



Revista Brasileira de Linguística Aplicada
ISSN: 1676-0786
rblasecretaria@gmail.com
Universidade Federal de Minas Gerais
Brasil

Di Felippo, Ariani

As abordagens teóricas e os formalismos para o tratamento computacional do significado lexical
Revista Brasileira de Linguística Aplicada, vol. 10, núm. 1, enero-marzo, 2010, pp. 43-69
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=339829612004>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

As abordagens teóricas e os formalismos para o tratamento computacional do significado lexical

Theoretical approaches and formalisms for the computational treatment of lexical meaning

Ariani Di Felippo*
UNESP

RESUMO: No âmbito do Processamento Automático de Línguas Naturais (PLN), o desenvolvimento de recursos léxico-semânticos é premente. Ao conceber os sistemas de PLN como um exercício de “engenharia da linguagem humana”, acredita-se que o desenvolvimento de tais recursos pode ser beneficiado pelos modelos de “representação do conhecimento”, desenvolvidos pela Engenharia do Conhecimento. Esses modelos, em particular, fornecem simultaneamente o arcabouço teórico-metodológico e a metalinguagem formal para o tratamento computacional do significado das unidades lexicais. Neste artigo, após a apresentação da concepção linguístico-computacional de léxico, elucidam-se os principais paradigmas de representação do conhecimento, enfatizando a abordagem do significado e a metalinguagem formal vinculadas a cada um deles.

PALAVRAS-CHAVE: semântica lexical, recurso linguístico-computacional, representação do conhecimento, processamento automático de línguas naturais.

ABSTRACT: The development of lexical-semantic resources for natural language processing is urgent. Conceiving the computational systems which process natural languages as an exercise of “human language engineering”, the development of such resources can benefit from the “knowledge representation models” provided by Knowledge Engineering. These models provide both the theoretical-methodological premises and the formal metalanguage for the computational treatment of word meaning. In this paper, we first present the linguistic-computational conception of “lexicon”. Then, we describe the main knowledge representation paradigms emphasizing the theoretical approach to word meaning and the formalism related to each of them.

KEYWORDS: Lexical Semantics, Computational-Linguistic resource, knowledge representation, natural language processing.

* arianidf@uol.com.br

Introdução

Os sistemas que processam (interpretam / geram) língua natural (registrada em meio escrito) são desenvolvidos no domínio do Processamento Automático de Línguas Naturais (PLN) e podem ser vistos, segundo Dias-da-Silva (1996, 2006), como um tipo especial de “sistema especialista”¹. Isso se baseia no fato de esses sistemas requererem uma parcela específica do conhecimento humano – o conhecimento linguístico – para realizar tarefas específicas como correção ortográfica, tradução automática, etc.

Em outras palavras, para as pesquisas que adotam a “concepção linguisticamente motivada de PLN”, o computador não poderá satisfatoriamente emular uma língua natural se não conseguir, em alguma medida, compreender o assunto que está em discussão. Logo, é preciso fornecer à máquina descrições e formalizações de dados linguísticos nas dimensões morfológica, sintática, semântico-conceitual e até mesmo pragmático-discursiva (ROCA, 2000).

Apesar de os sistemas de PLN realizarem satisfatoriamente os passos básicos de processamento da língua, eles não são capazes de “entender” o que os usuários dizem ou fazem (PALMER, 2001). Essa compreensão tem se tornado essencial para alguns sistemas que processam língua, particularmente para aqueles que processam duas ou mais línguas, como os sistemas de “tradução automática”. Para tanto, é notória a necessidade de se tratar o conhecimento de nível semântico. Neste caso, para “entender” ou “interpretar” as expressões linguísticas simples ou complexas (sintagmas e sentenças) de um texto, faz-se necessário o desenvolvimento de recursos bilíngues e multilíngues que armazenam informação semântica pertinente e formalizada (explícita) sobre as unidades lexicais (SAINT-DIZIER; VIEGAS, 1995; PALMER, 2001; HANKS, 2004).

Ao se assumir a concepção de Dias-da-Silva (2006), segundo a qual a construção de um sistema de PLN requer uma espécie de “engenharia do conhecimento linguístico”, assume-se, por conseguinte, que a construção

¹ No âmbito da Inteligência Artificial, o termo “sistema especialista” (do inglês, *expert system*) (MICROSOFT PRESS, 1998, p. 339) designa sistemas computacionais inteligentes, que tomam decisões e resolvem problemas referentes a um determinado campo de atuação, como finanças e medicina, utilizando conhecimento e regras analíticas definidas por especialistas no assunto (JACKSON, 1990; HAYES-ROTH, 1990; GIARRATAMO; RILEY, 2004). Um sistema de diagnóstico, por exemplo, necessita saber quais as características das doenças a serem diagnosticadas, pois, sem elas, é impossível elaborar um diagnóstico automaticamente.

desses recursos léxico-semânticos pode ser beneficiada pelas estratégias desenvolvidas na área da Engenharia do Conhecimento² (em inglês, *Knowledge Engineering*). Mais precisamente, a descrição e a formalização do significado lexical, tarefas da Semântica Lexical Computacional, podem ser beneficiadas pelos **modelos de representação do conhecimento** (DIAS-DA-SILVA, 2006). De modo geral, um modelo de RC pode ser caracterizado em função de vários papéis distintos e essenciais que ele desempenha (DAVIS *et al.*, 1993). Dentre eles, citam-se os papéis de: (i) substituto, funcionando como representação do conhecimento original; (ii) conjunto de compromissos ontológicos, fornecendo decisões sobre o que “ver” no original; (iii) meio para uma computação eficiente, fornecendo notação eficaz e expressiva para representar o original, e (iv) meio de expressão humana, propiciando a comunicação entre os humanos e os computadores. Assim, os modelos de “representação do conhecimento” (RC) fornecem simultaneamente o arcabouço teórico-metodológico e a tecnologia ou metalinguagem formal para a representação do significado subjacente às expressões linguísticas.

Neste trabalho, busca-se apresentar os três principais paradigmas de representação do conhecimento, enfatizando (i) a abordagem teórica do significado e (ii) o formalismo (metalinguagem ou formal) de representação fornecidos por cada um deles. Com isso, busca-se no geral familiarizar os leitores com as principais “ferramentas” disponíveis para o tratamento linguístico-computacional do significado lexical. Antes, porém, apresenta-se, logo abaixo, a concepção de léxico no domínio dos estudos linguísticos e computacionais da linguagem. Por fim, tecem-se algumas considerações finais sobre este artigo.

A arquitetura de um sistema de PLN padrão e a concepção de léxico

Teoricamente, as arquiteturas propostas para sistemas de PLN acabam por espelhar a arquitetura proposta para o sistema linguístico. Como decorrência, um sistema de PLN deve possuir (i) módulos autômatos, que realizam tarefas específicas, e (ii) módulos que armazenam um modelo de

² A área de Engenharia do Conhecimento visa ao desenvolvimento dos chamados “sistemas especialistas” ou “sistemas baseados em conhecimento” e agrega um conjunto de metodologias, técnicas e formalismos que suportam a construção desses sistemas.

conhecimento proposicional, os quais visam a criar simulacros de parcelas de mundo que lhe servem de referencial para interpretar os enunciados linguísticos (DIAS-DA-SILVA, 1996). A FIG. 1 ilustra esses dois grupos de componentes.

Os módulos de conhecimento podem ser divididos em três: o de análise, o especializado e o de síntese. As bases de conhecimento podem ser divididas em três bases: gramatical, conceitual e lexical. Com exceção do módulo especializado, os demais módulos de processamento e as bases de conhecimentos possuem estrutura e funcionamento semelhantes, embora os conteúdos possam variar em função da especificidade do sistema.

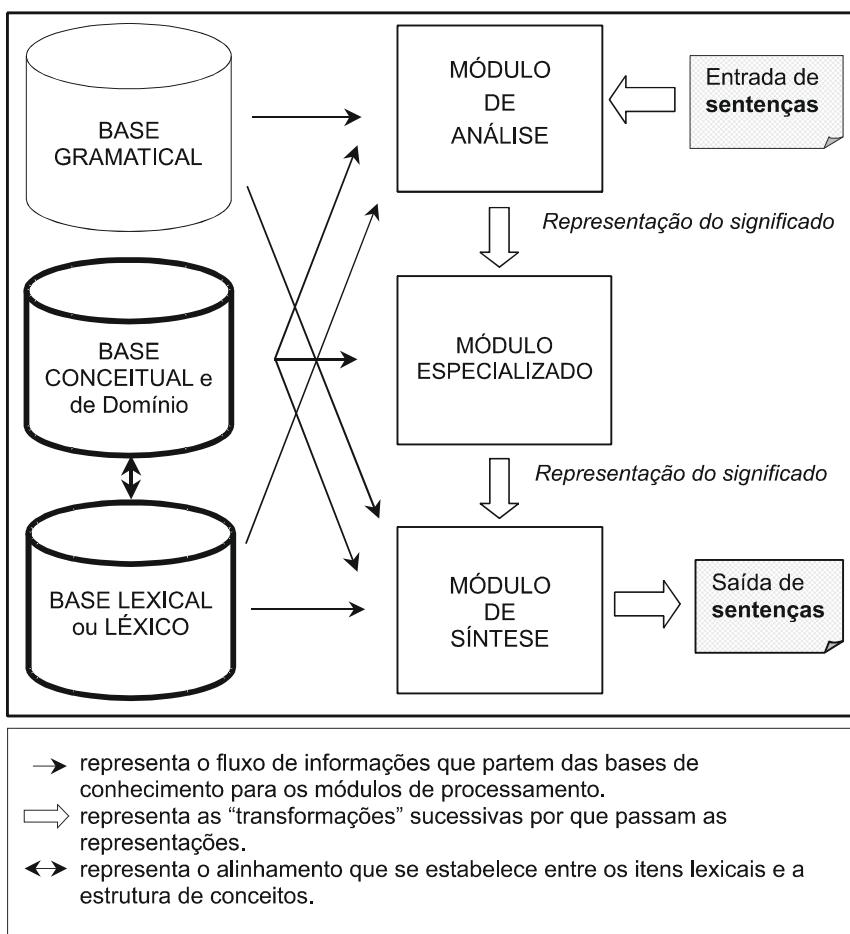


FIGURA 1 - A arquitetura padrão de um sistema de PLN de Dias-da-Silva (1996)

O Módulo de Análise (MA) é geralmente formado pelo analisador morfológico e pelo analisador sintático (ou *parser* sintático), além dos interpretadores semântico e pragmático-discursivo. Esse módulo é responsável pela construção de uma representação interna do significado das sentenças de entrada (no caso, digitadas via teclado).

O Módulo de Síntese (MS), por sua vez, transforma a representação abstrata gerada pelo MA em uma sequência de “frases contextualizadas”. Ao realizar a tarefa de construção de uma representação semântica, por exemplo, o MA utiliza-se, dependendo da sofisticação do sistema de que é parte, das bases grammatical, conceitual e lexical para executar todas ou parte das análises: morfológica, sintática, semântica e, até mesmo, pragmática.

A Base Gramatical fornece a representação das regras sintáticas da língua, que podem ser vistas como condições de admissibilidade de estruturas sintáticas bem-formadas; condições que servirão de referência para o módulo de análise – responsável pela construção das representações sintáticas, semânticas e pragmático-discursivas.

A Base Conceitual e de domínio fornece um modelo do mundo físico e conceitual, descrevendo tipos básicos de objetos, eventos, propriedades, relações e atributos em termos de representações hierarquicamente estruturadas, isto é, a sua estrutura consiste em uma rede de unidades conceituais interligadas em termos de relações de hiponímia / hiperonímia, entre outras.

Em particular, à Base Lexical, fica a tarefa de fornecer, aos MA e MS, a coleção de unidades lexicais, para as quais se faz necessária a especificação de conjuntos de traços morfológicos, sintáticos, semânticos e pragmático-discursivos. Esse tipo de base de dados, no domínio do PLN, é definido como o léxico do sistema.

Como mencionado, é premente, no PLN, o desenvolvimento de bases de dados lexicais que armazenem informação semântica sobre as unidades lexicais. Segundo a perspectiva linguisticamente motivada dos estudos do PLN, acredita-se que o tratamento computacional do significado lexical pode ser beneficiado pelos paradigmas de representação do conhecimento, os quais são apresentados na sequência.

Três paradigmas de RC

Com base em Cercone e McCalla (1987), Brachman e Levesque (1985, 2004) e Helbig (2006), foi possível identificar três paradigmas de RC: (i) RCs baseados na lógica (do inglês, *logic-oriented KR*); (ii) RCs com base em *frames*

(do inglês, *frame-oriented KR*); (iii) RCs baseadas em redes semânticas (do inglês, *network-oriented KR*). A seguir, os paradigmas são descritos em função da abordagem teórica do significado e da metalinguagem formal que englobam.

Os modelos de RC baseados na lógica

A abordagem teórica do significado: referencial

Os modelos de RC baseados na lógica foram os primeiros usados pelos pesquisadores em Inteligência Artificial para representar estruturas de conhecimentos nos computadores. Inclusive, foram considerados durante muito tempo como a resposta para todas as necessidades dos sistemas artificiais de compreensão das línguas naturais.

Nesses modelos, o significado é tratado sob o ponto de vista da **semântica referencial, extensional ou denotacional** (cf. DOWTY *et al.*, 1981; BARWISE; PERRY, 1983). Segundo essa abordagem, tal como se constituiu a partir daquele que costuma ser considerado seu fundador, G. Frege (1990)³, as expressões de uma língua adquirem significado ao denotarem objetos e eventos do mundo.

Segundo esse enfoque, o significado de expressões linguísticas (referenciais) divide-se em **sentido e denotação** (do alemão, *sinn* e *bedeutung*). Mais especificamente, diz-se que uma expressão linguística expressa um sentido e apresenta uma denotação ou denota algo.

O sentido, segundo Frege, é a forma de apresentação de um objeto, sendo que a denotação é o próprio objeto. O princípio da substituição define que sentidos diferentes podem remeter a uma mesma denotação. Por exemplo, em *O autor de Memórias Póstumas de Brás Cubas era mulato*, pode-se substituir a descrição definida *o autor de Memórias Póstumas de Brás Cubas* pela expressão *o fundador da Academia Brasileira de Letras*, sem alterar o valor de verdade da proposição expressa pela sentença em questão (MOURA, 2000). Ainda quanto à noção de sentido, diz-se que *O autor de Memórias Póstumas de Brás Cubas era mulato* e *O fundador da Academia Brasileira de Letras era mulato* são modos diferentes de apresentação (ou sentidos) de uma mesma proposição. A denotação, por sua vez, é o conjunto potencial de indivíduos a que uma expressão se refere no contexto de emissão. Categorias de expressões distintas têm denotações também distintas.

³ Tradução do texto original em alemão ‘Über Sinn und Bedeutung’, publicado em 1892.

Em primeiro lugar, apresenta-se a denotação dos predicados (lógicos e não gramaticais), como nomes comuns, adjetivos e verbos. Por exemplo, um nome comum, como *bicicleta*, denota cada um dos objetos que se classifica como bicicleta, ou seja, o conjunto das bicicletas do mundo. Na verdade, devido ao fato de que o conjunto das bicicletas do mundo muda constantemente, considera-se o conjunto de todas as bicicletas possíveis. Adjetivos, por sua vez, como *inteligente*, denotam cada um dos indivíduos que pertencem ao conjunto dos seres inteligentes. Verbos transitivos, como *querer*, denotam pares de indivíduos em que x quer y. Por fim, verbos intransitivos também denotam; pode-se dizer que *cantar*, por exemplo, denota os seres que cantam.

Em segundo lugar, apresenta-se a denotação dos nomes próprios, também denominados termos singulares. Os nomes próprios, especificamente, denotam certo indivíduo; por exemplo, *Machado de Assis* denota um indivíduo que escreveu *Memórias Póstumas de Brás Cubas* e que fundou a *Academia Brasileira de Letras*.

Por fim, apresenta-se a denotação de sentenças (ou frases). Nesse caso, diz-se que sentenças denotam valor de verdade, ou seja, a denotação de uma sentença é o seu valor de verdade. Na abordagem lógica, a ideia é a de que as sentenças declarativas denotam um valor de verdade, F(also) ou V(erdadeiro). Naturalmente, a proposição expressa por uma sentença pode ser F ou V, dependendo do contexto em que for emitida.

No âmbito da lógica intensional, as noções de sentido e denotação foram reformuladas e denominadas, respectivamente, **intensão** e **extensão** (ALLAN, 2001).

A intensão de uma expressão linguística é o conjunto de propriedades características de seu *denotatum* (aquel que é denotado). Por exemplo, as expressões *o autor de Memórias Póstumas de Brás Cubas* e *o fundador da Academia Brasileira de Letras* diferem em intensão; pois na primeira está presente a noção de escritor, enquanto que na segunda está presente a noção de fundador. As intensões são muitas vezes descritas como conceitos (ex.: CANN, 1993), no entanto, os conceitos são entidades cognitivas e as intensões, abstratas (PARTEE, 1979; ALLAN, 2001).

A extensão, por sua vez, pode ser definida como o objeto ou objetos aos quais a expressão linguística se aplica. Por exemplo, a extensão do termo singular *o autor de Memórias Póstumas de Brás Cubas* coincide com a extensão de *o fundador da Academia Brasileira de Letras*, pois ambos denotam o mesmo indivíduo, no caso, Machado de Assis. Desse ponto de vista, *pégaso* e *a fonte da vida* têm como extensões conjuntos vazios. Mais precisamente, diz-se que

extensão de um termo singular é o conjunto-unidade desse objeto; no exemplo, a extensão de ambos os termos é o conjunto-unidade Machado de Assis. A extensão de um predicado, como *admirar*, é o conjunto de todos aqueles pares ordenados de entidades x e y tais que x admira y. Vale ressaltar que a intensão de um predicado como esse é a propriedade P atribuível a indivíduos. A extensão de um predicado como *bicicleta*, por sua vez, é o conjunto de todos os objetos classificados como bicicleta, sendo que a intensão é o conjunto de propriedades características de um típico *denotatum* de bicicleta. Aplicando as noções de extensão e intensão às sentenças, diz-se que a extensão é o valor de verdade da proposição expressa pela sentença e a intensão é a própria proposição.

A intensão (ou significado) de um nome comum, como *bicicleta*, pressupõe o método clássico de categorização. Vale ressaltar, aqui, que o termo “clássico” está sendo empregado no mesmo sentido de Taylor (1985), Cruse (2004) e Croft e Cruse (2004): método que se baseia em Aristóteles e que dominou os estudos da Psicologia, Filosofia e Linguística durante quase todo o século XX.

Segundo esse método, o significado de uma expressão pode ser exaustivamente composto em um conjunto finito de primitivos conceituais que são condições necessárias e suficientes para determinar o sentido e a denotação da expressão. Assim, essa lista de propriedades precisa ser satisfeita para o uso correto de uma expressão. Em outras palavras, uma denotação pertence a uma categoria determinada se, e somente se, exibe todos e cada um dos primitivos que a definem; a falta de algum desses primitivos significa a sua exclusão automática da categoria.

Essa concepção de categorização baseia-se na distinção feita por Aristóteles entre a essência (isto é, partes imanentes responsáveis pela individualização) de uma coisa e seus acidentes (isto é, propriedades que não são definitórias). Para Aristóteles, o significado de *bicicleta*, por exemplo, é definido por uma fórmula da essência (TAYLOR, 1985). Em outras palavras, dizer que x é um y implica colocar uma entidade x na categoria y. Para isso, faz-se uma comparação das propriedades de x com os traços que definem a essência da categoria y, sendo que esse conjunto de propriedades caracteriza o significado de uma expressão.

Ilustra-se, aqui, a adoção desse método, tido clássico, com o seguinte exemplo: o significado de *bicicleta*. O significado de *bicicleta* pode ser representado, por meio da análise componencial, pelos traços [veículo], [duas rodas], [quadro], [selim] e [movida por pedais]. Esse conjunto de traços inclui

duas características: o “gênero próximo” e a “diferença específica”. No caso, *bicicleta* define-se pelo fato de pertencer à classe dos veículos (gênero próximo) e pelo fato de ter características específicas como duas rodas, selim, quadro e ser movida por pedais (diferença específica). Esses traços juntos formam a essência de “bicicletidade” ou de “ser uma bicicleta”. Segundo o princípio de que os traços da lista são necessários e essenciais, qualquer entidade no mundo que apresente esses traços pode ser corretamente designada pela expressão *bicicleta*. Caso contrário, se um dos traços não for encontrado ou se o valor de um dos traços for diferente, a entidade não apresenta a propriedade “ser uma bicicleta”.

Espera-se que o *denotatum* de um *pássaro*, por exemplo, seja bípede, tenha penas e seja capaz de voar. No entanto, há várias espécies de pássaros que não voam, por exemplo, os pinguins. Assim, vê-se que a noção de uma lista de propriedades necessárias nem sempre é suficiente para identificar ou individualizar um objeto.

Resumidamente, a abordagem referencial privilegia o aspecto informacional da língua, focalizando as conexões entre a língua e o mundo, seja ele real ou imaginário.

A metalinguagem formal dos modelos baseados na lógica

Do ponto de vista formal, as RC baseadas na lógica (clássica) utilizam os instrumentos de representação empregados na Lógica, que são os conjuntos e os modelos, buscando, assim, explicitar estruturas formais que descrevam a maneira como as expressões de uma língua natural denotam entidades do mundo.

Um **conjunto** é uma coleção de objetos, originários de um determinado domínio (ex.: números, indivíduos, etc.); o domínio de onde são originários os objetos de um conjunto é chamado de **universo de discurso** (U) (CHIERCHIA, 2003). O termo **modelo**, por sua vez, é, grosso modo, empregado para designar uma representação simplificada de uma realidade complexa que objetiva facilitar a compreensão dessa última. Do ponto de vista formal, entretanto, modelo é uma estrutura (um conjunto de objetos com propriedades e relações definidas sobre esses objetos) construída de tal forma que as expressões traduzidas para a lógica possam ser interpretadas. Por essa razão, a semântica referencial também recebe o nome de **semântica de modelo** (extensional) (ALLWOOD *et al.*, 1977; ALLAN, 2001; MÜLLER, 2003; HELBIG, 2006).

Como salienta Müller (2003), ao se adotar a lógica como metalinguagem, é preciso apresentar uma tradução sintática das expressões linguísticas para a

linguagem formal e explicar como as expressões traduzidas para a lógica são interpretadas no mundo. Em outras palavras, a representação da semântica segundo a tradição lógica requer a tradução da língua natural e a organização das entidades em um modelo de mundo logicamente organizado para que as expressões traduzidas possam ser interpretadas, posto que as expressões linguísticas são “vazias”, adquirindo significado na relação com um modelo.

Exemplificando-se como se dá a representação semântica nos modelos de RC de base lógica, considere a unidade lexical *bicicleta*. Sob o ponto de vista extensional, nomes comuns (concretos e contáveis), como *bicicleta*, denotam conjunto de indivíduos. Dessa forma, pode-se dizer que o significado da expressão *bicicleta* é “o conjunto de todas as coisas do mundo ao qual ela se aplica”. Suponha-se que o modelo de mundo possa ser reduzido ao conjunto U descrito na FIG. 2.

Usando-se noções da teoria dos conjuntos, pode-se identificar o significado de *bicicleta* pelo subconjunto B de U, que agrupa todos os indivíduos que pertencem à classe das bicicletas. Logo, B fica formalmente como: $B = \{x \mid B(x)\}$ (leia-se: B é o conjunto dos x que pertencem a U e que são bicicletas). Em outras palavras, diz-se que o significado de *bicicleta* é a sua extensão no mundo.

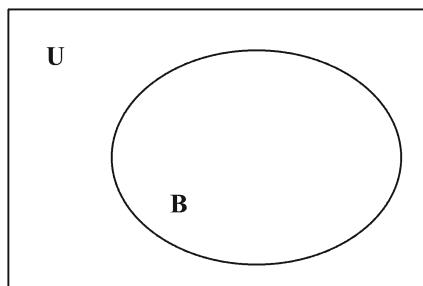


FIGURA 2 - Representação da denotação de *bicicleta* em termos de conjuntos

Alternativamente, pode-se dizer, sob o ponto de vista intensional, que o significado de *bicicleta* é o conjunto ou a classe dos indivíduos que têm a mesma propriedade, no caso, “ser bicicleta”. Nesse caso, o significado de *bicicleta* é a sua intensão, ou seja, o conjunto das propriedades que são comuns a todos os objetos dessa classe (extensão).

Suponha-se que a essência de “ser bicicleta” em U possa ser caracterizada informalmente como *<veículo que contém duas rodas, um quadro, um selim e é movido por pedais>*. Para representar a semântica de expressões

linguísticas como *bicicleta*, lança-se mão do recurso dos **postulados de significado**, introduzido por Carnap (1958). Segundo Allan (2001) e Cruse (2006), um postulado de significado é o vocabulário não-lógico (no caso deste trabalho, expressões do PB) usado na especificação semântica das expressões.

Assim, os componentes [veículo], [duas rodas], [quadro], [selim] e [movido por pedais], por exemplo, são incluídos na representação semântica de *bicicleta*, como ilustrado na FIG. 3. Nessa representação, o símbolo \wedge representa “e”. Dessa forma, leia-se: “B é o conjunto dos x que pertencem a U e que são veículos de duas rodas, com quadro, selim e movidos por pedais”.

$$B = \{ x \in U \mid \text{veículo}(x) \wedge \text{duas_rodas}(x) \wedge \text{quadro}(x) \wedge \text{selim}(x) \wedge \text{movido_por_pedais}(x) \}.$$

FIGURA 3 - Descrição formal da intensão de *bicicleta*

Utilizando-se o quantificado universal, (\forall), pode-se dizer, considerando U, que:

$\forall x | x \in U [\text{bicicleta}(x) \equiv \text{veículo}(x) \wedge \text{tem_duas_rodas}(x) \wedge \text{tem_quadro}(x) \wedge \text{tem_selim}(x) \wedge \text{movido_por_pedais}(x)]$
 (leia-se: para todo x, tal que x pertence a U, x é bicicleta se e somente se, x é um veículo que tem duas rodas, quadro, selim e é movido por pedais).

Exemplificando-se agora não a representação semântica da expressão de modo isolado, mas em contexto, considere a seguinte sentença: *A bicicleta quebrou*. Tal sentença pode ser traduzida, segundo o cálculo de predicados,⁴ na seguinte fórmula: $\exists x [B(x) \wedge Q(x)]$ (leia-se: existe um x tal que x é bicicleta e x quebrou).

⁴ A lógica de predicados (LPO), conhecida também como cálculo de predicados de primeira ordem, estende o sistema da lógica proposicional, em que proposições são descritas como entidades atômicas. Na LPO, as proposições são analisadas em termos de predicados e argumentos, tendo o formato P (t1,..., tn) (um predicado com um ou mais “argumentos”). A LPO também utiliza vários operadores que designam operações lógicas e relações e permitem a construção de expressões mais complexas. Os operadores mais importantes são os operadores de quantificação universal (\forall) (leia-se: “para todo x”) e existencial (\exists) (leia-se: “para algum x”) (ALLWOOD *et al.*, 1977; CRUSE, 2006).

Para que a interpretação dessa sentença seja possível, pode-se propor, mesmo verbalmente, um modelo ou estado-de-coisas em que haja pelo menos um indivíduo na intersecção entre os conjuntos das coisas que são bicicleta e o conjunto das coisas que quebram.

De acordo com o modelo de mundo U, o significado da referida sentença pode ser representado pela FIG. 4, sendo que o valor de verdade da proposição expressa por ela, $\exists x (B(x) \wedge Q(x))$, é verdadeiro.

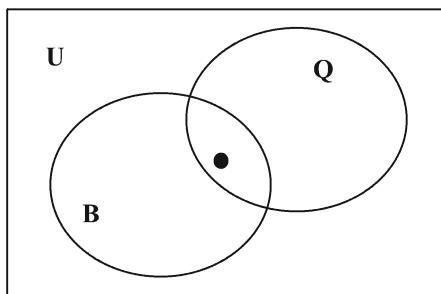


FIGURA 4 - Representação da denotação de *A bicicleta quebrou* em termos de conjuntos

Dessa forma, vê-se que as expressões recebem interpretações somente em um dado modelo ou universo de discurso. Para as sentenças em língua natural, a escolha de um modelo implica necessariamente a escolha de um contexto ou cenário. Assim, saber a denotação da sentença *A bicicleta quebrou* depende das condições fornecidas por U.

Os modelos de RC baseados em frames e em redes semânticas A abordagem teórica do significado: cognitiva

Os modelos de RC baseados em *frames* e em redes semânticas tratam o significado segundo a abordagem cognitiva, também denominada mentalista ou representacional.

Essa abordagem enquadra-se na subárea da Linguística Cognitiva, em que Lakoff (1987) e Langacker (1987, 1991) são figuras de destaque. A Linguística Cognitiva segue três princípios principais: (i) a língua não é uma faculdade cognitiva autônoma; (ii) a gramática é conceitualização; e (iii) o conhecimento da língua emerge do uso da língua (TAYLOR, 1985; CROFT; CRUSE, 2004). Assim, a Linguística Cognitiva se preocupa em determinar

como a mente humana funciona, ou seja, como a mente recebe as informações advindas de fontes diversas (visual, motora, etc.) e as processa.

Na abordagem da semântica cognitiva, o significado subjacente às expressões linguísticas é um **conceito**, o qual pode ser entendido como uma “descrição mental”, uma ideia (compartilhada pelos falantes) de um tipo de coisa (ex.: objeto, evento ou fenômeno do mundo real ou imaginário) que permite (aos falantes) “discriminar entidades desse tipo das entidades dos demais tipos”, ou seja, **categorizar** (ROSCH, 1973; ROSCH; MERVIS, 1975; TAYLOR, 1985; LÖBNER, 2002; CROFT; CRUSE, 2004; CRUSE, 2004). Os conceitos subjacentes às unidades lexicais, em especial, são denominados **conceitos lexicalizados**.

A FIG. 5 ilustra a reelaboração do triângulo semiótico de Ogden e Richards (1946) segundo as noções da abordagem cognitiva do significado, enfatizando a relação entre expressão linguística, conceito e categoria.

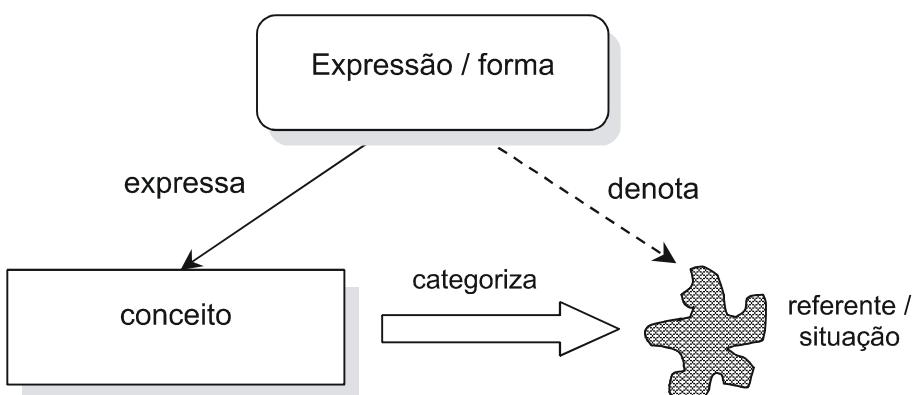


FIGURA 5 - Relação entre língua, mente e mundo

Mais especificamente, categorizar algo que se percebe é percebê-la como algo de certo tipo. Ao olhar uma bicicleta, por exemplo, reconhece-se esse objeto como sendo desse tipo. Assim, a categoria BICICLETA consiste de todas as coisas que se categorizam como sendo desse tipo. Aos elementos que pertencem a uma categoria, destina-se a denominação membros ou exemplares.

A categorização só é possível, então, se a respectiva categoria estiver disponível no sistema cognitivo, isto é, na mente. Assim sendo, a categorização requer uma representação mental de uma categoria, um conceito.

Quando o falante se depara com um objeto novo, o sistema cognitivo gera uma descrição preliminar com base nas características perceptíveis do objeto, por exemplo: cor, tamanho, etc. Se a descrição for compatível com o conceito, o objeto será categorizado como tal. Esse conceito, segundo Rosch (1973) e Rosch e Mervis (1975), baseia-se no **protótipo** da categoria, ou seja, no exemplar prototípico da categoria. Em outros termos, um membro prototípico é aquele que apresenta o maior número de atributos comumente julgados como mais característicos da categoria.

Se uma categoria é definida por um protótipo, as condições que definem o protótipo não precisam ser condições necessárias para o resto da categoria. Para a categoria AVE, por exemplo, ter-se-ia uma representação mental (conceito) de uma entidade prototípica com as características: [animal], [pequeno] e [capaz de voar]. No entanto, há membros dessa categoria que não são capazes de voar, por exemplo, os pinguins. De acordo com a teoria, isso não exclui pinguim da categoria AVE. Nesse caso, diz-se que os pinguins têm um grau de saliência menor ou são periféricos em relação ao protótipo da categoria AVE. Segundo Wittgenstein (1979), o agrupamento dos membros em uma categoria é feito por meio de similaridades parciais ou **semelhanças de família** (do inglês, *family resemblance*).

Com relação à definição de protótipo, ou seja, membro mais típico ou representativo de uma categoria, Lóbner (2002) salienta que seria mais adequado entendê-lo como um caso abstrato definido por um conceito que fixa certos traços. Assim, o conceito da bicicleta prototípica, por exemplo, conteria as propriedades mais salientes dos referentes. Neste ponto, é inevitável que se pergunte: quais são os traços que compõem o conceito? Segundo Lóbner (2002), os traços devem ser aplicáveis a uma grande parcela dos membros da categoria e a uma pequena parcela de não-membros. Por exemplo, [ter asas] e [botar ovos] são traços compartilhados por vários tipos de animais e, por isso, não são “bons” traços para caracterizar o conceito <ave>. Já os traços [ter asas] e [voar] são mais distintivos.

A noção de protótipo, aliás, fica bastante evidente quando se trata das **categorias de nível básico** (do inglês, *basic level categories*) (ROSCH, 1973) (FIG. 6). As categorias podem, então, ser vistas em níveis, por exemplo: VEÍCULO – BICICLETA – VELOCÍPEDE. As categorias, como VEÍCULO, são chamadas **superordenadas** e as categorias como VELOCÍPEDE, **subordinadas** (CRUSE, 2004).

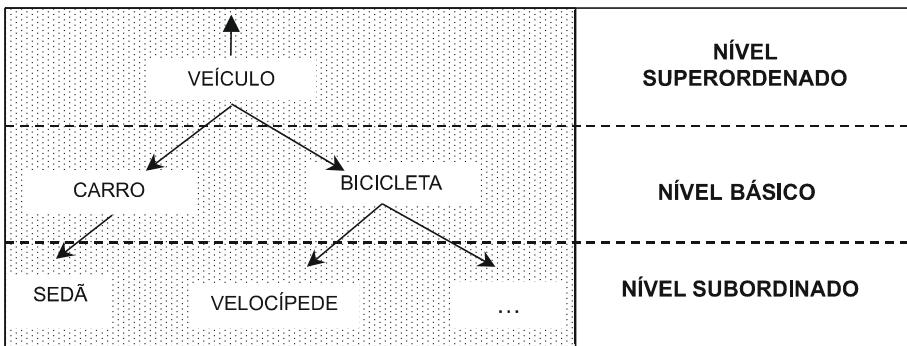


FIGURA 6 - Níveis de categorização

As categorias de nível básico, como BICICLETA, ocupam lugar de destaque no sistema cognitivo por várias razões: (a) os referentes são rapidamente reconhecidos; (b) os membros dessas categorias tem formato similar; (c) as categorias básicas visíveis permitem que se tenha uma figura ou imagem mental (ou seja, uma *gestalt*⁵) de seu conceito; (d) as categorias de nível básico são definidas pelo modo como se dá a interação com seus membros; para a categoria artefato, por exemplo, as partes características estão intimamente ligadas ao uso que se faz desses objetos (LAKOFF, 1987; CRUSE, 2004; CROFT; CRUSE, 2004).

O formato similar e o modo uniforme em que se dá a interação com os referentes / membros das categorias de nível básico visíveis facilitam a determinação do protótipo e a elaboração do conceito (TAYLOR, 1985; LAKOFF, 1987; ALLAN, 2001; LÔBNER, 2002; CRUSE, 2004; CROFT; CRUSE, 2004). Aliás, o papel de destaque das categorias básicas no sistema cognitivo tem reflexo no nível linguístico. As expressões linguísticas dessas categorias tendem a ser mais curtas e simples.

Vale ressaltar, por sim, que, similar à noção de protótipo como *gestalt*, encontra-se a noção de estereótipo, proposta por Putnam (1975). Para esse autor, o conceito subjacente a uma expressão linguística (tipicamente lexicalizada) não é composto por um conjunto bem-definido de propriedades necessariamente encontradas em todo referente, mas, sim, por um “conjunto mínimo de fatos estereotipados” sobre um referente típico. Como saliente Allan (2001), enquanto a noção de protótipo implica na escolha do “melhor

⁵ O termo “gestalt” é emprestado do alemão e significa grosso modo “imagem mental única” (ALLAN, 2001).

exemplar” da categoria, a noção de estereótipo implica na escolha de um “exemplar típico”.

Na abordagem representacional, defende-se amplamente a organização global dos conceitos. Em outras palavras, isso quer dizer que os conceitos não estão isolados na mente ou organizados aleatoriamente. Ao contrário, há relações entre os conceitos. Segundo Cruse (1986, 2004), por exemplo, os conceitos estariam organizados pelas relações lógico-conceituais. Dentre elas, destacam-se as relações de hiponímia⁶/ hiperonímia e meronímia⁷/ holonímia. As relações de hiponímia e meronímia e outras são a base para a estruturação dos conceitos nos modelos de representação do conhecimento baseados em redes semânticas. Tais relações estruturam, por exemplo, as redes lexicais em formato *wordnet*, que seguem o desenvolvimento da base da WordNet de Princeton⁸ (FELLBAUM, 1998). Nessa base, o conceito codificado pelo *synset* (isto é, conjunto de sinônimos ou quase sinônimos) {car; auto; automobile; machine; motorcar}, por exemplo, está relacionado ao conceito mais geral codificado pelo *synset* {motor vehicle; automotive vehicle}.

Segundo autores como Croft e Cruse (2004), por sua vez, os conceitos não estão organizados em função de relações como hiponímia e meronímia. Sob esse ponto de vista, aliás, Fillmore propôs a Semântica de Frames (FILLMORE, 1976). Segundo essa teoria, os conceitos organizam-se em função de **frames semânticos**, ou seja, esquemas de situações típicas, em que participantes, propriedades, papéis e outros elementos conceituais são estabelecidos (FILLMORE *et al.*, 2001). Os frames semânticos também podem ser entendidos como padrões globais de conhecimento de senso comum sobre conceitos centrais, de tal forma que um item lexical denotando o conceito evoca todo o frame semântico (BEAUGRANDE; DRESSLER, *apud* TAYLOR, 1985). No conhecido exemplo de Schank e Abelson (1977),

⁶ A hiponímia também pode se denominada de: subordinação / superordenação, subclasse / (super)classe, relação de classificação e relação *é um*.

⁷ A meronímia também recebe as seguintes denominações: “parte de” e “parte / todo”.

⁸ A WordNet de Princeton (<http://wordnet.princeton.edu/>) é uma base de dados em formato eletrônico em que unidades lexicais (do inglês), pertencentes às categorias dos substantivos, verbos, adjetivos e advérbios e codificadas sob a forma de *synsets* (abreviação do termo em inglês “*synonym set*”), estão conceitualmente relacionadas pelas relações de hiponímia / hiperonímia, antónimia, meronímia, acarretamento e causa. Essa base tem sido utilizada em várias aplicações de PLN, por exemplo: tradução automática, sumarização, recuperação de informação, entre outras.

o conceito *<restaurante>* está associado a outros conceitos, os quais são denominados elementos-frame, como *<freguês>*, *<pedido>*, *<conta>*, *<garçom>*, etc. Os frames podem representar também sequências de eventos. Nesse caso, os *frames* são nomeados **frames dinâmicos ou scripts**. Um *script* define uma sequência de eventos em função de protótipos para as pessoas, propriedades e objetos envolvidos na sequência de eventos (HELBIG, 2006). Em outras palavras, um *script* é a composição de uma sequência de ações que se repetem regularmente em situações similares e são representadas como um esquema padronizado (*frame*). Por exemplo, o significado de *comer em um restaurante* pode envolver os (i) eventos de *entrar*, *fazer o pedido*, *comer*, *pagar* e *sair* e os (ii) papéis de *freguês*, *garçom* e *caixa* (ALLAN, 2001). Os frames semânticos são responsáveis pela organização dos conceitos nos modelos de representação do conhecimento baseados em *frames*. No PLN, os frames semânticos constituem a base de organização de um recurso linguístico-computacional do inglês norte-americano bastante reconhecido, a base FrameNet⁹ (BAKER *et al*, 1998). Nessa base, o conceito *<frasco>* (do inglês, *bottle*) está associado a elementos-frame como [conteúdo], [material], [possuidor], [parte], entre outros.

A metalinguagem formal dos modelos baseados em frames

Do ponto de vista formal e interno, um *frame* é uma estrutura **atributo-valor**, composta por atributos (do inglês, *slots*) que são preenchidos com valores (do inglês, *fillers*) apropriados (MINSKY, 1975; HANDKE, 1995), assim como ilustrado na FIG. 7.

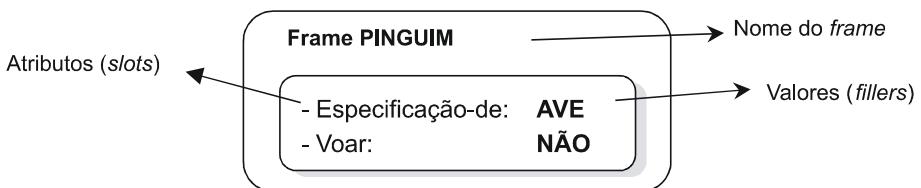


FIGURA 7 - Estrutura interna de um *frame*

⁹ <http://framenet.icsi.berkeley.edu/>

Do ponto de vista da organização global, um *frame* está relacionado a outros (subjacente, por exemplo, a um objeto, um fato ou um evento) em uma hierarquia de *frames*. A FIG. 8, baseada em Handke (1995, p. 100), ilustra o *frame* subjacente à expressão *pinguim* e sua relação com outro *frame*.

Mais especificamente, o *frame* ilustrado na FIG. 7 possui dois tipos de atributos. O atributo Especificação-de é usado para estabelecer a relação de hiponímia entre o *frame* em questão e seu *frame* hiperônimo, no caso, AVE. Esse atributo é necessário para a classificação de um *frame* e essencial para estabelecer a hierarquia entre os *frames*. Além desse atributo central, mais atributos podem ser especificados. No caso, o *frame* em questão descreve o atributo Voar, que apresenta um *valor padrão* (do inglês, *default values*) [não]. Os valores-padrão de um *frame* são especificados com base na noção de protótipos. Por essa razão, diz-se que os *frames* podem ser considerados como uma formalização da teoria dos protótipos (HANDKE, 1995).

Com base na noção de protótipo, pode-se especificar, por exemplo, que o *frame* AVE, ilustrado na FIG. 8, tem os seguintes atributos e valores-padrão: [Número de asas: 2; Bico: SIM; Pena: SIM; Voar: SIM]. O *frame* PINGUIM, enquanto hipônimo do *frame* AVE, herda os atributos e valores deste. O mecanismo de herança empregado nas representações baseadas em *frames* é dito não-monotônico, ou seja, a informação de um *frame* genérico é herdada até que não haja outra informação disponível nos *frames* mais específicos. O *frame* PINGUIM, por exemplo, especifica o seguinte par atributo-valor, [Voar: NÃO], cujo valor é conflitante com o valor do *frame* hiperônimo [Voar: SIM]. Quando ocorre esse tipo de conflito, considera-se a informação mais específica. No caso, sobrescreve-se o valor *padrão* do *frame* AVE pelo valor do *frame* PINGUIM.

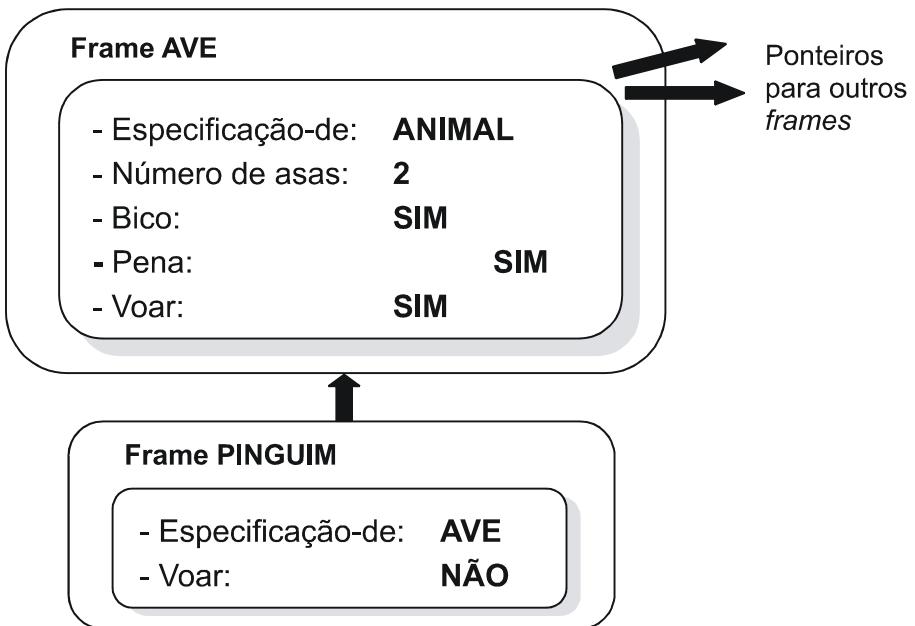


FIGURA 8 - Relação entre *frames*

Do ponto de vista da teoria dos protótipos, esse conflito é gerado pelo fato de os pinguins não serem membros prototípicos da categoria das aves e, por isso, não apresentarem todas as propriedades características das aves.

A metalinguagem formal dos modelos baseados em redes semânticas

As redes semânticas (RSs) constituem todo um paradigma de representação do conhecimento. Nesse paradigma, a representação dos conceitos é feita por **nós** e a das relações por **arcos** rotulados entre os nós (HANDKE, 1995). A FIG. 9 ilustra os construtos básicos de uma rede semântica, ou seja, os nós (ou conceitos, C) e os arcos (relações, REL).



FIGURA 9 - Ilustração dos construtos básicos de uma rede semântica

As redes semânticas têm uma longa tradição na área de representação do conhecimento. Elas se tornaram conhecidas por meio do trabalho do botânico Ritchens, que em 1956 propôs o formalismo (HUTCHINS, 1986).

As justificativas psicológicas para esse tipo de representação foram configuradas na década de 1960, especificamente por meio do trabalho colaborativo do cientista da computação A. M. Collins e do psicólogo M. R. Quillian (HANDKE, 1995).

O primeiro a considerar que esse tipo de representação reflete a organização da memória humana foi Quillian (1967, 1968), que propôs um modelo computacional da memória humana chamado **memória semântica**. Segundo o autor, esse modelo, em que conceitos são representados por nós, e relações entre conceitos, por arcos, explicaria diversos resultados experimentais sobre o comportamento da memória humana. Explicaria, por exemplo, o fato de que o reconhecimento de objetos que pertencem a classes mais numerosas requer mais tempo do que o reconhecimento dos pertencentes a classes menos numerosas (COLLINS; QUILLIAN, 1969; WILKINS, 1971; SMITH, 1978; MILLER; CHARLES, 1991).

Nos testes feitos por Quillian e reportados por Collins e Quillian (1969), assumiu-se o tempo de reação para indicar o número de níveis hierárquicos que separavam dois conceitos. Eles observaram, por exemplo, que os sujeitos dos experimentos levavam mais tempo para responder se é verdade que “um canário pode cantar” que para responder se “um canário pode voar”, e ainda mais tempo para responder se “um canário tem pele”. Nesse exemplo, assume-se que [cantar] é armazenado na memória como um traço de canário, [voar] é armazenado como traço de pássaro e [ter pele], como traço de animal. Segundo os autores, se esses três traços estivessem diretamente associados a canário, o tempo de resposta às perguntas seria o mesmo.

A FIG. 10 ilustra um fragmento de uma RS em que, por questão de simplicidade, as relações entre os nós são especificadas por 2 tipos de arcos, rotulados por “é um (tipo de)”, que expressa a relação taxonômica, e por “tem”, que expressa a relação de meronímia.

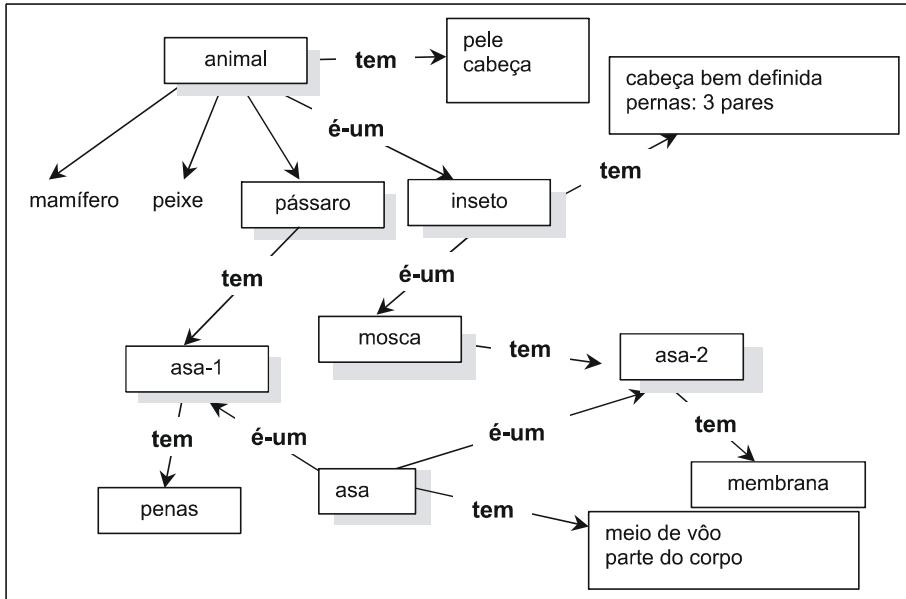


FIGURA 10 - Ilustração de um fragmento de uma rede semântica

Em uma RS, como a ilustrada na FIG. 10, os conceitos estão organizados hierarquicamente; nessa hierarquia, há um nó superior ao qual estão ligados os nós filhos, os quais, por sua vez, também têm outros conceitos como filhos e assim sucessivamente (BRACHMAN, 1979). Como salienta Dias-da-Silva (1996), cada nó pode representar um tipo ou subtipo semântico (ou conceitos) e cada arco direcionado pode não só representar relações “é-um (tipo de)” ou “é parte de”, mas também relações temáticas (agente, paciente, etc.) que se estabelecem entre predicados e seus argumentos (cf. FIG. 10). Outro aspecto importante desse formalismo é que, além de permitir a representação de tipos de estruturas, as redes semânticas permitem representar estruturas realizadas. Antes, porém, de se apresentar essa possibilidade, destacam-se os níveis de representação (ou domínios) propostos por Jackendoff (1983). O autor assume que a estrutura do pensamento humano organiza-se em função de quatro domínios – **mundo real**, **mundo projetado** (ou mundo da experiência), **estrutura conceitual** (ou conceitos) e **expressões linguísticas** (ou formas), a partir dos quais o autor discute as relações entre as expressões linguísticas e a realidade.

Ao se considerar que o “micromundo” a ser construído em um sistema de PLN corresponde a uma espécie de mundo projetado, é possível estabelecer

que a expressão linguística *bicicleta* (forma) é um tipo de <bicicleta> (conceito) e que uma realização concreta desse conceito como em, por exemplo, *Eu quero sua bicicleta*, é #bicicleta77# (referente) (DIAS-DA-SILVA, 1996). Dessa forma, as redes semânticas têm a possibilidade de suportar a especificação do significado intensional e extensional. Em outras palavras, é possível estabelecer a relação entre a expressão linguística, o conceito (significado potencial) e um possível referente, assim como ilustrado na FIG. 11, elaborada com base em Dias-da-Silva (1996).

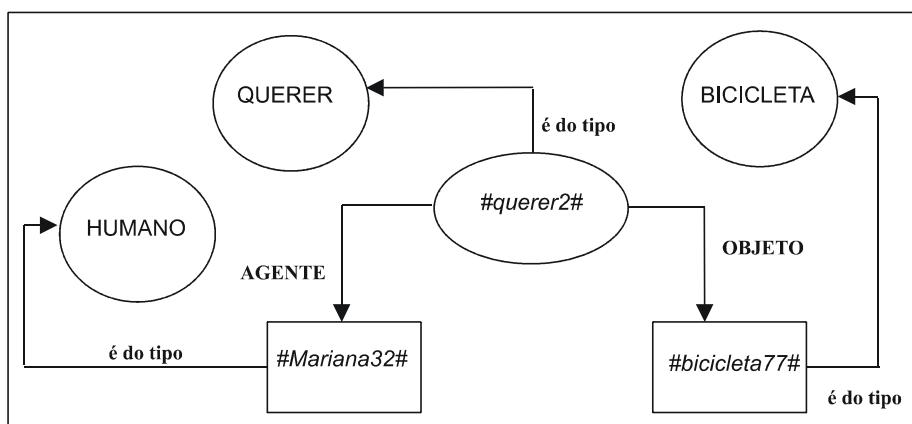


FIGURA 11 - Rede semântica que representa parte do significado expresso pela frase *Mariana quer a bicicleta*

Segundo Handke (1995), dois fatores foram fundamentais para a popularidade das redes semânticas como modelo de representação do conhecimento. O primeiro deles é elegância com que deduções podem ser feitas; por exemplo, para se deduzir que uma mosca tem 3 pares de pernas, é preciso apenas seguir a hierarquia “é um (tipo de)”, assumindo-se que os atributos associados ao nó superior, no caso, associados a <inseto>, são válidos para nós inferiores, no caso, o conceito <mosca>. O segundo fator de popularidade é a simplicidade e elegância com que as RSs podem ser implementadas.

Considerações finais

Ao assumir que um sistema de PLN é um tipo de sistema especialista, assume-se, consequentemente, que as pesquisas nessa área envolvem uma

espécie de “engenharia do conhecimento linguístico”. Como tal, as pesquisas em PLN podem usufruir dos modelos de RC desenvolvidos pela Engenharia do Conhecimento (DIAS-DA-SILVA, 2006), os quais fornecem simultaneamente a abordagem teórica segundo a qual o objeto é investigado e a metalinguagem formal para a descrição do conhecimento.

Quanto às abordagens teóricas do significado subjacentes aos paradigmas ou modelos de RC, salienta-se que a variedade de abordagens mostra que o estudo do significado pode ser feito de vários ângulos. Os modelos baseados na lógica clássica, por exemplo, focalizam a relação entre expressões linguísticas e mundo (abordagem extensional), enquanto os modelos baseados em redes semânticas e *frames* focalizam a relação entre expressões linguísticas e representações mentais (abordagem cognitiva). Provavelmente, essas perspectivas não são totalmente incompatíveis, pois o significado possui realmente várias facetas.

Ao se adotar determinado modelo ou paradigma de RC para o tratamento linguístico-computacional do significado das unidades lexicais, adota-se consequentemente um conjunto específico de princípios fundamentais que gerem o estudo do conhecimento, no caso, semântico, e uma tecnologia ou metalinguagem formal para a descrição desse conhecimento. Em outras palavras, selecionar uma determinada representação do conhecimento é, ao mesmo tempo, adotar um conjunto de decisões sobre o que “ver” no original (no caso, no significado das unidades lexicais). Nas palavras de Davis *et al.* (1993), selecionar uma representação do conhecimento significa assumir um conjunto de “compromissos ontológicos” (do inglês, *ontological commitments*), que pode ser considerado a contribuição mais importante de uma representação do conhecimento.

Assim, para o tratamento computacional do significado lexical, os pesquisadores do PLN precisam ter em mente que esses compromissos ontológicos, ao estabelecerem certa visão simplificada do objeto original, não são efeitos colaterais da escolha de uma representação do conhecimento; eles são, na verdade, a essência.

Assinala-se, por fim, que a inclusão de conhecimento léxico-semântico nos sistemas de PLN pode contribuir para melhorar o desempenho de tais sistemas. Se, por um lado, o PLN se beneficia com as informações providas pelos estudos linguísticos, por outro lado, a Linguística também pode ser beneficiada pelas pesquisas do PLN. Como evidenciou Winograd, pesquisar o PLN pode ser um modo acadêmico de investigação que pode auxiliar na compreensão dos próprios fatos da língua:

Todo mundo é capaz de compreender uma língua. A maior parte do tempo de nossas vidas é preenchida por atos de fala, leitura ou pensamentos, sem sequer notarmos a grande complexidade da linguagem. Ainda não sabemos como nós sabemos tanto [...] Os modelos [de PLN] são necessariamente incompletos [...] Mas, mesmo assim, constituem um referencial claro por meio do qual podemos refletir sobre o que é que fazemos quando compreendemos uma língua natural ou reagimos aos atos de fala nela codificados (WINOGRAD, 1972, tradução nossa).

Agradecimento

Ao CNPq (Processo: 140446/2005-2), que financiou a pesquisa de doutorado da qual este trabalho é parte.

Referências

- ALLAN, K. *Natural language semantics*. Oxford: Oxford Blackwell Publishers, 2001.
- ALLWOOD, J. et al. *Logic and linguistics*. Cambridge: CUP, 1977.
- BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The Berkeley FrameNet project. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS (COLING/ACL), 17, Montreal, Quebec (Canada). *Proceedings...* Montreal, Quebec (Canada), 1998, p. 86-90.
- BARWISE, J.; PERRY, J. *Situations and attitudes*. Cambridge: Cambridge University Press, 1983.
- BRACHMAN, R. J. On the epistemological status of semantic networks. In: FINDLER, N. V. (Ed.). *Associative networks*. New York: Academic Press, 1979, p. 3-50.
- BRACHMAN, R. J; LEVESQUE, H (Ed.). *Readings in knowledge representation*. San Mateo, Ca: Morgan Kaufmann, 1985.
- _____. *Knowledge representation and reasoning*. San Mateo, Ca: Morgan Kaufmann, 2004.
- CANN, R. *Formal semantics*. Cambridge: Cambridge University Press, 1993.
- CARNAP, R. *Meaning and necessity*. Chicago: Chicago Phoenix Books, 1958.
- CERCONE, N.; MCCALLA, G. (Ed.). *The knowledge frontier: essays in the representation of knowledge*. New York: Springer-Verlag, 1987.
- CHIERCHIA, G. *Semântica*. Campinas, SP: Editora da UNICAMP; Londrina, Pr: EDUEL, 2003. [Tradução]

- COLLINS, A. M.; QUILLIAN, M. R. Retrieval time from semantic memory. *Journal of verbal behavior and verbal learning*, 8, p. 240-247, 1969.
- CROFT, W.; CRUSE, A. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.
- CRUSE, A. *Lexical semantics*. Cambridge: Cambridge University Press, 1986.
- _____. *Meaning in language: an introduction to semantics and pragmatics*. Oxford: Oxford University Press, 2004.
- _____. *Glossary of semantics and pragmatics*. United Kingdom: Edinburgh University Press, 2006.
- DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What is knowledge representation? *AI Magazine*, v. 14, n. 1, p. 17-33, 1993.
- DIAS-DA-SILVA, B. C. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, 1996. 272p. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.
- _____. O estudo linguístico-computacional da linguagem. *Letras de Hoje*, v. 41, n. 2, p. 103-138, 2006.
- DOWTY, D. R. et al. *Introducion to Montague semantics*. Dordrecht, Reidel, 1981.
- FELLBAUM, C. (Ed.). *WordNet: an electronic lexical database*. Cambridge, MA: MIT Press, 1998.
- FILLMORE, C. Frame semantics and the nature of language. In: CONFERENCE ON THE ORIGIN AND DEVELOPMENT OF LANGUAGE AND SPEECH, 1976. *Annals...*, v. 280, p. 20-32, 1976.
- _____. WOOTERS, C.; BAKER, C. F. Building a large lexical databank which provides deep semantics. In: PACIFIC ASIAN CONFERENCE ON LANGUAGE, INFORMATION AND COMPUTATION – PACLIC, 15, 2001, Hong Kong. *Proceedings...* Hong Kong, 2001. Disponível em: <<http://www.icsi.berkeley.edu/ framenet/papers/dsemlex16.ps.gz>>. Acesso em: 20 Jul. 2007.
- FREGE, G. On sense and nominatum. In: MARTINICH, A.P., (Ed.). *The philosophy of language*. New York/ Oxford: Oxford University Press, 1990, p. 190-202. [Tradução]
- GIARRATANO, J.C.; RILEY, G.D. *Expert systems: principles and programming*. Boston: Course Techonology, 2004.
- HANDKE, J. *The structure of the lexicon: human vs machine*. Berlin: Mouton de Gruyter, 1995.

- HANKS, P. Lexicography. In: MITKOV, R. (Ed.). *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press, 2004, p. 48-69.
- HAYES-ROTH, F. Expert systems. In: SHAPIRO, E. (Ed.). *Encyclopedia of artificial intelligence*. New York, Wiley, 1990, p. 287-298.
- HELBIG, H. *Knowledge representation and semantics for natural language*. Berlin, Heidelberg: Springer-Verlag, 2006.
- HUTCHINS, W. J. *Machine translation*: past, presence, future. Ellis Horwwod/Wiley, Chichester/New York, 1986.
- JACKENDOFF, R. *Semantics and cognition*. Cambridge, Mass.: MIT Press, 1983.
- JACKSON, P. *Introduction to expert systems*. Wokingham: Addison-Wesley, 1990.
- LAKOFF, G. *Women, fire and dangerous things*: what categories reveal about mind. Chicago: University of Chicago Press, 1987.
- LANGACKER, R .W. *Foundations of cognitive grammar*, v. 1. Stanford: Stanford University Press, 1987.
- _____. *Foundations of cognitive grammar*, v. 2. Stanford: Stanford University Press, 1991.
- LÖBNER, S. *Understanding semantics*. Oxford: Oxford University Press, 2002.
- MICROSOFT PRESS. *Microsoft press dicionário de informática*. Rio de Janeiro: Editora Campus, 805 p., 1998.
- MILLER, G. A.; CHARLES, W. G. Contextual correlates of semantic similarity. *Language and cognitive processes*, 6, p. 1-28, 1991.
- MINSKY, M. A framework for representing knowledge. In: HAUGELAND, J., (Ed.). *Mind design*. Cambridge, Mass.: The MIT Press, 1975, p. 95-128.
- MOURA, H. M. de M. *Significação e contexto*: uma introdução a questões de semântica e pragmática. Florianópolis: Editora Insular, 2000.
- MÜLLER, A, L. et al., (Org.). *Semântica formal*. São Paulo: Contexto, 2003.
- OGDEN, C.K.; RICHARDS, I. A. *The meaning of meaning*. New York: Harcourt Brace Javanovitch, 1946.
- PALMER, M. Multilingual resources, multilingual information management: current levels and future abilities. *Linguistica Computazionale*, v. XIV-XV, p. 1-33, 2001.
- PARTEE, B. H. Semantics: mathematics or psychology. In: BÄUERLE, R. et al., (Ed.). *Semantics from different points of view*. Berlin: Springer-Verlag, 1979, p. 1-14.

- PUTNAM, H. The meaning of “meaning”. In: GUNDERSON, K. (Ed.). *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press, 1975, p. 131–193.
- QUILLIAN, M. R. Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12, p. 410-430, 1967.
- _____. Semantic Memory. In: MINSKY, M. (Ed.). *Semantic information processing*. Cambridge, Mass.: MIT Press, 1968, p. 227-270.
- ROCA, S. C. Individuación e información parte-todo. Representación para el procesamiento computacional del lenguaje. *Estudios de Lingüística Española*, v. 8, 2000. Disponível em: <<http://elies.rediris.es/elies8/>>. Acesso em: 10 jun. 2005.
- ROSCH, E. Natural categories. *Cognitive Psychology*, 4, p. 328-350, 1973.
- ROSCH, E.; MERVIS, C. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, 7, p. 573-605, 1975.
- SAINT-DIZIER, P.; VIEGAS, E. *Computational lexical semantics*. Cambridge: Cambridge University Press, 1995.
- SCHANK, R. C.; ABELSON, R. P. *Scripts, plans, goals and understanding*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.
- SMITH, E. E. Theories of semantic memory. In: ESTES, W.K. (Ed.). *Handbook of learning and cognitive processes*, 5. Hillsdale, N.J: Lawrence Erlbaum Ass., 1978.
- TAYLOR, J. R. *Linguistic categorization: prototypes in linguistic theory*. Oxford: Clarendon Press, 1985.
- WILKINS, A. J. Conjoint frequency, category size, and categorization time. *Journal of verbal learning and verbal behavior*, 10, p. 382-385, 1971.
- WINOGRAD, T. *Understanding natural language*. New York: Academic Press, 1972.
- WITTGENSTEIN, L. *Investigações filosóficas*. São Paulo: Ed. Abril Cultural, 1979.

Recebido em 22/07/08. Aprovado em 17/02/09.