noção geral ‘entidade ramificada’. É o que mostra a figura 12.

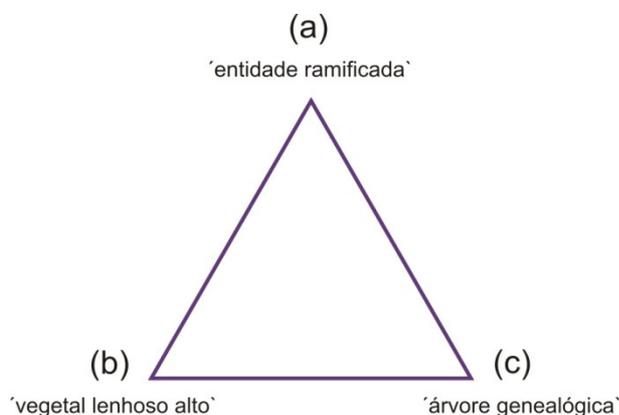


Figura 12 – Modelo de rede esquemática

O sentido *prototípico* (básico) é incorporado em uma rede esquemática como um caso especial e como um elemento útil para representar diferenças de incrustamento e de saliência entre os sentidos, não mais como o elemento centralizador da categoria como em redes radiais. Em relação horizontal com o elemento prototípico, está o sentido considerado a sua extensão, gerado por um processo em que o falante percebe alguma similaridade entre ele e o protótipo (LANGACKER, 2002). Entende-se haver uma conexão íntima entre o crescimento da rede horizontalmente (extensão) e verticalmente (esquematização) (LANGACKER, 2002). Nessa perspectiva, entende-se que o processo de extensão ocorre porque um falante percebe alguma similaridade entre um sentido básico e o sentido estendido. A percepção dessa similaridade representa o que há de comum entre os sentidos em questão, formando um conteúdo esquemático. O autor explica que o sentido relacionado por similaridade por si não é cognitivamente saliente e nem sempre pode ser considerado uma unidade de sentido estabelecida, isso porque pode ser somente uma ocorrência “efêmera”. Uma evidência da existência de saliência da extensão é quando ela se torna base para outras relações de extensão ou quando funciona como conteúdo esquemático para outras porções da rede. Do mesmo modo, a existência de um nível esquemático em seu modelo não é visto como um nível saliente e necessariamente acessado pelos falantes para a compreensão das variações contextuais.

Ao contrário do que acontece nas redes radiais, a rede esquemática foi pensada desde seu princípio para dar conta de propriedades dinâmicas da língua, podendo evoluir continuamente, de modo que o triângulo básico apresentado na figura 12 pode ser ampliado à medida que aumenta a quantidade de sentidos relacionados. É o que mostramos na figura 13, adaptada de Langacker (1987, p. 283), em que as linhas contínuas representam as relações entre o *sentido esquemático* (B) e suas *instanciações*, que são o *sentido prototípico*, inicialmente o A, e sua *extensão*, inicialmente o C; as linhas pontilhadas indicam relação entre *protótipos* e *extensões*.

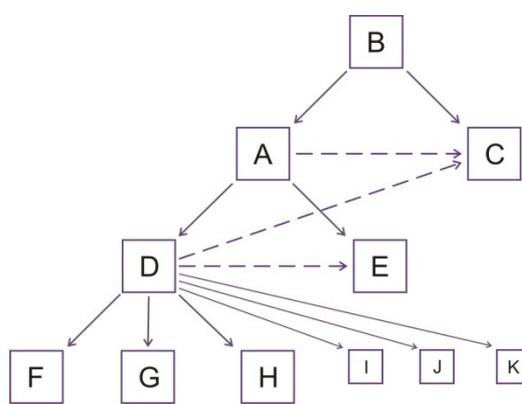


Figura 13 – Rede esquemática complexa

Langacker (2002) defende que a representação de redes esquemáticas fornece todas as informações necessárias: um inventário de sentidos que descrevem o conjunto de usos convencionais de uma expressão linguística; as relações que esses sentidos estabelecem com sentidos esquemáticos para expressar generalizações implicadas em um conjunto de sentidos e especificações de distância e de saliência cognitiva. Conforme Sandra e Rice (1995) e Taylor (1995), as redes esquemáticas apresentam uma grande vantagem para a estruturação de uma categoria polissêmica: a heterogeneidade, adequando-se à *estratégia do meio-termo*, sugerida na seção anterior. A presença de três tipos de nós (protótipo, esquema, extensão) concilia abordagens radicalmente monossêmicas, que procuram determinar um único sentido abstrato que cubra todos os usos de um item, e abordagens polissêmicas, que podem fazer uso de conteúdos esquemáticos altamente abstratos.. Além disso, os polissemistas, extremos ou não, podem também encontrar satisfação no fato de que os múltiplos sentidos, mais específicos ou mais gerais, podem ser posicionados de diferentes modos. Como

ilustração, Langacker constrói redes esquemáticas para *uncle*, em *Tom has an uncle* (Tom tem um tio), e para *ear*, em *Tom has two ears* (Tom tem duas orelhas/espigas). Se a rede de sentidos para *uncle* for composta de um nó esquemático *parent's brother*, isso não impede que ela contenha também nós mais específicos, como *mother's brother* e *father's brother*, todos elementos com graus de saliência distintos e, portanto, com potenciais de ativação distintos. Para situações em que o sentido é considerado vago (monossêmico), o que vai ocorrer, explica Langacker, é que o esquema será mais saliente do que suas instanciações; o inverso ocorre para sentidos claramente delimitados, sendo as instâncias mais salientes do que o conteúdo esquemático, conforme vemos na figura 14.

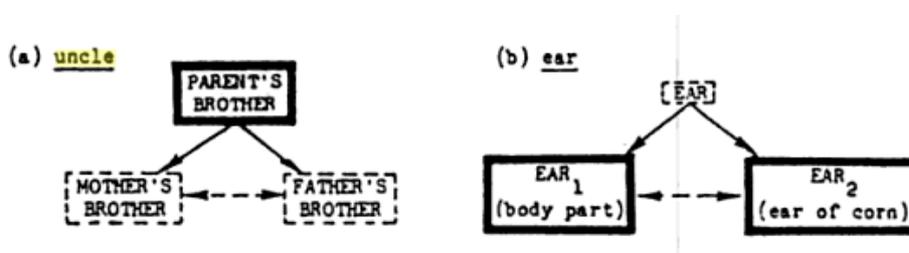


Figura 14 - Rede esquemática *uncle* e *ear*
(LANGACKER, 2002, p. 270)

Na figura 14, para representar possíveis diferenças de saliência, Langacker (2002) faz uso de linhas mais fortes em contraste com linhas pontilhadas. A rede esquemática (a) representa uma situação em que o elemento esquemático *parent's brother* é o mais saliente, de modo que, quando se fala em *uncle*, não é tipicamente relevante saber se é o irmão da mãe ou do pai. Na rede (b), o elemento esquemático está pontilhado e em tamanho menor, representando que ele está em posição de relevância inferior em relação a suas instanciações. Levando em consideração o fato de que falantes divergem sobre seus julgamentos de similaridade, Langacker considera importante incluí-lo, embora afirme que, em razão de seu alto grau de abstração, provavelmente sua construção seja pouco útil para a comunicação. Apresentados casos típicos dos dois extremos do *continuum* de multiplicidade de sentidos, o autor passa então a tratar de exemplo de polissemia e salienta que é para esse tipo de fenômeno – que está em posição intermediária do *continuum* – que sua rede tem mais contribuições a dar. Como ilustração, Langacker cita *talk*, em *Tom is talking*, que pode selecionar, por exemplo, um

sentido específico remetendo a uma situação de apresentação oral (palestra), um sentido que remete a uma conversa e, ainda, um sentido esquemático que abstrai o contexto e foca na atividade verbal em si. O sucesso de uma sentença como *Tom is talking and Bill too*, conforme explica Langacker, depende de qual dos contextos vai prevalecer na interpretação: se Tom estiver dando uma palestra e Bill estiver conversando com um amigo, a sentença poderá ter sucesso, visto que, em uma das ocorrências, é o esquema que é ativado para o processo de construção do sentido e, na outra, é uma de suas instâncias; por outro lado, se a ênfase for o aspecto social dos dois eventos, os conteúdos específicos são ativados em detrimento do esquema, fato que deixa a sentença problemática (seu grau de desvio reflete a distância cognitiva entre os dois sentidos).

A partir dos modelos radial e esquemático, Geeraerts (2006 [originalmente, 2001]) apresenta um modelo multidimensional que, fazendo uso de uma estrutura livre, incorpora as ligações entre elementos prototípicos e derivados, porém dá ênfase para as dimensões múltiplas e coocorrentes que constituem uma categoria polissêmica. Conforme o autor, seu objetivo é evitar as armadilhas do debate polissemia/monossemia e dar conta da natureza de influência mútua de análise do léxico sob as dimensões extensional e intensional de análise semasiológica e entre as dimensões responsáveis pela extensão do sentido *generalização, especialização, metáfora, metonímia*, etc. Nas palavras do autor, busca-se dar conta da “natureza multidimensional das extensões semânticas, [visto que] a estrutura de uma categoria polissêmica é caracterizada pela variação de combinatórias semânticas entre as múltiplas dimensões” (p. 178). Talvez a principal particularidade da proposta seja o fato de que, conforme Geeraerts (2006), ao contrário do que ocorre em uma rede radial à la Lakoff (1987), os nós em sua rede não constituem necessariamente sentidos distintos de uma categoria (embora isso possa ocorrer), mas componentes conceituais que coocorrem em várias conexões. Geeraerts enfatiza que o que está em questão em seu modelo não é a parte gráfica de uma rede, mas o estudo empírico envolvido, que é o fato de que as estruturas de unidades linguísticas polissêmicas requerem atenção específica à sua natureza multidimensional.

O tipo de relação de sentido identificado segue a terminologia e as noções exploradas desde a semântica diacrônica (ex.: similaridade, especialização, generalização, metáfora, metonímia) (GEERAERTS, 2006). Para entendermos a proposta

de Geeraerts, basta observarmos parte de sua análise para o verbo *vers*, que tem comportamento similar ao adjetivo *fresco*, em PB. Como exemplo, selecionamos as relações de similaridade e de especialização¹⁷: (a) a relação de similaridade é percebida entre os sentidos ‘*not previously known*’ (‘não conhecido anteriormente’) e ‘*newly made*’ (‘recentemente produzido’), com ‘*raw, inexperienced*’ (‘cru’, ‘sem experiência’); (b) a relação de especialização existente entre os sentidos ‘*retaining its original qualities*’ (‘que conserva as características originais’), ‘*new, in contradiction to being artificially preserved*’ (‘naturalmente conservado’), ‘*fit for drinking*’ (‘adequado para beber’), ‘*pure*’ (‘puro’), de um lado, e ‘*not previously known*’ (‘não conhecido anteriormente’) e ‘*in weaker sense: additional, different*’ (‘adicional’, ‘diferente’), de outro. A multiplicidade de sobreposições e a flexibilidade de dimensões podem ser observadas na análise que o autor apresenta para *fresh*: (i) é possível que um mesmo sentido nuclear se ligue a dois sentidos da rede, como em ‘novo, recente’ para os sentidos parafraseados como ‘ter existência recente’ e ‘diferente do que já apareceu antes’; e, ainda, pode-se identificar uma posição nuclear preenchida por dois sentidos simultaneamente, por exemplo, ‘condição ótima’ e ‘novo, recente’ - para uma situação como *The cake is fresh*. Observamos que é particular a Geeraerts essa concepção de protótipo múltiplo, que é distinta de uma concepção de duplo centro prototípico.

Para exemplificar um estudo que faz uso de redes de polissemia sob uma perspectiva predominantemente sincrônica, destacamos a proposta de Blank (2003), que denominaremos *rede sincrônica de polissemia*. Embora o autor siga a mesma terminologia de abordagens diacrônicas pré-estruturalistas comum à SLC, sua perspectiva é de fato olhar para o efeito sincrônico, elencando, conforme observamos em 3.2, relações sincrônicas entre sentidos associados a um item lexical polissêmico. Blank explica que, “quando dizemos que uma palavra é polissêmica, isso não significa que todos os seus sentidos estão interconectados ou que têm ‘algo em comum’”, ao invés disso, “a polissemia deve ser concebida como uma cadeia ou uma rede de sentidos”, exigindo inevitavelmente uma representação complexa que permita caracterizar precisamente as relações sincrônicas entre os sentidos (BLANK, 2003, p. 272). A figura 15 mostra a síntese da rede de polissemia que o autor constrói para o item lexical *man*, do inglês, a partir de uma série de análises de ocorrências em contexto.

¹⁷ Conforme análise original de Geeraerts (2006, p. 361).

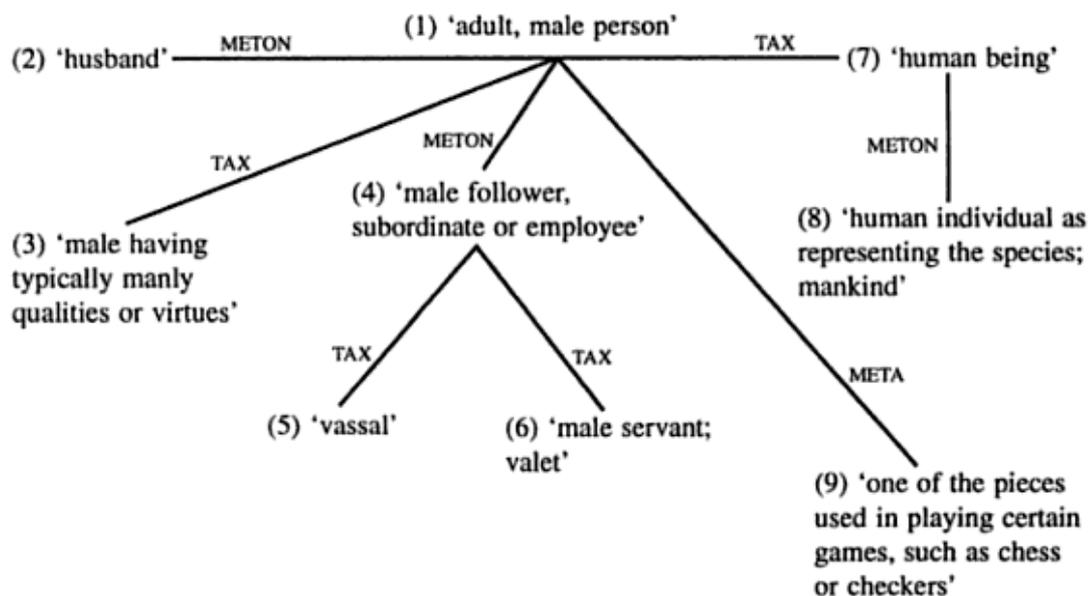


Figura 15 – Rede sincrônica de polissemia
(BLANK, 2003, p. 272)

A rede sincrônica de polissemia proposta por Blank para ilustrar seu inventário de relações tem uma estrutura bastante livre, aliás, o autor não faz qualquer menção a ela. Fica evidente que a estrutura tem um elemento central, de onde a maior parte dos *links* parte ou aonde eles chegam, visto que não são arcos direcionados; trata-se do elemento com maior peso estrutural. Sob o rótulo META, o autor representa relações de polissemia do tipo *metafórica*, que é entendida como derivada geralmente de processos diacrônicos de metáfora e caracterizada como baseada em “maior ou menor saliência de similaridade entre dois conceitos que pertencem a diferentes, e mesmo distantes, domínios conceituais” (BLANK, 2003, p. 268); na rede, META relaciona o sentido de *man* como ‘adulto, ser masculino’ e ‘uma peça de um jogo’, o que em PB chamaríamos de ‘homenzinho’, ‘peão’, ‘bonequinho’ (9). O rótulo METON é empregado de maneira bem ampla para representar relações sincrônicas do tipo *metonímico* baseadas em contiguidade conceitual, isto é, em “co-ocorrências típicas e salientes ou sucessão de elementos em *frames* ou cenários” (BLANK, 2003, p. 268), a *metonímia* é tida como sua contraparte diacrônica principal; na rede, a relação pode ser observada ligando ‘adulto, ser masculino’ (1) e ‘marido’ (2), também entre ‘ser humano’ e ‘indivíduo representante

da espécie humana' (8). O rótulo TAX, que representa a relação de polissemia do tipo *taxonômica*, definida como envolvendo relação de similaridade sincrônica entre dois sentidos relacionados taxonomicamente, de modo que um sentido é hiperônimo do outro; na rede da figura 15, destacam-se as relações entre 'adulto, ser masculino' (1) e 'ser masculino possuidor de qualidades e virtudes típicas' (3). Outro exemplo é a relação que ocorre em um nível mais específico que não diretamente a partir de um sentido central, que é entre 'seguidor masculino, subordinado ou empregado' (4) e 'vassalo' (5) ou, também, 'servente masculino, criado particular' (6).

Blank (2003) enfatiza que os rótulos empregados por ele são, em certa medida, imprecisos, tendo em vista que não descrevem o todo da polissemia de um item. Ao tentar estabelecer um paralelo entre os tipos de mudança semântica propostos por Bréal e os tipos de polissemia que deles podem ser derivados, Blank constata que não é possível se pensar em relação de consequência direta, de modo que a inovação semântica, ao contrário do que acreditava Bréal, é somente um tipo dentre um vasto conjunto de possibilidades de relações sincrônicas entre sentidos polissêmicos. Blank observa que as inovações semânticas se tornam convencionais no nível do discurso e que, eventualmente, a partir daí, elas são integradas no léxico (tornam-se lexicalizadas) como uma regra de língua, sem ou com restrições idiossincráticas, tanto de uso quanto de transferência. Exemplos típicos de transferências metonímicas, como *ham sandwich* para 'cliente que pediu um sanduíche', são considerados parte de um tipo de polissemia não-lexicalizada baseado em regras de discurso restritas a contextos discursivos muito específicos e aplicáveis a um número limitado de contextos, tal como o de nomear o cliente pela comida que ele pediu. Outro tipo de polissemia baseado em regras, mas agora lexicalizada, é o que ocorre com *book* e *school*, do inglês, que funcionam exatamente como *livro* e *escola* em PB. Esse tipo de polissemia é considerado por Blank como baseada em conhecimento enciclopédico de fundo (talvez universal) que nos permite dar foco, no caso de *livro*, ora para o 'conteúdo' (ex.: *Gostei de ler aquele livro*), ora para o "contêiner" (ex.: *Gostei da diagramação daquele livro*) e, no caso de *escola*, o foco pode ser na 'construção', no 'processo de aprendizagem', no 'corpo de alunos e professores', em um 'curso ou um conjunto de cursos' ou, ainda, para 'o departamento da universidade'.

Com a apresentação da rede sincrônica de polissemia de Blank (2003), encerramos esta seção. Iniciamos a seção trazendo perspectivas clássicas em SLC que

olham para o fenômeno da polissemia em busca da explicação para suas motivações cognitivas. As estruturas radial e esquemáticas seguem rigorosamente este objetivo e apresentam modelos mais fechados de estruturação dos sentidos em uma rede. Dessa parte de nossas discussões, é particularmente importante guardarmos informações sobre o tipo de relação percebida entre os elementos de uma categoria: horizontais e verticais, tal como propôs Langacker. Além disso, pensar em polissemia, homonímia e monossemia a partir da análise da relação entre esquema e instâncias é também altamente útil, e voltaremos a fazer uso dessas noções. Da proposta de Geeraerts, interessa-nos, especialmente, a noção de multidimensionalidade, que pode ser explorada a partir de uma estrutura de rede formalmente livre que represente a interação entre dimensões intensionais e extensionais, a partir das múltiplas e simultâneas operações cognitivas responsáveis pela construção das diferentes dimensões (pontos) das redes, e, ainda, podemos pensar na multiplicidade de dimensões que cabem nos pontos de uma rede multidimensional: sentidos, subsentidos, propriedades de sentidos. Por fim, da proposta de Blank, guardamos as seguintes informações (i) não há correspondência direta entre fenômenos sincrônicos (relações de polissemia) e diacrônicos (mudança semântica) envolvidos na variação de sentidos; (ii) os termos comuns à SLC, herdados da linguística histórica, são imprecisos para expressar as relações sincrônicas entre os sentidos; (iv) nem todas as variações de sentidos são lexicalizadas; (v) a estrutura de uma rede sincrônica de polissemia pode ser guiada pelo peso estrutural, de modo que o elemento central é o que possui o maior número de conexões – sobre os diferentes efeitos de saliência entre os sentidos, trataremos na próxima seção.

3.7 SÍNTESE DA SEÇÃO

As discussões realizadas nesta seção possibilitaram entender que a SLC é a abordagem teórica que mais respostas tem a dar para as questões que o estudo da polissemia suscita. A SLC adicionou novos *insights* para o tratamento da estrutura da polissemia: a natureza do agrupamento de uma estrutura polissêmica passou a ser analisada com mais profundidade do que nunca (GEERAERTS, 2006). A partir desse fato,

entende-se que representações lineares são problemáticas para a descrição de itens polissêmicos (ex. dicionários tradicionais, léxicos somente enumerativos) e passa-se a acreditar fortemente em uma estrutura de redes para representar a natureza estruturada das categorias de sentidos (GEERAERTS, 2006).

SEÇÃO 4 - OS SENTIDOS POLISSÊMICOS E SUAS RELAÇÕES

Neste estágio de nossas reflexões, vemos desenhado o cenário teórico da pesquisa, caracterizado o fenômeno semântico em foco, além de definidas as estratégias gerais de atuação para a representação da polissemia através de redes; falta-nos, ainda, caracterizar em detalhes os personagens principais que darão vida e forma à rede de polissemia; são eles: os sentidos e as relações. Ao longo desta Seção 4, investigaremos processos envolvidos na organização interna de uma categoria de sentidos polissêmicos, que, a partir da percepção, (i) delimitam porções de sentidos (4.1); (ii) constroem graus distintos de saliência entre os sentidos (4.2;) e, por fim, (iii) estabelecem relações entre os sentidos polissêmicos (4.3). Descrever os aspectos (i) e (ii) é descrever o que caracterizamos como a dimensão funcional do conteúdo lexical, e o aspecto (iii) como a dimensão estrutural do conteúdo lexical, conforme Geeraerts (2006) e Soares da Silva (2006).

4.1 A DELIMITAÇÃO DOS SENTIDOS

A noção de *delimitação de sentidos* é tema de debates em SLC. Para alguns autores (ex.: GEERAERTS, 2006; CRUSE, 1986; TUGGY, 2003), não se pode falar em delimitação de fronteiras entre sentidos, visto que elas são imprecisas e indefinidas na maioria das situações, justificando o rótulo de fronteiras difusas (*fuzzy*). Por outro lado, há autores (ex.: CROFT e CRUSE, 2004) que defendem a construção de fronteiras entre sentidos como uma das rotinas cognitivas responsáveis pelo processo de construção de significado (*construal of meaning*). Definir onde estão esses limites e como eles são é imprescindível para identificarmos quais devem ser os “pontos” da rede que estrutura uma categoria de sentidos polissêmicos e para guiar a representação das conexões e dimensões responsáveis por sua estruturação.

Em busca de estratégias para delimitar os sentidos polissêmicos, levaremos em consideração diferenças no *grau de antagonismo* e *de autonomia* identificados a partir dos critérios apresentados em Croft e Cruse (2004) e Cruse (1995) - texto muito semelhante ao de Cruse (1986). Levando em consideração o que observamos no

parágrafo anterior, podemos entender que Cruse (1986) é incompatível com Croft e Cruse (2004) por estes terem posições contrastantes sobre a existência de limites entre os sentidos. Embora não se possa negar essa oposição (salientada pelos próprios Croft e Cruse), entendemos que ambos os textos contribuem para a discussão que levaremos a efeito a partir desta seção por diferentes motivos.

O primeiro deles é o fato de que podemos claramente perceber nesses dois conjuntos de obras espaço para situações não típicas na construção do significado e para questionamentos sobre a validade de regras únicas e gerais para todas as situações. O segundo motivo é que, tendo em vista que a natureza do fenômeno da polissemia é múltipla, o tipo de tratamento que ele requer é também múltiplo. Precisamos olhar para os dois lados e, conforme vimos em Taylor (1995), buscar *o meio-termo*. Por fim, enfatizamos que a argumentação de Croft e Cruse (2004) a favor da existência de fronteiras claras não toma como base a negação da existência de flexibilidade entre os sentidos. O que os autores defendem é que essas variações podem ser explicadas como “construções de fronteiras diferentes das convencionais, não sendo necessário recorrer ao conceito de fronteiras difusas” (CROFT e CRUSE, p.95), noção que, segundo eles, vai contra a própria concepção de categoria como um constructo que demarca o que está *dentro* e o que está *fora* de um conjunto. Feito esse esclarecimento inicial, passaremos à caracterização da noção que adotaremos de *delimitação de sentidos*.

Explicaremos a natureza da *delimitação de sentidos* a partir de três tópicos diretamente associados aos princípios de fundo da SLC que apresentamos em 3.2. O primeiro deles remete ao fato de que o conteúdo lexical pode ser explicado em termos de um *continuum* de sentidos e de que esse *continuum* não é uniforme (CROFT e CRUSE, 2004). A justificativa é a de que, à medida que esse conteúdo vai sendo instanciado na superfície da língua, agrupamentos de sentidos vão sendo construídos no espaço conceitual, graus distintos de saliência, de similaridade, de antagonismo, de cooperação e de discretude vão sendo estabelecidos por convenção, explicam Croft e Cruse (2004). O segundo tópico envolve a relação entre esse *continuum* de sentidos e a polissemia. Quando o acesso a esse conteúdo semântico/conceitual se dá a partir de uma mesma estrutura fonológica, temos caracterizada a multiplicidade de sentidos; dependendo do espaço onde essa “porção densa” de conteúdo conceitual se posiciona no *continuum*, teremos homonímia, polissemia ou monossemia. O terceiro ponto a destacar remete ao *status* que as fronteiras entre sentidos adquirem nesse contexto. Segundo Croft e Cruse

(2004, p. 111), a delimitação dos sentidos é caracterizada a partir da percepção de espaços vazios entre eles, “linhas falhas”, que podem ser maiores ou menores, dependendo de como o sentido é construído em cada situação. Sob essa perspectiva, portanto, a polissemia é definida como uma questão de “construção de limites entre os sentidos” que compõem o espaço conceitual acessível por determinado item lexical (CROFT e CRUSE, 2004, p. 110), ou seja, é uma questão de delimitação de fronteiras entre porções do conteúdo semântico associado a um item lexical.

Tratamos, na Seção 3.5, da relação homonímia, polissemia e monosssemia do ponto de vista da representação do conhecimento, e a conclusão a que se chegou é a de que a opção por uma das abordagens depende do tipo de conteúdo lexical que se pretende registrar, se mais geral e estável ou se mais específico e mais dependente de elementos situacionais. Conforme já mencionamos, no entanto, essa definição não se resume a um trivial debate terminológico: o tipo de recorte a ser estabelecido entre os fenômenos não é uma tarefa objetiva que possa ser resolvida empiricamente; pelo contrário, depende, sobretudo, de decisões teóricas sobre, por exemplo, o papel reservado para a representação lexical e a função atribuída ao conhecimento enciclopédico para a determinação dos sentidos associados às unidades linguísticas. Em consonância com as investigações apresentadas até aqui, Cruse (1995) afirma que o que ocorre de fato na delimitação de sentidos homônimos, polissêmicos e monossêmicos é uma combinação de dois tipos de *continua*: (i) *continuum* de antagonismo, que vai de elementos totalmente antagônicos até elementos altamente similares, e (ii) um *continuum* de autonomia, que vai de elementos altamente autônomos (discretos) até elementos altamente cooperativos. A delimitação dos espaços nesses *continua* é tarefa que suscita outra série de discussões; como ilustração, reportaremos o que diz Cruse (1995) para o item *livro*.

Parece não haver dúvidas de que *livro* corresponde a duas leituras discretas, embora não antagônicas; entretanto, definir como dar conta de diferenças desse tipo não é algo livre de controvérsias, indo-se de abordagens monossêmicas a polissêmicas (cf. CRUSE, 1995). Pensando em relações de mapeamento entre formas lexicais e conteúdos semânticos, a resposta de Cruse (1995) é a de que leituras antagônicas terão mapeamentos separados relacionados a núcleos semânticos (ou *protótipos*) distintos, ao passo que, para leituras cooperativas, haverá um único mapeamento, com dois elementos centrais inclusos em um mesmo “envelope conceitual” (p.48). O *antagonismo*

– principal fator para definir a delimitação entre sentidos – é caracterizado quando é percebido um grau maior de saliência de uma leitura em relação às outras. Cruse (1995) traz a noção de “inibição lateral” para explicar esse mecanismo: trata-se da situação em que, em um conjunto de elementos, um deles responde mais fortemente aos estímulos, inibindo a resposta dos demais. Uma questão mais difícil de ser respondida, de acordo com o autor, é quando se tem caracterizadas unidades de sentidos ao mesmo tempo *autônomas e cooperativas*.

Olhando para sentidos *cooperativos*, Cruse (1995) salienta que o principal problema é explicar as diferenças de *status* da relação entre leituras que podem ser selecionadas separadamente (autônomas), mas que também podem ser unificadas (cooperativas). O exemplo do autor é *book*, e trataremos aqui do seu equivalente em PB, *livro* em sentidos como ‘objeto físico’ e ‘conteúdo’, que podem ser associados tanto a *livro* quanto a *romance*, por exemplo. Podemos dizer que compramos um livro ou que compramos um romance; é possível também dizer que o livro está muito bem escrito e que o romance está muito bem escrito; é possível ainda dizer que o livro está muito bem encadernado, e o mesmo pode ser dito para romance; por fim, podemos dizer que sentamos sobre o livro, mas, diferentemente, dizer que sentamos sobre o romance não é um uso muito comum. O que fazer nesses casos, sendo que ambos os itens podem ser considerados um tipo de publicação? Duas alternativas são previstas pelo autor. Uma delas é dizer que as duas leituras prototípicas associadas a *livro* não são igualmente acessíveis: elas exigem esforços de processamentos cognitivos distintos para ativar uma e não a outra. Isso pode justificar, por exemplo, as diferenças de preferência de leitura entre *um livro longo* e *um livro pesado* – são as diferenças de saliência de que falaremos na seção seguinte. Para explicar as diferenças entre *livro* e *romance* (*book* e *novel*, no original), no entanto, algo mais radical é sugerido por Cruse (1995) para a representação: para *romance*, somente a leitura prototípica ‘conteúdo’ é estabelecida, sendo ‘objeto físico’ gerado a partir de demandas do contexto.

Na tentativa de encontrarmos critérios para uma representação coerente para sentidos com diferentes graus de autonomia e de similaridade é que as seções seguintes serão desenvolvidas, tratando dos efeitos de saliência, da existência ou não de similaridade e do tipo de similaridade ou de relação, fatores que contribuem para delimitarmos e caracterizarmos os sentidos em uma rede de polissemia.

4.2 EFEITOS DE SALIÊNCIA ENTRE OS SENTIDOS

Refletir sobre as propriedades de *saliência* é essencial para entendermos as diferentes nuances que podem ser percebidas entre os sentidos, contribuindo para definirmos a disposição dos espaços em uma rede de polissemia e auxiliar futuros trabalhos voltados à geração e interpretação de textos. O conteúdo pressuposto nessas questões é o de que nem todos os elementos de uma categoria de sentidos têm o mesmo *status*. A partir daí, uma série de questões precisam ser respondidas: qual é a natureza das saliências? Em que medida um membro de uma categoria de sentidos é mais saliente do que outro? Qual é o membro mais saliente? Que critérios utilizar para definir essa saliência? O que significa dizer que os membros de uma categoria de sentidos têm *status* flexível? Em que termos essa flexibilidade se apresenta? Ela é total ou parcial?

Para compreendermos a natureza dos efeitos de saliência, é preciso voltarmos, agora com mais detalhes, à operação cognitiva da "percepção" que constrói o elo entre o mundo exterior e a representação mental. A percepção é uma das formas através das quais a cognição corporificada restringe o seu *input*. Sabe-se que a língua não é um espelho do que temos na mente, e a explicação para isso é que o conteúdo linguístico é construído via mecanismos de percepção, de modo que o que nós percebemos não é necessariamente o mesmo que nós experienciamos diretamente (EVANS e GREEN, 2006). Os mecanismos de percepção que facilitam nossa experiência foram formalizados nos anos 80 pelo movimento denominado Psicologia Gestalt (TYLER E EVANS, 2006), processo responsável pelas associações e agrupamentos que realizamos ao experienciar o mundo (ex.: princípio de similaridade, proximidade e contiguidade; figura e fundo; simetria; etc.). Perceber um sentido como mais saliente do que outro é um dos princípios gerais da nossa habilidade de percepção. Entre as formas de explicação dessa operação cognitiva de percepção, estão as noções gerais em LC de *figura* e *fundo* (TYLER E EVANS, 2006). O princípio de saliência baseado nessas noções refere-se à capacidade da percepção humana de automaticamente segregar qualquer cena em termos desse tipo de estrutura (TYLER E EVANS, 2006). A *figura* é a entidade que está em saliência, em destaque, por ser percebida como uma forma dominante. A *figura* está em contraste com o *fundo*, a parte da cena que é relegada ao *background*

(TYLER E EVANS, 2006). Na figura 16, vemos um exemplo de relação ambígua que “brinca” com a nossa percepção de *figura e fundo*.



Figura 16 - Figura e fundo
(O'REGAN e NOË, 2000)

A figura 16 ilustra um exemplo em que *figura e fundo* estão em competição por saliência, de modo que podemos fixar nosso olhar na imagem branca e não ver a preta, ou o contrário, mas ambas estão presentes na imagem.

Os diferentes efeitos de saliência exibidos pelos sentidos polissêmicos podem ser considerados equivalentes a efeitos de *centralidade*, *tipicalidade* ou *prototipicalidade* (GEERAERTS, 2006). O estudo sobre efeitos de prototipicalidade ou de saliência nascem no contexto da Teoria de Protótipos. Uma forma comum de entender a organização interna de uma categoria, sob essa visão, é pensar que sentidos menos prototípicos são derivados de sentidos mais prototípicos por meio de mecanismos cognitivos que facilitam a extensão de sentido (ex. metáfora conceitual, esquemas imagéticos) (EVENS e GREEN, 2006). O sentido básico é considerado o centro semântico de uma categoria, é ele que dá unidade/coesão à categoria, tornando acessíveis outras leituras. Lembramos que a posição central não precisa ser composta por apenas um elemento e nem por núcleos independentes. É possível que o sentido central seja constituído pela sobreposição de dois sentidos que podem também ocorrer independentemente (GEERAERTS, 2006). Para ilustrar, podemos voltar ao exemplo de *fresh* que

mencionamos em 3.6, de Geeraerts. O sentido central é formado pela combinação de ‘em ótimas condições’ e de ‘novo, recente’. Entender que há sobreposição entre porções do conteúdo associado a um item e que certos sentidos são ativados simultaneamente, embora um possa estar mais em destaque do que o outro por algumas das razões que mencionamos aqui, é a posição assumida por Geeraerts (2006) ao defender a estrutura multidimensional de uma categoria de sentidos. Essa postura, conforme destaca o autor, não está entre os princípios de uma teoria clássica de protótipos.

A definição do membro prototípico (não mais a partir de condições necessárias e suficientes fixas) pressupõe, de acordo com Geeraerts (2006a), levar em consideração a *gradiência* (grau de pertencimento a uma categoria), a *flexibilidade* (a inexistência de limites fixos entre os elementos de uma categoria) e a *saliência* (diferenças de peso estrutural). Como uma alternativa a abordagens baseadas em experimentos psicolinguísticos que realizam testes com falantes para a identificação do elemento prototípico, em diferentes trabalhos apresentados em Geeraerts (2006), encontramos critérios que possibilitam estabelecer diferenças no peso estrutural de várias instâncias de um item lexical. Apresentaremos abaixo aqueles critérios mais diretamente relacionados ao foco deste trabalho.

4.2.1 Saliência de Perspectiva (*Realce*)

A saliência de perspectiva, ou o realce, envolve diferenças de pontos de vista (“viewing arrangement”, Langacker 2001) na construção de determinada situação. Trata-se da seleção de determinada perspectiva em detrimento de outra (conforme ilustramos na figura 16). As noções de *perfil/base*, *figura/fundo*, *trajeto/marco* (“trajector”, “landmark”) são noções que aparecem em SLC para se referir a questões similares. O que está em questão é a saliência que ocorre pelo acesso ou a focalização de um aspecto particular de um sentido (uma *faceta* de sentido), por exemplo, *livro* pode ser usado com foco em sua parte física ou em sua parte textual. O olhar para a parte física realça o sentido de *livro* como um ‘objeto físico’, por exemplo, na sentença *Rasguei o livro*. O olhar para a parte textual salienta o ‘conteúdo’, por exemplo, na sentença *Este livro é muito agradável de ser lido*.

4.2.2 Saliência Semasiológica Paradigmática (*Prototypicalidade*)

O efeito de saliência semasiológica paradigmática, ou de prototypicalidade, é aplicado na identificação do melhor exemplar da categoria, ou seja, daquele elemento para quem a relação paradigmática *é um tipo de* melhor se aplica. Olhando para a relação entre os sentidos associados a um mesmo item, o sentido mais saliente será aquele que for considerado mais representativo da categoria de sentidos em questão. Geeraerts (2006) cita *bird* como exemplo que pode tanto significar ‘pássaro’ quanto, em linguagem informal, ‘garota’, sendo que o primeiro é tido como saliente paradigmaticamente, pois a ele mais prototipicamente se aplica a relação *é um* membro desta categoria rotulada como *pássaro*.

4.2.3 Saliência estrutural (*distintividade*)

O efeito de saliência estrutural, ou de distintividade, estará relacionado ao peso da dimensão distintiva que distingue um sentido do outro. Um dos pontos da rede (propriedades, dimensões, sentidos) será mais saliente do que o outro quando ele ocorrer com mais frequência na estrutura do léxico (GEERAERTS, 2006). Esse critério segue a hipótese estatística de Rosch (1975), que prevê, seguindo um critério simples, que os membros mais frequentes de uma categoria sejam os mais salientes. Como ilustração, o item *fruto* será apresentado a seguir.

4.2.4 Saliência Psicológica (*Funcional*)

A saliência psicológica, ou funcional, nasce da hipótese da vantagem *psicológica* de Rosch (1975), que salienta o acesso a determinado membro de uma categoria por ser cognitivamente mais vantajoso, mais econômico. Entende-se que esse princípio cognitivo é ativado para maximizar a riqueza conceitual de dada categoria através da incorporação de nuances de sentidos proximamente relacionadas em um único conceito,

tornando o sistema conceitual mais econômico (GEERAERTS, 2006). A organização das relações de sentido auxilia na identificação do elemento funcionalmente mais saliente. O fato de haver um maior número de relações apontando para um mesmo sentido é um forte indicativo de que ele pode ser considerado o central, a partir do qual outros podem ser derivados de modo mais econômico cognitivamente. Geeraerts (2006) aponta que esse princípio é útil para explicar diferentes agrupamentos de elementos que não puderam ser explicados a partir dos critérios anteriores.

Para encerrar esta seção, salientamos, seguindo Geeraerts (2006), que nenhum dos critérios pode ser tomado como infalível para testar a saliência. Seus estudos evidenciam, no entanto, que o uso combinado dos critérios tem apresentado resultados satisfatórios. A complementaridade dos efeitos de saliência pode ser percebida ao olharmos para a semântica de *fruit* (cf. Geeraerts, 2006, p. 349), item com comportamento similar ao seu equivalente em PB, *fruto*. O item é polissêmico, pois, ao lado de sua leitura mais básica e comum ('sweet and soft edible part of a plant, containing seeds'), há vários outros sentidos a ele convencionalmente associados: o sentido mais comum de *fruit* como 'sweet and soft edible part of a plant, containing seeds' ocupa posição mais central do que o sentido arcaico ('offspring') ou o técnico ('the seed-bearing part of a plant or tree'). A justificativa para a definição do sentido básico pode se dar a partir da percepção de três fatores de saliência a ele associados: (i) trata-se do sentido que veio mais rapidamente à mente das pessoas (saliência psicológica); (ii) é o sentido mais frequente (saliência estrutural); (iii) "edible part" é o sentido que pode ser considerado ponto de partida para descrever outros, isso porque, por exemplo, é o que mais facilita na compreensão da expressão *fruit of the womb*, se ela for nova para o leitor, quando já se entende o sentido de *edible part* (saliência paradigmática). Na base desse terceiro critério, está a noção da similaridade, tópico a ser detalhado nas próximas seções.

4.3 A SIMILARIDADE E A POLISSEMIA

Na Seção 2, tratamos da similaridade sob o olhar computacional e de sua utilidade ao PLN para tarefas de desambiguação - WSD. Vimos lá que há uma grande comunidade científica em busca de critérios e técnicas para extrair e codificar a *similaridade semântica*. Na Linguística teórica como um todo, não somente na SLC, a noção de similaridade está há muito tempo presente nos estudos sobre o léxico quando se trata de relações paradigmáticas e de polissemia. Em LC, seguindo pressupostos da Psicologia, a percepção da similaridade aparece como um dos princípios básicos da cognição. Até chegarmos neste estágio do trabalho, dialogamos tanto com a Linguística quanto com a Computação, e a noção de similaridade apareceu diversas vezes, porém a partir de um tratamento em certa medida vago e impreciso. Nesta seção, procuraremos corrigir esse fato.

A similaridade pode ser tomada como critério para a identificação de diferentes propriedades semânticas. Se olharmos para o léxico sob o ponto de vista onomasiológico, e pensando em relações paradigmáticas, o fenômeno típico a evidenciar similaridade é a sinonímia, que reflete a construção de máxima identidade semântica entre dois itens lexicais distintos. A relação de hiponímia também é comumente vista como fator que evidencia similaridade de algum tipo. Se pensarmos na semântica de *frames* (de Fillmore), a participação em um mesmo *frame semântico* pode ser também tomada como critério de similaridade entre itens distintos. Por outro lado, se o olhar para o léxico seguir a perspectiva semasiológica, é a polissemia o fenômeno que está diretamente relacionado à similaridade. A identificação de similaridade entre sentidos associados a um mesmo item lexical ou a uma mesma categoria lexical é tida como o principal critério para caracterizar a polissemia: quando dizemos que percebemos *relação* entre os sentidos, na verdade, estamos pretendendo dizer que há *similaridade*. Mas o que, de fato, isso significa?

4.3.1 Caracterização da Similaridade

Since speakers are very adept at perceiving semantic relationships, and since comparison is encouraged by common symbolization, it is seldom safe to assume that no connection whatever is established between the alternate senses associated with a form. Even a tenuous relationship, dimly perceived, is still a relationship, and though certain phenomena may presuppose a close or salient connection (LANGACKER, 2002, p. 268).

Em Linguística, a similaridade entre itens lexicais já foi explicada, por exemplo, em termos de traços semânticos. Conforme explica Lehrer (1973), Katz (1966), seguindo a perspectiva tradicional da semântica componencial, propôs que a similaridade entre sentidos polissêmicos seja medida através da contagem de componentes semânticos partilhados por dois sentidos. A sugestão pode parecer simples e útil, mas, assim como ocorre com outras aplicações da semântica de traços, a decisão sobre quais e/ou quantos descritores devem ser considerados relevantes para determinar se os sentidos são semelhantes ou diferentes não chega a um resultado considerado satisfatório. De acordo com esse critério, explica Lehrer, *bank* (instituição financeira) e *bank* (banco de areia) poderiam ser classificados como semelhantes, uma vez que ambos possuem os descritores [objeto físico] e [inanimado] (LEHRER, 1973 p.9), entretanto, o que é típico ocorrer é construirmos *bank* de modo a minimizar um possível conteúdo esquemático.

Na tentativa de não cair nessas armadilhas de técnicas binárias de identificação de similaridades, um critério utilizado é o julgamento intuitivo dos falantes. Julgamentos sobre a proximidade ou a relação entre os sentidos relacionados a um item lexical tomam como base uma distinção “pré-teórica”, ou seja, intuitiva. Soares da Silva (2006) realiza um estudo empírico para testar em que medida os falantes percebem similaridade entre sentidos associados a uma mesma forma lexical para o Português Europeu. Não só ele, mas também estudos mais antigos, para o inglês, como os de Sandra e Rice (1995) (com foco na semântica das preposições), Lehrer (1973) e Panman (1982), chegaram às seguintes conclusões: (i) é comum aos falantes a habilidade de distinção entre diferentes sentidos relacionados e sentidos completamente distintos associados por uma mesma forma; (ii) o fato de os falantes perceberem a existência ou não de relação entre sentidos justifica atribuir natureza semântica à distinção psicológica entre polissemia e homonímia; (iii) a multiplicidade de sentidos não se caracteriza em termos de uma dicotomia, mas como regiões salientes em um *continuum* de similaridade

(semelhança/diferença); (iv) os sentidos em graus máximos nos *continua* que organizam a multiplicidade de sentidos são percebidos de modo mais consensual do que aqueles situados em posições intermediárias; (v) os falantes têm menos dificuldades em identificar polissemia em sentidos com grau menor de autonomia e mais regulares (ex.: *livro* ‘objeto físico’ ‘conteúdo’) do que entre sentidos com grau maior de contraste e mais idiossincráticos (ex.: *posição, cabeça, boca*); (vi) os falantes estabelecem uma correlação direta entre similaridade semântica e identidade lexical, isto é, sentidos similares associados a um mesmo item lexical. No contexto da SLC, conforme já foi evidenciado, esse tipo de critério tem *status* privilegiado. Mesmo com seus resultados manifestamente distintos, no cômputo final, eles são considerados úteis. Para o problema que levantamos, no entanto, eles parecem ir ainda mais longe na imprecisão e vagueza, visto que não procuram definir a similaridade, apenas julgar sua presença ou ausência e quantificá-la em uma escala graduada.

Em SLC, quando se procura explicação para a natureza da similaridade, essa explicação, em sua quase totalidade, gira em torno da noção de *metáfora*. Cruse (1995), por exemplo, diz que a similaridade entre sentidos é gerada em função de processos de transferência metafórica e é caracterizada quando conseguimos perceber um dos sentidos como o *elemento básico* (ou o prototípico) e o outro como o elemento estendido (o “metafórico” ou “figurativo”), identificado em função de seu potencial mais restrito para ser construído em contextos neutros (ex.: *cabeça* como ‘parte do corpo humano/animal’ ou ‘quem está no comando’). Ainda que as definições de *metáfora* e de *figuratividade* fossem claras e livres de complicadores teóricos, essa explicação também não resolveria totalmente o nosso problema em busca de uma caracterização mais precisa da similaridade.

Mais elementos para a caracterização da similaridade encontramos ao olharmos para o fenômeno como um princípio cognitivo responsável pela construção de aproximações entre diferentes entidades. Nesse contexto, podemos explicar a similaridade em termos dos referidos princípios Gestalt, que explicam os mecanismos perceptuais inconscientes responsáveis pela construção de ‘todos’ ou de ‘gestalts’ a partir do processamento de *inputs* que são incompletos (EVANS & GREEN, 2006). De acordo com esse critério, podemos encontrar (se existir) um *elemento unificador* relevante para a interpretação dos sentidos em comparação. Entende-se que essa associação pode se dar em diferentes termos: baseando-se em fatores objetivos e/ou

fatores subjetivos. Os fatores objetivos pressupõem que as entidades percebidas como similares em uma cena compartilhem características objetivas, tais como tamanho, forma ou cor, e que sejam percebidas como pertencentes a um grupo (EVANS e RICE, 2006). Olhando para os sentidos, esse tipo de associação exige que as palavras denotem entidades similares objetivamente (HIRSCH, 1997). Para ilustrar, podemos pensar na similaridade de forma que pode ser construída entre *bala* 'projétil', em *Os moradores estão assustados com a quantidade de bala perdida na região*, e 'guloseima', em *Eu gosto de bala de uva*. Fatores subjetivos são considerados aqueles que criam associações de um "tipo especial" entre entidades a partir de experiências de mundo diversas (HIRSCH, 1997). Esse princípio pode ser interpretado como requerendo palavras que denotem associações subjetivas entre entidades particulares em função do espaço em que o falante vive (HIRSCH, 1997). Para ilustrar esse tipo de similaridade, vamos considerar a aproximação que pode ser construída entre *cabeça* como 'parte do corpo humano/animal', em *Minha cabeça dói*, e como 'chefe de uma empresa', em *Carlos agora é o cabeça da empresa*. O psicólogo Hirsch (1997) salienta que, mesmo na primeira situação (a de similaridade a partir de parâmetros objetivos), há espaço para subjetividade no julgamento sobre a existência ou não de similaridade entre os elementos que constituem o potencial conjunto. Além de variar em termos de objetividade/subjetividade e de intensidade, a similaridade pode ser construída com grau menor ou maior de linearidade (ou regularidade), conforme demonstraremos. Entendemos, a partir dessas observações, que olhar para a noção de similaridade como um princípio cognitivo responsável pela construção de relações de sentido de diferentes tipos ajuda-nos a entender de modo menos intuitivo o que se pretende comunicar quando se diz que é possível reconhecer a existência de *similaridade* entre os sentidos A e B.

A partir dessas observações iniciais, passaremos às próximas etapas de caracterização da similaridade de sentidos polissêmicos, a serem realizadas em dois momentos: (a) a identificação da existência ou não de similaridade entre sentidos e (b) a identificação do tipo de similaridade. Para investigar a existência de similaridade, vamos explorar os critérios linguísticos apresentados por Croft & Cruse (2004) e Cruse (1995). Mesmo que o foco de tais estudos não seja a similaridade, mas a inexistência dela (o antagonismo e o grau de autonomia), os critérios apresentados interessam por duas razões. A primeira deve-se ao fato de que, quando são construídos sentidos antagônicos,

se vê negado o acesso mental a qualquer conteúdo unificador, visto que ele precisaria ser muito abstrato. Sendo assim, avaliar o antagonismo entre os sentidos auxiliará a avaliar quão relevante ou não é buscarmos algum fator de coerência entre os sentidos em análise (ex. noção unificadora, conteúdo esquemático, relação de sentidos). A segunda justificativa é o fato de que investigar sobre quão autônomos são os sentidos será útil para avaliarmos quão autônoma deve ser a “entrada” para esse sentido no léxico ou como deve ser o “ponto” na rede de polissemia. Concluída a apresentação desses critérios, passaremos, na seção final desta parte do trabalho, a olhar para os elementos que vão dar forma à nossa rede de polissemia: as relações que explicitarão o tipo de similaridade percebido sincronicamente entre os sentidos.

4.3.2 A Existência ou Negação de Similaridade entre os Sentidos

Para apresentarmos cada tipo de similaridade, levaremos em conta as seguintes variáveis: quão autônomos e quão cooperativos são os sentidos? Quão antagônicos (e não similares) são os sentidos? Quão similares (e não antagônicos)? Dizer que dois sentidos são *antagônicos* significa dizer que eles estão em competição, portanto, exigirão que o leitor selecione apenas um dos sentidos e exclua o outro já no contexto linguístico imediato (o sintático) (CRUSE, 1995; CROFT e CRUSE, 2004); em contrapartida, dizer que dois sentidos são similares significa dizer que percebemos algum fator de coerência entre eles, mas isso não significa dizer que eles são idênticos, pois, conforme veremos, mesmo sentidos similares podem apresentar certo grau de antagonismo. A propriedade de autonomia dos sentidos é a responsável pela delimitação entre eles; veremos que eles podem estar em grau máximo de autonomia, caracterizando os “sentidos plenos” (cf. CROFT & CRUSE, 2004, p. 111), ou em graus de autonomia baixos, exigindo que olhemos para o sentido como um complexo formado por unidades de sentido menores, os “subsentes”, que são de dois tipos, *facetadas* e *microsentidos* (cf. CROFT & CRUSE, 2004, p. 116).

Leituras¹ alternativas podem ser consideradas: (1) completamente ambíguas por fazerem uso de sentidos com grau máximo de (a) antagonismo e (b) autonomia, de modo que – tipicamente – nenhum fator de coerência é percebido, caracterizando o que denominamos *grau insignificante de similaridade*; (2) *semidistintas* por serem percebidas como parcialmente (a) antagônicas e (b) autônomas; e, por fim, podem ser consideradas (3) cooperativas, tendo em vista que não é percebido grau relevante de (a) antagonismo, porém há certo grau de (b) autonomia. Para a apresentação de cada um desses graus de similaridade, os sentidos serão divididos conforme a sua capacidade de construção de fronteiras e de delimitação de unidades sob os rótulos *sentidos plenos* e subsentidos do tipo *facetas* e *microsentidos*.

Sob o rótulo *sentido pleno*, seguindo Croft e Cruse (2004), serão abarcadas aquelas unidades de sentido bem estabelecidas, incrustadas, no léxico de uma língua que são fortemente suportadas por restrições convencionais estáveis. São sentidos reconhecíveis a partir de distinções mais marcantes e menos sutis em razão de apresentarem importante grau de autonomia. Como exemplo, mencionamos os sentidos tipicamente construídos por *banco*, como ‘instituição financeira’ e ‘assento’, e por *cabeça*, como ‘parte do corpo humano/animal’ ou ‘quem está no comando’. A falta de similaridade (ou a presença de antagonismo) entre os sentidos é tipicamente tomada como um critério de grau alto de autonomia entre os sentidos. Existem, no entanto, unidades de sentido - *subsentidos* - que se parecem com os sentidos plenos em muitos aspectos, mas que não apresentam antagonismo ou que são construídos somente em contextos restritos (da mesma forma que os sentidos plenos podem ser unificados em contextos restritos) (CROFT & CRUSE, 2004). A ausência de antagonismo prevê a presença de similaridade e, portanto, a possibilidade de unificação, que pode ser de diferentes tipos, conforme veremos a seguir para os subsentidos do tipo *facetas* (ex.: o que ocorre com *book* ‘objeto físico’ e ‘conteúdo’) e *microsentidos* (ex.: o que ocorre com *faca* ‘utensílio’, ‘arma’).

¹ O termo “leitura” está sendo tomado como noção genérica, sem estabelecer diferença, por exemplo, entre sentido, sentido local, subsentido, faceta e microsentido - conceitos que passarão a fazer parte de nossas discussões a partir desta seção.

4.3.2.1 Grau insignificante de similaridade entre sentidos plenos

Sentidos plenos com *grau insignificante de similaridade* são sentidos com (a) alto grau de antagonismo e de (b) autonomia. O tipo de comparação entre os sentidos descritos aqui caracteriza situações típicas de *homonímia*. Uma das características principais de sentidos antagônicos típicos é a resistência à unificação. Um exemplo clássico ocorre em *banco*, no contexto (3)

3. Daniel parou em frente ao *banco* Banrisul.

No exemplo (3) o sentido ‘*banco* instituição financeira’ é apresentado como única porção relevante do conteúdo lexical de *banco*, ou seja, como única porção relevante acessada por seu potencial de significado. O fator de delimitação ativado foi o contexto linguístico imediato (CROFT e CRUSE, 2004).

Dizer que os sentidos são *autônomos* significa dizer que eles possuem capacidade para funcionar de modo independente em um mesmo contexto (CROFT e CRUSE, 2004). Sentidos que apresentam grau máximo de antagonismo apresentam também grau máximo de *autonomia* (ou de *discretude*). A combinação de grau máximo de antagonismo e grau máximo de discretude é que resulta no que Croft e Cruse (2004) denominam *sentidos plenos*, por revelarem fortes efeitos de delimitação (do inglês, *boundary effects*), demonstrados a partir das propriedades mencionadas. A propriedade da autonomia serve como critério para a identificação de uma das principais características dos sentidos plenos, que é a resistência à unificação, revelando um alto grau de antagonismo. Como *fatores de autonomia*, os seguintes critérios podem ser tomados (CROFT e CRUSE, 2004):

(a) *Autonomia de atenção*

A autonomia de atenção é considerada como critério mais forte para a identificação de antagonismo. Trata-se da habilidade de exclusão mútua entre dois sentidos em termos de focos de atenção. Nessas situações, as unidades de sentido estão em competição: se uma estiver em foco, a outra precisa ser excluída. Em PB, essa propriedade pode ser percebida no exemplo 4.

4. Chegamos ao *banco* que procurávamos.

Fica claro que em (4) o leitor precisa escolher exclusivamente entre um dos sentidos de *banco* armazenados em seu léxico mental, por exemplo, ‘instituição financeira’ ou ‘assento’. Croft e Cruse (2004) destacam dois fatores como principais consequências dos efeitos de antagonismo e de autonomia de atenção que podem ser revelados a partir de testes linguísticos, que são baseados no “julgamento sobre a aceitabilidade e restrições semânticas em frases com duas ocorrências de um dado item lexical em análise (uma das quais pode estar implícita ou subjacente)” (SOARES DA SILVA, 2006); são eles: *teste de identidade* e *teste de zeugma*.

O teste de restrição de identidade, inicialmente proposto por Lakoff (1970), prevê que a polissemia seja detectada quando as duas ocorrências coordenadas do item exigirem identidade semântica. Em PB, esse efeito pode ser percebido em 5.

5. Isabel foi correndo até aquele *banco*; e o mesmo fez Cristina.

Em (5) há fortes restrições favorecendo a construção de dois sentidos: *banco* como ‘instituição financeira’ e como ‘assento’. Antagonismo e autonomia de atenção são percebidos no fato de que somos fortemente constrangidos a construir o mesmo sentido para ambas as porções da sentença (CROFT e CRUSE, 2004). Dois pontos são destacados pelos autores a partir desse exemplo. O primeiro é o fato de que as restrições de identidade não são absolutas: é possível construirmos (5) com diferentes sentidos para *banco* para cada uma das partes da sentença, mas essa leitura requer mais trabalho cognitivo e a penalidade é gerar uma interpretação que pareça um jogo de palavras. O segundo ponto destacado pelos autores é que, na construção da sentença sem a relação anafórica verbal, a pressão por um sentido único é mais fraca, como mostra 6.

6. Isabel foi correndo até aquele *banco*; Cristina foi correndo até aquele *banco* também.

A segunda consequência das propriedades de antagonismo e de autonomia de atenção em grau máximo aparece quando as duas leituras antagônicas e discretas são

simultaneamente evocadas e ligadas pela supressão do segundo elemento; trata-se do efeito definido zeugmático², demonstrado em 7.

7. Isabel e Cristina foram correndo até o *banco*.

O *teste de zeugma* tem sido criticado por não gerar resultados consistentes quando um mesmo contraste de sentidos em alguns contextos produz zeugma, mas em outros não (CROFT e CRUSE, 2004). Uma situação considerada ainda mais grave pelos autores que contribui para pensarmos na invalidade do teste é quando exemplos prototípicos de sentidos geradores de ambiguidade não produzem zeugma, caso ilustrado em (8).

8. Nem todos os bancos são bancos de dinheiro³.

A primeira menção de *banco* em (8) representa uma unificação dos sentidos convencionalmente associados ao item. O contexto, nesse caso, faz com que encontremos uma propriedade unificadora suficientemente distintiva para a categoria *banco* como um todo (CROFT e CRUSE, 2004). Essa inconsistência, no entanto, não é vista por Croft e Cruse como razão para anular a validade do teste, e a justificativa é a própria flexibilidade inerente ao significado lexical, que só é especificado por completo em uso, seguindo-se a SLC. Os autores enfatizam que esperar “consistência” perfeita de testes como esse seria esperar que o sentido fosse uma herança dos itens lexicais, fato que contraria os princípios da SLC. Situações não prototípicas como essa, que contrariam as regras de convenção, são sempre possíveis.

(b) *Autonomia relacional*

A *autonomia relacional* é apresentada como um critério menor para a identificação de antagonismo. Um dos sintomas desse tipo de autonomia é a propriedade que dois sentidos possuem de ativar conjuntos distintos de relações de sentido (CROFT e CRUSE, 2004). As dificuldades encontradas para a aplicação deste critério de maneira confiável são justificadas, conforme Croft e Cruse, pela própria sensibilidade que as relações paradigmáticas têm ao contexto; por essa razão, quanto mais estáveis forem as

² O efeito zeugmático é considerado por Croft e Cruse (2004) como equivalente ao efeito de fenômenos como *silepse* e *punning*, visto que todos têm a mesma origem.

³ Exemplo adaptado de Croft e Cruse (2004, p. 115).

relações e mais amparadas por restrições convencionais, mais fortes serão as evidências que elas apresentarão de autonomia e antagonismo nos sentidos em análise. Entre as relações mencionadas pelos autores, estão a antonímia, a hiponímia/hiperonímia e a meronímia/holonímia. Os autores destacam que, quando as relações em questão são as duas últimas, para que os sentidos sejam considerados autônomos, é importante que os sentidos pertençam a campos lexicais distintos. Os dois sentidos de *bank* ('instituição financeira' e 'margem do rio') ilustram bem a situação: *bank* em *bank of the river* (margem do rio) tem *source* (fonte), *bed* (leito) e *mouth* (boca) como co-merônimos, mas não há um hipônimo ou hiperônimo evidente para todos eles; já para *bank* em *the bank in the High Street* (o banco da High Street), co-hipônimos como *building society* (edifício de uma sociedade), *insurance company* (companhia de seguros) e *financial institution* podem ser identificados como um hiperônimo, e diferentes departamentos do *banco* seriam os merônimos.

(c) *Autonomia composicional*

O critério da autonomia composicional assumirá maior importância em sentidos com níveis mais baixos de autonomia, mas ele também implica claramente unidades delimitadas por fronteiras bem definidas. Trata-se de um critério para análise semasiológica com base em fatores onomasiológicos. Esse tipo de autonomia caracteriza situações em que, em um processo composicional – combinação sintagmática entre nome e verbo, nome e adjetivo –, um dos elementos seleciona somente uma porção do sentido do outro elemento (CROFT e CRUSE, 2004). Os autores explicam que o estabelecimento de fronteiras entre os sentidos sensíveis a esse tipo de autonomia se dará em torno da porção selecionada pelo conteúdo semântico do item em questão, separando-o do restante. Como exemplo para o PB, podemos citar *banco confortável*, em que o adjetivo ignora completamente o sentido 'instituição financeira'. Por outro lado, para uma situação como *banco prático*, o critério de autonomia composicional não iria ser útil, tendo em vista os múltiplos sentidos também do adjetivo *prático*.

Os seguintes critérios podem ser tomados como os *fatores prototípicos de antagonismo*, causando grau insignificante de similaridade entre os sentidos (CROFT e CRUSE, 2004):

- (i) Dificuldade para inclusão dos sentidos como co-hipônimos em uma organização hierárquica;

- (ii) Dificuldade para a aplicação da propriedade do *englobamento*, isto é, para a assimilação dos sentidos como partes de um mesmo todo relacionadas a uma *noção unificadora* (Gestalt global) válida. O máximo que é possível é encontrarmos como unificador um elemento altamente genérico, como *coisa, entidade, local*, noções que não são suficientemente distintivas para delimitarmos o que é e o que não é, por exemplo, *banco*. Nos termos de Langacker (1987), dizemos que o conteúdo esquemático nessas situações é pouco informativo ou distante.

4.3.2.2 Grau relevante de similaridade entre sentidos plenos

Sentidos com *grau relevante de similaridade* são sentidos que devem ser considerados (a) antagônicos e (b) autônomos, porém que não se encontram em graus máximos em nenhum desses *continua* (*cf.* CRUSE, 1995). Por ser uma situação intermediária de antagonismo e de autonomia, é inevitável percebermos tanto fatores de autonomia e discretude entre os sentidos quanto fatores de coerência.

Dizer que dois sentidos são parcialmente antagônicos e autônomos significa dizer que entre eles há certo grau de competição, fazendo com que o leitor precise selecionar um deles para a interpretação do enunciado, mas significa dizer também que entre eles há fatores de coerência (similaridade). Sendo assim, em razão do seu potencial de autonomia, esse tipo de sentido pode ser considerado também um *sentido completo* (*full sense*, nos termos de CROFT e CRUSE, 2004), porém tem propriedades de antagonismo menores do que o que vimos na seção anterior.

Como principal fator de coerência, que nega e/ou neutraliza certas propriedades de antagonismo, destacamos a possibilidade de identificação de alguma “noção geral” (CROFT e CRUSE, 2004, p. 110). No PB, podemos considerar como exemplo novamente o item *banco*, porém agora explorando outras porções do conteúdo conceitual a ele associado. Há vários usos desse item relacionados à noção de ‘coleção ou custódia’ de dinheiro ou de alguma outra ‘*commodity*’, como ocorre em contextos como *o banco da rua principal, Banco Mundial, o banco de sangue, o banco de olhos, o banco de dados, etc.*⁴

⁴ Exemplos traduzidos dos originalmente apresentados em Croft e Cruse (2004, p. 110).

Esse fator de coerência justifica entendermos ser possível formar um grupo (*cluster*, cf. CROFT e CRUSE, 2004) entre esses usos de *banco*, isto é, uma categoria de sentidos. A coerência entre os membros desse grupo, no entanto, não é total: percebe-se a existência de uma marcante descontinuidade (fator de autonomia) delimitando cada um dos elementos desse grupo ‘custódia’, que formarão *subgrupos* (ex. ‘custódia de dinheiro’, ‘armazenamento e custódia de órgãos’). Para explicar esse tipo de delimitação de fronteiras entre sentidos é que Croft e Cruse (2004) fazem uso da noção de *linha falha*, que facilita a construção de fronteira/delimitação e justifica percebermos esses sentidos como autônomos ou discretos em maior ou menor grau, podendo ser percebidos como membros de uma mesma categoria ou de categorias distintas, conforme a distância entre eles. Como evidência dessa discretude, Cruse (1989) destaca sua típica habilidade para participação em domínios altamente distintos, explicada a seguir.

Segundo Cruse (1989), embora não sejam completamente distintos, eles não geram problemas no uso porque fazem parte de domínios diferentes, exigindo que um dos sentidos seja selecionado a partir do contexto linguístico imediato. A noção de domínio é entendida aqui meramente como um conjunto de itens lexicais inter-relacionados por determinada relação semântica, tal como antonímia, hiponímia, meronímia (CRUSE, 1989). Dentro do seu próprio domínio, cada sentido local funciona como um sentido plenamente delimitado. Para os sentidos associados a *banco* que mencionamos acima, isso ficou claro ao observarmos que *banco* responsável pela custódia de dinheiro pode ser considerado um tipo de ‘instituição financeira’ (relação de hiponímia), já *banco* responsável pela custódia de sangue pode ser considerado um tipo de ‘instituição de saúde’ e, ainda, *banco* de dados pode ser considerado um tipo de recurso computacional.

Outro tipo de leitura parcialmente distinta é o que ocorre em *boca*, nos contextos *a boca do cavalo*, *a boca da garrafa*, *a boca de uma caverna*, *a boca do rio*, etc. (cf. CRUSE, 1995; 1989)⁵, que estão unidos pelo fator de coerência da similaridade. Esse tipo de variação de sentido é considerado o efeito sincrônico típico de extensões metafóricas a partir de um sentido comum com base na relação de similaridade. Como evidência dessas contiguidades, Cruse (1995) apresenta o critério da falta de efeito zeugmático:

⁵ Exemplo traduzido de Cruse (1995, p. 338).

sentidos não completamente distintos são contíguos em um *continuum*, portanto, tipicamente não geram zeugma, conforme exemplo 9.

9. A boca da caverna lembra a da garrafa⁶.

O exemplo (9) sugere a existência de algum elemento esquemático unindo os dois usos de *boca* em questão. Usos mais separados por pontos mais distantes no *continuum* de similaridade são ilustrados em (10); entretanto, também não são zeugmáticos.

10. O chocolate envenenado entrou na boca da Condessa exatamente quando seu iate entrou na do rio.⁷

Do mesmo modo que os usos de *banco* mencionados, os sentidos associados a *mouth* também fazem parte de domínios diferentes e exigem seleção a partir do contexto linguístico imediato – esses são os fatores de autonomia entre esse tipo de sentidos parcialmente distintos e significativamente similares. Por exemplo, o sentido básico, ou “literal”, de *boca* é ‘boca de um animal/humano’ (a julgar por sua relativa liberdade de contexto), conforme argumenta Cruse (1989; 1995). As extensões desse sentido são relativamente frágeis porque requerem condições precisas para seu aparecimento. Somente o sentido, de acordo com o autor, tem potencial para aparecer em um contexto neutro tal como *Fizemos um projeto sobre bocas*. O sentido ‘boca do rio’ é dificilmente possível, a menos que o item lexical *rio* ocorra na vizinhança próxima em uma porção anterior do discurso. O autor entende que, mesmo quando é montado o contexto para uma viagem em um rio, ele julga mais plausível entender (11) como mais básico ou comum do que (12) se nenhuma menção recente de *rio* tiver ocorrido.

11. Vamos descer a boca do rio?

12. Vamos descer a boca?

Vimos dois tipos de sentidos que potencialmente geram leituras parcialmente ambíguas em razão da possibilidade de percebermos efeitos de coerência (negando efeitos de antagonismo, mas não totalmente) e efeitos de autonomia (negando efeitos de coerência, mas não totalmente). O tipo de multiplicidade de sentidos a ser tratado aqui é

⁶ Exemplo traduzido de Cruse (1995, p. 39).

⁷ Exemplo traduzido de Cruse (1995, p. 39).

aquele a que nos referimos como *polissemia irregular*. O exemplo mais clássico desse tipo de fenômeno é o de *posição*, quando usado em sentidos ‘local ocupado’ ou ‘atitude mental’. Uma possibilidade de explicar a similaridade que pode ser percebida entre os sentidos é dizer, fazendo associação a mecanismos cognitivos (cf. LAKOF e JOHNSON, 1980), que ocorre em razão do empréstimo de certas propriedades semânticas de um domínio como *lugar* a um domínio como *opinião*, entendendo-se haver a seguinte associação mental: “*declarar opinião*” é “*ocupar um lugar*”.

Os seguintes critérios podem ser tomados como *fatores prototípicos de coerência* entre sentidos parcialmente distintos:

- (i) Possibilidade de aplicar a propriedade do *englobamento*, sendo possível encontrarmos uma *noção unificadora* válida ou um *conteúdo esquemático*;
- (ii) Falta de efeito zeugmático;
- (iii) Grau médio de antagonismo;
- (iv) Percepção da existência de um sentido mais neutro (“básico” ou literal) e de um sentido aplicável a contextos mais restritos (“extensão”)
- (v) Possibilidade de reconhecer fatores de continuidade revelados pela *relação de similaridade* entre os sentidos.

Os seguintes critérios podem ser tomados como *fatores prototípicos de autonomia* entre os sentidos parcialmente distintos:

- (i) Há diferentes graus de saliência entre os sentidos (ex.: sentido mais básico e sentido menos básico);
- (ii) Os sentidos são altamente discretos;
- (iii) Autonomia relacional e associação a diferentes campos semânticos;
- (iv) A relação entre os sentidos é assimétrica, configura-se pela transferência de certas propriedades de um sentido (que pertence a um domínio A) para outro sentido (pertencente a um domínio B); trata-se de uma relação de A para B, que não se aplica, portanto, de B para A, não havendo empréstimo mútuo.
- (v) Delimitação do sentido a partir do contexto linguístico imediato;

4.3.2.3 A similaridade entre facetas de sentidos

O foco agora são os *subsentes* do tipo *faceta*. Suas principais características são (a) certo grau de autonomia, mas (b) inexistência de antagonismo. As facetas são percebidas como *componentes distintos* (subsentes do tipo *faceta*) de um único sentido global (sentido formado pela cooperação entre as facetas). A principal evidência de coerência entre as facetas é o fato de que o leitor não precisa selecionar apenas uma delas para a interpretação de um enunciado, apenas colocar uma em evidência. A principal evidência de autonomia entre as facetas é o fato de que elas não são capazes de estabelecer relação de hiponímia com um único elemento.

Esse tipo de variação de leitura gera o que referimos como *polissemia regular*. O rótulo *regular* dado a esse tipo de polissemia justifica-se por diferentes razões. A primeira razão é o fato de que os sentidos relacionados, embora tenham superordenados distintos em uma rede hierárquica, são elementos que podem ser considerados “perspectivas” de um mesmo sentido. Trata-se do acesso ou do enfoque de um aspecto particular de um sentido, conforme vimos em *livro* em 4.1, que pode ser usado com foco em sua parte física ou em sua parte textual. A segunda razão para o rótulo *regular* é o fato de que o mesmo tipo de alternância semântica é recorrente do mesmo modo no léxico. Como exemplo, além da alternância bidimensional ‘objeto físico’/‘conteúdo’, uma nova dimensão pode ser percebida, que é a faceta ‘texto físico’ (ex.: para tratar do tipo de impressão ou do *lay-out*), caracterizando um tipo de alternância de tridimensional de facetas percebida também em *agenda*, *polígrafo*, *apostila*, etc. Croft e Cruse (2004) enfatizam que, “embora as facetas associadas a um mesmo item sofram fortes restrições de convenção, em princípio, não há razão para acreditar que as facetas de sentidos formam um conjunto fechado” (p. 117). Esse tipo de relação entre facetas foi mencionado na Seção 2 deste trabalho sob o rótulo primo (*cousin*), evidenciando ser possível encontrarmos uma classe de sentidos que apresenta o mesmo tipo de alternância - não podemos, no entanto, deixar de observar que há exceções (ex.: *novel* ou *romance*).

Os seguintes critérios podem ser tomados como *fatores prototípicos de autonomia* entre sentidos parcialmente distintos, conforme os autores:

(a) *Autonomia relacional*

As facetas mostram completa independência relacional, de modo que cada faceta pode participar de relações de sentidos próprias, independentemente das demais. Por exemplo, em muitos contextos, o hipônimo de *livro* forma duas taxonomias paralelas, que não se relacionam:

13. Alguns dos *livros* que nós lemos foram romances, e outros, biografias.⁸

Em (13), *livro* é construído como a faceta ‘conteúdo’, que se apresenta em relação hiperonímica com os hipônimos incompatíveis *romance* e *biografia*.

14. Alguns dos *livros* eram de capa dura outros brochura.⁹

Em (14), *livro* é construído como a faceta ‘objeto físico’, que se apresenta em relação hiperonímica com os hipônimos incompatíveis *brochura* e *capa dura*.

15. *Alguns dos *livros* eram de capa dura outros brochura.

O contexto (15) ilustra que *romance* e *capa dura* não são combinados em um contexto que espera que os itens sejam construídos como incompatíveis em um único domínio.

16. Um romance é um tipo de *livro*.

17. Um *livro* de capa dura é um tipo de *livro*.

O fato de ser normal que sentenças descontextualizadas como (16) e (17) mostram que a potencialidade para a construção das facetas mencionadas para *livro* é bem suportada por restrições convencionais.

(b) *Autonomia composicional*

Os predicados podem ser aplicados às facetas independentemente:

18. *Livro* vermelho; *livro* empoeirado.

⁸ Exemplo traduzido de Croft e Cruse (2004, p. 117).

⁹ Exemplo traduzido de Croft e Cruse (2004, p. 117).

Nas construções em (18), o adjetivo modifica somente a faceta ‘objeto físico’ de *livro*. Em (19), no entanto, o adjetivo modifica somente a faceta ‘conteúdo’, propriedade considerada por Croft e Cruse (2004) como indicação relativamente fraca de autonomia.

19. *Livro* interessante; *livro* difícil.

Em um contexto apropriado, como (20), uma pergunta de *sim* e *não* contendo *livro* pode ter potencial para ser verdadeiramente respondida de ambas as maneiras, o que significa que elas remetem, alternativamente, a uma das facetas de cada vez.

20. Você gostou do *livro*?

Fora de contexto, é provável que a pergunta seja respondida com foco na faceta ‘conteúdo’ somente, mas não é difícil visualizar uma situação em que se possam construir as facetas ‘objeto físico’ ou ‘texto físico’, por exemplo.

(c) “Núcleos” autônomos

Outra manifestação de autonomia entre as facetas apresentada pelos autores pode ser percebida em (21) e (22).

21. Não estou interessado no conteúdo, etc., estou interessado no *livro* em si.

22. Não estou interessado na encadernação, etc., estou interessado no *livro* em si.

Os contextos acima podem ser normalmente interpretados e envolvem diferentes leituras da expressão *do livro em si*: em (21), ‘objeto físico’, em (22), ‘conteúdo’. Esse fato é tomado como indicação de que *book* é associado a dois “núcleos” semânticos autônomos (CROFT e CRUSE, p. 119).

(d) Outros fatores de autonomia

O item *livro* pode ser ativado mesmo se uma de suas facetas for deixada de fora.

23. Eu ganhei um *livro* para escrever durante os minutos em que eu estiver na reunião.

24. A: Como está o seu *livro*?

B: Ah, está todo aqui [apontando para a cabeça], mas eu não escrevi nada ainda.

Por fim, conforme mostra (25), cada faceta pode ser alvo independente para extensão metafórica:

25. Um *livro* de cruzadinhas.

Vimos que somente a faceta ‘objeto físico’ foi selecionada. Por fim, os autores destacam o fato de que cada faceta pode receber nome próprio independentemente: *O pequeno príncipe* é o nome do conteúdo, não do objeto físico, nem de ambas as facetas texto e objeto físico combinadas.

Passaremos agora a tratar dos critérios que podem ser tomados como *fatores prototípicos de unificação* das facetas. Assim como Croft e Cruse (2004) mostram ocorrer com *book*, percebemos que há usos de *livro* que podem ser unificados em um sentido global, o que justifica o argumento de que as facetas não têm o *status* de sentido pleno. Eles enfatizam que, por vezes, a autonomia das facetas é superestimada, e a evidência principal é o fato de que as facetas não podem ser unificadas como hipônimos de uma única categoria: não há categoria que agrupe ‘conteúdo’ e ‘objeto físico’; além disso, elas não são partes de um todo em sentido normal. Seguem, na sequência, os critérios para unificação apresentados por Croft e Cruse (2004):

(a) *A coocorrência é a leitura prototípica*

Os autores defendem que a leitura prototípica de *livro* é a leitura combinada com ambas as facetas, embora uma possa ser selecionada sem a outra, fato que, quando acontece, é considerado periférico.

(b) *Propriedades composicionais unificadoras*

Um aspecto para unir as propriedades composicionais é a existência de predicados que combinam com ambas as facetas simultaneamente:

26. Publicar um *livro*

Em (26), temos uma situação que impossibilita separar as facetas, visto que não é possível publicar um *livro* sem haver ‘conteúdo’ e ‘objeto físico’.

Outra situação de unificação composicional das facetas é ilustrada em (27):

27. Este é um *livro* muito interessante, mas é horrível de carregar.

Observa-se que a situação de zeugma não se aplica a contextos como esses, evidenciando a inexistência de antagonismo, conforme Croft e Cruse (2004). *Interessante* modifica a faceta 'texto', e *pesado de carregar* modifica a faceta *objeto físico*.

Por fim, para justificar o status de subsentidos às facetas, Croft e Cruse (2004) retomam o que ocorre em *banco* e *teacher*. No caso de *banco*, o que ocorre é que a representação conceitual provê forte pressão para separar as leituras; no caso de *criança*, em (28) e (29), ocorre exatamente o contrário; não havendo sintomas de autonomia, a representação conceitual provê forte pressão para unificar as leituras.

28. O nome da *criança* será Izabela.

29. O nome da *criança* será Antônio.

Vemos que *criança* em (28) é interpretada como feminino e em (29) como masculino. Nos exemplos de *livro*, os autores salientam que as restrições de estabilidade não nos induzem veementemente a seguir em nenhuma das direções, então, somos relativamente livres para construir as facetas tanto de maneira autônoma quanto como unificadas em resposta a outras restrições, tais como contextuais e comunicativas. E Croft e Cruse enfatizam que isso não ocorre por alguma falha das restrições conceituais, mas por elas agirem em direções contrárias, efetivamente cancelando umas às outras.

As seguintes propriedades podem ser tomadas como *fatores prototípicos de coerência* entre as facetas de sentidos:

- (i) Ausência de antagonismo;
- (ii) Destaque de apenas uma das facetas;
- (iii) Detecção de um conjunto de sentidos que apresenta o mesmo tipo de alternância (ex.: 'objeto físico' e 'conteúdo').
- (iv) Possibilidade de identificação de um Gestalt para as facetas;

As seguintes propriedades podem ser tomadas como *fatores prototípicos de autonomia* entre as facetas:

- (i) Certo grau de autonomia entre as facetas, sendo elas percebidas como componentes distintos de um único sentido;

- (ii) Não existência de um único hiperônimo para as duas facetas;
- (iii) Seleção de outras relações paradigmáticas distintas pelas facetas.

4.3.2.4 A similaridade entre os microssentidos

Os microssentidos estão em relação regular se um for uma especialização do outro, o que acarreta, evidentemente, que o último seja uma generalização do primeiro (CRUSE, 2000a). O autor explica que é possível distinguir especialização de generalização se reconhecermos um dos sentidos como mais básico do que outro: se A é mais básico do que B, e B é mais especializado do que A, então B é uma especialização de A.

Na seção anterior, tratamos do primeiro tipo de subsentido apresentado por Croft e Cruse (2004). Passaremos agora a tratar dos subsentidos classificados como *microssentidos* pelos autores. São unidades de sentido distintas de um item que ocorrem em diferentes contextos e cujas interpretações se estabelecem em relação de incompatibilidade mútua no mesmo nível hierárquico, assim como, por exemplo, nomes de animais, como *gato*, *cachorro*, *ovelha*, *vaca*, *porco*, *cavalo*. Os microssentidos são unidades de sentido que apresentam (a) significativo grau de autonomia, embora possam ser unificados como hipônimos de uma mesma categoria e (b) não há antagonismo, visto que funcionam como *co-autohipônimos* de um sentido geral. Além dessas características, é marcante o fato de que o hiperônimo não tem *status default*. Croft e Cruse (2004) apresentam *knife* como exemplo típico. Em razão de seu comportamento muito semelhante ao de *faca*, apresentaremos o item do PB para ilustrar, visto que ele pode ser usado como uma arma, como um instrumento, como um talher, sendo cada um desses usos um microssentido do sentido global de *faca*.

Croft e Cruse destacam diferentes fontes para as evidências de autonomia dos microssentidos:

(a) *Autonomia relacional*

Cada microssentido constrói seu próprio conjunto de relações de sentido independente. Os microssentidos de *faca* (30), por exemplo, geram interpretações que

pertencem a diferentes taxonomias e que, portanto, têm diferentes *hiperônimos*, *hipônimos* e *co-hipônimos*.

30. Talher: faca, garfo, colher

Arma: faca, revólver, granada

(b) *Autonomia de condições de verdade*

Os microsentidos exibem autonomia de condições de verdade. O teste das condições de verdade através de perguntas do tipo sim/não que apresentam respostas independentes para cada microsentido. Como exemplo, podemos citar (31)

31. Mãe (na mesa; Eduardo está partindo uma carne com seus dedos): Usa a faca para cortar a carne, Eduardo.

Eduardo (está com uma faca em seu bolso, mas não com uma faca do tipo que corte): Eu não tenho.

(c) *Autonomia com base na restrição de identidade*

O teste de restrição de identidade é também considerado importante evidência de autonomia entre os microsentidos, visto que, tipicamente, há forte pressão para que anaforizante e anaforizado tenham a mesma leitura, conforme mostra (32)

32. Cícero precisa de uma *faca*; Daniel também.

Cícero enviou uma *carta*; Daniel também.

Cícero tem um *equipamento*; Daniel também.

A restrição parece ser mais forte em certas situações, o que é visto pelos autores como reflexo de graus de autonomia, e isso fica mais claro quando contrastamos (32) com situações em que a inexistência de contraste é clara, como em (33).

33. Cícero tem um *carro*; Daniel também.

Cícero tem um *primo*; Daniel também.

Para encerrar a caracterização dos *microsentidos*, consideremos as seguintes questões:

- (i) *O que faz, então, os microsentidos não poderem ser caracterizados como sentidos plenos?* Os microsentidos não são sentidos distintos porque eles são unificáveis sob um mesmo hiperônimo e não são antagônicos, de acordo com os critérios usuais (CROFT e CRUSE, 2004).

- (ii) *Por que os microsentidos não são facetas?* A razão pela qual os microsentidos não são facetas é que as facetas são de tipos ontológicos distintos (ex.: concreto e abstrato) e não podem ser unificadas sob um mesmo hiperônimo (CROFT e CRUSE, 2004).

- (iii) *Por que os microsentidos não são simples enriquecimentos contextuais de sentidos superordenados?* Há diversas razões (CRUSE, 1995): (a) têm grau maior de discretude do que simples especificações contextuais (ex. *criança* – feminino ou masculino); (b) microsentidos, ao contrário de enriquecimentos contextuais, têm suas próprias relações de sentido (autonomia relacional) e participam, ao contrário de enriquecimentos contextuais, de domínios independentes (autonomia de domínio); (c) dentro do seu próprio domínio, os microsentidos não precisam ser adjetivados para que seu conteúdo básico seja percebido, ao contrário do que ocorre em casos como o de *criança*; (d) o superordenado é marcado e não aparece somente por pressão contextual, embora varie de domínio para domínio (ex.: *faca* como *talher*, *faca* como *arma*); isso não ocorre com situações de modulações contextuais como a de *criança*, em que, em um contexto neutro, a leitura dominante é também neutra em relação ao sexo.

Ao longo desta seção, foram descritas uma série de propriedades que servem como evidências para o delineamento de fronteiras entre sentidos plenos, facetas e microsentidos. Essas noções são importantes para entendermos sobre o grau de antagonismo, similaridade e de autonomia existente entre os diferentes sentidos e para que possamos propor uma metodologia de análise dos sentidos polissêmicos

(lembramos que, quando fazemos uso do termo *sentido*, pretendemos nos referir tanto a sentidos plenos quanto a subsentidos) baseada em critérios mais explícitos do que a intuição do falante. Em razão da flexibilidade inerente à construção dos sentidos, não consideramos definitivos ou determinantes os resultados de quaisquer testes. Eles servirão como recursos auxiliares à delimitação do tipo de polissemia por serem representativos de usos estabelecidos construídos por cada item.

4.3.3 Tipos de similaridade: relações de polissemia

Em 4.3.2, tratamos das questões implicadas na delimitação dos sentidos, percebendo diferentes graus de similaridade revelados a partir de critérios de antagonismo e de autonomia entre os sentidos, fato importante para caracterizar o tipo de polissemia (e de multiplicidade de sentidos como um todo). Nesta seção, abordaremos os diferentes efeitos de saliência que os múltiplos sentidos associados a um mesmo item lexical podem apresentar, facilitando a ativação de um ou de outro no uso. Concluídas as reflexões sobre os aspectos funcionais implicados na descrição da semântica de um item polissêmico, nesta seção, o objetivo é caracterizar a similaridade entre sentidos em termos de tipo.

As relações que selecionamos para a apresentação tomaram como critério o viés sincrônico de similaridade entre sentidos, e procuramos, para enfatizar essa perspectiva, fazer uso de uma terminologia distinta daquela comumente empregada em SLC já marcada de conotação diacrônica, por ter como objetivo explicar a motivação cognitiva de determinada extensão de sentido. Para este inventário de tipos, levamos em consideração as informações que inventariamos na Seção 3 especialmente, aquelas apresentadas por Langacker (ex.: relações verticais e horizontais), Geeraerts (ex.: generalização, especialização, metáfora, metonímia) e Blank (ex.: polissemia metafórica, polissemia metonímica, polissemia por co-hiponímia, polissemia taxonômica). Acreditamos que o inventário de relações que apresentaremos a seguir dê conta do tipo de conteúdo que percebemos como mais relevante para a construção de uma rede sincrônica de polissemia: (i) explicitação de informações que abarcam relações de

polissemia entre sentidos institucionalizados em uma língua particular (PB)¹⁰; (ii) distinção entre relações regulares e irregulares de polissemia; (iii) explicitação de relações lineares entre elemento geral e elemento específico; (iv) explicitação de relações de similaridade lineares que deem conta de alternâncias do tipo que vimos ocorrer em *livro* ‘objeto físico’ e ‘conteúdo’, além de outras alternâncias, como *parte-todo*, *instituição-prédio*, entre outras tipicamente denominadas alternâncias de *figura e fundo*; (v) explicitação de relações horizontais irregulares baseadas em similaridade menos previsíveis e lineares porque envolve sentidos com importante grau de autonomia e que se associam a domínios altamente distintos.

A apresentação das relações de polissemia seguirá o seguinte critério: parte-se do maior grau de regularidade ou linearidade e chega-se ao menor. Iniciaremos pelas relações regulares entre microssentidos, *autohiponímia/autosuperordenação*; em seguida, trataremos da relação regular que ocorre entre unidades de sentidos nomeadas *facetas*, a *automeronímia/autosuperordenação*; por fim, trataremos do tipo clássico de relação de polissemia, que é a irregular, que se manifesta entre sentidos plenos.

4.3.3.1 Relação de Polissemia Regular

O foco aqui é o tipo de relação ilustrada nas ocorrências de *livro*, *cachorro* e *faca* nas seções anteriores. Trata-se da *relação de polissemia regular*. O rótulo *regular* é dado a essa relação de polissemia especialmente pelo fato de que o mesmo tipo de alternância semântica é recorrente do mesmo modo no léxico. Entre os tipos de polissemia reconhecidos como lineares, sistemáticos ou regulares, destacaremos, a seguir, os mais proeminentes: a polissemia por *parte* e *todo* (*automeronímia/autoholonímia*) e a polissemia por *é-um* (*autohiponímia/autosuperordenação*) (CRUSE, 2000).

¹⁰ Uma situação oposta à de sentidos fortemente estabelecidos por restrições de uso ocorre em nuances de sentido geradas no momento do uso (discurso). São sentidos que não fazem parte do potencial de sentido acessível por um item. Conforme Blank (2003), trata-se de uma distinção entre sentidos “lexicalizados” (ex. *livro* ‘objeto físico’ e ‘conteúdo’) e “não-lexicalizados” (ex. *sanduíche de presunto*, expressão considerada comum no discurso de garçons). Embora esse uso possa ser considerado comum e recorrente, seria um erro dizer que, entre os possíveis sentidos da expressão, está “pessoa que pediu sanduíche”. Essa leitura é gerada no discurso (BLANK, 2003).

A seguinte frase lógica¹¹ pode ser tomada como critério geral para a identificação da polissemia regular: a polissemia de uma unidade lexical A com o sentido a_1 e a_j é chamada regular em uma dada língua se existir pelo menos uma outra unidade lexical B com sentidos b_1 e b_j que sejam semanticamente distintos um do outro exatamente da mesma maneira (APRESJAN, 1973). O tipo de distinção/similaridade será determinado por cada uma das relações indicadas abaixo:

4.3.3.1.1 Relação de polissemia por automeronímia/autoholonímia

O tipo de polissemia regular mais explorado na literatura é o que relaciona *parte* e *todo*, considerando-se essa relação entre sentidos polissêmicos uma relação tão geral quanto é a meronímia.

Esse tipo de relação regular ocorre entre facetas de sentidos, elementos com alto grau de similaridade. Esse rótulo representa facetas de sentidos que podem ser simultaneamente acessadas em dado contexto, mas também se pode dar saliência para uma apenas. Embora as facetas de sentidos tenham superordenados distintos em uma rede hierárquica, elas são elementos que podem ser considerados regulares por serem “perspectivas” de um mesmo sentido geral¹², portanto, a similaridade é, de certo modo, previsível (CRUSE, 2000a). Como exemplo, mencionamos a alternância *objeto físico/conteúdo*, que é percebida também em itens lexicais como *agenda, polígrafo, apostila*, etc., e que, ainda, tem certo grau de universalidade, visto que ocorre em várias línguas do mesmo modo (CRUSE, 2000) (ex. *book - inglês, Buch - alemão, libro - espanhol e italiano, livre - francês*).

Tendo em vista que as alternâncias de sentidos relacionados por polissemia regular já foram bastante exploradas na literatura, é possível acessarmos um vasto material sobre nominais tipicamente associados a facetas de sentido que formam uma classe. A similaridade de sentidos percebida entre *automerônimos/autoholônimos* é equivalente à que ocorre entre itens lexicais merônimos. Desse modo, podemos estudar essa relação de polissemia fazendo uma analogia à relação paradigmática clássica. A partir dessa visão, vemos a relação como uma família de relações. Fazendo uma adaptação do que propõem Saint-Dizier e Viegas (1995) para a *meronímia*, aproveitando

¹¹ Adaptada de Apresjan (1973).

¹² Tais alternâncias semânticas são frequentemente consideradas como reflexo da metonímia, relação baseada em correlação, contiguidade ou proximidade conceptual. Dentro de um mesmo domínio (não “cross-domain”, por isso contiguidade conceptual), possibilita que uma entidade seja usada pela outra porque ambas coexistem nesse mesmo domínio” (EVANS & GREEN, 2006)

algumas das classes clássicas e gerais de polissemia regular (PUSTEJOVSKY, 1995), apresentamos, para encerrar a seção, o quadro 4, com um inventário de tipos gerais de relação de polissemia regular por *automeronímia* com exemplos em PB. Salientamos que esse quadro abarca apenas tipos gerais e, conforme mostramos na Seção 2, com a proposta do CoreLexII (quadro 2), esses tipos podem ser mais específicos, facilitando sua extração automática. Nossa intenção aqui, no entanto, é registrar uma tipologia que sirva como guia para nossa análise manual.

| |
|--|
| <p>1. Todo/parte, todo/porção
 Ex.: <i>janela</i>
 Encomendei a <i>janela</i> para a sala./Passei pela <i>janela</i>.</p> |
| <p>2. Produto/substância, objeto/material
 Ex.: <i>chocolate</i>
 Comprei <i>chocolate</i>./Esse bolo é de <i>chocolate</i>.</p> |
| <p>3. Continente e conteúdo, objeto/substância
 Ex.: <i>colher</i>
 Entortei a <i>colher</i>./Quero uma <i>colher</i> de açúcar no meu café.</p> |
| <p>4. Objeto físico/conteúdo
 Ex.: <i>livro</i>
 Li o <i>livro</i>./O <i>livro</i> é pesado para carregar.</p> |
| <p>5. Pessoas/funcionamento/instituição/estrutura física
 (tipo mais específico do que os anteriores)
 Ex.: <i>universidade</i>
 A <i>universidade</i> está em festa.
 A <i>universidade</i> inicia em fevereiro.
 A <i>universidade</i> foi fundada há 31 anos.
 A <i>universidade</i> está em obras.</p> |
| <p>6. Animal/alimento
 (tipo mais específico do que os anteriores)
 Ex.: <i>porco</i>
 Comprei um <i>porco</i> para o sítio./Comi <i>porco</i> no almoço.</p> |

Quadro 4 – Inventário de tipos de *automeronímia*

4.3.3.1.2 Relação de polissemia por autohiponímia/autossuperordenação

As relações de polissemia do tipo *autohiponímia* ou *autossuperordenação* são apresentadas por Cruse (2000a) e constituem relações verticais que ocorrem entre sentidos gerais e específicos em diferentes graus de detalhamento.

A autohiponímia/autossuperordenação envolve relação entre unidades de sentido que, seguindo Croft & Cruse (2004), denominamos *microsentidos*. Trata-se de uma relação entre um sentido construído em um grau menor de detalhamento e um sentido construído em grau maior de detalhamento. É o que ocorre, por exemplo, entre os sentidos de *cachorro* como ‘macho da raça canina’ (maior grau de detalhamento) e como ‘membro da raça canina’ (menor grau de detalhamento).

Na relação de *autohiponímia*, conforme Cruse (2000a), o sentido geral é geralmente construído como padrão e está havendo uma especificação em dado contexto. A relação entre o elemento geral e o específico pode ser aplicada repetidamente, marcando diferentes graus de detalhamento dos sentidos. Cruse (2000a) especifica essa relação entre especialização e generalização como *linear*, visto que o fato de o sentido **A** ser generalização do sentido **B** acarreta que **B** seja especialização de **A**. Para ilustrar, vamos detalhar os microsentidos tipicamente associados a *cachorro*, exemplo adaptado de *dog*, apresentado por Cruse (2000a, p. 111): o item *cachorro* é frequentemente associado ao sentido geral (padrão) “membro da raça canina”, como em *Os donos de gatos e cachorros devem registrar seus animais de estimação*, e uma leitura mais específica, como em *Aquele não é um cachorro, mas uma cadela*. Cruse (2000a) chama a atenção para o fato de que a leitura mais específica demonstra autonomia, visto que a segunda oração contradiz a leitura mais geral de *cachorro*, conforme mostra a figura 17.

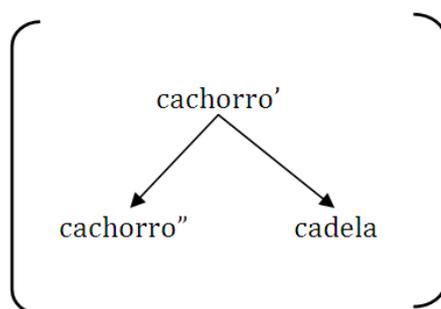


Figura 17 - Autohiponímia *cachorro*

A figura 17 mostra que *cachorro''* É-AUTOHIPÔNIMO de *cachorro'* e que *cadela* É-UM-TIPO-DE *cachorro'*, construído como ‘animal da raça canina’.

A *autossuperordenação* ocorre quando um item tem como padrão um sentido mais específico, contextualmente restrito, e está havendo uma generalização em dado contexto. Um exemplo de *autossuperordenação*, adaptado de Cruse (2000a, p. 111), é *homem*. Esse item lexical envolve a generalização do masculino, sendo usado para denotar raça humana, em contraste com o uso de *homem* denotando o ser masculino da raça humana, como em *João é um homem de caráter sem igual*, e em uma leitura mais geral, como em *O homem precisa pôr em prática alternativas ecológicas*. A figura 18 ilustra esses exemplos.

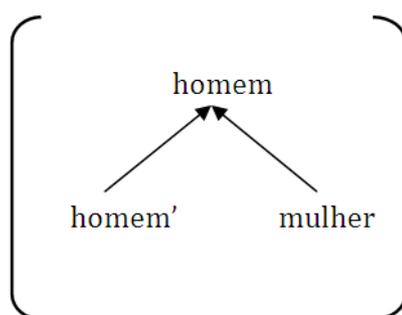


Figura 18 – Autossuperordenação *homem*

A figura 18 mostra que *homem* “É-AUTOSSUPERORDENADO-DE *homem*” e que *mulher* “É-UM-TIPO-DE *homem*, como ‘*ser humano*’.

Quando esse tipo de relação ocorre entre itens lexicais distintos, no contexto das relações semânticas nos moldes estruturalistas, o termo usualmente empregado para o elemento mais geral é *superordenado* ou *hiperônimo* e para o termo mais específico é *subordinado* ou *hipônimo*. Para identificarmos a relação hiponímica entre os nomes, podemos fazer uso do seguinte esquema: X é um Y, ou X é um tipo de Y; a mesma frase lógica é válida para sentidos polissêmicos.

A terminologia empregada aqui foi proposta em Cruse (2000a). Entretanto, essa relação entre um elemento mais geral e outro mais específico aparece na literatura também como relação de *esquema-instância* (LANGACKER, 1987; 2002) e *especialização-generalização* (denominação que tem origem na linguística diacrônica, empregada, por exemplo, por BRÉAL, e que é também utilizada por autores como Geeraerts (2000a)) e como relação de *polissemia taxonômica*, por Blank (2003). Para enfatizar nosso foco sincrônico, não faremos uso do vocabulário empregado para expressar relações de

motivação de polissemia, mas de termos cunhados para descrever sincronicamente relações entre sentidos, conforme Cruse. Outra questão importante a destacar é que olharemos para as relações de autohiponímia e autossuperordenação como relações inversas, não sendo relevante para nossa aplicação pensar em uma distinção estrita entre ambas.

4.3.3.2 Relação de Polissemia Irregular

A *relação de polissemia irregular* é caracterizada por um tipo de similaridade e pode ser aplicada a sentidos plenos que são construídos de formas diferentes ou mesmo conflitantes, porém, é possível percebermos entre os elementos alguma nuance de sentido comum, que pode ser descrita como uma “noção geral”, tal como vimos em “custódia” para *banco*. Trata-se de um princípio de agrupamento do tipo de similaridade a partir da percepção de uma *noção geral* ou como um Gestalt global. É essa noção geral que será a responsável por justificar pensarmos em certos sentidos como constituintes, mesmo com alto grau de um conjunto formado por elementos com certo grau de antagonismo e de autonomia. Essa é a relação que caracteriza o fenômeno clássico de polissemia e que foi apresentada por Langacker como uma relação gerada por extensão do tipo *similaridade* a partir do protótipo e por Blank (2003) como *polissemia metafórica*, em razão da sua motivação típica ser transferências desse tipo.

O rótulo *irregular* é aplicado porque esse tipo de relação de polissemia, embora altamente recorrente no léxico, é mais difícil de ser inferido ou previsto a partir de padrões de geração, tendo em vista que a similaridade a partir de associações com alto grau de subjetividade. Conforme temos comentado ao longo do trabalho, a similaridade que pode ser construída entre dois ou mais sentidos pode ser de diferentes tipos (forma, função, aparência, etc.), que se organizam a partir de diferentes graus de intensidade e de convencionalização. Mesmo associações objetivas, conforme já mencionamos, têm profundas marcas de similaridade. É essa percepção de similaridades baseadas em fatores subjetivos que possibilita construirmos relação entre sentidos com importante grau de discretude e antagonismo que podem ser associados a domínios discursivos diferentes ou mesmo distantes (BLANK, 2003). Além do exemplo de *cabeça* que

referimos ao longo do trabalho, podemos destacar o item *posição*, que tipicamente constrói sentidos como ‘local ocupado’ (em *Estou em posição confortável*) ou ‘atitude mental’ (em *Preciso tomar uma posição*). Uma forma de explicar a similaridade entre os sentidos é pensar em empréstimo de certas propriedades semânticas de um domínio como *lugar* a um domínio como *opinião*. Pode-se dizer que se faz a seguinte associação mental: “*declarar opinião*” é “*ocupar um lugar*”¹³; a partir daí, poderíamos dizer que é possível construir um conteúdo esquemático *localização*, englobando os dois sentidos mencionados.

A relação entre sentidos irregularmente polissêmicos será indicada em nosso trabalho pelo rótulo *é_similar_a*, entretanto, não podemos deixar de chamar a atenção para a não-especificidade do rótulo. Para demonstrar a dificuldade de encontrarmos um rótulo mais específico para a relação, vamos recuperar os exemplos de polissemia irregular que apresentamos ao longo do trabalho, procurando explicitar o tipo de similaridade que conseguimos construir em cada caso: (a) *bala* (similaridade de forma), (b) *boca* (similaridade de forma e função), (c) *cabeça* (similaridade de função e posição), (d) *banco* (no sentido de ‘custódia’, similaridade de função), (e) *posição* (similaridade por associação de processo cognitivo e processo físico). Além de haver uma variedade de associações, fica evidente que algumas das associações são mais fáceis de ser construídas do que outras e, por consequência, serão mais consensuais entre os falantes. Como exemplo de similaridade simples de ser construída, destacamos *cabeça* ‘parte do corpo humano’ e ‘quem está no comando’. Como exemplo mais difícil de ser construído, destacamos *posição*.

Os critérios para a caracterização da *polissemia irregular* são os seguintes: (i) há diferentes graus de saliência entre os sentidos, mas todos são altamente discretos; (ii) observa-se um grau médio de antagonismo entre os sentidos, já que é possível reconhecer alguma similaridade; (iii) a relação de similaridade tem grau alto de irregularidade e baixo de previsibilidade, já que a associação se dá entre sentidos

¹³ Numa perspectiva diacrônica, a polissemia assistemática é frequentemente associada ao fenômeno da metáfora (entendida como “palavras e expressões convencionais produzidas a partir de co-relações, especificadas como mapeamentos entre dois domínios”, conforme Lakoff (1987) e Lakoff & Johnson (1980)), considerada sua principal motivação. Na análise de *posição*, nota-se, de fato, que uma associação metafórica é o que está na base da similaridade existente entre os sentidos denominados irregularmente polissêmicos, visto que ela ocorre entre sentidos que mapeiam para domínios distintos.

evocadores de domínios distintos; (iv) a relação entre os sentidos é *assimétrica*¹⁴, configura-se pela transferência de certas propriedades de um sentido (que pertence a um domínio A) para outro sentido (pertencente a um domínio B); trata-se de uma relação de A para B, que não se aplica, portanto, de B para A, não havendo empréstimo mútuo.

A seguinte fórmula lógica pode ser tomada como critério geral para a identificação da polissemia irregular: uma unidade lexical A com o sentido a_1 e a_2 é chamada irregular em uma dada língua se **NÃO** existir outra unidade lexical B com sentidos b_1 e b_2 que sejam semanticamente distintos um do outro exatamente da mesma maneira da distinção entre a_1 e a_2 .

4.4 QUADRO SÍNTESE DA SEÇÃO

¹⁴ A propriedade de assimetria é destacada por Evans & Green (2006) ao descreverem a polissemia gerada por metáfora.

| | | Formam <i>polyset</i> * | | |
|--------------------------------------|---|---|--|---|
| Multiplicidade de sentidos | Homonímia | Polissemia irregular | Polissemia regular | |
| Tipo de sentido | Sentidos plenos 1 | Sentidos plenos 2 | Subsentidos | |
| | | | Microsentidos | Facetas |
| Exemplos | Manga, <i>banco</i> (inst./assento) | Posição, cabeça, boca, <i>banco</i> (inst.=custódia) | Cachorro, faca | Livro, janela |
| Similaridade | Grau insignificante | Leituras semidistintas (grau relevante) | Grau funcional | Grau alto |
| Tipo de relação | - | Assimétrica (irregular)
<i>tem_similaridade_com</i> | Simétrica (regular - recorrente)
<i>é_autohipônimo_de</i> <i>é_automerônimo_de</i> | |
| Autonomia | Grau máximo | Grau significativo | Grau significativo - Há incompatibilidade mútua no mesmo nível hierárquico – são co-autohipônimos. | Grau mínimo |
| Antagonismo | Grau máximo | Grau médio | Grau mínimo | Inexistente |
| Capacidade de unificação | Altamente resistente | Ocorre.
<i>Tipo de unificação:</i> Possibilidade de construção de uma ‘noção geral’. | Ocorre.
<i>Tipo de unificação:</i> um subsentido é uma subespecificação do outro. | Ocorre.
<i>Tipo de unificação:</i> os subsentidos são componentes distintos de um único sentido. Partes de um mesmo todo. |
| Seleção de atenção | Grau máximo | Grau máximo | Algum grau. O subsentido hiperônimo não tem <i>status</i> de atenção <i>default</i> . | Grau baixo - Seleção de perspectiva. Não é necessário delimitar apenas uma das facetas, mas é possível (núcleos autônomos). |
| Efeito zeugmático | Grau máximo | Difícilmente ocorre | Grau significativo. | Difícilmente ocorre |
| Autonomia relacional | Grau máximo. Campos semânticos distintos. | Grau máximo. Campos semânticos distintos. | Grau alto. | Grau máximo. Campos semânticos similares, tipos ontológicos distintos, mas não são unificadas sob um mesmo hiperônimo. |
| Autonomia composicional | Completa (não muito relevante) | Completa (não muito relevante) | Completa (não muito relevante) | Ocorre, mas também pode servir como evidência de unificação. |
| Outros critérios de autonomia | | | a. Autonomia de condição de verdade.
b. A ativação do co-hipônimo não exige adjetivação. | a. “Núcleos” autônomos.
b. Independência de ativação das facetas. |

Quadro 5 – Síntese da seção 4.

SEÇÃO 5 - DELINEAMENTO LINGUÍSTICO DO MODELO DE REPRESENTAÇÃO: O CONSTRUCTO *POLYSET*

As discussões empreendidas nas seções anteriores convergem neste ponto para a proposição de uma metodologia de análise e representação linguístico-computacional da polissemia de nominais que complemente as informações léxico-conceituais já codificadas em uma *wordnet*. Nesta seção, o objetivo é apresentar o delineamento linguístico desse modelo. Para isso, como já observamos na introdução, adotamos a estratégia de pesquisa *task-driven* (cf. NIRENBURG e RASKIN, 2004), de modo que procuraremos manipular os constructos teóricos propostos no âmbito da SLC, sem perder de vista o objetivo de o constructo a ser proposto contribuir para a ampliação da utilidade das *wordnets* para o desenvolvimento de sistemas de PLN. Essa meta exigiu, ao longo de todo o trabalho, um direcionamento da pesquisa para a proposição de uma representação linguística que possa ser codificada computacionalmente. Salientamos que essa orientação aplicada não exige - nem serve de justificativa para - um olhar leviano à teoria linguística e, conforme procuraremos demonstrar, essa tarefa interdisciplinar pode ser realizada com sucesso.

A metodologia geral da tese, como argumentamos na introdução, seguiu aquela proposta por Dias-da-Silva (1996, 1998; 2003; 2006), e, portanto, previu a integração dos três domínios mutuamente complementares: o linguístico, o linguístico-computacional e o computacional. Para a análise linguística, foram necessárias reflexões referentes aos dois primeiros domínios. O **domínio linguístico-computacional** tem o papel de estabelecer a ligação entre o primeiro e o terceiro domínios. Assim, a discussão que se fez no âmbito desse domínio forneceu o cenário computacional da pesquisa, ao mapear as lacunas e as exigências da aplicação-alvo. A conclusão dessa etapa, descrita na Seção 2, evidenciou que uma representação formal da polissemia de nominais em uma base de dados lexicais possibilita que um sistema de PLN manipule níveis distintos de generalidade de sentidos. Desse modo, a base se torna mais útil, por exemplo, tanto para sistemas de tradução automática e de processamento de informação, que exigem um

grau de granularidade mais fina, quanto para sistemas de classificação de textos, que necessitam de um grau de granularidade mais grosseiro.

Conhecido o cenário de aplicação, trabalhamos a segunda etapa da pesquisa: o **domínio linguístico**. Embora ele esteja, em certa medida, subordinado a necessidades computacionais, procuramos explorar o modelo escolhido em profundidade, de maneira a tornar a representação linguístico-computacional também adequada a um contexto linguístico-cognitivo. A flexibilidade e a irregularidade do significado lexical, tão enfatizadas na SLC, parecem, em um primeiro momento, colocar uma limitação à sua codificação computacional. Conforme procuramos demonstrar nas Seções 3 e 4, trata-se apenas de uma impressão. A abordagem de redes, mais especificamente, a de redes de polissemia, revelou ser a solução ideal, por elas possibilitarem uma representação que abrange dois tipos de codificação: a que codifica a representação de sentidos gerais e abstratos e a que codifica a representação de níveis de instanciação dessas abstrações, fornecendo uma estratégia para a representação do significado lexical em diferentes níveis de generalidade e de estabilidade. Considerar a polissemia um fenômeno categorial possibilitou entendermos que os sentidos podem ser decompostos em termos de “categorias” e, desse modo, são explicados como entidades complexas e a partir de uma estrutura multidimensional. Com base nas investigações relatadas na Seção 4, seguindo o paradigma da SLC, o modelo de representação deve apresentar as seguintes características: ser capaz de codificar (i) os tipos de sentido (sentidos plenos e subsentidos do tipo facetas e microsentidos), (ii) os graus distintos de saliência, (iii) os graus distintos de delimitação de sentidos (maior ou menor autonomia entre os sentidos) e (iv) os fatores de similaridade, que caracterizam as relações de sentido e os tipos de polissemia.

Tomando as *wordnets* como a aplicação alvo e a SLC como referência teórica, esta seção visa a desenvolver a primeira etapa de aplicação da teoria estudada, dedicando-se à caracterização do modelo representacional para a polissemia de nominais: os *polysets*. Propomo-nos a representar fundamentalmente a realidade linguística sincrônica dos nominais polissêmicos do PB. Entretanto, a representação proposta tem *status* múltiplo conforme descreveremos.

Esta quinta Seção está organizada em duas grandes partes: em 5.1, abordaremos *o polyset como um constructo de status múltiplo* e, em 5.2, *os componentes e a estrutura do polyset*.

5.1 O *POLYSET* COMO UM CONSTRUCTO DE *STATUS* MÚLTIPLO

O *polyset*¹⁵ deve ser entendido como um “conjunto de sentidos polissêmicos”. Trata-se de um constructo destinado à representação estruturada do conteúdo semântico associado a itens lexicais polissêmicos. Os critérios propostos para a construção dos *polysets* possibilitam a representação da polissemia de nominais porque (i) representam o sentido em diferentes níveis de generalidade; (ii) explicitam o tipo de relação existente entre os sentidos associados a uma mesma estrutura fonológica; (iii) explicitam diferentes graus de saliência entre os sentidos que formam um conjunto polissêmico. Desse modo, a partir da aplicação dos critérios propostos para a construção dos *polysets*, a base da WN.Br terá representações distintas para itens homônimos e itens polissêmicos (de diferentes tipos). Destacamos, mais uma vez, que a base teórico-metodológica para a proposição do *polyset* é a SLC e a sua estruturação interna é descrita em termos de uma rede de polissemia com enfoque sincrônico nos moldes propostos por Geeraerts (2006) e Blank (2003), por esse tipo de rede ser multidimensional e livre de formato fixo.

Os *polysets* são considerados constructos com *status* múltiplo porque são construídos em termos científico, pois ancoram-se teoricamente na SLC, e tecnológico, pois se prestam ao PLN, precisando, portanto, dar conta dos requisitos e das exigências de um modelo de representação em dois domínios: linguístico-cognitivo e linguístico-computacional.

Para dar conta desses papéis, a representação do *polyset* conforma-se a critérios. Para nos ajudar na formulação desses critérios, tomaremos os propostos por Helbig (2006),¹⁶ para a apresentação do seu robusto modelo de representação do conhecimento baseado em redes e organizado em múltiplas camadas, o MultiNet. O modelo de Helbig pretende representar o conhecimento linguístico como um todo e de maneira independente de língua (nível conceitual) e de aplicação computacional. Tendo em vista que, ao contrário do MultiNet, os *polysets* são construídos para servir a propósitos específicos, descrição linguística e aplicação computacional, nem todos os

¹⁵ A denominação *polyset* é motivada pela denominação *synset*. Enquanto este é um constructo criado para representar um conjunto de itens lexicais que compartilham num mesmo conceito, aquele é um constructo que propomos para representar um conjunto de sentidos que se aproximam por polissemia.

¹⁶ Di Felippo (2008) explorou o modelo MultiNet para a construção de ontologias aplicada à WN.Br.

critérios propostos por Helbig (2006) serão aproveitados. Feita a necessária adaptação, chegamos ao seguinte perfil a que o *polyset* deve atender:

1. Requisitos gerais:

- a) **Adequação cognitiva:** as estratégias representacionais devem possibilitar a modelagem de estruturas conceituais humanas (na medida em que são conhecidas) e sua manifestação na semântica das línguas naturais.
- b) **Interoperabilidade:** os meios de representação devem ser aplicáveis tanto a investigações linguísticas quanto linguístico-computacionais;
- c) **Homogeneidade:** os meios de representação devem ser capazes de expressar os diferentes tipos de sentido polissêmico e, além disso, devem possibilitar a integração da descrição semântica de um item específico à descrição semântica como um todo;
- d) **Reutilização:** os critérios de construção e os constructos resultantes devem ser explícitos e precisos a fim de possibilitar sua reutilização por outras equipes de pesquisadores;
- e) **Viabilidade:** os constructos devem ter potencial de tratamento técnico e implementacional;
- f) **Automaticidade:** os meios descritivos devem permitir o processamento automático do conhecimento.

2. Requisitos internos

- a) **Completeness:** os sentidos polissêmicos instanciados em uma língua natural devem ser representados pelos recursos descritivos previstos pelo modelo;
- b) **Granularidade ótima:** diferentes níveis de granularidade de sentidos devem ser inseridos no modelo;

- c) **Consistência:** deve-se cuidar para que não sejam derivadas informações inadequadas ou logicamente contraditórias a partir de outras inseridas no modelo;¹⁷
- d) **Multidimensionalidade:** os sentidos devem ser descritos de maneira a expressar sua estrutura multidimensional, que é influenciada por diferentes níveis de representação do conteúdo semântico;
- e) **Interpretabilidade local:** os construtos, embora devam ser logicamente interpretáveis por si, de maneira independente de sua inclusão em uma base de conhecimento sobre o léxico da língua, devem ser compatíveis à estrutura geral desse léxico.

É com esse perfil que os componentes e a estrutura do *polyset* serão propostos. Ao final da Seção 7, quando tivermos concluído a apresentação da codificação computacional dos dados, será possível refletirmos sobre a adequação do *polyset* a esse perfil.

5.2 OS COMPONENTES E A ESTRUTURA DO *POLYSET*

Ao longo desta seção passaremos a descrever os componentes e a estrutura do constructo *polyset*. A construção do *polyset* prevê a manipulação dos dois elementos estruturais básicos de uma rede de polissemia: os *nós* e os *arcos*. A construção da rede prevê, por sua vez, a identificação dos itens lexicais polissêmicos e a representação de seus sentidos com base em *cópus*. Os componentes do *polyset* são itens lexicais, conjuntos de sentidos e relações de polissemia, conforme descreveremos a seguir.

¹⁷ Para isso, recomenda-se a aplicação combinada dos testes e das etapas de análise dos sentidos polissêmicos.

5.2.1 Os nós e a caracterização dos sentidos polissêmicos

Os nós da rede representam os sentidos polissêmicos e os *arcos*, as relações de polissemias que os unem. O nó pode instanciar uma unidade simples ou complexa, podendo ser descrito por uma breve glosa, por um hiperônimo, ou por um sinônimo. Para a descrição do conteúdo dos nós, identificamos um item lexical polissêmico e analisamos o conteúdo veiculado por ele, com base em córpus e em recursos léxicos. Como córpus, utilizamos, preferencialmente, a *web* “como córpus”, através de buscas simples no sistema Google¹⁸. Em casos especiais, se necessário, recorreremos ao Córpus NILC¹⁹. Como recursos léxicos, usamos dicionários monolíngues (WEISZFLOG, 1998, WEBSTER’S, 1997), FERREIRA, 1999, HOUAISS, 2001, BORBA, 2002; “Random House Webster’s”, 1994), bilíngues (HOUAISS, CARDIM, 1982), AULETE (2009) e *wordnets* (WN.Br, WN.Pt, WN.Pr).

A seguir, caracterizaremos cada componente da estrutura do *polyset*, são eles: *item lexical*, *constante*, *sentidos* e *conjuntos de sentidos*.

5.2.1.1 O item lexical

No modelo que estamos propondo, os itens lexicais instanciados em uma língua particular são, conforme discutimos na Seção 3, entidades simbólicas que estabelecem o pareamento entre forma (estrutura fonológica) e sentido (estrutura semântica/conceitual). Como já assinalara Saussure, os critérios dessa associação são definidos por padrões de convenção. É preciso lembrar, também, que um item lexical com múltiplos sentidos é aquele que ativa um determinado tipo de rotina cognitiva, de modo que uma mesma estrutura fonológica provê acesso a diferentes porções relacionadas do espaço conceitual. Classificar um item lexical como polissêmico pressupõe, assim, entender que seu conteúdo é formado por sentidos múltiplos que apresentam algum tipo de similaridade.

¹⁸ Disponível em <http://www.google.com.br/>

¹⁹ Disponível em <http://www.linguateca.pt/CETENFolha/>

Os itens lexicais substantivos do PB foram extraídos, fundamentalmente, dos próprios textos teóricos estudados, tanto no domínio linguístico quanto no computacional; poucos foram selecionados diretamente no *cópus*. Fizemos essa opção com o objetivo de testar a validade do modelo representacional proposto para resolver questões discutidas ao longo do trabalho e que são frequentes na literatura sobre o tema.

5.2.1.2 A constante

A denominação *constante* é empregada para descrever a *estrutura fonológica/grafêmica* de um item lexical, por esta ser a porção formal comum aos sentidos que compõem o *polyset*. A constante tem papel fundamental no modelo, visto que o reconhecimento da polissemia passa pela identificação de uma mesma estrutura fonológica/grafêmica (simples ou complexa) associada a uma estrutura semântica complexa, ou seja, a múltiplos sentidos relacionados. Em termos práticos, a constante tem aqui duas funções principais: servir como rótulo ou identificador do *polyset* e como chave de busca nas *wordnets*, nos dicionários e nos *cópus*.

5.2.1.3 Os sentidos

Sob a denominação *sentido polissêmico*, incluímos tanto sentidos plenos quanto subsentidos (facetadas e microsentes). Como dissemos, eles são os *nós* da rede de polissemia que estrutura o *polyset*. Sem perder de vista os princípios da SLC, consideraremos os sentidos entidades cognitivas convencionalmente associadas aos itens lexicais e construídos no uso que se faz da língua. Dessa maneira, qualquer registro ou representação da semântica lexical adquire aqui o *status* de abstração a partir dos dados e, portanto, um "constructo" (KILGARRIFF, 2006, p. 29).

Conforme mencionamos, os itens lexicais polissêmicos analisados foram extraídos fundamentalmente dos textos estudados. A seleção, por sua vez, se deu em dois momentos: a análise de ocorrências mais ilustrativas de polissemia selecionadas da

literatura e a análise da frequência dessas ocorrências no *cópus* de referência. A identificação dos sentidos foi feita a partir da análise combinada de diferentes fontes. A primeira etapa previu a análise (a) do inventário de sentidos polissêmicos extraídos dos dicionários monolíngues do PB (WEISZFLOG, 1998, WEBSTER'S (1997), FERREIRA, 1999, HOUAISS, 2001, BORBA, 2002); (b) dos *synsets* da WN.Br²⁰ e (c) das informações disponíveis na WN.PT, ambos os recursos são limitados por estarem em construção. A segunda etapa previu a identificação dos itens lexicais do inglês que são equivalentes aos do português. Os itens lexicais do inglês desempenham o papel de chave de busca para a consulta à WN.Pr. Para esse processo, tomamos como referência o dicionário inglês-português português-inglês "Dicionário Eletrônico Webster's" (HOUAISS, CARDIM, 1982) e o Dicionário Inglês-Ingês Eletrônico "Random House Webster's" (1994). O terceiro passo foi a consulta à WN.Pr, incluindo a análise das glosas nela propostas para cada sentido e das relações semânticas nela incluídas. O uso dos dicionários e das bases *wordnets* como fontes de busca dos sentidos se justifica por serem esses registros de usos bem estabelecidos, fornecendo um repertório desejável e controlado de informações para os fins deste trabalho. Por fim, enfatizamos que, concomitantemente às consultas aos dicionários e às *wordnets*, consultas ao *cópus* de referência (Internet via Google e *CópusNilc*) foram feitas para verificarmos os usos de determinado item lexical, para atestarmos os sentidos registrados nos dicionários ou, ainda, para identificarmos possíveis sentidos novos.

A partir dessas fontes de busca de sentidos, identificamos informações semânticas de tipos diferentes: usos reais registrados no *cópus*, diferentes acepções registradas nos dicionários, glosas e relações semânticas codificadas nas *wordnets*. A partir dessa variedade de informações, definimos estratégias para o registro de cada sentido analisado. Nessa análise, a opção foi pela clareza, objetividade, precisão, além da praticidade. O conteúdo dos *nós*, portanto, pode ser registrado a partir de um dos seguintes elementos:

²⁰ Trata-se da base derivada do *Thesaurus eletrônico do Português*: um dicionário eletrônico de sinônimos e antônimos formado por cerca de 44 mil unidades lexicais (Dias-Da-Silva, 2003a; Dias-Da-Silva e Moraes 2003).

(a) Uma glosa reduzida

A glosa da rede de polissemia deve ser entendida como uma “adaptação reduzida” da tradução livre da glosa da WN.Pr²¹, combinada com a definição dos dicionários consultados. A construção da glosa que propomos para as unidades de sentido da rede de polissemia observa o seguinte critério: usar o número mínimo de itens lexicais necessários para expressá-la. Por exemplo, para o *polyset*, a glosa da WN.Pr exemplificada na nota 7 é reduzida para “objeto físico”. Como a informação da glosa não é manipulável pelo sistema computacional, julgamos que essa redução não compromete o desempenho do sistema e é suficiente para auxiliar o analista a identificar o sentido por ela codificado.

(b) Um hiperônimo suficientemente específico

A informação do hiperônimo de um item lexical orienta a análise de delimitação do sentido lexical. Por essa razão, ele é usado, como conteúdo do *nó* da rede, quando ele contribuir de modo mais preciso e informativo para a delimitação do sentido. Por exemplo, um dos sentidos do item lexical *base* registrado nos dicionários e na WN.Pr é assim especificado: “tudo o que serve de apoio ou suporte”; o seu hiperônimo direto na WN.Pr, em PB, é “suporte”. Notamos que o hiperônimo identifica, em apenas uma palavra, o que a especificação identifica por meio de uma frase. Logo, nesse exemplo, a opção é pelo hiperônimo.

(c) Um sinônimo representativo

Um sinônimo é usado para preencher o *nó* da rede de polissemia quando ele puder ser substituído, com o mínimo de alteração semântica, em um número grande de contextos²². Por exemplo, uma das glosas da WN.Pr associadas ao item lexical *posição* é “the act of putting something in a certain place or location” e o *synset*

²¹ Nas *wordnets*, a glosa é entendida como uma definição²¹ informal que representa o sentido codificado no *synset*, ou seja, o *nó* da rede. Um exemplo de glosa da WN.Pr é, para o *synset* {*book, volume*}, “physical objects consisting of a number of pages bound together”.

²² É importante que se diga que não se pretende com isso considerar que o sinônimo escolhido é absoluto, fenômeno raro ou inexistente.

correspondente é {placement, location, locating, position, positioning, emplacement}. Nesse exemplo, usar *localização*, tradução de *location*, componente do *synset*, é suficiente para descrever o conteúdo do *polyset*.

(d) Uma noção geral

Um elemento que usamos para representar o registro do sentido de um item lexical que exhibe polissemia irregular é uma *noção geral*, que codifica um conteúdo semântico esquemático e não tem o mesmo *status* dos mencionados acima, pois serve como uma noção globalizante que representa a similaridade entre os sentidos e não os sentidos propriamente ditos. Lembramos que esse tipo de noção, embora generalizante, não tem *status* de hiperônimo e deve permitir a delimitação de grupos de sentidos. Para a identificação da noção geral, consideraremos as informações fornecidas pelos dicionários, pelas *wordnets*, pelo *corpus*, além do conhecimento do analista sobre usos convencionalmente associados ao item em análise. Esse tipo de informação tem a função de rotular um agrupamento de sentidos, explicitando o tipo de similaridade percebida entre eles e serve, portanto, como elemento auxiliar na análise e estruturação do *polyset*. O exemplo citado ao longo do trabalho foi a noção de “custódia”. Essa noção indica um grupo de sentidos associados a *banco*. Já as noções de “entidade, coisa ou local”, visto que não são capazes de distinguir o que é um *banco*, não são adequadas.

Para avaliarmos o impacto quantitativo do modelo de *polysets* na organização da WN.Br, fez-se uma filtragem automática dos dados disponíveis na base²³. A estatística mostrou que de um total de 16.925 entradas para nominais, 5.957 são codificados como polissêmicos, segundo a noção grosseira de polissemia adotada pelos idealizadores da base. Esses números são importantes para revelar o impacto quantitativo do modelo proposto para a organização geral da base.

Este nível inicial de análise permitiu que construíssemos o primeiro inventário dos múltiplos sentidos associados a um item lexical. Nesta fase, situações de monossemia (ex.: criança, tio, meia-idade) foram desconsideradas. Esse primeiro

²³ A filtragem dos dados da base da WN.Br tem sido feita pela equipe do NILC-SC, grupo parceiro do projeto de construção da WN.Br.

inventário exigiu outra filtragem: eliminação dos itens lexicais homônimos, tarefa que seguiu esta análise inicial.

5.2.1.4 Um conjunto de sentidos

Identificados os sentidos que marcam os *nós* da rede de polissemia, o próximo passo é construir o *conjunto de sentidos* de acordo com a percepção de alguma similaridade. Os múltiplos sentidos delimitados na análise inicial constituem apenas uma lista (estrutura homonímica). A partir dessa lista, tendo em vista que a estratégia polissêmica é a preferencial em um *polyset*, em razão das diversas vantagens computacionais e cognitivas observadas ao longo do trabalho, o objetivo é especificar os fatores de coerência que se estabelecem entre os sentidos. Sempre que eles forem especificados, constrói-se um conjunto de sentidos, o *polyset* propriamente dito. A partir desse segundo nível de construção do *polyset*, temos a indicação de alguma similaridade, mas ainda não se sabe com precisão de que tipo ela é.

Podemos assim dizer que os *polysets* complementam as informações associadas aos *synsets*, refinando-os semanticamente e são, ao mesmo tempo, mais livres em sua estrutura, uma vez que o sentido pode ser composto de elementos que fazem parte de níveis distintos de generalidade: podemos ter uma glosa, um hiperônimo ou um sinônimo. Um exemplo de diferenças entre os componentes do *polyset* é dado na comparação dos conjuntos de sentidos que identificamos para os itens lexicais *livro*, enquanto ‘objeto físico’ e ‘conteúdo’ (glosa simplificada) e *posição*, enquanto ‘localização’ e ‘opinião’ (sinônimos). O processo de construção dos agrupamentos de *nós* da rede previu a consulta a todas as fontes de sentidos mencionadas acima e levou, também, em consideração os fatores de coerência e de autonomia, além do tipo de relação de sentido e dos efeitos de saliência.

5.2.1.5 A construção dos *polysets*: a delimitação dos tipos sentidos

A tarefa de delimitação dos tipos sentidos constitui a primeira fase específica na construção do *polyset*. Nas duas etapas anteriores, a tarefa foi a de organizar e registrar as informações disponibilizadas em diferentes fontes. Assim, a partir deste ponto é que passamos a apresentar a proposta de representação dos diferentes tipos de sentido identificados nas seções 3 e 4.

Entre os sentidos polissêmicos, na seção 4, identificamos sentidos semi-distintos e sentidos cooperativos, conforme o grau de delimitação. Os sentidos semi-distintos, os irregulamente polissêmicos, são os sentidos plenos que têm um grau médio de antagonismo e um grau significativo de autonomia. Os sentidos cooperativos apresentam certo grau de autonomia, com um grau de antagonismo menor ou inexistente. Conforme são delineadas as fronteiras entre os diferentes sentido, as diferentes leituras podem ser classificadas como *sentidos plenos* ou subsentidos do tipo *facetadas* e *microsentidos*. Sob o rótulo *sentido pleno*, incluímos aqueles que apresentam distinções marcantes, bem estabelecidos, incrustados no léxico da língua e fortemente restritos por restrições convencionais estáveis (CROFT e CRUSE, 2004). Sob o rótulo de *subsentidos*, incluímos aqueles que apresentam grau de antagonismo baixo ou inexistente. Há, portanto, a possibilidade de unificação, que pode ser de diferentes tipos: enquanto as facetadas são entendidas como partes de um mesmo sentido geral, os microsentidos são entendidos como especializações desse sentido geral.

Feitas essas contextualizações, passamos ao foco da seção: inventariar, no quadro 6, os fatores de coerência, ou seja, as evidências de similaridade, e os fatores de discretude, ou seja, as evidências de autonomia entre os sentidos. Para o inventário apresentado aqui, selecionamos aqueles critérios da Seção 4 que julgamos os mais relevantes para a construção de um *polyset*.

| TIPOS DE SENTIDO | EVIDÊNCIAS DE COERÊNCIA | EVIDÊNCIAS DE AUTONOMIA |
|------------------|---|---|
| Sentidos plenos | 1. Presença da noção geral
2. Ocorrência marginal do efeito zeugmático | (1) Grau de seleção de atenção máximo (Como o leitor é obrigado a selecionar apenas um dos sentidos no âmbito do enunciado, a delimitação entre eles ocorre já no contexto linguístico sentencial)
(2) Possibilidade de identificação de um sentido mais neutro e outro menos neutro, aplicável a contextos mais restritos
(3) Presença de autonomia relacional |

| | | |
|--|--|--|
| | | (4) Presença de autonomia composicional |
| Subsentidos do tipo facetas | (1) Capacidade de unificação (as facetas do sentido são partes de um mesmo sentido)
(2) Inexistência de antagonismo
(3) Grau de autonomia marginal.
(4) Grau de seleção de atenção marginal.
(5) Ocorrência marginal do efeito zeugmático.
(6) Coocorrência de leitura prototípica.
(7) Presença de propriedades composicionais unificadoras | (1) Presença de autonomia relacional (as facetas de sentido têm hiperônimos distintos a observação sobre o tipo de hiperônimo possibilita a percepção de classes regulares de polissemia)
(2) Presença de núcleos autônomos, isto é, ocorrência de independência na ativação das facetas
(3) Grau de seleção de atenção marginal (não há necessidade de seleção de apenas uma das facetas) |
| Subsentidos do tipo microsentidos | (1) Presença de hiperônimo comum (importante fator unificador)
(2) Grau mínimo de antagonismo (os microsentidos são co-autohipônimos) | (1) Microsentidos em relação de incompatibilidade mútua (co-autohipônimos).
(2) Presença de autonomia relacional
(3) Presença da autonomia de condições de verdade.
(4) Presença de autonomia de domínio. |

Quadro 6 - Evidências de coerência e de autonomia

5.2.1.6 Efeitos de saliência entre os sentidos

A análise dos efeitos de saliência entre os sentidos constitui a segunda fase de construção de um *polyset*. Sua identificação precisa, no entanto, somente é concluída a partir da terceira atividade, quando se os rotulam arcos. Aqui, destacamos duas questões fundamentais: a identificação do núcleo da categoria e a identificação dos efeitos de saliência, tanto a psicológica como a estrutural.

No contexto das discussões teóricas das seções 3 e 4, entende-se que, de acordo com os critérios semânticos sincrônicos de delimitação do núcleo de uma categoria de sentidos, podemos ter os seguintes cenários típicos: (a) sentido único – quando há um sentido saliente que se sobrepõe aos outros, caracterizando a polissemia irregular; (b) sentidos duplos e independentes – quando a independência é parcial, caracteriza-se a polissemia irregular ou a polissemia regular do tipo autohiponímia; (c) sentidos duplos e sobrepostos – quando há a ocorrência de um modo independente e caracterizando a polissemia regular do tipo automeronímia.

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 *cópus*, conforme descreveremos.

5.2.1.4.1. Saliência psicológica

A detecção do efeito da saliência psicológica funcional para a delimitação do núcleo de uma categoria de sentido pressupõe a análise do(s) sentido(s) que ocorrem(m) com um número maior de ligações. 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. Embora a primeira razão seja suficiente para os propósitos da análise sincrônica, é relevante mencionar que o critério é também válido para a identificação dos sentidos primários em abordagens que fazem interface entre a sincronia e a diacronia, mais comuns na SLC. O fato de o sentido ocorrer com um número maior de relações (os *links*) pode ser indicativo de que os seus componentes estão mais frequentemente presentes em outros sentidos.

5.2.1.4.2 Saliência estrutural

O efeito de saliência estrutural, ou de distinção, está relacionado à frequência de ocorrência de um sentido em relação a outro. Para a contagem do peso estrutural de um sentido integrante de um *polyset*, é feita uma busca pelo item lexical sob análise na web como cópula, analisando-se as 10 primeiras ocorrências trazidas pelo motor de busca do Google™. Esse exercício tem o objetivo de ilustrar a metodologia, visto que o número de ocorrências analisadas para cada item é pequeno. É importante notar também que o resultado das buscas é subordinado à ocorrência do sentido entre as páginas mais consultadas.

5.2.2 Os arcos da rede de sentidos polissêmicos

Os *arcos* de uma rede de sentidos polissêmicos representam as *relações* entre os diferentes sentidos polissêmicos. São as relações as responsáveis por dar forma e por justificar o agrupamento dos sentidos em um *polyset*. É somente após a conclusão dessa fase de análise que o *polyset* é caracterizado por completo. De maneira análoga ao que fizemos na apresentação da teoria na Seção 4, o processo de especificação das relações na modelagem dos dados prevê duas fases. A primeira fase, descrita anteriormente, é mais geral, e consiste na detecção da existência ou não de relação entre os sentidos para

a composição do conjunto de sentidos. A segunda fase é mais específica, e consiste na explicitação do tipo de relação que rotula os arcos. O tipo de codificação da relação determina (a) o tipo de grau de generalidade ou de especificidade entre os sentidos, (b) o sentido mais saliente da rede (isto é, aquele que tem maior peso estrutural por ocorrer com um número maior de *links*) e (c) o grau de produtividade da combinação de sentidos no léxico como um todo. Nas subseções seguintes, descreveremos as propriedades essenciais de cada uma das relações já discutidas na Seção 4.

5.2.2.1 A relação de polissemia regular

Há sentidos que são relacionados pela polissemia regular do tipo *automeronímia/autoholonímia* (relação de PARTE-TODO). A relação ocorre quando os subsentidos (as facetas) associam-se por relação de *parte-todo* (ex.: porção/substância, objeto físico/conteúdo). A similaridade entre as facetas de sentido é alta, de maneira que ambas as facetas são componente distintos de um mesmo sentido geral. Como a principal característica das facetas de sentido é a de elas apresentam grau mínimo de autonomia, não é necessário delimitar apenas uma delas em um contexto, embora seja comum que a cada contexto uma esteja em relevo. Dos exemplos que referimos ao longo do trabalho, destacamos *livro* ‘objeto físico’ e *livro* ‘conteúdo’. O rótulo da relação é *é_automerônimo_de* e o do seu reverso é *é_autoholônimo_de*. Para a identificação da relação regular de polissemia do tipo 2, a automeronímia, a aplicação da seguinte fórmula lógica precisa ser verdadeira: A é *parte de* B, e A e B são facetas de sentido. A regularidade da relação se dá na medida em que podemos perceber outros pares de sentidos relacionados também por automeronímia. Lembramos que a automeronímia abarca uma família de relações que não constituem partes em sentido estrito, tais como *objeto-substância*, *continente-conteúdo*, *objeto físico-conteúdo*, etc. O rótulo da relação é *é_automeronímia_de* e o do seu reverso é *é_autoholonímia_de*.

Há os sentidos que são relacionados pela polissemia regular do tipo *autohiponímia/autosuperordenação* (relação de *É-UM*). Essa relação ocorre quando um dos subsentidos (microsentidos) é construído de modo mais específico em relação aos outros. A similaridade entre os microsentidos sentidos é funcional, de maneira que o

microsentido subordinado é um subtipo do superordenado e herda todas as suas características, acrescentando, porém, outras que o distinguem dos seus co-autohipônimos. Como ocorre em relações hierárquicas, os co-autohipônimos em um mesmo nível hierárquico são mutuamente incompatíveis. Dos exemplos que mencionamos ao longo do trabalho, destacamos *faca* como ‘utensílio doméstico’, ‘arma branca’ e ‘instrumento cirúrgico’. O rótulo da relação é *é_automerônimo_de* o do seu reverso é *é_autosuperordenado_de*. Para a identificação da relação regular de polissemia do tipo 1, a autohiponímia, a aplicação da seguinte frase lógica precisa ser verdadeira: *A é um B*, ou *A é um tipo de B*, e *A* e *B* são microsentidos. A regularidade da relação se dá na medida em que detectamos outros pares de sentidos relacionados também por autohiponímia.

5.2.2.2 A relação de polissemia irregular

Há sentidos com alto grau de independência uns dos outros, mas, mesmo assim, é possível que se perceba alguma relação de similaridade (objetiva ou subjetiva) entre eles. Ao contrário das relações que se detectam na polissemia regular, trata-se de uma relação horizontal, não havendo herança entre os sentidos. O que ocorre é uma comparação entre os sentidos de modo que podemos dizer que há algum grau de similaridade entre eles. Esse tipo de similaridade caracteriza a polissemia irregular, situação em que dois ou mais sentidos, com grau médio de antagonismo e significativo grau de autonomia, associam-se de algum modo a uma mesma “noção geral”. Entre os exemplos apresentados ao longo do trabalho, destacamos *cabeça* ‘membro do corpo humano’ e *cabeça* ‘chefe’ e *posição* ‘local ocupado’ e *posição* ‘opinião’. O rótulo para essa relação é *é-similar-a*. A aplicação da seguinte frase lógica precisa ser verdadeira: Um item lexical *A* com os sentidos a_1 e a_2 apresenta polissemia irregular se **NÃO** existir nenhuma outra unidade lexical *B* com sentidos b_1 e b_2 que sejam semanticamente distintos um do outro exatamente de maneira diferente da distinção entre a_1 e a_2 .

Para sintetizar a seção 5.2.2, apresentamos o quadro 7, que resume as propriedades centrais de cada relação.

| PROPRIEDADES | POLISSEMIA REGULAR | | POLISSEMIA IRREGULAR |
|----------------------|--|--|--|
| | Rótulo | | Rótulo |
| | <i>é_automerônimo_de</i> ou
<i>é_autoholônimo_de</i> | <i>é_autohipônimo_de</i>
ou <i>é_autosuperordenado_de</i> | <i>é_similar_a</i> |
| Tipo de sentido | Faceta | Microsentido | Pleno |
| Tipo de Relação | Regular | Regular | Irregular |
| Tipo de similaridade | Hierárquica
Relação entre parte-todo e
variantes | Hierárquica
Relação entre geral e específico | Horizontal
Relação objetiva ou subjetiva a
uma noção geral |
| Grau de similaridade | Alto | Funcional | Intermediário |
| Grau de autonomia | Mínimo | Incompatibilidade mútua de
microsentidos (co-
autohipônimos) | Alto |
| Grau de antagonismo | Inexistente | Mínimo | Médio |
| Fórmula lógica | [<i>A é parte de B</i> , e A e B são facetas
de sentido] | [<i>A é um B</i> , ou <i>A é um tipo de B</i> , e A
e B são microsentidos] | Um item lexical A com os sentidos
a1 e a2 apresenta polissemia
irregular se NÃO existir nenhuma
outra unidade lexical B com
sentidos b1 e b2 que sejam
semanticamente distintos um do
outro exatamente de maneira
diferente da distinção entre a1 e
a2 |

Quadro 7 – Propriedades centrais da relação de polissemia.

5.2.3 Síntese da seção

Ao longo desta seção, caracterizamos cada componente de um *polyset*, apresentando sua função e funcionamento na ao longo dos três etapas gerais de construção do modelo. O quadro 8 resume as principais etapas de sua construção

| A CONSTRUÇÃO DO <i>POLYSET</i> | | | |
|---------------------------------|---|-------------------------------------|----------------------------------|
| | Etapas preliminares | | Etapas de construção |
| | Definição dos <i>nós</i> da rede | Indicação dos conjuntos de sentidos | Construção da rede de polissemia |
| 1 Item lexical polissêmico | Identificar | x | x |
| 1.2 Constante | Identificar | x | x |
| 1.3 Sentidos polissêmicos | Delimitar potencial de sentidos e <i>status</i> | Analisar | Analisar |
| 1.3.1 glosas | Construir | Analisar | Analisar |
| 1.3.2 noção geral | Identificar | Analisar | Analisar |
| 1.3.3 <i>córpus</i> | Selecionar e analisar | Analisar | Analisar |
| 1.3.4 <i>wordnets</i> | Consultar | Consultar | Analisar |
| 1.3.5 dicionários | Consultar | Consultar | Analisar |
| 2. Grupos de sentidos | x | Formar | Rotular |
| 3. <i>Status</i> de delimitação | Analisar | Analisar | Especificar |
| 3.1 Sentido pleno | x | Analisar | Identificar |
| 3.2 Subsentido | x | x | x |
| 3.2.1 Faceta | x | Analisar | Identificar |
| 3.2.1 Microsentido | x | Analisar | Identificar |
| 4. Efeitos de saliência | x | Analisar | x |
| 4.1 Psicológica | x | x | Identificar |
| 4.2 Estrutural | x | x | Identificar |
| 5. Tipo de relação de sentido | x | Identificar a existência | Especificar |
| REGULAR | x | x | Indicar |
| 5.1 automeronímia | x | x | Indicar |
| 5.2 autohiponímia | x | x | Indicar |
| IRREGULAR | x | x | Indicar |
| 5.3 similaridade | x | x | Indicar |

Quadro 8 – Níveis gerais de estruturação de um *polyset*

Conforme descreve o quadro 8, um *polyset* é construído em dois grandes níveis: etapas preliminares de construção do *polyset* e etapas de construção do *polyset*. É importante destacar que ambos os momentos têm um funcionamento integrado e mutuamente cooperativo. Em síntese, as tarefas desenvolvem-se conforme a descrição que segue.

As **etapas preliminares de construção do *polyset*** têm dois objetivos gerais: definir os *nós* da rede e indicar os conjuntos de sentidos. Para determinar os *nós* que comporão a rede de polissemia que estrutura o *polyset*, é necessária a realização das seguintes tarefas: (i) a identificação do item lexical polissêmico no PB; (ii) a delimitação do seu potencial de sentido (os *nós* da rede). Como exemplo, para o item lexical *livro*, selecionamos os sentidos *livro* ‘objeto físico’ e *livro* ‘conteúdo’. Para indicar os conjuntos de sentidos relacionados, com base nas informações delimitadas no nível anterior e nas informações fornecidas pela WN.Pr, são construídos, para cada sentido identificado,

conjuntos de sentidos sem a indicação ainda do tipo de relação. Cada conjunto de sentidos é identificado por uma constante que representa o item lexical polissêmico em análise. Essa análise é útil não só para a tarefa interna ao processo de construção de *polyset*, mas também para a indicação de uma possível integração entre a análise semasiológica do *polyset* e a análise onomasiológica do *synset*, pois o sentido do item lexical em questão precisará ser expresso linguisticamente através de um sinônimo, de um hiperônimo ou de uma glosa simplificada. Voltando ao exemplo de *livro*, o conjunto de sentidos pode ser representado, do seguinte modo $\{\{\text{livro, conteúdo}\}, \{\text{livro, objeto físico}\}, \}_{\text{livro}}$.

As **etapas de construção do *polyset***. Quando chegamos nesse nível, já construímos os *nós* (sentidos) da rede e os agrupamentos entre os sentidos (conjuntos de sentidos). Falta-nos identificar as seguintes informações: (i) o *status* dos sentidos, percebendo (a) a ocorrência de sentido pleno, de faceta ou de microsentido e (b) as diferenças de saliência, e (ii) os rótulos das relações, explicitando o motivo dos agrupamentos. Será neste nível que a rede de polissemia tomará forma, sendo enriquecida com os *nós* e *arcos* rotulados. Os rótulos com os quais trabalhamos são *é-similar-a*, *é-autohipônimo-de*, *é-autosuperordenado-de*, *é-automerônimo-de*, *é-autoholônimo-de*. Além disso, são também indicadas as propriedades regular ou irregular para cada relação. Em formato gráfico, o exemplo de *livro* é ilustrado na figura 19.

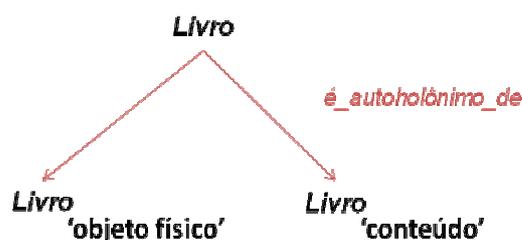


Figura 19 - Representação gráfica dos automerônimos de *livro*

SEÇÃO 6 – CONSTRUÇÃO E IMPLEMENTAÇÃO DO *POLYSET*

Na seção anterior, apresentamos o constructo *polyset* como um modelo de análise e representação linguístico computacional da polissemia de nominais que tem a finalidade de servir à complementação das informações léxico-conceituais disponíveis em uma *wordnet*. O objetivo desta seção é apresentar um exercício de aplicação dessa proposta, relatando os processos desenvolvidos para a construção e implementação do *polyset*. Sendo assim, dividimos a seção em três momentos principais: no primeiro, apresentamos um exercício de análise e representação linguística; no segundo, um exercício computacional; e, no terceiro, destacaremos as principais contribuições do modelo *polyset* para as *wordnets*, em especial, para a WN.Br. Lembramos que a descrição proposta tem *status* múltiplo: o de representação linguística, pois caracteriza um fenômeno que se manifesta na superfície da língua; o de representação linguístico-computacional, pois se configura como uma descrição formalizável das informações linguísticas; e, ainda, o de representação linguístico-cognitiva, pois tem o *status* de “isomorfismo funcional”, entendendo-se que a realidade psicológica de uma representação não é imediata, mas pode ser representativa do que temos na mente (SANDRA e RICE, 1995, p. 102).

Os itens lexicais em análise serão *fonte*, *banco* e *café*. Os exemplos foram escolhidos por (a) serem recorrente na literatura, recebendo tratamentos diversos, (b) por ilustrarem as relações de polissemia estudadas e, sobretudo, (c) por gerarem diferentes configurações para a rede que estrutura o *polyset*. Os sentidos que recheiam os nós da rede de polissemia proposta para cada item lexical não têm o objetivo de esgotar suas possibilidades de uso, mas o de ilustrar o potencial representacional do *polyset*, a partir de sentidos registrados nos dicionários e *wordnets* pesquisados que tiveram ocorrência no *corp*us. A partir desse critério, entendemos que o *polyset* registrará sentidos convencionalizados e relevantes sincronicamente. Para a delimitação e registro dos sentidos, conforme observamos na Seção 5, levamos em consideração as acepções dos recursos léxicos analisados. A descrição dos resultados da etapa linguística foi organizada em quadros síntese, registrando as *etapas preliminares*, e em *redes de polissemia*, representando a configuração do *polyset* para cada um dos itens lexicais

investigados. Os resultados da etapa de implementação computacional serão apresentados a partir das telas do editor Protégé-OWL.

A seção está organizada do seguinte modo: em 6.1, apresentaremos a *face linguística de construção do polyset*; em 6.2, o foco é a *a face computacional de construção do polyset*; em 6.3, são sintetizadas as *contribuições do modelo polyset para a WN.Br*; por fim, em 6.4, sintetizaremos a seção.

6.1 A FACE LINGUÍSTICA DE CONSTRUÇÃO DO *POLYSET*

O objetivo desta seção é descrever, de acordo com o ferramental teórico delimitado nas Seções 3 e 4, as ocorrências de polissemia em nominais do português. Trata-se da segunda etapa do domínio linguístico desta tese responsável pela análise e representação dos dados. Essa tarefa é realizada em dois momentos, conforme referimos anteriormente. O primeiro deles é explicitado no quadro 9, que serve como modelo para a descrição das etapas preliminares de construção do *polyset*, que buscam a definição dos *nós* da rede de polissemia e a indicação dos conjuntos de sentidos.

| DEFINIÇÃO DO NÓ DE UM <i>POLYSET</i> | |
|--------------------------------------|--|
| 1. Constante | Indicar a constante |
| 2. Exemplo | Indicar um exemplo para cada sentido em análise |
| 3. Chave de busca em inglês | Indicar a chave de busca correspondente a cada sentido |
| 4. Glosa reduzida | Indicar (fonte: <i>wordnets</i> e dicionários) |
| 5. Hiperônimo relevante | Entre as possíveis traduções dos hiperônimos fornecidos pela WN.Pr, indicar o hiperônimo mais representativo, quando relevante. O hiperônimo representará o nó da rede somente se for considerado o componente mais relevante para descrever o sentido em análise e suficientemente específico. |
| 6. Sinônimo relevante | Entre as possíveis traduções dos sinônimos fornecidos pela WN.Pr, indicar o sinônimo mais representativo, quando relevante. O sinônimo representará o nó da rede somente se for considerado o componente mais representativo do sentido em análise. |
| 7. Nó do <i>polyset</i> | O conteúdo do nó da rede de polissemia deve servir como referência ao analista humano, que contará também com a frase-exemplo para facilitar/complementar sua delimitação. Seu conteúdo pode ser composto por: <ul style="list-style-type: none"> a. um sinônimo representativo b. um hiperônimo suficientemente específico c. uma glosa reduzida |
| 8. Conjunto de sentidos | Agrupar os sentidos em conjuntos rotuladas pela <i>constante</i> . |

Quadro 9 - Resultado das etapas preliminares de construção do *polyset*

Concluídas as etapas preliminares, passamos ao segundo momento da análise e representação linguística mencionado anteriormente. Elaboraremos, como resultado das etapas de construção do *polyset*, uma rede de sentidos polissêmicos, representando graficamente as diferentes relações de polissemia e os efeitos de saliência entre os sentidos. Lembramos que o formato da rede não segue nenhum padrão pré-definido (tal como as redes radiais e esquemáticas analisadas na Seção 3), pois, conforme mencionamos na seção anterior, ela é inspirada nos modelos de Geeraerts (2006) e Blank (2003). O quadro 10 apresenta uma síntese das tarefas dessa fase.

| CONSTRUÇÃO DA REDE DE POLISSEMIA | |
|----------------------------------|---|
| 1. Organização da rede | Nesta etapa é construída uma rede manual que ilustra a distribuição dos sentidos. |
| 2. Rótulo das relações | Os rótulos com os quais trabalharemos são <i>é-similar-a</i> , <i>é-autohipônimo-de</i> , <i>é-autosuperordenado-de</i> , <i>é-automerônimo-de</i> , <i>é-autoholônimo-de</i> . |
| 3. Noção geral | *Componente identificado somente em itens irregularmente polissêmicos. Ele serve como especificação do tipo de similaridade. |

| | |
|----------------------------------|--|
| 4. Status dos sentidos I | Os tipos são: sentido pleno, faceta, micros sentido. O tipo de sentido ficará evidente no modelo a partir da organização hierárquica e das relações entre sentidos. |
| 5. Status dos sentidos II | |
| a. saliência psicológica | Determinada pela predominância. |
| b. saliência estrutural | Frequência no cópuz.
Lembramos que, para essa análise, fizemos um consulta aos primeiros 10 resultados de uma busca no Google. Tarefa que serve como ilustração da saliência, mas que é quantitativamente limitada. |

Quadro 10 - Resultado das etapas de construção do *polyset*

6.1.1 Análise do item lexical fonte

| A DEFINIÇÃO DOS NÓS | |
|-------------------------------------|--|
| 1. Constante | Fonte |
| 2. Exemplos | <p>1. O aproveitamento de uma fonte de água requer o estudo geológico completo de toda a área, no qual se deve levar em conta as condições de poluição dos arredores, da área de captação e da área de infiltração, e o isolamento da fonte, para evitar qualquer contaminação bacteriana.²⁴</p> <p>2. A presente fonte ornamental cibernética é provida de um número adequado de bocais variados de saída de água constando de meios ejetores de água, meios reguladores de caudal e meios de pilotagem e ligação elétrica com um microprocessador de controle individual dos bocais e complementada por uma série de elementos decorativos e espetaculares complementares coordenados entre si, facilmente aplicável às fontes instaladas em praças, parques e outros sítios com fins decorativos.²⁵</p> <p>3. As formigas invadem o videocassete e vão fazer ninho na fonte, peça que engloba o transformador de voltagem.²⁶</p> <p>4. Os exemplos mostram que a multiplicidade de fontes científicas que a web oferece representa uma dificuldade para os usuários, quando eles têm de avaliar e selecionar as informações.²⁷</p> <p>5. "A experiência brasileira com biocombustíveis e fontes renováveis de energia mais limpas e baratas abre a perspectiva de mais emprego e renda no campo e é nossa resposta ao desafio da mudança do clima", afirmou Lula a Koroma, primeiro líder de Serra Leoa a visitar o Brasil²⁸.</p> <p>6. A televisão continua a ser a principal fonte de informação para a maioria dos europeus, mesmo tendo se tornado menos informativa e mais sensacionalista.²⁹</p> |
| 1. Chaves de busca em inglês | <p>1-2 <i>fountain</i></p> <p>3-6. <i>source</i></p> <p>*7. <i>font (typeface)</i></p> |
| 2. Hiperônimos relevantes | <p>1. formação geológica</p> <p>2. estrutura, construção</p> <p>3. -</p> <p>4. publicação</p> <p>5. -</p> |

²⁴ Fonte: <http://www.emdiv.com.br/pt/mundo/asmaravilhas/1364-fonte-nascente.html>

²⁵ Fonte: www.patentesonline.com.br/fonte-ornamental-cibernetica-77535.html

²⁶ Fonte: CórpusNilc.

²⁷ Fonte: www.comciencia.br/reportagens/cultura/cultura10.shtml

²⁸ Fonte: <http://g1.globo.com/Noticias/Mundo/0,,MUL1273371-5602,00-LULA+OFERECE+AJUDA+A+SERRA+LEOA+PARA+PRODUZIR+BIOCOMBUSTIVEIS.html>

²⁹ Fonte: <http://www.observatoriodaimprensa.com.br/artigos.asp?cod=351MON005>

| | |
|---|---|
| | 6. comunicador |
| 3. Sinônimos relevantes | <ol style="list-style-type: none"> 1. nascente 2. chafariz 3. cabo de energia 4. bibliografia, documentação 5. origem, causa 6. informante |
| 4. Glosas WN.Pr/acepções de dicionários (relevantes) | <ol style="list-style-type: none"> 1. Local em que brota a água proveniente do solo (AULETE, 2009) 2. Instalação para aproveitar a água de nascente natural, ou mesmo água encanada. (HOUAISS, 2001) 3. Circuito de eletricidade (BORBA, 2002) 4. Texto de apoio ou referência. (BORBA, 2002) 5. Aquilo que dá origem, matriz, procedência. (combinação de sentidos apresentados no dicionário Houaiss, 2001) 6. Pessoa que fornece informações a repórteres. (HOUAISS, 2001) |
| 5. Nós do <i>polyset</i>_{fonte} | <ol style="list-style-type: none"> 1. Nascente de água 2. Instalação para aproveitar a água de nascente natural, ou mesmo água encanada 3. Circuito de eletricidade 4. Texto de referência 5. Origem, causa, procedência 6. Que ou quem transmite informações |
| 6. Conjuntos de sentidos | <p>{ ['nascente de água', 'instalação para aproveitar a água de nascente natural, ou mesmo água encanada']₁₋₂, ['origem', causa, procedência, 'circuito de eletricidade', 'texto de referência']₃₋₆ }_{fonte}</p> |

Quadro 11 – Resultado da etapa preliminar de construção do *polyset*_{fonte}

6.1.1.2 ETAPAS DE CONSTRUÇÃO DO *POLYSET*_{fonte}:

A rede de polissemia para os sentidos associados a *fonte* está representada na figura 20.

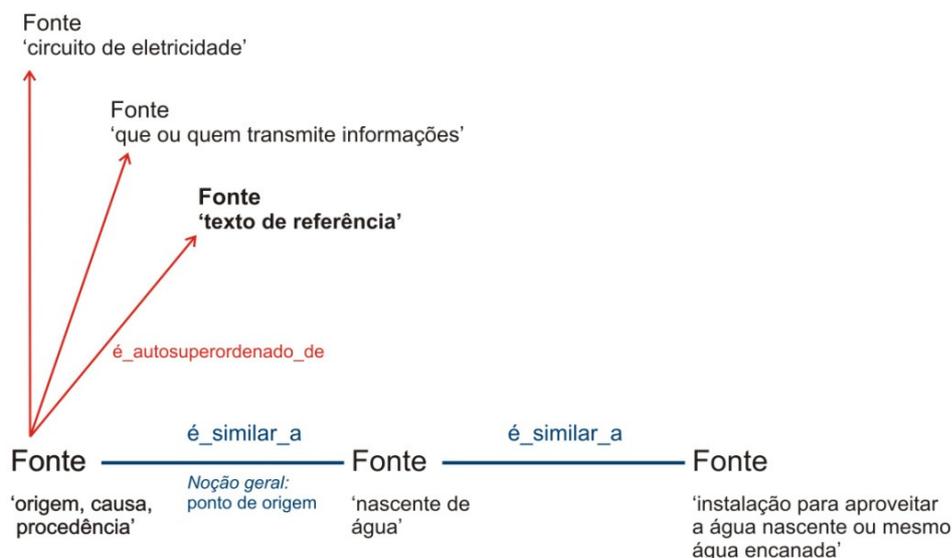


Figura 20 - Polyset fonte

Os sentidos associados a *fonte* selecionados pelos exemplos (1-6) ilustram uma situação em que uma mesma unidade fonológica provê acesso a um conteúdo conceitual delimitado de formas distintas. Conforme representado na figura 20, os sentidos estão associados tanto por polissemia irregular (em azul) quanto e por polissemia regular (em vermelho). A relação de polissemia irregular, expressa através do rótulo *é semelhante a*, caracteriza uma aproximação subjetiva entre 'nascente de água', 'origem, causa, procedência' e 'instalação para aproveitar a água de nascente natural, ou mesmo água encanada'. A similaridade que percebemos entre esses sentidos plenos é subjetiva e pode ser descrita por meio da noção geral *ponto de origem*: no exemplo (1), tem-se ilustrado um ponto natural de origem de água, no (2), o foco é um ponto artificial de origem de água e, no (5), salienta-se a origem de energias renováveis. Ilustrando a relação de polissemia regular, vemos o nó 'origem, causa, procedência' que pode funcionar como conteúdo esquemático para uma série de leituras percebidas como micros sentidos por se encaixarem nos critérios de Croft & Cruse (2004) para a identificação de uma relação entre sentido especializado e sentido geral. Entre os critérios, conforme mostramos no quadro 5, destacamos o grau significativo de autonomia, a seleção de atenção e a possibilidade de unificação dos subsentidos como subespecificação de outro. Nos exemplos em análise, em (3), o foco é a origem da energia elétrica, em (4), a origem do conteúdo científico; em (6), a origem de informações sobre

atualidade. Sendo assim, o nó que representa o exemplo (5) estabelece com os nós (3), (4) e (6) uma relação de autosuperordenação.

Em negrito na figura estão os sentidos de *fonte* como ‘origem, causa, procedência’ e ‘texto de referência’. O primeiro é o que apresenta maior predominância na rede, portanto, o que tem maior saliência psicológica. O segundo é o que tem maior saliência estrutural, considerando os resultados da busca no Google.

Além dos sentidos compõem o *polyset*_{fonte}, entre os sentidos mais frequentes no corpúsculo associados à unidade fonológica *fonte*, está ‘conjunto de caracteres do mesmo tipo e tamanho’ (BORBA, 2002), em contextos como *preparamos uma coleção de fontes de computador para fazer convite de casamento*³⁰. Conforme demonstra a figura 20, no entanto, esse sentido ficou de fora do *polyset* por não conseguirmos construir nenhuma relação de similaridade com os demais.

6.1.2 Análise do item lexical *banco*

| A DEFINIÇÃO DOS NÓS | |
|---------------------|--|
| 1. Constante | Banco |
| 2. Exemplos | <p>1a. “Ao digitar sua senha do banco para confirmar a operação, o Itaú Shoptline se comunica com a operadora do cartão que autoriza o pagamento ao estabelecimento comercial onde a compra foi feita.”³¹</p> <p>1b. “Nos demais Estados do país a greve nos bancos privados será suspensa a partir desta quinta, segundo a Contraf”.³²</p> <p>1c. “Nos dias 25, 31 e 1º de janeiro, os bancos ficarão fechados.”³³</p> <p>1d. “Fácil comodidade para fazer consultas, efetuar pagamentos, transferências e outras transações sem precisar ir ao banco.”³⁴</p> <p>2a. “Voluntários fazem campanha por banco de órgãos do ABC.”³⁵</p> <p>2b. “O banco de sangue de São Paulo obedece normas nacionais e internacionais de segurança.”³⁶</p> |

³⁰ Fonte: <http://fonte.gratuita.com.br/>

³¹ Fonte: http://www.itaubr.com.br/conveniencia/serv_comprasnet_pgto.htm

³² Fonte: www1.folha.uol.com.br/folha/dinheiro/ult91u333681.shtml

³³ Fonte: http://g1.globo.com/Noticias/Economia_Negocios/0,,MUL931724-9356,00.html

³⁴ Fonte: <http://www.bb.com.br/portalbb/page22,136,3524,0,0,2,8.bb?codigoNoticia=3261>

³⁵ Fonte: <http://www.metodista.br/rroonline/cidades/voluntarios-querem-opo-no-abc>

³⁶ Fonte: <http://www.bssp.com.br/doacao.html>

| | |
|--|---|
| | <p>2c. “O banco de olhos não escolhe e nem tem preferência de qualquer espécie, pois a pessoa que irá receber os olhos entrará numa lista de espera seguindo uma ordem cronológica de inscrição.”³⁷</p> <p>3a. O objetivo principal de um sistema de banco de dados é possibilitar um ambiente que seja adequado e eficiente para uso na recuperação e armazenamento de informações.³⁸</p> <p>3b. O banco de imagens <i>on-line</i> que disponibiliza o trabalho de fotógrafos de todo o mundo, e que se destaca pelos preços que pratica para imagens de elevadíssima qualidade: entre um e cinco dólares por imagem.³⁹</p> <p>4. Jovem versátil sai do banco e garante vitória no fim do jogo.⁴⁰</p> <p>5. Sua ergonomia é ótima, pois o controle de altura do banco e a regulagem da direção proporcionam uma boa posição para dirigir.⁴¹ As doações diárias de sangue não completam as necessidades do banco.⁴²</p> <p>6. As andorinhas vêm das ilhas costeiras da costa da América do Norte e Europa e utilizam o banco como parada obrigatória todos os anos, entre setembro e março.⁴³</p> |
| 7. Chave de busca em inglês | <i>bank</i> |
| 8. Hiperônimos relevantes | 1a. instituição financeira
6. formação geológica |
| 9. Sinônimos relevantes | 1a. companhia bancária |
| 10. Glosas WN.Pr/acepções de dicionários (relevantes) | <p>1a-d. Estabelecimento de crédito, para transação de fundos públicos ou particulares (BORBA, 2002)</p> <p>2. Local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas (<i>banco</i> de sangue/ leite/ olhos/ esperma etc.) (AULETE, 2009)</p> <p>3. Depósito ou conjunto de coisas armazenadas para uso ou distribuição futuros, certos serviços ou operações, etc. (<i>banco</i> de dados, <i>banco</i> de imagens). (AULETE, 2009)</p> <p>4. (não consta em nenhum dos recursos léxicos analisados, mas é frequente no <i>cópus</i>)</p> <p>5. Móvel usado como assento com ou sem encosto, rústico, feito de madeira, pedra ou concreto. (BORBA, 2002)</p> <p>Qualquer assento improvisado. (BORBA, 2002)</p> <p>6. Elevação especialmente dentro do mar. (WN)</p> |

³⁷ Fonte: <http://boasaude.uol.com.br/lib/ShowDoc.cfm?LibDocID=3611&ReturnCatID=1776>

³⁸ Fonte: <ftp://ftp.unicamp.br/pub/apoio/treinamentos/bancodados/cursodb.pdf>.

³⁹ Fonte: www.sobresites.com/design/imagens.htm

⁴⁰ Fonte:

<http://www.sidneyrezende.com/noticia/49932+lateral+improvisado+sai+do+banco+e+garante+vitória+no+fim+do+jogo>

⁴¹ Fonte: <http://www.parana-online.com.br/canal/automoveis/news/35531/>

⁴² Fonte: <http://3planassiste.pgr.mpf.gov.br/informes/pr-banco-de-sangue-necessita-de-doadores>

⁴³ Fonte: <http://www.nectonsub.com.br/item/845/>

| | |
|--|--|
| <p>11. Nós do <i>polyset</i>_{café}</p> | <p>1. Instituição
 1a. Instituição financeira
 1b. Pessoas
 1c. Funcionamento
 1d. Sede física onde funciona a instituição financeira
 2. Local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas.
 2a <i>banco</i> de sangue
 2b <i>banco</i> de olhos
 3. Depósito ou conjunto de coisas armazenadas para uso ou distribuição futuros, certos serviços ou operações, etc.
 3.1 coleção de dados
 3.2 coleção de imagens
 4. Reserva, suplente
 5. Assento
 6. Aglomeração de areia, conchas, fósseis e detritos de rochas</p> |
| <p>12. Conjuntos de sentidos</p> | <p><i>Polyset</i>¹ { ['instituição financeira', 'pessoas', 'funcionamento', 'sede física onde funciona a instituição financeira']_{a-d}, ['local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas' (<i>banco</i> de sangue)_{2a} (<i>banco</i> de olhos)_{2b}]₂, ['depósito ou conjunto de coisas armazenadas para uso ou distribuição futuros, certos serviços ou operações, etc. (<i>banco</i> de dados, <i>banco</i> de imagens)'. ('coleção de dados')_{3a} ('coleção de imagens')_{3b}]₃, ('reserva, suplente')₄ }_{banco1}</p> <p><i>Polyset</i>² { ('assento')₅, ('aglomeração de areia, conchas, fósseis e detritos de rochas')₆ }_{banco2}</p> |

Quadro 12 – Resultados das etapas preliminares de construção do *polyset*_{banco}

6.1.2.1 ETAPAS DE CONSTRUÇÃO DO *POLYSET*_{banco}:

O item lexical *banco* é frequentemente citado na literatura como exemplo de homonímia, conforme mencionamos ao longo deste trabalho, por não poder ser percebida qualquer similaridade entre *banco* ‘instituição financeira’, sentido (1), e ‘assento’, sentido (8). Esses não são, no entanto, os únicos sentidos convencionalmente associados ao item, de maneira que diferentes agrupamentos podem ser percebidos em seu conteúdo semântico. Conforme indicamos no quadro 13, *banco* é associado a dois *polysets* independentes construídos a partir da análise dos sentidos referentes aos exemplos (1-6): o primeiro é formado pelos sentidos (1-4) e o segundo pelos sentidos (5 e 6).

sentido pleno (2), 'local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas', serve como conteúdo esquemático para as leituras codificadas como microssentidos '*banco* de sangue' e '*banco* de córnea', relação de autohiponímia/autosuperordenação.

Entre os sentidos exemplificados em (1a-d), respectivamente, 'instituição financeira', 'pessoas', 'funcionamento' e 'sede física onde funciona a instituição financeira', pode ser percebida uma relação de polissemia regular do tipo *automeronímia*. Trata-se da classe mencionada em Buitelaar (2000) e que retomamos no quadro de fechamento da Seção 4. 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 22 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.

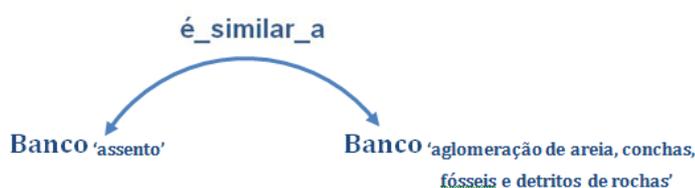


Figura 22 - *Polyset2* *banco*

A similaridade que construímos entre os sentidos (8), *banco* como 'assento' e (9), 'aglomeração de areia, conchas, fósseis e detritos', foi motivada especialmente pela combinação entre o exemplo (9) e a acepção geral apresentada em (8), em que *banco* é definido como 'qualquer assento improvisado' (BORBA, 2002). A partir da combinação dessas informações, entendemos que *banco* nos sentidos (8) e (9) associam-se em termos da 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. Não percebemos, no entanto, os sentidos exemplificados em (8) e (9) como microssentidos de um sentido mais geral em razão do fato de que o uso de um *banco* de areia, por exemplo, como apoio ou assento constrói-se apenas em situações especiais, como foi o que ocorreu em (9); fato que impossibilita preenchermos um dos

critérios principais da relação de *autohiponímia* que é a capacidade de unificação do tipo *é-um*. Há alto grau de autonomia e antagonismo entre esses sentidos, configurando-se como sentidos plenos relacionados por similaridade.

6.1.3 Café

| A DEFINIÇÃO DOS NÓS | |
|-------------------------------------|---|
| 1. Constante | café |
| 2. Exemplos | <p>1a. É toda essa composição que faz do café uma bebida natural e saudável.⁴⁴</p> <p>1b. Tomei uma água, um café. Não demorou muito. Dali eu saí e fui para a sala da ministra conduzida pela Erenice. Nós conversamos amenidades.⁴⁵</p> <p>2. Estas são as bases do cultivo orgânico do café que pode, entretanto, variar dependendo se o cultivo é mais ou menos intensivo.⁴⁶</p> <p>3. O café não é remédio, mas a comunidade médico-científica já considera a planta como funcional (previne doenças mantendo a saúde) ou mesmo nutracêutica (nutricional e farmacêutico).⁴⁷</p> <p>4. Durante a torra do café, esses ácidos clorogênicos formam novos compostos bioativos: os quinídeos.⁴⁸</p> <p>5. O uso da cor café na pintura provoca efeitos especiais e personaliza ambientes com estilo.⁴⁹</p> <p>6. Você é da turma que pula o café da manhã pensando em economizar calorias e acha que está fazendo um belíssimo negócio?⁵⁰</p> <p>7. Vamos ao café, fazer um lanchinho para conseguir ir até o final do expediente?⁵¹</p> |
| 3. Chaves de busca em inglês | (1-5) <i>coffee</i> (6) <i>bar</i> (7) <i>breakfast</i> |
| 4. Hiperônimos relevantes | 1a. alimento > bebida
alimento > bebida > porção |

⁴⁴ Fonte: http://www.abic.com.br/sabor_cafe.html

⁴⁵ Fonte: <http://www1.folha.uol.com.br/folha/brasil/ult96u612562.shtml>

⁴⁶ Fonte: <http://www.cnpab.embrapa.br/publicacoes/sistemasdeproducao/cafe/cafe.htm>

⁴⁷ Fonte: http://www.abic.com.br/sabor_cafe.html

⁴⁸ Fonte: http://www.abic.com.br/sabor_cafe.html

⁴⁹ Fonte: <http://www.revistacafeicultura.com.br/index.php?tipo=ler&mat=17104>

⁵⁰ Fonte: http://boaforma.abril.com.br/edicoes/189/fechado/Dieta/conteudo_130.shtml

⁵¹ Fonte: <http://twitter.com/ArantesGM/status/3181969279>

| | |
|---|---|
| | <ul style="list-style-type: none"> 2. planta > árvore 3. objeto natural > semente 4. substância 5. cor 6. refeição 7. local |
| 5. Sinônimos relevantes | 6. bar, boteco |
| 6. Glosas WN.Pr/acepções de dicionários relevantes | <ul style="list-style-type: none"> 1.a Bebida feita com a infusão de grãos de café (WN.Pr).
b Porção individual dessa bebida (HOUAISS, 2001). 2. Árvore cafeeiro (HOUAISS, 2001). 3. Fruto do cafeeiro (HOUAISS, 2001). 4. Produto resultante do beneficiamento, torrefação, resfriamento, descanso e moagem dos grãos de café (AULETE, 2009).
Pó de café. (BORBA, 2002). 5. A cor marrom do café ('grão') seco ou torrado, de matiz amarelado ou avermelhado e tonalidade entre o pardo ou um tom muito escuro, quase preto, a que ger. tb. corresponde a cor da bebida (HOUAISS, 2001). 6. Refeição rápida onde se toma a bebida (HOUAISS, 2001). 7. Estabelecimento comercial onde se vende essa bebida (BORBA, 2002). |
| 7. Nós do <i>polyset</i>_{café} | <ul style="list-style-type: none"> 1a. bebida b. porção da bebida 2. cafeeiro 3. fruto do cafeeiro 4. pó de café 5. cor 6. refeição rápida 7. bar |
| 8. Conjuntos de sentidos | {('bebida', 'porção da bebida', 'cafeeiro', 'fruto do cafeeiro', 'pó de café') ₁₋₄ , ('cor') ₅ , ('refeição rápida') ₆ , ('bar') ₇ } _{café} |

Quadro 13 – Resultado das etapas preliminares de construção do *polyset*_{café}

6.1.3.1 ETAPAS DE CONSTRUÇÃO DO *POLYSET*_{café}:

A rede de polissemia para os sentidos associados a *café* está representada na figura 23.

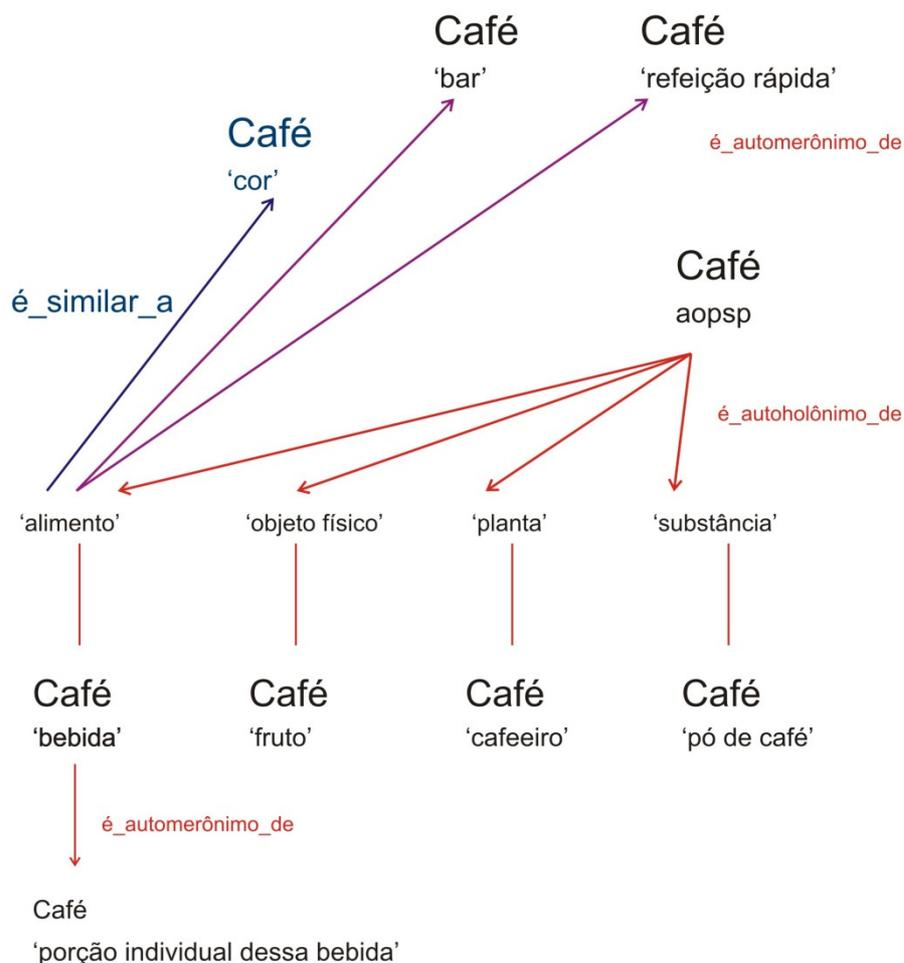


Figura 23 - Polysset *Café*

Conforme ilustra a figura 23, os sentidos associados a *café* podem ser agrupados em dois grandes conjuntos: polissemia irregular (em azul) e polissemia regular (em vermelho). A relação de polissemia irregular, expressa através do rótulo *é_similar_a*, caracteriza um conjunto de sentidos plenos relacionados por similaridade: entre *café* no sentido de 'cor', exemplo (5), e *café* no sentido de 'alimento', exemplo (1), percebemos um tipo de similaridade objetiva, baseada na percepção visual que associa a cor do café com a cor de objetos em geral. A relação de polissemia irregular agrupa diferentes conjuntos de facetas de sentido. O primeiro grupo que podemos identificar é aquele que relaciona uma classe regular de polissemia descrita em Buitelaar (1998), a classe que abreviamos, seguindo os moldes do autor, em "aopsp", para representar um sentido composto pelas facetas 'alimento', 'fruto', 'cafeeiro', 'pó_de_café'. Esse é o tipo clássico de polissemia regular rotulado como autoholonímia. A regularidade, nesse caso, justifica-se

por podermos perceber itens como *chocolate*, *erva-doce*, *anis*, *amendoim* que podem delimitar o mesmo tipo de relação. O segundo grupo de sentidos relacionados por polissemia regular é entre *café* ‘bebida’ (exemplo 1), ‘bar’ (exemplo 7) e ‘refeição rápida’ (6). Rotulamos a relação como automeronímia, visto que a bebida café é tipicamente parte do cenário de um bar, por ser vendida no bar, e é parte também típica do evento refeição, como é o café da manhã, ilustrado em (6). A distribuição das conexões entre os sentidos mostra que o sentido com maior saliência psicológica é *café* como ‘bebida’ e os resultados da busca no Google mostraram também esse como o sentido mais recorrente, portanto, com maior saliência estrutural.

6.2 A FACE COMPUTACIONAL DE CONSTRUÇÃO DO *POLYSET*

Nesta seção, o objetivo é equacionar o domínio computacional da tese, tarefa que compreende a apresentação de uma estratégia para a codificação do *polyset*. Para a realização de tal tarefa, utilizaremos o editor Protégé-OWL, que possibilita a criação da rede de polissemia, sua visualização e pesquisa de termos, além da codificação e exportação da base de conhecimento para arquivos em diferentes formatos (OLW, RDF, CLIPS, HTML, N-TRIPLE, N3, TURTLE). Para incluir no editor o conteúdo descrito em 6.1, adaptaremos, conforme as necessidades representacionais do fenômeno da polissemia, as recomendações da World Wide Web Consórcio (W3C) para a codificação da base WN.Pr em Ontology Web Language (OWL). Conforme mencionamos na introdução, o *status* que pretendemos dar à representação proposta no domínio computacional é o que Nirenburg e Raskin (2004, p. 14) denominam “tese fraca de IA”. Esse tipo de representação de conhecimento nega que um sistema computacional modele diretamente o conhecimento humano sobre semântica. Ao invés disso, sugere haver equivalência funcional, isto é, sugere que os programas computacionais podem alcançar resultados com qualidade análoga à humana, apesar de não usarem exatamente os mesmos métodos. Passaremos à explicitação das estratégias para a codificação de um *polyset* no editor Protégé-OWL⁵².

⁵² A mesma versão do Protégé-OWL foi testada e aprovada por em Di Felippo (2008) para o alinhamento de conceitos das *wordnets* WN.Pr e WN.Br, fato que demonstra a facilidade de integração das informações representadas no modelo *polyset* às demais informações em construção para a WN.Br.

6.2.1 A CODIFICAÇÃO DE UM *POLYSET* EM OWL ATRAVÉS DO PROTÉGÉ-OWL

Reconhecendo a importância das *wordnets* para sistemas de PLN, no contexto da W3C, Gangemi (2004)⁵³ propôs padrões para a codificação da WN.Pr no formato Ontology Web Language (OWL) com o objetivo de criar um esquema representacional comum aos diferentes projetos, facilitando o intercâmbio de dados. A opção pela linguagem OWL deve-se ao fato de ela ser a linguagem recomendada pela W3C para a representação de ontologias⁵⁴ em ambientes de Web Semântica (do Inglês: *Semantic Web*)⁵⁵. A OWL é recomendada por não ser uma simples linguagem de marcação de hipertexto com informações estruturais apenas (como é o HTML), mas por possuir semântica e sintaxe próprias para a modelação do conteúdo semântico de textos, contribuindo para a realização de inferências (ALVES, 2005). Tendo em vista as vantagens da codificação de conteúdo semântico em OWL, faremos uso do editor Protégé-OWL que, além de possibilitar a modelagem dos dados nessa linguagem, possibilita que eles sejam exportados para outros formatos, mostrando-se um recurso de grande utilidade (ex. CLIPS, HTML, N-TRIPLE, N3, TURTLE).

Seguindo as especificações propostas por Gangemi (2004), e sua equipe, para a codificação dos constructos da WN.Pr (ex.: formas lexicais, *synset*, sentidos e relações) em OWL, apresentaremos a seguir os critérios para representação dos *polysets* no Protégé-OWL, recurso programado para abarcar todos os constructos previstos pela linguagem. Seguiremos o mesmo esquema de codificação das informações semânticas para facilitar a integração do *polyset* à estrutura geral das *wordnets*. Seguindo a proposta da W3C (GANGEMI, 2004), os componentes que estruturam um *polysets* em OWL são representados como a figura 24⁵⁶.

⁵³ Disponível em <http://www.w3.org/2001/sw/BestPractices/WNET/tf.html>

⁵⁴ Uma definição clássica de ontologia é a de Gruber (1993, p.199), que afirma que “uma ontologia é uma especificação explícita de uma conceitualização (CHISHMAN, et. al). A definição do termo não é consensual na literatura, mas, para os fins desta tese, ela é suficiente.

⁵⁵ Padrão de representação de dados na web a fim de possibilitar que sistemas de PLN processem o conteúdo semântico de bases de dados lexicais criado pela W3C. O conteúdo semântico de documentos disponíveis na Internet estão codificados fundamentalmente em na linguagem padrão para a marcação de hipertexto, HTML (HyperText Markup Language) (ALVES, 2005).

⁵⁶ Figura adaptada da proposta de codificação da *wordnet* em OWL disponível em <http://www.w3.org/2001/sw/BestPractices/WNET/wordnet-sw-20040713.html>

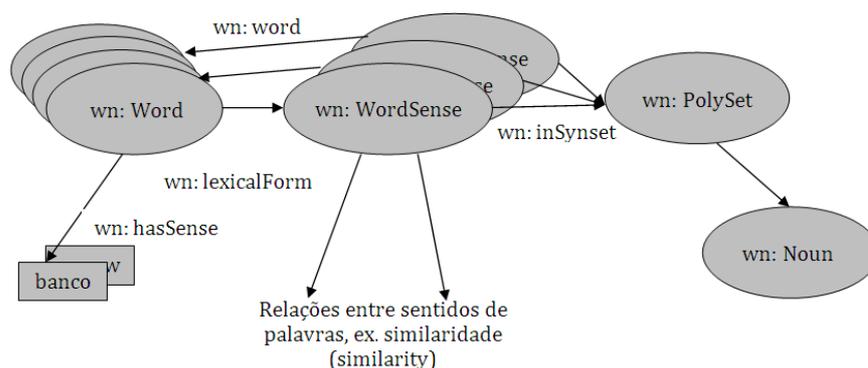


Figura 24 - Os componentes do *polyset* em OWL

Conforme retratamos na figura 24, há dois elementos que estruturam uma representação em OWL que são centrais ao modelo de *polyset*, são eles: *classes* e *propriedades*. As classes estão representadas nas esferas cinza e as propriedades nos rótulos das setas que ligam cada esfera. No esquema padrão do Protegé-OWL, as classes são responsáveis pela definição das entidades codificadas (ex.: *polyset*, *word*, *banco*). As classes genéricas do esquema de Gangemi (2004) para uma *wordnet* são (i) *Synsets*, que podem ser do tipo *synsets* para adjetivos, advérbios, nomes e verbos, (ii) *Word*, que pode ter como subclasse *collocation*, para incluir unidades complexas (na definição do autor, *string* composta por duas ou mais palavras conectadas por espaços ou hífen, por exemplo *man-eating*), (iii) *WordSense*, que pode se referir ao significado de adjetivos, advérbios, nomes ou verbos. Por recomendação de Knublauch et al. (2004), as classes são grafadas com iniciais maiúsculas e sem espaço entre as palavras. A figura 25 mostra esses elementos em estrutura hierárquica no Protégé-OWL.

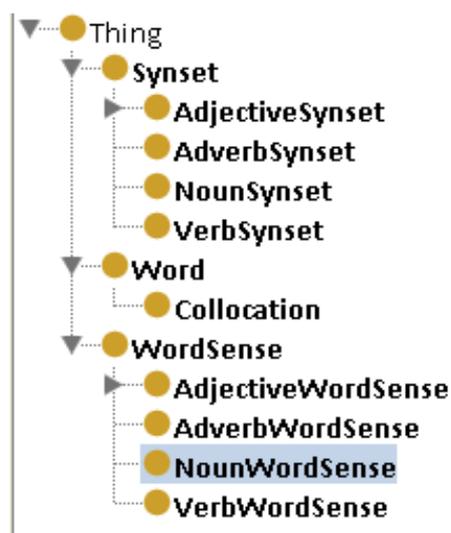


Figura 25 - Classes genéricas de uma *wordnet* em OWL

Adaptadas ao modelo *polyset*, essas classes genéricas são *polyset*, *Noun_Polyset*, *Word* e *Word_Sense*, conforme mostra a figura 26.

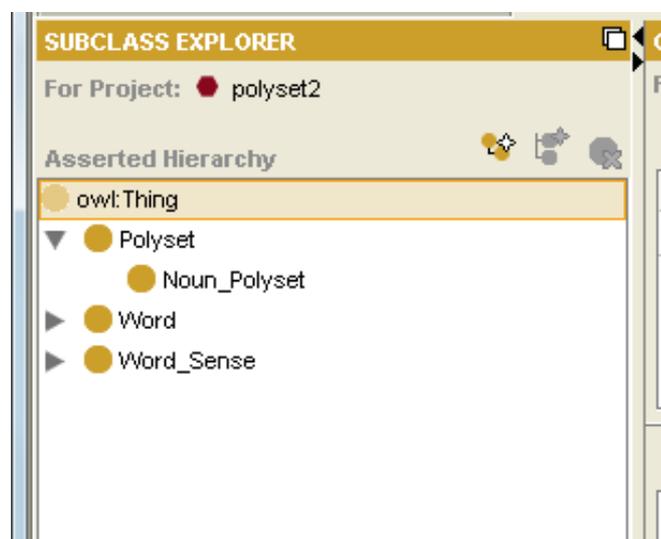


Figura 26 - Classes genéricas do modelo *polyset*

Dois tipos de propriedades codificadas pelo editor interessam ao nosso modelo: *object property*, para definir as relações de sentido, e *datatype property*, para definir outros tipos de características entre os sentidos, tais como *regular* e *noção geral*. Para as propriedades, a recomendação é inicial minúscula e as demais palavras com iniciais

maiúsculas, também sem espaço. A seguir, descreveremos como cada um dos componentes do *polyset* são codificados no Protégé-OWL, seguindo Gangemi (2004).

Os **nós das redes de polissemia** são os sentidos e estão representados do mesmo modo que os nós da rede *wordnet*: em termos de *classe* e estruturados hierarquicamente, conforme figura 27.

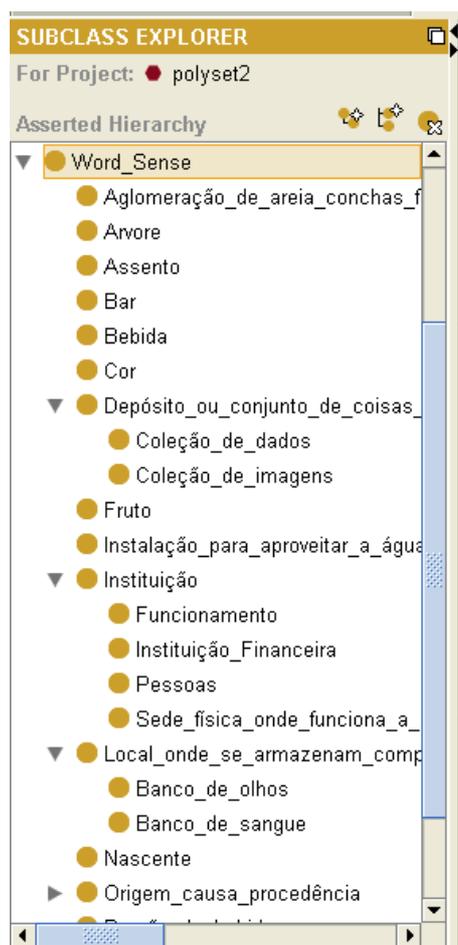


Figura 27 - Os nós como classes em estrutura hierárquica

Conforme apresentamos na Seção 5, os **nós** que compõem um *polyset* são os sentidos polissêmicos, e sua identificação pressupõe perceber sua associação convencional com uma mesma estrutura fonológica. Esses dois polos que integram um item lexical ficam evidentes no modelo de codificação proposto por Gangemi (2004), pois são inseridos como classes separadas. Essa estratégia faz com que não seja

necessário inserir mais de uma vez um mesmo item lexical com índices que os diferenciem e tornem possível a representação das relações que estruturam o modelo *polyset* do modo com descrevemos na seção teórica desta tese. Conforme indicamos na figura 28, o polo formal (Word) é representado pela classe *word* e recebe a propriedade, do tipo *datatype*, *lexicalForm*, e o polo semântico (Word Sense) é representado pela classe *WordSense*, e é especificado por diferentes propriedades do tipo *object property*. A conexão entre os polos se dá por meio da propriedade *hasSense* (*object property*), conforme a figura 28.

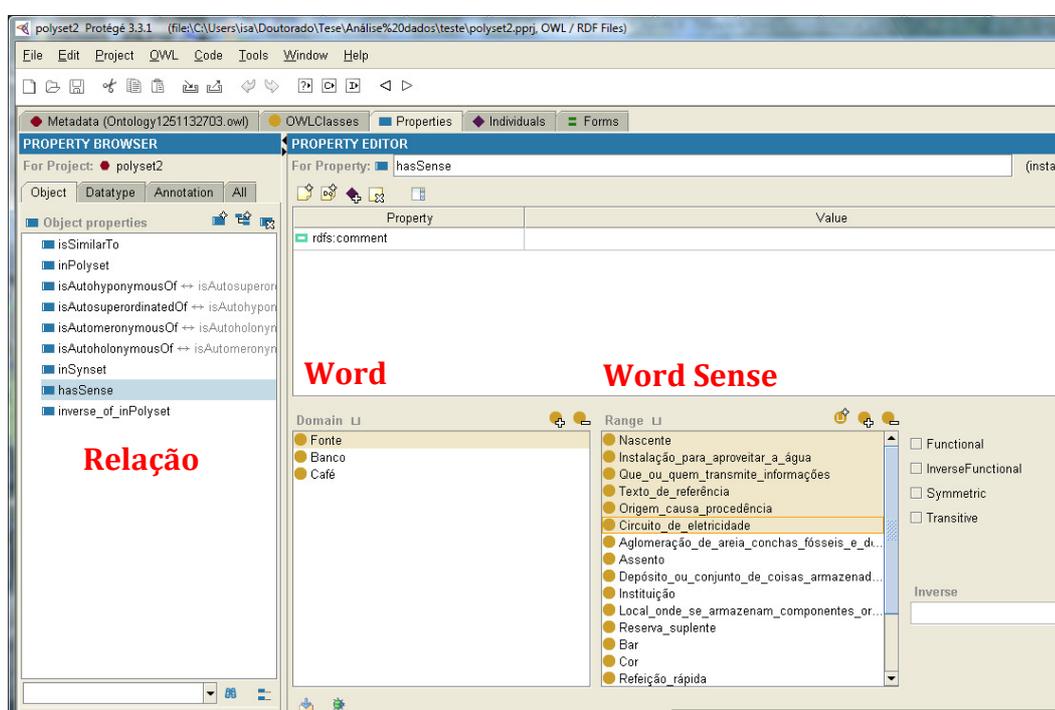


Figura 28 - A relação entre estrutura fonológica e semântica como uma *object property* do tipo *hasSense*

A codificação da relação entre uma mesma *word* e mais de uma *WordSense* significa que determinada *word* tem mais de um sentido. Na WN.Pr, conforme vimos na Seção 2, não se indica a existência ou não de conexões entre os diferentes sentidos associados a uma mesma *word*. O que as *wordnets* tipicamente codificam é quando *words* distintas associam-se a um mesmo *WordSense* (sinonímia), informação que, na proposta de Gangemi (2004) é explicitada por meio da relação *inSynset*. A proposta do modelo *polyset* é, justamente, codificar essas relações que têm ficado de fora das

wordnets. Desse modo, quando dois ou mais sentidos estiverem associados a uma mesma *word* receberão a propriedade (*object property*) ***inPolyset***, por analogia à relação *inSynset*. Um *polyset*, portanto, deve ser entendido, conforme descrito em 5, como um **conjunto de sentidos polissêmicos**. Essa informação, ilustrada na figura 29, é um importante acréscimo do modelo *polyset* ao paradigma *wordnet*. Na figura vemos que os sentidos de *fonte* como ‘nascente’, ‘instalação_para_aproveitar_água’, ‘origem_causa_procedência’ fazem parte de um *polyset* de nominais (Noun_Polyset).

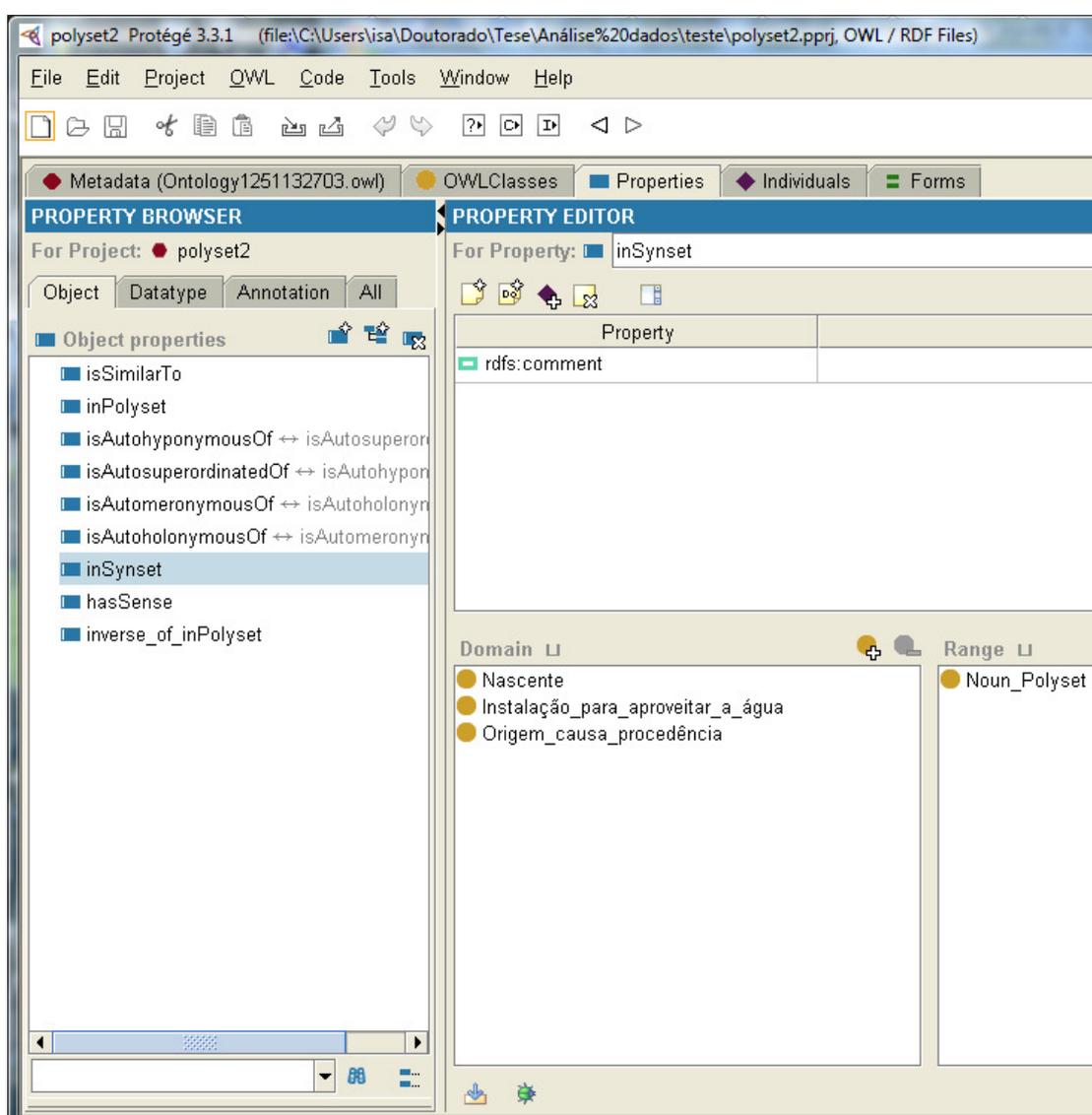


Figura 29 - Conjunto de sentidos polissêmicos como *object property* do tipo *inSynset*

Os **arcos da rede de polissemia** que estruturam um *polyset* são, conforme caracterizamos na Seção 5, as relações que estruturam a rede. Defendemos, nas seções anteriores, uma representação dos sentidos polissêmicos em diferentes níveis de generalidade e especificidade. É, conforme descrevemos, a partir dessa estruturação de sentidos que os *polysets* estão organizados. Vimos também, na Seção 2, que essa organização hierárquica entre os sentidos polissêmicos não tem sido, até onde temos notícias, codificada sistematicamente na WN.Pr, de modo que não há indicação no esquema representacional de Gangemi (2004). Fazendo uso da mesma lógica de Gangemi para a codificação das relações léxico-conceituais das *wordnets*, representaremos as relações entre sentidos polissêmicos por meio de propriedades do tipo *property object*, conforme a figura 30.

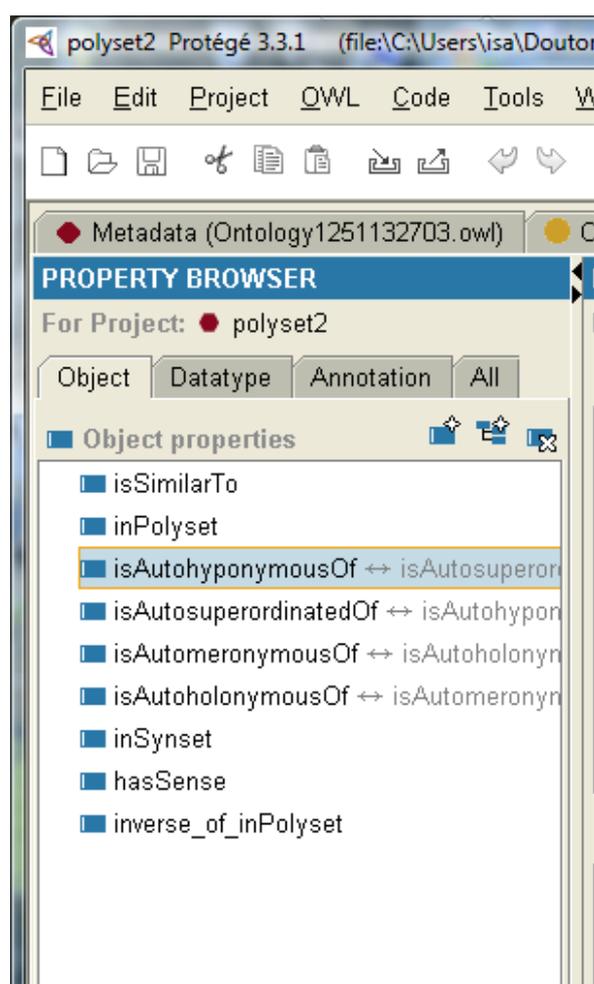


Figura 30 - Relações como *object properties*

Para codificar cada relação de sentido na aba de *object property*, é necessário definir um sentido de base (*domain*, no Protégé-OWL) e um alvo (*range*, no Protégé-OWL). O quadro 14 lista as diferentes relações codificadas no modelo *polyset* e indica o domínio e o alvo de cada relação

| Property (propriedade) | Domain (domínio) | Range (alvo) |
|-------------------------------|-------------------------|---------------------|
| OBJECT PROPERTY | | |
| <i>inPolyset</i> | WordSense | WordSense |
| <i>hasWordSense</i> | Polyset | WordSense |
| <i>Word</i> | WordSense | Word |
| <i>isSimilar</i> | WordSense | WordSense |
| <i>isAutohiponimousOf</i> | WordSense | WordSense |
| <i>isAutosuperordinatedOf</i> | WordSense | WordSense |
| <i>isAutomeronymousOf</i> | WordSense | WordSense |
| <i>isAutoholonymousOf</i> | WordSense | WordSense |
| DATATYPE PROPERTY | | |
| <i>globalNotion</i> | WordSense | |

Quadro 14 – Domínio e alvo nas relações de sentido

Para a estruturação dos sentidos plenos e dos subsentidos que compoem um *polyset*, informação que não está presente no padrão para *wordnets*, seguiremos a lógica de estruturação de informações do Protégé-OWL: hierarquização dos sentidos e indicação de *object properties*. A relação de autohiponímia (e seu inverso, a autosuperordenação) é expressa na figura 31.

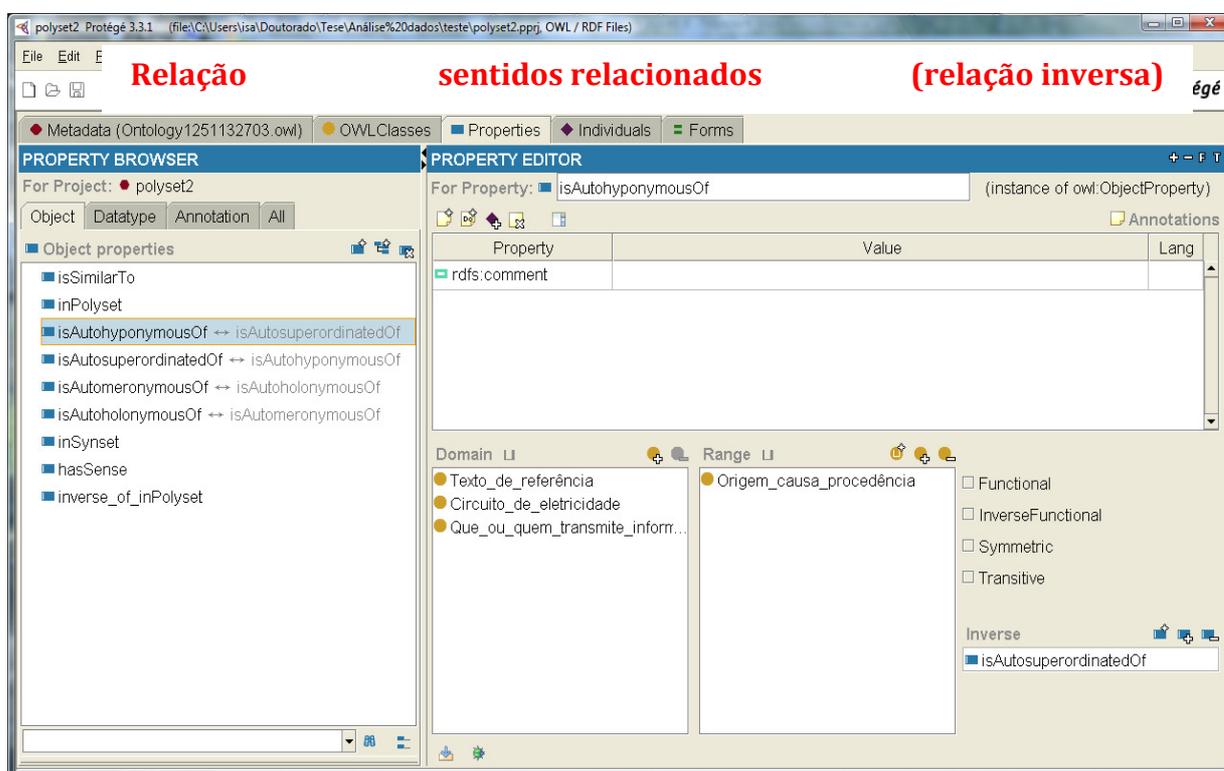


Figura 31 - Relação de autohiponímia/autosuperordenação

A relação de autoholonímia (e seu inverso, a automeronímia) é representada do mesmo modo, através da codificação de uma *object property*, conforme mostra a figura 32.

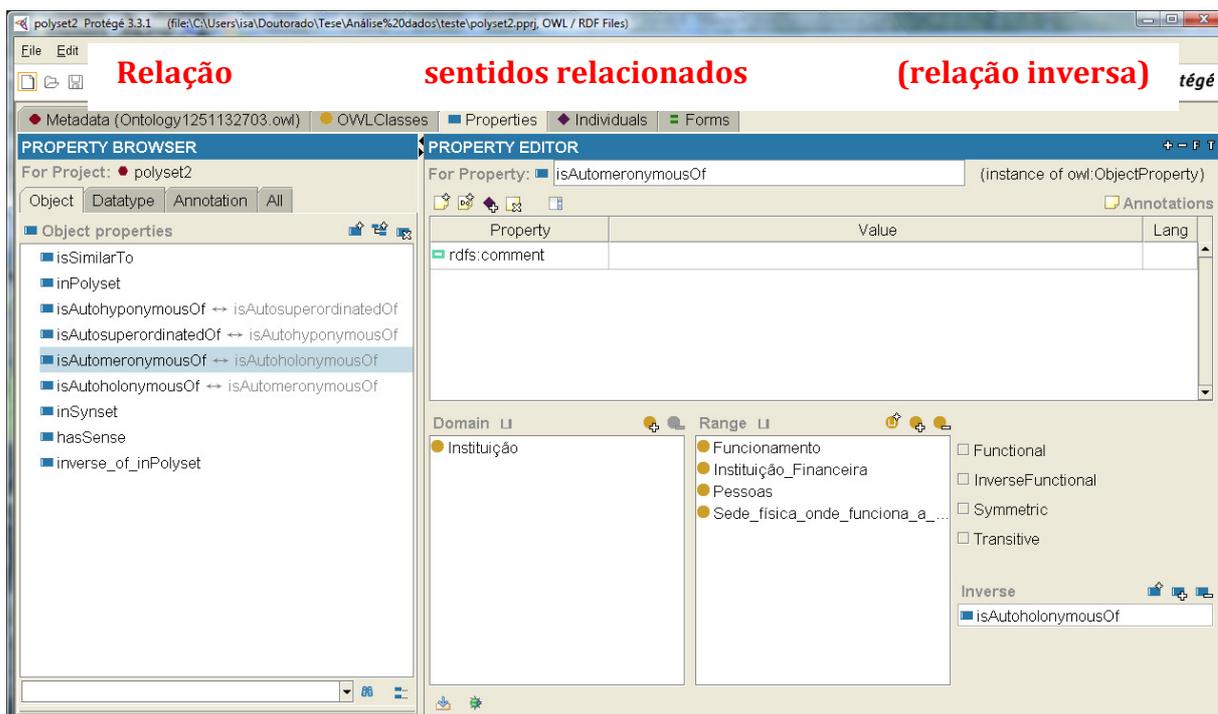


Figura 32 - Relação de automeronímia/autoholonímia

Para concluir a codificação de um *polyset*, é preciso ainda explicitar distintos graus de saliência entre os sentidos. Essa é mais uma informação que não está prevista no esquema de Gangemi (2004) por não constar na WN, configurando-se como um acréscimo do modelo *polyset*. Os efeitos de saliência entre os sentido são codificados no através de relações do tipo *datatype property*: para indicar o sentido com maior saliência psicológica, o rótulo da relação é *psycSalience*; para indicar o sentido com maior saliência estrutural ou de frequência, o rótulo da relação é *freqSalience* – conforme mostra a figura 33.

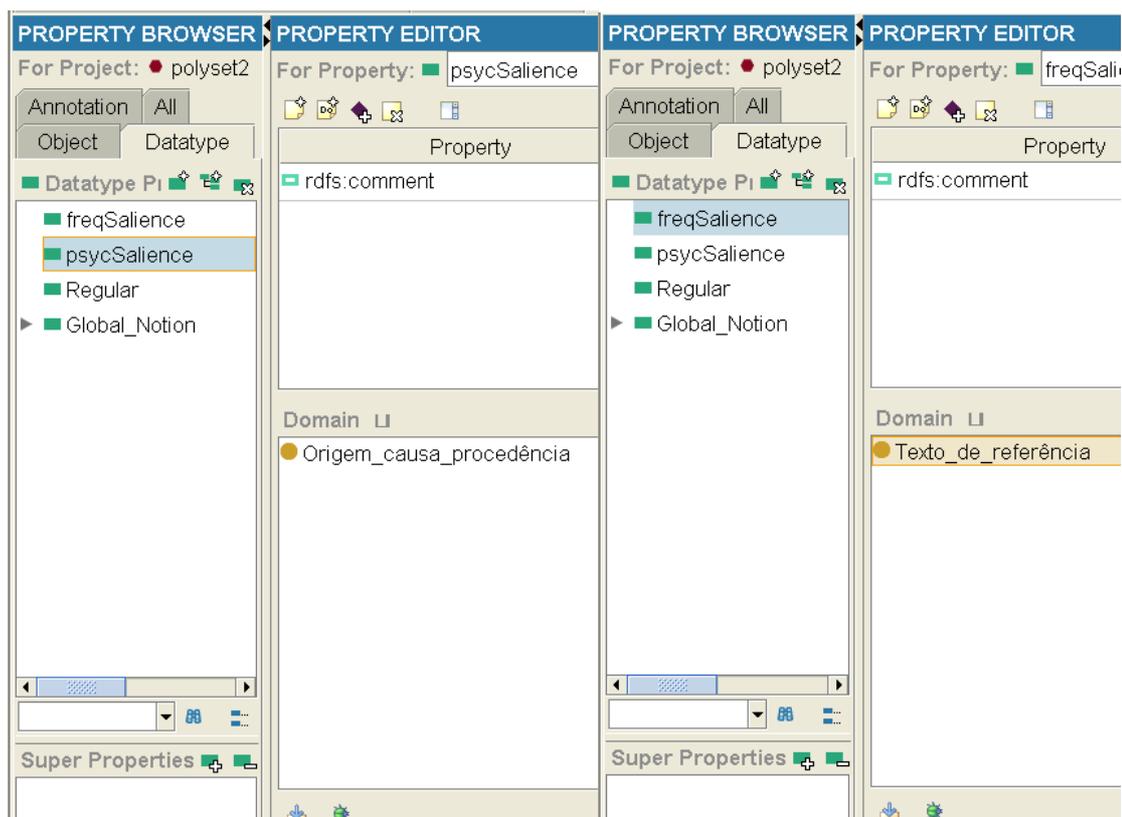


Figura 33 - Graus de saliência

Para concluir a seção, apresentamos, na figura 34, uma visualização completa do exercício realizado no formato de rede construído a partir de um *plug-in* que pode ser habilitado no Protégé-OWL, o OWL-Viz.

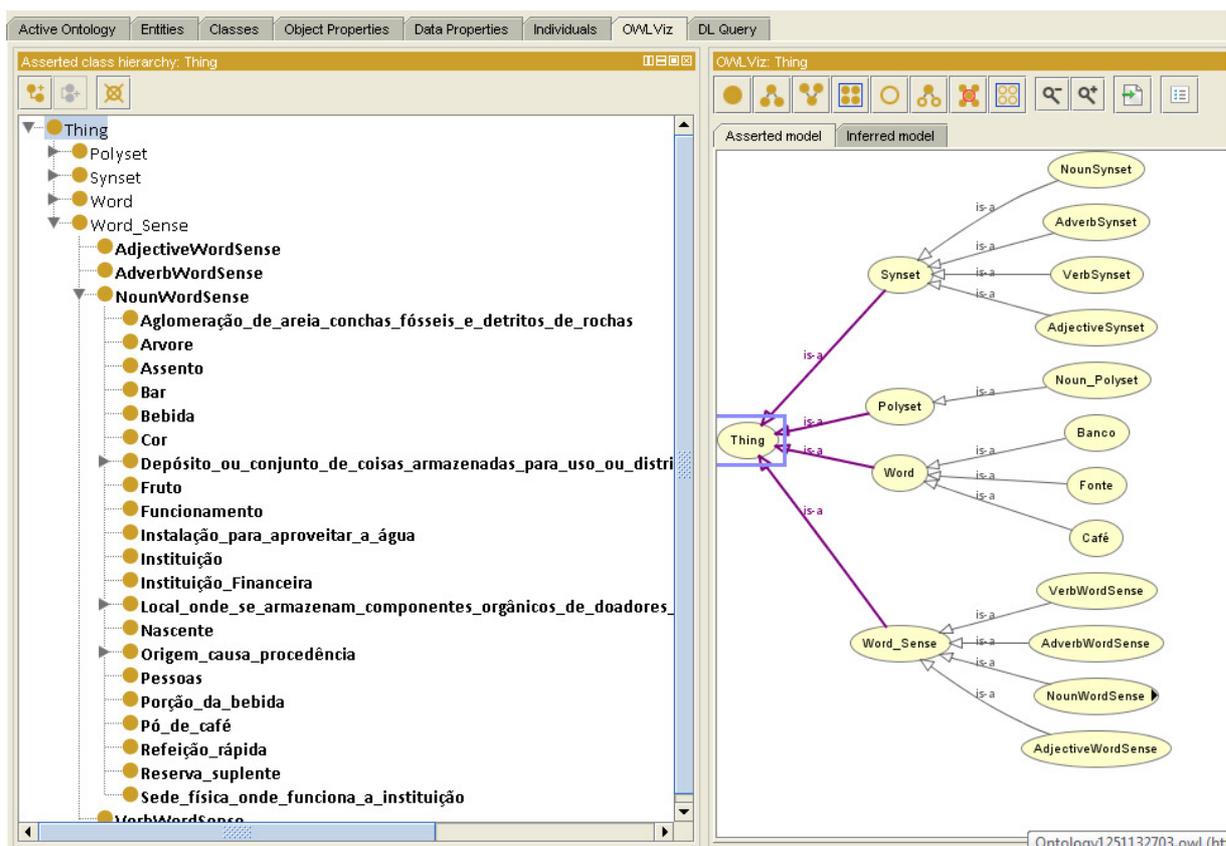


Figura 34 - Visão geral dos *polysets* codificados no Protégé-OWL

6.3 SÍNTESE DAS CONTRIBUIÇÕES DO MODELO *POLYSET* PARA A WORDNET.BR

A WordNet.Br (WN.Br) (DIAS-DA-SILVA, 2005; 2006), o foco de aplicação desta tese, está em desenvolvimento. Desde 2002, e, em sua situação atual, contém mais de 20 mil unidades lexicais analisadas, somando mais de 20 mil *synsets*, entre verbos, substantivos, advérbios e adjetivos. As relações identificadas até o momento para as formas verbais, adverbiais e adjetivais foram *sinonímia* e *antonímia*; além disso, foi feita a indexação de parte desses *synsets* com seus equivalentes na WN2.0. A metodologia de desenvolvimento adotada pelo grupo é por abrangência, e não por profundidade.

A inclusão dos nomes, foco desta tese, está em fase embrionária, e o tratamento da polissemia segue, por enquanto, a regra geral da WN.Br, apresentando apenas estatísticas sobre itens com sentido múltiplo. Trata-se de uma quantificação do número

de *synsets* de que um item lexical participa. Por exemplo, se uma mesma forma ocorre em três *synsets*, estipula-se que seu grau de polissemia é 3, independentemente do tipo de relação que se estabelece entre os sentidos. Nesse cenário, temos já uma base numerosa e com uma grande quantidade de substantivos polissêmicos. Conforme mencionamos na Seção 5, de um total de 16.925 entradas para nominais, 5.957 são codificados como polissêmicos. A organização desses conjuntos de sinônimos em termos de *polyset* terá, portanto, grande impacto na estruturação da WN.Br.

Destacamos duas como as contribuições centrais do modelo proposto para uma *wordnet* e, em especial, à WN.Br. A primeira delas é a proposição de uma metodologia para a representação estruturada para a polissemia de nominais, complementando as informações léxico-conceituais já disponibilizadas pela base. A segunda contribuição é a proposição de uma metodologia para a codificação computacional de tais informações, de maneira compatível com a proposta de codificação da WN.Br em OWL apresentada pela W3C. Para demonstrar visualmente essa compatibilidade, integramos a representação de *polysets* à base WN em OWL disponibilizada pela W3C⁵⁷ no Protégé-OWL, conforme a figura 35.

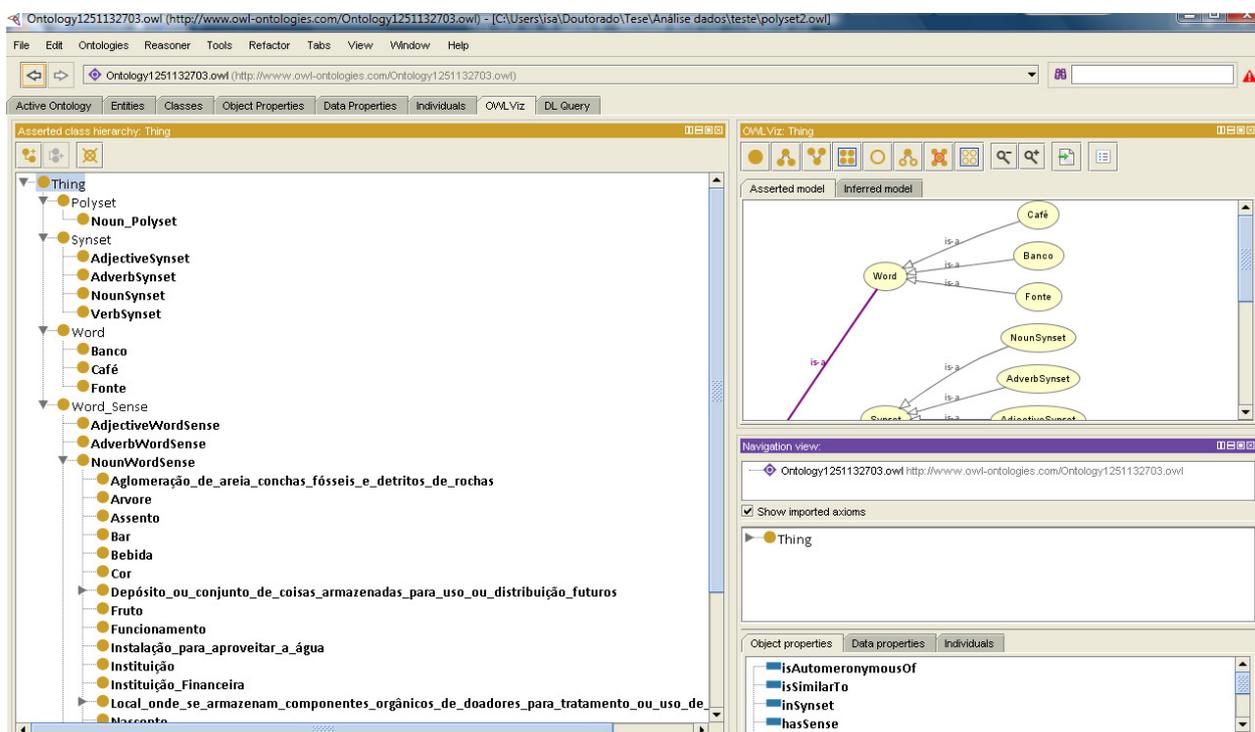


Figura 35 - wordnet e polysets em OLW no Protégé-OWL

⁵⁷ Disponível em <http://www.w3.org/2006/03/wn/wn20/download/>

Vimos na Seção 5 que há uma série de requisitos (cf. HELBIG, 2006) importantes para uma representação linguística que pretende ter *status* cognitivo e computacional. Para encerrar esta seção de apresentação das contribuições do modelo ao desenvolvimento da WN.Br, enfatizamos que incluir uma descrição de nominais polissêmicos nos moldes propostos fará com que a base esteja de acordo com os seguintes requisitos:

- (a) *Adequação cognitiva* – a representação da polissemia em termos de rede, conforme descrevemos na Seção 3 atribui o *status* cognitivo ao modelo;
- (b) *Interoperabilidade* – conforme ficou evidenciado na Seção 6, todas as propriedades previstas para o modelo *polyset* puderam ser representadas computacionalmente com auxílio do Protégé-OWL e do esquema de representação de Grangemi (2004);
- (c) *Homogeneidade e interpretabilidade local* – Estruturar os *polysets* em termos de rede de polissemia possibilitou representar as diferentes relações de sentido, conforme vimos em 6.1, e integrar essa representação à estrutura do léxico como um todo, conforme vimos em 6.3;
- (d) *Reutilização* – A adequação do modelo a esse requisito geral evidencia-se por dois motivos. O primeiro deles deve-se ao fato de as informações estarem codificadas em um editor que possibilita a exportação de sua base de conhecimentos para diferentes formatos. O segundo motivo deve-se ao fato de os *polysets* terem sido estruturados em OWL seguindo o padrão proposto pela W3C;
- (e) *Viabilidade e automaticidade* – Os *polysets* têm potencial para tratamento técnico e implementacional, conforme vimos evidenciado na Seção 6;
- (f) *Completude e consistência* - a possibilidade de representar polissemia regular e irregular a partir de um mesmo constructo, sem que sejam derivadas informações inadequadas;
- (g) *Granularidade ótima* - a possibilidade de representar sentidos polissêmicos em diferentes níveis de generalidade e estabilidade;
- (h) *Multidimensionalidade* – o constructo *polyset*, em sua codificação no Protégé-OWL está apto a representar informações provenientes de diferentes níveis de generalidade e de diferentes tipos, conforme vimos nos tipos de

propriedades que podem ser evidenciada a partir da combinação das diferentes propriedades codificadas (*object property* e *datatype property*).

6.4 SÍNTESE DA SEÇÃO

Para finalizar a Seção 6, com o objetivo de sintetizar as etapas de construção de um *polyset*, elaboramos o que chamaremos de regras gerais para construção de um *polyset*. Para sua aplicação, é necessário estar de posse das informações elencadas na Seção 5, pois aqui apresentamos apenas um guia geral que registra os tipos de relação de polissemia inventariados nesta tese. Tais regras foram construídas procurando seguir estilo similar ao algoritmo apresentado em Buitelaar (2000), na tentativa de evidenciar possíveis momentos para automatização dos processos de construção de um *polyset*. Tendo em vista esse objetivo de automatização, indicaremos somente uma *wordnet* como referência para os sentidos. Para a aplicação de tais critérios ao PB, não se pode esquecer que, enquanto a WordNet.Br estiver em processo de construção, é necessário seguir as etapas preliminares sintetizadas no quadro 9, realizando estudo de dicionários, *wordnets* e cópulas. As regras gerais são as seguintes:

Regra geral A – Um item lexical é considerado com sentido múltiplo sempre que participar de mais de um *synset*.

Regra geral B* - Por *default*, cada novo *synset* (sentido) deve ser listado na base de dados lexicais de maneira independente.

* Esse valor *default* é anulado sempre que possível aplicar a regra C.

Regra geral C - Sempre que alguma relação de polissemia for identificada, o padrão *default* é quebrado e uma das seguintes regras é aplicada:

Regra 1 – Se forem identificados sentidos plenos ligados por alguma *noção geral unificadora*, indicar a relação *é_similar_a* para rotular a relação entre eles.

Regra 2 – Se a relação de polissemia for *regular*, aplicar um dos seguintes rótulos:

2.1 Se forem identificadas as relações do tipo *é_um* entre os sentidos (microsentidos), aplicar o rótulo *é_autohipônimo_de* (a relação reversa é *é_autosuperordenado_de*).

Regra 3 - Se forem identificadas as relações do tipo *é_parte_de* entre os sentidos (facetas), aplicar o rótulo *é_automerônimo_de* (a relação reversa é *é_autosuperordenado_de*).

SEÇÃO 7 - CONCLUSÃO

O objetivo geral desta tese foi propor uma estratégia de representação linguístico-computacional da semântica de nominais polissêmicos que complemente as informações léxico-conceituais codificadas em *wordnets*, com vistas a ampliar a utilidade das bases para sistemas de PLN. O estudo foi desenvolvido seguindo uma metodologia de pesquisa que integra três domínios mutuamente complementares: o linguístico-computacional, o linguístico e o computacional, conforme propõe Dias-da-Silva (1996, 1998, 2003, 2006). O domínio linguístico-computacional ocupou a posição de “pivô” entre os dois outros domínios. Foi ele o responsável por fornecer o tema para esta pesquisa e por articular a relação entre os domínios linguístico e computacional. Respeitando essa estratégia de pesquisa, o conteúdo desta tese foi distribuído em cinco seções, cujos resultados serão sintetizados a seguir.

A Seção 2 ficou responsável pelo equacionamento da primeira etapa desenvolvida no domínio linguístico-computacional: a investigação sobre a relação polissemia, PLN e *wordnets*. A partir das reflexões empreendidas nessa seção, foi possível perceber que a falta de representação explícita da polissemia em bases do tipo *wordnet* tem sido apontada como uma das principais limitações para a maior utilização do recurso em sistemas de PLN. Vimos que diversas técnicas automáticas têm sido propostas para extrair similaridades entre os sentidos, e as técnicas que fazem uso das *wordnets* têm sido as que apresentam melhores resultados. Por não haver nesses recursos informações para as relações de polissemia, relações do tipo *sinonímia* e *hiponímia* têm sido exploradas para fins de desambiguação. Ao estudarmos o potencial descritivo do paradigma *wordnet*, ficou evidente que, em sua base, a polissemia não está prevista; isso porque, o significado lexical é representado sob a ótica onomasiológica, tendo a *sinonímia* como a relação estruturadora fundamental. A investigação sobre a WN.Pr, a WN.Pt e a GN comprovou essa dificuldade e demonstrou que, embora os projetos reconheçam a importância de representar o fenômeno, não apresentam uma estratégia adequada à descrição da polissemia como um todo. Vimos que há a preocupação com o fenômeno, porém, dada as dificuldades de sua delimitação por exigir um minucioso estudo linguístico, ele tem sido deixado em segundo plano. Concluímos a

Seção 2 cientes da importância da codificação da polissemia e informados sobre a situação atual de desenvolvimento das diferentes *wordnets*, portanto, informados sobre as possibilidades e as necessidades das *wordnets*.

As Seções 3 e 4 tiveram como foco o desenvolvimento da fase teórica do domínio linguístico. Na Seção 3, caracterizamos a polissemia sob o viés da SLC, tornando possível lançarmos mão de um arcabouço de constructos adequados à descrição das principais questões envolvidas na representação da polissemia. Olhar para o fenômeno do significado múltiplo sob o viés cognitivo, possibilitou descrever os sentidos como uma entidade complexa, estruturada em termos de redes. As redes de polissemia sincrônicas, em sua configuração livre e multidimensional, conforme propõem Blank (2003) e Geeraerts (2006), demonstraram ser a estratégia descritiva mais adequada à representação da flexibilidade do sentido para os propósitos desta tese. Nessa configuração de rede, há espaço para, em uma mesma estrutura, representar sentidos mais específicos e instáveis - particulares a determinada situação de comunicação -, e sentidos mais gerais, convencionais e estáveis - comuns a diversas situações de comunicação. Essa flexibilidade é especialmente interessante para a construção de uma base de dados lexicais que pretende ser útil a sistemas computacionais diversos como é uma *wordnet*.

A Seção 4 foi importante para compreendermos os processos envolvidos na organização interna de uma categoria de sentidos polissêmicos, investigando questões como delimitação de sentidos, graus de saliência e relações de sentido. No contexto linguístico-cognitivo, as relações de polissemia representam rotinas cognitivas que o falante realiza para delimitar sentidos compreendendo a língua em suas diversas situações. De acordo com o paradigma cognitivo da semântica, entende-se que essas rotinas cognitivas realizam-se baseadas em habilidades de percepção e em padrões de convenção. No contexto linguístico-computacional, a representação de diferentes relações entre os sentidos possibilita que o sistema perceba diferentes graus de similaridade, de saliência e de estabilidade entre os sentidos.

A Seção 5 foi dedicada à apresentação do delineamento linguístico do modelo de representação proposto: os *polysets*. Defendemos que a construção de um *polyset* seja realizada em dois momentos principais. O primeiro momento, denominado *etapas preliminares*, é dedicado à definição dos *nós* da rede de polissemia e à indicação dos agrupamentos entre os sentidos. Entre as atividades realizadas nessa etapa estão a

consulta ao *cópus*, a consulta aos recursos léxicos e a indicação de um conteúdo para cada um dos *nós* do *polyset*. O segundo momento, denominado *etapas de construção do polyset*, é dedicado à definição propriamente dita da rede de polissemia. É nessas fases que são rotulados os *arcos* da rede de polissemia, identificado o *status* de delimitação dos sentidos e seus efeitos de saliência. Do ponto de vista linguístico, o *polyset* é considerado construído quando se tem estruturada uma rede manual que representa as múltiplas relações entre os sentidos polissêmicos associados a um item lexical. Do ponto de vista linguístico-computacional e computacional, a construção do *polyset* se dá quando é finalizado o processo de inclusão dos dados na ferramenta de edição, etapa realizada na Seção 6.

Reservamos para a seção final da tese, a Seção 6, o espaço para a construção e implementação do *polyset*, respondendo à etapa aplicada dos domínios linguístico, linguístico-computacional e computacional desta pesquisa. Do ponto de vista linguístico, destacamos que os exemplos analisados revelaram a riqueza do fenômeno da polissemia. O item lexical *fonte* é tipicamente tratado na literatura como um exemplo de polissemia irregular, *banco* é exemplo típico de homonímia e *café* é exemplo de polissemia regular. Conforme ficou evidente nas redes de polissemia construídas para cada um dos itens lexicais, o conjunto de sentidos associados a um item é, de fato, múltiplo e precisa ser organizado em termos de uma estrutura multidimensional. Somente com esse tipo de configuração de rede é possível representar as diferentes relações, os diferentes níveis de saliência, de especificidade e de sensibilidade ao contexto situacional que são percebidos no espaço semântico acessível a uma unidade do léxico. Do ponto de vista computacional, destacamos a importância do esquema de codificação que seguimos para a inclusão do *polyset* no Protégé-OWL. Seguir um modelo padrão facilita a reutilização da representação e a interação entre grupos de pesquisa. A possibilidade e facilidade de interação entre representações que seguem os mesmos princípios ficou evidente ao apresentarmos a tela do Protégé-OWL que integra os *polysets* à estrutura de uma *wordnet* que contém informações clássicas.

Conforme mencionamos ao longo do trabalho, não há consenso sobre qual é a melhor abordagem para a análise e a representação da polissemia. A definição se dá em termos de sua adequação aos objetivos da pesquisa. Sem perder de vista esse fato, entendemos que o modelo adotado aqui é suficiente para os nossos propósitos por representar o significado lexical de um modo flexível, podendo expressar as diferentes

conexões entre eles, bem como diferentes níveis de especificidade e de generalidade em uma estrutura hierárquica.

Como trabalhos futuros, prevemos a expansão da aplicação do modelo a outras classes de palavras. Considerando as reflexões teóricas apresentadas, ficou evidente que o modelo de *polysets* foi proposto para substantivos, entretanto, sua estrutura de base, as redes de polissemia, não é restrita a eles, de modo que a abrangência do estudo pode ser expandida em trabalhos futuros. Uma outra expansão prevista para este trabalho, para facilitar a replicação da análise para um número maior de dados, é a automatização de algumas das etapas, por exemplo, a busca nos dicionários e no córpus.

REFERÊNCIAS BIBLIOGRÁFICAS

ALANI, H. TGVizTab: An Ontology Visualisation Extension for Protégé. In: **Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering**, October 26, Sanibel Island, Florida, USA, 2003.

ALLWOOD, J. Meaning potentials and context: Some consequences for the analysis of variation in meaning. In: CUYCKENS, H.; DIRVEN, R.; TAYLOR, J. (Ed.). **Cognitive approaches to lexical semantics**. Berlin: Mouton de Gruyter, p. 29-66, 2003.

ALVES, I. M. da R. **O uso da Semântica Verbal em Sistemas de Extração de Informação: a Construção de uma Ontologia de Domínio Jurídico**. Dissertação de Mestrado em Lingüística e Língua Portuguesa – Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, 2005.

ALVES, I. M. da R. **Os Tradutores Automáticos e a Ambigüidade: Uma Abordagem Crítica**. Trabalho de Conclusão de Curso de Graduação em Letras (Português-Inglês) – Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, 2003.

APRESJAN, J. **Regular Polysemy**. *Linguistics*, p. 142, 5-32, 1973.

AULETE, C. **Dicionário contemporâneo da língua portuguesa**. Edição brasileira original: Hamílcar de Garcia. Versão Eletrônica. Atualizada em 2009.

BLANK, A. 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, 2003. p. 267-293.

BORBA, F. S. (Coord.). **Dicionário de usos do português do Brasil**. São Paulo: Ed. da UNESP, 2002.

BOYD-GRABER, J.; FELLBAUM, C.; OSHERSON, D.; SCHAPIRE, R. Adding dense, weighted connections to WordNet. In: **Proceedings of the 3rd International WordNet Conference**, Jeju Island, Korea. 2006.

BRÉAL, M. **Ensaio de Semântica: Ciência das Significações**. São Paulo: EDUC, 1992.

BRISARD, F. The English tense-system as an epistemic category: the case futurity. In: VERSPOOR, M.; LEE, K. D.; SWEETSER, E. (Eds.). **Lexical and syntactical constructions and the construction of meaning**. Amsterdam: John Benjamins, p. 271-286, 1997.

BROCCIAS, C. Approachs to grammar. In: KRISTIANSEN, G.; ACHARD, M.; DIRVEN, R.; IBÁÑEZ, F. J. R. M. (Eds.) **Cognitive Linguistics: Current Application and Future Perspectives**. Mouton de Gruyter: Berlin, p. 81-118, 2006.

BROWN, S. W. Polysemy in the mental lexicon. **Colorado Research in Linguistics**, Boulder, v. 21, p. 4, June 2008. Disponível em: <http://www.colorado.edu/ling/CRIL/Volume21_Issue1/paper_BROWN.pdf>. Acesso em: 15 mar. 2009.

BRUGMAN, C. **Story of over**. M.A. thesis--University of California, Berkeley, 1981.

BRUGMAN, C.; LAKOFF, G. Cognitive typology and lexical networks. In.: GEERAERTS, D. (Ed.). **Cognitive Linguistics: Basic Readings**, Berlin: Mouton de Gruyter, p. 109-139, 2006.

BUITELAAR, P. **Corelex: systematic polysemy and underspecification**. PhD Thesis--Department of Computer Science, Brandeis University, Waltham, 137p., 1998.

BUITELAAR, P. Reducing lexical semantic complexity with systematic polysemous classes and underspecification. In: **Proceedings of the ANLP2000 Workshop on Syntactic and Semantic Complexity in Natural Language Processing Systems**, Seattle, USA, p. 14-19, 2000.

CANÇADO, M. **Manual de Semântica – Noções Básicas e Exercícios**. Belo Horizonte: Editora UFMG, p. 183. 2005.

CHISHMAN, R. L. de O. **A teoria do léxico gerativo: uma abordagem crítica**. 2000. 185 f. Tese (Doutorado em Lingüística e Letras)--Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2000.

CHISHMAN, R. L. de O; BERTOLDI, A.; ALVES, I. M. da R. **Integrando léxicos semânticos e ontologias: uma aproximação a favor da web semântica** (*no prelo*)

COPESTAKE, A.; BRISCOE, T. Semi-productive polysemy and sense extension. **Journal of Semantics**, Nijmegen, v. 12, n. 1, p. 15-67, 1995. Disponível em: <<http://www.cl.cam.ac.uk/~aac10/papers/jsem.ps.gz>>. Acesso em: 20 jan. 2006.

CROFT, W.; CRUSE, D. A. **Cognitive linguistics**. Cambridge: Cambridge University Press, 2004.

CRUSE, D. A. **Lexical semantics**. Cambridge: Cambridge University Press, 1986.

CRUSE, D.A. Aspects of the micro-structure of word meanings. In: RAVIN, Y.; LEACOCK, C. (Ed.). **Polysemy: theoretical and computational approaches**. Oxford: Oxford University Press, p. 30-51, 2000.

CRUSE, D.A. **Meaning in language: an introduction to semantics and pragmatics**. Cambridge: Cambridge University Press, 2000.

CRUSE, D.A. Polysemy and related phenomena from a cognitive linguistic viewpoint. In: SAINT-DIZIER, P.; VIEGAS, E. (Eds.). **Computational lexical semantics**. Cambridge: Cambridge University Press, p. 33-49, 1995.

CUYCKENS, H.; ZAWADA, B. (Ed.). **Polysemy in cognitive linguistics: selected papers from the Fifth International Cognitive Linguistics Association**, Amsterdam, 1997. Amsterdam: Benjamins, 2001.

DEANE, P. D. Polysemy and cognition. *Língua*, 75, p. 325-361, 1988.

Di FELIPPO, A. **Delimitação e alinhamento de conceitos lexicalizados no inglês norte-americano e no português brasileiro**. Tese de Doutorado em Linguística e Língua Portuguesa. Faculdade de Ciências e Letras (FCL), Universidade Estadual Paulista (UNESP). Araraquara, São Paulo, 253p., 2008.

DIAS-DA-SILVA, B. C. A construção da base da Wordnet.Br: conquistas e desafios. In. **III WORKSHOP em Tecnologia da Informação e da Linguagem Humana. XXV Congresso da Sociedade Brasileira de Computação**. São Leopoldo: Universidade Federal do Rio Grande do Sul, 2005.

DIAS-DA-SILVA, B. C. **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, Universidade Estadual Paulista Júlio de Mesquita Filho, Araraquara, 1996.

DIAS-DA-SILVA, B. C. A rede WordNet e a compilação de um thesaurus eletrônico. **Fórum Lingüístico**, Florianópolis, v. 3, n. 2, p. 157-176, 2003.

DIAS-DA-SILVA, B. C. Bridging the gap between linguistic theory and natural language processing. In: INTERNATIONAL CONGRESS OF LINGUISTICS, 16., 1997, Paris. **Proceedings...** Oxford: Elsevier Sciences, 1998. p. 1-10.

DIAS-DA-SILVA, B. C. Human language technology research and the development of the brazilian portuguese WordNet. In: HAJCOVÁ, E.; KOTESOVCOVÁ, A.; MÍROVSKY, J. (Eds.): **Proceedings of the 17th International Congress of Linguists**. Prague: Matfyzpress, MFF UK, 12p., 2003.

DIAS-DA-SILVA, B. C. O estudo lingüístico-computacional da linguagem. **Letras de Hoje**, Porto Alegre, v. 41, p.103-138, 2006.

EVANS, V.; GREEN M. **Cognitive Linguistics: An Introduction**. Mahwah, NJ: Lawrence Erlbaum Associates. 2006.

EVENS, M. W. **Relational Models of the Lexicon**. Cambridge: Cambridge University Press, 1988.

FAUCONNIER, G.; TURNER, M. Blending as a central process of grammar. In: GOLDBERG, A. (Ed.). **Conceptual structure, discourse and language**. Standford: CSLI Publications, p.113-130. 1996.

FAUCONNIER, G.; TURNER, M. Polysemy and conceptual blending. In: NERLICK, B.; TODD, Z.; HERMAN, V.; CLARKE, D.D. (Eds.). **Polysemy and Flexible Patterns of Meaning in Mind and Language**, p. 79-94, 2003.

FELLBAUM, C.; MILLER, G.A. Whither WordNet. Power Point Presentation, 2007.

FERREIRA, A. B. de H. **Dicionário Aurélio eletrônico século XXI (Versão 3.0)**. São Paulo: LexiKon Informática, 1999. 1 CD-ROM.

FILLMORE, C. J.; JOHNSON, C. R; PETRUCK, M. Background to FrameNet. **International Journal of Lexicography**. v.16, n.3, p.235-250, 2003.

GEERAERTS, D. (Ed.). **Cognitive linguistics: basic readings**. Berlin: Mouton de Gruyter, 2006.

GEERAERTS, D. **Words and other wonders: papers on lexical and semantic topics**. Berlin: Mouton de Gruyter, 2006.

GOLDBERG, A. **Constructions: a construction grammar approach to argument structure**. Chicago: The University of Chicago Press, 1995.

GUARINO, N. Understanding, Building and Using Ontologies. **KAW96 Knowledge Acquisition Workshop**. Banff, Alberta, Canada, 1997.

HAMP, B.; FELDWEG, H. GermaNet: a lexical-semantic net for German. In: **PROCEEDINGS of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications**. Madrid, 1997.

HANDKE, J. **The structure of the lexicon: human versus machine**. Berlin: Mouton de Gruyter, 1995.

HIRSCH, E. **Dividing reality**. New York: Oxford University Press, 1997.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa**. Rio de Janeiro: FL Gama Design, 2001. 1 CD-ROM.

HOUAISS, A.; VILLAR, M. de S. **Dicionário eletrônico Houaiss da língua portuguesa (versão 1.0)**. Rio de Janeiro: Objetiva, 2001. 1 CD-ROM.

JANSSEN, T. Monosemy versus polysemy. In: CUYCKENS, H.; DIRVEN, R.; TAYLOR, J. (Ed.). **Cognitive approaches to lexical semantics**. Berlin: Mouton de Gruyter, p. 93-122, 2003.

KATZ, J., FODOR, J. A. The structure of a semantic theory. **Language**, 39, p. 170-210, 1963.

KAY, P.; FILLMORE, C. Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. **Language**, 75(1), p. 1-33, 1999.

KILGARRIFF, A. **Polisemy**. PhD thesis--University of Sussex, Brighton, UK, 1992. Disponível em: <<http://www.kilgarriff.co.uk/Publications/1992-K-thesis.pdf>>. Acesso em: 15 mar. 2009.

KLEIN, D. E.; MURPHY, G. L. Paper has been my ruin: Conceptual relations of polysemous senses. **Journal of Memory and Language**, 47, p. 548-570. 2002.

KLEPOUSNIOTOU, E. Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. **Journal of Neurolinguistics** v. 20:1, p. 1-24. 2007.

KNUBLAUCH, H.; FERGERSON, R. W.; NOY, N. F.; MUSEN, M. A. The protégé OWL plugin: an open development environment for Semantic Web Applications. In: **Third International Semantic Web Conference - ISWC 2004**, Hiroshima, Japan, p. 1-15, 2004.

KROVETZ R; CROFT, W.B. Lexical Acquisition and Information Retrieval. In: **Fifth Conference on Applied natural language processing**, p.380-387. Disponível em <<http://portal.acm.org/citation.cfm?id=146802.146810>>. Acesso em 30 maio 2006.

KRYMOLOWSKI, Y.; ROTH, D. Incorporating Knowledge in Natural Language Learning: A Case Study. In: **COLING-ACL'98 workshop on the Usage of WordNet in Natural Language Processing Systems**. 1998.

LAKOFF, G. **Women, fire, and dangerous things: what categories reveal about the mind**. Chicago: University of Chicago Press, 1987.

LAKOFF, G.; JOHNSON, M. **Metaphors We live by**. Chicago: University of Chicago Press, 1980.

LANGACKER, R. W. **The cognitive basis of grammar**. Berlin: Mouton de Gruyter, ed. 2, 395p., 2002.

LANGACKER, R.W. **Concept, image, and symbol: the cognitive basis of grammar**. Berlin: Mouton de Gruyter. 1990.

LANGACKER, R.W. **Foundation of cognitive grammar**. Stanford: Stanford University Press, v.1: Theoretical prerequisites, 516p., 1987.

LANGACKER, R.W. **Grammar and Conceptualization**. Berlin: Mouton de Gruyter, 2000.

LEHRER, A. Homonymy and polysemy: measuring similarity of meaning. **Language Sciences**, 3, p. 33-39, 1974.

LEHRER, A. Polysemy in derivational affixes. In: NERLICH, B. et al (Ed.). **Polysemy: flexible patterns of meaning in mind and language**. Berlin: Mouton de Gruyter, p. 217-232, 2003.

LEHRER, A. **Semantic fields and lexical structure**. Amsterdam : North Holland, 1974.

LEVELT, W. J. M. **Speaking: to intention to articulation**. Cambridge, Mass.: The MIT Press, 1992.

LYONS, J. **Semantics**. Cambridge: Cambridge University Press, 1977.

MARRAFA, P. **WordNet do Português: uma base de dados de conhecimento lingüístico**. Lisboa: Instituto Camões, 2001.

MENDES, S.; CHAVES, R. P. Enriching WordNet with qualia information. In: **Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources**, Pittsburgh, June 2001. Disponível em : <<http://www.seas.smu.edu/~rada/mwnw/papers/WNW-NAACL-212.pdf.gz>>. Acesso em: 15 mar. 2009.

MICHAELIS moderno dicionário da língua portuguesa. Disponível em: <<http://michaelis.uol.com.br/moderno/portugues/index.php>>. Acesso em: 15 mar. 2009.

MILLER, G.A. Nouns in WordNet. In: FELLBAUM, C. (Ed.). **WordNet: an electronic lexical database**. Cambridge: The MIT Press, p. 23-46, 1998.

MILLER, G.A. WordNet: an online lexical database. Special issue of **International Journal of Lexicography**, 3(4), 1990.

MILLER, J.; JOHNSON-LAIRD, P. N. **Language and perception**. Cambridge, Mass.: Belknap Press of Harvard University Press, 1976.

NIRENBURG, S. e RASKIN V. **Ontological Semantics**. Mass., Cambridge: The MIT Press, 2004.

NIRENBURG, S.; RASKIN, V. **Ontological semantics**. Cambridge, Mass.: MIT Press, 2004.

NIRENBURG, S.; RASKIN, V. Ten Choices for Lexical Semantics. In. **Computer and Cognitive Science** MCC-96-304, New Mexico State University: Computing Research Laboratory, 1996.

PETERS, W.; PETERS, I.; VOSSSEN, P. Automatic sense clustering in Euro WordNet. In: **Proc. First Intl Conf on Language Resources and Evaluation**. Granada, Spain, p. 409-416, 1998.

PINKER, S. **Do que é feito o pensamento**. São Paulo: Companhia das Letras, 561p., 2008.

PUSTEJOVSKY, J. **The generative lexicon**. Cambridge, Mass.: MIT Press, 1995.

RAVIN, Y.; LEACOCK, C. (Eds.). **Polysemy: Theoretical and Computational Approaches**. Oxford: O.U.P. 2000.

RAVIN, Y.; LEACOCK, C. (Ed.). **Polysemy: theoretical and computational approaches**. Oxford: Oxford University Press, 2000.

RICE, S. Growth of a lexical network: Nine English preposition in acquisition. In: Cuyckens, H., Dirven, R. e Taylor J.R. (Eds.) **Cognitive Approaches to Lexical Semantics**. Berlin/New York: Mouton de Gruyter, p. 243-280, 2003.

RICE, S. Growth of a lexical network: nine English prepositions in acquisition. In: CUYCKENS, H.; DIRVEN, R.; TAYLOR, J. R. (Ed.). **Cognitive approaches to lexical semantics**. Berlin: Mouton de Gruyter, p. 243-280, 2003.

ROSCH, Eleanor. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104, p. 192-233, 1975.

SAINT-DIZIER, P.; VIEGAS, E. **Computational lexical semantics**. Cambridge: Cambridge University Press, 1995.

SANDRA, D.; RICE, S. Network analyses of prepositional meaning: mirroring whose mind – the linguist's or the language user's?. **Cognitive Linguistics**, v. 6, n. 1, p. 89-130, 1995.

SILVA, A.S. **A Semântica de Deixar: Uma Contribuição para a Abordagem Cognitiva em Semântica Lexical**. Braga: Empresa do Diário do Minho, Ltda., 1999.

SILVA, A.S. **O Mundo dos Sentidos em Português: Polissemia, Semântica e Cognição**. Coimbra: Edições Alamedina, p.392, 2006.

SILVA, A.S. O sentido múltiplo: polissemia, semântica e cognição. In: Feltes, H.P.M. (Ed.), **Produção de Sentido. Estudos transdisciplinares**. Caxias do Sul: Editora da Universidade de Caxias do Sul, p.91-115, 2003.

TALMY, L. Lexicalization patterns: semantic structure in lexical forms. In: SHOPEN, T. (Ed.) **Language typology and syntactic description: grammatical categories and the lexicon**. (v.3). Cambridge: Cambridge University Press, p. 57-149, 1985.

TALMY, L. **Toward a cognitive semantics**. Cambridge, Mass.: MIT Press, 2 v., 2000.

TAYLOR, J. R. **Cognitive grammar**. Oxford: Oxford University Press, 2002.

TAYLOR, J. R. Cognitive models of polysemy. In: NERLICH, B. et al (Ed.). **Polysemy: flexible patterns of meaning in mind and language**. Berlin: Mouton de Gruyter, p. 31-47, 2003b.

TAYLOR, J. R. **Linguistic categorization: prototypes in linguistic theory**. London: Clarendon Press; New York: Oxford University Press, 1995.

TAYLOR, J. R. Polysemy's paradoxes. **Language Sciences**, v. 25, p. 637-655, 2003a.

TAYLOR, J.R. **Polysemy and the lexicon**. In: G Kristiansen, M. Achard, R. Dirven, & F. J.R. de Mendonza Ibanez (Eds.), *Cognitive Linguistics: Current Application and Future Perspectives*. Berlin- New York: Mouton de Gruyter. 2006, p. 51-80.

TAYLOR, J.R. Prototypes in cognitive linguistics. In: Robinson, P.; Elis, N.C. **Handbook of cognitive linguistics and second language acquisition**. New York: Routledge, p. 39-65, 2008.

TUGGY, D. Ambiguity, polysemy, and vagueness. In: Dirk G. **Cognitive Linguistics: Basic Readings**. 2006. P. 167-184.

TYLER, A; EVANS, V. Reconsidering prepositional polysemy networks: the case of over. In: NERLICH, B. et al. (Ed.). **Polysemy: flexible patterns of meaning**. Berlin: Mouton de Gruyter, p. 95-160, 2003b.

ULLMAN, S. **Semantics: an introduction to the science of meaning**. Oxford: Blackwell, 1962.

VANDELOISE, C. Representation, prototypes, and centrality. In: TSOHATZIDIS, S. (Ed.). **Meanings and prototypes: studies in linguistic categorization**. London: Routledge, p. 403-437, 1990.

VOSSSEN, P et al. Compatibility in interpretation of relations in EuroWordNet. **Computers and the Humanities**, Dordrecht: Kluwer Academic Publishers, v. 32, p. 153-184, 1998.

VOSSSEN, P. EuroWordNet: a multilingual database with semantic networks. **Computers and the Humanities**, Dordrecht, v. 32., n. 2-3, p. 73-89, 1998.

VOSSSEN, P.; FELLBAUM C. Universals and idiosyncrasies in multilingual WordNet. In: BOAS, H.C. (ed). **Trends in linguistics: multilingual FrameNets in computational lexicography**. Berlin: Mouton de Gruyter, P. 319-346, 2009.

WEINREICH, U. Webster's third: a critique of its semantics. **International Journal of American Linguistics**, New York, v. 30, p. 405-409, 1964.

WEISZFLOG, W. (Ed.). **Michaelis português: moderno dicionário da língua portuguesa: versão 1.0**. São Paulo: DTS Software Brasil, 1998. 1 CD-ROM.

ZLATEV, J. Polysemy or generality? Mu. In: CUYCKENS, H.; DIRVEN, R.; TAYLOR, J. R. (Ed.). **Cognitive approaches to lexical semantics**. Berlin: Mouton de Gruyter, p. 447-494, 2003.