

RELAÇÕES PARADIGMÁTICAS E SINTAGMÁTICAS NA CONSTRUÇÃO DE UM *DICIONÁRIO ANALÓGICO*

Felipe Iszlaji de ALBUQUERQUE
Bento Carlos Dias da SILVA

ABSTRACT: This paper is inserted in the context of drawing up a model linguistic-computational for a thesaurus of the Brazilian Portuguese language, with the criteria of relevance, available on the global network of computers. Our proposal for a thesaurus on the internet aims to structure semantic relations both paradigmatic as syntagmatic. Our hypothesis is that the semantic relations paradigmatic (synonymy, antonymy, hyperonymy, etc.) are worked better in structures such as the wordnets and semantic relations syntagmatic are adequately dealt by semantics of frames, that is, in structures such as the framenets.

KEYWORDS: thesaurus; semantic relations; paradigm; syntagm.

1. Introdução

O presente artigo se insere no contexto da elaboração de um modelo lingüístico-computacional para um dicionário analógico da língua portuguesa, com critérios de relevância, que irá integralizar uma plataforma digital de ferramentas lingüísticas, disponível na rede mundial de computadores. Tal plataforma tem como objetivo central instrumentalizar publicitários, jornalistas, escritores, poetas, roteiristas, redatores, estudantes etc. (público composto pelas profissões e atividades ditas “criativas”) no momento em que estão produzindo suas criações.

À maneira dos tradicionais Dicionários Analógicos, como o Dictionnaire idéologique: recueil des mots, des phrases, des idiotismes et des proverbes de la langue française classés selon l’ordre des idées (ROBERTSON, 1859), nossa proposta é nos aprofundarmos I) nos problemas de natureza ontológica, ou seja, na organização em função de conceitos lexicalizados (Crystal, 1997) e na representação de estruturas conceituais que sirvam de ancoragem para a estruturação do léxico (Jackendoff, 1990; Pustejovsky, 1995; Miller & Fellbaum, 1991); II) nas relações que se instituem entre estas estruturas e as expressões idiomáticas e os provérbios do português brasileiro. Em especial, neste artigo, discutiremos as relações lexicais paradigmáticas e sintagmáticas como organizadoras deste tipo de obra lexicográfica. A nossa proposta para um dicionário analógico na internet objetiva estruturar relações semânticas tanto paradigmáticas – em ausência, das quais os sinônimos são o melhor exemplo – quanto sintagmáticas, em presença. Nossa hipótese é de que as relações semânticas paradigmáticas (sinonímia, antonímia, hiperonímia, etc.) são melhor trabalhadas em estruturas como a das wordnets (estruturas em forma de rede, onde as unidades lexicais são organizadas sob a forma de synsets – do inglês, synonym set) (MILLER, 1998), e as relações sintagmáticas são adequadamente tratadas pela semântica de frames, ou seja, em estruturas como a das framenets (FILLMORE et al., 2003; BAKER, FILLMORE e LOWE, 1998).

Na seção 2 faremos uma breve apresentação do objeto Dicionário Analógico, ou Thesaurus – como é conhecido este tipo de obra lexicográfica na língua inglesa. Faremos uma exposição de sua origem, história e obras recentes, além de abordarmos a distinção clássica dentro da Lexicografia entre semasiologia e onomasiologia, uma questão importante quando tratamos de organização lexical. Na seção 3 abordaremos outros tipos de organização lexical

que são referência quando o assunto é o tratamento computacional do léxico. No caso específico, vamos explicitar algumas informações e conceitos básicos sobre os projetos da Wordnet e da Framenet. Por fim, na seção 4, vamos apresentar a nossa proposta de organizar um dicionário analógico na internet tendo em conta as relações lexicais paradigmáticas e sintagmáticas.

2. Dicionários Analógicos/ Thesaurus

Ao contrário dos *dicionários semasiológicos*, os *dicionários analógicos* partem de uma determinada ideia e examinam as diversas maneiras com as quais essa ideia encontrou expressão na palavra (unidades lexicais), sendo considerados, portanto, de ordem *onomasiológica* (BIDERMAN, [1978] 2001, CORREIA, 2009). Os *dicionários analógicos* nasceram na segunda metade do século XIX e o médico inglês Peter Mark Roget é considerado o iniciador desse tipo de repertório, com a publicação, em 1852, do seu *Thesaurus of english words and phrases, classified and arranged so as to facilitate the expression of ideas and assist in literary composition* (ROGET, 1852). Seu trabalho foi traduzido para diversas línguas e utilizado como modelo para várias obras lexicográficas do tipo. A denominação *dicionário analógico* ganhou a preferência dos países de língua latina, enquanto que nos países de língua inglesa a tendência foi a de chamar de *thesaurus*¹. No Brasil, uma das obras de referência nesse campo é o *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, [1950] 2010) do professor goiano Francisco dos Santos Azevedo. Publicado pela primeira vez em 1950, a obra do professor Azevedo acaba de ser relançada (2010), em edição revista e atualizada, pela editora Lexikon.

O percurso que permite encontrar uma unidade lexical (BORBA, 2003) tendo como ponto de partida seu conteúdo semântico (conceito) faz-se, geralmente, (1) pelo sistema conceitual ou pelo plano de classificação das ideias, normalmente apresentados no início das obras; (2) pela sinonímia; (3) pela antonímia; (4) por analogia.

A analogia pode ser entendida como um processo cognitivo que transfere informações de um domínio-alvo para um domínio-fonte. Neste sentido, o conceito de analogia se aproxima da concepção de metáfora conceitual da Linguística Cognitiva (Gentner et al., 2001).

Na visão clássica de metáforas, estas eram compreendidas como um recurso de superfície, utilizadas intencionalmente na retórica e na poética (RICOEUR, [1975] 2000). No entanto, a abordagem da Linguística Cognitiva (LAKOFF & JOHNSON, op. cit.) revolucionou o estudo das metáforas ao propor que a metáfora é conceitual, relacionada às funções de pensamento e de linguagem sendo, desta forma, inerente à cognição humana. Analisando por essa perspectiva, nosso sistema conceitual é, em grande medida, metaforicamente estruturado. Ou seja, a metáfora tem um importante papel na formação dos conceitos e da linguagem.

Por outro lado, a estruturação de conceitos pode também se dar não pela analogia, pautada na ideia de similaridade, mas pela maneira como as coisas estão dispostas, lado a lado, no mundo, ou seja, pelo princípio da contiguidade. Segundo Jakobson, são específicas da contiguidade as organizações mentais decorrentes da causalidade, que cumprem uma lógica linear, de causa e efeito, ou seja, associações estruturadas por subordinação, onde os signos mantêm entre si relações de proximidade (JAKOBSON, 1973: 34-62).

¹ “dicionário organizado em função de conceitos lexicalizados” (CRYSTAL, 1997, p. 158 *apud* Dias-da-Silva et al., 2000, p. 3-6).

3. Wordnet e Framenet

O termo Wordnet começou designando a base de dados de conhecimento lexical (MARRAFA, 2001) desenvolvida na universidade de Princeton (WN.Pr) por um grupo liderado por George Miller, sob a justificativa teórica de experimentos psicolinguísticos realizados nas décadas de 60 e 70 (MILLER, 1998). Atualmente, fala-se em Wordnets, no plural, que são as equivalentes da WN.Pr para as diversas línguas ao redor do mundo, como a Wordnet.Br (WN.Br) (DIAS-DA-SILVA et al., 2002), em construção para o Português Brasileiro (PB), por exemplo. A WN.Pr foi construída, inicialmente, com base no *Brown Corpus* (Standard *Corpus* of Presente Day Edited English) (ver FRANCIS, 1965) e no *Thesaurus* do Roget. No entanto, ao longo do tempo, outras fontes foram incorporadas. A WN.Pr é estruturada como uma rede, em que as unidades lexicais, que correspondem aos lemas nos dicionários ou aos lexemas na semântica lexical (FELLBAUM, 1998), são organizadas sob a forma de *synsets* (do inglês, *synonym set*) e constituem os seus **nós**; as relações semântico-conceituais que se estabelecem entre os *synsets* constituem os seus **arcos**.

A Framenet (FILLMORE et al., 2003; BAKER, FILLMORE e LOWE, 1998) é um projeto de lexicografia computacional baseado na teoria da semântica de frames (*frame semantics*). De acordo com FILLMORE (1982), um *frame* é qualquer sistema de conceitos relacionados de forma que, para se compreender qualquer conceito, é necessário compreender toda a estrutura em que ele se insere. Um *frame semântico* constitui, portanto, uma representação esquemática que formaliza o resultado das relações sintáticas e semânticas de uma unidade lexical que representa uma situação envolvendo vários participantes, propriedades e outros papéis conceituais que constituem cada elemento do *frame*. Através de processos de anotação semi-automática de sentenças, o projeto extrai informações semântico-sintáticas de unidades lexicais retiradas de extensos *corpora* eletrônicos. O projeto FrameNet de Berkeley² possui atualmente mais de 10.000 unidades lexicais, das quais mais de 6.000 estão exaustivamente anotadas com informações semântico-sintáticas e arroladas em aproximadamente 800 frames. No Brasil, há um projeto para a construção de uma FrameNet para o PB em fase inicial de desenvolvimento, sob coordenação da Profa. Dra. Maria Margarida Martins Salomão (SALOMÃO, 2009).

4. Relações paradigmáticas e sintagmáticas

Como esboçamos anteriormente, a hipótese modernamente aceita é a de que o sistema conceptual humano é, por um lado, construído pelos princípios da analogia, da similaridade, da metáfora e, por outro, por princípios metonímicos, de contigüidade (JACKOBSON, 1973; LAKOFF & JOHNSON, 1980; 2003; CROFT & CRUSE, 2004; BARCELONA, 2003; DIRVEN e PÖRINGS, 2003).

Como exemplo, extraído da história do cinema, pode-se notar a orientação manifestamente metonímica das produções de D. W. Griffith, que apresentam uma gama espetacular de grandes planos sinedóquicos e de montagens metonímicas em geral. Em contraposição, o princípio da montagem eisensteiniana – na qual a montagem deixa de ser a simples ligação de planos à mercê da narrativa e transforma-se em “verdadeiras comparações fílmicas” (JACKOBSON, op. cit.: 57) – apresenta um novo tipo de montagem, de intenções metafóricas. A junção das tomadas, no cinema de Eisenstein, não está motivada pela causalidade, mas pelo conflito entre similaridades internas aos planos e dissimilaridades contextuais.

² Disponível em <http://framenet.icsi.berkeley.edu>, acessado em 01 ago 2010.

Na associação por similaridade, instituem-se as analogias, relações entre signos de domínios ou contextos distantes, mas que mantêm entre si alguma semelhança conceitual. Se fizermos um paralelismo com as ideias saussurianas (SAUSSURE, 1969), contiguidade e similaridade passam, respectivamente, a sugerir os dois eixos da linguagem: sintagma (subordinação/cominação) e paradigma (justaposição/seleção). A nossa proposta entende que um *dicionário analógico* deve estruturar relações semânticas tanto **paradigmáticas** quanto **sintagmáticas**. Nossa hipótese é de que as relações semânticas paradigmáticas são melhor trabalhadas em estruturas como a das *wordnets*, e as relações sintagmáticas são adequadamente tratadas pela semântica de frames, portanto, estruturas como a das *framenets*. Nossa proposta se insere nesse contexto de pesquisas e se nutre dele.

Os princípios que organizam uma Wordnet são essencialmente paradigmáticos: sinônimos, antônimos, hipônimos e hiperônimos. Os conjuntos de sinônimos (synsets) são a base organizacional de qualquer rede Wordnet. Dentre as potenciais unidades lexicais sinônimas, o falante deverá escolher apenas uma para incorporar à sua fala ou para utilizar no seu texto. Na sentença (1), por exemplo, a unidade lexical **pé** poderia ser substituída pela unidade lexical **base** sem que o sentido da frase fosse alterado. Qualquer aplicação diferente desta, ou é um recurso estilístico de redundância, ou uma sentença metalinguística, como em (2).

(1) Ao **pé** desse monte realizavam-se as competições esportivas entre eles, as Olimpíadas³.

(2) Os **pés** são a **base** que **sustenta** o corpo humano.

Ao contrário, as *framenets* trabalham com o princípio organizacional da semântica de frames. Como dissemos na seção anterior, *frame* é uma representação esquemática de uma situação envolvendo vários participantes, os chamados elementos do *frame* (*frame elements*). Os elementos do *frame* são os papéis semânticos dos participantes de um evento, sendo específicos a cada frame e por isso também chamados de papéis situacionais. Desta forma, dizemos que o conhecimento de um frame é um conhecimento “experencial”, na medida em que o entendimento de um frame passa pelo entendimento completo do evento que esse *frame* representa (Fillmore, 1977).

Por exemplo, a *Framenet* de Berkeley possui um frame chamado *Observable_body_parts* que possui como elementos centrais (core) os FEs (*frame elements*) denominados *Body_part* e *Possessor* e, dentre outros elementos não centrais (non-core), o FE chamado *Orientalional_location*. Na sentença (3), **pé** é a unidade lexical instanciada em PB que preenche o FE *Body_part* (parte do corpo) da mesma forma que Marcelo é o participante do frame que corresponde ao FE *Possessor* (possuidor). Na sentença (4), a unidade lexical **direito** instancia no PB uma das duas possibilidades que pode preencher a FE *Orientalional_location*.

(3) **Marcelo** se feriu no **pé** ao ajudar nas buscas de corpos e começou a ter convulsões.

(4) Acho que torci o **pé direito**, contou.

Segundo Fillmore e Petruck (2003), o *frame* é uma categoria conceitual e os elementos do frame são instanciados na sentença através das unidades lexicais. Uma unidade lexical é o cruzamento entre um lema e um frame, ou seja, é uma palavra considerada segundo um frame específico. Nessa perspectiva, as palavras na *Framenet* não são polissêmicas, pois cada novo sentido está associado a um *frame* diferente, caracterizando uma nova unidade lexical.

³ Sentenças extraídas do corpus do Nilc:
<http://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>

Enquanto a Wordnet desfaz a polissemia principalmente por meio do conjunto de sinônimos e, portanto, por meio de relações paradigmáticas, a Framenet resolve a polissemia ao preencher o contexto da unidade lexical com conceitos sintagmaticamente relacionados.

Na Wordnet, **piano** significa o instrumento musical de teclas pretas e brancas, e não um adjetivo para leveza e suavidade, quando aparece no *synset* {piano, pianoforte, forte-piano}. Diferentemente, na Framenet, **piano** significará o mesmo instrumento musical se outros elementos de contexto estiverem associados a ele para nos induzir a esta interpretação. Elementos como existir um executor (pianista), existir um fabricante (Steinway), um produto (sonata), estar associado a um evento (concerto) etc.

Se pensarmos pelo viés do sistema conceptual humano, podemos dizer que este modelo nos ajuda a enxergar que quando pensamos em conceitos do mundo físico ativamos outros conceitos relacionados metonimicamente, em termos de *frames*, no sentido de “fotografias do mundo”. Por exemplo, quando pensamos na ideia de ‘pé’ todo um grupo de conceitos metonimicamente relacionados são ativados: pé, cabeça, mão, perna etc. Por outro lado, se pensarmos em ‘pé’ metaforicamente, selecionando apenas seus traços [extremidade] + [inferior], e propositadamente ignorando o traço [corpo humano], todo um grupo de conceitos psíquicos que compartilham desses traços são simultaneamente ativados: base, fundamento, apoio, alicerce etc. Da mesma forma, a ideia de ‘teatro’ (enquanto ‘construção’) é um todo constituído de partes menores, partes metonimicamente ativadas quando se pensa em uma delas: platéia, palco, coxia etc. Ao passo que a ideia de ‘teatro’ enquanto ‘arte’ e/ou ‘técnica’ ativa paradigmaticamente todo um conjunto de atividades que possuem esses traços: teatro, dança, cinema, música etc.

5. Conclusão

Hoje, existem vários recursos lexicais disponíveis na Web. Não sofremos com carência de conteúdo, mas com seu excesso e sobreposição em ambientes e plataformas diversas. O desafio hoje é organizar esse conteúdo de maneira que o usuário possa intuitivamente alcançar o mais rápido possível a informação que deseja. Com este artigo defendemos que, dentre outros tipos de abordagem, a estruturação que aproxima as relações paradigmáticas e sintagmáticas devem contribuir para esta tarefa. Aproximando as abordagens das wordnets e das framenets propomos um modelo de dicionário analógico onde as relações metafóricas e metonímicas entre os conceitos estejam adjacentes.

Referências

- AZEVEDO, F. F. S. *Dicionário analógico da língua portuguesa*. Brasília: Thesaurus, 1983.
- BIDERMAN, M. T. C. *Teoria Lingüística*. São Paulo: Martins Fontes, 1978 (2001).
- BOISSIERE, P. *Dictionnaire analogique de la langue française: répertoire complet des mots par les idées, des idées par les mots*. Paris: Aug. Boyer, 1862.
- BORBA, F.S. *Organização de dicionários: uma introdução à lexicografia*. São Paulo: Unesp, 2003.
- CRUSE, D. A. *Lexical semantics*. New York: Cambridge University Press, 1986.
- DESCLÉS, J.P. et al. Sémantique Cognitive de l'action: contexte théorique, *Langages*, 132, p. 28-47, 1998.

- DI SCIULLO, A. M. & WILLIAMS, E. *On the Definition of Word*. Cambridge: MIT Press, 1987.
- DIAS-DA-SILVA, B. C. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, 1996, 272 f. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, UNESP, Araraquara, 1996.
- _____. Os domínios lingüístico e tecnológico do estudo do processamento automático das línguas naturais. *Estudos Lingüísticos*, v. 26, p. 612-7, 1998.
- DIAS-DA-SILVA, B. C. et al. *A construção de um thesaurus eletrônico para o português do Brasil*. In: ENCONTRO PARA O PROCESSAMENTO COMPUTACIONAL DA LÍNGUA PORTUGUESA ESCRITA E FALADA, 5., 2000, São Carlos. Anais... São Carlos: ICMSC, 2000. p. 1-10.
- DUBOIS, J. & DUBOIS, C. *Introduction à la lexicographie: le dictionnaire*. Paris: Larousse, 1971.
- FELLBAUM, C. (Ed.) *WordNet: an electronic lexical database*. Cambridge, Mass.: MIT Press, 1998.
- FILLMORE, C. J., The case for case, In E. BACH & R. HARMS (Eds). *Universals in linguistic theory*, p 1-90. New York: Holt, Rinehart and Winston, 1968.
- GIBBS, R. W., Jr. *The poetics of mind: figurative thought, language and understanding*. Cambridge: University Press, 1994.
- JACKENDOFF, R. *Semantics and cognition*. Cambridge, Mass.: The MIT Press, 1983.
- _____. *Semantics structures*. Cambridge, Mass.: The MIT Press, 1990.
- JAKOBSON, R. *Lingüística e comunicação*. São Paulo: Cultrix, 1973.
- JOHNSON, M. *The Body in the Mind: The bodily basis of meaning, imagination and reason*. Chicago: The University of Chicago Press, 1987.
- LAKOFF, G. *Women, fire and dangerous things*. Chicago: University of Chicago Press, 1987.
- LAKOFF, G.; JOHNSON, M. *Metaphors We Live By*. Chicago: The Un. of Chicago Press, 1980.
- LYONS, J. *Semantics*, v. 1 e 2, Cambridge: Cambridge University Press, 1977.
- MARRAFA, P. *WordNet do Português: uma base de dados de conhecimento lingüístico*. Lisboa: Instituto Camões, 2001.
- MILLER, G. A., FELLBAUM, C. Semantic networks of English. *Cognition*, 41, n. 1-3, p. 197-229, 1991.
- ROBERTSON, T. *Dictionnaire idéologique: recueil des mots, des phrases, des idiotismes et des proverbes de la langue française classés selon l'ordre des idées*. Paris: A. Derache, 1859.
- ROGET, P. M. *Thesaurus of English Words and Phrases: classified and arranged so as to facilitate the Expression of Ideas and assist in Literary Composition*. London: Longmans, 1852.
- ROSCH, E.; LLOYD, B. B. *Cognition and Categorization*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978.
- SALOMÃO, M. M. M. *FrameNet Brasil: Um trabalho em progresso*. In *Calidoscópio*, Vol. 7.3, 2009.
- SAUSSURE, F. *Curso de lingüística geral*. São Paulo: Cultrix; USP, 1969.