

PROPOSTA PARA A CONSTRUÇÃO DE UM MODELO DE DICIONÁRIO ANALÓGICO NA INTERNET

Felipe Iszlaji de ALBUQUERQUE¹

RESUMO: Este artigo dá o ponta-pé inicial para uma série de artigos que deverão elucidar o processo de construção do modelo de *dicionário analógico* na Internet que estamos desenvolvendo, em nível de doutorado, sob orientação do prof. Dr. Bento Carlos Dias da Silva. Para tanto, faremos primeiramente uma apresentação deste tipo de obra lexicográfica e apresentaremos, resumidamente, sua origem e história. Abordaremos também outros tipos de organização lexical que são referência quando o assunto é o tratamento computacional do léxico. Por fim, apresentaremos a nossa proposta para um *dicionário analógico* na Internet.

Palavras-chave: dicionário analógico; thesaurus; wordnet; framenet

ABSTRACT: It is the article that gives the kick-off for a series of articles that should elucidate the construction process of the model of *thesaurus* on the internet that is being developed, on the doctoral level, under the supervision of Bento Carlos Dias da Silva, PhD. To this end, we first present this type of lexicographic work and present, in short, its origin and history. We will also include other types of lexical organizations that are reference when the subject is the treatment computational of the lexicon. Finally, we will present our proposal for a *thesaurus* on the Internet.

Keywords: thesaurus; wordnet; framenet

1. O que é um *dicionário analógico*?

Ao contrário dos chamados dicionários padrões de língua, ditos *semasiológicos* (p. ex., Aurélio, 2009; Michaelis, 1998; Houaiss, 2009), que partem da palavra para chegar aos seus diversos significados, os *dicionários analógicos* partem de uma determinada ideia e examinam as muitas maneiras pelas quais essa ideia encontrou expressão na palavra (unidades lexicais; unidade mínima do léxico mental, cf. Biderman, 1999), sendo considerados, portanto, de ordem *onomasiológica* (Biderman, [1978] 2001).

No âmbito das obras lexicográficas, o adjetivo *analógico* é frequentemente empregado como sinônimo de *onomasiológico*. 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). Esse trabalho foi traduzido para diversas

¹ Doutorando do programa de pós-graduação de Linguística e Língua Portuguesa da UNESP de Araraquara. Endereço eletrônico: felipealbuquerque2006@gmail.com.

línguas e utilizado como modelo para várias obras lexicográficas. A denominação *dicionário analógico* ganhou a preferência dos países de língua latina, enquanto que, nos países de língua inglesa, a tendência foi a adoção da denominação *thesaurus*. No entanto, é preciso estar cauteloso com os diferentes objetos que o termo *thesaurus* pode denominar atualmente. Por meio de um levantamento realizado por Dias-da-Silva et al. (2000, p. 3-6), foi possível reconhecer seis diferentes tipos de objetos denominados *thesaurus*, separados e ordenados conforme sua natureza, conforme mostra a Fig. 1.

Objetos	Natureza
THESAURUS 1	Um tipo de inventário exaustivo que procura registrar o vocabulário de uma determinada língua, um “tesouro” vocabular.
THESAURUS 2	Um “dicionário organizado em função de conceitos lexicalizados” (CRYSTAL, 1997, p. 158), ou seja, um dicionário onomasiológico, cujo precursor foi Roget.
THESAURUS 3	Dicionário ancorado, de modo geral, nas relações de sinonímia e antonímia na organização dos lexemas.
THESAURUS 4	Dicionário empregado no domínio da Informática e Documentação.
THESAURUS 5	Outro emprego, também motivado pelo advento da Informática, no qual um arquivo contendo sinônimos exibe alternativas de correção durante uma verificação ortográfica.
THESAURUS 6	Um tipo específico de ferramenta de auxílio à expressão lingüística, parte integrante de vários processadores de textos.

Figura 1 – As noções de *thesaurus*

Os *dicionários analógicos* equivalem ao *Thesaurus 2*, em que o léxico se organiza em termos de conceitos semântica e ontologicamente estruturados. O *Thesaurus* de Roget teve grande sucesso de público nos países de língua inglesa e foi também “traduzido” para diversas línguas. Na adaptação da obra de Roget para o francês, Robertson (1859) dá a ela o título de *Dictionnaire idéologique: recueil des mots, des phrases, des idiotismes et des proverbes de la langue française classés selon l’ordre des idées*. Por conta deste título, nos países de língua latina e alemã, os *thesaurus* e *dicionários analógicos* são também conhecidos como *dicionários ideológicos*, ou seja, de ‘ideias’. No entanto, no Brasil, a grande maioria dos lexicógrafos e pesquisadores da área evita esta denominação pela evidente confusão que o termo *ideológico* pode provocar. No Brasil, a obra de referência desse tipo de iniciativa é o *Dicionário Analógico da Língua Portuguesa* (Azevedo, 1983), meticulosamente compilado, em vinte anos de trabalho, pelo professor goiano Francisco dos Santos Azevedo, sendo que sua última edição data do ano de 1983.

2. Como se estruturam os dicionários analógicos?

A macroestrutura dos dicionários analógicos define-se por quatro componentes características. A primeira é o chamado *plano de classificação das ideias*, que é a ossatura do dicionário e onde estão listados os conceitos principais da obra. O plano de classificação das ideias corresponde ao *sistema conceitual* dos dicionários terminológicos e tem origem nos primeiros dicionários da humanidade, seguindo uma ontologia de bases aristotélicas. A segunda componente, em que os conceitos listados no plano de classificação são desmembrados, é constituída pelos *quadros sinópticos de categorias* (ver Fig.2 e 3), que contêm subconjuntos ou subtópicos.

QUADRO SINÓPTICO DE CATEGORIAS
CLASSE I. RELAÇÕES ABSTRATAS

Seção I — EXISTÊNCIA.

1.) Abstrata.....	1. Existência	2. Inexistência
2.) Concreta.....	3. Subsistencialidade	4. Insustancialidade
3.) Formal.....	5. Intrinsecabilidade	6. Extrinsecabilidade
4.) Modal.....	7. Estado	8. Circunstância

Seção II — RELAÇÃO.

9. Relação	10. Não relação
11. Consangüinidade	12. Correlação
13. Identidade	14. Contraste
15. Diferença	16a. Desconformidade
16. Uniformidade	17. Semelhança
18. Dessemelhança	19. Imitação
20. Originalidade	20a. Variabilidade
21. Cópia (resultado da imitação)	22. Protótipo (coisa que serve de cópia)
23. Acordo	24. Desacordo

Seção III — QUANTIDADE.

	<i>Aboluta</i>	<i>Relativa</i>
1.) Simples.....	25. Quantidade	26. Grau
	27. Igualdade	28. Desigualdade
	29. Média	
	30. Compensação	
	<i>Por comparação</i>	
2.) Comparativa.....	31. Grandeza	32. Pouquidão
	<i>Por comparação com um objeto semelhante</i>	
	33. Superioridade	34. Inferioridade
	<i>Variações de quantidade</i>	
	35. Aumento	36. Diminuição
	37. Adjução	38. Subtração
	39. Adjunto (coisa acrescentada)	40. Resto (coisa restante)
3.) Subordinada.....	41. Mistura	42. Pureza
	43. Junção	44. Disjunção
	45. Vínculo	
	46. Coesão	47. Incoesão
	48. Combinação	49. Decomposição

- IX -

PRIMEIRA PARTE
CLASSE I
PALAVRAS QUE EXPRESSAM RELAÇÕES ABSTRATAS
Seção I. EXISTÊNCIA

1.) Ser — abstrato

<p>1. Existência, existibilidade, tudo que vibra e palpita, realza, realidade, positividade, atualidade, subsistência, presença, forma, silhueta, pegadas, vulto; permanência, estabilidade, conservação das coisas; vida, objetividade, palpabilidade, fato, fato consumado; realidade gélida, pura, palpável, fria, indiscutível; existência 494; concretização, materialização, corporificação; concrecibilidade, materialidade; presença 166; coexistência 120; (ciência da existência): ontologia.</p> <p>V. Existir, ser, haver, subsistir, estar, permanecer, viver 359; conservar sua força ou ação, continuar a ser, ficar em pé, perdurar, fazer-se sentir, prevalecer, persistir, desenvolver-se, crescer; medrar, pulular, florescer, reinar, imperar, dominar, aparecer 446; conegar 66; respirar, palpitar, vibrar, ter vida, ocorrer 151; vegetar, produzir 161; vigorar, vigor, vingar, estar de pé, estar em vigor; sorrir, alvejar, negrejar, correr, soprar, enguar-se; contemplar-se, admirar-se, encontrar-se, notar-se, ver-se, ostentar-se em.</p> <p>Adj. existente & V.; existencial, vital, real, positivo, efetivo, absoluto, verdadeiro, prevalente, substancial, substantivo, exato 494; cuja existência não se contesta, concrecível, inextinto, ativo, em erupção, em movimento, vivo 359;</p>	<p>2. Inexistência, não existência, negação da existência, falta de existência, ausência 187; falta, desconhecimento, nulidade, vacuidade, nulidade, insubsistência, nada, coisa nula, irrealdade, <i>fabula rasa</i>, lacuna, falha, omissão, espaço em branco, vácuo, carência; ilusão, fantasia, fantasmagoria, produto da imaginação, ficção, fábula, mito, pilhéria, farsa, burla, mentira 646; arremedo, simulacro, figura de retórica, rótulo falsificado, sonhada, fantasia, fantasmagórico, problema, nihilismo, aniquilamento, destruição 162.</p> <p>V. inexistir, não (existir & V. 1); estar no rol das coisas inexistentes, brilhar pela ausência, ainda estar por vir ao mundo, ser nulo e vazio, ser fruto da imaginação, desertar, não se conhecer, perecer, extinguir-se, acabar-se, findar, evaporar-se, dissolver-se, diluir-se, morrer 360; tornar inexistente & adj., aniquilar, extirpar, anular, nulificar, ab-rogar, extinguir, suprimir, cancelar (<i>destruir</i>) 162; fantasmagorizar.</p> <p>Adj. inexistente, irreal, insubstancial, imaterial, negativo, nenhum, nulo, omissivo, aniquilado, extirpado, extinto, exaustivo, esgotado, diluído, perdido, extraviado, desconhecido, morto 360; incriado, ingerado, inato, inefável, que não teve nascença, que Deus ainda não criou,</p>
--	---

- 1 -

Figuras 2 e 3 – Quadro Sinóptico de Categorias e os grupos de palavras; imagens extraídas do Dicionário Analógico da Língua Portuguesa (AZEVEDO, 1983, p. IX e p. 1)

A terceira componente é o dicionário propriamente dito, composto dos grandes grupos de palavras (ver Fig. 3) que se arrolam para cada um dos conceitos presentes nos quadros sinópticos. Em cada grupo de palavras são indicados em negrito o conceito central listado no quadro sinóptico, seguida por sinônimos e analogias. Cada página é dividida em duas colunas; cada uma delas apresenta grupos de palavras antagônicas (por

exemplo, EXISTÊNCIA/INEXISTÊNCIA). Cada grupo é também subdividido pelas quatro grandes categorias gramaticais, nesta ordem: substantivo, verbo, adjetivo e advérbio. Alguns conceitos possuem também, ao final, uma lista de expressões e frases correspondentes.

O critério adotado para a reunião das palavras em grupos é o princípio da *analogia*. Por analogia, entendem-se as palavras da língua ligadas uma à outra "seja por uma comunidade de ideias, seja por relações de emprego frequentes de causa, de meio, de efeito, etc." (Boissiere, 1862). No *Dicionário Analógico* do professor Azevedo, o autor da apresentação, Bernardo Élis, escreve sobre o alcance almejado para cada grupo de palavras: "não se arrolam apenas sinônimos, mas a imensa gama de palavras, termos, vocábulos ou expressões que se inscrevem nessa ampla e meio nebulosa área do campo semântico".

A quarta componente dos dicionários analógicos contém uma lista, em ordem alfabética, de todas os conceitos presentes no dicionário. Essa componente estrutural do dicionário varia de obra para obra. Em alguns dicionários ela se caracteriza como um *índice remissivo*; em outros, além dessas informações, nela são apresentadas definições das diferentes acepções que a unidade lexical em questão possui na língua.

Nos *dicionários analógicos*, o principal ponto de partida das consultas de tipo onomasiológico são as primeira e segunda componentes. Como dissemos, o princípio de funcionamento deste tipo de obra é o de encontrar palavras, partindo dos conceitos por elas designados, seguindo uma ordenação de sentido que vai do mais geral ao mais específico. Em seguida, com base na analogia, o percurso pode assumir características diferentes, permitindo a passagem de um grupo de palavras a outro, indicadas ao consulente por *remissivas*.

3. Para que serve um *dicionário analógico*?

Baldinger (1966) diz que, enquanto o percurso semasiológico adota a perspectiva do interpretante, a estrutura onomasiológica veste a perspectiva daquele que fala, "daquele que deve escolher entre diferentes meios de expressão" (p. 30). Nessa mesma orientação, com relação à função dos dicionários, Correia (2009) nos oferece uma distinção entre *dicionários de compreensão* e *dicionários de produção*. Segundo a autora, os dicionários de compreensão "privilegiam a função de descodificação", "realizando um percurso semasiológico". Já os dicionários de produção, "incorporam informação

gramatical e sobre contextos de uso das palavras, sobre combinatórias, sinônimos e antônimos, remissões para palavras morfológicas e semanticamente relacionadas” (p. 44). Este tipo de obra é também chamada, por vezes, de dicionários ativos ou *codificadores*, na medida em que assumem a perspectiva de quem codifica o enunciado.

O subtítulo do *Thesaurus* de Roget expressa bem o que o autor pretendia com sua obra que, em grande medida, é também a finalidade de todas as obras lexicográficas deste tipo: *classificado e ordenado de modo a facilitar a expressão de ideias e auxiliar na composição literária*². Roget expõe a questão da seguinte forma: “a revisão de um catálogo de palavras de significado análogo vai sugerir, com frequência, por associação, outras sucessões de pensamento. A apresentação dos assuntos sob aspectos novos e variados pode expandir grandemente a esfera de nossa visão mental” (Roget, 1925, p. XVIII).

Fica evidente, portanto, que os *dicionários analógicos* são de grande utilidade para aqueles que querem produzir um texto consistente, bem elaborado, seja com pretensões retóricas ou poéticas. O *Dicionário Analógico* do professor Azevedo, por exemplo, é livro de cabeceira de Chico Buarque, um dos maiores letristas da língua portuguesa. Segundo seu biógrafo, Chico recebeu o exemplar das mãos do pai, Sérgio Buarque de Hollanda, e cuidou de tê-lo em duplicata, reservando um para o seu apartamento parisiense e outro para a sua residência no Rio de Janeiro (Hollanda, 2006).

4. Dicionários analógicos e outras bases de conhecimento lexical na internet

Na Fig. 3, apresenta-se uma lista de *thesaurus* e *dicionários analógicos* disponíveis na internet para os idiomas *inglês, francês e português brasileiro (PB)*. Foram identificados por meio de extensa pesquisa e selecionados, dentre outros existentes, por critérios de representatividade. Entre eles, há grande heterogeneidade com relação à forma de produção, critérios, layout, abrangência etc. Não especificaremos aqui cada uma destas características, por não ser esse o enfoque do artigo.

² “[...] *classified and arranged so as to facilitate the expression of ideas and to assist in literary composition* [...]” (ROGET, 1925)

Idioma	Nome/Endereço	Observações
ING	The Free Dictionary www.thefreedictionary.com	<i>Thesaurus</i> , dicionário, enciclopédia e outras ferramentas lingüísticas em língua inglesa; disponível gratuitamente.
ING	Thesaurus Reference www.thesaurus.reference.com	<i>Thesaurus</i> digital adaptado diretamente da obra de Roget; disponível gratuitamente.
ING	SynonymFor www.synonymfor.com	Interface para a base de dados da WordNet de Princeton; disponível gratuitamente.
ING	Visual Thesaurus www.visualthesaurus.com	<i>Thesaurus</i> visual que se utiliza da base de dados da WordNet de Princeton; disponível para testar.
ING	Snappy Words www.snappywords.com	<i>Thesaurus</i> visual que também se utiliza da base de dados da WordNet de Princeton; disponível gratuitamente.
FRA	Dictionnaire des synonymes du CRISCO www.crisco.unicaen.fr	<i>Dicionário de Sinônimos</i> com características de <i>dicionário analógico</i> ; desenvolvido pela Universidade de Caen; disponível gratuitamente.
FRA	Dictionnaire Sensagent www.dictionnaire.sensagent.com	<i>Dicionário Analógico</i> estruturado em rede semântica; multi-idíomas; disponível gratuitamente.
PB	Thesaurus da Língua Portuguesa www.alcor.concordia.ca/~vjorge/Thesaurus/	<i>Dicionário Analógico</i> ; representativo apenas por ser o único da língua portuguesa do Brasil disponível na internet; não é atualizado desde 2003.

Figura 4 – Lista de *thesaurus* e dicionários analógicos disponíveis na internet, acessados em: 01 mar 2010.

O *Thesaurus Reference* e o *SynonymFor* são, respectivamente, uma adaptação do *Thesaurus* de Roget e uma interface para a base de dados da WordNet de Princeton (WN.Pr)³. Um tipo diferente de apresentação de um *thesaurus* é o representado pelos *visual thesauri*, que inovam na forma de apresentação visual, organizando uma rede semântica preconcebida, computacionalmente tratável (p. ex., a WN.Pr), por meio de cores e de diferentes tipos gráficos de linhas e nós. O *Visual Thesaurus* e o *Snappy Words* são dois exemplos desse tipo; eles organizam visualmente a rede semântica estruturada na base de dados da WN.Pr.

O termo WordNet começou designando a base de dados de conhecimento lexical (Marrafa, 2001) desenvolvida na universidade de Princeton por um grupo liderado por George Miller, sob a justificativa teórica de experimentos psicolingüísticos realizados

³ Disponível em: <www.wordnet.princeton.edu/> Acesso em: 01 mar 2010.

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 PB, por exemplo.

Apesar de existirem muitos pontos de contato, as WordNets não são simplesmente *thesauri* digitais, uma vez que a sua arquitetura e a sua natureza computacional é mais complexa que a do *dicionário analógico* e possibilita um grande número de aplicações em tecnologias linguísticas. 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*. Podemos dizer que o *synset* é um conjunto de unidades lexicais de uma mesma categoria sintática que representam um mesmo conceito em um determinado contexto, p.ex.: {*actor, histrion, player, thespian, role player*} que ocupa o espaço conceitual definido informalmente pela glosa “*a theatrical performer*”. Todo *synset* é construído de forma a codificar apenas um único conceito. Perceba que, para fazer isso, os *synsets* não podem ser muito numerosos; não podem conter o grande número de itens que são listados nos grupos de palavras dos *thesauri* ou *dicionários analógicos*. Se procurarmos pela unidade lexical ‘*actor*’ no *Thesaurus* de Roget, teremos o numeroso grupo de palavras {*amateur, artist, barnstormer, bit player, character, clown, comedian, entertainer, extra, foil, ham, hambone, headliner, idol, impersonator, ingénue, lead, mime, mimic, pantomimist, performer, play-actor, player, soubrette, stand-in, star, stooge, straight person, thesp, thespian, trouper, understudy, ventriloquist, villain, walk-on*}. Na verdade, observamos que, no *Thesaurus*, o campo conceitual que aglutina as palavras em um mesmo grupo é bastante genérico. No exemplo, poderia ser definido pela glosa “*person who performs*”. Isto porque os *thesauri* e *dicionários analógicos* têm essa função de aproximar conceitos gerais por analogias. Cumpre esclarecer, assim, que os *synsets* das WordNets representam

conceitos bem mais específicos, recortando os campos conceituais com um grau de refinamento muito além daquele permitido pelos agrupamentos por analogia.

Outra base de conhecimento lexical de interesse para esta discussão é a FrameNet. A FrameNet (Baker, Fillmore e Lowe, 1998; Fillmore et al., 2003) é um projeto de lexicografia computacional baseado na teoria da *semântica de frames*. 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 extraídas de extensos *corpora* eletrônicos. O projeto FrameNet de Berkeley⁴ 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 de construção de uma FrameNet para o PB em fase inicial de desenvolvimento, sob coordenação da Profa. Dra. Maria Margarida Martins Salomão (Salomão, 2009).

5. Proposta para um *dicionário analógico da língua portuguesa na internet*

Antes de tudo, é preciso dizer que nossa proposta para um *dicionário analógico* na internet se insere no contexto da construção de uma plataforma digital de grande magnitude – provisoriamente denominada *Brainstorm*, termo em inglês que pode ser livremente traduzido como “tempestade de ideias” –, idealizada com o objetivo de propiciar aos seus usuários uma série de ferramentas linguísticas. Essas ferramentas irão oferecer relações linguísticas de natureza léxico-semântica, como as relações de sinonímia e antonímia; de natureza lógico-conceitual, como as relações de hiponímia/hiperonímia, meronímia/holonímia (relações entre a parte e o todo), troponímia (isto é, a hiponímia verbal), acarretamento e causa (cf. Cruse, 1986; Fellbaum, 1998; Lyons, 1977); de natureza fonética (rimas); e, ainda, relações que se estabelecem entre as unidades lexicais (e seus conceitos) e as expressões idiomáticas e os provérbios do PB. Há ainda um espaço

⁴ Disponível em <http://framenet.icsi.berkeley.edu>, acessado em: 01 mar 2010.

reservado para fotografias e imagens em geral, em parceria com grandes bancos de imagem (p. ex., *Getty Image*⁵), que estarão associadas ao conceito buscado. Tal plataforma tem como finalidade instrumentalizar publicitários, jornalistas, escritores, poetas, roteiristas, redatores etc. (público composto pelas profissões e atividades ditas “criativas”) no momento da produção das suas criações. O objetivo prioritário da plataforma não é sanar as dúvidas do consultante com relação ao significado das palavras, como o fazem os dicionários tradicionais da língua portuguesa (p. ex., Aurélio, 2009; Michaelis, 1998; Houaiss, 2009). Apesar de estar previsto um espaço dentro da plataforma para a veiculação também deste tipo de obra lexicográfica, não é esta a sua principal função. A plataforma, como dissemos, deverá ser, portanto, mais robusta: disponibilizar um número significativo de ferramentas lingüísticas a partir da entrada de determinada palavra-chave e, assim, oferecer ao usuário um leque de relações possíveis entre unidades lexicais e conceitos, de forma a instrumentalizá-lo no momento de sua criação textual.

Na proposta aqui apresentada, não pretendemos dar conta de todo o conhecimento lingüístico apontado acima, mas apenas parte dele – exatamente a parte que deve corresponder a um tipo de *dicionário analógico*. O modelo de *dicionário analógico* que estamos desenvolvendo em nível de doutorado deverá se tornar o coração da plataforma. A proposta de investigação neste projeto é, portanto, aprofundarmos o estudo i) dos problemas de natureza ontológica, ou seja, da organização do léxico em função de conceitos lexicalizados (Crystal, 1997) e da representação de estruturas conceituais que sirvam de ancoragem para a estruturação deste léxico, que compreendam relações de natureza léxico-semântica e lógico-conceitual (Jackendoff, 1990; Pustejovsky, 1995; Miller & Fellbaum, 1991); ii) das relações que se instituem entre estas estruturas e as expressões idiomáticas (EIs) e os provérbios do PB.

5.1. *Quadro teórico-metodológico*

Há uma tendência em destacar que o princípio que rege a categorização é a analogia⁶ ou raciocínio analógico. Por outro lado, a estruturação de conceitos pode também se dar não pela analogia, pautada na ideia de similaridade, mas pela maneira

⁵ Disponível em www.gettyimages.com.br, acessado em: 01 mar 2010.

⁶ Trabalhamos neste artigo com o conceito de analogia como um processo cognitivo que transfere informações de um domínio-alvo para um domínio-fonte. Neste sentido, o conceito de analogia se aproxima bastante da concepção de metáfora conceitual da Linguística Cognitiva (GENTNER *et al.*, 2001)

como as coisas estão dispostas, lado a lado, no mundo, ou seja, pelo princípio da contiguidade. Segundo Jakobson (1973, 34-62), 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. Já 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).

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 sinédóquicos 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” (Jakobson, op. cit., p. 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.

A nossa proposta entende que um *dicionário analógico* deve 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*, e as relações sintagmáticas são adequadamente tratadas pela semântica de frames, ou seja, estruturas como a das *framenets*. Nossa proposta se insere nesse contexto de pesquisas e se nutre nele, cf. Fig. 5. Em artigos futuros explicitaremos exata e adequadamente como essa relação se dará dentro do *Brainstorm*.

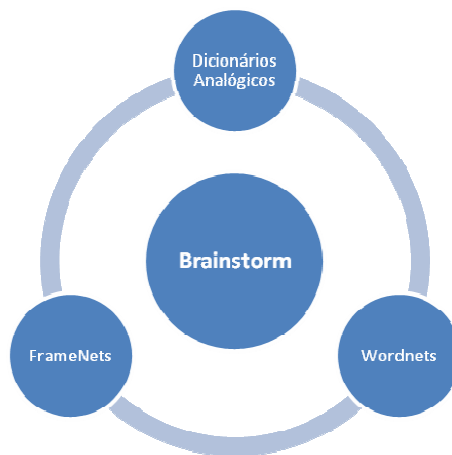


Figura 5 – Contexto de projetos léxico-computacionais onde o *Brainstorm* está inserido.

Ao transferir conceitos de um domínio semântico para outro, pode-se produzir a metáfora. Na visão clássica de metáforas, estas eram compreendidas como um recurso de superfície, utilizadas intencionalmente na retórica e na poética. No entanto, a abordagem da linguística cognitiva (Lakoff & Johnson, 1980) revolucionou o estudo das metáforas, ao propor que a metáfora é conceitual, relacionada às funções do pensamento e da linguagem sendo, portanto, parte da cognição humana. Além disso, nessa perspectiva, o sistema conceitual é, em grande parte, metaforicamente estruturado, ou seja, a metáfora tem uma importante função na formação dos conceitos e da linguagem. Como esboçamos acima, a hipótese modernamente aceita é a de que o sistema conceptual é, por um lado, construído pelos princípios da analogia, da similaridade, da metáfora e, por outro, por princípios metonímicos, de contigüidade (Jakobson, 1973; Lakoff & Johnson, 1980; 2003; Croft & Cruse, 2004; Barcelona, 2003; Dirven & Pörings, 2003).

Utilizaremos como objeto de estudo deste trabalho três diferentes campos conceituais (Trier *apud* Lyons, 1977), de forma a abranger conceitos distantes e, portanto, ilustrar os diferentes problemas que possam surgir na construção do nosso modelo lingüístico. Trabalhando com os campos conceituais do AMOR, da ARTE e do CORPO HUMANO⁷, pretendemos dar conta de entidades de 1ª e 2ª ordem, conforme caracterizadas em Lyons (1977). Segundo o autor, as entidades de 1ª ordem denotam entidades concretas – conceitos do tipo “objeto concreto discreto” – que são canonicamente expressas por nomes concretos. Conceitos deste tipo categorizam referentes perceptíveis pelos sentidos,

⁷ Não pretendemos dar conta de modelar os 3 campos conceituais de forma exaustiva, mas apenas de maneira a suscitar os principais desafios teóricos e técnicos para a construção do modelo que, desta forma, deverá servir como fundamento para futuros trabalhos de expansão, complementação e implementação.

localizados no tempo e no espaço, e que são contáveis e indivisíveis. Por outro lado, as entidades de 2ª ordem denotam atos, eventos, estados relacionados a seres, coisas ou a estados de coisas e são canonicamente expressas por verbos, nomes e adjetivos. A necessidade de trabalharmos com mais de um campo conceitual se faz, igualmente, com vistas a suscitar e explicitar os princípios da analogia, como nas locuções *boca de cena*, *pé de valsa* e *menina dos olhos*, por exemplo.

Ao longo do trabalho, os conceitos dos domínios selecionados serão estruturados em categorias, subcategorias e, por fim, representados, quando possível, em unidades lexicais. Em seguida, faremos a distribuição dos conceitos lexicalizados nas classes gramaticais (*substantivo*, *verbo*, *adjetivo* e *advérbio*) ou, no caso das expressões, especificaremos as distinções entre *locuções*, *expressões idiomáticas* e *ditos e provérbios populares*.

Como critério de relevância para a disposição dos conceitos linguisticamente expressos, utilizaremos a teoria dos protótipos e como critério de relevância para a ordenação das unidades lexicais, faremos uso de estatísticas de frequência em *corpus*. Ao considerar as categorias como estruturas instáveis e não claramente definidas, a *Teoria dos Protótipos* (TP) as concebe como estruturas de atributos (traços) graduáveis e com limites difusos (*Fuzzy categories*; Taylor, 1989). Os graus de relevância dos atributos dentro das categorias correspondem ao fato de que estas se organizam em torno de um centro cognitivo exemplar, dentro do qual as entidades são ordenadas e também incluídas ou excluídas pelos falantes, produzindo o que se conhece como efeitos de protótipo [*Prototype effects*] (Lakoff, 1987; Taylor, op. cit.). A concepção da estrutura prototípica das categorias linguísticas tem fundamentos na Psicologia Cognitiva (Rosch, 1978) e na filosofia de Wittgenstein (1953).

Nossa proposta é, portanto, a elaboração de um modelo lingüístico-computacional para um *dicionário analógico*, com critérios de relevância – baseado na TP (Rosch, op. cit.; Lakoff, op. cit.; Taylor, op. cit.) e na *Linguística de Corpus* (Beber Sardinha, 2004) –, em plataforma digital, disponível na *internet*, utilizando-se, para tanto, não só de conhecimentos da *Lexicologia* (Biderman, op. cit.) e da *Lexicografia* (Borba, 2003; Hartmann, 1983; Atkins & Rundell, 2008), mas também, das concepções da *Semântica Cognitiva* (Lakoff & Johnson, 1980; 2003; Lakoff, 1987; 1993; Croft & Cruse, 2004) a respeito dos problemas de *categorização* e *conceptualização*. Por fim, é necessário produzir o alinhamento, ou seja, a equivalência lingüístico-computacional com a plataforma do

Brainstorm que o *dicionário analógico* irá integrar, de forma a apontar o seu posicionamento e a sua importância dentro deste contexto maior.

REFERÊNCIAS

- ATKINS, B. T. S.; RUNDELL, M. **The Oxford Guide to Practical Lexicography**. Oxford: Oxford University Press, 2008.
- AZEVEDO, F. F. S. **Dicionário analógico da língua portuguesa: idéias afins**. Brasília: Thesaurus, 1983.
- BALDINGER, K. Semasiologia e onomasiologia. Tradução de Ataliba T. de Castilho. **Alfa**. São Paulo, v. 9, 1966, p. 7-36.
- BARCELONA, A. **Metaphor and metonymy at the crossroads: a cognitive perspective**. Berlin, New York: Mouton de Gruyter, 2003.
- BEBER SARDINHA, A. P. **Linguística de Corpus**. São Paulo: Manole, 2004.
- BIDERMAN, M. T. C. **Teoria Lingüística**. São Paulo: Martins Fontes, 2001 [1978].
- _____. Conceito lingüístico de palavra. In: **Revista Palavra**. Petrópolis: Vozes, 1999.
- BOISSIERE, P. **Dictionnaire analogique de la langue française: répertoire complet des mots par les idées, des idées par les mots**. Paris : Aug. Boyer, 1862.
- CORREIA, M. **Os Dicionários Portugueses**. Lisboa: Caminho, 2009.
- CROFT, W.; CRUSE, D.A. **Cognitive linguistics**. Cambridge University Press, 2004.
- CRUSE, A. **Lexical semantics**. Cambridge: University Press, 1986.
- CRYSTAL, D. **The Cambridge encyclopedia of the English language**. Cambridge: Cambridge University Press, 1997.
- 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.
- _____; Oliveira, M.F.; Moraes, H.R. Groundwork for the development of the Brazilian Portuguese Wordnet. In: RANCHHOD, E.M.; MAMEDE, N.J. (org.) **Advances in natural language processing**. Berlin: SpringerVerlag, 2002. p.189-196.
- _____. Os domínios lingüístico e tecnológico do estudo do processamento automático das línguas naturais. **Estudos Lingüísticos**, v. 26, 1998, p. 612-617.
- DIRVEN, R.; PÖRINGS, R. (org.). **Metaphor and metonymy in comparison and contrast**. Berlin, New York: Mouton de Gruyter, 2003.

FELLBAUM, C. (org.) **WordNet: an electronic lexical database**. Cambridge, Mass.: The MIT Press, 1998.

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

FRANCIS, W. N. A standard corpus of edited present-day American. English. **College English**, 26, 1965, p. 267-273.

GENTNER, D., HOLYOAK, K.J., KOKINOV, B. (org.). **The Analogical Mind: Perspectives from Cognitive Science**. Cambridge: MIT Press, 2001.

HARTMANN, R. R. K., (org.) **Lexicography: principles and practice**. London: Academic Press, 1983.

HOLLANDA, C. B. **Tantas Palavras**. São Paulo: Companhia das Letras, 2006.

JACKENDOFF, R. **Semantics and cognition**. Cambridge, Mass.: The MIT Press, 1983.

_____. **Semantics structures**. Cambridge, Mass.: The MIT Press, 1990.

JAKOBSON, R. **Linguística e comunicação**. São Paulo, Cultrix, 1973.

KITTAY, E.F. **Metaphor: its cognitive force and linguistic structure**. Oxford: Oxford University 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 University 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 linguístico**. Lisboa: Instituto Camões, 2001.

MILLER, G. A., FELLBAUM, C. Semantic networks of English. **Cognition**, v. 41, n. 1-3, 1991, p. 197-229.

PUSTEJOVSKY, J. **The generative lexicon**. Cambridge, Mass.: The MIT Press, 1995.

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 **Calidoscópico**, v. 7.3, 2009.

SAUSSURE, F. **Curso de lingüística geral.** Trad. de A. Chelini , J. P. Paes e I. Blikstein. São Paulo: Cultrix; USP, 1969.

TAYLOR, J. R. **Linguistic Categorization:** Prototypes in linguistic Theory. Oxford : Clarendon Press, 1989.

VISUAL THESAURUS. Version 3.0.0.Thinkmap Inc., 2005.

WITTGENSTEIN, L. **Philosophical Investigations.** Oxford : Basil Blachwell, 1953.