

Isa Mara da Rosa Alves
ialves@unisinos.br

Rove Luiza de Oliveira Chishman
rove@unisinos.br

Redes de polissemia: um mapeamento entre sentidos polissêmicos em ontologias semiformais

Polysemy networks: Mapping between polysemous sense in semi-formal lexical ontology

RESUMO - Fundamentando-se no paradigma cognitivo de descrição de língua, este trabalho tem como objetivo representar a polissemia de nominais de modo útil a sistemas computacionais. A tarefa desenvolveu-se em três domínios mutuamente complementares: linguístico, linguístico-computacional e computacional. Olhar para o fenômeno do significado múltiplo sob o viés da Semântica Lexical Cognitiva possibilitou descrever os sentidos como uma entidade complexa, estruturada em termos de uma rede de polissemia.

Palavras-chave: polissemia, redes de polissemia, relações de polissemia, ontologias, *wordnet*.

ABSTRACT - Based on the cognitive framework of language description, this paper aims at representing noun polysemy so that it can be useful to computational systems. The task is tackled in three mutually complementary domains: linguistic, computational-linguistic, and computational domains. Regarding the phenomenon of the multiple meaning from the Cognitive Lexical Semantics perspective has allowed for describing meanings as a complex entity, structured in terms of a polysemic network.

Key words: polysemy, polysemy network, polysemic relations, ontology, *wordnet*.

Introdução

Este trabalho visa a contribuir para o desenvolvimento de aplicações de Processamento Automático de Língua Natural (PLN), área que requer uma representação robusta e precisa do significado. Representar as relações de polissemia nesse cenário é ainda um grande desafio, tanto para a Linguística quanto para a comunidade de PLN. Neste artigo, apresentamos um modelo de rede como uma estratégia vantajosa para modelar a polissemia em ontologias do tipo lexical semiformal e usamos as *wordnets*, em especial, a WordNet.Br (Dias-da-Silva, 2010), para ilustrar a aplicabilidade da proposta. As *wordnets*, em linhas gerais, podem ser entendidas como recursos de semântica lexical que, conforme Pévot *et al.* (2010), são primeiramente léxicos, tendo em vista que suas entradas são expressões linguísticas, mas também podem ser entendidas como ontologias linguísticas, tendo em vista que são estruturadas a partir de conjuntos de sinônimos, nomeados *synsets*, que agrupam conceitos lexicalizados. Seguindo o paradigma cognitivo de descrição de língua, adotamos a posição de Pévot *et al.* (2010), que concebem o léxico como um local de armazenamento de conhecimento pessoal que pode ser facilmente recuperado a partir de dicas lexicais.

A metodologia do trabalho segue a proposta de Dias-da-Silva (1996, 2006), que prevê atividades organizadas em três domínios mutuamente complementares para o desenvolvimento de sistemas e recursos de PLN: 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 utilização de uma representação semântica múltipla, possibilitando a introdução de níveis distintos de generalidade entre os sentidos em uma base de dados de modo a reduzir a quantidade de processamento lexical a ser realizada pelo sistema. Esse tipo de tarefa tem desafiado os pesquisadores que se dedicam a construir ontologias linguísticas ou formais. Neste trabalho, nosso foco será mostrar as soluções propostas, entretanto, é preciso dizer que elas nasceram de uma revisão detalhada das problematizações envolvidas no âmbito teórico e da análise de trabalhos relacionados realizados previamente por outros grupos de pesquisa, tais como WordNet.Pr (Fellbaum, 1998), WordNet.Pt (Marrafa, 2001), GermaNet (Hamp e Feldweg, 1997).

O modelo representacional proposto foi denominado *polyset*. *Polysets* são constructos estruturados em

termos de redes de polissemia que possibilitam a representação de diferentes níveis de generalidade de sentido, diferentes graus de proeminência de sentido e diferentes tipos de relações de polissemia.

A Semântica Lexical Cognitiva (SLC) mostrou ser o paradigma teórico e descritivo mais adequado para os propósitos do modelo, tendo em vista que seu ferramental é rico o suficiente para descrever o fenômeno da multiplicidade de sentidos em termos de redes de sentidos, uma estratégia suficientemente flexível para nossos propósitos.

No domínio computacional, seguindo a metodologia de van Assem *et al.*, propomos uma estratégia para a codificação dos *polysets* em Ontology Web Language (OWL) com auxílio do editor de ontologias Protégé-OWL.

O restante do artigo está estruturado como segue. A seção *Como ontologias semiformais como wordnets lidam com a polissemia* situa o cenário no qual o estudo foi projetado. As seções *A polissemia como uma rede de sentidos* e *O modelo de rede polyset* delimitam o ferramental teórico que fundamenta o constructo *polyset*. As seções *A face linguística de construção do polyset* e *A face computacional de construção do polyset* ilustram a proposta de construção do *polyset*, a primeira focaliza o planejamento linguístico do modelo e a última sua implementação. A seção *Conclusão* sinaliza a relevância da inclusão do modelo *polyset* para as *wordnets* no contexto dos recursos lexicais.

Como ontologias semiformais como wordnets lidam com a polissemia

Representar relações entre sentidos polissêmicos em ontologias linguísticas do tipo semiformal, como podem ser entendidas as *wordnets*, exige, de início, desafiar os princípios globais da estrutura dessa rede. Para entender o desafio envolvido, mesmo sem envolver as problemáticas da definição do fenômeno, basta observarmos a organização da base: os sentidos são agrupados em *synsets* (conjuntos de sinônimos cognitivos), e suas relações são estabelecidas entre os *synsets*. Isso quer dizer que 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 para o léxico sob a perspectiva semasiológica.

Embora soluções não tenham sido inteiramente implementadas, a inclusão da polissemia é reconhecida como tarefa relevante por pesquisadores envolvidos com a construção das diversas *wordnets* – por exemplo, WordNet.Pr (Fellbaum, 1998), WordNet.Pt (Marrafa, 2001), GermaNet (Hamp e Feldweg, 1997). As propostas já apresentadas constam ainda no plano da teoria e são parciais, pois focam fundamentalmente na representação da polissemia sistemática.

Conforme veremos na próxima seção, propomos que as redes *wordnets* tenham em sua estrutura dois tipos

fundamentais de agrupamentos: conjuntos de sinônimos (*synsets*) e conjunto de sentidos polissêmicos (*polysets*), que codifiquem o conteúdo associado aos itens lexicais em termos de “categorias de sentidos.

A polissemia como uma rede de sentidos

Sabe-se que a polissemia 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 forma lexical que pode ser associada a mais de um sentido. Além disso, é convencional dizer que sentidos polissêmicos são aqueles que apresentam algum tipo de similaridade. Embora essas considerações possam parecer relativamente claras e simples, diversas questões surgem ao tentarmos demarcar os limites entre um único sentido e mais de um e ao tentarmos identificar semelhanças entre os sentidos. Para possibilitar tal delimitação, neste trabalho, evocaremos os princípios da Semântica Lexical Cognitiva (SLC), teoria que segue os pressupostos gerais da Linguística Cognitiva.

Opondo-se à concepção de que *língua*, *cognição* e *cultura* funcionam de forma independente e integrando conhecimentos adquiridos em áreas como Psicologia, Linguística, Antropologia, Filosofia e Ciências da Computação, a SLC 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 enfrentadas pela Linguística Cognitiva não são novas, mas que algumas das respostas dadas pela teoria o são e, entre elas, está o tratamento da polissemia.

É a partir de um olhar curioso, atento e flexível que a Linguística Cognitiva como um todo estuda o significado múltiplo, investigando tanto suas regularidades quanto suas irregularidades. Criticando a representação linear (ex.: dicionários tradicionais; léxicos enumerativos) como uma boa alternativa para a descrição do léxico mental humano, Langacker (1987, 2002), por exemplo, argumenta que um item lexical instancia uma categoria complexa, isto é, uma categoria de sentidos, que pode ser descrita como um complexo de sentidos relacionados que apresentam diferentes graus de incrustamento. A representação em redes aparece então como modelo alternativo suficientemente adequado para descrever o sentido como uma categoria (Geeraerts, 2006; Taylor, 1995, 2006).

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 – (Taylor, 2003), a SLC argumenta que o sentido é, por si só, uma entidade complexa, isto é, trata-se de uma categoria (Lakoff, 1987) que descreve os sentidos (potenciais) convencionalmente associados a

uma estrutura fonológica (forma lexical) estruturada em formato de uma *rede*. É importante notar ainda que, conforme argumenta Blank (2003), quando dizemos que um item lexical é polissêmico, não significa dizer que todos os sentidos associados àquele item estão relacionados ou que possuem algo em comum. Essa é a origem da compreensão da polissemia como uma rede (ou corrente) de sentidos.

Para a SLC, uma representação em redes facilita o funcionamento de rotinas cognitivas de acesso a diferentes porções de conteúdo conceitual a partir de diferentes pontos de ativação (Croft e Cruse, 2004). Para Sandra e Rice (1995), o mínimo que pode ser dito é que o modelo de redes é comprometido com o processo psicológico da categorização humana. Do ponto de vista metodológico, uma alta flexibilidade é percebida nessa construção, conforme Taylor (1995), na medida em que o foco desce para sentidos específicos, o item vai sendo visto como mais polissêmico e à medida que o foco passa para os níveis mais altos, mais esquemáticos, o item é visto como menos polissêmico, ou até monossêmico. O estudo da polissemia está, portanto, intimamente ligado ao estudo da monosssemia e da homonímia; não há uma fronteira bem definida entre esses três fenômenos. Assim, os problemas relacionados ao fenômeno remetem aos limites imprecisos entre polissemia e monosssemia, de um lado, e entre polissemia e homonímia, de outro.

Distinguir polissemia de monosssemia implica reconhecer a existência de diferentes sentidos ou de simples modulação contextual. Portanto, conceber 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 de (i) reduz os múltiplos sentidos como instância de homonímia por não serem percebidos como um conjunto (uma categoria), fato que pode ter como consequência considerar que estão sendo percebidos dois ou mais itens sem relação (e não um), instanciados na superfície da língua a partir da mesma estrutura fonológica. A negação da parte (ii) leva a entendermos a variação semântica como uma simples especificação contextual da estrutura semântica (categoria de sentidos), caracterizando a monosssemia.

Entendemos que as redes de polissemia constituem representações que possibilitam modelar em um nível suficientemente profundo as sutilezas das semelhanças de sentido que as relações de polissemia expressam. Por fim, destacamos que representar a polissemia como uma rede de sentidos pressupõe entender que: (a) os sentidos são 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); (c) a escolha por uma das estratégias - monossêmica, polissêmica ou homonímica - precisa levar em conta, de um lado, a (in)existência de relação entre os sentidos e, de outro, a participação mais ou menos efetiva do contexto 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, tal como é o modelo de redes.

O modelo de rede *polyset*

O *polyset* é um constructo lexical de base semântica. Ele representa uma rede de sentidos que define agrupamentos de sentidos polissêmicos (tanto regulares quanto irregulares). Nas redes de polissemia sincrônicas, um item lexical é concebido como uma categoria conceitual que agrupa sentidos relacionados (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. São várias as propostas de redes para tratar do léxico: *rede radial* (Brugman, 1981; Lakoff, 1987), *rede esquemática* (Langacker, 1987, 2002); *rede multidimensional* (Geeraerts [2006] [2001]); e *rede de polissemia sincrônica* (Blank, 2003). As redes sincrônicas de polissemia, 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 do modelo de redes *polyset*.

A construção de um *polyset* requer a identificação dos itens lexicais polissêmicos e a representação de seus sentidos com base em *corpus*. A seguir, caracterizaremos cada componente da estrutura do *polyset* que contribui para a delimitação dos *nós* (sentidos) e dos *arcos* (relações) que constituem rede. São eles: o *item lexical polissêmico*, a *constante*, os *sentidos polissêmicos*, os *conjunto de sentidos*, os *tipos de sentidos* e seus *efeitos de saliência*.

O item lexical polissêmico

No modelo proposto, os itens lexicais instanciados em uma língua particular são entidades simbólicas que estabelecem o pareamento entre forma (estrutura fonológica) e sentido (estrutura semântica/conceitual). Isso significa 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 (Croft e Cruse, 2004). Classificar um item lexical como

polissêmico pressupõe, assim, entender que seu conteúdo é formado por sentidos múltiplos que apresentam algum tipo de similaridade.

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 duas funções principais: servir como rótulo ou identificador do *polyset* e como chave de busca nas *wordnets*, nos dicionários e nos *corpora*.

Os sentidos polissêmicos

Sob a denominação *sentido polissêmico*, incluímos tanto sentidos plenos quanto subsentidos (facetas e microsentidos), noções propostas por Croft e Cruse (2004) que serão delimitadas a seguir. Eles são os *nós* da rede de polissemia que estrutura o *polyset*. A identificação dos sentidos foi feita a partir da análise combinada de diferentes fontes (dicionários, *wordnets*, *corpora*). O conteúdo dos *nós*, portanto, pode ser registrado a partir de um dos seguintes elementos, seguindo critério de clareza: (a) uma glosa reduzida; (b) um hiperônimo suficientemente específico; (c) um sinônimo representativo; (d) uma noção geral, que codifica um conteúdo semântico esquemático, explicitando o tipo de similaridade percebida entre eles.

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. A partir dessa lista, o objetivo é especificar os fatores de coerência que se estabelecem entre os sentidos, quando houver. Sempre que eles forem identificados, constrói-se um conjunto de sentidos, formando um conjunto polissêmico de sentidos, ou seja, um *polyset*. A partir desse segundo nível de construção do *polyset*, temos a indicação da existência de alguma similaridade, mas ainda não se sabe com precisão de que tipo ela é.

Olhando para a relação *polyset* vs. *wordnets*, podemos dizer então que os *polysets* complementam as informações associadas aos *synsets* (que já compõem as *wordnets*), refinando-os semanticamente e são, ao mesmo

tempo, mais livres em sua estrutura, uma vez que o sentido (nó da rede) 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 (Croft e Cruse, 2004; Cruse, 1995), além do tipo de relação de sentido e dos efeitos de saliência.

A delimitação dos tipos sentidos para a construção dos polysets

A especificação dos tipos de sentidos é o próximo passo e segue o critério de Croft e Cruse (2004) para a identificação de coerência e de autonomia entre sentidos. Conforme são delineadas as fronteiras entre os diferentes sentidos, as diferentes leituras podem ser classificadas como *sentidos plenos* ou subsentidos do tipo *facetas* e *microsentidos*. Sob o rótulo *sentido pleno*, estão aqueles sentidos que apresentam distinções marcantes, são bem estabelecidos, estão incrustados no léxico da língua e são fortemente delimitados por restrições convencionais estáveis. Sob o rótulo de *subsentidos*, estão aqueles sentidos considerados *facetas*, pois constituem-se partes de um mesmo sentido geral, e aqueles considerados *microsentidos*, pois são sentidos que são entendidos como uma especialização de um sentido mais geral.

Os sentidos do tipo *facetas* são percebidos como componentes distintos de um único sentido global (o sentido é formado pela cooperação entre diferentes facetas de sentido). São aqueles sentidos polissêmicos que apresentam (a) significativo grau de autonomia, mas não são (b) antagonísticos. A principal evidência de coerência entre as facetas de sentido é o fato de o leitor não precisar selecionar apenas uma delas para a interpretação de um enunciado, mas apenas colocar uma delas em evidência. O principal indício de autonomia é o fato de que as facetas não são capazes de estabelecer relação de taxonomia (*é um tipo de*) com um único elemento. Esse tipo de leitura gera o que entendemos como *polissemia regular*. Como exemplo, destacamos os sentidos convencionalmente associados a *livro*, que são ‘objeto físico’ (ex.: *Levarei o livro amarelo*), ‘conteúdo’ (ex.: *Adorei o livro indicado pela professora, a narrativa é ótima!*) e ‘texto físico’ (ex.: *O livro está com uma diagramação péssima.*). Os sentidos do tipo *microsentidos*, por sua vez, (a) apresentam um significativo grau de autonomia, embora possam ser unificados como hipônimos de uma mesma categoria e (b) baixo grau de antagonismo, tendo

em vista que funcionam como *co-autohipônimos* de um sentido geral. Além de tais propriedades, há o fato de que o hiperônimo não tem *status* padrão. Por exemplo, *faca*, que pode ser concebido como uma ‘arma’, como um ‘utensílio doméstico’, ou como um ‘instrumento cirúrgico’.

Efeitos de saliência entre os sentidos

Para a identificação do núcleo de uma categoria de sentidos representada em termos de um *polyset*, consideraremos dois critérios fundamentais: (a) a predominância na rede de polissemia e (b) a frequência de ocorrência nos *corpora*. O sentido considerado com grau maior de saliência funcional, primeiro critério, é o sentido que tem um número mais alto de conexões na rede. 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 em 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. O sentido mais frequente no *corpus* é aquele que possui maior saliência estrutural, segundo critério. Para a contagem do peso estrutural de um sentido integrante de um *polyset* foi feita uma busca orgânica no Google™.

Os arcos da rede de polissemia

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. O processo de especificação das relações na modelagem dos dados prevê duas fases. A primeira fase 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 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.

A relação de polissemia regular

Há sentidos que apresentam relação regular de polissemia do tipo *automeronímia/autoholonímia* (relação de PARTE-TODO) ou do tipo *autohiponímia/autosuperordenação* (relação É-UM) (Cruse, 2000). A relação ocorre quando os subsentidos (as facetas) associam-se por relação de *parte-todo* (ex.: porção/substância, objeto físico/conteúdo) (Cruse, 2000). A similaridade entre as facetas de sentido é alta, de maneira que ambas as facetas são componentes distintos de um mesmo sentido geral. Como a principal característica das facetas de sentido destaca-se o fato de que elas apresentam um grau mínimo de autonomia, de modo que não é necessário delimitar apenas uma delas em um contexto, embora seja comum que a cada contexto uma delas esteja em relevo. Como exemplo, destacamos *livro* ‘objeto físico’ e *livro* ‘conteúdo’. O rótulo da relação é *é_automerônimo_de* e o seu reverso é *é_autoholônimo_de*. Para a identificação da relação regular de polissemia do tipo 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 seu reverso é *é_autoholonímia_de*.

A polissemia regular do tipo *autohiponímia/autosuperordenação* (relação de É-UM) ocorre quando um dos subsentidos (microsentidos) é construído de modo mais específico em relação aos outros (Cruse, 2000). A similaridade entre os microsentidos 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. Como exemplo, destacamos *faca* como ‘utensílio doméstico’, ‘arma branca’ e ‘instrumento cirúrgico’. O rótulo da relação é *é_automerônimo_de* e o do seu reverso é *é_autosuperordenado_de*. Para a identificação da relação regular de polissemia do tipo 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.

A relação de polissemia irregular

Há sentidos com alto grau de independência entre si, mas, ainda assim, com alguma 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”. Destacamos como exemplo *cabeça* ‘membro do corpo humano’ e *cabeça* ‘chefe’ e *posição* ‘local ocupado’ e *posição* ‘opinião’. O rótulo a ser utilizado para codificar 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, apresentamos o Quadro 1, que resume as propriedades centrais de cada relação.

A face linguística de construção do *polyset*

As etapas linguísticas de construção de um *polyset* são realizadas em dois momentos. O primeiro deles é explicitado no Quadro 2, 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.

O segundo momento é ilustrado no Quadro 3. Ele consiste na especificação da rede de polissemia propriamente dita, através da representação gráfica de diferentes relações de polissemia e dos efeitos de saliência entre os sentidos.

Para ilustrar a seção, apresentaremos a seguir os dois passos descritos acima em funcionamento para a descrição do item lexical *banco*. O item *banco* é um exemplo clássico de homonímia, por não haver similaridade entre seus dois sentidos nucleares: ‘instituição financeira’ e ‘assento’. Esses não são, no entanto, os únicos sentidos convencionalmente associados ao item: conforme as Fi-

Quadro 1. Propriedades centrais da relação de polissemia.

Chart 1. Central properties of the polysemy relationship.

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	[A é parte de B, e A e B são facetas de sentido]	[A é um B, ou A é um tipo de B, e A e B são microsentidos]	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

Quadro 2. Resultado das etapas preliminares de construção do *polyset*.

Chart 2. Result of the preliminary stages of the *polyset* construction.

Constante	Indicar a constante
Exemplo	Indicar um exemplo para cada sentido em análise
Chave de busca em inglês	Indicar a chave de busca correspondente a cada sentido
Glosa reduzida	Indicar (fonte: <i>wordnets</i> e dicionários)
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.
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.
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: a. um sinônimo representativo b. um hiperônimo suficientemente específico c. uma glosa reduzida
Conjunto de sentidos	Agrupar os sentidos em conjuntos rotulados pela <i>constante</i> .

Quadro 3. Resultado das fases de construção do *polyset*.

Chart 3. Result of the stages of the *polyset* construction.

(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) <i>Status</i> dos sentidos I	Os tipos são: sentido pleno, faceta, microsentido. O tipo de sentido ficará evidente no modelo a partir da organização hierárquica e das relações entre sentidos.
(5) <i>Status</i> dos sentidos II	
(a) saliência psicológica	Determinada pela predominância.
(b) saliência estrutural	Frequência no <i>corpus</i> .

Quadro 4. Exemplos de *banco*.**Chart 4.** Bank examples.

Banco	
Exemplos	<p>1a. “Ao digitar sua senha do banco para confirmar a operação, o Itaú Shopline se comunica com a operadora do cartão que autoriza o pagamento ao estabelecimento comercial onde a compra foi feita.”</p> <p>1b. “Nos demais Estados do país a greve nos bancos privados será suspensa a partir desta quinta, segundo a Contraf”.</p> <p>1c. “Nos dias 25, 31 e 1º de janeiro, os bancos ficarão fechados.”</p> <p>1d. “Fácil comodidade para fazer consultas, efetuar pagamentos, transferências e outras transações sem precisar ir ao banco.”</p> <p>2a. Voluntários fazem campanha por banco de órgãos do ABC.</p> <p>2b. O banco de sangue de São Paulo obedece normas nacionais e internacionais de segurança.</p> <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.</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.</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.</p> <p>4. Jovem versátil sai do banco e garante vitória no fim do jogo.</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.</p> <p>6. As doações diárias de sangue não completam as necessidades do banco.</p> <p>7. 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.</p>

Para ilustrar o agrupamento polissêmico 1 associado a *banco*, observemos a Figura 1.

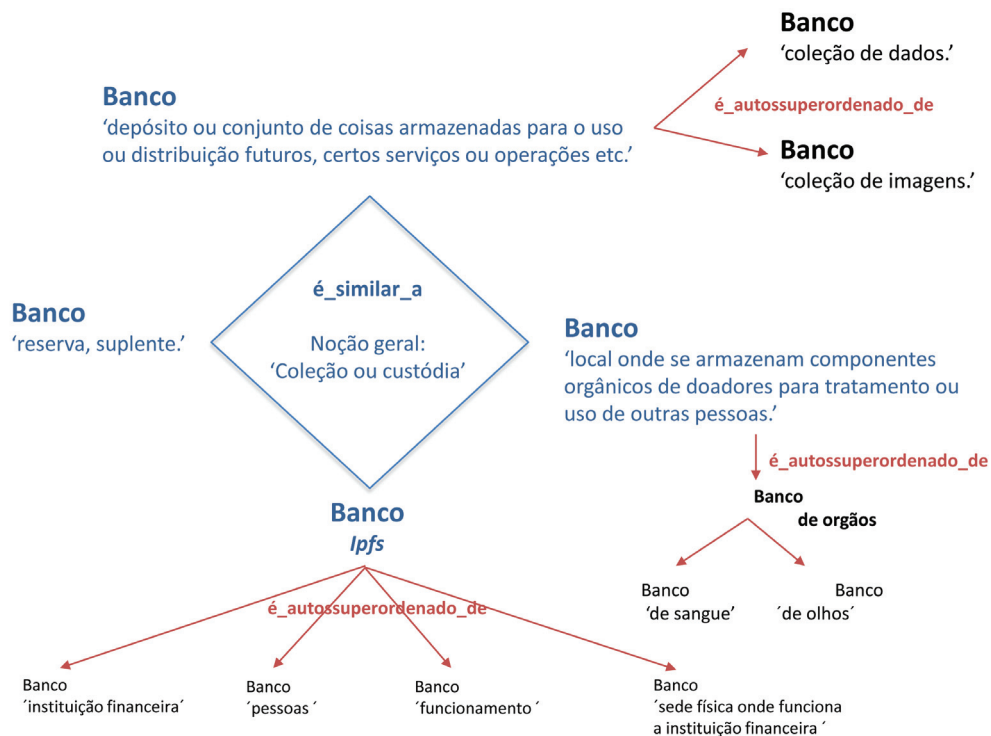


Figura 1. *Polysetl*_{banco}.
Figure 1. *Polysetl*_{bank}.

guras 1 e 2 ilustram, ele pode ser associado a dois *polysets* independentes, como os exemplos do *corpus* mostram (Quadro 4). A Figura 1 representa o primeiro *polyset* associado a *banco*. A rede é estruturada em termos de polissemia regular e irregular. Considerando os exemplos do *corpus*, optamos por organizar o conjunto de sentidos identificados para *banco* associados à noção geral ‘coleção ou custódia’ de algo de valor em diferentes níveis. Na figura, a noção está relacionada a três sentidos, de cima para baixo, da direita para a esquerda: (a) ‘depósito ou conjunto de coisas armazenadas para uso ou distribuição futuros, certos serviços ou operações, etc.’; (b) ‘reserva, suplente’; (c) ‘local onde se armazenam componentes orgânicos de doadores para tratamento ou uso de outras pessoas’; (d) uma ‘Instituição Financeira’ (representada pela figura pelas iniciais das facetas de sentidos, *ipfs*, respectivamente, ‘instituição’, ‘pessoas’, ‘funcionamento’ e ‘sede física onde funciona a instituição financeira’).

Um segundo nível de relações associa os sentidos por polissemia regular, no qual os sentidos em (c) funcionam como conteúdo esquemático para as leituras codificadas como microsentidos: ‘banco de sangue’ e ‘banco de córnea’, uma relação de polissemia taxonômica por autohiponímia/autosuperordenação. As facetas em (d) (‘instituição’, ‘pessoa’, ‘operação’ e ‘sede física onde funciona a instituição financeira’) estão relacionadas por *automeronímia*. Há regularidade porque outros itens lexicais podem ser descritos como sendo associados a facetas de sentido similares: *escola, creche, universidade*.

O agrupamento de sentidos ilustrado na figura 2 constitui um segundo *polyset*. O *polyset2* relaciona os sentidos (e) ‘assento’ e (f) ‘aglomeração de areia, conchas, fósseis e detritos de rochas’ por polissemia irregular. A similaridade que construímos entre tais sentidos associados a *banco* teve como motivação principal a acepção de *banco* fornecida por Borba (2002): ‘qualquer assento improvi-

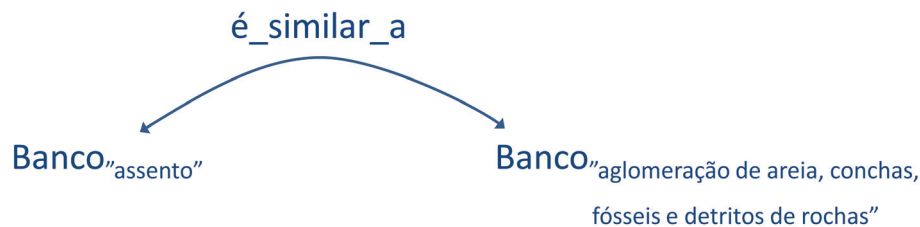


Figura 2. *Polyset2*
banco.

Figure 2. *Polyset2*
bank.

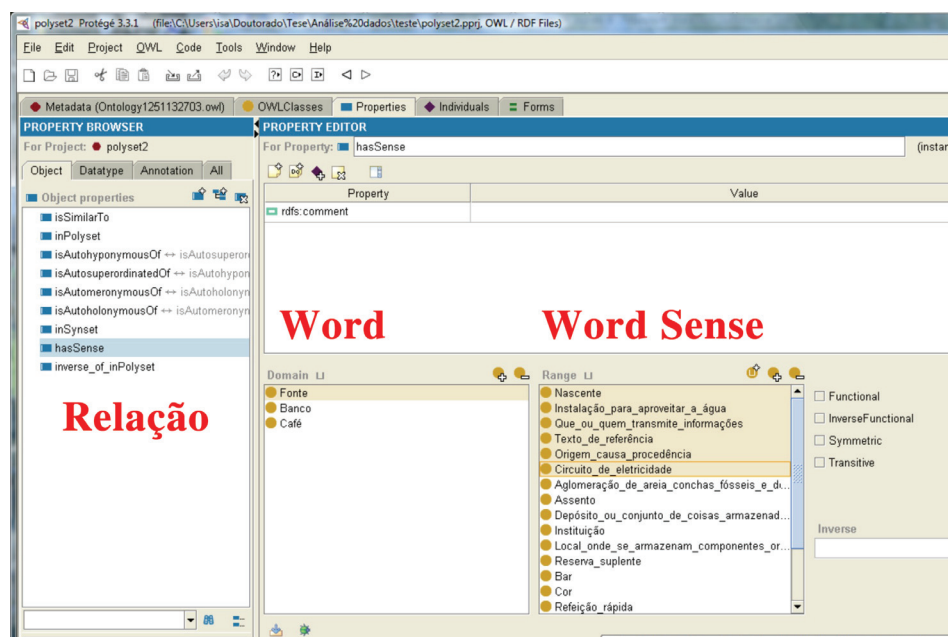


Figura 3. Protégé-OWL.

Figure 3. Protégé-OWL.

sado'. A partir daí, olhamos para o *corpus* (Quadro 4) e percebemos similaridade entre os sentidos selecionados pelos enunciados 5, que pode ser associado ao sentido (e), e 7, que pode ser associado ao sentido (f). Entendemos que tais sentidos de *banco* associam-se de duas formas: em termos da similaridade subjetiva, relacionada à noção geral *função de apoio*, ou de similaridade objetiva, baseada na forma de um banco do tipo móvel e um banco formado a partir de areia ou de fósseis e detritos.

Não percebemos os sentidos acima mencionados como microsentidos 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 no exemplo (7); fato que impossibilita preenchermos um dos critérios principais da relação de *autohiponímia*, que é a capacidade de unificação do tipo *é-um*. Além disso, há um alto grau de autonomia e antagonismo entre esses sentidos, evidenciando a ocorrência de sentidos plenos relacionados por polissemia irregular.

Ressaltamos, por fim, que, para o conteúdo semântico convencionalmente associado a *banco*, percebemos dois tipos de agrupamentos. Dois grandes agrupamentos polissêmicos múltiplos puderam ser percebidos, um deles foi ilustrado no conjunto polissêmico 1 (*polyset 1*) e o outro no conjunto polissêmico 2 (*polyset 2*). Entre os conjuntos *polysets 1* e *2*, no entanto, não percebemos qualquer similaridade semântica, logo, tem-se caracterizada a homonímia, havendo apenas uma coincidência na forma *b-a-n-c-o*.

A face computacional de construção do *polyset*

O *polyset* é facilmente implementado com o editor de ontologias Protégé-OWL. A escolha pela linguagem Ontology Web Language (OWL) deve-se ao fato de que é a linguagem recomendada pela World Wide Web Consórcio (W3C) para a codificação de ontologias em ambientes de Web Semântica. De acordo com essa proposta, as seguintes categorias devem ser codificadas (i) *words*, equivalente à constante; (ii) *wordsense* e *noun wordsense*, equivalente a sentido e sentido de nominais; e (iii) *synsets* e *noun synsets*, categoria que será substituída pelas categorias *polyset* e *noun polyset*. Além dessas informações, o editor permite que sejam codificadas propriedades e relações entre as entidades do tipo (i) e do tipo (ii), possibilitando a codificação das seguintes relações: diferentes sentidos a um mesmo *polyset*, diferentes sentidos a *polysets* distintos, sentidos a sentidos e a propriedade de saliência de sentido.

A Figura 3 ilustra uma das telas do editor Protégé-OWL. Nela está representada a implementação dos principais componentes do *polyset*, são eles, da esquerda para a direita, as *relações* (**Relação**), as *palavras* (**Word**) e os sentidos (**Word Sense**).

Conclusão

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. 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). Tal fato justifica a importância de enriquecer sistemas de recuperação de informação com estratégias de desambiguação fundamentadas em um léxico robusto, tal como as *wordnets*, que possibilitam extensões como aquelas discutidas neste trabalho.

Procuramos mostrar aqui que uma abordagem flexível como a de redes de polissemia é uma alternativa útil à estruturação de ontologias linguísticas semiformais por permitir a representação dos sentidos polissêmicos em seus diferentes níveis de generalidade e, portanto, servir a diferentes tipos de sistemas de PLN. Destaca-se, por fim, que oferecer uma representação linguístico-computacional da polissemia de nominais útil a sistemas de PLN via *wordnet*, especialmente, WN.Br, é relevante para contribuir para a solução de uma das limitações identificadas nas *wordnets* que, embora robusta, apresenta uma relativa baixa densidade de relações.

Referências

- BLANK, A. 2003. Polysemy in the lexicon and discourse. In: B. NERLICH; Z. TODD; V. HERMAN; D.D. CLARKE (eds.). *Polysemy: flexible patterns of meaning in mind and language*. Hawthorne, Mouton de Gruyter, p. 267-293.
<http://dx.doi.org/10.1515/9783110895698.267>
- BORBA, F.S. (coord.). 2002. *Dicionário de usos do português do Brasil*. São Paulo, Ed. da UNESP, 1674 p.
- BRUGMAN, C. 1981. *Story of over*. Berkeley, CA. Dissertação de mestrado. University of California, 220 p.
- CROFT, W.; CRUSE, D.A. 2004. *Cognitive linguistics*. Cambridge, Cambridge University Press, 356 p.
- CRUSE, D.A. 1995. Polysemy and related phenomena from a cognitive linguistic viewpoint. In: P. SAINT-DIZIER; E. VIEGAS (eds.), *Computational lexical semantics*. Cambridge, Cambridge University Press, p. 33-49.
<http://dx.doi.org/10.1017/CBO9780511527227.004>
- CRUSE, D.A. 2000. *Meaning in language: an introduction to semantics and pragmatics*. Cambridge, Cambridge University Press, 497 p.
- DIAS-DA-SILVA, B.C. 1996. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, SP. Tese de Doutorado. Universidade Estadual Paulista Júlio de Mesquita Filho, 197p.
- DIAS-DA-SILVA, B.C. 2010. Brazilian Portuguese WordNet: a computational-linguistic exercise of encoding bilingual relational lexicons. *International Journal of Computational Linguistics and Applications*, 1:137-150.

- DIAS-DA-SILVA, B.C. 2006. O estudo linguístico-computacional da linguagem. In: *Letras de Hoje*, 41:103-138.
- FELLBAUM, C. (ed.). 1998. *WordNet: an electronic lexical database*. Cambridge, The MIT Press, 423 p.
- GEERAERTS, D. 2006. *Words and other wonders: papers on lexical and semantic topics*. Berlin, Mouton de Gruyter, 493 p.
- HAMP, B.; FELDWEG, H. 1997. GermaNet - a Lexical-Semantic Net for German. In: *ACL WORKSHOP AUTOMATIC INFORMATION EXTRACTION AND BUILDING OF LEXICAL SEMANTIC RESOURCES FOR NLP APPLICATIONS*, Madrid, 1997. *Anais...* Madrid, p. 7-5.
- LAKOFF, G. 1987. *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago, University of Chicago Press, 614 p.
- LANGACKER, R.W. 1987. *Foundation of cognitive grammar*. Stanford, Stanford University Press, vol. 1, 628 p.
- LANGACKER, R.W. 2002. *The cognitive basis of grammar*. 2ª ed., Berlin, Mouton de Gruyter, 395 p.
- MARRAFA, P. 2001. *WordNet do Português: uma base de dados de conhecimento linguístico*. Lisboa, Instituto Camões, 77 p.
- PÉVOT, L.; HUANG, C.R.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A. 2010. Ontology and the lexicon: a multi-disciplinary perspective. In: C.R. HUANG; N. CALZOLARI; A. GANGEMI (orgs.) *Ontology and the lexicon*. Cambridge, University Press, p. 3-24.
<http://dx.doi.org/10.1017/CBO9780511676536.002>
- SANDRA, D.; RICE, S. 1995. Network analyses of prepositional meaning: mirroring whose mind – the linguist's or the language user's? *Cognitive Linguistics*, 6(1):89-130.
<http://dx.doi.org/10.1515/cogl.1995.6.1.89>
- TAYLOR, J.R. 1995. *Linguistic categorization: prototypes in linguistic theory*. London/New York, Clarendon Press/Oxford University Press, 308 p.
- TAYLOR, J.R. 2003. Polysemy's paradoxes. *Language Sciences*, 25:637-655. [http://dx.doi.org/10.1016/S0388-0001\(03\)00031-7](http://dx.doi.org/10.1016/S0388-0001(03)00031-7)
- TAYLOR, J.R. 2006. Polysemy and the lexicon. In: G. KRISTIANSEN; M. ACHARD; R. DIRVEN; F. J. R. M. IBÁÑEZ (eds.). *Cognitive Linguistics: Current Application and Future Perspectives*. Berlin/New York, Mouton de Gruyter, p. 51-80.
<http://dx.doi.org/10.1515/9783110197761.1.51>
- TYLER, A.; EVANS, V. 2003. Reconsidering prepositional polysemy networks: the case of over. In: B. NERLICH; Z. TODD; V. HERMAN; D. D. CLARKE (eds.), *Polysemy: flexible patterns of meaning*. Berlin, Mouton de Gruyter, p. 95-160.
<http://dx.doi.org/10.1515/9783110895698.95>
- VAN ASSEM, M.; GANGEMI, A.; SCHREIBER, G. 2006. Conversion of WordNet to a standard RDF/OWL representation. In: *LREC2006*, Genova, 2006. *Anais...* Genova, p. 165-169.

Submissão: 15/09/2011

Aceite: 24/11/2011

Isa Mara da Rosa Alves

Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950, Cristo Rei
93022-000, São Leopoldo, RS, Brasil

Rove Luiza de Oliveira Chishman

Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950, Cristo Rei
93022-000, São Leopoldo, RS, Brasil