

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Faculdade de Ciências e Letras – UNESP/Araraquara

FELIPE ISZLAJI DE ALBUQUERQUE

**Modelo linguístico-computacional para um
Dicionário Analógico Digital**

Araraquara
2013

FELIPE ISZLAJI DE ALBUQUERQUE

**Modelo linguístico-computacional para um
Dicionário Analógico Digital**

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

Linha de Pesquisa: Estudos do Léxico

Orientador: Prof. Dr. Bento Carlos Dias da Silva

Araraquara
2013

FELIPE ISZLAJI DE ALBUQUERQUE

**Modelo linguístico-computacional para um
Dicionário Analógico Digital**

TESE PARA A OBTENÇÃO DO GRAU DE DOUTOR

BANCA EXAMINADORA

Presidente e orientador: Prof. Dr. Bento Carlos Dias da Silva
1º Examinador: Profa. Dra. Maria Margarida Martins Salomão
2º Examinador: Profa. Dra. Ariani Di Felippo
3º Examinador: Profa. Dra. Sandra Maria Alúísio
4º Examinador: Prof. Dr. Heronides Maurilio de Melo Moura

Araraquara
2013

à Claudia,
Chico, Otto e Anna.

Agradeço ao CNPq por tornar
possível a realização deste trabalho;

Agradeço a todos que contribuíram para a
realização deste trabalho, especialmente ao
prof. Dr. Bento Carlos Dias da Silva;

Agradeço o apoio dos meus pais, dos meus
irmãos, dos meus amigos e, em especial, agradeço
o apoio da minha mulher.

“E por um bom tempo aquele livro me ajudou no acabamento de romances e letras de canções, sem falar das horas em que eu o folheava à toa; o amor aos dicionários, para o sérvio Milorad Pavic, autor de romances-enciclopédias, é um traço infantil no caráter de um homem adulto.”

– Chico Buarque, prefácio 2ª ed. do Dicionário Analógico (AZEVEDO, 2010)

RESUMO

Esta pesquisa se insere no contexto da construção do *website* **Dicionário Criativo** (www.dicionariocriativo.com.br), idealizado com o objetivo de oferecer ferramentas linguísticas que, relacionadas, instrumentalizam o usuário para a produção de textos de caráter criativo (poemas, letras de canções, anúncios publicitários etc.). Para estruturar e organizar as relações entre os diferentes conteúdos linguísticos, a pesquisa em questão propõe um modelo linguístico-computacional para um *dicionário analógico digital*, tendo em vista que a estrutura *onomasiológica* dos dicionários analógicos é mais adequada às necessidades impostas. Neles, o usuário parte de uma determinada ideia e examina as diferentes maneiras pelas quais essa ideia encontrou expressão na palavra. Ou seja, eles organizam o léxico dentro de uma estrutura conceitual predefinida. O resultado desta pesquisa deverá ser parte integrante do conteúdo do Dicionário Criativo, ao mesmo tempo em que estrutura os demais conteúdos linguísticos previstos (locuções, expressões idiomáticas, provérbios e citações). Tendo em vista que esse trabalho se insere no campo de pesquisa do Processamento Automático das Línguas Naturais (PLN), nossa investigação utiliza a metodologia de Dias-da-Silva (2006), que propõe que as pesquisas em PLN sejam sistematizadas em três domínios complementares: linguístico, linguístico-computacional e computacional. No domínio linguístico, trabalhamos com a abordagem teórica da Semântica Lexical Cognitiva, especialmente com a Semântica de Frames (FILLMORE, 1976) e a Teoria de Protótipos (ROSCH, 1975). No domínio linguístico-computacional, trabalhamos com a Teoria das Redes (BARABÁSI, 2002) e a Teoria dos Conjuntos Difusos (ZADEH, 1965) para o tratamento formal dos resultados obtidos na fase linguística. O domínio computacional não faz parte do escopo desta tese. O *corpus* de onde foram extraídas as unidades lexicais e de onde se inferem parte dos conceitos é composto pelas letras de músicas do álbum *O Grande Circo Místico* (1983) de autoria de Chico Buarque.

Palavras-chave: dicionário analógico digital, ferramenta de auxílio à escrita, processamento de línguas naturais, redes semânticas, teoria dos conjuntos difusos, teoria dos protótipos.

ABSTRACT

This thesis is inserted in the context of the construction of the website *Dicionário Criativo* (www.dicionariocriativo.com.br), which was conceived to offer linguistic tools to help users in creative writing (such as poems, lyrics, advertisements, etc.). In order to structure and organize the relationships between the different linguistic contents, this research proposes a linguistic-computational model to a digital thesaurus, once the onomasiological structure of the onomasiological dictionaries is the more suitable to the use described. With this kind of dictionary, which organizes the lexicon in a predefined conceptual structure, the user works with an idea and examines the ways it is expressed in words. This research results will be part of the content of the *Dicionário Criativo* and will structure the other linguistic contents (phrases, idioms, proverbs, and quotes). We are in the field of Natural Language Processing, approached here according to Dias-da-Silva's (2006) methodology which proposition structures the researches in this field in three complementary domains: the linguistic domain, the linguistic-computational domain, and the computational domain (this later does not take part in the scope of our thesis). In the linguistic domain we work with the Cognitive Lexical Semantics theoretical approach, focusing on Frame Semantics (Fillmore, 1976) and on Theory of Prototypes (ROSCH, 1975), and in the linguistic-computational domain we work with Theory of Networks (BARABÁSI, 2002) and with Theory of Fuzzy Sets (ZADEH, 1965) for formal treatment of the results obtained in the linguistic phase. The corpus from which we extract the lexical unities and from which we infer part of the concepts is composed by lyrics of Chico Buarque's *O Grande Circo Místico* (1983).

keywords: digital thesaurus; writing tool; natural language processing; semantic networks; theory of fuzzy sets; theory of prototypes.

LISTA DE FIGURAS

Figura 1.1. Interface gráfica do Dicionário Criativo.....	5
Figura 1.2. Macroestrutura léxico-conceitual do Dicionário Analógico Digital.....	6
Figura 1.3. Componentes lexicográficos do <i>Dicionário Analógico Digital</i>	7
Figura 1.4. Representação esquemática de uma cauda longa gerada a partir dos CdSAs que instanciam as categorias conceituais de AFEIÇÃO e de SIMPATIA.	12
Figura 2.1 Quadro Sinóptico de Categorias; <i>Dicionário Analógico da Língua Portuguesa</i> (AZEVEDO, 1983, p. IX).....	20
Figura 2.2. Os grupos de palavras; <i>Dicionário Analógico da Língua Portuguesa</i> (AZEVEDO, 1983, p. 1).	21
Figura 2.3 e 2.4. Associações altamente hierarquizadas e associações planas (ZUGMAN, 2008, p. 70).....	26
Figura 2.5. Interface do TeP 2.0 para o usuário final	33
Figura 2.6. Resultado do <i>Visual Thesaurus</i> para a busca pela palavra <i>mouth</i> ; acessado em 01 mar 2010.....	34
Figura 2.7. Resultado do <i>Dictionnaire des Synonymes du CRISCO</i> para a busca pela palavra <i>amour</i> ; acessado em 21 maio 2010.	36
Figura 2.8. Recorte para o resultado do dicionário analógico do <i>Dictionnaire Sensagent</i> para a busca pela palavra <i>amour</i> ; acessado em 21 maio 2010.	37
Figura 2.9. Plano de classificação das ideias do <i>Diccionario Ideológico de la Lengua Espanola</i> (CASARES, 1959).....	41
Figura 2.10. Árvores de categorias do <i>Cadê</i> , um dos primeiros sistemas de buscas da internet.....	42
2.8 Síntese da seção	47
Figura 3.1. <i>Frame</i> DORMITÓRIO DE SOLTEIRO extraído do <i>Dicionário Multilíngue</i> (Reader's Digest Brasil, 1998).	58
Figura 3.2. Distribuição dos nós em Curva de sino e Lei de potência e sua correlação com redes randômicas (mapa rodoviário) e redes livres de escala (mapa aeroviário)	61
Figura 3.3. Redes sociais: exemplo de redes altamente clusterizadas (BARABÁSI, 2002)	63
Figura 3.4. Rede semântica de relações sintagmáticas a partir da UL <i>palco</i>	64
Figura 3.5. Rede semântica de relações paradigmáticas a partir do item lexical <i>amor</i>	65
Figura 3.6. Rede semântica difusa para as relações paradigmáticas do domínio AMOR... ..	66
Figura 3.7. Relações de peso das ULs do CdSA AFETO (domínio AMOR).....	67
Figura 3.8. Exemplo de interface para os CdSAs NAMORADO e NAMORADA que se relacionam sintagmaticamente no interior do <i>frame</i> Relação_Amorosa (<i>anaset</i> AMOR).	68
Figura 4.1. Árvore de Porfírio (ECO, 1984, p. 463).	75

Figura 4.2. Top Ontology da EuroWordNet	84
Figura 4.3. Grandes dimensões da experiência humana.	89
Figura 4.4. “ <i>Ceci n’est pas une pipe</i> ” (“Isto não é um cachimbo”), pintura de René Magritte	91
Figura 4.5. A relação entre forma e significado nos três tipos de signos piercianos (DELBECQUE, 2008, p. 21).....	91
Figura 4.6. Exemplo ilustrativo de uma nuvem de palavras com ULs dos domínios AMOR, FOGO e CASAMENTO.	98
Figura 5.1. Semelhanças de família da categoria JOGO (GLOCK, 1996).	106
Figura 5.2. Radialidade da categoria <i>Balan</i> : mulheres, fogo e coisas perigosas – adaptado de FELTES (2007)	109
Figura 5.3. Representação Geral de uma função de pertinência (NICOLETI;CAMARGO, 2009, p. 25).....	117
Figura 7.1. Exemplo da categoria conceitual BOCA extraído do <i>Dicionário Visual do Corpo Humano SBS</i>	171
Figura 7.2. Resultados do Dicionário Criativo, no módulo de citações, para a busca pela UL <i>amor</i>	208
Figura 7.3. Exemplo de interface do DAD para uma busca por <i>amor</i>	211
Figura 7.4 Modelo de representação dos <i>anaset</i> s, CdSAs e ULs no banco de dados	213

LISTA DE QUADROS

Quadro 1.1. Matriz léxico-conceitual para a construção de um <i>anaset</i>	13
Quadro 2.1. As noções de <i>thesaurus</i>	19
Quadro 2.2 – Lista de <i>thesaurus</i> e <i>dicionários analógicos</i> disponíveis na internet, acessados em: 01 mar 2010.....	33
Quadro 3.1. Exemplo de matriz lexical com o item lexical <i>amar</i>	51
Quadro 3.2 Relações léxico-conceituais paradigmáticas de uma <i>wordnet</i>	52
Quadro 3.3 Relações léxico-conceituais sintagmáticas de uma <i>wordnet</i>	52
Quadro 4.1. Diferentes objetivos e artefatos ontológicos para cada área do conhecimento.....	77
Quadro 4.2. Plano de classificação das ideias.....	79
Quadro 4.3. Grandes Categorias (<i>major headings</i>) em comparação.....	81
Quadro 4.4. <i>Major headings</i> do HTOED.....	88
Quadro 5.1. Domínios conceituais e nível de base.....	111
Quadro 5.2. CdSA para o conceito de <i>HOMEM QUE AMA OU É AMADO</i> do domínio <i>AMOR</i>	119
Quadro 5.3. Exemplo de banco de dados relacional <i>fuzzy</i>	120
Quadro 6.1 – Componentes do <i>anaset</i> na matriz léxico-conceitual.....	123
Quadro 6.2. ULs de entrada do <i>anaset AMOR</i>	126
Quadro 6.3. Matriz preenchida com a Grande Categoria onde se inclui o <i>anaset</i> <i>TEATRO:</i> <i>MUNDO SOCIAL</i>	128
Quadro 6.4. Exemplo de duas categorias conceituais do <i>anaset</i> <i>TEATRO</i>	129
Quadro 6.5. Exemplo de CSs extraídos do <i>Dicionário Houaiss de Sinônimos e Antônimos</i> para o <i>anaset AMOR</i>	132
Quadro 6.6. Exemplo da matriz léxico-conceitual do <i>anaset AMOR</i> preenchida com CSs e CSAs.....	134
Quadro 6.7. Exemplo da matriz léxico-conceitual do <i>anaset AMOR</i> consolidada com CSAs_difusos_sem.....	136
Quadro 6.8. Exemplo da matriz léxico-conceitual do <i>anaset AMOR</i> consolidada com CdSAs.....	139
Quadro 7.1. ULs do domínio <i>AMOR</i> extraídas do <i>corpus</i>	144
Quadro 7.2. ULs de entrada do <i>anaset AMOR</i>	145
Quadro 7.3. Matriz preenchida com a Grande Categoria onde se inclui o <i>anaset AMOR:</i> <i>MUNDO PSÍQUICO</i>	148
Quadro 7.4. Matriz do <i>anaset AMOR</i> preenchida com a estrutura conceitual.....	152
Quadro 7.5. Estrutura conceitual do <i>anaset AMOR</i>	152

Quadro 7.6. Exemplo do <i>anaset</i> AMOR preenchido com CSs e CSAs.....	160
Quadro 7.7. Resultados consolidados para o <i>anaset</i> AMOR.....	161
Quadro 7.8. ULs do domínio CORPO HUMANO extraídas do <i>corpus</i>	165
Quadro 7.9. ULs de entrada do <i>anaset</i> CORPO HUMANO	166
Quadro 7.10. Matriz do <i>anaset</i> CORPO HUMANO preenchida com a estrutura conceitual.	171
Quadro 7.11. Estrutura conceitual do <i>anaset</i> CORPO HUMANO.....	172
Quadro 7.12. Exemplo do <i>anaset</i> CORPO HUMANO preenchido com CSs e CSAs.....	184
Quadro 7.13. Resultados consolidados para o <i>anaset</i> CORPO HUMANO.....	185
Quadro 7.14. ULs do domínio TEATRO extraídas do <i>corpus</i>	188
Quadro 7.15. ULs de entrada do <i>anaset</i> TEATRO.....	189
Quadro 7.16. Matriz preenchida com a Grande Categoria onde se inclui o <i>anaset</i> TEATRO: MUNDO SOCIAL.....	191
Quadro 7.17. Matriz do <i>anaset</i> AMOR preenchida com a estrutura conceitual.....	194
Quadro 7.18. Estrutura conceitual do <i>anaset</i> TEATRO	195
Quadro 7.19. Exemplo do <i>anaset</i> TEATRO preenchido com CSs e CSAs	204
Quadro 7.20. Resultados consolidados para o <i>anaset</i> TEATRO	205
Quadro 7.21. Pequena amostra de EI's do DOEIB, separadas por temas e com seus respectivos significados.	210
Quadro 7.22. Exemplo de preenchimento da tabela UL_CSA_difuso	214
Quadro 7.23. Exemplo da tabela UL_CSA_difuso preenchida com os dados	215
Quadro I.1 – ULs extraídas por faixa do álbum <i>Grande Circo Místico</i>	232

ÍNDICE DE ABREVIATURAS

CS	Conjunto de Sinônimos
CSA	Conjunto de Sinônimos e Analógicos
CdSA	Conjunto difuso de Sinônimos e Analógicos
DAD	Dicionário Analógico Digital
EF	Elementos do Frame
IA	Inteligência Artificial
LC	Linguística Cognitiva
NILC	Núcleo Interinstitucional de Lingüística Computacional
PB	Português Brasileiro
PLN	Processamento de Línguas Naturais
RLE	Rede Livre de Escala
RR	Rede Randômica
TeP	<i>Thesaurus</i> eletrônico para o Português do Brasil
TCD	Teoria dos Conjuntos Difusos
TP	Teoria dos Protótipos
TR	Teoria das Redes
SC	Semântica Cognitiva
UL	Unidade Lexical
HTOED	Historical Thesaurus of the Oxford English Dictionary

SUMÁRIO

SEÇÃO 1 – INTRODUÇÃO	1
1.1 Delimitação e contextualização do tema	1
1.2 Objetivos e estratégias metodológicas	9
1.3 Organização do trabalho	13
SEÇÃO 2 – OS DICIONÁRIOS ANALÓGICOS	16
2.1 O que é um dicionário analógico?	18
2.2 Como se estruturam os dicionários analógicos?	20
2.3 Para que serve um dicionário analógico?	23
2.4 Dicionários Analógicos e Criatividade	24
2.5 Dicionários Analógicos <i>versus</i> Dicionários de Sinônimos	27
2.5.1 O <i>Thesaurus</i> Eletrônico para o Português do Brasil (TeP).....	32
2.6 Dicionários Analógicos na internet.....	33
2.7 Análise do Estado da Arte dos Dicionário Analógicos	37
2.8 Síntese da seção	37
SEÇÃO 3 – RELAÇÕES PARADIGMÁTICAS E SINTAGMÁTICAS NA ORGANIZAÇÃO DO DAD	49
3.1 Wordnets	49
3.2 Framenets	53
3.3 O papel das relações paradigmáticas e sintagmáticas no DAD	55
3.4 Equacionamento linguístico-computacional: Teoria das redes e Redes semânticas ..	60
3.4.1 Rede semântica para o DAD	65
3.5 Síntese da seção	70
SEÇÃO 4 - RELAÇÕES HIERÁRQUICAS E CATEGORIAS SUPREMAS NA CONSTRUÇÃO DO DAD	72
4.1 As Categorias Supremas de Aristóteles	73
4.2 Hierarquias, Taxonomias, e Ontologias Linguísticas.....	75
4.3 O Plano de Classificação das Ideias	79
4.4 Grandes Categorias em comparação	81
4.5 Hierarquias e componentes semânticos.....	84
4.6 Linguística Cognitiva: conhecimento enciclopédico e realismo corporificado	87
4.7 Proposta de Grandes Categorias para o DAD.....	88
4.8 Equacionamento linguístico-computacional: Rizoma	93
4.9 Síntese da seção	96
SEÇÃO 5 – EFEITOS DE SALIÊNCIA COMO CRITÉRIOS DE RELEVÂNCIA DO DAD	100
5.1 Categorias conceituais e categorias linguísticas	101
5.2 Saliência psicológica: o caráter difuso das categorias conceituais	105
5.2.1 Wittgenstein e a semelhança de família	105
5.2.2 Rosch e a Teoria dos protótipos.....	106
5.2.3 Redes radiais.....	109
5.2.4 Proeminência conceitual e nível de base.....	110
5.3 Determinando sentidos e referentes prototípicos: aspectos metodológicos	111

5.4 Equacionamento linguístico-computacional: Teoria dos Conjuntos Difusos para modelar categorias léxico-conceituais difusas	115
5.4.1 Banco de dados relacional <i>fuzzy</i>	119
5.5 Síntese da seção	120
SEÇÃO 6 – MODELO DE REPRESENTAÇÃO E METODOLOGIA DE TRABALHO	122
6.1 O constructo <i>anaset</i>	122
6.1.2 Etapas de construção do <i>anaset</i>	124
6.1.2.1 Identificação e coleta das ULs - Etapa 1	124
6.1.2.2 Identificação e delimitação das categorias conceituais - Etapa 2	127
6.1.2.3 Identificação, coleta e arranjo dos CSs e CSAs - Etapa 3	130
6.1.2.4 Consolidação dos resultados - Etapa 4	134
6.2 Domínios léxico-conceituais que serão modelados: CORPO HUMANO, AMOR e TEATRO	141
6.3 Síntese da seção	142
SEÇÃO 7 – CONSTRUÇÃO E IMPLEMENTAÇÃO DE UM MODELO PARA O DAD	143
7.1 Modelagem do <i>anaset</i> AMOR	143
7.1.1 Etapa 1 - Identificação e coleta das ULs do domínio AMOR	144
7.1.2 Etapa 2 - Estruturação conceitual do <i>anaset</i> AMOR	146
7.1.3 Etapa 3 - Identificação, coleta e arranjo dos CSs e CSAs do <i>anaset</i> AMOR	153
7.1.4 Etapa 4 - Consolidação dos resultados do <i>anaset</i> AMOR	161
7.2 Modelagem do <i>anaset</i> CORPO HUMANO	165
7.2.1 Etapa 1 - Identificação e coleta das ULs do domínio CORPO HUMANO	165
7.2.2 Etapa 2 - Estruturação conceitual do <i>anaset</i> CORPO HUMANO	167
7.2.3 Etapa 3 - Identificação, coleta e arranjo dos CSs e CSAs do <i>anaset</i> CORPO HUMANO	174
7.2.4 Etapa 4 - Consolidação dos resultados do <i>anaset</i> CORPO HUMANO	184
7.3 Modelagem do <i>anaset</i> TEATRO	188
7.3.1 Etapa 1 - Identificação e coleta das ULs do domínio TEATRO	188
7.3.2 Etapa 2 - Estruturação conceitual do <i>anaset</i> TEATRO	190
7.3.3 Etapa 3 - Identificação, coleta e arranjo dos CSs e CSAs do <i>anaset</i> TEATRO	196
7.3.4 Etapa 4 - Consolidação dos resultados do <i>anaset</i> TEATRO	204
7.4 Integração com o Dicionário Criativo e propostas de interface do DAD	207
7.5. Aspectos Computacionais	212
7.5.1 Banco de Dados	213
7.5.2 Back-end	215
7.5.3 Front-end	215
7.6 Síntese da seção	216
SEÇÃO 8 – CONCLUSÃO	218
REFERÊNCIAS BIBLIOGRÁFICAS	226
APÊNDICE I – Corpus da pesquisa	232
APÊNDICE II – Fontes de informação léxico-conceitual	234
ANEXO I - Canções de Chico Buarque e Edu Lobo para o Balé <i>Grande circo místico</i> (1982)	241
ANEXO II – Lista de arquivos do CD em anexo	249

SEÇÃO 1 – INTRODUÇÃO

Esta Seção introdutória está dividida em três subseções. A Subseção 1.1 aborda o tema da tese – *Modelo linguístico-computacional para um Dicionário Analógico Digital* – no contexto da tradição dos dicionários analógicos e, ao mesmo tempo, no contexto da construção do *website Dicionário Criativo*. Ainda na mesma Subseção delimitamos o escopo da tese dentro do campo de pesquisas em Processamento da Linguagem Natural (doravante **PLN**) e no âmbito de investigações que projetam e desenvolvem **ferramentas de auxílio à escrita** (*computer-assisted writing system*) (DIAS-DA-SILVA, DI FELLIPO, 2008). A Subseção 1.2 esclarece os objetivos da tese e os procedimentos metodológicos que são necessários para alcançá-los. A Subseção 1.3 explicita a organização da tese.

1.1 DELIMITAÇÃO E CONTEXTUALIZAÇÃO DO TEMA

“O *Dicionário Analógico* fornece um apanhado global da língua portuguesa, racionalmente distribuído; dá a palavra ou locução que se ignora ou saiu da memória; evita repetições de palavras e as impropriedades resultantes do desconhecimento de inúmeros vocábulos e das locuções analógicas, sinonímicas e antonímicas; obriga o escritor que procura uma palavra ou locução a percorrer o grupo das que se lhe acham associadas por uma ligação ideológica.” (SPITZER, 1952)

O trecho acima, extraído da apresentação para o *Dicionário Analógico da Língua Portuguesa* do padre Carlos Spitzer (1952), introduz a finalidade da obra: alargar e distender o vocabulário daqueles que dela farão uso, principalmente estudantes e profissionais da escrita. No sentido inverso aos dicionários padrões de língua, de ordem *semasiológica*, em que as entradas são **itens lexicais** ordenados alfabeticamente e aos quais são conferidos os diferentes significados pertinentes, as entradas nos *dicionários analógicos* são **categorias conceituais**, ou mesmo **domínios conceituais** inteiros, sob os quais se agrupam as **unidades lexicais** (doravante **UL**) que lhes correspondem. Essa direção do nível conceitual para o nível léxico é denominada, em lexicografia, de **percurso onomasiológico** (BIDERMAN, [1978] 2001). Enquanto os dicionários semasiológicos possuem a função decodificadora e, portanto, são conhecidos como **dicionários de compreensão** (CORREIA, 2009, p. 44), os dicionários onomasiológicos

possuem a função codificadora e são denominados **dicionários de produção**. Os dicionários analógicos possuem esta última função, ou seja, auxiliar o seu usuário na produção do texto. Quando, por exemplo, um poeta ou jornalista tem uma ideia em mente, mas não sabe ou não lembra a palavra que melhor expressa o conteúdo dessa ideia, ele pode recorrer ao dicionário analógico e encontrar dezenas de palavras que, com diferentes matizes semânticos, lexicalizam aquela ideia.

Dessa maneira, os dicionários analógicos são valiosos para aqueles que trabalham com a escrita, como jornalistas, redatores publicitários, letristas, poetas, escritores e roteiristas, além de intelectuais e acadêmicos. No entanto, mesmo assim, os dicionários analógicos estiveram muito tempo fora de catálogo no Brasil. Até 2010, quando a Editora Lexikon publicou uma nova edição, revista e ampliada, do *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, [1950] 2010) do professor goiano Francisco dos Santos Azevedo, a obra esteve fora de catálogo por quase três décadas. Não há dúvidas de que essa nova edição preenche uma lacuna no mercado editorial brasileiro. O seu sucesso de vendas, com mais de 10 mil obras vendidas em menos de seis meses, demonstra que existe demanda para esse tipo de conteúdo. Porém, as tecnologias da informação têm propiciado um crescente e acelerado processo de digitalização da informação e os dicionários e obras de referência ganham em funcionalidade e usabilidade quando se utilizam das possibilidades da tecnologia digital. Além da limitação física de tamanho imposta por um dicionário impresso, fato que não existe no ambiente digital, as pesquisas por palavra-chave em gigantescos bancos de dados relacionais, a estrutura de hiperlinks e as possibilidades de interação do usuário com uma interface gráfica tornam o acesso à informação contida nos dicionários e obras de referência muito mais fácil e produtivo. O armazenamento digital potencializa a rede de relações morfológicas, sintagmáticas, semânticas e paradigmáticas entre as diferentes ULs e possibilita o acesso à informação por caminhos hiperlinkados. Além disso, o ambiente digital permite a interligação da informação dos diferentes tipos de dicionários e obras de referência. Um dicionário de compreensão, por exemplo, pode se complementar com um dicionário de produção e tê-los em um mesmo ambiente, retornando respostas simultâneas a uma mesma pesquisa, tem a capacidade de potencializar suas funções.

Por conta dessa mudança e da necessidade cada vez maior de construção e manipulação do léxico em grandes bases de dados estruturados, a tradicional indústria

de dicionários enfrenta dificuldades nessa transição. A academia tem protagonizado esse processo e vem desenvolvendo aplicações computacionais com diferentes níveis de sofisticação dentro do campo de pesquisas em PLN, campo que se dedica exatamente a investigar e desenvolver sistemas computacionais em que a língua natural é o objeto primário (GRISHMAN, 1986). Ainda que a construção dessas aplicações nem sempre seja o foco das investigações acadêmicas, elas muitas vezes surgem como resultado de processos, métodos e recursos necessários à construção dos sistemas de PLN e a compilação de dicionários, *thesaurus* e enciclopédias eletrônicas destaca-se hoje como atividade do campo, assim como a construção de ferramentas de auxílio à escrita. Como exemplo desse protagonismo, podemos citar o *Thesaurus Eletrônico do Português do Brasil* (doravante **TeP**)¹. Desenvolvido por uma equipe de pesquisadores vinculados ao Núcleo Interinstitucional de Linguística Computacional (NILC), o TeP (DIAS-DA-SILVA *et al.*, 2000) é o único *thesaurus*² de língua geral computacionalmente implementado e disponível na internet para o português brasileiro (doravante **PB**).

O projeto do Dicionário Criativo está inserido nesse contexto de ferramentas digitais para o auxílio à escrita disponíveis na internet. Sua particularidade, contudo, é direcionar atenção para o auxílio à escrita criativa. Para isso, duas características são essenciais para o sucesso da ferramenta: i) uma grande variedade de conteúdos linguísticos e não linguísticos (p. ex. imagens) e ii) a interligação dos conteúdos por meio de conexões analógicas. Conforme o professor de psicologia norte-americano Sarnoff Mednick, no artigo *The associative basis of the creative process* (1962), “se o trabalho criativo depende das analogias e das conexões que conseguimos fazer, quanto mais variado for nosso estoque de matéria-prima, de ideias e de conhecimento, mais chances teremos de chegar a uma novidade”.

Para satisfazer (i), o Dicionário Criativo foi concebido em módulos, em que os diferentes conteúdos linguísticos e não linguísticos são organizados. A Figura 1.1 ilustra a interface do Dicionário Criativo. Com a ajuda da figura, apresentamos, no parágrafo abaixo, a sua macroestrutura em termos não técnicos e incluímos, entre chaves, as fontes de onde foram extraídos os conteúdos de cada módulo. Em seguida, incluiremos a presente pesquisa no contexto do Dicionário Criativo e faremos uso da terminologia lexicográfica que permanecerá ao longo da tese.

¹ Disponível em <http://www.nilc.icmc.usp.br/tep2/index.htm>.

² Dicionário de sinônimos e antônimos.

No topo da *webpage*, abaixo da logomarca, é exibido o campo de pesquisa da palavra-chave **(a)** e, conforme se faça a pesquisa por alguma palavra, desenrola-se logo abaixo um *dicionário de língua portuguesa* padrão [iDicionário Aulete³] **(1)**, um *dicionário analógico* [ampliação do modelo resultante desta tese] **(2)**, um *dicionário de expressões idiomáticas* **(3)**, outro de *provérbios* **(4)** e um terceiro de *citações* [diversas fontes⁴] **(5)**; uma *enciclopédia* [wikipedia⁵] **(6)** e um *banco de imagens* [GettyImages⁶] **(7)**.

³ Disponível em <http://aulete.uol.com.br>.

⁴ Digitalização de dicionários impressos e pesquisa em dicionários on-line específicos de expressões idiomáticas, de provérbios e de citações.

⁵ Disponível em <http://pt.wikipedia.org>.

⁶ Disponível em <http://www.gettyimages.com.br>.

.di.
cionário
rio
criativo

a

amor (a.mor)[6]

sm.

1 Sentimento que faz alguém querer o bem de outrem ou de alguma coisa

2 Afeto profundo, devoção de uma pessoa a outra (amor materno)

3 Sentimento terno e caloroso de uma pessoa por outra, inclusive de natureza física e sexual

4 Relação amorosa

5 O ato sexual (fazer amor)

6 Inclinação, apego ao que desperta prazer ou empatia (amor à música)

Amor	<i>subs.</i>	carinho idolatria afeto amoricos amorosidade	amorio inclinação desejo dileção predileção	preferência simpatia estremecimento benquerença afeição	dedicação querença admiração apego aferro
Sentimento Concórdia Deleite Desejo Amizade Favorito Carícias Benevolência Aprovação	<i>v.</i>	amar apaixonar-se por estar apaixonado por estar enamorado	por estimar muito benquerer gostar de ser devoto de	cultuar ter em preço querer a dar a vida por	estremecer simpatizar com sentir ternura por entranhar-se estimar
	<i>adj.</i>	amante lamecha babadinho amador amativo	namorador cupidíneo cupidinoso vênereo apaixonado	apaixonante louco de amor perdido de amor	doído de amor bajoujo namorado enamorado amorisado

Amor ao próximo: Sentimento ou exercício da caridade.

Amor à primeira vista: Instantâneo interesse amoroso por alguém logo ao primeiro encontro.

Amor carnal: O mesmo que amor físico.

Amor cortês: *Liter.* Na literatura medieval, amor nobre, puro e leal de um cavaleiro por uma mulher.

Amor físico: O que busca a satisfação sexual; amor carnal.

Amor livre: O que prescinde de, e mesmo repudia, convenções morais ou legais.

Amor com amor se paga.
Amor e dinheiro não querem parceiro.
Amor é a gente querendo achar o que é da gente.
Amor é sede depois de se ter bebido.
Pancada de amor não dói.

"Sempre há um pouco de loucura no amor, porém sempre há um pouco de razão na loucura."
Friedrich Nietzsche

"O amor é a única coisa que me conserva o juízo."
Sue Townsend

"Um grande amor não necessita de muitas palavras."
Pink

"E não importa para onde vamos seguindo, entre nós sempre haverá lembrança de um olhar, de um carinho, e da integridade de momentos sinceros."
Mário Quintana

Amor

A palavra amor (do latim amor) presta-se a múltiplos significados na língua portuguesa. Pode significar afeição, compaixão, misericórdia, ou ainda, inclinação, atração, apêlito, paixão, querer bem, satisfação, conquista, desejo, libido, etc. O conceito mais popular de amor envolve, de modo geral, a formação de um vínculo emocional com alguém, ou com algum objeto que seja capaz de receber este comportamento amoroso e enviar os estímulos sensoriais e psicológicos necessários para a sua manutenção e motivação.

gettyimages | br

Figura 1.1. Interface gráfica do **Dicionário Criativo**.

Para satisfazer (ii), ou seja, a interligação dos conteúdos por meio de conexões analógicas, levamos a cabo o desenvolvimento da presente pesquisa, intitulada *Modelo linguístico-computacional para um Dicionário Analógico Digital*. Propusemo-nos a pesquisar, elaborar e construir um modelo de dicionário analógico digital, por entendermos que sua confecção estrutural completa (ainda que não exaustiva) seria um trabalho hercúleo e incompatível com as limitações de tempo, recursos, e mesmo de escopo, de uma tese de doutorado. O resultado desta pesquisa deverá, portanto, ser posteriormente ampliado para integrar o conteúdo do Dicionário Criativo, especificamente do módulo (2), ao mesmo tempo em que sua macroestrutura de ordem onomasiológica servirá para estruturar grande parte dos outros conteúdos, especificamente os módulos (3), (4), (5) e (7), além de potenciais módulos novos.

Na Figura 1.2, isolamos o módulo correspondente ao *Dicionário Analógico Digital* (2) (doravante **DAD**) e destacamos os principais termos provenientes da Semântica Lexical (CRUSE, 1986; 2006) para explicitar a sua macroestrutura e funcionamento. Ao mesmo tempo, a Figura 1.3 faz a correspondência entre os termos provenientes da Semântica Lexical com a função que lhes é atribuída dentro do dicionário, usando para isso a terminologia lexicográfica.

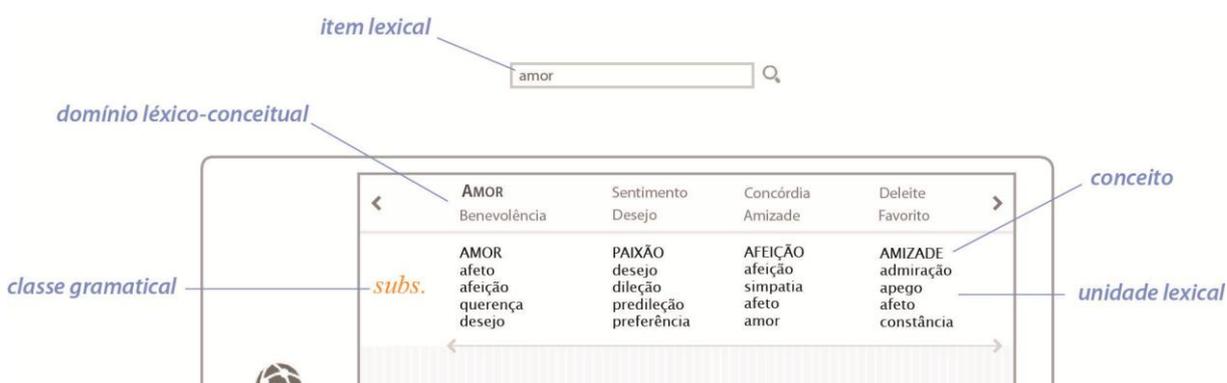


Figura 1.2. Macroestrutura léxico-conceitual do DAD.

O DAD será organizado de forma que cada **entrada** corresponda a um **domínio léxico-conceitual**. Cada domínio é, então, dividido entre as quatro principais **classes gramaticais**⁷ (substantivo, verbo, adjetivo e advérbio). Dentro de cada classe gramatical estarão as **unidades lexicais**, organizadas em diferentes conjuntos difusos

⁷ Para o modelo em desenvolvimento nesta tese, escolhemos trabalhar exclusivamente com a classe dos substantivos.

de sinônimos e análogos, em que a unidade lexical de maior relevância – segundo os critérios que serão explicitados na Seção 5 – estará em destaque para representar o **conceito** ou **categoria conceitual** que aquele conjunto de sinônimos e analógicos **EVoca**.

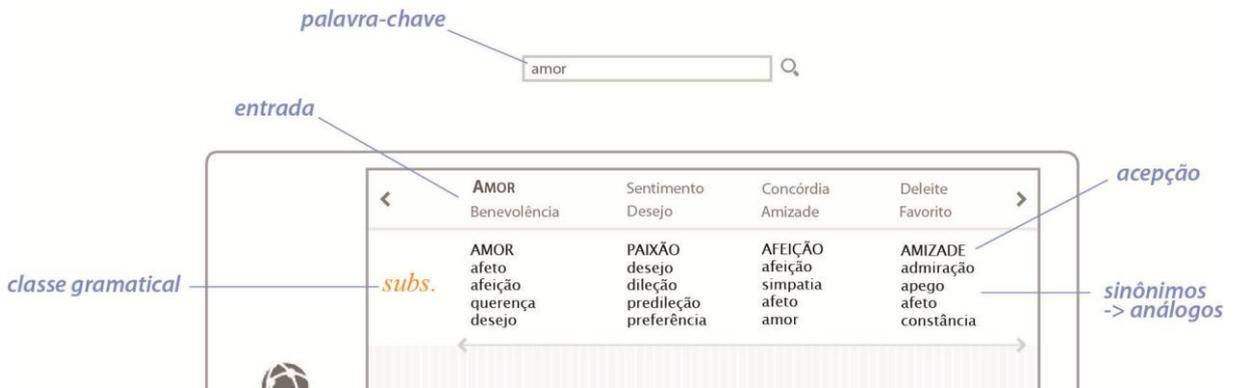


Figura 1.3. Componentes lexicográficos do DAD.

O funcionamento básico do DAD cumpre as seguintes etapas: i) o usuário digita uma **palavra-chave** (que deve ser um **item lexical**) no campo de pesquisa (no exemplo, a palavra-chave *amor*); ii) o sistema identifica se a palavra é ou não uma **forma flexionada**; caso o seja, o sistema recupera a **forma canônica** correspondente e efetua a busca à partir dela; iii) o sistema percorre todos os verbetes existentes e, como resultado, exibe no cabeçalho todas as entradas dos verbetes em que a forma canônica aparece (no exemplo, as entradas AMOR, BENEVOLÊNCIA, SENTIMENTO, DESEJO, CONCÓRDIA, AMIZADE, DELEITE e FAVORITO); iv) se a forma buscada corresponder a uma entrada, o verbete relativo àquela entrada é o que será exibido primeiro; v) a entrada a qual corresponde o verbete exibido abaixo é destacada em negrito (no exemplo, a entrada AMOR); vi) dentro dos verbetes, cada conjunto difuso de sinônimos e analógicos (doravante **CdSA**) é uma **acepção**, encabeçada por uma unidade lexical (doravante **UL**) que, segundo critérios de relevância, melhor lexicaliza o conceito expresso pela acepção (no exemplo, as acepções AMOR, PAIXÃO, AFEIÇÃO e AMIZADE). O exemplo acima apresenta resultados para ULs em relações paradigmáticas, inspirado, por um lado, pelos dicionários analógicos impressos e, por outro lado, pela base de dados lexicais **WordNet** de Princeton (doravante WN.Pr) (FELLBAUM, 1998).

No entanto, uma importante característica dos dicionários analógicos é listar dentro dos seus **verbetes** (= **grupos de analógicos**) não apenas ULs em relações

paradigmáticas, mas também ULs em relações sintagmáticas. Nos dicionários analógicos impressos, os dois tipos de relação não são explicitados. No nosso modelo, no entanto, sempre que possível, cada entrada possui uma janela para as relações paradigmáticas e uma janela para as relações sintagmáticas. Assim como nos inspiramos na WordNet para implementar as relações paradigmáticas, buscamos inspiração na **FrameNet** (FILLMORE *et al.*, 2003) para a implementação das relações sintagmáticas (cf. 3.2).

Os dicionários analógicos devem seu nome à relação de **analogia**. De acordo com Gentner (1983), a analogia é um processo cognitivo que transfere informações de um domínio-alvo para um domínio-fonte. Nesse sentido, o conceito de analogia se aproxima bastante da concepção de **metáfora conceitual** da Linguística Cognitiva (GENTNER *et al.*, 2001). As pessoas usam domínios concretos e familiares para compreender domínios desconhecidos ou abstratos. No entanto, os autores dos *dicionários analógicos* utilizam a relação de analogia de maneira ampla. Além da relação metafórica, são contempladas relações sinonímicas, metonímicas, de especialização e de generalização. Na tradição dos dicionários analógicos, a analogia se dá como a relação existente entre as palavras "seja por uma comunidade de ideias, seja por relações de emprego frequentes de causa, de meio, de efeito" (BOISSIERE, 1862). Nesse sentido, a analogia é, portanto, uma relação entre "palavras que são coligadas pelo sentido"⁸. Para os fins desta tese, portanto, os termos analogia, analógico e análogos se referem a toda sorte de aparentados semânticos minimamente convencionalizados. No verbete **897. Amor** do *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, 1950[2010]) temos, por exemplo, as seguintes ULs e suas relações de sentido com o conceito AMOR:

- (1) *afeto, afeição* (relação sinonímica)
- (2) *ardor, fervor, chamas, calor* (relação metafórica)
- (3) *sexo, beijo, abraço, carícia* (relação metonímica)
- (4) *sentimento, emoção* (relação de generalização)
- (5) *amor materno, amor paterno* (relação de especialização)

Na Seção 2, que trata dos dicionários analógicos, daremos continuidade a essas questões e elucidaremos como as diferentes relações de sentido serão tratadas no nosso modelo. A próxima Subseção (1.2) esclarece os objetivos da tese e os procedimentos

⁸ *Le petit Robert: dictionnaire de la langue française*. Paris: Dictionnaires Le Robert, 1996. (CD-Rom)

metodológicos que são necessários para alcançá-los tendo em vista o contexto delineado nesta Subseção (1.1).

1.2 OBJETIVOS E ESTRATÉGIAS METODOLÓGICAS

Nosso objetivo geral é construir a parte léxico-conceitual de uma aplicação linguístico-computacional que será útil como ferramenta de auxílio à escrita criativa. Na tese aqui apresentada, não pretendemos dar conta de todo o conhecimento linguístico necessário para a construção do *website* Dicionário Criativo, mas apenas da parte correspondente ao DAD. Nossa proposta pretende melhorar potenciais inconsistências dos dicionários analógicos tradicionais à luz dos avanços nas pesquisas em Semântica Lexical e, dentro do contexto da tecnologia digital, progredir em direção aos avanços das pesquisas em PLN. Para tanto, faz-se necessários os seguintes objetivos específicos:

I) explicitar a diferença entre relações semânticas paradigmáticas e sintagmáticas no interior dos grupos de analógicos. Nossa hipótese é a de que as relações paradigmáticas (sinonímia, antonímia, hiperonímia etc.) são melhor trabalhadas em estruturas como as das redes *wordnets* (MILLER *et al.*, 1990; FELLBAUM, 1998), e as relações sintagmáticas (agente, instrumento, lugar etc.) são adequadamente tratadas em estruturas semelhantes às das redes *framenets* (FILLMORE *et al.*, 2003; BAKER, FILLMORE e LOWE, 1998);

II) propor e implementar um sistema conceitual de **Grandes Categorias** (*major headings*) e uma relação hierárquica entre os **conceitos lexicalizados** (CRYSTAL, 1997) que satisfaça os objetivos de um dicionário de analogias com a função de estimular a criatividade. Nossas pesquisas nos levaram a um alinhamento com a proposta do *Historical Thesaurus of the Oxford English Dictionary* (KAY, C. *et al.*, 2009), que organiza todos os conceitos dentro de apenas três grandes categorias conceituais: MUNDO EXTERNO, MUNDO MENTAL e MUNDO SOCIAL⁹;

III) propor e implementar i) critérios de relevância no interior dos grupos de analógicos baseados nos princípios de **saliência semântica**, ou seja, o papel preponderante que os sentidos **centrais** ou **prototípicos** (ROSCH, 1975) representam dentro de um domínio léxico-conceitual; e ii) critérios de relevância das ULs no interior

⁹ No original: The external world, The mental world e The social world.

dos conjuntos de sinônimos e analógicos baseado na frequência de emprego das ULs extraídas de diferentes dicionários de sinônimos e dicionários analógicos;

Diante da intrínseca interdisciplinaridade da pesquisa em PLN, Dias-da-Silva (1996, 1998, 2006) propõe a organização da investigação em três grandes domínios complementares: **Linguístico**, **Linguístico-Computacional** e **Computacional**. No primeiro domínio, é sistematizado (reunido e descrito) o conhecimento linguístico necessário para o desenvolvimento da aplicação; no segundo domínio, são propostas as modelagens e representações formais para essa sistematização; por fim, no terceiro domínio, as formalizações da etapa anterior são implementadas em programas de computador em interfaces especificamente criadas para a codificação.

Por conseguinte, essa é a metodologia que adotamos tanto para organizar as discussões em cada uma das seções da tese, quanto para desenvolver e concluir a pesquisa com a proposição do dicionário almejado. Assim, para cada desafio posto em cada seção, a nossa tarefa é a de explicitar o caminho de investigação em cada um dos domínios.

No domínio linguístico, trabalhamos no nível de análise semântico e lexical, dentro do paradigma de pesquisa da *Linguística Cognitiva*. A *Semântica Lexical Cognitiva* provisiona o embasamento necessário para o estudo das relações semânticas tanto paradigmáticas quanto sintagmáticas e fornece o ferramental necessário para o estudo do processo de categorização humana, indispensável para superar os desafios discutidos nas seções da tese. Um dos objetivos principais da *Semântica Lexical Cognitiva* é delinear uma abordagem geral para a teoria da categorização humana que, em oposição à visão estruturalista do significado, defende que os seres humanos compreendem o mundo por meio de diferentes modelos de categorização (LAKOFF, 1987). Um dos principais modelos teóricos desse paradigma de pesquisa é a **Teoria dos Protótipos** [*Prototype Theory*] (ROSCH, 1975), que postula que as categorias não são estruturas homogêneas, mas antes 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, 1989). Tal proposta abarca uma dupla concepção de categoria e de categorização. Por um lado, distingue a estruturação interna das categorias conceituais (a dimensão horizontal) e, por outro lado, estabelece as linhas gerais da estruturação intercategoriais (dimensão vertical). Ainda no domínio linguístico, mas em um nível

diferente, nossa proposta nos leva aos estudos em *Lexicografia*, entendida como a “técnica de feitura de dicionários” (BORBA, 2003), tendo em vista que a tarefa fim desta tese é a elaboração de um tipo específico de dicionário. Dessa forma, mesmo sem a necessidade de dedicar uma Seção ao assunto ou explicar longamente sobre os princípios que norteiam a confecção de dicionários, estaremos constantemente nos utilizando dos métodos e da metalinguagem oriunda desse campo de estudos.

No domínio linguístico-computacional, a formalização do significado lexical, ocupação pertinente à *Semântica Lexical Computacional*, pode ser beneficiada pelos modelos de **Representação do Conhecimento** (doravante **RC**) (DIAS-DA-SILVA, 2006). Os modelos de RC fornecem o arcabouço teórico-metodológico e a metalinguagem formal para a representação do significado lexical levantado na fase linguística. Diante desse arcabouço, elegemos como referência os formalismos (i) com base em **redes semânticas** (QUILLIAN, 1967; 1968) e (ii) com base em **frames semânticos** (MINSKY, 1975; FILLMORE, 1976; SCHANK, ABELSON, 1977). As redes semânticas têm uma extensa tradição na área de RC e são amplamente utilizadas, principalmente, pela simplicidade e elegância com que podem ser implementadas (HANDKE, 1995). Ao mesmo tempo, como forma de modelar as categorias e os CdSAs (Seção 5), utilizaremos a **Teoria dos Conjuntos Difusos** (*fuzzy set theory*) (ZADEH, 1987), que serve para representar e inferir conhecimento sobre informações imprecisas.

Por fim, questões relacionadas ao domínio computacional, como a arquitetura de bancos de dados relacionais MySQL e a camada de aplicação em linguagem de programação PHP, apesar de fora do escopo desta tese, serão brevemente sugeridas como uma etapa complementar a este trabalho na última seção. Contemplaremos, ainda, na mesma seção, questões pertinentes à interface gráfica de consulta para o usuário final, desde que relevantes para as discussões acerca da disposição visual dos dados – em termos de *design* da informação.

Em analogia ao constructo *synset* da WN.Pr (cf. 3.1), criamos, para os fins deste trabalho, o constructo *anaset*. O *anaset* é a estrutura por detrás de cada verbete do DAD e equivale aos grupos de palavras dos dicionários analógicos tradicionais. Assim como um *synset* é um conjunto de unidades lexicais sinônimas que evocam um mesmo conceito, o *anaset* é um conjunto de categorias conceituais relacionadas que evocam um mesmo domínio conceitual. Cada categoria conceitual será instanciada por um conjunto difuso de unidades lexicais sinônimas e análogas, fruto do achatamento entre conjuntos

de sinônimos (doravante **CS**) e conjuntos de sinônimos e analógicos (doravante **CSA**) equivalentes extraídos, respectivamente, de dicionários de sinônimos e dicionários analógicos. Ou seja, as ULs não terão o mesmo *peso* dentro do conjunto. Portanto, a categoria conceitual glosada por “atração muito intensa de uma pessoa por outra” poderá ser representada pelo conjunto abaixo, em que os números representam o peso de determinada UL no interior do conjunto:

{AFEIÇÃO (1), amor (0.958), afeto (0.896), simpatia (0.813), ternura (0.667), apego (0.604), carinho (0.563), inclinação (0.5), benquerença (0.458), benevolência (0.458), querença (0.438), dileção (0.396), amizade (0.396), dedicação (0.333), admiração (0.292), estima (0.292), predileção (0.25), preferência (0.229), derretimento (0.229), meiguice (0.229), idílio (0.229), aferro (0.208), constância (0.208), intimidade (0.208), derriço (0.208), conchego (0.208), estremecimento (0.208), idiopatia (0.208), fraternidade (0.208), chamego (0.188), boa vontade (0.188), caridade (0.188), ênfatuação (0.188), comunhão de sentimentos (0.188), gosto (0.188), desvelo (0.188), zelo (0.188), cuidado (0.188), adoração (0.125), idolatria (0.104), amorosidade (0.104), [...]}

O conjunto acima, se representado em um plano cartesiano, toma a forma de uma cauda longa (*long tail*, cf. 3.4), ou seja, poucas ULs possuem um peso grande, enquanto muitas ULs detêm um peso pequeno. Isso deverá significar, conforme a nossa intenção, que, no interior de um domínio léxico-conceitual *A*, as ULs com maior peso lexicalizam melhor um conceito *S* do que as ULs com menor peso. No entanto, conforme a representação esquemática da Figura 1.4, as ULs de menor peso (que formam a cauda longa) servirão como ligação entre as diferentes categorias conceituais, explicitando, dessa forma, que as fronteiras entre categorias muito próximas não possuem, na verdade, contornos muito bem definidos. Dentro do domínio léxico-conceitual AMOR, por exemplo, as categorias conceituais de AFEIÇÃO e de SIMPATIA compartilham muitas ULs.

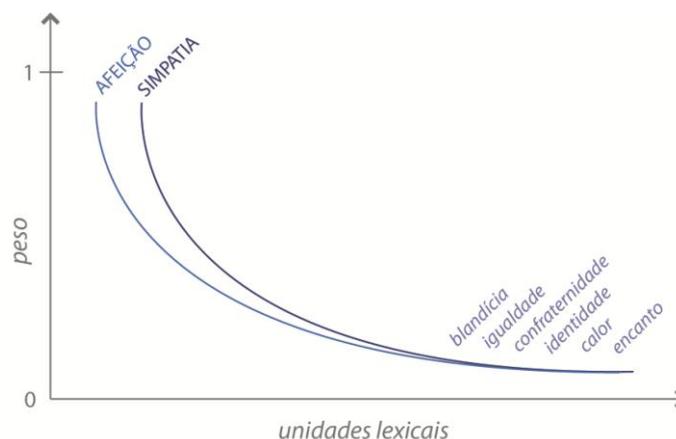


Figura 1.4. Representação esquemática de uma cauda longa gerada a partir dos CdsAs que instanciam as categorias conceituais de AFEIÇÃO e de SIMPATIA.

Como forma de organizar os procedimentos metodológicos necessários para a construção de um *anaset*, foi elaborada uma matriz léxico-conceitual (Quadro 1.1) em que as ULs extraídas do *corpus* e das obras de referência ocupam a coluna da extrema esquerda e as demais colunas correspondem às categorias conceituais pertinentes ao domínio léxico-conceitual em questão. No cruzamento, por exemplo, entre a linha **UL1** e a coluna **S1** (em azul), a célula será preenchida com os CSs e CSAs – extraídos dos dicionários de sinônimos e dicionários analógicos *X, Y, Z* –, que possuírem a unidade lexical *UL1* e ao mesmo tempo instanciarem a categoria conceitual *S1*.

Quadro 1.1. Matriz léxico-conceitual para a construção de um *anaset*.

Anaset A	S1	S2	S3	S4	S5	S6
UL 1	{CS1} [X] {CS2} [Y] {CSA3} [Z]					
UL 2		{CS4} [X] {CSA5} [Y]				
UL 3						

Os detalhes de cada componente e a complementação teórico-metodológica da matriz léxico-conceitual que serve de construção para os *anaset*s serão explicitados e tratados ao longo da tese, conforme os temas e objetivos de cada seção. A Seção 6 mostrará como a matriz será utilizada na construção e implementação do modelo do DAD, ou seja, como e quais resultados podem ser obtidos a partir da matriz para serem incorporados ao banco de dados do dicionário e exibidos conforme a interface gráfica do Dicionário Criativo.

1.3 ORGANIZAÇÃO DO TRABALHO

Levando em consideração a delimitação do tema, os objetivos e as estratégias metodológicas descritas em 1.1 e 1.2, respectivamente, a proposta de tese *Modelo linguístico-computacional para um Dicionário Analógico Digital* organiza-se em sete seções, que passamos a descrever a seguir.

A segunda Seção abordará a tradição dos *dicionários analógicos*, explicitando informações que serão relevantes para todo o desenvolvimento posterior da tese, tais

como sua origem, seu funcionamento e sua utilidade. Na mesma seção, faremos uma análise crítica do estado da arte desse tipo de obra lexicográfica para identificar potenciais inconsistências e propor soluções, tendo sempre em vista a construção de um modelo digital. Para tanto, procederemos a uma análise da forma e do conteúdo de dicionários analógicos do português, inglês, francês e espanhol, publicados em papel ou disponíveis em meio digital, com vistas à delimitação do objeto e suas características, bem como da avaliação de seus conteúdos como fonte de conhecimento lexical.

As seções seguintes examinarão, em detalhes, os problemas apontados na Seção 2 e formularão hipóteses para a organização da informação léxico-conceitual dentro de um dicionário analógico digital, com vistas a encontrar soluções nos domínios linguístico e linguístico-computacional, que inclui o design e a arquitetura da informação em ambiente digital¹⁰.

A organização se dará em torno de três eixos, cada qual representando um desafio com relação à modelagem de um dicionário analógico digital, conforme os objetivos especificados em 1.2. Por fim, cada uma dessas seções estará subdividida em duas grandes partes, sendo as primeiras sempre dedicadas ao domínio linguístico e as últimas ao domínio linguístico-computacional, seguindo a metodologia adotada e descrita em 1.2.

A Seção 3 propõe estruturar a informação léxico-conceitual em termos de relações paradigmáticas e sintagmáticas. Para tanto, utilizaremos as bem-sucedidas experiências adquiridas pelos proponentes das redes WordNet (MILLER *et al.*, 1990; FELLBAUM, 1998) e FrameNet (FILLMORE *et al.*, 2003), ainda que com elas conservemos importantes diferenças que oportunamente serão apontadas. Na segunda metade da seção, no domínio linguístico-computacional, vamos apresentar as noções que aproveitaremos da Teoria das Redes (BARABÁSI, 2002) e mostraremos como elas podem ser utilizadas como referência para formalizar as relações semânticas paradigmáticas e sintagmáticas em termos de redes semânticas.

A Seção 4 propõe um sistema conceitual de Grandes Categorias (*major headings*¹¹) diferente daquele consagrado no *Roget's Thesaurus* e reproduzido no *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, [1950] 2010), de fortes raízes

¹⁰ Em Tecnologia da Informação, o termo **arquitetura da informação** refere-se ao design dos dados armazenados pelos sistemas de informação, ao passo que o termo **design da informação** refere-se à interface gráfica da informação, ou seja, é a parte do sistema que interage diretamente com o usuário final.

¹¹ Refere-se ao nível mais alto de uma hierarquia conceitual.

aristotélicas. Como veremos, nossas pesquisas nos levaram a um alinhamento com a proposta do *Historical Thesaurus of the Oxford English Dictionary* (KAY, C. et al. [eds.], 2009), que organiza sua estrutura léxico-conceitual sob três grandes categorias: *THE EXTERNAL WORLD*, *THE MENTAL WORLD*, *THE SOCIAL WORLD*. Esse tipo de organização não apenas se demonstra filosoficamente mais alinhado ao nosso trabalho – na medida em que coloca o homem no centro do processo de conceptualização, como intersecção dos mundos físico, psíquico e social –, como também, e isso é o mais importante, demonstrou trazer grandes benefícios lexicográficos, como veremos ao final da seção. Na parte dedicada ao domínio linguístico-computacional, resgataremos o paradigma da Teoria das Redes com outro foco, mais amplo, relacionando-o com o conceito de **rizoma** (DELEUZE, GUATTARI, 1976).

A Seção 5 propõe a ordenação dos grupos de analógicos por meio de critérios de relevância. Ou seja, dado um domínio léxico-conceitual ou uma categoria de objetos, quais são os conceitos lexicalizados que devem ser apresentados em primeiro lugar, em detrimento de outros? Para enfrentar esse desafio, trabalharemos com a **Teoria dos Protótipos** (ROSCH, 1975) no domínio linguístico e, no domínio linguístico-computacional, apresentamos a **Teoria dos Conjuntos Difusos** (ZADEH, 1965), mostrando como ela pode ser aplicada na modelagem dos CdSAs.

Na Seção 6, propomos e defendemos a elaboração do constructo *anaset*. A primeira parte discorre sobre as componentes e a estrutura do constructo, elaborado especificamente para os fins desta tese. A segunda parte apresenta os domínios léxico-conceituais a serem modelados como exemplo.

Seguindo o que foi definido nas seções anteriores na construção do modelo para o DAD, a Seção 7 executa os procedimentos para chegar aos resultados projetados, comentando cada etapa. Ao final da Seção 7 vamos discorrer sobre a integração do DAD com o Dicionário Criativo e discutir questões de interface e implementação computacional.

SEÇÃO 2 – OS DICIONÁRIOS ANALÓGICOS

Nesta seção, faremos uma análise crítica do estado da arte dos *dicionários analógicos*, procurando elucidar a origem, o funcionamento e a utilidade desse tipo de obra lexicográfica e, dessa maneira, prover o leitor com as informações necessárias sobre o estágio de desenvolvimento do objeto que deu origem a esta pesquisa. Sendo assim, esta Seção dá início à construção de hipóteses, mas é anterior ao processo de resolução de problemas que caracterizarão o restante da tese. Dessa forma, esta é a única Seção que faz parte do desenvolvimento da tese que não está dividida entre os domínios linguístico e linguístico computacional.

Na Subseção 2.1, os *dicionários analógicos* são apresentados como obras que possuem um caráter **onomasiológico**. Dentro da tradição lexicográfica, um dicionário onomasiológico é aquele em que cada entrada é o rótulo de uma categoria ou domínio conceitual e, no interior da entrada, registram-se as ULs que, de alguma forma, são consideradas instâncias do conceito rotulado, em oposição à estruturação dos dicionários semasiológicos, em que cada entrada é um item léxico e, no seu interior, registram-se os diferentes conceitos instanciados pelo item léxico. Em seguida, apresentamos a origem dos *dicionários analógicos* e a sua designação equivalente em outros idiomas. Em especial, faremos a distinção entre os significados do termo *thesaurus* que, tanto em inglês quanto em português, é empregado em diferentes áreas do conhecimento com diferentes nuances de significado.

A Subseção 2.2 discorre sobre a estrutura típica de um dicionário analógico, com suas quatro componentes características: o **plano de classificação das ideias**, o **quadro sinóptico de categorias**, os **grupos de palavras** (ou **grupos de analógicos**) e o **índice remissivo**.

A Subseção 2.3 discorre sobre a aptidão de um *dicionário analógico* e ilustra sua função como obra lexicográfica. Faz-se a distinção entre **dicionários de compreensão** e **dicionários de produção** (CORREIA, 2009) e inscreve-se os *dicionários analógicos* nesta última categoria, dado que sua função principal é auxiliar o processo de produção da escrita.

Na esteira dessa discussão, a Subseção 2.4 expõe a relação entre os *dicionários analógicos* e a escrita criativa; e demonstra como esses dicionários têm sido amplamente

utilizados por escritores, poetas e letristas no momento em que estão produzindo suas composições literárias.

Na Subseção 2.5, vamos analisar as diferenças e semelhanças entre dicionários analógicos e dicionários de sinônimos – em termos tanto macroestruturais como de conteúdo lexical. Essa discussão será de grande importância para a nossa proposta para um dicionário analógico digital organizado com critérios de relevância, apresentada com profundidade na Seção 6.

Os dicionários analógicos e *thesaurus* que já se encontram disponíveis na internet, aproveitando os recursos do meio digital, serão analisados na Subseção 2.6, em separado das obras impressas, pois entre eles existe grande disparidade com relação à forma de produção, critérios, interface gráfica e abrangência; muito mais do que se verifica nos produtos em formato de livro. É de se ressaltar que dentre os idiomas inglês, francês, espanhol e português, que fazem parte do nosso escopo nesta seção, apenas os dois primeiros possuem versões eletrônicas disponíveis na internet com algum grau de relevância acadêmica ou de tradição lexicográfica.

No final da seção, em 2.7, faremos uma análise crítica do estado da arte desse tipo de obra lexicográfica e, a partir dessa análise, iremos sugerir avanços, tendo sempre em vista a construção de um modelo digital para o PB. Para tanto, procederemos a uma análise da forma e do conteúdo de dicionários analógicos do português, inglês, francês e espanhol publicados em papel ou disponíveis em meio digital, com vistas à delimitação do objeto e suas características, bem como da avaliação de seus conteúdos como fonte de conhecimento¹².

Os dicionários analógicos (ou *thesaurus*) impressos que foram analisados para esta Seção são: i) em inglês, o *Thesaurus of English words and phrases*¹³ (ROGET, [1852] 1911), por ser o precursor desse tipo de obra, e o *Roget's Internacional Thesaurus* (KIPFER, B. A. [ed.], 2010), 70ª edição do *Roget's Thesaurus* original que, além de ter seu conteúdo revisto e atualizado, teve sua macroestrutura sensivelmente modificada; ii) em francês, o *Dictionnaire idéologique: recueil des mots, des phrases, des idiotismes et des proverbes de la langue française classés selon l'ordre des idées*¹⁴ (ROBERTSON, 1859),

¹² A seleção dos dicionários analógicos como fontes de conhecimento léxico-conceitual e o estabelecimento de critérios de filtragem da informação serão, de fato, abordadas na Seção 6, junto com as demais obras de referência que servirão a esse fim. Na Seção 2, faremos uma análise geral e uma filtragem parcial.

¹³ Versão digitalizada pelo Projeto Gutenberg a partir da obra publicada em 1911 pela editora *Crowell Co.*

¹⁴ Versão digitalizada pela *Google* a partir da obra original publicada em 1859 pela editora *A. Derache*; disponível no endereço www.books.google.com.

pioneiro na França; iii) em espanhol, o *Diccionario Ideológico de la Lengua Espanola* (CASARES, 1959), principal referência desse tipo de obra para o idioma espanhol; e iv) em português, o *Dicionário Analógico da Língua Portuguesa*¹⁵ (SPITZER, [1936] 1952), precursor no Brasil, e o *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, [1950] 2010), o mais representativo em língua portuguesa e com edição revista e atualizada pela editora Lexikon.

2.1 O QUE É UM DICIONÁRIO ANALÓGICO?

Ao contrário dos dicionários padrões de língua¹⁶, que partem dos itens lexicais para o registro de seus significados em diferentes acepções, os *dicionários analógicos* partem de uma estrutura conceitual, sob a qual organizam o léxico. Ou seja, o percurso se faz do nível conceitual para o nível lexical, sendo considerado, portanto, de ordem **onomasiológica** (CORREIA, 2009). Vejamos, abaixo, por exemplo, a introdução do *Dictionnaire idéologique* escrita pelo próprio autor, na ocasião do lançamento da obra em 1859:

Le problème que les dictionnaires ordinaires se chargent de résoudre est celui-ci : — Un mot étant donné, trouver sa signification, ou l'idée qu'il représente. Le problème dont le *Dictionnaire idéologique* doit fournir la solution est exactement l'inverse de celui qui précède :— Une idée étant donnée, trouver le mot qui l'exprime le plus convenablement. A cet effet, les mots et les phrases de la langue sont classés dans ce dictionnaire, non selon leur prononciation ou leur orthographe, mais strictement selon leur *signification*.¹⁷ (ROBERTSON, 1859, *introduction*)

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 adaptado para

¹⁵ Fora de catálogo e sem versão digitalizada, o volume aqui utilizado foi adquirido em um sebo.

¹⁶ Por exemplo, os dicionários *Novo Aurélio Século XXI* (FERREIRA, 1999); *Michaelis*: moderno dicionário da língua portuguesa (WEISZFLOG, 1998); e *Dicionário Houaiss da Língua Portuguesa* (HOUAISS, VILLAR, 2001).

¹⁷ O problema que os dicionários comuns são responsáveis por resolver é o seguinte: para uma palavra dada, encontrar o seu significado ou a idéia que ela representa. O problema para o qual um *Dicionário Ideológico* deve fornecer a solução é exatamente o inverso: - Para uma idéia que está dada, encontrar a palavra que a expresse mais adequadamente. Para esse fim, as palavras e frases da língua são classificadas nesse dicionário, não de acordo com sua pronúncia ou grafia, mas estritamente de acordo com seu *significado* (disponível em <http://projects.chass.utoronto.ca/langueXIX/robertson/>, acessado em julho de 2011).

diversas línguas, e é utilizado como modelo para várias obras lexicográficas. A denominação *Dicionário Analógico* ganhou a preferência nos países de língua latina, enquanto que, nos países de língua inglesa, a tendência foi adotar a 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 segundo sua natureza, conforme mostra o quadro 2.1.

Quadro 2.1. As noções de *thesaurus*.

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 exhibe 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.

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. 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 desse 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 maioria dos lexicógrafos e pesquisadores da área evita essa 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, [1950] 2010), que também é uma adaptação do *Roget's Thesaurus*. Meticulosamente compilado, em vinte anos de trabalho, pelo

professor goiano Francisco dos Santos Azevedo, a obra foi publicada pela primeira vez em 1950 e relançada em 2010, em edição revista e atualizada, pela editora Lexikon, sob a coordenação editorial do lexicógrafo Paulo Geiger.

2.2 COMO SE ESTRUTURAM OS DICIONÁRIOS ANALÓGICOS?

A macroestrutura dos dicionários analógicos define-se por quatro componentes características. A primeira é o **plano de classificação das ideias**, que é a ossatura do dicionário e em que estão listadas as categorias conceituais mais altas dentro da hierquia conceitual da obra. O plano de classificação das ideias é uma estrutura conceitual elaborada conforme uma ontologia de bases aristotélicas. A segunda componente, em que as categorias conceituais listadas no plano de classificação são desmembradas em categorias menores e mais específicas, é constituída pelo **quadro sinóptico de categorias** (Figura 2.1).

QUADRO SINÓPTICO DE CATEGORIAS

CLASSE I. RELAÇÕES ABSTRATAS

Seção I — EXISTÊNCIA.

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

Seção II — RELAÇÃO.

1.) Absoluta....	9. Relação	10. Não relação
	11. Consanguinidade	
	12. Correlação	
	13. Identidade	14. Contraste
2.) Contínua....	16. Uniformidade	15a. Desconformidade
	17. Semelhança	18. Desequilíbrio
3.) Parcial....	19. Imitação	20. Originalidade
	21. Cópia (resultado da imitação)	20a. Variabilidade
4.) Geral....	23. Acordo	22. Protótipo (coisa que serve de cópia)
		24. Desacordo

Seção III — QUANTIDADE.

1.) Simples....	25. Quantidade	26. Grau
	27. Igualdade	28. Desigualdade
		29. Média
		30. Compensação
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. Adição	38. Subtração
	39. Adjunto (coisa acrescentada)	40. Resto (coisa restante)
3.) Subordinada	41. Mistura	42. Pureza
	43. Junção	44. Disjunção
	45. Vínculo	46. Incoerência
	48. Coesão	47. Incoerência
	49. Combinação	49. Decomposição

- IX -

Figura 2.1 **Quadro Sinóptico de Categorias**; *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, 1983, p. IX).

A terceira componente é o dicionário propriamente dito, composto dos grandes **grupos de palavras** (Figura 2.2) que se arrolam para cada uma das categorias

conceituais que figuram como entrada no quadro sinóptico de categorias. Em cada grupo de palavras são indicados, em negrito, o conceito central do grupo como um todo, seguido por sinônimos e analogias. Cada página é dividida em duas colunas, em que cada uma apresenta grupos de palavras antagônicos. Na Figura 2.2, por exemplo, temos os grupos de palavras cujas entradas são EXISTÊNCIA, na coluna à esquerda, e INEXISTÊNCIA, na coluna à direita. 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.

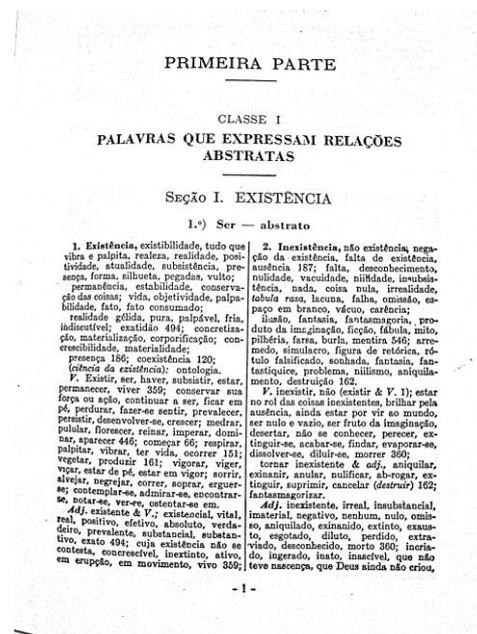


Figura 2.2. Os grupos de palavras; *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, 1983, p. 1).

O critério adotado para a reunião das palavras em grupos é o princípio da **analogia**. 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 as palavras presentes no dicionário. Essa componente estrutural do dicionário muda de obra para obra. Em alguns dicionários ela se caracteriza como um **índice remissivo**; em outros, além do índice, são apresentadas definições das diferentes acepções da palavra em questão.

Nos dicionários analógicos, o principal ponto de partida das consultas deveria ser, complementarmente, a primeira e a segunda componentes, ou seja, o plano de classificação e o quadro sinóptico. Como dissemos, o princípio de funcionamento desse 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. No entanto, cf. veremos em 2.7, constatou-se que a consulta por meio do índice remissivo é a mais comum entre os consulentes desse tipo de obra. Procura-se por uma palavra no índice que o levará para um grupo de analógicos em que a palavra ou expressão buscada possa ser encontrada via sinonímia, analogia ou mesmo antonímia, que estará em um grupo de analógicos vizinho. Assim, continuamente, o percurso pode assumir características diferentes, permitindo a passagem de um grupo de palavras a outros adjacentes.

Há ainda as microestruturas e estruturas de nível intermediário. Além dos grupos de analógicos propriamente ditos, uma UL pode estar contida dentro de uma **lista** (*word lists*). As listas podem tanto estar no nível estrutural de um grupo de analógicos (com numeração própria), como pode estar subordinada a um deles. Elas não arrolam sinônimos ou analógicos, mas grupos de objetos nos quais os membros se relacionam paradigmaticamente de maneira mais ou menos estruturada. Uma lista pode inventariar todos os feriados nacionais ou a estrutura de parentescos por consanguinidade conforme o Direito Civil Brasileiro. Pode, igualmente, catalogar os instrumentos musicais ou elencar os termos pejorativos de cada profissão. Conforme a apresentação da 70ª edição do *Roget's Thesaurus*, “The **lists** can save you many excursions to specialized reference books”¹⁸.

Algumas ULs também são anotadas com rubricas para situá-la adequadamente com informação relativa a especificidades de uso. Na maioria das obras analisadas, as rubricas podem indicar: i) **estrangeirismos** (<lat> Latim, <fr> Francês); ii) **frequência de uso** (<desus> Desusado, <p. us.> Pouco usado); iii) **área do conhecimento** (<quim> Química, <vet> Veterinária); **variação temporal** (<ant> Antigo, <neol> Neologismo); **figuras de linguagem** (<euf> Eufemismo, <fig> Figurado); e **contextos de uso** em geral (<pop> Popular, <pej> Pejorativo).

No nível microestrutural têm-se também a separação de alguns agrupamentos semânticos pelo sinal gráfico do **ponto-e-vírgula** (;) e pela **quebra de linha**. Nas obras

¹⁸ “As **listas** podem poupar-lhe muitas excursões a obras de referência especializadas”.

que seguem a tradição do *Roget's Thesaurus*, as palavras com forte relação semântica se separam de outros agrupamentos com alguma variação de sentido pelo sinal do ponto-e-vírgula (*semicolons*) e pelas quebras de linha e formação de novos parágrafos. Tais recursos sinalizam uma ligeira mudança no sentido ou na aplicação das palavras. O uso do ponto-e-vírgula e as quebras de linha parece ser uma questão de gradação. Se a mudança de sentido é suave, procurou-se utilizar o ponto-e-vírgula, ao passo que uma mudança um pouco mais contundente, optou-se pela quebra de linha. No entanto, cf. veremos em 2.7, os critérios para a formação de tais agrupamentos, e mesmo para o uso de cada recurso gráfico, não estão explicitados na apresentação de nenhuma das obras e, mesmo depois de uma análise cuidadosa, não foi possível reconhecer um padrão estável e confiável para os agrupamentos. O tópico “como usar este dicionário”, da nova edição do Dicionário Analógico (AZEVEDO, 2010, p. xi), diz apenas que “os grupos não têm uma estrutura lógica, embora as palavras estejam, geralmente, agrupadas por proximidade semântica”.

2.3 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, p. 44) 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”. Esse tipo de obra é também chamada, por vezes, de **dicionários ativos** ou **codificadores** (CORREIA, 2009), 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 desse tipo: *classificado e ordenado de modo a facilitar a expressão de ideias e auxiliar na*

*composição literária*¹⁹. Na introdução da obra, 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, [1852] 1911).

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 (WERNECK *apud* HOLLANDA, 2006). Em prefácio para a nova edição da obra, Chico Buarque escreveu: “E por um bom tempo aquele livro me ajudou no acabamento de romances e letras de canções [...]. Palavra puxa palavra, e escarafunchar o dicionário analógico foi virando para mim um passatempo (desenfado, esparecimento, entretém, solaz, recreio, filistria)”.

2.4 DICIONÁRIOS ANALÓGICOS E CRIATIVIDADE

A literatura sobre criatividade é bastante ampla e rica de perspectivas. No entanto, não faz parte do escopo desta tese confrontar os diferentes pontos de vista ou aprofundar-se em algum deles. Tomamos aqui, apenas, alguns paradigmas consensuados ao longo dos estudos sobre o tema – que, inclusive, se perpetuam dentro da visão da psicologia cognitivista mais moderna – para falar sobre a relação entre criatividade e analogia.

O matemático francês Henry Poincaré dizia que toda ideia é uma analogia, ou seja, uma ligação nova entre conceitos antes desconexos. Parece consenso entre os estudiosos da criatividade que as pessoas criativas demonstram uma particular habilidade em encontrar semelhanças escondidas e associar ideias que, para o olhar rotineiro, parecem impossíveis de se combinar (ZUGMAN, 2008). Os dicionários

¹⁹ “[...] classified and arranged so as to facilitate the expression of ideas and to assist in literary composition [...]”

análogos favorecem a criatividade justamente por terem sido concebidos tendo por base o conceito de analogia. Retomando a apresentação do *Roget's Thesaurus*, “a revisão de um catálogo de palavras de significado análogo vai sugerir, com frequência, por associação, outras sucessões de pensamento”. Esse encadeamento de ideias é o motor da criatividade. Pessoas criativas tendem a não enxergar barreiras entre dois ou mais domínios distintos. Conforme Zugman (2008, p. 65), “Pessoas criativas parecem boas em encontrar semelhanças ocultas, criar incongruências controladas, associar coisas que, antes delas, pareciam impossíveis de se misturar”.

No livro *De onde vêm as boas ideias*, de Steven Johnson (2011), o autor nos conta sobre o costume iluminista de se manter um “livro de citações”. Praticamente todos os pensadores, eruditos e homens das letras dos séculos XVII e XVIII tinham um livro desse tipo. Segundo Johnson (p. 73), “em sua forma mais comum, a prática envolvia a transcrição de passagens interessantes ou inspiradoras das obras lidas, reunindo uma enciclopédia personalizada de citações”. Grandes nomes desse período, como Francis Bacon e John Locke, eram entusiastas dos livros de citações como uma ferramenta para avivar a memória. No entanto, o sistema de indexação de um livro de citações, como se pode imaginar, guarda uma tensão entre ordem e caos, “entre o desejo de arranjo metódico e o de surpreender novos elos de associação” (p. 74). Locke sugeriu um método de indexação que, ao mesmo tempo em que proporcionava a ordem suficiente para encontrar as citações buscadas, ao mesmo tempo favorecia um ruído randômico que suscitava novas conexões. Johnson dá o exemplo do método de Locke para concluir:

Impor ordem demais é correr o risco de deixar uma intuição promissora órfã num projeto mais amplo que morreu, e torna difícil para essas ideias misturarem-se e procriarem quando revisitamos. Precisamos de um sistema para capturar intuições, mas não necessariamente categorizá-las, porque categorias podem erigir barreiras entre ideias díspares, restringi-las às suas próprias ilhas conceituais. (p. 75)

Grande parcela dos indivíduos tende a reagir de forma semelhante a um mesmo influxo. No entanto, eles diferem na maneira como organizam as informações de *input* na sua estrutura mental. Há quem as disponha em hierarquias restritas, ou seja, para esses indivíduos, determinado estímulo possui ligações fortes e privilegiadas a uma quantidade fixa e finita de outros estímulos. Isso quer dizer que as informações são rigorosamente alocadas em diferentes compartimentos (em termos cognitivos, podemos falar em categorias fechadas e altamente hierarquizadas). Segundo Sarnoff Mednick (1962), professor de psicologia norte-americano, essa é a diferença fundamental que

separa pessoas muito criativas das pouco criativas. Ou seja, enquanto a pouco criativa possui uma íngreme hierarquia entre seus conceitos, a mais criativa desenvolve uma arquitetura conceitual relativamente mais plana, mais horizontalizada, em que os conceitos possuem muitas associações diretas uns com os outros. Nas Figuras 2.3 e 2.4, extraídas de Zugman (2008, p. 70), as letras representam conceitos de um mesmo domínio. Na Figura 2.3 vemos alguns grupos de conceitos associados (ABFG e MQRS, por exemplo). As linhas grossas representam uma forte associação entre esses conceitos e, portanto, quando o conceito A é ativado, evoca, de imediato, os conceitos B e F e, por meio de F, acessa G. No entanto, partindo de A, não temos acesso, por exemplo, ao grupo de conceitos CDI ou ao conceito L. Ou seja, em ordenamentos altamente hierarquizados, os conceitos são mantidos em diferentes compartimentos, o que dificulta os processos criativos.

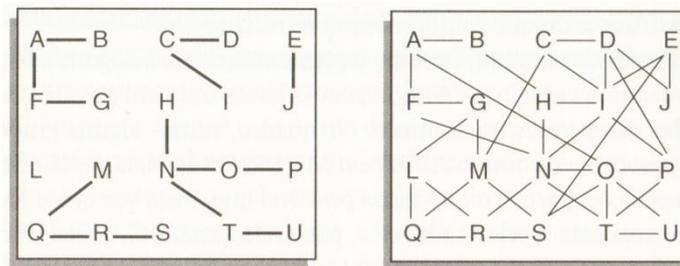


Figura 2.3 e 2.4. Associações altamente hierarquizadas e associações planas (ZUGMAN, 2008, p. 70).

Na Figura 2.4, pelo contrário, as associações entre os conceitos são mais fracas – representadas pelas linhas finas –, mas o número de associações entre os conceitos são muito maiores. Conceitos anteriormente distantes agora estão mutuamente acessíveis, facilitando assim conexões imprevistas e criativas.

Esse aspecto particular dos estudos sobre a criatividade humana é fundamental para o que será discutido e proposto na Seção 4 com relação à estrutura hierárquica do nosso modelo. É também o mesmo apanágio que diferencia com clareza o trabalho desenvolvido nesta tese de propostas como a das *wordnets*. A arquitetura das *wordnets*, fortemente hierarquizada, tem a vantagem de ser mais facilmente manipulável por máquina, o que a torna reutilizável em um grande número de aplicações tecnológicas. No entanto, essa mesma configuração perde força diante do objetivo de estimular conexões entre conceitos que propiciem a criatividade do usuário humano. É nesse sentido que na Seção 4 fazemos a opção por aplicar a estrutura de **rizoma** (DELEUZE,

GUATTARI, 1976) no nosso modelo de dicionário analógico digital, em detrimento da estrutura em árvore taxonômica das *wordnets*.

2.5 DICIONÁRIOS ANALÓGICOS *VERSUS* DICIONÁRIOS DE SINÔNIMOS

Dicionários analógicos e dicionários de sinônimos possuem utilidades parecidas. De maneira geral, os dois tipos de obra servem para que o consulente encontre palavras que melhor expressem o teor de seu pensamento, procurando por variações de forma e matizes de significado, seja por motivos de precisão conceitual, seja por ensejos estilísticos. Um poeta que esteja terminando um verso pode consultar qualquer uma das duas obras para procurar uma palavra que expresse o significado desejado e que, ao mesmo tempo, possua uma forma que seja sonoramente adequada para o pé do estribilho, por exemplo. Da mesma forma, um advogado que esteja redigindo um contrato poderá consultar uma das obras para encontrar, dentre as matizes de significado, aquele que expresse o conteúdo conceitual pretendido com menor ambiguidade e que, ao mesmo tempo, a palavra lhe pareça adequadamente formal ou erudita.

No entanto, tanto o conteúdo quanto a estrutura desses dicionários possuem diferenças. Elas serão analisadas a seguir. Começamos destacando as semelhanças e diferenças reconhecidas pelos próprios autores ou organizadores dos dois tipos de obra. Vejamos, por exemplo, o que diz Francisco Fernandes na apresentação da primeira edição do seu *Dicionário de Sinônimos e Antônimos da Língua Portuguesa* (1945, sem página):

É sabido que os sinônimos perfeitos são pouco numerosos; muitas vezes a sinonímia se estabelece apenas entre determinadas acepções translatas comuns a duas ou mais palavras; outras vezes o termo que se dá como sinônimo de outro a esse não se liga senão por simples vínculo analógico. Assim, a um dicionário como esse mais propriamente se deveria chamar ANALÓGICO; mas esse adjetivo designa hoje outra espécie de léxico, motivo por que se manteve o título DICIONÁRIO DE SINÔNIMOS. Para justificá-lo, entretanto, procurei, sempre que possível, dispor os sinônimos de cada verbete em grupos mais intimamente ligados entre si, o que facilita o trabalho da consulta e é o critério adotado pelos melhores autores de obras do gênero.

De fato, a constatação de que sinônimos perfeitos – ou absolutos – são raros no léxico de uma língua é compartilhada pela maioria dos linguistas e lexicógrafos que estudam o assunto ou que com ele se deparam. No seu livro *Introdução à linguística*

teórica, por exemplo, Lyons afirma que “há poucos sinônimos perfeitos nas línguas naturais” (LYONS, 1979, p. 476). Cruse (1986) vai mais longe e assinala que, para o bem da verdade, é impossível provar que duas palavras possuam significados idênticos e identidade conceitual em todos os contextos em que podem ser aplicadas, pelo simples fato de que é impossível atestar e confrontar os significados atualizados em um número teoricamente infinito de contextos.

O número reduzido de sinônimos perfeitos é o motivo pelo qual um dicionário de sinônimos que levasse a cabo o seu nome seria de muito pouca utilidade. Daí a necessidade de construir um repertório que não arrole apenas pares de sinônimos perfeitos, mas que introduza também relações sinonímicas parciais, tais como o que Cruse (op. cit., p. 270) chama de **sinônimos cognitivos** (unidades lexicais que compartilham certas propriedades semânticas) e as unidades lexicais de sentido próximo ou vizinho, o que Francisco Fernandes chamou de vínculos analógicos. Dessa forma, podemos dizer que a fronteira que separa os dois tipos de obras, principalmente com relação aos critérios para incluir esta ou aquela palavra nos seus verbetes, é uma fronteira difusa. Ao justificar o seu dicionário de sinônimos, contrapondo-o a um dicionário analógico, o professor Fernandes diz que seu empenho foi no sentido de “dispor os sinônimos de cada verbeito em grupos mais intimamente ligados entre si”. Evidentemente, esse critério é maleável, graduável e completamente dependente da subjetividade daqueles que se aventuram a compilar obras desse tipo, já que são os autores que, em última instância, vão determinar quais palavras e sentidos lhe parecem mais relacionados entre si.

Essa superposição entre o conteúdo e o aspecto dos dois tipos de dicionários foi também uma preocupação do responsável pelo projeto do *Dicionário Houaiss de Sinônimos e Antônimos*, como reconhecido por ele em determinado momento do texto que introduz a obra (HOUAISS, 2008, p. xii):

Evitamos incluir em nossos repertórios as chamadas *palavras análogas*. [...] O dicionário foi projetado para não se aproximar demais da metodologia dos léxicos ideológicos, também ditos analógicos ou de ideias afins, que o fazem extensivamente.

Tal preocupação reforça o nosso argumento colocado acima de que a distinção entre sinônimos e análogos, em certo sentido, é nebulosa. No entanto, como atestam os dois autores, é possível estabelecer critérios que distingam os dois tipos de obra. Para tanto, a maioria dos dicionários de sinônimos adotam o conceito de **sinonímia**

contextual. De acordo com essa noção de sinonímia, “duas unidades lexicais são sinônimas em um contexto C, se a substituição de uma pela outra em C não altera o valor de verdade denotado por C” (FELLBAUM, 1998). Tal noção atrela a condição de permutabilidade entre as palavras que se pretendem sinônimas a um contexto específico, o que amplia enormemente a abrangência de um dicionário de sinônimos ao mesmo tempo em que restringe os casos em que duas palavras possuem sentidos muito próximos, mas que não podem nunca, em contexto nenhum, trocarem-se uma no lugar da outra. Como benefício suplementar, em decorrência da adoção da sinonímia contextual, os verbetes passam, evidentemente, a se organizar internamente em contextos. Dessa forma, cada conjunto de sinônimos aponta, idealmente, para um conceito unívoco. Vejamos, por exemplo, o verbete **amor** no *Thesaurus Eletrônico para o Português do Brasil* (doravante **TeP**)²⁰:

1. amor, paixão
2. amor, adoração, culto, devoção, tenção, veneração
3. amor, aplicação, atenção, carinho, cuidado, dedicação, desvelo, diligência, vigilância, vigília, zelo
4. amor, afeição, afeiçoamento, afetividade, afeto, amizade, querença, querência

No entanto, ao comparar os CSs de cada verbete entre os diferentes dicionários de sinônimos, vemos que as acepções raríssimas vezes são equivalentes. Ou seja, as acepções – que deveriam equivaler a um conceito unívoco – também são difusas. O que é para uma obra apenas uma acepção pode ser dividida em duas ou mais acepções em outras obras. Um mesmo domínio conceitual é recortado de maneiras diferentes a partir de metodologias e concepções diferentes.

A maioria dos dicionários de sinônimos da língua portuguesa disponíveis reproduz os mesmos sinônimos já contemplados pelos dicionários tradicionais, de significados. Ou seja, ficam de fora as variações sutis e os significados similares que poderiam, com maior criatividade e/ou fidelidade, expressar uma ideia ou conceito. Como exemplo, tomemos a unidade lexical *afeição*, primeiramente dentro dos principais dicionários de significados da língua portuguesa, destacando os termos que podem ser reconhecidos como sinônimos:

UL: afeição

²⁰ Como veremos na próxima Seção o TeP é um *thesaurus* no sentido de dicionário de sinônimos que usa o conceito de sinonímia contextual.

iDicionário Aulete²¹ = 1 Sentimento de **afeto**, **carinho**, **apego** por algo ou alguém 2 **Inclinação**, **tendência**, **pendor** para algo.

Michaelis²² = 1 **Afeto** 2 **Simpatia** 3 Sentimento de **amor** 4 **Amizade** 5 **Pendor** para alguém ou alguma coisa.

Houaiss²³ = 1 ligação afetiva; sentimento amoroso em relação a; **afeto**, **afeiçoamento** 2 **inclinação**, **pendor** para alguma coisa

Agora, vejamos como a mesma UL é contemplada pelos principais dicionários de sinônimos disponíveis²⁴ e no TeP:

{afeição, afeto, amor, simpatia, admiração, estima} [MIC]

{afeição, afeto, amizade, amor, benquerença, carinho, inclinação, propensão, simpatia} [FER]

{afeição, adoração, afeto, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU]

{afeição, admiração, afeto, inclinação, simpatia} [TEP]

De fato, podemos observar que os dicionários de sinônimos se limitam a elencar, a cada entrada, os sinônimos mais próximos e costumeiros, na sua maioria, já contemplados pelos dicionários de significados. Se somarmos esses conjuntos, vamos obter 21 sinônimos diferentes para a UL em questão. Em contrapartida, um dicionário analógico, como o *Dicionário Analógico da Língua Portuguesa* do professor Francisco dos Santos Azevedo [AZE], não possui entrada específica para a UL *afeição*, mas tem ela inserida no grupo de analógicos referente ao verbete **897. Amor**, que arrola 53 sinônimos e analógicos mesmo antes da primeira quebra de linha²⁵, como podemos ver a seguir:

{afeição, carinho, idolatria, afeto, amoricos, amorosidade, amorio, inclinação, dileção, predileção, preferência, simpatia, estremecimento, benquerença, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência, agarramento, entranha, aspiração, galanteio, galanice, namoro, namorico, namorisco, flerte, amizade colorida, paixão, adoração, ardor, fervor, chamas, calor, devoção, atração, êxtase, enlevamento, arroubamento, enlevo, feitiçaria, namoramento, xaveco, ...} [AZE]

²¹ Disponível em <http://aulete.uol.com.br>, acessado em 28 de novembro de 2011.

²² Disponível em <http://michaelis.uol.com.br>, acessado em 28 de novembro de 2011.

²³ Dicionário Houaiss da Língua Portuguesa, versão eletrônica 3.0.

²⁴ Principais dicionários de sinônimos impressos e disponíveis em língua portuguesa que se valem de sinonímia contextual: i) *Michaelis Dicionário de Sinônimos e Antônimos* (POLITO, 2009) = [MIC]; ii) *Dicionário de Sinônimos e Antônimos da Língua Portuguesa* (FERNANDES, 1945) = [FER]; iii) *Dicionário Houaiss: Sinônimos e Antônimos* (HOUAISS, 2008) = [HOU].

²⁵ As quebras de linha, cf. 2.2 e 2.7, separam as ULs em conjuntos com maior afinidade semântica. No entanto, os critérios utilizados para essa divisão parecem ser elásticos e tem sido difícil encontrar um padrão. Tarefa dificultada pela omissão dos autores com relação a informações desse respeito.

Esse grupo de analógicos compreende quase todas as ULs presentes nos CSs somados. Dos 21 sinônimos resultantes da reunião de todos os conjuntos, apenas três não estão contempladas no grupo de analógicos. No entanto, daí não se pode concluir que os dicionários analógicos se comportem como dicionários de sinônimos melhorados ou mesmo mais completos. Eles são objetos diferentes. Como dissemos, da maneira como existem hoje, podem ser usados em conjunto ou cada qual a sua hora, de acordo com as necessidades do usuário. Em certa medida, os dois possuem vantagens e desvantagens.

Enquanto os dicionários de sinônimos – que adotam a noção de **sinonímia contextual** – agrupam ULs que lexicalizam um e apenas um conceito, os grupos de palavras de um dicionário como o do Azevedo misturam diferentes conceitos lexicalizados em seu interior. A desvantagem é que esses grupos de palavras agigantam-se e não se estruturam internamente, o que dificulta a consulta do usuário. Se o usuário pretende apenas substituir a palavra *afeição* dentro de um determinado contexto, correr os olhos por uma listagem de dezenas de termos e topar com palavras como *namoro* e *flerte* – que efetivamente não substituem a palavra *afeição* em provavelmente nenhum contexto – torna-se uma experiência pouco produtiva. Com relação aos dicionários de sinônimos, a desvantagem que se apresenta é diametralmente oposta. *Namoro* e *flerte* não podem trocar de lugar com a palavra *afeição* numa frase, mas lexicalizam conceitos fortemente relacionados com ela. Como os dicionários de sinônimos são organizados em ordem alfabética, e não por domínios léxico-conceituais, ao se consultar por *teatro*, p. ex., não se tem acesso a *ator* ou *figurino* ou *cenário*. Ou seja, unidades lexicais fortemente relacionadas não estão mutuamente acessíveis, recurso importante para o caso do usuário não estar apenas procurando uma palavra substituta, mas estar a construir toda uma paisagem.

A solução de apenas utilizar os dois diferentes objetos em concomitância acaba por gerar uma inconveniente duplicação da informação (*overlap*). Dessa forma, a solução mais arrazoada para o problema parece ser mesclar o mais vantajoso de cada estrutura, usando os CSs para recortar e organizar os grupos de analógicos. Desse encontro, nasce

o constructo *anaset*. O desenvolvimento do constructo e de seu procedimento de construção serão tema da Seção 6.

2.5.1 O *THESAURUS* ELETRÔNICO PARA O PORTUGUÊS DO BRASIL (TEP)

O TeP é o único *thesaurus*²⁶ de língua geral computacionalmente implementado e disponível na internet para o português brasileiro (PB). Ele já está em sua segunda versão, TeP 2.0, e pode ser acessado pelo endereço <http://www.nilc.icmc.usp.br/tep2/index.htm>.

O TeP pode ser considerado, em última instância, como uma ferramenta de auxílio à escrita, na medida em que oferece ao usuário final, por meio de interface própria (Figura 2.5), a conveniência de procurar por palavras sinônimas e antônimas que ele, por questões de estilo e de precisão, deseja substituir durante a redação de um texto em língua portuguesa (DIAS-DA-SILVA, 2000). Por outro lado, sua base de dados lexicais, distribuída gratuitamente para fins de pesquisa, tem se provado uma importante fonte de conhecimento lexical estruturado, contribuindo assim para o desenvolvimento de novas aplicações em PLN, como é o caso desta tese.

The screenshot shows the TeP 2.0 beta web interface. At the top, there is a header with the TeP 2.0 beta logo and logos for NILC, ICMC USP, and UNESP. Below the header, there is a search bar with the word "amor" entered and a "Buscar" button. Underneath the search bar, there are options to show examples and antonyms. The search results for "amor" are displayed, including the word "amor" (Substantivo) and four numbered definitions with their respective synonyms.

amor

Buscar

Todas Mostrar Exemplo Mostrar Antônimos

Quis dizer [amornar](#), [amorado](#), [amoral](#), [amorfo](#), [amoroso](#), [amorfia](#)?

amor (Substantivo)

1. **amor**, paixão

2. **amor**, adoração, culto, devoção, tenção, veneração

3. **amor**, aplicação, atenção, carinho, cuidado, dedicação, desvelo, diligência, vigilância, vigília, zelo

4. **amor**, afeição, afeiçoamento, afetividade, afeto, amizade, querença, querência

EXPANDIR TRIPLOS

Contato Ajuda Download da Base

²⁶ *Thesaurus* tipo 3: dicionário de sinônimos e antônimos.

Figura 2.5. Interface do TeP 2.0 para o usuário final

O TeP foi desenvolvido segundo os pressupostos da WN.Pr e, por isso, pauta-se na metodologia de construção de *synsets*, que já pressupõe o conceito de sinonímia contextual. A base de dados do TeP é utilizada na nossa pesquisa como fonte de conhecimento lexical e vai integrar a construção dos CdSAs que, relacionados, formarão os *anaset*s. Esse percurso é contemplado, no interior desta tese, na Seção 6.

Atualmente, o TeP 2.0 contém 44.678 unidades lexicais organizadas em 19.888 conjuntos de sinônimos, além de 4.276 relações de antonímia entre os conjuntos da base (DIAS-DA-SILVA *et al.*, 2008).

2.6 DICIONÁRIOS ANALÓGICOS NA INTERNET

O Quadro 2.2 apresenta uma lista de *thesaurus* e *dicionários analógicos* disponíveis na internet para os idiomas **inglês, francês e português brasileiro**. Foram identificados por meio de extensa investigação e selecionados, dentre outros existentes, por critérios de representatividade. Entre eles, há grande heterogeneidade com relação à forma de produção, critérios, interface gráfica, abrangência etc. Não especificaremos aqui cada uma dessas características, por não ser esse o enfoque da pesquisa.

Quadro 2.2 – Lista de *thesaurus* e *dicionários analógicos* disponíveis na internet, acessados em: 01 mar 2010.

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 teste.
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-idiomas; 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.

O *The Free Dictionary* se vale do conteúdo digitalizado da segunda edição do *Collins Thesaurus of the English Language* (Harper-Collins Publishers, 1995 [2002]). Já o *Thesaurus Reference* e o *SynonymFor* são, respectivamente, uma adaptação do *Thesaurus* de Roget e uma interface para a base de dados da WN.Pr²⁷. 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.

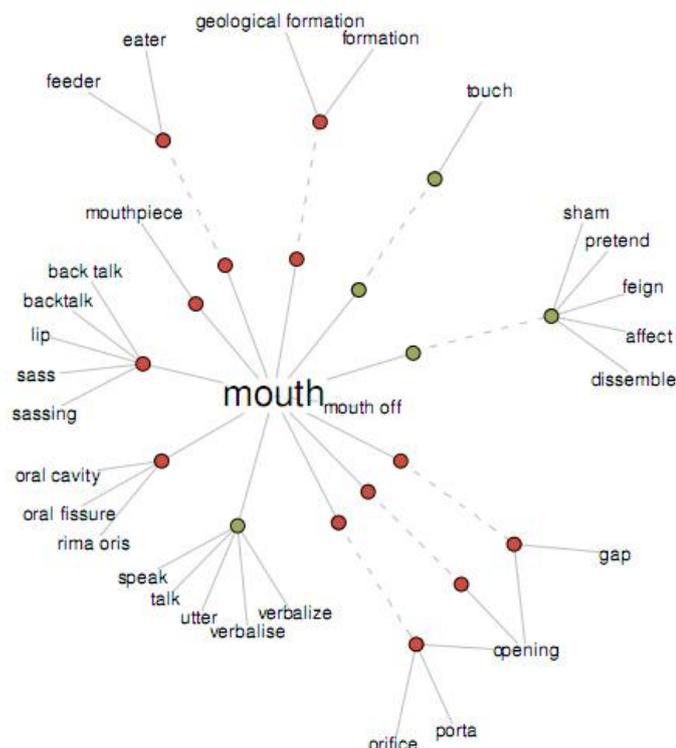


Figura 2.6. Resultado do *Visual Thesaurus* para a busca pela palavra *mouth*; acessado em 01 mar 2010.

²⁷ Disponível em: www.wordnet.princeton.edu, acessado em 01 mar 2010.

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 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 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. Falaremos mais sobre as redes *wordnets* na Seção 3.1.

Na língua francesa, se destaca o *Dictionnaire des Synonymes du CRISCO* (Centre de Recherche Inter-langues sur la Signification en Contexte). Apesar do nome, o *Dictionnaire des Synonymes* apresenta características de dicionário analógico, contendo 49.160 entradas e 200.649 relações entre elas. É também o único da lista que explicita uma ordem de relevância entre os sinônimos – sendo, inclusive, representada visualmente, como mostra a Figura 8.

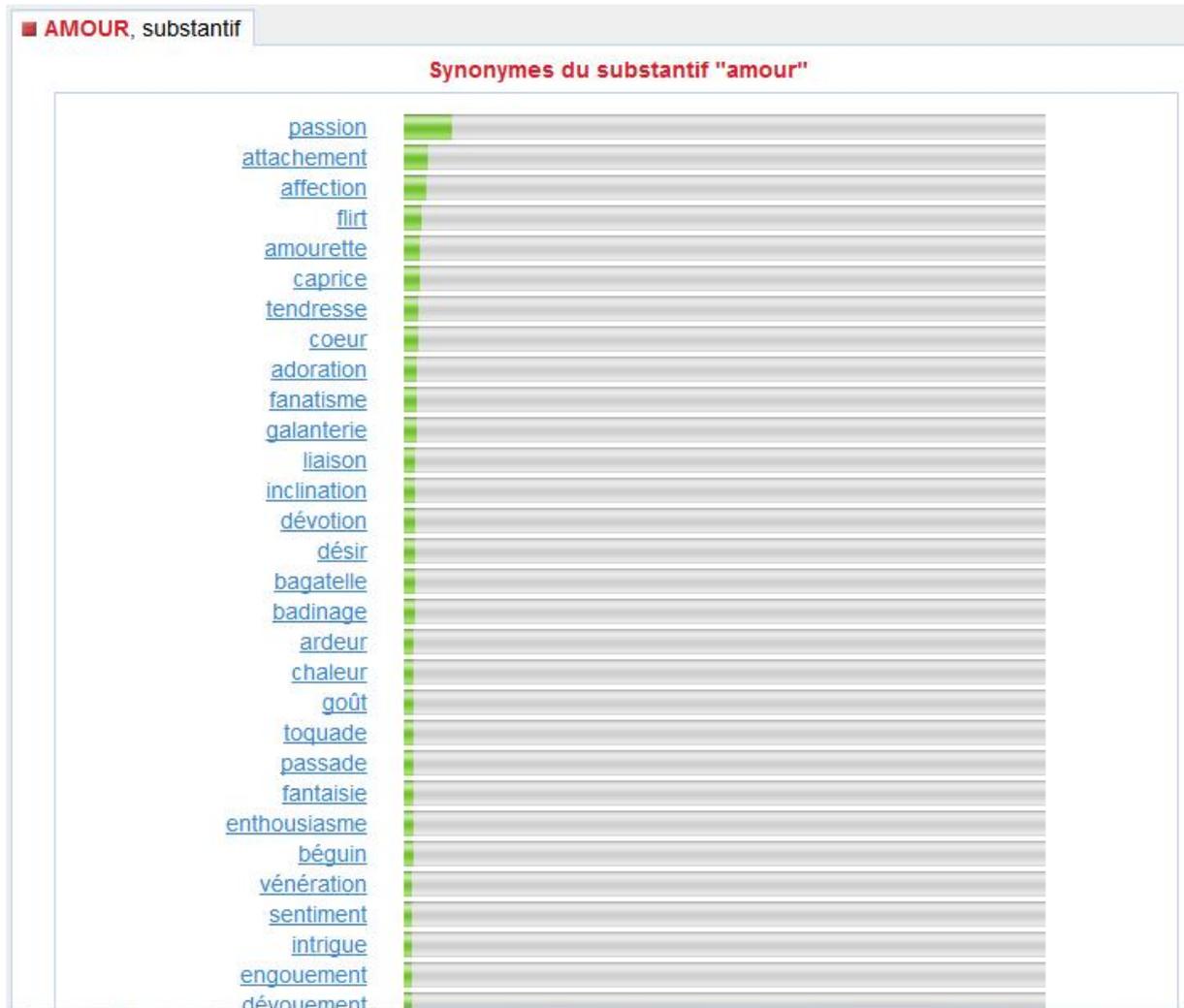


Figura 2.7. Resultado do *Dictionnaire des Synonymes du CRISCO* para a busca pela palavra *amour*; acessado em 21 maio 2010.

O *Dictionnaire Sensagent*, da Memodata (França), é uma plataforma multilíngue que se define como “uma enciclopédia *on-line*, um *thesaurus*, um dicionário de definições e mais”. O coordenador do projeto é o ex-professor da universidade de Caen nas áreas de PLN e Semântica Lexical, Dominique Dutoit. O dicionário é baseado em um conjunto de cerca de 45.000 conceitos que se relacionam entre si em até 200 tipos de relações para descrever o significado aproximado das palavras de cada idioma. Para o francês, cerca de 200 mil palavras estão contempladas. Para o inglês, são outras 160 mil e, para o alemão, mais 75 mil palavras abarcadas. Há, ainda, 80 mil palavras árabes e 72 mil palavras japonesas. As fontes utilizadas para a construção dessa base de dados lexicais foram os *synsets* a WN.Pr e outras estruturas léxico-conceituais como o SUMO (*Suggested Upper Merged Ontology*) (NILES, PEASE, 2001). Na Figura 9, podemos ver a

parte da plataforma dedicada ao dicionário analógico (*dictionnaire analogique*) exibindo resultados para uma busca pela palavra *amour*.

The screenshot shows the search results for the word "amour" in the "dictionnaire analogique". The results are organized into several categories, each with a small icon and a label in brackets:

- Comportement et mécanisme comportemental, Comportement et mécanismes comportementaux - Activité sexuelle, Comportement sexuel, Comportement sexuel avant le mariage, Comportement sexuel préconjugal, Comportement sexuel préuptial, Cunnilingus, Cunnilingus, Fellation, Orientation sexuelle, Rapports sexuels par voie anale, Rapports sexuels par voie orale, Relations sexuelles buccogénitales, Relations sexuelles bucco-génitales, Relations sexuelles par voie anale, Relations sexuelles par voie orale, Sexe anal, Sexe oral, Sodomie [Hyper.]**
- Amour [Analogie]**
 - Émotions, Regret, Sentiments [Hyper.]
 - Cour (hommage), Cour amoureuse, Parade sexuelle (animal) [Analogie]
 - Amour (n.) [Classe]
- dieu et déesse : par origine [Classe...]**
- créature imaginaire d'aspect humain [Classe]**
- mythologie romaine [Thème]**
- ange [Thème]**
- amour entre êtres humains [Thème]**
 - dieu du panthéon romain [Classe]
 - ange [Classe]
 - amour entre êtres humains [termes liés]
 - Amour (n. pr.)
- prénom féminin [Classe...]**
- ancien [Caract.]**
 - prénom féminin (peu usité de nos jours) [Classe...]
 - Amour (n. pr.)
- cours d'eau frontalier séparant des États [Classe]**
- les noms de cours d'eau, mer, océan [Classe...]**
- détroit (bras de mer) [Classe...]**
- pays socialiste et assimilé [Classe]**
- État d'Extrême-Orient [Classe...]**

Figura 2.8. Recorte para o resultado do dicionário analógico do *Dictionnaire Sensagent* para a busca pela palavra *amour*; acessado em 21 maio 2010.

O *Thesaurus da Língua Portuguesa*, apresentado na última linha do Quadro 2.2, não possui relevância acadêmica, na medida em que foi elaborado, segundo o próprio autor, à revelia de métodos científicos ou critérios editoriais que mereçam confiança. É fruto, portanto, de puro exercício introspectivo de um falante médio da língua portuguesa. Ele é contemplado no Quadro 2.2 e nessas linhas por ser o único *thesaurus* em língua portuguesa assemelhado com a estrutura e o conteúdo de um dicionário analógico e disponível na internet. Essa constatação, em contraste com o que acabamos de ver para as demais línguas, corrobora para o propósito desta tese.

2.7 ANÁLISE DO ESTADO DA ARTE DOS DICIONÁRIO ANALÓGICOS

Nesta Subseção, passamos em revisão o que foi dito na Seção 2 de forma a examinar o estado da arte dos dicionários analógicos (impresso/ *on-line*) com vistas a explicitar limitações passíveis de serem suplantadas no escopo desta tese.

De imediato, cf. 2.6, constatou-se que não há, ainda, para o PB, nenhuma coleção ou estrutura na internet que satisfaça os critérios mínimos para que se possa dizer que ali funcione um dicionário analógico digital confiável. Diferentemente dos idiomas inglês e francês, o PB carece de um dicionário analógico digital que seja resultado de pesquisa científica, segundo critérios acadêmicos ou, no mínimo, de tradição lexicográfica. Tal constatação justifica nossa proposta de construção de um modelo para um dicionário analógico digital a ser disponibilizado na internet.

Primeiramente, cf. 2.2, as relações semânticas paradigmáticas e as relações semânticas sintagmáticas não são explicitadas no interior dos grupos de analógicos em nenhuma das obras analisadas. Ainda que existam agrupamentos mais ou menos coerentes separando os dois tipos de relação, os arranjos são confusos e não tornam explícita a distinção semântica que ali ocorre. Como exemplo, abaixo, fragmentos do verbete **903. Casamento** do *Dicionário Analógico* do professor Francisco dos Santos Azevedo (AZEVEDO, 2010, p. 437), conforme sua diagramação original:

903. Casamento, matrimônio, enlace (matrimonial), consórcio, himineu, conúbio, recebimento, ligação, união, maridança, maridagem, laço conjugal, *vinculum matrimonii* = vínculo matrimonial/conjugal, boda, desposório, esposório, esponsais ou enponsálias, núpcias, mistura, casório, o facho do himineu, bênção nupcial; banho de igreja (pop.), conjúgio; tálamo(s) (fig.); beco sem saída = casamento (pop.); [*quebra de linha*]

[...]

noivo, noiva, prometida, damas de honra ou honor (ant.), nubente, contraente, homem casado, esposo, marido, companheiro, consorte, cônjuge, neógamo, papel queimado = homem casado, madame, esposa, senhora, matrona, costela, cara-metade, companheira, gamologia, gamomania, casal, jovem par; [*quebra de linha*]

padrinho, madrinha = tambeira ou tameira, paraninfo, tessemunha; [*quebra de linha*]

O primeiro enxerto elenca palavras e expressões sinônimas para a UL *casamento*, ao passo que o segundo trecho arrola palavras que designam participantes envolvidos no **frame semântico** de CASAMENTO – conforme a Semântica de Frames (FILLMORE, 1982) (cf. 3.2). Isso não está explícito em qualquer lugar da obra, havendo apenas quebras de linha para separar não apenas esse tipo, mas qualquer tipo de agrupamento. Além disso, as quebras de linha muitas vezes não seguem uma lógica afiada, como essa do segundo enxerto. Se o parágrafo deveria contemplar todos os participantes de um casamento, o parágrafo abaixo (padrinho, madrinha etc.) não deveria existir em separado. Se, no entanto, a intenção era de que o parágrafo considerasse apenas os

participantes principais da cena (noivo, noiva, marido, esposa etc.), então o elemento *dama de honra* não deveria pertencer ao conjunto, estando separado em parágrafo isolado ou em conjunto com os demais participantes no parágrafo abaixo. E as ULs *gamologia* e *gamomania* são ainda mais estranhas ao agrupamento, já que nenhuma das duas palavras designam participantes. Segundo o *Dicionário Houaiss Eletrônico*, a primeira designa um “tratado a respeito do casamento” e a segunda nomeia uma psicopatologia “que se caracteriza pela monomania do casamento”.

Há, ainda, uma questão estrutural, de igual importância e teor, que é interna aos parágrafos. Os agrupamentos semânticos – a gradação entre sinônimos e análogos – não estão explicitados. O segundo parágrafo, por exemplo, poderia ser reescrito, utilizando-se a notação de *synsets*, como abaixo:

{noivo, prometido}

{noiva, prometida}

{nubente, contraente}

{recém-casado, neógeno}

{homem casado, esposo, marido, companheiro, consorte, cônjuge, papel queimado}

{esposa, madame, senhora, matrona, costela, cara-metade, companheira}

{casal, jovem par}

Noivo e *prometido* dizem respeito ao mesmo participante, estando as duas ULs, portanto, em uma relação paradigmática, na medida em que, em contraste com as demais ULs, são comutáveis entre si com o menor grau de ambiguidade. *Nubente* e *contraente* poderiam igualmente se referir ao mesmo participante, sendo eles, no entanto, não marcados em termos de gênero. Já os termos *recém-casado* e *neógeno* não podem designar o mesmo participante (*noivo/prometido*) no mesmo momento *t* do enquadramento em questão. Os dois conjuntos fazem parte de diferentes estágios do **script** (SCHANK & ABELSON, 1977, cf. 3.2) de CASAMENTO, assim como o conjunto {homem casado, esposo, marido, companheiro, consorte, cônjuge, papel queimado} pertence a um terceiro período de tempo. Os três conjuntos se encontram, portanto, em relação sintagmática.

t_1 {noivo, prometido} + {nubente, contraente}

t_2 {recém-casado, neógeno}

t_3 {homem casado, esposo, marido, companheiro, consorte, cônjuge, papel queimado}

Como resposta a esse tipo de limitação das obras analisadas, acreditamos ser possível: i) agrupar e relacionar as ULs em grupos de análogos que se distinguem

claramente entre relações paradigmáticas e sintagmáticas; ii) utilizar conjuntos de sinônimos para organizar os grupos de analógicos; iii) explicitar na interface de usuário o tipo de relação e, quando for o caso, os tipos de participantes do *frame* semântico. O desenvolvimento dessas questões será tratado na Seção 3.

Passemos a analisar agora a estrutura hierárquica dos dicionários analógicos destacados ao longo desta seção, com especial atenção para o topo da hierarquia, onde estão as categorias mais altas. Constatamos que o *Plano de Classificação das Ideias* e os níveis hierárquicos dos dicionários analógicos, cf. 2.2, são idiossincráticos, não havendo por parte dos autores uma justificativa ou argumentação para que o sistema conceitual proposto se organize dessa forma e não de outra. No entanto, estudos de pesquisadores independentes procuraram encontrar as raízes que dão origem à macroestrutura conceitual utilizada nos principais dicionários do gênero. Lyons (1977), por exemplo, acredita que o *Roget's Thesaurus* tenha sido bastante influenciado pela filosofia do século XVII (fortemente marcada pelas ideias de Francis Bacon, Descartes e Leibniz), que acreditava ser possível a construção de “uma linguagem ideal para a sistematização e desenvolvimento do conhecimento científico”. Já Babini (2006) sugere que a arquitetura conceitual proposta por Roget tem fortes raízes aristotélicas. Como o *Roget's Thesaurus* deu origem a toda uma tradição de dicionários analógicos, cf. 2.2 e 4.3, o seu sistema conceitual está replicado nas principais obras do gênero.

Em termos mais práticos do que filosóficos, tendo em vista que estamos lidando com uma obra destinada ao uso geral e não apenas aos iniciados, o *Plano de Classificação das Ideias* possui a função de guiar o usuário do conceito por ele conhecido até a palavra ou expressão que ele desconhece ou que não lhe vem facilmente à memória. Sobre isso, vejamos o que diz certa passagem da Seção “como usar esse dicionário” da última edição do *Dicionário Analógico da Língua Portuguesa* do professor Azevedo (2010, p. x):

O acesso às sugestões apresentadas pode ser feito por dois caminhos de busca. Um, no modelo do *Thesaurus* de Roget, identificando a área conceitual na qual se encaixa a palavra ou expressão que se quer encontrar, e buscando nessa área o grupo analógico mais próximo daquele que provavelmente conteria o termo procurado, como se verá abaixo. Outro, a partir de um termo ou expressão que se conhece, para buscar no(s) grupo(s) analógico(s) onde ele se encontra outras alternativas de expressão. No primeiro caso, a busca se faz pela árvore classificatória dos **grupos analógicos**. No segundo, pelo **índice geral** que relaciona cada um dos quase 100 mil termos e expressões do dicionário ao(s) grupo(s) em que se encontra.

No primeiro caso, se tivermos em mente, por exemplo, o conceito AMOR, devemos percorrer o quadro sinóptico pelo caminho **classe VI. Afeições> divisão III.**

Afeições Simpáticas> 1º) *Sociais*> 897. **Amor**, desviando-se, intuitivamente, de ramificações entre outras seis classes (*Relações abstratas, Espaço, Matéria, Entendimento, Vontade*), outras quatro divisões (*Em geral, Pessoais, Morais, Religiosa*) e, por fim, outras três subdivisões (*Altruístas, Especiais, Retrospectivas*). Não utilizamos a notação reservada aos conceitos para nenhum dos níveis de bifurcação, visto que, de fato, não se trata de uma árvore de conceitos (como a *árvore de Porfírio* ou uma *wordnet*), mas sim de um sistema de classes/categorias com restrições, cf. 4.3. Como foi dito, o mesmo vale para as demais obras analisadas, com exceção do *Historical Thesaurus of the Oxford English Dictionary* (KAY, C. et al. [eds.], 2009), que será citado adiante e analisado com profundidade na Subseção 4.7.

O *Diccionario Ideológico de la Lengua Espanola* (CASARES, 1959), apesar de esclarecer em seu prefácio que os precursores da classificação ideológica do léxico foram Roget (1852) e Boissière (1862), acabou por construir uma classificação própria, como se vê na Figura 2.9.

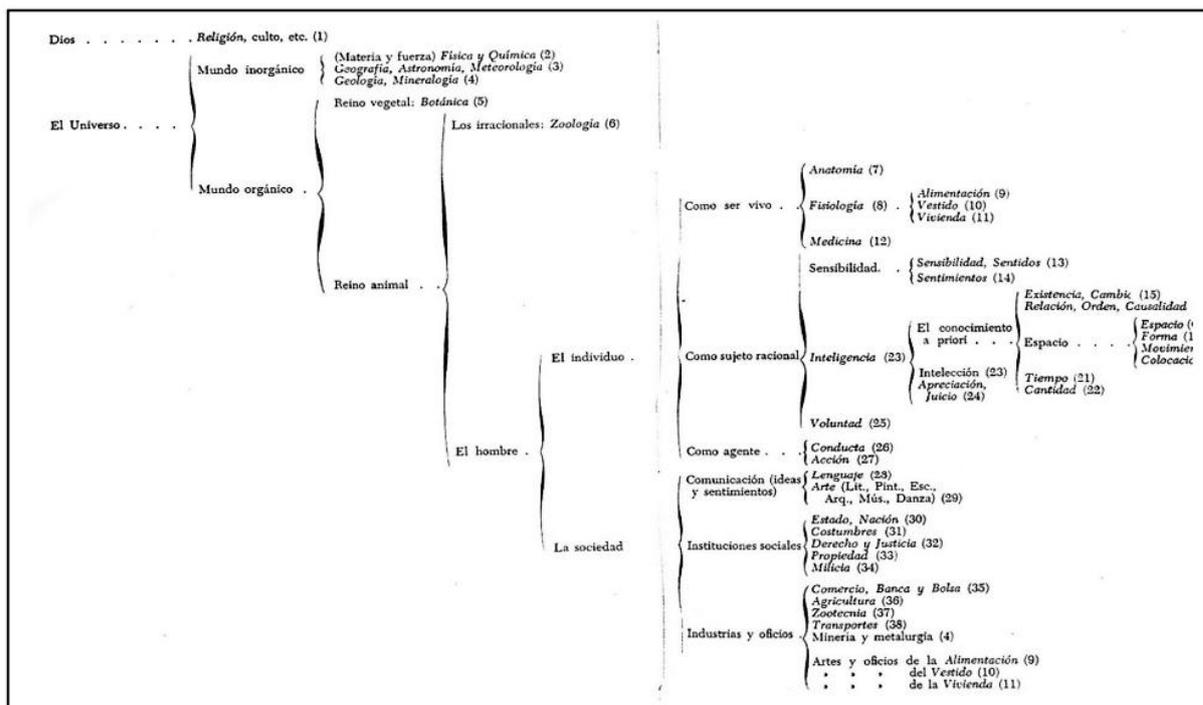


Figura 2.9. Plano de classificação das ideias do *Diccionario Ideológico de la Lengua Espanola* (CASARES, 1959)

Inúmeras tentativas de classificar e organizar as ideias e o léxico foram levadas a cabo desde a antiguidade e, especialmente, nos últimos 500 anos, impulsionadas pelo florescimento da razão e a invenção da prensa. Com o advento da tecnologia digital e da

internet, os sistemas de organização e recuperação da informação ganharam novo fôlego. Nos primeiros anos da internet, foram criados sistemas de indexação e recuperação da informação que se baseavam tanto na consulta por palavras-chave quanto em índices organizados em árvores de categorias, cf. Figura 2.10, da mesma forma que um dicionário analógico possui um índice remissivo e um plano de classificação das ideias.



Figura 2.10. Árvores de categorias do *Cadê*, um dos primeiros sistemas de buscas da internet.

Hoje, no entanto, permanece apenas o percurso de consulta por palavra-chave, que se tornou o padrão das ferramentas do gênero – vide o famoso motor de buscas da *Google* –, da mesma maneira que a consulta em um dicionário analógico é muito mais corrente e efetiva quando executada via índice remissivo, onde estão listadas todas as palavras e expressões que integram o dicionário, do que pelo plano de classificação das ideias. Exercício esse que será ainda mais conveniente no ambiente digital, na medida em que o usuário não precisará correr os olhos por uma lista de palavras em ordem alfabética, bastando apenas digitar a palavra dentro de um campo de pesquisas e, dessa

forma, obter como resposta uma pequena lista com os domínios léxico-conceituais onde aquela palavra aparece.

De fato, a implementação de um dicionário analógico em meio digital deve levar essas questões em consideração, procurando extrair da tecnologia disponível todo o seu potencial. Tal recurso, no entanto, não descaracteriza a propriedade onomasiológica de um dicionário analógico, na medida em que o ordenamento por domínios léxico-conceituais permanece e o artifício proposto destina-se apenas a referenciar, de maneira mais ágil, os domínios nos quais o usuário poderá encontrar os sinônimos e analógicos que esteja procurando.

Essas considerações são importantes para esclarecer que i) conceitualmente os planos de classificação das ideias são idiossincráticos e *ad hoc*; e ii) sua aplicação prática como sistema de recuperação da informação é ineficaz e, em meio digital, parece mesmo ser desnecessária. No entanto, há ainda dois esclarecimentos a serem feitos. O primeiro é o de que não estamos negando que a organização da informação lexical em estruturas hierárquicas possua validade tanto cognitiva quanto prática. Como veremos na Seção 4, esse é um longo e intrincado debate e tudo dependerá dos propósitos e aplicações do sistema que se está a criar. O segundo é que, para os fins desta tese, discutiremos aprofundadamente apenas a camada mais alta da hierarquia, ou seja, as **Grandes Categorias** (*major headings*). Nosso modelo abdica de categorias intermediárias entre os domínios conceituais e as Grandes Categorias, no topo da hierarquia.

Nossa proposta, enfim, sugere uma classificação no topo da hierarquia muito distinta daquela consagrada no *Roget's Thesaurus*. Nossas pesquisas, cf. veremos em 4.7, nos levam a um alinhamento com a proposta do *Historical Thesaurus of the Oxford English Dictionary*, que organiza todos os diferentes grupos de conceitos lexicalizados abaixo de apenas três Grandes Categorias: THE EXTERNAL WORLD, THE MENTAL WORLD, THE SOCIAL WORLD²⁸. Na Seção 4, vamos mostrar que esse tipo de sistema conceitual não apenas demonstra-se filosoficamente mais apurado ao nosso trabalho, como também, demonstrou possuir benefícios lexicográficos. Conceitos que prototipicamente exibem internamente uma estrutura de *frame*, com mais relações sintagmáticas do que paradigmáticas, costumam pertencer às categorias MUNDO FÍSICO e MUNDO SOCIAL, ao

²⁸ Conforme veremos na Seção 4 (cf. 4.7.1), a nossa designação para as mesmas três categorias foi levemente alterada, buscando uma tradução mais apropriada e transparente aos nossos propósitos. Passam-se, portanto, respectivamente, a denominar-se: MUNDO FÍSICO, MUNDO PSÍQUICO e MUNDO SOCIAL.

passo que conceitos prototipicamente abstratos, com mais relações paradigmáticas do que sintagmáticas, tendem a se incluírem na categoria MUNDO PSÍQUICO.

Por fim, com relação às questões macroestruturais, constatou-se ainda que os grupos de analógicos possuem muitas palavras e expressões, às vezes, centenas delas, que são organizadas internamente apenas pelas classes gramaticais, quebras de linha e ponto-e-vírgulas, como foi apontado anteriormente. Vejamos, por exemplo, o caso do grupo de analógicos encabeçado pela UL **amor** no *Dicionário Analógico* do professor Francisco dos Santos Azevedo (2010), seguindo a pontuação original do livro:

897. Amor, carinho, idolatria, afeto, amoricos, amorosidade, amorio (ant.), inclinação (desejo) 865; dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência 906; agarramento, entranha, aspiração, galanteio, galanice; namoro, namorico, namorisco, flerte, amizade colorida, paixão, adoração, ardor, fervor, chamas, calor, devoção, atração, êxtase, enlevamento, arroubamento, enlevo, feitiçaria, namoramento, xaveco (gír.);

chama inédita; chama de cintilações desconhecidas/ de efeitos maravilhosos; amatividade, Cupido, Afrodite, Vênus, Ondim ou Ondina;

mirto, murta, setas do amor;

história/ laços/ caso/ negócio de amor; dedo de Cupido (namoro) 902; olhar amoroso, transportes amorosos, filoginia, erotídeas; amor inextinguível/ verdadeiro/ irremediável/ inabalável/ acendrado/ sincero/ ardente/ profundo/ indestrutível/ imaculado/ platônico/ paterno/ materno/ filial, conjugal/ incondicional/ inabalável/ sem fim/ infinito; piedade filial, favorito 899; popularidade, prestígio, influência; amante, proco (desus.), namorado, namorador, namoradeiro, pretensor, pretendente, admirador, vegete = amante velho, apaixonado, adorador, galanteador, cortejador, adorante, galã, amoroso, jacaré (pop.), marrancho (pop.), derriçador, quebra-esquinas, babão, Lotário, bandoleiro, conquistador, D. João (libertino) 962, Casanova;

chichisbéu, caro, esposo, bem, amigo, querido, derriço, predileto, cujo, zinho (bras.), frecheiro, beijocador, beijoqueiro, flerte, namorido (pop.), ficante (bras. gír.);

namorada, arrojada (ant.), cupida, apaixonada, querida, amorzinho, amada, Dulcineia, derriço, benzinho, predileta, anjo, querubim, serafim, ídolo, deusa, inclinação, objeto da simpatia, cotó, namoradeira, janeleira, frança, pau de cabeleira;

noivo, noiva, fiancée, pretendida, futura, nubente; casal de pombinhos, morada de amor, ninho, dois corações num só, Romeu e Julieta, Abelardo e Heloísa;

tesão, sensualidade, amor carnal, lascívia, erotismo, luxúria, volúpia, voluptuosidade, transa, sexo, cópula, sexualidade, beijo, abraço, carícia, orgasmo, gozo.

V. amar, apaixonar-se por, estar apaixonado/ enamorado por, estimar muito, benquerer, gostar de, adorar, ser devoto de, cultuar, ter em preço, querer a, dar a vida por, estremecer, simpatizar com, sentir ternura por, entranhar-se, estimar, ter amores, encher-se de afetos, pender para, agradecer-se de, preferir, ter preferências, entranhar-se de amor, apegar-se a, aferrar-se a, antepor, afeiçoar-se;

ter/ nutrir amor por ou a; adorar com imaculado amor, dar o seu coração a alguém, sentir bem de alguém, possuir o coração de, ser o beliz de alguém, concentrar todas as esperanças em, dedicar todo o afeto a;

adorar, idolatrar, amar como perdido, amar perdidamente; amar deveras/ idolatradamente/ até a adoração/ até o sacrifício/ com furor/ com exaltação/ com todos os enternecimentos/ com delírio/ como um louco; gostar de alguém a morrer;

querer a alguém como as meninas dos olhos/ como os seus olhos; estar amartelado de amores por, estar a dar até a última gota de sangue por;

beber os ares/ os ventos por; devotar a alguém o culto de verdadeira estima, alucinar-se, desvairar-se, arder por alguém, morrer por, babar-se por, querer comer alguém aos bocados, amoriscar-se, enamorar-se de, trazer nos olhos e no coração, ter no coração de alguém um altar, ser louco/doido por alguém, suspirar por, embeijar com alguém, queimar-se nos olhos de, engar (fig.), possuir o coração de, reinar no coração de, ser fanático por alguém, ter por alguém verdadeiro fanatismo, trazer alguém nas palmas da mão, corresponder ao amor de alguém, entregar-se à vertigem do amor, abajoujar-se, ter coração de estalagem, despertar paixão, fazer conquista, tornar-se o mimoso de, seduzir, atrair, cativar, ganhar a simpatia de, encantar, deslumbrar, enfeitiçar, prender o coração, embeijar, abrasar de amor, atrair de modo irresistível, inclinar, tornar afeiçãoado, benquistar; insinuar-se, tornar-se simpático;

granjear/ captar a simpatia; cair em graça, cair na graça de, serem dois corações num só, servir de pau de cabeleira, desejar 865; ser a tampa da panela de alguém (pop.);

transar, fazer amor, beijar, acariciar, abraçar, gamar, gozar, ter orgasmo.

Adj. amante & v.; lamecha, babadinho, amator, amativo, namorador, cupidíneo, cupidinoso, vénero, venéreo, apaixonado; apaixonante;

perdido/ doido/ louco de amor; bajoujo, louco, doido, namorado, enamorado, amoriscado, coamante, dedicado, afeiçãoado = adicto, desvelado, benquerente, devotado, férvido, sapeca, namoradeira, terno, meigo, faceiro, propenso, amorudo (burl.), amoroso, requebrado, lânguido, mimoso, suave, voluptuoso, fino, extremoso, constante, querençoso, afetuoso, maternal, paternal, fraternal, cordial, simpático, insinuativo, amorável, amigável;

impudico, amatório, derretido, erótico, velhaquesco;

encantador, apegadiço, bem-amado, rendido, amado, estremeado, querido, quisto, benquisto, bem-visto, dileto, estimado = preçado, prezado, predileto, favorito, caro, precioso, preferido, invejável, invejado, amável, caroável, adorável, adorando (ant.), sedutor, encantador, interessante, cativante, insinuante, fascinante, feiticeiro, querubínico, seráfico, angélico;

sensual, lúbrico, lascivo, luxurioso, orgásmico, sexual.

Adv. ternamente & *adj.*; com todo o ardor de uma paixão; apaixonadamente.

O grande número de palavras e expressões listadas sem importantes critérios de agrupamento não facilita as buscas do usuário. Nesse exemplo, só a classe dos substantivos possui mais de duas centenas e o grupo inteiro chega perto de quinhentas sugestões de palavras ou expressões. Por um lado, cf. 2.5, o volume de ULs sob um

mesmo domínio léxico-conceitual é o grande diferencial desse tipo de obra em relação aos dicionários de sinônimos. No entanto, a ausência de critérios para agrupar e dispor a informação léxico-semântica dentro dos grupos de analógicos faz com que esse volume se torne um problema para o usuário.

É preciso, portanto, criar critérios de organização e de relevância para que o usuário possa ter acesso facilitado às palavras e expressões que ele esteja buscando dentro de um mesmo domínio léxico-conceitual. A solução é, mais uma vez, utilizar CSs para organizar os grupos de analógicos, com a diferença de que esses conjuntos não podem ser fechados e estáticos como o são nas *wordnets*. Para os nossos fins, conjuntos de sinônimos e analógicos que sejam difusos e dinâmicos são mais apropriados, cf. Seção 5. Nesse caso, o critério de relevância para determinar o peso das ULs dentro de cada CdSA será a sua frequência de aparição em agrupamentos semânticos equivalentes extraídos de diferentes dicionários de sinônimos e dicionários analógicos, como no exemplo abaixo.

{**afeição**, admiração, afeto, inclinação, simpatia} [TEP]

{**afeição**, afeto, amor} [MIC]

{**afeição**, simpatia, admiração, estima} [MIC]

{**afeição**, afeto, amizade, amor, benquerença, carinho, inclinação, propensão, simpatia} [FER]

{**afeição**, adoração, afeto, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU]

{**afeição**, amor, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfatuação} [IDS]

{dileção, predileção, preferência, simpatia, estremecimento, benquerença, **afeição**, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência} [AZE]

Consolidado:

{afeição (7), simpatia (6), afeto (5), amor (4), admiração (3), apego (3), benquerença (3), carinho (3), inclinação (3), ternura (3), amizade (2), benevolência (2), dedicação (2), dileção (2), estima (2), querença (2), adoração (1), aferro (1), caridade (1), chamego (1), conchego (1), constância (1), cuidado (1), derretimento (1), derriço (1), comunhão de sentimentos (1), desvelo (1), ênfatuação (1), estremecimento (1), fraternidade (1), gosto (1), idílio (1), idiopatia (1), intimidade (1), meiguice (1), predileção (1), preferência (1), propensão (1), vontade (1), zelo (1)}

Conjuntamente, vamos utilizar a Teoria dos Protótipos (ROSCH, 1975), cf. 5.2.2, e o conceito de radialidade (LAKOFF, 1987), cf. 5.2.3, para ajudar a organizar os CdSAs no interior de seus respectivos *anaset*s. Por exemplo, o conceito AMOR ROMÂNTICO é prototípico do domínio AMOR. Ele ajuda a explicar as interpretações ligadas aos conceitos PAIXÃO e SIMPATIA, que são sentidos bifurcados em relação paradigmática com o protótipo. É uma relação de especialização em que PAIXÃO corresponde a um sentimento mais intenso e potencialmente de interesse sexual. Ao passo que SIMPATIA denota um afeto mais brando, entre pessoas que se admiram e que não almejam expressar esse amor fisicamente. Já os conceitos ADORAÇÃO e CUIDADO, dentro do domínio do AMOR, parecem derivar respectivamente dos conceitos PAIXÃO e SIMPATIA, com a diferença de que o sentimento nesse caso se dirige a objetos, eventos ou ideais abstratos.

2.8 SÍNTESE DA SEÇÃO

As investigações apresentadas nesta Seção serviram para explicitar o estado da arte dos dicionários analógicos – ou *thesaurus*, como são denominados em inglês. Definimos que os dicionários analógicos são obras em que o léxico se organiza em termos de conceitos semântica e ontologicamente estruturados e que o percurso se faz do nível conceitual para o nível lexical, sendo considerados, portanto, de ordem onomasiológica.

A macroestrutura dos dicionários analógicos foi definida como possuindo quatro componentes características: i) o plano de classificação das ideias, ii) os quadros sinópticos de categorias, iii) os grupos de palavras e iv) o índice remissivo. Ainda que o ponto de partida das consultas de tipo onomasiológico deveria proceder por meio da primeira e da segunda componentes, constatou-se que a consulta via índice remissivo é a mais utilizada. Isso porque os planos de classificação das ideias e os quadros sinópticos de categorias são idiossincráticos, não permitindo o manejo prático e intuitivo por parte do consulente. Tal constatação motivou a nossa proposta de simplificação do sistema hierárquico conceitual do nosso modelo em relação àquele consagrado na tradição do *Roget's Thesaurus*, conforme veremos na Seção 4 da presente tese.

Com relação aos grupos de palavras, constatamos que eles carecem de critérios explícitos para a sua organização interna. Além de não ficar clara a distinção entre os conceitos sob um mesmo domínio conceitual, os grupos de palavras não explicitam a divisão, a nosso ver de grande importância, entre agrupamentos conceituais em relação paradigmática e agrupamentos conceituais em relação sintagmática. Tal constatação motivou a nossa proposta de organizar as ULs em grupos de analógicos que, internamente, distingam quais agrupamentos conceituais estão em relação paradigmática e quais estão em relação sintagmática, conforme veremos na Seção 3.

A presente Seção evidenciou também que os dicionários analógicos possuem uma função primordialmente codificadora, sendo considerados, portanto, dicionários de produção. Tal constatação reforçou nossa hipótese de que este tipo de obra, desde que adequadamente adaptada para o ambiente digital, deverá ocupar um papel central quando integrada ao *website* Dicionário Criativo, cujo objetivo é auxiliar os seus usuários na produção de textos criativos. No mesmo sentido, a Subseção dedicada à criatividade aponta que associações entre conceitos mais fracas, porém mais numerosas, estimulam o pensamento criativo. Conceitos anteriormente distantes se tornam acessíveis, o que propicia conexões imprevistas e criativas. Motivados por essa ideia, apresentaremos, no final da Seção 4, a nossa opção por uma estrutura conceitual em forma de rizoma, em oposição à estrutura em forma de árvore taxonômica das *wordnets*.

Por fim, a comparação entre dicionários analógicos e dicionários de sinônimos, levada a cabo na presente seção, reforçou a nossa hipótese de que a distinção entre sinônimos e análogos, em certo sentido, é nebulosa. Um mesmo domínio conceitual é recortado em categorias conceituais de maneiras diferentes a partir de metodologias e concepções diferentes. Dessa forma, abandonamos a ideia de trabalhar com um constructo representacional para o agrupamento de ULs que possua limites bem definidos e que pretenda instanciar um conceito unívoco. O que motivou a proposição do constructo CdSA (conjunto difuso de sinônimos e analógicos) e sua ordenação prototípica dentro de *anaset*s, conforme veremos nas Seções 5 e 6, respectivamente.

SEÇÃO 3 – RELAÇÕES PARADIGMÁTICAS E SINTAGMÁTICAS NA ORGANIZAÇÃO DO DAD

Foi demonstrado em 2.7 que os dicionários analógicos disponíveis revisados não explicitam a diferença entre relações léxico-semânticas paradigmáticas e sintagmáticas no interior dos seus grupos de analógicos. Por isso, nesta seção, discutiremos as relações lexicais e lógico-conceituais paradigmáticas e sintagmáticas como organizadoras desse tipo de obra lexicográfica. Nossa hipótese é a de que as relações paradigmáticas (sinonímia, antonímia, hiperonímia, etc.) são melhor trabalhadas em estruturas como as das redes *wordnets* (MILLER *et al.*, 1990; FELLBAUM, 1998), e as relações sintagmáticas são adequadamente tratadas em termos das estruturas semelhantes às das redes *framenets*, que implementam computacionalmente a semântica de frames (FILLMORE *et al.*, 2003; BAKER, FILLMORE e LOWE, 1998).

Nas subseções 3.1 e 3.2, vamos discorrer sobre informações e conceitos básicos dos projetos das *wordnets* e das *framenets*, respectivamente, tendo em vista a importância de alguns de seus conceitos e métodos para o nosso modelo. Na Subseção 3.3 – que propõe o equacionamento do domínio linguístico –, vamos argumentar pela proposta de organizar o *Dicionário Analógico Digital* a partir da explicitação das relações paradigmáticas e sintagmáticas na estruturação da informação léxico-conceitual dentro dos grupos de analógicos. Na Subseção 3.4 – que propõe o equacionamento do domínio linguístico-computacional –, vamos apresentar as noções que aproveitaremos da Teoria das Redes (BARABÁSI, 2002) e mostraremos como elas podem ser utilizadas para formalizar as relações semânticas paradigmáticas e sintagmáticas em termos de redes semânticas.

3.1 WORDNETS

As *wordnets* (“redes de palavras”) são um tipo especial de base de dados lexicais²⁹, em que unidades do léxico se agrupam por meio do significado que elas expressam e esses conjuntos se organizam dentro de uma rede conceitual. O termo *WordNet* começou designando a rede WordNet de Princeton (WN.Pr), desenvolvida por um grupo de pesquisa conduzido por George Miller dentro na universidade de Princeton

²⁹ Bases de dados lexicais são grandes repositórios estruturados de informação lexical que fornecem diferentes tipos de informação a diferentes sistemas de PLN (MARRAFA, 2001).

ao longo dos anos 90 (MILLER, 1998). Hoje já existem, ou estão em desenvolvimento, *wordnets* em diferentes línguas, que seguem os mesmos pressupostos que a WN.Pr – como, por exemplo, a WordNet.Br (WN.Br) (DIAS-DA-SILVA *et al.*, 2002), em construção para o PB.

A WN.Pr é 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** (*synonym set*). Os *synsets* constituem os **nós** da rede e as relações léxico-conceituais que se estabelecem entre os *synsets* constituem os seus **arcos** (ou arestas). A construção de um *synset* se vale da noção de **sinonímia contextual**. Tal noção define que duas ULs são sinônimas num contexto linguístico C se a substituição de uma pela outra em C não altera o significado de C (ULLMANN, 1963; LYONS, 1981; MILLER; FELLBAUM, 1991). O emprego do construto *synset* pressupõe que o falante tem acesso aos conceitos expressos pelos itens lexicais de sua língua. Ou seja, se o falante desconhece o significado de um determinado item lexical, uma forma sinônima vai ajudá-lo a identificar o conceito correspondente. Os *synsets* são sempre construídos a partir dessa possibilidade de intersubstituição de itens lexicais em contextos mínimos. Dessa forma, por meio da relação léxico-semântica paradigmática da sinonímia os itens lexicais de uma língua deixam de ser polissêmicos e passam a lexicalizar apenas um conceito particular. Assim, por exemplo, o item lexical polissêmico *amar* está distribuído em pelo menos dois *synsets* diferentes dentro da WN.Br. Em (1) o verbo *amar* lexicaliza o conceito glosado por “gostar de; ser ligado a”, enquanto que em (2) o verbo *amar* lexicaliza o conceito glosado por “ter relação sexual com”.

(1) {amar, bem-querer, estimar, estremecer, gostar, prezar, querer}

(2) {amar, copular}

No âmbito das *wordnets*, a polissemia é descrita em função de uma **matriz lexical** (FELLBAUM, 1998). A noção de matriz lexical possibilita a construção de *synsets* independentemente da especificação explícita do conceito por ele lexicalizado. Segundo Dias-da-Silva (2010), esse modelo tem como fundamento:

(a) a adoção do modelo relacional de representação do significado lexical, também conhecido como método diferencial de representação, que parte do princípio de que a ativação de um conceito lexicalizado na mente do falante realiza-se por meio da ativação do conjunto de formas lexicais que o lexicaliza;

(b) a noção de matriz lexical, que formaliza a correspondência biunívoca que se estabelece entre a forma e o significado das unidades lexicais de uma língua.

A matriz lexical é representada por meio de um plano cartesiano em que o eixo das abscissas contém as formas lexicais separadas em colunas e o eixo das ordenadas contém os *synsets* que representam os conceitos lexicalizados, separados em linhas. Quando duas células da mesma coluna são preenchidas, a forma lexical correspondente àquela coluna é **polissêmica**, pois isso indica que a mesma forma lexicaliza dois conceitos distintos. Quando duas células da mesma linha são preenchidas, as formas lexicais correspondentes são **sinônimas**, já que isso indica que essas diferentes formas lexicalizam o mesmo conceito.

O Quadro 3.1 é uma matriz lexical para o exemplo dado anteriormente, com os dois *synsets* extraídos da WN.Br para o verbo *amar*. Por lexicalizar os conceitos codificados em S1 e S2, o item lexical *amar* (F1), é considerado polissêmico. Ao mesmo tempo, as formas *amar* (F1) e *copular* (F8) são unidades sinônimas, pois lexicalizam o mesmo conceito S2.

Quadro 3.1. Exemplo de matriz lexical com o item lexical *amar*.

SYNSETS	FORMAS LEXICAIS							
	F1	F2	F3	F4	F5	F6	F7	F8
	amar	bem-querer	estimar	estremecer	gostar	prezar	querer	copular
S1 {amar, bem-querer, estimar, estremecer, gostar, prezar, querer}	S1<F1	S1<F2	S1<F3	S1<F4	S1<F5	S1<F6	S1<F7	
S2 {amar, copular}	S2<F1							S2<F8

Podemos dizer, portanto, que o *synset* é um conjunto de ULs de uma mesma categoria sintática que lexicalizam 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*”. Para codificar apenas um único conceito, 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 UL *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 mais genérico. No exemplo, poderia ser definido pela glosa “*person who performs*”. Isso 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. Já os dicionários analógicos se estruturam em categorias conceituais mais amplas.

Ao passo que a relação de sinonímia se estabelece entre ULs, as outras relações de sentido presentes nas *wordnets* são relações de natureza léxico-conceitual e se estabelecem exclusivamente entre *synsets*.

Os *synsets* são as unidades mínimas e, portanto, os nós de uma rede de relações léxico-conceituais. Os arcos da rede correspondem a **relações léxico-conceituais paradigmáticas** (*antonímia*, *hiperonímia/hiponímia* e *troponímia*) e **relações léxico-conceituais sintagmáticas** (*meronímia*, *causa* e *acarretamento*) – conforme os Quadros 3.2 e 3.3, adaptado de Dias-da-Silva (2005). No entanto, a expressiva maioria das informações codificadas em uma rede *wordnet* é de natureza paradigmática (FELLBAUM, 1998).

Quadro 3.2 Relações léxico-conceituais paradigmáticas de uma *wordnet*

Relações léxico-conceituais paradigmáticas		
Tipo de Relação	Classe Lexical Relevante	Exemplo
Antonímia	Substantivo	homem/mulher
	Verbo	entrar/sair
	Adjetivo	bonito/feio
Hiponímia/Hiperonímia	Substantivo	rosa/flor
Troponímia	Verbo	caminhar/mover

Quadro 3.3 Relações léxico-conceituais sintagmáticas de uma *wordnet*

Relações léxico-conceituais sintagmáticas		
Meronímia (parte-todo)	Substantivo	cabeça/nariz
Acarretamento	Verbo	comprar/pagar
Causa	Verbo	matar/morrer

A base da WN.Br está sendo construída a partir das listas de conjuntos de sinônimos da base do TeP 2.0 e fundamenta-se nas metodologias desenvolvidas para a construção da WN.Pr e redes particulares da *EuroWordNet* (doravante **EWN**) (VOSSSEN, 1998). Atualmente, a WN.Br contém 44 mil ULs e 18,5 mil *synsets*, sendo 11 mil verbos (4 mil *synsets*), 17 mil substantivos (8 mil *synsets*), 15 mil adjetivos (6 mil *synsets*) e mil advérbios (500 *synsets*) (DIAS-DA-SILVA, 2010, p. 139).

A EWN é uma base de dados lexicais multilíngue, em que *wordnets* de diferentes línguas, estruturadas segundo os mesmos princípios da rede americana, interligam-se por meio de relações de correspondência. A conexão entre as diferentes *wordnets* é intermediada por um Índice Inter-Lingual (*Inter-Lingual-Index*) (doravante **ILI**). É ele que garante a correspondência entre os *synsets* de línguas diferentes que lexicalizam um mesmo conceito.

3.2 FRAMENETS

Outra base de conhecimento lexical de interesse para nossa discussão é a FrameNet³⁰ (FILLMORE *et al.*, 2003; BAKER, FILLMORE e LOWE, 1998). A FrameNet é um projeto de lexicografia computacional baseado na teoria da **Semântica de Frames** (FILLMORE, 1982). De acordo com Fillmore, um **frame semântico** (doravante *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. Segundo Fillmore (1977), significações são relativizadas em “enquadramentos” (*frames*), de maneira a expressar o *continuum* entre os domínios da linguagem e da experiência. Segundo ele, o *frame* é uma estrutura conceitual culturalmente definida que atua na organização de segmentos da realidade. Há, portanto, uma relação direta entre a conceptualização de um *frame* e as experiências sócio-culturais dos falantes.

Embora os *frames* sejam esquemas conceituais – portanto, não exclusivamente linguísticos –, expressões da língua evocam *frames* e, ao mesmo tempo, é a partir deles que elas adquirem significado. Voltando ao nosso exemplo anterior, os itens da língua *noivo* e *prometido* evocam o *frame* de CASAMENTO e é só a partir dele que podemos considerá-los sinônimos e, portanto, instâncias do mesmo conceito; o conceito glosado

³⁰ <https://framenet.icsi.berkeley.edu/>

por “indivíduo que se comprometeu em casamento”. Na Linguística, um *frame* constitui, portanto, uma estrutura conceitual que formaliza o resultado das relações sintáticas e semânticas de uma UL e representa uma situação envolvendo vários participantes, propriedades e outros papéis conceituais que constituem cada **elemento do frame** (doravante **EF**). Cada *frame* especifica os EFs que dele participam. Segundo Salomão (2009, p. 173), “os Elementos de um Frame emergem na valência lexical e construcional das expressões evocadoras desse *frame* e correspondem, na designação mais tradicional da literatura, às Funções Temáticas da grade argumental dos núcleos lexicais”. No entanto, a autora chama a atenção para a especificidade dos EFs na versão mais atual da teoria em oposição à imprecisão das abordagens clássicas das Funções Temáticas.

Os Elementos do Frame, na versão mais recente da teoria, são de fato Funções Microtemáticas, postuladas em relação ao frame a que se referem. Assim os lexemas *vender* e *comprar*, que evocam ambos, com diferentes perspectivas, o frame do COMÉRCIO DE MERCADORIAS, têm como seus Elementos de Frame, respectivamente, as funções VENDEDOR e COMPRADOR: nisso diferem das abordagens mais tradicionais, que atribuiriam às mesmas funções o idêntico rótulo de agente. (SALOMÃO, 2009, p. 173)

Podemos ter também uma ordenação cronológica dos EFs, dentro de uma situação cultural estereotipada que se desenrola no tempo, que recebe o nome de *script* (SCHANK & ABELSON, 1977). O exemplo mais famoso na literatura foi dado por Schank e Kass (1988), o *script* do RESTAURANTE, que pode ser resumido como abaixo, em que os termos destacados são os EFs:

- (1) **Ator** vai ao **restaurante**
- (2) **Ator** senta
- (3) **Ator** pede uma **refeição** ao **garçom**
- (4) **Garçom** traz a **refeição** para o **ator**
- (5) **Ator** come a **refeição**
- (6) **Ator** dá dinheiro ao **restaurante**
- (7) **Ator** sai do **restaurante**

Segundo Schank e Abelson (1977, p. 136), um *script* é uma sequência de ações ordenadas e apropriadas para um dado contexto espaço-temporal e organizadas em torno de um objetivo, revelando aspectos do conhecimento de mundo do sujeito em questão. Um *script* pode ser elaborado com diferentes níveis de granulação e, assim como o *frame*, ele é perfilado a partir de algum EF. No exemplo acima, o *script* de restaurante foi perfilado a partir do EF **ator** e teria sua configuração alterada se fosse perfilado, por exemplo, pelo EF **garçom**. Um *script* de CASAMENTO CATÓLICO poderia ser descrito como a seguir:

- (1) os **convidados**, os **padrinhos** e o **noivo** chegam à **igreja**
- (2) o **noivo** espera a **noiva** no **altar**
- (3) as **damas de honra** entram na **igreja**
- (4) o **organista** toca a **marcha nupcial** no **órgão**
- (5) a **noiva** entra na **igreja** com o pai e segurando o **buquê**
- (6) o **padre** dá início à **cerimônia religiosa**
- (7) os **noivos** fazem **juramento** e trocam as **alianças**
- (8) o **noivo** levanta o **véu** da **noiva** e a **beija**
- (9) o **padre** os declara **marido** e **mulher (esposa)**
- (10) os **recém-casados** saem da **igreja**

Em termos gerais, a finalidade da Framenet é identificar e descrever *frames*, a partir de unidades básicas de análise, que são os Elementos do Frame – que podem ser centrais (*core*) ou periféricos (*non-core*) – e a Unidade Lexical (UL), definida como um pareamento entre um item lexical e um *frame* específico. Por meio 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 Berkeley³¹ 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 o projeto de construção de uma FrameNet para o PB, a FrameNet Brasil³², em desenvolvimento, sob coordenação da Profa. Dra. Maria Margarida Martins Salomão (SALOMÃO, 2009). A FrameNet Brasil é um projeto de pesquisa ligado ao PPG em Linguística da Universidade Federal de Juiz de Fora (UFJF). Seu objetivo é desenvolver recursos lexicais e sintáticos para o PB, tendo como base as teorias da Semântica de Frames e da Gramática das Construções.

3.3 O PAPEL DAS RELAÇÕES PARADIGMÁTICAS E SINTAGMÁTICAS NO DAD

A hipótese modernamente aceita é a de que o sistema conceitual humano é, por um lado, construído por princípios de similaridade, de inferência metafórica e, por outro, por princípios de contiguidade, de inferência metonímica (JACKOBSON, 1973; LAKOFF, JOHNSON, 1980; 2003; CROFT, CRUSE, 2004; DIRVEN, PÖRINGS, 2003).

³¹ Disponível em <http://framenet.icsi.berkeley.edu>

³² Disponível em <http://framenetbr.ufjf.br>, acessado em 12/01/2013

Segundo Lakoff & Johnson (1980, p. 3), “nosso sistema conceitual comum, em termos de o que pensamos e como agimos, é de natureza fundamentalmente metafórica”. Ainda que a vida mental se inicie com experiências que não são metafóricas, como as sensações e emoções, que fazem parte do nosso corpo e que interagem com o mundo físico, daí em diante as metáforas conceituais (LAKOFF, JOHNSON, 1980) são adquiridas por condicionamento associativo. Ao mesmo tempo, o sistema conceitual humano reflete a estrutura metonímica de um mundo externo disposto em segmentos adjacentes no plano espacial e de uma textura causal no plano temporal (PINKER, 2008).

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 investigações saussurianas (SAUSSURE, 1969), similaridade e contiguidade passam, respectivamente, a sugerir os dois eixos da linguagem: paradigma (justaposição/ seleção) e sintagma (subordinação/ combinação). A nossa proposta entende que um *dicionário analógico* deve, com efeito, estruturar as relações semânticas em torno dos dois eixos: o **paradigmático** e o **sintagmático**.

Os princípios que organizam a rede WN.Pr são essencialmente paradigmáticos: sinônimos, antônimos, hipônimos e hiperônimos. Conforme 3.1, os conjuntos de sinônimos (*synsets*) são a base organizacional de qualquer rede *wordnet*. Dentre as potenciais ULs sinônimas, o falante deverá selecionar apenas uma para incorporar à sua fala ou para utilizar no seu texto. Na frase (1), por exemplo, a UL *pé* pode ser substituída pela UL *base* sem que o sentido da frase seja alterado, constituindo assim uma instância da relação de sinonímia. Uma aplicação diferente dessa passa a ser uma extensão do sentido e gerar um efeito estilístico ou instaurar uma frase metalinguística, como em (2), em que *pé* é hipônimo de *base*.

- (1) Ao **pé** desse monte realizavam-se as competições esportivas entre eles, as Olimpíadas³³.
 (2) Os **pés** são a **base** que sustenta o corpo humano.

Ao contrário, as redes *framenets* fundamentam-se, como já se mencionou, no princípio organizacional da semântica de *frames*. Os EFs são os papéis semânticos dos participantes do evento/situação linguisticamente expresso, sendo específicos a cada *frame* e por isso também chamados de papéis situacionais. Dessa forma, dizemos que o

³³ Frases adaptadas do *corpus* do Nilc: <http://www.linguateca.pt/aceso/corpus.php?corpus=SAOCARLOS>. Acesso: 10 set. 2010.

conhecimento codificado no *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).

A base da FrameNet de Berkeley, por exemplo, contém o *frame* *Observable_body_parts*, composto pelos EFs centrais (*core elements*), *Body_part* e *Possessor* e, não centrais (*non-core elements*), que inclui, dentre outros, *Orientalional_location*. Na frase (3), *pé* é a UL no PB que instancia o EF *Body_part* (parte do corpo) da mesma forma que *Marcelo* é o participante do *frame* que corresponde ao EF *Possessor* (possuidor). Na frase (4), a UL *direito* instancia no PB uma das duas possibilidades que pode preencher a EF *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.

Como pode ser observado, o *frame* é uma categoria conceitual e os EFs são instanciados na frase por meio das ULs (FILLMORE *et al.*, 2003). Uma unidade lexical é o cruzamento entre um lema e um *frame*, ou seja, é o sentido da palavra fixado por um *frame* específico. Assim, a polissemia das palavras é resolvida na rede FrameNet, pois cada novo sentido está associado a um *frame* específico, caracterizando uma nova UL. Enquanto a WN.Pr desfaz a polissemia distribuindo os diferentes sentidos de uma palavra em *synsets* distintos e, portanto, por meio de diferentes paradigmas, a FrameNet resolve a polissemia fornecendo o contexto específico, e, portanto, por meio de configurações sintagmáticas, para a instanciação do sentido específico de um item léxico polissêmico.

Na WN.Pr, os sentidos da UL do inglês *piano* instanciam-se nos *synsets* nominais {*piano, pianoforte, forte-piano*}, glosado por “*a stringed instrument that is played by depressing keys that cause hammers to strike tuned strings and produce sounds*”, e {*piano, pianissimo*}, glosado por “*low loudness*”. Já na rede FrameNet, os sentidos da UL do inglês *piano* definem-se em termos do *frame* *Noise_makers*, que contém os EFs *Noise_maker* (central), *Creator*, *Ground*, *Material*, *Name*, *Time_of_Creation*, *Type* e *Use* (não centrais), além de a UL *piano* se associar a outras que também são ativadas pelo *frame* *Noise_makers*.

Os dicionários analógicos, tipicamente, arrolam grupos de palavras com ênfase nas relações paradigmáticas. Talvez por servirem, inicialmente, à composição textual de

discursos, conferências e palestras, entendeu-se que esse tipo de obra deveria priorizar ideias abstratas e as relações paradigmáticas entre as ULs correspondentes. Vejamos novamente abaixo o verbete **897. Amor** extraído do *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, [1950] 2010):

897. Amor. *Subst.* carinho, idolatria, afeto, amoricos, amorosidade, amorio (ant.), inclinação (desejo) 865; dileção, predileção, preferência, simpatia, estremeimento, benquerença, afeição, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência (...)

Para contrastar com essa ideia, pense nos livros didáticos dos cursos de idiomas, em que o objetivo é ensinar ao aluno o nome de objetos e situações cotidianas em outra língua. Neles, é comum vermos fotografias ou ilustrações de cenários estereotipados com flechas apontando para objetos, personagens, paisagens e estados de coisas com o respectivo nome, tipicamente atribuído a eles na língua estrangeira em questão. Ou seja, o outro idioma nos é apresentado por meio de *frames* e *scripts*. As ULs *cômoda*, *gaveta*, *colcha*, *cama (de solteiro)*, *mesa de cabeceira*, *despertador* etc. se relacionam, metonimicamente, para compor o *frame* de um DORMITÓRIO DE SOLTEIRO, por exemplo, como na Figura 3.1.

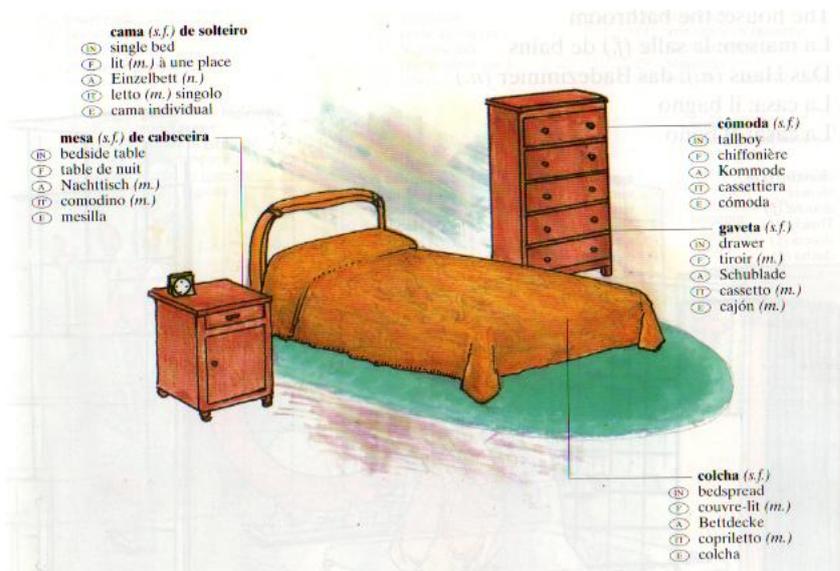


Figura 3.1. *Frame* DORMITÓRIO DE SOLTEIRO extraído do *Dicionário Multilíngue* (Reader's Digest Brasil, 1998).

Nesse tipo de material, é fornecido também, por vezes, uma série de *scripts* de como pegar um avião, como reservar um hotel, como comprar jornais etc. A relação com o mundo, e não com ideias abstratas, é muito mais premente. Na verdade, esse é o melhor exemplo de organização por relações sintagmáticas. Mesmo a FrameNet,

enquanto plataforma, não possui essa abordagem. Uma página de resultados da FrameNet não aponta as ULs que podem instanciar cada EF. A proposta da FrameNet é agrupar em categorias as ULs que compartilham os mesmos EFs³⁴. Por esse motivo, temos categorias como a apontada anteriormente, *Noise_makers*, que arrola as ULs *drum, guitar, piano, rattle, siren* etc., mas não arrola, por exemplo, as ULs *drummer, guitarist, pianist*, nem as ULs *listener, audience, auditorium*. De fato, *Noise_makers* é uma categoria de objetos agrupados pelo traço [emissor de som] e não uma cena culturalmente estereotipada. Nossa proposta distancia-se da proposta da FrameNet nesse sentido. Propomos instanciar os EFs com as ULs correspondentes dentro do domínio léxico-conceitual correspondente.

Categorias de objetos são construídas na associação por similaridade, por conceitos que mantêm maiores ou menores semelhanças conceituais entre si. Dessa forma, podemos construir a categoria INSTRUMENTOS MUSICAIS com os conceitos lexicalizados no PB por *piano, violão, bateria, saxofone* etc. Ao contrário, os *frames* se constroem por meio de associações metonímicas, de contiguidade, fortemente pautados na experiência de mundo. Um *frame* evocado pela UL *piano*, lexicalizado pelo PB, deve arrolar as ULs *piano, pianista, recital, partitura, auditório* etc. Da mesma forma, no nível puramente conceitual, o conceito TEATRO (como hipônimo de EDIFÍCIO) é um todo constituído por conceitos-parte metonimicamente ativados: PLATÉIA, PALCO, COXIA, BOCA DE CENA entre outros. Já o conceito TEATRO (como hipônimo de ARTE) ativa todo um conjunto de conceitos paradigmaticamente relacionados: TEATRO, DANÇA, CINEMA, MÚSICA entre outros, ou seja, tipos de manifestação artística. São todos conceitos hipônimos da categoria ARTE, assim como os conceitos ATOR, ATRIZ e FIGURANTE são hipônimos da categoria ARTISTA ou, mais especificamente, ARTISTA DRAMÁTICO.

No nosso modelo, a categoria conceitual TEATRO (como hipônimo de ARTE), glosada por “o ofício ou a arte teatral”, é uma categoria com estrutura interna, cujo *frame* correspondente pode ser rotulado como *Espetáculo_teatral*. Na granulação escolhida por nós para a modelagem das categorias conceituais – tendo em vista a função de um dicionário analógico –, *ator, atriz e figurante* são ULs que instanciam o EF

³⁴ Each lexical unit is linked to a semantic frame, and hence to the other words [lexical items] which evoke that frame. This makes the FrameNet database similar to a thesaurus, grouping together semantically similar words. (RUPPENHOFER et al., 2006, Seção 1.1). Grifos nossos.

central ARTISTA DRAMÁTICO, dentro do *frame* Espetáculo_teatral, do *anaset* correspondente ao domínio léxico-conceitual TEATRO.

3.4 EQUACIONAMENTO LINGUÍSTICO-COMPUTACIONAL: TEORIA DAS REDES E REDES SEMÂNTICAS

Imaginemos uma pessoa metódica que deseja estender as roupas no seu varal de maneira organizada. Ela poderá organizá-las seguindo uma ordem de *contiguidade*: primeiro uma camisa, depois uma calça e a seguir as meias, por exemplo, combinando as peças de acordo com o uso que de fato fazemos delas no dia-a-dia. Dessa forma, a organização pode ter apenas uma direção, com dois sentidos (camisa x , calça y , meia z – meia z , calça y , camisa x). Agora, e se a mesma pessoa quiser organizar suas peças, lado a lado, por categorias, como poderá proceder? Todas as camisas primeiro, depois todas as calças, depois todas as meias? Ou todas as roupas brancas de um lado e as coloridas de outro? Ou por tamanho? etc. Imaginemos, ainda, que ela deseje, em um mundo abstrato, organizar todos esses mundos possíveis. Quantos varais se entrecruzando em diferentes direções e sentidos ela deveria ter?

Esse tipo de questão nos leva à Teoria das Redes Complexas – ou apenas **Teoria das Redes** (*Network Theory* [BARABÁSI, 2002]). Uma rede pode ser caracterizada segundo algumas propriedades estruturais e topológicas (HANNEMAN, RIDDLE, 2005) herdadas do formalismo matemático da **Teoria dos Grafos**. Formalmente, um grafo G é composto por 2 conjuntos: um conjunto (V) de objetos (**vértices**) e um conjunto (A) de pares relacionados (**arestas**) (DIESSEL, 2000). Dois vértices são adjacentes ou vizinhos se existe alguma aresta entre eles. O **grau** de um vértice corresponde ao número total de relacionamentos que esse vértice possui.

Na Teoria das Redes (doravante **TR**), os vértices são chamados **nós** e as arestas são chamadas **links**. A distribuição dos graus de um nó (*node-degree*) pode ser representada por uma função e exibida graficamente em um plano cartesiano. Em uma **rede randômica** (ERDÖS, RÉNYI, 1959), os *links* são colocados de maneira aleatória, então a maioria dos nós possui aproximadamente o mesmo número de *links*, seguindo uma **distribuição de Poisson**. Já em muitas redes complexas, observou-se que a distribuição de graus desvia de uma distribuição de Poisson e segue uma **lei de**

potência. Estas redes são chamadas de **redes livres de escala** (*scale-free network* [BARABÁSI, 2002]). A Figura 3.2 mostra, na linha de cima, dois planos cartesianos em que o eixo das abscissas refere-se ao número de *links* (k) e o eixo das ordenadas refere-se ao número de nós com k *links*. O plano cartesiano da esquerda ilustra a distribuição de graus de uma rede randômica, que segue uma distribuição de Poisson e, por isso, possui o aspecto visual de uma **Curva de sino**. Já o plano cartesiano da direita ilustra a distribuição de graus de uma rede livre de escala, que segue uma lei de potência e, por isso, possui o aspecto visual mais próximo de uma **Cauda longa** (*long tail*).

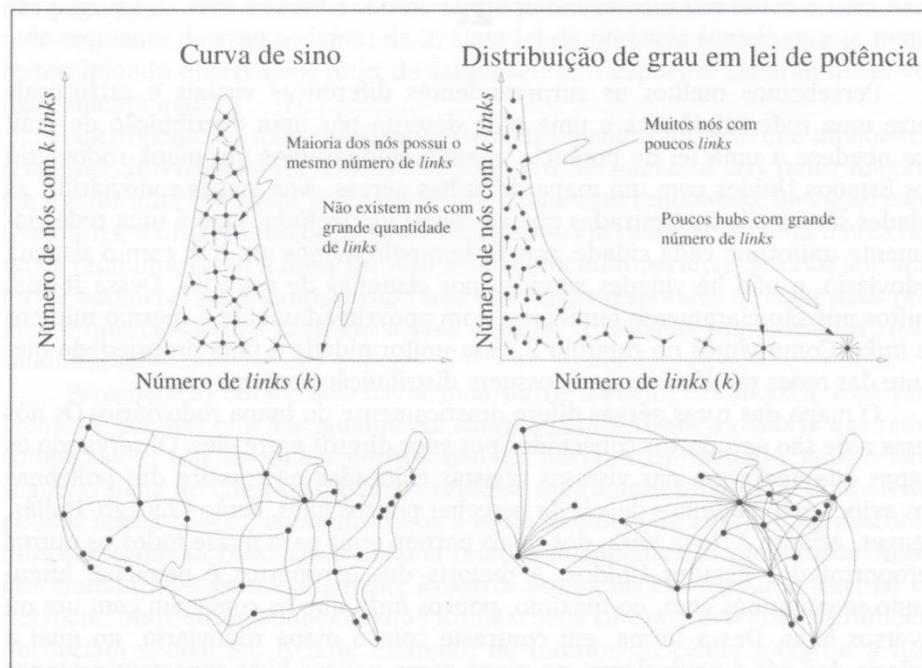


Figura 3.2. Distribuição dos nós em Curva de sino e Lei de potência e sua correlação com redes randômicas (mapa rodoviário) e redes livres de escala (mapa aeroviário)

A Figura 3.2 mostra a correlação da distribuição de Poisson (*acima à esquerda*) com uma rede randômica dando como exemplo um mapa rodoviário (*abaixo à esquerda*) e, para a correlação da lei de potência (*acima à direita*) com uma rede livre de escala, ela dá como exemplo um mapa aeroviário (*abaixo à direita*). Ou seja, em um mapa em que as cidades representam os nós da rede e as rodovias representam os *links* entre esses nós, a rede resultante é claramente uma rede randômica, já que a maioria dos nós possui aproximadamente o mesmo número de *links* e, portanto, não existem nós com muitos *links*. Ao contrário, um mapa aeroviário, em que as cidades (nós) são ligadas por rotas de aviões (*links*), a configuração resultante é de uma rede livre de escala, onde a maioria dos nós possuem poucos *links* e alguns poucos nós possuem muitos *links*.

Rede Livre de Escala (doravante **RLE**) é o nome dado às redes que evoluem a partir de um mecanismo de *conexão preferencial* (BARABÁSI, ALBERT, 1999). Como as RLEs são *dinâmicas*, e não *estáticas* como as redes randômicas (doravante **RR**), a cada período de tempo, novos nós com um número fixo de *links* são adicionados à rede. Esses *links* se conectam preferencialmente aos nós da rede com maior grau (quantidade de conexões), também chamados de **hubs** ou **conectores**. O mesmo número de rodovias, mais ou menos, chegam e partem de uma grande cidade como Nova York e de uma cidade pequena como Springfield (Massachusetts). No entanto, o mesmo não acontece em relação às rotas aeroviárias. Grandes cidades possuem um número muito maior de vôos chegando e saindo dos seus aeroportos. Em um mapa aeroviário global, cidades como Nova York, Paris e São Paulo, p. ex., são os *hubs* de uma RLE. Essas cidades recebem conexões preferenciais quando um novo nó é incorporado à rede. Quando uma pequena cidade do interior constrói o seu aeroporto, uma de suas primeiras conexões é com a capital que, sendo uma cidade importante da região, já é um *hub* com muitas conexões. No entanto, quando uma nova cidade se liga à rede por meio de rodovias, ela obrigatoriamente será conectada às cidades mais próximas, adjacentes no espaço geográfico.

Outra medida que podemos obter ao se estudar uma determinada rede é o seu **coeficiente de agrupamento** (*clustering coefficient*). O coeficiente de agrupamento é o indicador de conectividade de um nó e é obtido calculando-se a razão entre o número de relacionamentos existentes e o total de relacionamentos possíveis entre o nó e os seus vizinhos. Uma rede é tão mais *clusterizada* quanto mais os seus nós se conectam uns aos outros. Um bom exemplo são as relações sociais. Uma cidade A ligada a uma cidade B por uma rodovia *x* não precisa construir uma rodovia *y* para chegar a uma cidade C se C estiver próxima de B e ligada a ela por uma rodovia *z*. Para chegar de A até C basta usar as rodovias *x* e *z*, passando por B. No entanto, as relações sociais não possuem a mesma lógica. Uma pessoa é ligada ao pai e ligada à mãe, independentemente de o pai e mãe serem ligados entre si. Assim como os colegas de trabalho de um mesmo departamento são todos ligados entre si. A Figura 3.3 ilustra quatro grupos de pessoas, no qual todos os membros são coligados entre si. Cada um desses grupos possui, portanto, um coeficiente de agrupamento máximo.

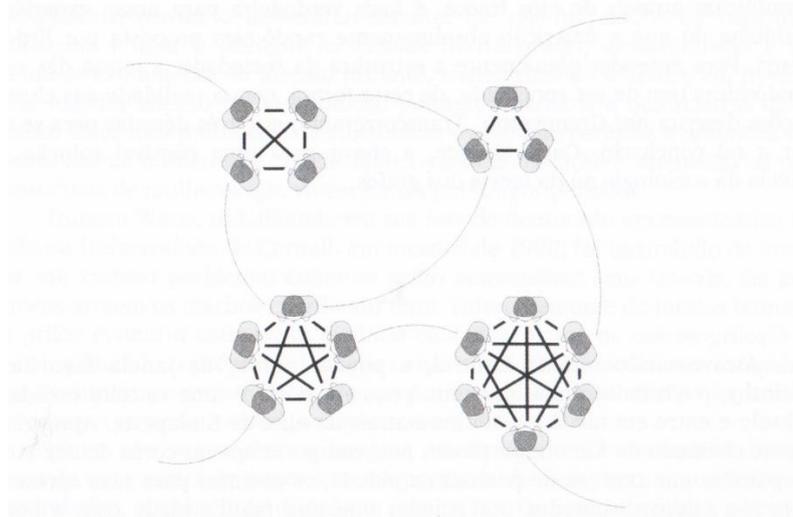


Figura 3.3. Redes sociais: exemplo de redes altamente clusterizadas (BARABÁSI, 2002)

A Figura 3.3 ilustra também os **vínculos fortes** e **vínculos fracos** de uma rede. A rede social é uma rede formada por pequenos círculos de amigos totalmente conectados (*clusters*) e que se unem por vínculos fortes (linhas grossas); e, ao mesmo tempo, possui vínculos fracos (linhas finas) que conectam membros desses círculos fechados com outros círculos. Os vínculos fracos desempenham papel crucial na sociedade. Segundo Granovetter (1973), quando se trata de atividades como arranjar emprego, propagar ideias e saber das novidades, os nossos conhecidos e colegas (vínculos fracos) são mais importantes do que os fortes laços familiares e de amizade (vínculos fortes), na medida em que são os vínculos fracos que nos conectam com experiências diferentes. É possível perceber uma relação entre a Figura 3.3 e as Figuras 2.3 e 2.4 da Seção 2.4, em que tratamos da conexão entre ideias e o seu papel para a criatividade. Algumas ideias e conceitos se relacionam fortemente, mas é exatamente por meio dos vínculos fracos que as conexões entre ideias podem ser mais interessantes e criativas.

Redes onde os nós representam palavras ou conceitos e os arcos correspondem a relações semânticas são chamadas de redes semânticas. A utilização do formalismo de nós e arcos para a representação do conhecimento foi proposta por Collins e Quillian (1969). Os arcos **é-um** (relação paradigmática) e **é-parte** (relação sintagmática) são os mais comuns em sistemas de redes semânticas. Durante os anos setenta, um artigo de Minsky (1975) introduziu a noção de nós com estrutura interna, propondo assim uma nova forma de representação do conhecimento, derivada das redes semânticas, chamada *frames*. As duas propostas de representação do conhecimento (redes semânticas e

frames) são, respectivamente, os alicerces teóricos para a construção, cerca de duas décadas mais tarde, dos léxicos-computacionais da WN.Pr e da FrameNet.

Nossa ideia é que, quando formalizadas em grafos, as relações sintagmáticas formam RRs, enquanto que relações paradigmáticas geram RLEs. Ou seja, esses “nós com estrutura interna”, os *frames*, podem ser representados, na verdade, como pequenos agrupamentos altamente *clusterizados* (aglomerados) dentro da rede. Na Figura 3.4, temos o exemplo de uma rede construída com conceitos que mantêm entre si relações sintagmáticas. O princípio de construção da rede foi adicionar um arco de relação entre os conceitos que, em suas definições, citassem uns aos outros em relações do tipo parte-todo e de contiguidade (proximidade espacial). Por exemplo, a definição para o conceito de COXIAS extraída do *Aulete Digital*: “espaço situado entre o palco e as paredes adjacentes a esse, onde os atores aguardam a hora de entrar em cena, fora das vistas do público; bastidores”. Dessa forma, foi adicionado um arco entre os conceitos de COXIAS e de PALCO, e assim sucessivamente para os demais conceitos. Estamos considerando os conceitos pertencentes à categoria conceitual PARTES DO TEATRO, dentro do domínio léxico-conceitual TEATRO.

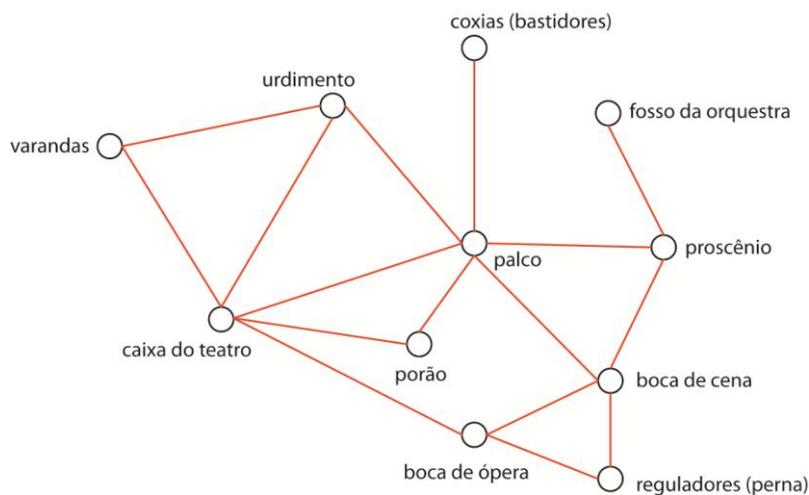


Figura 3.4. Rede semântica de relações sintagmáticas a partir da UL *palco*.

Nitidamente, a rede derivada é uma RR, em que o número de ligações de um nó não é significativamente discrepante dos demais. Mesmo o conceito PALCO que possui o maior número de conexões (6) possui apenas uma conexão a mais que o conceito CAIXA DE TEATRO (5). Mas o mais importante para a topologia desse tipo de rede é o fato de que os conceitos mantêm relações entre si. Ou seja, independente de um conceito central que sirva como conexão preferencial. Veremos esse mesmo tipo de topologia sempre

que estivermos lidando com conceitos do mundo físico (como as cidades ligadas por estradas) ou social (como no caso das relações de parentesco).

Diferentemente, uma rede resultante de relações paradigmáticas entre conceitos e ULs, normalmente, se apresenta como uma RLE. Vejamos, na figura 3.5, o aspecto de uma rede de relações paradigmáticas, inspirada na metodologia do *Visual Thesaurus*, que utiliza conjuntos de sinônimos como base de dados. Dessa vez, estamos considerando as ULs e os conceitos pertencentes ao domínio léxico-conceitual AMOR, em que cada nó representa uma categoria conceitual diferente.

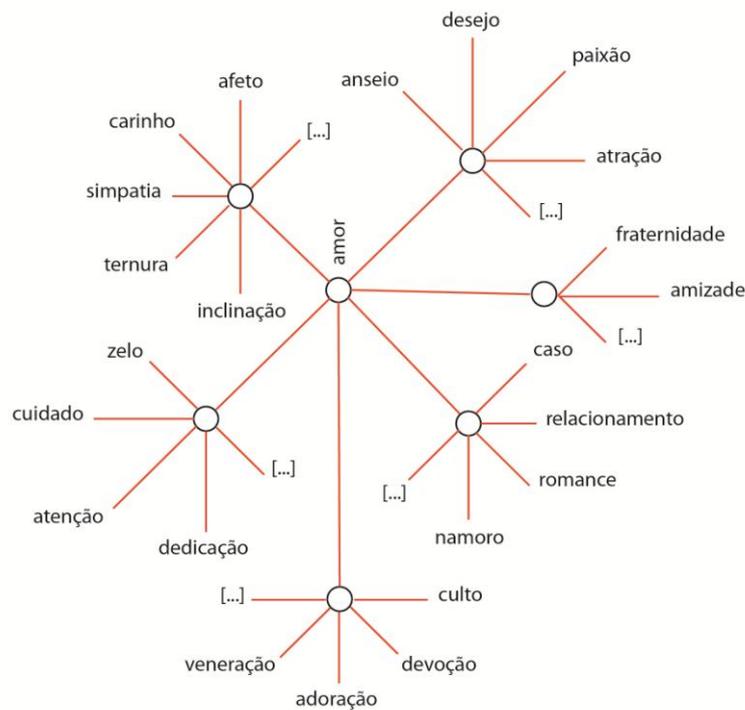


Figura 3.5. Rede semântica de relações paradigmáticas a partir do item lexical *amor*.

Como podemos perceber, a configuração das redes ilustradas em 3.4 e 3.5 são bem diferentes. A rede derivada de relações paradigmáticas tende a uma topologia de RLE. Veremos esse mesmo tipo de topologia sempre que estivermos lidando com conceitos do mundo psíquico, ou seja, conceitos abstratos.

3.4.1 REDE SEMÂNTICA PARA O DAD

Como dissemos em 3.1, a WN.Pr é uma rede semântica em que as unidades lexicais são organizadas sob a forma de *synsets*. Na WN.Pr os *synsets* são os nós e as

relações léxico-conceituais que se estabelecem entre os *synsets* constituem os arcos da rede. Nosso modelo é inspirado na WN.Pr, mas mantém com ela algumas importantes diferenças. Isso porque os objetivos das duas redes são distintos. Os *synsets* da WN.Pr representam conceitos em uma granularidade muito refinada, o que a torna uma base de dados léxico-conceitual útil a diversas outras aplicações computacionais como tradução automática e mineração de dados. No entanto, o modelo da WN.Pr não propicia a estrutura adequada para um dicionário de analogias que pretende integrar um dicionário cuja a função é estimular a criatividade e servir de ferramenta para a escrita criativa. Isso porque seus *synsets* são conjuntos fechados (*crisp*) e a relação entre eles segue uma arquitetura rigidamente hierarquizada e estável. Nosso objetivo, ao contrário, é propor um modelo de representação que satisfaça as necessidades de um dicionário criativo. O que significa construir uma rede em que os nós base são ULs – e não *synsets* –, mas que também possua nós agregadores parecidos com um *synset*, que são os CdSAs (Cf. Seção 5.5). Cada UL tem um arco de ligação para cada CdSA a que pode pertencer e essa ligação tem um peso, que é o grau de pertença ao conjunto. Dessa forma, teremos CdSAs que são os nós agregadores da rede e representam categorias conceituais, tal como ilustrado na Figura 3.6.

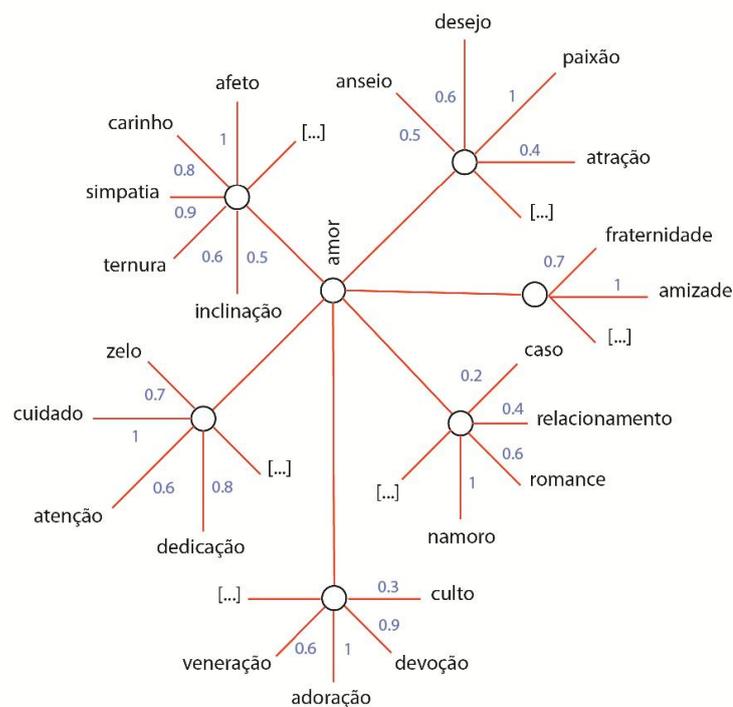


Figura 3.6. Rede semântica difusa para as relações paradigmáticas do domínio AMOR.

Diferentes cortes (dentro da escala 0–1) podem proporcionar diferentes graus de refinamento dos campos conceituais. No topo da curva de pesos as ULs estarão mais semanticamente próximas entre si instanciando a relação léxico-semântica da sinonímia; ao passo que na cauda da curva estarão ULs que evocam conceitos um pouco mais distantes do centro de sentido do CdSA, mas semanticamente relacionados, tal como ilustrado na Figura 3.7.

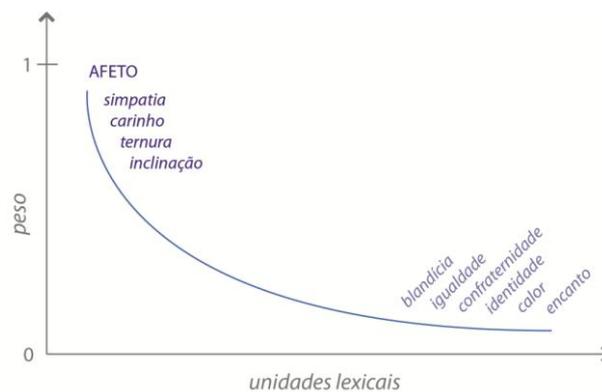


Figura 3.7. Relações de peso das ULs do CdSA AFETO (domínio AMOR).

Os conceitos evocados pelas ULs da cauda podem se relacionar com o núcleo conceitual do CdSA por meio de diferentes relações de sentido (metofóricas, metonímicas, de especialização, de generalização etc.), mas nenhuma delas será explicitada. Ou seja, não existem arcos de relação semântica tal como em uma *wordnet*. Todas as relações, no nosso modelo, são consideradas sob o mesmo rótulo de **relações analógicas**, tal como a tradição dos dicionários analógicos (Cf. Seção 2). Dessa forma, mantemos a rede com uma arquitetura difusa e fluida, o que satisfaz as necessidades e objetivos da aplicação computacional proposta. Nada impede, no entanto, que trabalhos futuros venham a incorporar alguns atributos das redes *wordnets* ao nosso modelo, produzindo assim uma rede intermediária em que, por exemplo, CdSAs possuam arcos de relação semântica iguais aos que hoje existem na WN.Pr. Por hora, nossa opção foi deixar os CdSAs o mais elásticos possível, organizando-os apenas dentro de seus respectivos domínios conceituais.

Cada CdSA estará em relação paradigmática ou sintagmática à outros CdSAs dentro de um mesmo domínio léxico-conceitual. Na verdade, o próprio método de construção de um CdSA é dependente de um domínio conceitual específico (Cf. 6.1.3). Assim como um *synset* se situa dentro da teia de relações semânticas de uma *wordnet*,

um CdSA só existe ancorado a um dado domínio conceitual. O conjunto de CdSAs dentro de um mesmo domínio equivale ao constructo que batizamos de *anaset* (Cf. Seção 6.1). A relação de pertença de um CdSA dentro do *anaset* é também uma relação difusa, o que deverá representar, segundo nossa intenção, que dentro de cada domínio conceitual existem categorias conceituais prototípicas e periféricas.

Os CSAs que se relacionam sintagmaticamente dentro de um mesmo domínio conceitual estarão explicitados como tal e terão, inclusive, uma interface própria. O CdSA {namorado (1), querido (0.81), bem (0.66), amor (0.63), amado (0.59), caro (0.55), ...} que evoca o conceito de HOMEM QUE AMA OU É AMADO está em relação sintagmática ao CdSA {namorada (1), amada (0.78), querida (0.72), amante (0.51), ídolo (0.48), amor (0.48), ...} que evoca o conceito de MULHER QUE AMA OU É AMADA. Os dois CdSAs evocam conceitos que preenchem EFs centrais do *frame* *Relação_Amorosa* (no *anaset* AMOR), tal como exibido na Figura 3.8. Neste exemplo, o EF central poderia ser rotulado como ‘amantes’ e aparece na Figura 3.8 ilustrado pelo ícone de um homem e uma mulher.

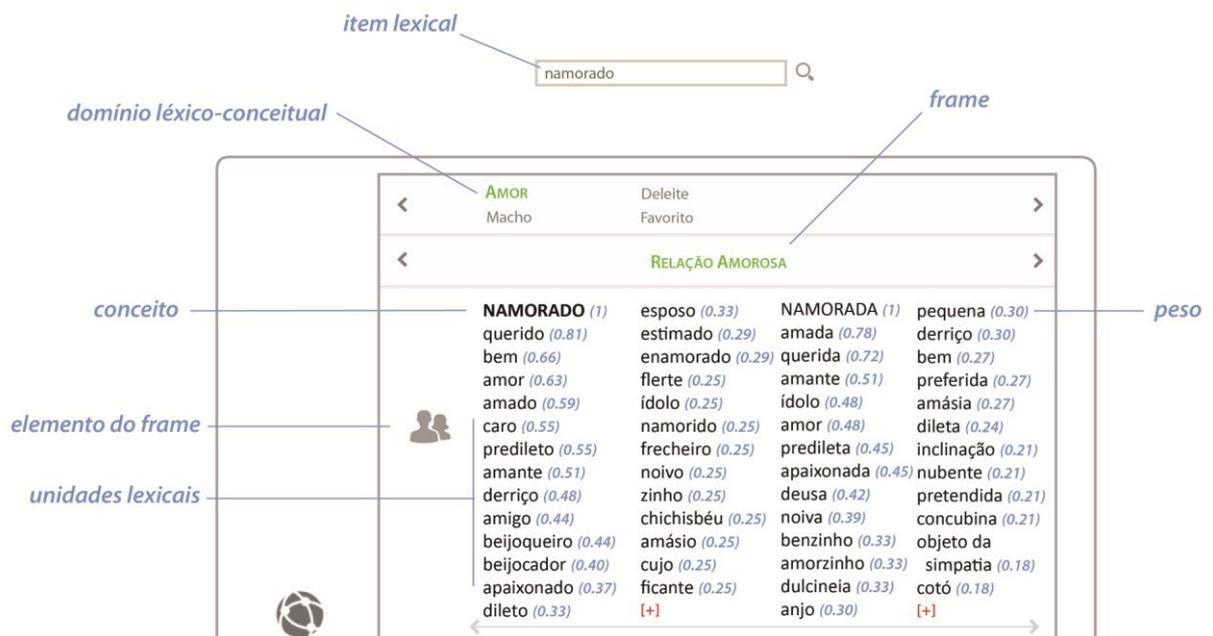


Figura 3.8. Exemplo de interface para os CdSAs NAMORADO e NAMORADA que se relacionam sintagmaticamente no interior do *frame* *Relação_Amorosa* (*anaset* AMOR).

A Figura 3.8 ilustra, portanto, que o item lexical *namorado* é a UL mais relevante do CdSA que evoca o conceito de HOMEM QUE AMA OU É AMADO, que preenche o EF central ‘amantes’ do *frame* *Relação_Amorosa* do *anaset* AMOR. Como mostra o cabeçalho dos domínios léxico-conceituais da Figura 3.8, o item lexical *namorado* pode

aparecer ainda em outros três domínios: MACHO, DELEITE e FAVORITO³⁵. Será necessário, em trabalhos futuros, modelar todos os possíveis domínios léxico-conceituais que aparecerão no DAD para saber i) se esses domínios, inicialmente propostos por Roget, de fato devem ser mantidos e ii) qual seria o posicionamento e o grau de relevância da UL *namorado* em cada um deles. Provavelmente, em nenhum deles a UL *namorado* teria a mesma relevância que tem dentro do *frame* *Relação_Amorosa*. Para incluí-la no domínio MACHO, por exemplo, a característica que está sendo levada em conta é o traço [masculino] e não os traços [que ama] ou [é amado] e, por isso, a relevância de *namorado* provavelmente será baixa, muito atrás da relevância de ULs como *homem*, *varão* e *rapaz*. Já para o domínio conceitual FAVORITO, o traço [é amado] é que está sendo considerado para arrolar a UL *namorado* ao lado de outras ULs relacionadas como *protegido*, *predileto* e *preferido*. Provavelmente, no entanto, estas estarão entre as mais relevantes do domínio, enquanto *namorado* terá uma relevância mais baixa. Talvez não tão baixa quanto quando vinculada ao domínio MACHO, mas certamente não tão alta quanto no caso de quando aparece no *frame* *Relação_Amorosa* do domínio AMOR. Essas observações são importantes para esclarecer que um dicionário analógico digital como o que estamos propondo pode organizar ULs e conceitos em diferentes configurações, segundo diferentes domínios conceituais e conforme relações paradigmáticas ou sintagmáticas. Serve também para chamar a atenção para o fato de que uma mesma UL que evoque um mesmo conceito pode ser colocada dentro de domínios léxico-conceituais distintos em razão de suas propriedades. E, ainda assim, podemos alcançar seus diferentes posicionamentos e navegar pelas diversas configurações da rede com apenas um clique ou dois (cf. Seção 7.4). É essa característica do nosso modelo que nos permite dizer que a rede semântica por detrás do DAD possui, além de uma estrutura difusa (cf. Seção 5.5), uma arquitetura em forma de rizoma (cf. Seção 4.8).

³⁵ Conforme as mil categorias do *Roget's Thesaurus* fielmente espelhadas no Dicionário Analógico do professor Azevedo.

3.5 SÍNTESE DA SEÇÃO

Nesta seção, as investigações operadas acerca do equacionamento do domínio linguístico argumentam no sentido da proposta de organizar o DAD por meio da explicitação das relações paradigmáticas e sintagmáticas na estruturação dos agrupamentos conceituais no interior de cada domínio léxico-conceitual. Para tanto, investigamos os modelos das *wordnets* e das *framenets* como modelos de referência para a arquitetura de relações paradigmáticas e sintagmáticas, respectivamente. Vimos que quando pensamos em conceitos do mundo físico, ativamos primeiramente outros conceitos relacionados metonimicamente. É dessa maneira, por exemplo, que se estrutura todo o domínio léxico-conceitual CORPO HUMANO, visto que o corpo humano é uma entidade do mundo físico. Por exemplo, quando escrevemos *pé*, acionamos um grupo de conceitos metonimicamente relacionados: PÉ, CABEÇA, MÃO, PERNA entre outros. No entanto, podemos também acionar conceitos relacionados metaforicamente. Por exemplo, quando focalizamos traços como [extremidade] + [inferior] e ignoramos o traço [corpo humano], todo um grupo de conceitos psíquicos que compartilham esses traços são simultaneamente ativados: BASE, FUNDAMENTO, APOIO, ALICERCE entre outros. É interessante perceber que não temos dificuldade em reconhecer os conceitos em relação sintagmática – como PÉ, CABEÇA, MÃO e PERNA – como conceitos independentes. Apesar de fortemente relacionados, eles não são intercambiáveis em nenhum contexto, já que denotam entidades claramente diferentes. O que os une é o fato de fazerem parte de um mesmo objeto físico: o corpo humano. Já os conceitos de BASE, FUNDAMENTO, APOIO e ALICERCE são intercambiáveis em inúmeros contextos. Ou, na verdade, o certo seria dizer que as ULs *base*, *fundamento*, *apoio* e *alicerce* é que são intercambiáveis em diferentes contextos, justamente por evocarem um mesmo conceito, ou seja, o conceito de EXTREMIDADE INFERIOR. E, portanto, a relação paradigmática se estabelece entre ULs e não entre conceitos, a saber, a relação de sinonímia, tal como no *synset* {alicerce, apoio, base, fundamento, sustentáculo} presente no TeP. No entanto, não há nos dicionários uma acepção para cada uma das ULs *alicerce*, *apoio*, *base* e *fundamento* que seja de fato equivalente. Segundo o Dicionário de Sinônimos de Antenor Nascentes (2011), *alicerce* é “a parte sólida, de alvenaria, sobre a qual assenta uma construção; *base* é “a parte inferior da construção, aquela em que ela começa a erguer-se do solo”; e *fundamento* é “o conjunto

formado pela área do solo e pelos alicerces, sobre os quais assenta uma construção; neles se firma a base”. Ou seja, em separado, nem conceitos psíquicos eles são. Só o são quando em conjunto (como no caso do *synset* supracitado) ou quando o contexto de uso claramente seleciona apenas os seus traços [extremidade] e [inferior].

As investigações perpetradas acerca do equacionamento do domínio linguístico-computacional avançaram sobre noções essenciais da TR e mostraram como tal teoria pode servir para a formalização da informação léxico-conceitual do nosso modelo, no estágio intermediário entre o domínio linguístico e o computacional. Vimos que redes resultantes de relações paradigmáticas entre conceitos e ULs, normalmente, se apresentam como RLEs, ao passo que redes resultantes de relações sintagmáticas formam RRs.

É preciso ter em mente que, apesar de nossos resultados finais não serem visualmente apresentados em forma de rede, como em um *Visual Thesaurus*, sua lógica e formalismo é uma camada intermediária entre os dados linguísticos brutos e sua codificação computacional, tal como na WN.Pr. No entanto, a arquitetura do DAD é um pouco diferente, como vimos, da arquitetura das *wordnets*. Ou seja, uma rede em que os nós base são ULs – e não *synsets* –, mas que também possui nós agregadores parecidos com um *synset*, que são os CdSAs. Sendo assim, cada UL tem um arco de ligação para cada CdSA a que possa pertencer e essa ligação tem um peso, que é o grau de pertença ao conjunto. E, conforme o que foi proposto nesta seção, os CdSAs serão organizados conforme relações paradigmáticas e relações sintagmáticas no interior de um mesmo *anaset*.

SEÇÃO 4 - RELAÇÕES HIERÁRQUICAS E CATEGORIAS SUPREMAS NA CONSTRUÇÃO DO DAD

Em 2.2 e 2.7 foi dito que a estrutura hierárquica dos dicionários analógicos, formalizadas no chamado plano de classificação das ideias, é confusa e de pouco proveito prático. Por isso, esta Seção descreve uma proposta diferente para o sistema conceitual a ser utilizado na construção do modelo para o DAD. Nossa proposta sugere uma relação hierárquica distinta daquela consagrada no *Roget's Thesaurus* e replicada no *Dicionário Analógico da Língua Portuguesa* do professor Azevedo. Nossas pesquisas nos levaram a um alinhamento com a proposta do *Historical Thesaurus of the Oxford English Dictionary* (KAY, C. et al. [eds.], 2009), que organiza todos os conceitos abaixo de apenas três grandes categorias conceituais: THE EXTERNAL WORLD, THE MENTAL WORLD, THE SOCIAL WORLD. Esse tipo de organização no topo da hierarquia conceitual, como foi dito em 2.7, demonstra-se filosoficamente mais alinhado ao nosso trabalho, ao mesmo tempo em que tem demonstrado benefícios lexicográficos, como veremos ao longo desta seção.

Tendo em vista uma plataforma que visa estimular a criatividade e que tem como usuário final falantes da língua em geral – ao invés de linguistas e filósofos – buscamos uma alternativa para superar a complexidade do sistema de classificação das palavras de tradição rogetiana. O quadro sinóptico de categorias do *Roget's Thesaurus* é uma estrutura conceitual que tem um sentido prático: ajudar o consulente a encontrar, partindo das ideias, as palavras e expressões que melhor imprimam o teor do seu pensamento. No entanto, na esteira de uma longa tradição epistemológica que vai de Aristóteles a Leibniz, Roget estrutura suas categorias e subcategorias crendo na possibilidade de construção e sistematização de uma língua ideal – que provou-se infrutífera. Se, por um lado, seu sistema conceitual não demonstrou ser menos *ad hoc* que outros esforços de representação do conhecimento, por outro, também não parece facilitar a vida do consulente (o índice remissivo é muito mais prático e, por isso, muito mais utilizado que o quadro sinóptico). O próprio *Roget's Thesaurus*, em sua última versão (KIPFER, B. A. [ed.], 2010), já traz um quadro sinóptico diferente e simplificado. O forte da sua estrutura, no entanto, são as 1000 categorias centrais, que funcionam como verbetes e que serão mantidas na nossa proposta, inclusive para manter um ponto de contato com essa tradição.

A adaptação para um modelo computacional não deve ser apenas uma mudança de suporte, precisa contemplar adequações perante as novas tecnologias e à luz das modernas teorias. O ambiente digital – com seus bancos de dados relacionais, buscas por palavras-chave, hyperlinks etc. – sugere outro tipo de relação com a informação, muito diferente daquela existente com a obra impressa. Frente às modernas concepções das ciências cognitivas – por exemplo, a ideia de *categorias de nível base* e a *teoria dos protótipos* –, todo um novo caminho de pesquisa para a estruturação do léxico se abriu. Nossa proposta entende que a adaptação para um modelo computacional encerra algumas críticas ao que foi desenvolvido e que vem sendo replicado para o modelo impresso. Para além dessa mudança de suporte tecnológico, faremos nesta Seção uma revisão de caráter filosófico das categorias aristotélicas e demais propostas para sistemas conceituais frente às modernas concepções das ciências cognitivas e em comparação com os demais esforços de classificação das ideias e do léxico das línguas. Por fim, apresentaremos nossa proposta de estruturação hierárquica para o DAD, explicitando os benefícios lexicográficos que motivaram a nossa abordagem em contraste com as abordagens tradicionais.

4.1 AS CATEGORIAS SUPREMAS DE ARISTÓTELES

Segundo Angioni (2006), *katêgoria* significa, em muitos contextos da obra aristotélica, “predicação” entendida abstratamente como “ato de predicar” ou, ainda, “denominação”, “designação”. Mas, na obra em questão, o termo adquiriu um sentido mais estrito, o mesmo consagrado pela tradição, o de *categoria*, i.e., os “gêneros supremos do ser”. Para obter o quadro das categorias supremas, Aristóteles se coloca a seguinte questão: “a que tipo de coisa dada no mundo, ou a que tipo de situação ou estado de coisas, tal e tal termo se reportam?”. A abstração máxima dessa questão, que pode ser expressa pelas fórmulas “o que é x?”, “x é de que tipo?” etc. e sua solução “x é um tipo de y”, é a essência de qualquer hierarquia conceitual. Ela formula a principal relação lógico-conceitual de uma ontologia³⁶, a relação “is-a” (é-um), como em “Sócrates

³⁶ Como demonstram Guarino e Giaretta (1995), o termo “ontologia” é controverso e faz proliferar uma grande variedade de definições distintas, ainda que compartilhem inúmeros traços e flutuem todos à volta de um mesmo núcleo duro. O termo nasce na filosofia (em grego, *ontos* e *logos* traduzem “conhecimento do ser”) e

é um homem” e “o homem é um animal”. Ao lado da relação “part-of” (é-parte-de), essa é a principal relação e a base dos diferentes tipos de ontologia. Conforme o domínio do conhecimento e o segmento da realidade que a ontologia pretende modelar, essa relação possui nomes específicos. No domínio Linguístico chamamos essa relação de *hiperonímia* ou *hiponímia*, conforme a direção do vetor. O termo mais específico é um hipônimo de um termo mais geral, que é, portanto, seu hiperônimo. Dado o exemplo acima, dizemos que *homem* é hiperônimo de *Sócrates* e hipônimo de *animal*. Em outros domínios do conhecimento, essa mesma relação pode ser chamada de subordinação/superordinação, subconjunto/superconjunto etc. Em termos lógicos a hiponímia é entendida como inclusão de classes: se X é a classe dos *animais* e Y é a classe dos *homens*, podemos dizer que X inclui estritamente Y.

Depois de reduzir seus conceitos unívocos uns aos outros, em espécies e gêneros sucessivos, Aristóteles chegou a dez gêneros supremos, considerados irreduzíveis, que ficaram conhecidas como **categorias aristotélicas** (ANGIONI, 2006). A lista proposta por Aristóteles é famosa:

- 1a.) *substância* - o que se atribui como existente em si, como em seu sujeito;
- 2a.) *quantidade* - determinação atribuída a um ser, pela qual ele se distribui em partes ao lado das partes;
- 3a.) *qualidade* - determinação que afeta intrinsecamente o objeto, fazendo-o uma tal e qual coisa;
- 4a.) *relação* - determinação que atribui ao objeto o estar em relação para outro;
- 5a.) *tempo* - o existir enquanto dura;
- 6a.) *lugar* - determinação de um ser, em que ele incorre enquanto está em referência a um outro;
- 7a.) *posição* - determinação resultante das partes de um corpo com referência ao lugar de outro;
- 8a.) *ação* - determinação resultante do exercício de causar;
- 9a.) *paixão* - determinação resultante do exercício de causar, correspondente à direção inversa, pela qual o sujeito padece como receptor

abarca o estudo da natureza do ser, da realidade, do que existe e de questões metafísicas em geral. Ainda que o termo não apareça no texto das *Categorias* de Aristóteles, podemos dizer que ele contém o primeiro esforço, que se tem registro, na direção de se modelar uma ontologia. Falaremos mais sobre ontologias na próxima seção.

10a.) *posse* (ou *hábito*) - determinação que se cria com a adjacência de outro objeto.

4.2 HIERARQUIAS, TAXONOMIAS, E ONTOLOGIAS LINGUÍSTICAS

Na obra intitulada *Introductio in Praedicamenta*, o filósofo neoplatônico Porfírio comenta a obra *Categorias*, de Aristóteles, e incorpora a lógica aristotélica ao neoplatonismo. Nesse mesmo livro, encontra-se a famosa "Árvore de Porfírio" (*Arbol porphyriana*), que ilustra sua classificação lógica da substância em termos de uma cadeia de conceitos subordinados, partindo dos mais gerais para os mais específicos. A "Árvore de Porfírio" pode ser vista como uma antecessora das modernas classificações taxonômicas e pode ser esquematizada como na Figura 4.1.

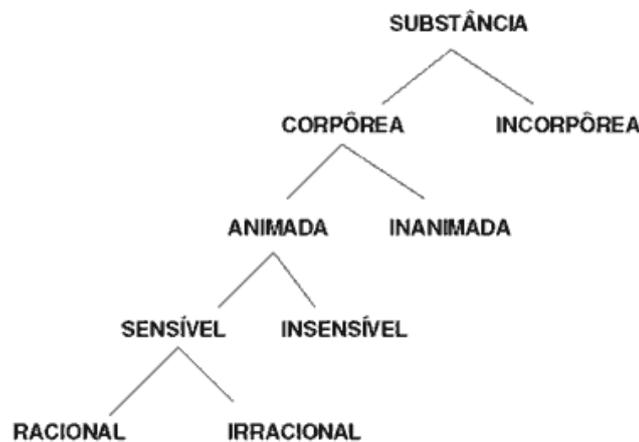


Figura 4.1. Árvore de Porfírio (ECO, 1984, p. 463).

Os *thesauri*, como o *Roget's Thesaurus*³⁷, também estruturam relações léxico-conceituais hierárquicas. No entanto, a hierarquia por detrás de um *thesaurus* ou de um dicionário analógico (cf. 4.3) é uma hierarquia altamente informal. Lyons (1977, p. 242) cita o *thesaurus* de Roget exatamente na sua Seção dedicada à "estrutura hierárquica do vocabulário". O autor acredita que, embora o *Roget's Thesaurus* tenha sido inicialmente

³⁷ 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.

idealizado para *facilitar a expressão das ideias e auxiliar na composição literária*, ele foi bastante influenciado pela filosofia do século XVIII (fortemente marcada pelas ideias de Francis Bacon, Descartes e Leibniz), que acreditava ser possível a construção de “uma linguagem ideal para a sistematização e desenvolvimento do conhecimento científico”. Lyons cita igualmente, na Seção supracitada, o **sistema conceitual** (*Begriffssystem*) de Hallig e Wartburg (1952), um dos mais ambiciosos *thesauri* com uma estrutura hierárquica de categorias e subcategorias conceituais. Sob a justificativa de que não estão a classificar as palavras, mas sim os signos linguísticos que representam os conceitos, Hallig e Wartburg (1952, p. 96) apresentam a divisão tripartite – *O Universo, O Homem e O Homem e o Universo* – que defendem convir a qualquer sistema linguístico, na medida em que distingue conceitos que podem ser reconhecidos universalmente, por todas as línguas.

As categorias aristotélicas e os demais sistemas conceituais construídos ao longo dos séculos procuram estruturar tudo o que existe no mundo, por meio do exercício filosófico dessas relações basilares, de hierarquia e de pertença. Hoje, outras áreas do conhecimento, especialmente as da Computação e a da Ciência da Informação, procuram estruturar, a partir das mesmas relações, hierarquias de diferentes domínios e que sejam computacionalmente tratáveis. Junto com a relação ‘part-of’ (meronímia), a relação ‘is-a’ (hiperonímia) é a principal relação lógico-conceitual utilizada na construção dos diferentes tipos de ontologias computacionais.

Nesse sentido computacional, as ontologias podem também ser chamadas de *bases de conhecimento* e ocupam lugar central no campo de estudos denominado **Representação do Conhecimento**. As *bases de conhecimento*, ou ainda *bases conceituais*, são como um “modelo de mundo”, onde estão descritos tipos de objetos, eventos, propriedades e os relacionamentos entre eles (ALLEN, 1995). No âmbito da Computação, uma das definições mais usuais para ontologia é a de que ela é “uma especificação de uma conceitualização caracterizada por propriedades formais e propósitos específicos” (GRUBER, 1993). Ontologias formais e computacionalmente tratáveis possuem diversas aplicações específicas em Tecnologia da Informação (TI). Desempenham, por exemplo, um papel fundamental nos sistemas de PLN porque limitam a “visão de mundo” simulada pelo sistema (DIAS-DA-SILVA, 1996). O Quadro 4.1, adaptado de Vossen (2003) por Marcellino, Giroto e Dias-da-Silva (2010), ilustra o

objetivo e os tipos de artefatos ontológicos propostos a partir de cada área do conhecimento:

Quadro 4.1. Diferentes objetivos e artefatos ontológicos para cada área do conhecimento.

Tradição	Objetivo	Artefato criado
Filosofia	Categorização de entidades em termos de tipos lógicos e de padrões	Ontologias formais
Ciência Cognitiva	Construção manual de pequenos sistemas para domínios específicos, usando lógicas de descrição que têm uma semântica mais precisa, são mais expressivas e suportam a implementação de funções	Redes conceituais e <i>frames</i>
Inteligência Artificial	Desenvolvimento e construção de ontologias para tarefas e sistemas específicos, moldados para permitir decisões num sistema de informação	
Semântica Lexical	Construção de léxicos em que as unidade léxicas são definidas em termos de suas relações umas com as outras	Léxicos
Lexicografia	Definir unidade léxicas em dicionários para consulta por usuários humanos	Dicionários
Ciência da Informação	Desenvolver, geralmente sem o uso de técnicas linguísticas, a categorização de informação em termos de tópicos para sua recuperação em, por exemplo, bancos de dados	<i>Thesaurus</i>

Fonte: Marcellino, Giroto e Dias-da-Silva (2010)

Na esfera do PLN e da Inteligência Artificial (doravante IA), dois tipos de ontologia podem ser identificados com nitidez: as chamadas *ontologias linguísticas* e as *ontologias conceituais* (VOSSSEN, 1998a; PALMER, 2001). As primeiras caracterizam-se por conter apenas conceitos lexicalizados, ou seja, conceitos expressos por uma ou mais unidades lexicais de uma língua. As *ontologias conceituais*, ao contrário, caracterizam-se pelo armazenamento de conceitos que não estão lexicalizados como, por exemplo, os conceitos COISA PARCIALMENTE TEMPORAL e PARTES DO CORPO HUMANO (VOSSSEN, 1998a; PALMER, 2001). Para determinados propósitos, esses níveis não-lexicalizados são importantes para uma estruturação mais controlada dos conceitos. Ao mesmo tempo, as ontologias conceituais podem ignorar conceitos lexicalizados que não sejam suficientemente relevantes para os seus fins.

Uma ontologia conceitual também pode servir de embasamento para a construção de um léxico estruturado: partindo de conceitos, e suas inter-relações, investiga-se as palavras que lexicalizam esses conceitos (percurso onomasiológico). Do mesmo modo, um léxico enriquecido de uma hierarquia semântica automaticamente desenha uma base sólida para a construção de uma ontologia conceitual. Nada impede

que esses dois constructos – ontologias linguísticas e ontologias conceituais – sejam complementares e ocupem diferentes camadas de informação semântica. A base de um modelo como esse está em destacar os sentidos linguísticos e as relações de sentido lexical (sinonímia, hiponímia e antonímia) de suas contrapartes lógico-conceituais (conceito, inclusão e disjunção) preservando as duas camadas em paralelo e ao mesmo tempo relacionadas. Tal modelo, também chamado de **ontoléxico**, se mostra “um sistema que é ontologicamente robusto e linguisticamente motivado” (PRÉVOT, BORGO, OLTRAMARI, 2005, p. 2).

Entre as ontologias linguísticas, as mais estudadas em PLN são *Mikrokosmos* (VIEGAS *et al.*, 1996), *SENSUS* (HOVY, 1998) e a WN.Pr (FELLBAUM, 1998), sendo esta última a mais difundida e a única com versão projetada para o português do Brasil (DIAS-DA-SILVA *et al.*, 2002), conforme 3.1. A WN.Pr foi construída, inicialmente, com base no *Brown Corpus (Standard Corpus of Presente Day Edited English)* (FRANCIS, 1965) e no *Thesaurus* do Roget. No entanto, diferentemente dos sistemas conceituais citados anteriormente, o topo da ontologia linguística da WN.Pr é ocupado por apenas um conceito, o de ENTIDADE (*entity*), que subdividi-se, posteriormente, em ENTIDADE FÍSICA (*physical entity*), ENTIDADE ABSTRATA (*abstract entity*) e COISA (*thing*).

Derivada da WN.Pr, a EuroWordNet (EWN) (VOSSSEN, 1998) é uma base de dados lexicais multilíngue, estruturada segundo os mesmos princípios da rede americana. No entanto, a estrutura hierárquica no topo da ontologia da EWN é muito diferente da estrutura supracitada da base lexical de Princeton. A *Top Ontology* da EWN corresponde à divisão em tipos de entidades, conforme proposta em Lyons (1977) e é assim descrita por Vossen *et al.* (1997):

The first level of the Top Ontology is divided into three types:

- a. *1stOrderEntity* (roughly corresponding to concrete, perceivable objects and substances)
- b. *2ndOrderEntity* (states, situations and events)
- c. *3rdOrderEntitiy* (mental entities such as ideas, concepts, knowledge)

Segundo Lyons (1977), as entidades de 1ª ordem denotam entidades concretas – conceitos do tipo “objeto concreto discreto” – que são canonicamente expressas por nomes concretos. Conceitos desse tipo categorizam referentes perceptíveis pelos sentidos, 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. Já as entidades de 3ª ordem são sempre expressas por nomes abstratos e denotam entidades mentais (ideias, teorias, doutrinas, conceitos etc.).

4.3 O PLANO DE CLASSIFICAÇÃO DAS IDEIAS

A hierarquia por detrás de um *thesaurus* ou de um dicionário analógico é uma hierarquia altamente informal. Conforme 2.2, a macroestrutura dos dicionários analógicos de tradição rogetiana possui quase sempre quatro componentes características, das quais a primeira é o chamado **plano de classificação das ideias** (Quadro 4.2) – que é o esqueleto conceitual do dicionário. O plano de classificação das ideias contém i) na coluna à esquerda, as seis classes que estão no topo da hierarquia rogetiana; ii) na coluna do meio, as divisões de cada classe; e iii) na coluna da direita, os números correspondentes a cada verbete do dicionário. A mesma macroestrutura – com as mesmas classes e as mesmas divisões – está presente no *Roget's Thesaurus* (ROGET, 1852), no *Dictionnaire Idéologique* (ROBERTSON, 1859) e no *Dicionário Analógico* (AZEVEDO, [1950] 2010).

Quadro 4.2. Plano de classificação das ideias

<i>Classes</i>	<i>Divisões</i>	<i>Números</i>
I. Relações Abstratas	I. Existência	1-8
	II. Relação	9-24
	III. Quantidade	25-57
	IV. Ordem	58-83
	V. Número	84-105
	VI. Tempo	106-139
	VII. Mudança	140-152
	VIII. Causa	153-179
II. Espaço	I. Em Geral	180-191
	II. Dimensões	192-239
	III. Forma	240-263
	IV. Movimento	264-315
III. Matéria	I. Em Geral	316-320
	II. Inorgânica	321-356

	III. Orgânica	357-449
IV. Entendimento	I. Formação das ideias	450-515
	II. Comunicação das Ideias	516-599
V. Vontade	I. Individual	600-736
	II. Com referência à sociedade	737-819
VI. Afeições	I. Em Geral	820-826
	II. Pessoais	827-887
	III. Simpáticas	888-921
	IV. Morais	922-975
	V. Religiosas	976-1000

Os dicionários que seguem essa macroestrutura do *Roget's Thesaurus*, possuem mil verbetes organizados numa taxonomia com apenas três níveis de ramificação. No Dicionário Analógico do professor Azevedo, temos a seguinte estrutura hierárquica para uma das acepções do item léxico *amor*:

palavra: amor
grupo analógico: amor
divisão: afeições simpáticas/ sociais
classe: afeições

São menos níveis do que na WN.Pr. O *synset {love}* na WN.Pr – glosado por “*a strong positive emotion of regard and affection*” –, por exemplo, possui 7 níveis de ramificação até o *synset {entity}*, o mais alto da hierarquia:

S: (n) love
S: (n) emotion
S: (n) feeling
S: (n) state
S: (n) attribute
S: (n) abstraction, abstract entity
S: (n) entity

4.4 GRANDES CATEGORIAS EM COMPARAÇÃO

O Quadro 4.3 compara as Grandes Categorias (*major headings*) – ou seja, as mais altas categorias dentro da hierarquia – dos diferentes exemplos de sistemas conceituais que foram citados até o momento:

Quadro 4.3. Grandes Categorias (*major headings*) em comparação.

<i>Categorias</i> ARISTÓTELES	Substância, Quantidade, Qualidade, Relação, Lugar, Tempo, Posição, Estado, Atividade e Passividade.
<i>Thesaurus of English Words and Phrases</i> ROGET (1852)	Abstract Relations, Space, Matter, Intellect, Volition, Affections.
<i>Dictionnaire Idéologique</i> ROBERTSON (1859)	Rapports Abstracts, Espace, Matière, Intelligence, Volonté, Affections.
<i>Dicionário Analógico da Língua Portuguesa</i> AZEVEDO ([1950] 2010)	Relações Abstratas, Espaço, Matéria, Entendimento, Vontade, Afeições.
<i>Begriffssystem</i> (Sistema Racional de Conceitos) HALLIG & WARTBURG (1952)	O Universo, O Homem, O Homem e o Universo.
<i>WordNet (WN.Pr)</i> Princeton University (MILLER <i>et al.</i> , 1990) http://wordnet.princeton.edu/	Entity (physical entity, abstract entity, thing).
<i>EuroWordNet (EWN)</i> University of Amsterdam (VOSSSEN <i>et al.</i> , 1998)	1 st Order Entity, 2 nd Order Entity, 3 rd Order Entity

Não podemos equiparar *pari passu* essas diferentes estruturas e suas categorias, mas o fato é que, excluindo-se da lista as categorias aristotélicas, todas as demais foram utilizadas como categorias estruturantes de obras que procuram organizar o léxico. E é por esse viés que pretendemos compará-las nesta seção, com o objetivo de confrontá-las e, dessa forma, extrair os melhores modelos para os propósitos do nosso trabalho.

Podemos ver que o maior número de categorias foi proposto por Aristóteles. Dez diferentes categorias, com a particularidade de que o estagirita reserva lugar de destaque para a categoria das SUBSTÂNCIAS (ANGIONI, 2006). Para Aristóteles, as substâncias são sempre o sujeito de uma predicação. Para ser considerada uma autêntica predicação, uma sentença deve ter como sujeito uma entidade que seja uma substância. As demais categorias contêm aquelas entidades com as quais se pode predicar uma substância: “x é branco”, “x é justo”, “x é musical”. O item x não pode ter a mesma natureza ontológica de seus predicados. Aristóteles diz que esse tipo de

predicação exprime um *ente por concomitância*. No entanto, existe um tipo diverso de predicação em que o sujeito e o predicado possuem a mesma natureza ontológica, ou seja, um dos fatores pode ser tomado como elemento do outro. Se o resultante da composição entre sujeito e predicado não é diferente da unidade dada anteriormente no sujeito, diz Aristóteles que se trata de uma predicação que exprime um *ente em si mesmo*. “Sócrates é um homem” e “o homem é um animal” são exemplos desse tipo de predicação – exatamente o tipo de predicação que edifica uma hierarquia.

Apesar de ser consenso afirmar que Roget inspirou-se nas categorias aristotélicas para derivar suas seis grandes categorias (BABINI, 2006), o seu *plano de classificação das ideias* não preserva o lugar de destaque que Aristóteles atribuiu à categoria das SUBSTÂNCIAS. Tal categoria sequer aparece nominalmente citada entre as seis *major headings* do *Roget's Thesaurus*, tendo como equivalente mais próxima a categoria MATÉRIA, recortada pelas divisões *em geral, inorgânica e orgânica*. Robertson (1859) e Azevedo (1950) se utilizam da mesma macroestrutura geral de Roget, como podemos observar no Quadro 4.3. As três obras possuem 1000 grupos analógicos subordinados a 24 divisões das seis grandes categorias. Apesar de ser o mais conhecido, o mais replicado e o mais estudado dicionário conceitual dos tempos modernos, Roget nunca explicitou as razões teóricas e filosóficas que justificassem o sistema conceitual do seu *Thesaurus* tal como ele é (OLD, 2003). Portanto, para Lyons (1977, p. 242), é difícil justificar, em termos hierárquicos, as seis classes principais de Roget, assim como é difícil justificar a divisão tripartite do sistema conceitual de Hallig e Wartburg.

Na introdução do trabalho de Hallig e Wartburg (1952) temos mais detalhes dos princípios norteadores da obra. Por exemplo, os autores dizem que a obra i) deve ser o mais geral possível, não se restringindo ao vocabulário de uma língua ou dialeto; ii) os conceitos devem ser pré-científicos; iii) a escolha e classificação dos conceitos devem partir de um princípio em que o todo se constitua em um conjunto organizado. No entanto, essa organização não parece ser menos *a priori* do que a proposta por Roget, sendo, contudo, completamente diversa desta. Os autores, como vemos no Quadro 4.3, apresentam uma divisão tripartite no topo de seu sistema conceitual. Na primeira parte, O UNIVERSO, tem-se os conceitos referentes à natureza orgânica e inorgânica – tal como a categoria MATÉRIA de Roget –, subdividida em quatro partes: *I - O céu e a atmosfera; II - A terra; III - As plantas; IV - Os animais*. Na segunda parte, O HOMEM, apresentam-se os conceitos referentes ao homem, subdividida em quatro partes: *I - O homem, ser físico; II -*

A alma e o intelecto; III - O Homem, ser social; IV - A organização social. Na última parte, O HOMEM E O UNIVERSO, figuram os conceitos relacionados ao homem, em face de si mesmo e do mundo, organizados sob as divisões: *I - O a priori; II - A ciência e técnica.*

A ontologia por detrás do léxico-computacional da WN.Pr adota uma estrutura arbórea que, de fato, parte de apenas uma raiz – a categoria suprema ENTITY (entidade). Em outras palavras, o que essa ontologia está dizendo é que tudo o que existe, física ou abstratamente, é uma entidade. Colocamos entre parênteses, na nossa tabela, também, o nível imediatamente abaixo da *major heading* ENTITY: PHYSICAL ENTITY, ABSTRACT ENTITY e THING. Não há, no entanto, consenso com relação a essa hierarquia de um só lexema na raiz. Lyons (1977, p. 240) já chamava a atenção para o fato de que o conceito ENTIDADE (lexicalizado em língua portuguesa pela UL *entidade*) não ocupa satisfatoriamente essa posição de ascendência de maneira exclusiva. Ainda que não façamos a crítica com relação às partes distintas do discurso que, por si só, impediriam a hierarquização a partir de uma única UL, Lyons considera que mesmo se atendo à categoria gramatical dos substantivos não são plausíveis as hierarquias desse tipo, que se originam de um único ponto:

Consideremos, em primeiro lugar, os nomes em português: não há, nesta língua, nenhum lexema que seja super-ordenado em relação a todos os nomes. Mesmo a palavra mais ou menos técnica ‘entidade’ falha nesse aspecto, uma vez que cobre apenas os nomes numeráveis, e os seus equivalentes mais próximos no português quotidiano, ‘coisa’ e ‘objeto’, são ainda mais restritos. Não há nenhum lexema que seja super-ordenado a todos os nomes abstratos, ou a todos os nomes concretos, ou a todos os nomes massivos, ou a todos os membros de qualquer das subclasses principais de nomes que normalmente se reconhecem em português.

Coerente com esse pensamento, Lyons (1977) sugere uma hierarquia tripartite – que, no entanto, é absolutamente diversa daquela de Hallig e Wartburg (1952). Como foi apontado no final da Seção 4.2, a divisão tripartite de Lyons em entidades de 1ª, de 2ª e de 3ª ordem é hoje utilizada no projeto da EuroWordNet. A ontologia da EWN foi concebida para tentar uniformizar as mesmas relações lógico-conceituais da WN.Pr para as línguas européias. O caminho adotado leva em conta outros importantes trabalhos em Semântica Lexical junto com o trabalho de Lyons. Os *synsets* de conceitos concretos, por exemplo, são classificados sob a categoria das entidades de 1ª ordem e podem ser especificados segundo a **estrutura Qualia** (PUSTEJOVSKY, 1995), que explicita os quatro papéis fundamentais do significado de uma UL: i) **Constitutivo** (*Constitutive*), que exprime a relação entre um objeto e suas partes; ii) **Formal** (*Formal*), que distingue

o objeto em um domínio mais amplo; iii) **Télico** (*Telic*), que revela a função do objeto; e iv) **Agentivo** (*Agentive*), que considera fatores envolvidos na origem do objeto. Na EWN, por exemplo, a UL *veículo* é classificada como Artefato (agentivo) + Objeto (formal) + Veículo (télico), dentro da categoria das entidades de 1ª ordem. Tal estrutura gera uma trama de 63 características que podem ser conjugadas, como na Figura 4.2. É, portanto, uma estrutura extremamente robusta e teoricamente bem embasada, mas que, por isso mesmo, encerra também uma terminologia bastante técnica.

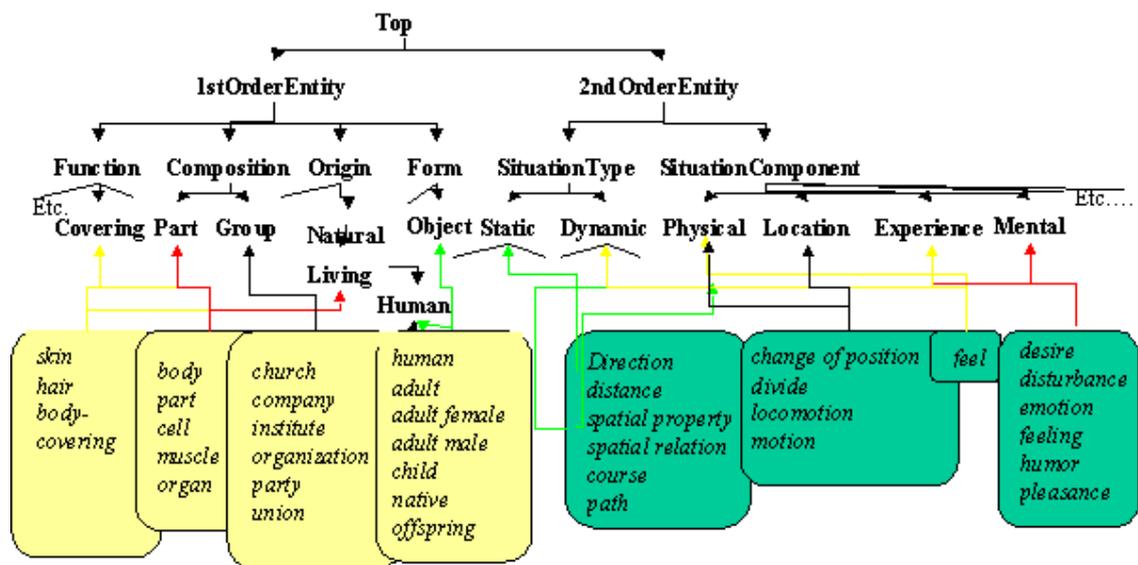


Figura 4.2. Top Ontology da EuroWordNet

4.5 HIERARQUIAS E COMPONENTES SEMÂNTICOS

Segundo Lyons (1977, p. 256), as ideias para uma linguagem ideal ou um sistema conceitual racional, estruturado em categorias e hierarquias, tais como os abordados nesta seção, se relacionam, até certo ponto, com o método estruturalista de descrição do significado conhecido como **análise componencial**. As duas abordagens compartilham inconsistências e problemas, tanto de caráter prático quanto teórico.

Os primeiros estruturalistas a proporem a análise componencial, ainda que com perspectivas diferentes, foram Jakobson e Hjelmslev. Sustentavam-se na tese de que o sentido de um lexema pode ser analisado em termos de um conjunto de componentes de sentidos mais gerais (também chamados de componentes, traços ou primitivos semânticos). O principal problema era então definir quais seriam esses primitivos

semânticos. Ou seja, determinar quais eram os conceitos atômicos dos quais os conceitos moleculares eram derivados. Historicamente, foram sugeridos três modos (HAIMAN, 1980 *apud* ECO, 1984): i) os primitivos deveriam ser os conceitos mais simples; ii) dependeriam diretamente da nossa experiência do mundo e incorporados por ostensão; iii) são idéias inatas de caráter platônico. A seguir, vamos analisá-los um a um.

Segundo Eco (1984, p. 88), não é nada fácil definir o que é um conceito simples, além do risco dos conceitos simples se apresentarem em maior número do que os conceitos complexos, na medida em que tomemos o caminho, por exemplo, de Wierzbicka (1972, p. 21) para as palavras simples:

nomes para as partes do corpo e para objetos que ocorrem na natureza - como mar, rio, campo, bosque, nuvem, montanha, vento etc. -, para artefatos humanos como mesa, casa, livro, papel etc. As expressões que num certo sentido não podem ser explicadas são as palavras para as 'espécies' (no sentido lato do termo): gato, rosa, maçã, cana, ouro, sal etc.

Ainda conforme Eco, “para um falante comum é mais simples, no sentido de que é mais facilmente compreensível, o conceito de 'homem' do que o de 'mamífero’”, ainda que, teoricamente, o segundo seja um conceito mais basilar do que o primeiro, exatamente por ser mais geral. O segundo modo também não é consistente. Dizer que os conceitos atômicos são dados de modo empírico e por ostensão, admite a inconsistência de que uma determinada experiência seja rara para a maioria dos falantes e, portanto, não produza conceitos atômicos, ao passo que a mesma experiência seja corrente para algum indivíduo ou conjunto de indivíduos em especial, o que a tornaria uma fonte de primitivos para esse grupo. A terceira e última proposta de como identificar os primitivos semânticos é diametralmente oposta a esse segundo modo e, no entanto, também é incapaz de esclarecer a questão satisfatoriamente. Se os primitivos são ideias universais inatas, de caráter platônico,

Ou há uma idéia para cada gênero natural (a cavalinidade) e então a lista é aberta. Ou há poucas idéias muito mais abstratas (como o Uno e o Múltiplo, o Bem, os conceitos matemáticos) e então não bastam para distinguir os significados dos termos lexicais. (ECO, 1984: 88)

Esse último modo, intrinsecamente conceptualista, é de fato o mais frágil deles. Poucos autores modernos, entre eles Katz (1981), movem-se nesta direção. Ao passo que os dois primeiros estão, de alguma forma, presentes na Semântica Cognitiva, por meio da teoria das **categorias de nível básico** (ROSCH, 1978; ROSCH, MERVIS, 1975) e do experiencialismo ou, em sua versão mais recente, o **realismo corporificado**

(*embodied realism*) (LAKOFF, JOHNSON, 1999), respectivamente. Ainda que não estejam aí de maneira a endossar ou reformular a tese do atomismo semântico – negado pelo paradigma cognitivista –, mas contribuindo para modernas teorias de categorização e conceptualização, conforme veremos em 4.6.

Ainda com relação ao segundo modo, Eco diz que existe, nesse caso, um paradoxo de natureza “teórica”, na medida em que toda a discussão em torno de uma lista de primitivos está fundamentada na ideia de explicar uma competência linguística separada do conhecimento de mundo, o que de fato não acontece nesse caso. É a clássica dicotomia entre conhecimento definicional (conhecimento das propriedades essenciais) e conhecimento enciclopédico (conhecimento de propriedades contingentes). O conhecimento enciclopédico está para a enciclopédia assim como o conhecimento definicional está para o dicionário. Identificar propriedades essenciais e relações sistemáticas entre elas é o que está no horizonte da visão estruturalista. Para tanto, segundo Eco, resta apenas uma quarta possibilidade – um sistema de primitivos tal que a relação de encaixamento recíproco entre eles tenha que ser finito, gerando uma árvore – com começo, meio e fim – de relações hierárquicas. E aqui estamos de volta à árvore porfiriana. Diz Eco (1984, p. 99):

Quando Aristóteles falava de inventário finito (*Secondi Analytici*, 83a et seqs.), partia das substâncias primeiras e procurava defini-las inventando, por assim dizer, árvores quase *ad hoc*, enquanto Porfirio não evita a tentação neoplatônica de conceber (ainda que em sentido lógico) uma 'cascata dos seres'. O fato é que em toda teoria dos inventários finitos funciona uma forma *mentis* neoplatônica, embora totalmente secularizada.

Eco vai argumentar que quando se introduz a diferença específica em uma árvore de hipônimos e hiperônimos, a árvore deixa de ser um exemplo de dicionário e torna-se inevitavelmente uma enciclopédia.

Mas podemos dizer sem simulação que a árvore dos gêneros e das espécies, de qualquer modo que seja construída, explode numa poeira de diferenças, num turbilhão infinito de acidentes, numa rede não hierarquizável de qualia. O dicionário (porque é como tal que a árvore nos interessa hoje, e podemos olhar com distanciamento para a fissão de um universo neoplatônico) dissolve-se necessariamente, por força interna, numa galáxia potencialmente desordenada e ilimitada de elementos de conhecimento do mundo. Em consequência, torna-se uma enciclopédia e o faz porque de fato era uma enciclopédia que se ignorava ou um artifício idealizado para mascarar a inevitabilidade da enciclopédia. (ECO, 1984, p. 110)

Sendo assim, não há nenhuma garantia de que a árvore definicional seja finita. Seus primitivos, gêneros e espécies, devem ser interpretados como conjunções de diferenças. Nas palavras de Eco (1984, p. 111), “o dicionário é uma enciclopédia mascarada”. Isso não quer dizer que a árvore de hipônimos e hiperônimos deva ser

inteiramente abandonada, assim como a semântica componencial também não o foi completamente. Porém, há uma evidente inconsistência teórica na árvore definicional, tal como reconhecidamente existe na análise via primitivos e traços semânticos.

4.6 LINGUÍSTICA COGNITIVA: CONHECIMENTO ENCICLOPÉDICO E REALISMO CORPORIFICADO

A Linguística Cognitiva (doravante **LC**) é uma abordagem teórica que perspectiva a linguagem como uma faceta totalmente integrada à cognição humana que reflete a interação de fatores sociais, culturais, comunicacionais, funcionais e psicológicos. As unidades e as estruturas da linguagem são estudadas, pela ótica dessa teoria, não como se fossem entidades independentes, mas sim como manifestações de capacidades cognitivas gerais e que somente podem ser compreendidas no interior de um conjunto realista de aquisição, desenvolvimento cognitivo e processamento mental.

A LC fundamenta-se em três princípios básicos (GEERAERTS, 1995, p. 113): o da primazia da semântica na análise linguística e os da natureza enciclopédica e perspectivada do significado linguístico. A primazia da semântica decorre do próprio ponto de vista cognitivista adotado: se a categorização é a função basilar da linguagem, então a significação será o fato linguístico primário. Os dois outros postulados especificam o caráter do fenômeno semântico. Assim, se a linguagem serve para categorizar o mundo, então o significado linguístico não pode ser dissociado do conhecimento de mundo. Não se pode, dessa forma, postular a existência de um nível estrutural de significação separado do nível em que o conhecimento de mundo está coligado às formas linguísticas. Já não faz mais sentido, dentro dessa perspectiva, a dicotomia entre “conhecimento linguístico” e “conhecimento enciclopédico”. Ao invés de ser a efígie de um mundo projetado, a língua é a interpretação e a construção desse mundo. Ela o organiza na medida das necessidades, dos interesses e das experiências dos indivíduos e culturas. Nesses princípios assenta a essência da posição filosófica e epistemológica do movimento cognitivo: o **experencialismo** ou **realismo corporificado** (*embodied realism*).

Existem duas hipóteses básicas definidoras do experencialismo. Por um lado, a ideia de que o ser humano não tem acesso a estruturas objetivas independentes da realidade. Por outro, a hipótese de que o homem não é dotado de qualquer espécie de

razão transcendental e que, portanto, as operações mentais estão fortemente embasadas e derivam da interação sensório-motora do nosso corpo com o ambiente ao redor. No *The Oxford Handbook of Cognitive Linguistics*, Rohrer (2007) sintetiza a hipótese do experiencialismo corporificado dessa forma: “*the embodiment hypothesis is the claim that human physical, cognitive, and social embodiment ground our conceptual and linguistic systems.*”

4.7 PROPOSTA DE GRANDES CATEGORIAS PARA O DAD

Dentre os exemplos de sistemas conceituais e Grandes Categorias falados até aqui, faltou mencionarmos o modelo do *Historical Thesaurus of the Oxford English Dictionary* (doravante **HTOED**). Ele é organizado em três *major headings*: **I THE EXTERNAL WORLD**, **II THE MENTAL WORLD** e **III THE SOCIAL WORLD**³⁸. Esses, por sua vez, estão divididos em 354 subcategorias. No geral, o HTOED contém quase 800 mil significados, organizados em mais de 236 mil categorias e subcategorias. No exemplo abaixo, a numeração mostra que o conceito lexicalizado por *terms of endearment*, no quarto nível da hierarquia semântica, está abaixo do conceito LOVE, que pertence à categoria EMOTION que, por sua vez, alinha-se abaixo da Grande Categoria *THE MENTAL WORLD*. No entanto, em alguns casos, a árvore chega a descer até sete níveis.

02 The mental world
 02.02 Emotion
 02.02.22 Love
 02.02.22.04 Terms of endearment

No Quadro 4.4 separamos as principais categorias imediatamente abaixo das três *major headings* que organizam o topo da hierarquia do HTOED:

Quadro 4.4. *Major headings* do HTOED.

The external world	The mental world	The social world
1. The Earth	1. Soul, spirit, mind	1. Society/life in association with others
2. Life	2. Emotion/feeling	2. Inhabiting/dwelling
3. Physical sensibility	3. Judgement, opinion	3. Relations between social groups

³⁸ I Mundo Externo, II Mundo Mental e III Mundo Social

4. Matter	4. Aesthetics	4. Authority
5. Existence	5. Will/faculty of will	5. Law
6. Relative properties	6. Expectation	6. Education
7. The Supernatural	7. Having/possession	7. Religion
	8. Languages	8. Communications
		9. Travel/travelling
		10. Work / Serious occupation
		11. Leisure/The Arts

Dentre todos os sistemas conceituais citados nesta seção, o HTOED é o mais recente. Lançado em 2010, o modelo de sua estrutura veio endossar nossa tese para uma divisão tripartite baseada nas três grandes dimensões da experiência humana. O HTOED abandona a ideia de trabalhar com grandes categorias de predicáveis. Assim como a nossa proposta, sua classificação coloca o homem no centro da conceptualização do mundo, como intersecção das dimensões físico, psíquica e social (Figura 4.3).



Figura 4.3. Grandes dimensões da experiência humana.

Como vimos na Subseção 4.6, são estas também as três grandes dimensões da experiência humana na visão da Linguística Cognitiva. Por tudo isso, defendemos que uma ontologia encabeçada pelas Grandes Categorias MUNDO FÍSICO, MUNDO PSÍQUICO e MUNDO SOCIAL seria a mais apropriada para o nosso modelo conceitual. Acrescentemos ainda a esse modelo a intersecção entre cada duas dessas Grandes Categorias: PSÍQUICO-

FÍSICO, PSÍQUICO-SOCIAL e uma certa realidade FÍSICO-SOCIAL – tudo aquilo que é produto do trabalho humano e possui uma existência física: ferramentas, construções etc.

Esse modelo reflete estruturas mais profundas da língua e do sistema conceitual humano e, ainda, como temos dito, trás benefícios lexicográficos. Antes de passarmos a explorar as vantagens práticas do modelo proposto, vejamos como ele reflete princípios gerais presentes em outras teorias.

A semiótica peirciana diferencia três tipos de signos – os *índices*, os *ícones* e os *símbolos* (PEIRCE, [1932] 2010). O índice é um signo que indica, aponta, reenvia; o ícone é um sinal que representa, guardando similitudes com a coisa representada; e o símbolo é um sinal construído socialmente, por convencionalidade. Esses três tipos de signos estão apoiados em três diferentes princípios gerais que determinam a ligação entre uma forma e um significado. O índice depende do ambiente imediato, do *hic et nunc*. Nas palavras de Pierce, “um índice é um signo que se refere ao Objeto que denota em virtude de ser realmente afetado por esse Objeto” (p. 52). Há, portanto, uma relação de **contiguidade** entre o signo e a coisa representada, tal como, por exemplo, assevera a sabedoria popular: “onde há fumaça, há fogo” e “fumaça na serra, chuva na terra”. Há uma relação de causa e efeito – ou de parte-todo – absolutamente direta e limitada ao espaço-tempo presente. O índice é o signo do qual nos valem na maior parte de nossa comunicação mímica e gestual; por exemplo, quando apontamos a direção da rodoviária com o dedo indicador em riste. Não à toa o dedo *indicador* tem esse nome. Já os signos icônicos preservam uma relação de **similaridade** com a coisa representada. Segundo Pierce, “qualquer coisa, seja uma qualidade, um existente individual ou uma lei, é Ícone de qualquer coisa, na medida em que for semelhante a essa coisa e utilizado como seu signo”. O termo *ícone* é cotidianamente utilizado nos dias de hoje para descrever os pequenos desenhos de pastas, documentos etc. nas interfaces gráficas dos nossos sistemas operacionais. Esse é um bom exemplo, já que todo desenho figurativo é um signo icônico da coisa representada. Quando o pintor Magritte intitula o quadro da Figura 4.4 com a frase “Isto não é um cachimbo”, ele produz um estranhamento que explicita o fato de que ali não há realmente um cachimbo, apenas a representação pictórica desse objeto. E sua pintura só é capaz de evocar um cachimbo por guardar com ele similitudes de forma, cor, textura etc. Diferentemente da palavra *pipe* (*cachimbo*, em português), que evoca o objeto sem manter com ele qualquer similaridade, apenas por

convenção, já que tanto a palavra *pipe* quanto a palavra *cachimbo* são signos linguísticos e, portanto, signos simbólicos.



Figura 4.4. “Ceci n’est pas une pipe” (“Isto não é um cachimbo”), pintura de René Magritte

Os símbolos são signos altamente abstratos e, por isso, apenas os seres humanos são capazes de manipulá-los. Por ter a necessidade de comunicar a propósito de coisas abstratas – como acontecimentos passados e futuros, sentimentos, expectativas etc. – o homem é obrigado a utilizar sistemas de signos simbólicos, dos quais o mais elaborado é o da língua natural, seja em sua forma falada ou escrita. Conforme a definição dada por Pierce, “um Símbolo é um signo que se refere ao Objeto que denota em virtude de uma lei, normalmente uma associação de ideias gerais que opera no sentido de fazer com que o Símbolo seja interpretado como se referindo àquele Objeto”.

É possível perceber, dessa forma, que a tripla índice/ ícone/ símbolo é derivada, respectivamente, dos princípios gerais de contiguidade/ semelhança/ convenção, como ilustrado na Figura 4.5.

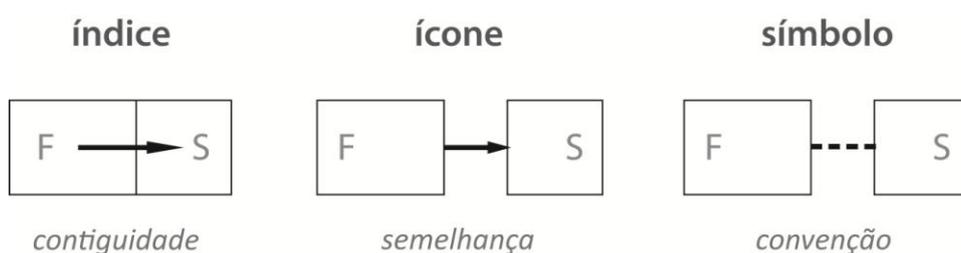


Figura 4.5. A relação entre forma e significado nos três tipos de signos piercianos (DELBECQUE, 2008, p. 21)

Se pensarmos pelo viés do sistema conceitual humano, podemos dizer que esse modelo de ontologia nos ajuda a enxergar que quando pensamos em conceitos do mundo físico ativamos outros conceitos relacionados metonimicamente, em termos de *frames* físicos, no sentido de “fotografias do mundo”. Conforme já citamos anteriormente, quando pensamos no conceito PÉ, todo um grupo de conceitos metonimicamente relacionados são ativados: CABEÇA, MÃO, PERNA etc. Por outro lado, se selecionarmos apenas seus traços [extremidade] + [inferior], e, propositadamente, ignorarmos 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, o conceito TEATRO enquanto ‘construção’ (que pertenceria, dentro do nosso modelo, à categoria que chamamos de FÍSICO-SOCIAL) é um todo constituído de partes menores, partes metonimicamente ativadas quando se pensa em uma delas: PLATÉIA, PALCO, COXIA etc. Ao passo que o conceito de TEATRO enquanto ‘arte’ e/ou ‘técnica’ ativa paradigmaticamente todo um conjunto de atividades que possuem esses traços: TEATRO, DANÇA, CINEMA, MÚSICA etc.

Como veremos ao longo da Seção 7, a vantagem lexicográfica desse modelo está no fato de que, com tal divisão no topo da hierarquia, é fácil perceber que um domínio léxico-conceitual prototipicamente inserido, por exemplo, na Grande Categoria MUNDO FÍSICO (como o domínio CORPO HUMANO) exhibe, entre os seus conceitos, um predomínio da relação sintagmática. Ou seja, todo domínio inserido na Grande Categoria MUNDO FÍSICO se constrói mediante princípios metonímicos. O mesmo se pode afirmar para os domínios inseridos nas Grandes Categorias MUNDO SOCIAL e FÍSICO-SOCIAL. No caso destes últimos, no entanto, a relação de contiguidade que vincula os seus conceitos não é universal (e, conseqüentemente, imutável), mas sim dependente de um determinado contexto sócio-histórico – ou seja, são relações convencionalizadas, nesse sentido. Para modelar um domínio do MUNDO SOCIAL, como o domínio CASAMENTO, por exemplo, é necessário vinculá-lo a um determinado tipo de modelo cultural estereotipado que lhe dá sustentação. Já nos domínios inseridos na Grande Categoria MUNDO PSÍQUICO prevalece, entre seus conceitos, a relação paradigmática, cf. veremos na Seção 7.1. No geral, os conceitos desta Grande Categoria são conceitos abstratos e, potencialmente, mais universais. No entanto, um domínio conceitual deve ser prototipicamente incluído em uma das Grandes Categorias, mas se estende quase sempre para fora dela, na medida em que nenhuma das três realidades se basta. Elas são

dimensões entrecruzadas. O conceito AMOR, por exemplo, dentro do domínio léxico-conceitual AMOR (MUNDO PSÍQUICO), evoca os domínios conceituais SEXO (PSÍQUICO-FÍSICO) e CASAMENTO (MUNDO SOCIAL).

4.8 EQUACIONAMENTO LINGUÍSTICO-COMPUTACIONAL: RIZOMA

Conforme Seção 3.4, a arquitetura léxico-conceitual do DAD se faz com base no formalismo das redes semânticas e, mais especificamente, por meio de uma rede semântica difusa em que CdSAs são organizados dentro de domínios conceituais específicos. A "Árvore de Porfírio" é uma rede onde os nós são os gêneros (e diferenças específicas) e os *links* representam a relação de hiperonímia ou hiponímia – conforme se esteja subindo ou descendo na árvore. As *wordnets* também são estruturadas em forma de árvore (FELLBAUM, 1998). Nelas, as ULs são organizadas sob a forma de *synsets* e cada *synset* equivale a um nó na rede. As relações léxico-conceituais de hiperonímia/hiponímia se dão entre os *synsets* e constituem os seus *links*. No entanto, tendo em vista o que foi discutido nesta seção, chegamos à mesma conclusão que muitos autores cognitivistas:

[...] é impossível representar o léxico como uma árvore taxonômica única dividida em ramificações cada vez mais numerosas. Ele é antes composto por uma multiplicidade de hierarquias que se cruzam e sobrepõem. (DELBECQUE, 2008, p. 73)

Portanto, nosso modelo abdica da estrutura em árvore e mantém o formalismo em rede seguindo a topologia de **rizoma**, proposta por Deleuze e Guattari (1976). Não raro, um conceito colocado dentro de uma categoria ou domínio conceitual por razão de suas propriedades pode entrar também em uma segunda ou terceira categoria por meio de outros atributos que adquirem relevância de acordo com contextos e necessidades outras. Nada mais apropriado para um dicionário de analogias que um sistema que abarque e estimule estas diferentes ligações em diferentes sentidos e contextos.

[...] cada ponto do rizoma pode ser unido e deve sê-lo com qualquer outro ponto e, com efeito, no rizoma não há pontos ou posições, mas apenas linhas de conexão; um rizoma pode ser quebrado num ponto qualquer e recomeçar seguindo a própria linha; é desmontável, invertível; uma rede de árvores que se abrem em toda direção pode produzir rizoma, o que equivale a dizer que em cada rizoma pode ser retalhada uma série indefinida de árvores parciais; o rizoma não tem centro. A idéia de uma enciclopédia como rizoma é consequência direta da inconsistência de uma árvore de Porfírio. (ECO, 1984, p. 116)

O modelo para o Dicionário Analógico Digital foi formalizado, portanto, em uma estrutura de rede descentralizada onde as relações léxico-conceituais sejam dinâmicas e adaptáveis, contendo apenas uma demarcação difusa entre as diferentes dimensões da experiência humana: mundos físico, psíquico e social. Para tanto, será incorporado ao modelo computacional princípios da lógica difusa (*fuzzy logic*), de forma a valorar o grau de relação entre os nós da rede, conforme Seção 5.5.

A primeira resposta do DAD para uma busca no Dicionário Criativo pelo item lexical *namorado* é semelhante ao que aparece na Figura 3.8 da Seção 3.4.1. Isto porque, estamos supondo, é nesse tipo de configuração que o item *namorado* aparecerá como mais relevante. Ou seja, como a UL mais relevante (valor 1) do CdSA que evoca o conceito HOMEM QUE AMA OU É AMADO, que preenche o EF central ‘amantes’ do *frame* *Relação_Amorosa* do domínio léxico-conceitual AMOR. No entanto, por força da UL *namorado* evocar uma entidade que é do sexo masculino, ela poderá também aparecer como resultado dentro do domínio MACHO que, conforme a interface sugerida, aparecerá ao lado do domínio AMOR, no cabeçalho de domínios, como um *link*. Bastará o usuário clicar no *link* para ter acesso à UL *namorado* dentro do novo contexto. Da mesma forma, por força da UL *namorado* evocar uma entidade que “é amada”, e tudo que é amado é preferido, ela poderá também aparecer listada dentro do domínio FAVORITO. FAVORITO será outro link no cabeçalho, ao lado dos domínios AMOR e MACHO. Perceba que não é a polissemia da UL *namorado* que nos permite organizá-la dentro de diferentes domínios, mas tão somente as facetas do seu significado. A repetição, em diferentes contextos, de uma mesma UL que evoca um mesmo conceito não é inútil nem trivial. A prova disso é que em cada domínio em que potencialmente a UL puder ser listada ela terá uma diferente relevância dentro de CdSAs diferentes, hora organizada dentro de um *frame* específico, hora não. Permitir que o usuário navegue por essa estrutura reticulada e sinuosa é condição fundamental para um dicionário de analogias. Não há apenas uma ou duas maneiras de ingressar e trafegar pela rede – ao contrário, há inúmeras. Como salientado por Deleuze & Guattari (1995, p. 24), “uma das características mais importantes do rizoma talvez seja a de ter sempre múltiplas entradas”.

Na WN.Pr, a UL *boyfriend* aparece no *synset* {*boyfriend, fellow, beau, swain*} que evoca o conceito glosado por “*a man who is the lover of a man or woman*”³⁹. Ela aparece apenas nesse *synset*, que é hipônimo direto dos *synsets* {*man, adult male*} e {*lover*}. E

³⁹ tradução livre: “um homem que é o amante de um homem ou uma mulher”.

{lover} é hipônimo direto de {person, individual, someone, somebody, mortal, soul}. Para a WN.Pr, o conceito BOYFRIEND (NAMORADO) é um tipo de MAN (HOMEM) ou um tipo de LOVER (AMANTE). E LOVER é um tipo de PERSON (PESSOA). Partindo do *synset* {boyfriend, fellow, beau, swain}, e por meio dos recursos oferecidos pela interface da **WordNet Search 3.1**⁴⁰, só é possível subir na árvore de conceitos (seguindo os *links direct hypernym* ou *inherited hypernym*) até os conceitos PHYSICAL ENTITY (ENTIDADE FÍSICA) ou simplesmente ENTITY (ENTIDADE) – o topo da árvore conceitual. Ou abrir uma categoria de conceitos em relação paradigmática (seguindo o *link sister term*). Ainda que haja um esforço de se incluir outros arcos de relação semântica, prevalece, como dissemos anteriormente, a relação lógico-conceitual “é um tipo de”. Ou seja, o tipo de relação que dá forma a uma estrutura arbórea. Nesse tipo de arquitetura léxico-conceitual, estarão sempre distantes (em número de *links*, por exemplo) os conceitos de AMOR e NAMORADA; e ainda mais distantes os conceitos de BEIJO, CARINHO e SEXO. Já em uma estrutura rizomática, o tecido de vinculação de ULs e conceitos extrapola a relação “é um tipo de” e pode instanciar relações inusitadas como “se relaciona com” ou “se parece com”, sem necessariamente explicitá-las.

Um rizoma não começa nem conclui, ele se encontra sempre no meio, entre as coisas, inter-ser, intermezzo. A árvore é filiação, mas o rizoma é aliança, unicamente aliança. A árvore impõe o verbo “ser”, mas o rizoma tem como tecido a conjunção “e... e... e...” Há nesta conjunção força suficiente para sacudir e desenraizar o verbo ser. (DELEUZE, GUATTARI, 1995, capa)

Apenas complementar uma estrutura léxico-conceitual fortemente paradigmática (como as das *wordnets*) com relações sintagmáticas (como as das *framenets*) permitiria, na metáfora dos aeroportos e estradas (cf. Seção 3.4), saltar de um aeroporto internacional (grandes *hubs*), viajar por estradas e vicinais (*clusters*) até um aeroporto regional (pequenos *hubs*), voar até um segundo aeroporto deste tipo e dele prosseguir por estradas, vicinais, avenidas e ruelas até o destino desejado. No entanto, ensejar uma topologia rizomática na rede de relações léxico-conceituais permitirá trilhar caminhos, sendas e atalhos improváveis, análogos somente, para ficar na metáfora espaço-temporal, aos **buracos de minhoca** da física moderna. Conforme o modelo proposto evolua de um protótipo limitado a apenas três domínios léxico-conceituais em direção a uma base de dados mais ampla, novas conexões analógicas, antes impensáveis, serão

⁴⁰ <http://wordnetweb.princeton.edu/perl/webwn>, acessado em 04 de janeiro de 2013.

incorporadas à rede, estimulando o pensamento e a escrita criativa, o que satisfaz os objetivos inicialmente aventados.

Em uma arquitetura difusa e rizomática, os próprios usuários do DAD poderão, futuramente, incluir ou fazer surgir novas relações analógicas. Seja por ações voluntárias ou involuntárias do usuário, a rede poderá ser refinada ao longo do tempo pelo seu próprio uso, mantendo sua estrutura elástica e flexível. Em trabalhos futuros, poderão ser criados, por exemplo, i) um sistema de monitoramento dos cliques em *links* que ajudará a criar trilhas ou mesmo clareiras na topologia da rede; e ii) ferramentas de inclusão ou exclusão de nós e arcos respaldados na inteligência coletiva de milhares de usuários. Não se trata, portanto, de um sistema simplesmente flexível e resiliente, mas verdadeiramente adaptável. No entanto, para manter a consistência da rede, ela deve preservar essa estrutura composta de CdSAs, organizados em *frames* ou diretamente dentro de um *anaset* específico, protipicamente arranjados no interior de alguma das três Grandes Categorias: MUNDO FÍSICO, MUNDO PSÍQUICO e MUNDO SOCIAL.

4.9 SÍNTESE DA SEÇÃO

As discussões realizadas nesta Seção tiveram por objetivo propor uma estrutura hierárquica para o DAD distinta daquela consagrada no *Roget's Thesaurus* e replicada no *Dicionário Analógico da Língua Portuguesa* do professor Azevedo. As investigações levadas a cabo nessa seção, em que comparamos os diferentes sistemas conceituais historicamente propostos para a organização do léxico, ratificam um alinhamento com a proposta do HTOED. Nele, todos os conceitos se organizam dentro de apenas três grandes categorias conceituais: THE EXTERNAL WORLD, THE MENTAL WORLD, THE SOCIAL WORLD. Em consonância com os preceitos da LC, como vimos, foi defendida ao longo desta Seção uma estrutura hierárquica encabeçada pelas Grandes Categorias MUNDO FÍSICO, MUNDO PSÍQUICO e MUNDO SOCIAL. Além, ainda, da intersecção entre cada duas dessas Grandes Categorias: PSÍQUICO-FÍSICO, PSÍQUICO-SOCIAL e FÍSICO-SOCIAL. Essa seção, no entanto, não propõe níveis intermediários entre os domínios léxico-conceituais e o topo da hierarquia. Para esta tese, estaremos trabalhando com apenas três níveis conceituais. Uma UL pertence a uma categoria conceitual (formalizada em um CdSA), que pertence a

um domínio léxico-conceitual (formalizado em um *anaset*) que, por sua vez, integra uma das Grandes Categorias propostas nesta seção.

Como vimos, um *anaset*, com seus CdSAs, não se limitará a elencar ULs relacionadas a uma determinada Grande Categoria. No entanto, ele deve ser prototipicamente incluído em uma delas. Como dissemos, as Grandes Categorias são, na verdade, dimensões entrecruzadas. Partindo, por exemplo, do conceito AMOR e estendendo para a esquerda o conceito mais prototipicamente FÍSICO e para a direita o mais prototipicamente SOCIAL, temos:

FOGO ←————— AMOR —————→ CASAMENTO

A relação entre o conceito de amor e de fogo está eternizada no talvez mais famoso verso da tradição poética da língua portuguesa: “amor é fogo que arde sem se ver”⁴¹. Por muito tempo essa relação foi considerada uma metáfora poética, literária, no nível puramente linguístico. Com a Teoria da Metáfora Conceitual (LAKOFF & JOHNSON, 1980), essa relação metafórica passou a ter seu valor cognitivo reconhecido, mudando seu entendimento como simples figura retórica para o de operação cognitiva fundamental. De maneira simplificada, entende-se que a metáfora propicia o entendimento de um domínio abstrato em termos de um domínio mais concreto. Ela produz o mapeamento de correspondências sistemáticas entre um domínio fonte e um domínio alvo. Sendo o primeiro predominantemente concreto e o segundo predominantemente abstrato. As metáforas conceituais nos fazem “conceber eventos, atividades, emoções, ideias, etc. como entidades e substâncias.” (LAKOFF, 2002, p. 76).

Apesar da metáfora conceitual O AMOR É UMA VIAGEM ter ficado famosa – é uma das metáforas mais estudadas na literatura lakoffiana –, a metáfora O AMOR É FOGO (KÖVECSES, 2003) é encontrada com maior frequência em *corpus* (PIRES, 2008), com atestações como *seu coração estava em fogo, a febre do amor, o mais ardente calor do nosso amor* etc. Já o casamento é prototipicamente visto exatamente como a efetivação social, legal e/ou religiosa, da união entre dois indivíduos que se amam. A evidência de

adaptável, preservando apenas uma demarcação difusa entre as diferentes dimensões da experiência humana.

SEÇÃO 5 – EFEITOS DE SALIÊNCIA COMO CRITÉRIOS DE RELEVÂNCIA PARA O DAD

Em 2.5 foi dito que os dicionários de sinônimos agrupam ULs que lexicalizam apenas um conceito, ao passo que os grupos de palavras dos dicionários analógicos agrupam diferentes conceitos lexicalizados em seu interior. Quase sem qualquer estruturação semântica interna, esses grupos possuem, muitas vezes, dezenas ou centenas de palavras. Tal fato dificulta a consulta do usuário. Já a desvantagem dos dicionários de sinônimos é que conceitos fortemente relacionados não estão mutuamente acessíveis, na medida em que esse tipo de dicionário é organizado em ordem alfabética, e não por domínios conceituais. Sugerimos, portanto, mesclar o mais vantajoso da estrutura de cada tipo de obra, usando os conjuntos de sinônimos para organizar os grupos de analógicos. Ainda assim, pelo volume de ULs envolvidas em um mesmo domínio léxico-conceitual, outras estratégias de organização da informação léxico-conceitual parecem necessárias. Nossa proposta, nesta seção, é elaborar critérios de relevância no interior dos grupos de palavras baseados nos princípios de **saliência semântica**, ou seja, no papel preponderante que os sentidos **prototípicos** (ROSCH, 1975) representam dentro de um domínio léxico-conceitual.

Para tanto, em 5.1 vamos discorrer sobre princípios e teorias da categorização humana até chegar à concepção cognitivista de como o homem apreende e organiza o mundo a sua volta. A Linguística Cognitiva postula que as categorias léxico-conceituais não são estruturas homogêneas. Ao contrário, exibem uma estrutura prototípica, em que alguns dos seus elementos são mais representativos da categoria do que outros. A discussão sobre saliência psicológica e categorias difusas será tratada na Subseção 5.2. Em 5.3 vamos nos debruçar sobre os procedimentos metodológicos que podem ser aplicados para determinar a estrutura prototípica de uma categoria léxico-conceitual. Por fim, na Seção 5.4, vamos tratar do equacionamento linguístico-computacional das categorias léxico-conceituais que exibem uma estrutura prototípica por meio da modelagem da Teoria dos Conjuntos Difusos.

5.1 CATEGORIAS CONCEITUAIS E CATEGORIAS LINGUÍSTICAS

Podemos definir a noção de **conceito** como “a ideia que temos de qualquer coisa, da sua forma de existir no mundo” (DELBECQUE, 2006). Ele pode estar relacionado com uma entidade individual ou um grupo de entidades. Todo conceito que, dessa forma, projeta a realidade de acordo com a nossa experiência constitui uma **categoria conceitual**. Não é possível perceber um objeto, um evento, uma ação ou um sentimento sem o impulso involuntário de categorizá-los de imediato dentro de uma categoria conceitual.

As categorias conceituais e os conceitos possuem, dentro de uma comunidade linguística, uma correspondência em categorias e signos linguísticos. Essa correspondência não é exata, dado que a língua contém apenas parte dos conceitos que o homem é capaz de manipular. As categorias linguísticas procuram lexicalizar as categorias conceituais de maneira que os seres humanos possam compartilhar suas ideias e experiências entre si. O objetivo da análise onomasiológica é evidenciar a estrutura subjacente a um conjunto de ULs que evocam conceitos semanticamente próximos ou relacionados. Ou seja, organizar um conjunto de ULs que designam entidades pertencentes a um mesmo domínio conceitual. As ULs *amor*, *carinho*, *zelo* e *namorado*, por exemplo, remetem todas para o domínio conceitual AMOR. Elas não evocam todas o mesmo conceito ou categoria conceitual, mas conceitos distintos que pertencem a um mesmo domínio da experiência humana. Chamamos *namorado* o indivíduo que devota seu *amor* a alguém e que esse amor normalmente é expresso por meio de *carinho* e demonstração de *zelo* para com o objeto amado. Um **domínio conceitual** pode ser definido, portanto, como um campo coerente do nosso universo constituído de concepções e experiências. No decorrer deste trabalho utilizamos o termo **domínio léxico-conceitual** para expressar a ideia de domínios conceituais estruturados por meio de categorias conceituais lexicalizadas. Um dicionário de teatro, por exemplo, deve listar entre os seus verbetes as ULs que designam entidades (pessoas, objetos, ações, acontecimentos, estados de coisas, qualidades, propriedades etc.) pertencentes ao domínio léxico-conceitual TEATRO.

Em todos os domínios, principalmente nos domínios não especializados, a correspondência entre ULs e categorias conceituais é disforme. Às vezes exibindo o que é chamado de **buraco lexical** (ausência de uma UL específica para uma categoria

conceitual saliente) e às vezes apresentando superposições. Pense na sombra de um objeto iluminado por diferentes refletores; a imagem é de uma sobreposição de formas aproximadas do objeto (cada qual com uma deformação angular dependente da posição de sua fonte luminosa). Da mesma forma, uma mesma UL pode evocar diferentes conceitos relacionados – o que é chamado **polissemia** – e um mesmo conceito ou categoria conceitual pode ser lexicalizada por mais de uma UL.

Coloca-se então a questão, dentro da análise onomasiológica, de saber o estatuto que uma determinada UL possui em relação a outras ULs para evocar uma mesma entidade ou conceito. É o problema de identificar entre ULs sinônimas aquelas que, dentro de um dado contexto, melhor designam o conceito em questão. Em seguida, coloca-se a questão de saber a relevância de uma determinada entidade ou conceito dentro de uma categoria conceitual. E por fim, dado um domínio da experiência (ou do conhecimento) específico, saber quais são as suas categorias conceituais salientes; ou seja, que melhor representam ou evocam o domínio como um todo.

O **modelo clássico de categorização** (ANGIONI, 2006), tal como formulado por Aristóteles (SANTOS, 1995) e perpetuado ao longo de séculos pela filosofia ocidental, é baseado numa estrutura de atributos necessários e suficientes. Ou seja, um ente pertence a uma categoria determinada se, e somente se, exhibe todos os traços característicos que a definem e, portanto, a ausência de qualquer um desses traços significa automaticamente a sua exclusão da categoria em questão. Como consequência dessa abordagem, as entidades possuem ou não possuem um traço e, portanto, não é possível haver casos ambíguos. Dessa forma, as categorias possuem limites claramente definidos e não existem membros centrais ou marginais, todos possuem o mesmo status.

A **semântica estruturalista** (SAUSSURE, 1916; TRIER, 1934) não incorporou todos os pressupostos clássicos de origem aristotélica, mas manteve-se fiel a maioria deles. Como se sabe, uma das ideias principais do estruturalismo é a concepção da linguagem como um sistema autônomo, em que os traços do significado de uma entidade linguística estão relacionados entre si, sistematicamente (LYONS, 1977).

Pode-se dizer que o mesmo é válido para a versão da **análise componencial**, que se ampara fortemente nessa perspectiva binária de análise semântica. Essa modalidade de análise é essencial ao método tradicional de definição por gênero próximo e diferença específica.

Esse método de descrição do significado das palavras e dos sintagmas baseia-se na tese segundo a qual o sentido de cada lexema pode ser analisado em termos de um conjunto de componentes de sentido mais gerais (ou traços semânticos), algumas das quais, ou todas, serão comuns a diferentes lexemas do vocabulário. Na medida em que a análise componencial está associada ao conceptualismo, as componentes de sentido (...) podem ser concebidas como conceitos atômicos, e os sentidos dos lexemas particulares como conceitos moleculares. (LYONS, 1977, p. 255)

Dessa forma, mais uma vez, a categorização é concebida em termos de traços essenciais, sob a presunção de que espelham as distinções do mundo no qual o sujeito se desenvolve. No entanto, a realidade provou-se demasiado complexa para caber nesse método de análise do significado (cf. 4.5). Segundo Taylor (1989), as categorias são inumeráveis e expansíveis e, na mesma medida, o seriam os traços semânticos ou as componentes de sentido que alimentam esse tipo de análise. Lyons (1977) reforça essa desconfiança, não só com relação à pretensa validade universal de certas componentes de sentido ou com o grau de validade cognitiva do método, mas também com relação à fragilidade, incompletude e vagueza desse modelo de análise semântica.

A realidade psicológica das componentes de sentido tem sido frequentemente posta em causa, o mesmo acontecendo com a sua universalidade. Contudo, o que nas discussões de carácter genérico sobre os méritos da análise componencial não é frequente mencionar, é o fato de, mesmo nas áreas em que parece relativamente convincente, deixar por explicar pelo menos tanto quanto consegue explicar. (LYONS, 1977, p. 268)

A verdade é que a análise componencial não é um método arquitetado por cima de uma teoria semântica de interesse cognitivo. Como assinala Lakoff (1987), os pressupostos que sustentam grande parte do método componencial foram herdados de teorias (estruturalistas) acerca de como são as coisas no mundo e não de teorias cognitivas, que procuram abordar os fenômenos semânticos se perguntando como a mente faz com que o mundo tenha sentido.

Em oposição à visão estruturalista do significado, a **Semântica Cognitiva** (doravante **SC**) defende que os seres humanos compreendem o mundo por meio de diferentes modelos de categorização, mesmo que algumas entidades possam ser compreendidas por meio de modelos idealizados (LAKOFF, 1987). Um dos objetivos principais da SC é delinear uma abordagem geral para a teoria da categorização humana. Segundo Lakoff,

a maioria de nossas palavras e conceitos designam categorias [...] Categorização não é um processo que deve ser estudado superficialmente. Não há nada mais básico do que a categorização para o nosso pensamento, percepção, ação, e discurso. Cada vez que nós vemos algo como “um tipo” de coisa, por exemplo, uma árvore, nós estamos categorizando. [...] A compreensão de como categorizamos é o ponto central para a

compreensão de como nós pensamos, funcionamos e, conseqüentemente, um ponto central para a compreensão daquilo que nos faz humanos. (p. 5)

A SC é uma subárea da LC. Conforme vimos na Seção 4.6, a LC é uma abordagem teórica que perspectiva a linguagem como uma faceta totalmente integrada aos fatores sociais, culturais, comunicacionais, funcionais e psicológicos. Dessa forma, a linguagem é estudada como manifestação de capacidades cognitivas gerais e não como uma competência autônoma.

Given this perspective, the analysis of the conceptual and experiential basis of linguistic categories is of primary importance within Cognitive Linguistics: the formal structures of language are studied not as if they were autonomous, but as reflections of general conceptual organization, categorization principles, processing mechanisms, and experiential and environmental influences (GEERAERTS & CUYCKENS, 2007, p. 34).

Talmy (2000) sustenta que a LC examina propriedades formais da língua a partir de uma perspectiva conceitual, preocupada com “os padrões nos quais os processos pelos quais o conteúdo conceitual está organizado na linguagem” (p. 2); e tem como meta relacionar suas descobertas com as abordagens psicológicas sobre as estruturas cognitivas, incluindo a análise da memória semântica, a associatividade de conceitos, a estrutura de categorias, a geração de inferências e o conhecimento contextual.

São objetos da SC as características estruturais da categorização linguística, tais como prototipicidade, polissemia, modelos cognitivos, *frames*, metáfora, metonímia, categorias radiais, imagens mentais etc. São também do interesse dela os princípios funcionais da organização linguística, a interface conceitual entre sintaxe e semântica, a base pragmática e ligada à experiência da linguagem-no-uso e, evidentemente, os debates em torno da relação entre linguagem e pensamento.

A introspecção é tida como metodologia científica legítima e de crucial importância pelos linguistas cognitivistas na abordagem dos temas acima colocados. Tal como propõe Talmy, deve-se incluir procedimentos tais como “a manipulação de material linguístico cujos significados possam ser acessados” (p. 5), considerando, no entanto, que “os achados resultantes a partir da introspecção devem ser correlacionados com aqueles resultantes de outras metodologias”, tais como: i) análise de relatos originários da introspecção de outros sujeitos; ii) análise de discursos e corpora; iii) análise translinguística e diacrônica; iv) avaliação do contexto e do arcabouço cultural; v) técnicas experimentais da psicolinguística; vi) estudos provenientes da neuropsicolinguística; e vii) exames instrumentais da neurociência.

É dentro dessa perspectiva teórica que estamos trabalhando para resolver aspectos relevantes de como o homem categoriza os conceitos e a linguagem. Na Seção seguinte vamos falar sobre a estrutura prototípica das categorias linguísticas, que tem fundamentos na filosofia de Wittgenstein (1953) e na Psicologia Cognitiva de Rosch (1975).

5.2 SALIÊNCIA PSICOLÓGICA: O CARÁTER DIFUSO DAS CATEGORIAS CONCEITUAIS

Nesta Subseção, dividida em quatro partes, serão explicadas as noções de *semelhança de família* (5.2.1), *Teoria dos Protótipos* (5.2.2), *Redes Radiais* (5.2.3) e categorias de *nível de base* (5.2.4). Esses conceitos possuem destaque dentro da SC e serão indispensáveis para a nossa proposta de modelagem do DAD.

5.2.1 WITTGENSTEIN E A SEMELHANÇA DE FAMÍLIA

Em seu livro *Investigações Filosóficas* (1953), Ludwig Wittgenstein refere-se à extensão do significado da categoria JOGO e conclui que seus diferentes membros não possuem nenhum traço que seja comum a todos os demais membros da categoria. Não há sequer um traço que a multiplicidade de atividades designadas pela palavra *jogo* compartilhem. Em lugar de traços comuns que garantam uma definição analítica para a categoria como um todo, Wittgenstein defende que o que existe é uma rede entrecruzada de similaridades, o que ele denominou como **Semelhanças de Família** (*Familienähnlichkeit*) (WITTGENSTEIN, [1953] 1975). Wittgenstein argumenta que o que confere unidade ao conceito não é um “fio único” que transpassa todos os casos, mas uma sobreposição de diferentes fibras, como em uma corda. Pode-se ilustrar essa ideia como na Figura 5.1, em que as letras equivalem aos membros da categoria JOGO (*futebol, amarelinha, xadrez*) e os números correspondem aos atributos da categoria (*em grupo, competitivo, divertido*).

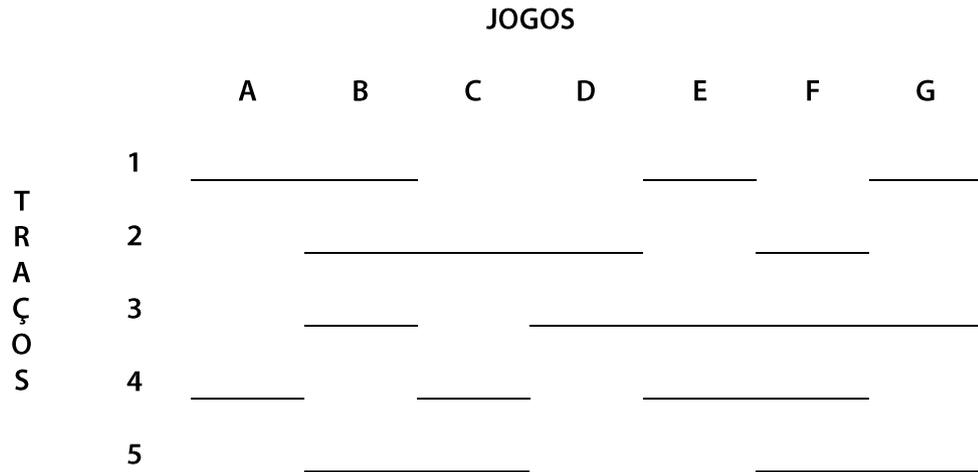


Figura 5.1. Semelhanças de família da categoria JOGO (GLOCK, 1996).

Se a estrutura das categorias consiste em um conjunto de semelhanças de família e não em traços mínimos essenciais compartilhados e se, como muitas vezes ocorre, as categorias tendem a se fundir em outras, o conhecimento de uma categoria não implica que possamos estabelecer, com certeza, que membros pertençam ou não a ela. Dessa forma, os limites das categorias são difusos (*fuzzy*). Wittgenstein sugere ainda que os conceitos determinados por semelhanças de família se desenvolvem em torno de um ou mais “centros de variação”, ou seja, membros exemplares – como o ‘futebol’ no caso da categoria JOGO.

As ideias e a certeza de Wittgenstein de que a teoria clássica do significado (e da categorização) é inadequada para prever a classe denotacional das palavras foram confirmadas empiricamente nos estudos que se originaram na Psicologia Cognitiva, o que favoreceu o surgimento de uma perspectiva mais versátil para o estudo da estrutura das categorias e que poderia cobrir os casos para os quais a semântica clássica, de fundamento aristotélico, resultava inadequada.

5.2.2 ROSCH E A TEORIA DOS PROTÓTIPOS

As considerações tecidas anteriormente, a respeito do caráter difuso e da estrutura em termos de atributos e semelhanças de família das categorias lingüísticas, têm validade dentro do enfoque cognitivista. Em contraste com o modelo clássico, a **Teoria dos Protótipos** (doravante **TP**) (ROSCH, 1975) postula que as categorias não

são estruturas homogêneas. De acordo com evidência experimental (LABOV, 1973; ROSCH, 1973, 1975; KEMPTON, 1981; TAYLOR, 1989), as categorias exibem melhor uma estrutura prototípica, ou seja, há bons e maus exemplos. As **categorias difusas** (*fuzzy categories*) 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 passou a ser chamado **efeitos de protótipo** (*prototype effects*) (LAKOFF, 1987; TAYLOR, 1989).

Dessa forma, uma categoria léxico-conceitual é constituída por membros prototípicos, por membros menos prototípicos e por membros marginais, de maneira gradual. O membro mais representativo, ou seja, aquele que os falantes primeiro evocam ao tomar contato com o nome de uma categoria é chamado o **protótipo**, ou membro **prototípico** da categoria, ao passo que os membros menos representativos ocupam a margem da categoria e são chamados de membros **periféricos** ou **marginais**. De maneira geral, o centro de uma categoria léxico-conceitual se mostra claramente definido, enquanto que os limites dela são difusos, com elementos que tendem a invadir outras categorias léxico-conceituais.

Para se distinguir do enfoque clássico da categorização e do significado, a semântica dos protótipos reempregou a noção de traço ou componente pela de **atributo**. Enquanto os traços se caracterizam por serem binários e, em consequência, por ter o mesmo estatuto analítico, os atributos têm efeito. A existência de membros mais representativos implica que existem atributos mais centrais (prototípicos) que outros.

Uma investigação pioneira, que serviu de argumento, não só para a noção de protótipo, mas também para a LC em geral, é o estudo dos termos que denotam cores básicas (BERLIN, KAY, 1969). Os resultados contradizem a hipótese estruturalista da arbitrariedade das categorias linguísticas, assim como a concepção de sua organização em traços essenciais. O estudo de Berlin e Kay, contrariamente à visão estruturalista, apontou que a divisão e organização do *continuum* da cor em categorias não se constitui em termos de unidades discretas, mas em torno de entidades focais. Cada categoria de cor tem uma cor focal, um exemplar central primário, de cuja generalização depende a classe de denotação completa da categoria e cuja existência está determinada por fatores biológicos, cognitivos e ambientais.

A TP introduz, assim, uma metodologia alternativa de análise e apresentação da estrutura do significado. Uma descrição categorial deve considerar, como fonte dos atributos a incluir, tanto os bons e os maus exemplos quanto os membros marginais (ou seja, aqueles de pertinência duvidosa). Ao oferecer uma categoria difusa e variável em sua distância em relação ao protótipo central (não equidistante, como na teoria clássica) são introduzidos novos fatores de organização. O protótipo corresponderá, assim, a um “modelo mental de base”, por meio do qual configuramos determinada parte da realidade que sistematizamos em uma dada UL.

Os estudos prototípicos já demonstraram seu valor teórico, tendo em vista que, cada vez mais, se têm encontrado evidências empíricas de que o sistema conceitual humano é prototipicamente organizado.

Podemos resumir as formulações da TP como abaixo:

- a) categorias não representam divisões arbitrárias dos fenômenos do mundo, mas devem ser compreendidas na relação com outras capacidades cognitivas da mente humana como, p. ex., a memória;
- b) nem todas as categorias são prototipicamente organizadas; o conceito de prototipicidade pode ser aplicado ao próprio conjunto de categorias a serem analisadas, sendo algumas mais prototipicamente prototípicas do que outras; as categorias cognitivas de cores, formas, assim como de organismos e objetos concretos são bons exemplos de categorias prototipicamente estruturadas (LABOV *apud* UNGERER E SCHMID, 1996), assim como categorias para conceitos psicológicos (WITTGENSTEIN, 1953);
- c) as fronteiras das categorias cognitivas são difusas e, portanto, categorias vizinhas compartilham membros marginais umas com as outras; entre os membros prototípicos e os membros marginais, as categorias cognitivas possuem membros que podem ser avaliados numa escala de tipicidade;
- d) membros centrais de categorias cognitivas têm o maior número de atributos em comum com outros membros da categoria e um número menor de atributos é compartilhado com membros de categorias vizinhas; o que equivale a dizer que membros de categorias marginais compartilham somente um pequeno número de atributos com outros membros de suas categorias e possuem atributos que pertencem a outras categorias.

A TP é frequentemente citada nos estudos de categorias de objetos como, por exemplo, as categorias FRUTA, PÁSSARO, VEÍCULO, ROUPA, MÓVEL ou INSTRUMENTOS MUSICAIS. Essas categorias, de objetos físicos e do cotidiano, são prototipicamente prototípicas. No entanto, categorias abstratas – em que as entidades são atividades, qualidades, estados de coisas ou conceitos psicológicos – também são estruturadas dessa maneira. O amor romântico, por exemplo, é prototípico da categoria AMOR (KÖVECSES, 1988).

5.2.3 REDES RADIAIS

A noção de **redes radiais** (ou **categorias radiais**) foi proposta por Lakoff (1987) ao discutir o sistema de classificadores do Dyrbal, uma língua aborígine da Austrália. O Dyrbal possui um sistema de classificadores que marca a categoria a que os substantivos pertencem. As palavras *bayi* (machos humanos, animais), *balan* (fêmeas humanas, água, fogo, combate), *balam* (alimentação carnívora) e *bala* (tudo o que não está nas outras classes) são utilizadas na frente dos substantivos e servem para categorizar todos os objetos do universo Dyrbal. A segunda categoria (*balan*) é a responsável pelo nome do livro: *Woman, fire and dangerous things* (Figura 5.2). Segundo Lakoff, cada um desses classificadores (com exceção de *bala*) se organizam de maneira radial. As redes radiais operam com princípios de extensão de significado a partir do conceito mais central ou prototípico da categoria, incluindo “modelos metafóricos, modelos metonímicos, relações de esquema de imagens etc.” (p. 204).

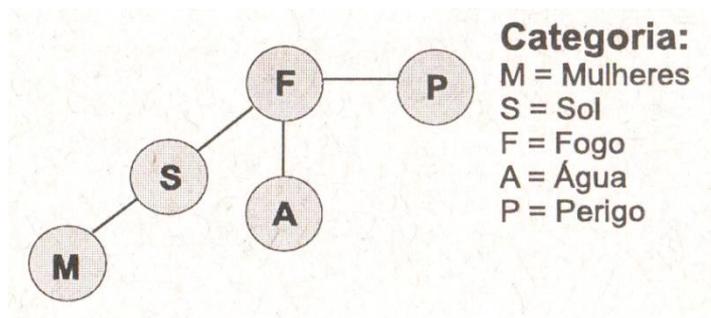


Figura 5.2. Radialidade da categoria *Balan*: mulheres, fogo e coisas perigosas – adaptado de FELTES (2007)

Dentre todos os significados possíveis para um item lexical, há sempre um (ou mais de um) que identificamos como sendo o mais central ou prototípico, cf. 5.2.2. Esse significado ajuda a explicar os demais. O significado prototípico de *amor* pode ser definido pela glosa “afeição muito intensa de uma pessoa por outra”. A partir desse significado, associado ao conceito AMOR ROMÂNTICO, podemos chegar aos outros significados de *amor*. Isso porque, mesmo distintos por um grau de proeminência, os significados de um item lexical polissêmico continuam interligados por relações sistemáticas.

Segundo Feltes (2007), o AMOR ROMÂNTICO motiva o AMOR FAMILIAR pelo fato de que, prototipicamente, as pessoas que se amam se casam e constituem família. A

FORTE AMIZADE, assim como o conceito AMOR FAMILIAR, pode ser definida em termos de uma “forte afeição sem envolver desejo sexual”. Expressões como “ele é como um irmão pra mim” e “ele é praticamente da família” são empregues com frequência para expressar uma forte relação de amizade. Ao mesmo tempo, o conceito AMOR ROMÂNTICO evoca, por extensão metonímica, o conceito de RELAÇÃO AMOROSA. Há uma relação de contiguidade entre os conceitos, pois onde há um AMOR ROMÂNTICO, provavelmente há uma RELAÇÃO AMOROSA, a menos que seja um AMOR PLATÔNICO. Por conseguinte, uma RELAÇÃO AMOROSA possui como elementos centrais a PESSOA QUE AMA e a PESSOA AMADA que, prototipicamente, é uma relação recíproca, ou seja, a PESSOA QUE AMA é uma PESSOA AMADA e vice-versa.

Os processos cognitivos de metonímia, metáfora, generalização e especialização caracterizam as ligações entre os significados tanto não centrais quanto com o centro da rede radial de conceitos e podem produzir extensões semânticas quase infinitas em diferentes direções. Na verdade, só seremos capazes de identificar o final da extensão de uma rede como esta quando ela já estiver invadindo outro domínio conceitual, conforme delimitações difusas que por necessidades teóricas ou práticas sejamos obrigados a distinguir. Ao mesmo tempo, por outro lado, a contribuição da noção de radialidade se dá pela alternativa que ela coloca diante da dificuldade de se encontrarem propriedades essenciais e comuns a toda uma gama de significações flutuantes em torno de um conceito. Dentro do nosso modelo, a noção de categorias radiais será utilizada para ordenar as categorias conceituais dentro de cada *anaset*.

5.2.4 PROEMINÊNCIA CONCEITUAL E NÍVEL DE BASE

Rosch e Mervis (1975) também fizeram importantes investigações sobre as categorias de nível básico ou **nível de base**. Para elas esse é “um nível de abstração em que objetos concretos do mundo se dividem mais naturalmente em categorias” (p. 586). Os domínios da experiência são hierarquizados (cf. Seção 4), ou seja, vão do nível mais *genérico* ao nível mais *específico*, apresentando pelo menos três níveis diferentes. Rosch observou que o nível psicologicamente mais básico está próximo ao meio da taxinomia hierárquica.

Entre os diferentes conceitos que se encontram no nível de base, destaca-se o conceito normalmente designado como **termo médio**. É o termo mais utilizado, aquele que a criança aprende primeiro. Na aquisição da língua, os nomes *árvore*, *cachorro*, *calças* e *maçã* (em destaque no Quadro 5.1) precedem claramente nomes mais genéricos como *planta*, *animal*, *roupa* e *fruta*. E só posteriormente surgem nomes mais específicos como *pinheiro*, *labrador*, *jeans* e *gala*.

Quadro 5.1. Domínios conceituais e nível de base

Níveis	Domínios conceituais			
Nível Genérico	<i>planta</i>	<i>animal</i>	<i>roupa</i>	<i>fruta</i>
Nível de base	<i>árvore</i>	<i>cachorro</i>	<i>calças</i>	<i>maçã</i>
Nível específico	<i>pinheiro</i>	<i>labrador</i>	<i>jeans</i>	<i>gala</i>

É no nível básico de categorização que as pessoas conceitualizam coisas como *gestalten* perceptuais e funcionais. Segundo Rosch *et al.* (1976), as categorias de nível básico (como *cadeira*) são as primeiras com as quais podemos interagir corporeamente (ao sentar, dobramos as pernas de um modo específico). Em um nível superordenado (como *mobília*) não é possível descrever uma interação específica. Já no nível subordinado, as interações com o objeto (como *cadeira de barbeiro*) são praticamente iguais às que ocorrem com os objetos de nível básico, acrescido apenas de uma experiência contextual particular (*cadeira de barbeiro*, *cadeira de escritório*, *cadeira de cozinha*, *cadeira de praia*).

5.3 DETERMINANDO SENTIDOS E REFERENTES PROTOTÍPICOS: ASPECTOS METODOLÓGICOS

Vimos que uma categoria léxico-conceitual é constituída por membros centrais ou prototípicos e membros menos centrais (cf. 5.2.2) e vimos também que uma palavra com muitos sentidos (polissemia) também possui sentidos mais prototípicos do que outros (cf. 5.2.3). A questão que se coloca agora, nesta seção, é como determinar quais membros de uma categoria ou quais sentidos de uma palavra são centrais e quais deles são periféricos.

Existem três maneiras em uso corrente nos estudos do gênero que se relacionam entre si (TAYLOR, 1992; TALMY, 2000; DELBECQUE, 2006). Podemos produzir questionários e testes e aplicá-los aos falantes de uma comunidade linguística de

maneira que os seus resultados possam nos evidenciar qual é o sentido que primeiro vem à cabeça de um falante quando em contato com uma palavra ou, no mesmo sentido, qual é a posição de um membro em uma dada categoria. Podemos também tentar encontrar o sentido que, por questões de evolução histórica, esteja na base dos outros sentidos. Por fim, podemos proceder a estatísticas de frequência de palavras e de sentidos a partir de um *corpus* previamente tratado e tornado equilibrado segundo determinados critérios.

O problema com os dois primeiros métodos é o custoso número de horas-homem necessárias para dar cabo à tarefa. E horas-homem de especialistas, profissionais experimentados em psicolinguística, linguística e/ou lexicografia. Para alguns poucos domínios seria uma tarefa exequível, mas para mil diferentes domínios não seria plausível proceder dessa maneira. Já o método de frequência em *corpus* é uma empreitada que, com os algoritmos corretos, pode ser executada automaticamente, por computadores.

Segundo Taylor (1992), um fator importante para a identificação do protótipo é a frequência de uso, sendo ela um indício de prototipicidade. A frequência de uso pode ser averiguada e estudada a partir de pesquisas em grandes corpora eletrônicos que, coletados conforme critérios específicos, servirão como uma amostra devidamente equilibrada de uma determinada língua em determinado período de tempo. No entanto, não basta quantificar a frequência em *corpus* do item lexical, é necessário identificar em qual de suas acepções ele está sendo utilizado. Ou mesmo desfazer os casos de homonímia. Isso porque a frequência de emprego de um item lexical, isolado do seu contexto, não corrobora para identificação de efeitos prototípicos. Uma UL é prototípica ou não em relação a uma categoria conceitual específica. A dificuldade que se coloca neste caso é que computacionalmente é muito simples contabilizar a frequência de um item lexical dentro de uma base de dados, mas é, ao mesmo tempo, extremamente difícil identificar em qual acepção aquele item está sendo utilizado. As frases “embora o amor se baseie mais em fatores mentais e emocionais do que em pura sexualidade” e “tem pela educação um amor desbragado”, extraídas do *corpus* do NILC⁴², instanciam diferentes acepções do item lexical *amor*. Ou, por exemplo, as atestações “O coração é o órgão central do sistema circulatório” e “esta esperança íntima e secreta entrou a invadir o coração de minha mãe”. No primeiro caso, *coração* está sendo utilizado em sua acepção

⁴² <http://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>, acessado em 02 de setembro de 2012.

de “órgão muscular dos animais vertebrados” e, no segundo caso, no sentido de “sede das emoções e dos sentimentos”⁴³. E os computadores, por enquanto, são incapazes e calcular essas diferenças.

Conforme 3.1 e 3.2, o sentido de um item lexical pode ser apreendido a partir de i) determinado enquadramento conceitual ou ii) por meio da relação de sinonímia; eliminando-se assim a polissemia inerente ao item lexical. Os dicionários analógicos, ainda que de maneira imperfeita, procedem pela primeira maneira e os dicionários de sinônimos pela segunda. Além disso, como foi dito em 2.5, um dos principais procedimentos metodológicos deste trabalho é utilizar os CSs extraídos dos dicionários de sinônimos para organizar os grupos de analógicos extraídos dos dicionários analógicos. Dessa forma, o resultado obtido mantém o caráter expansivo dos dicionários analógicos ao mesmo em que agrupa ULs com maior proximidade semântica.

Dicionários construídos diretamente a partir de grandes corpora exigem grandes equipes e demandam muito tempo, como foi dito anteriormente. Uma alternativa para essa prática, proposta no âmbito da lexicografia computacional, é a reutilização de dicionários já existentes como *corpus* de referência (BRISCOE, BOGURAEV, 1989). Na construção do TeP, por exemplo, foram utilizados como *corpus* de referência os principais dicionários disponíveis em língua portuguesa (DIAS-DA-SILVA, MORAES, 2003). O projeto Onto.PT, que tem como objetivo a construção automática de uma rede léxico-semântica para o português europeu, estruturada de forma semelhante à WN.Pr, também se valeu da reutilização de dicionários (Dicionário Aberto e Wikcionário) e *thesaurus* (*Open Thesaurus* e o próprio TeP) como fontes primárias de informação lexical.

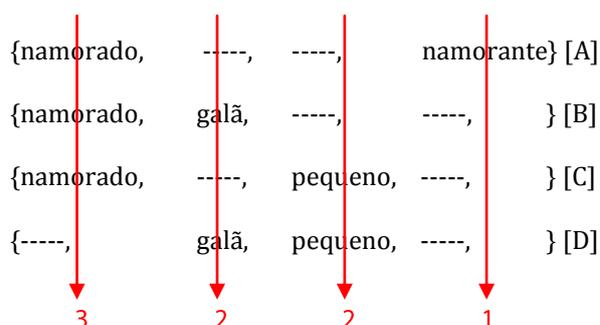
Parece-nos que a melhor solução para o problema, portanto, é utilizar as próprias obras que compõem nosso *corpus* de referência diretamente como fonte primária de informação não apenas para a extração de ULs e CSs, mas também para a frequência de uso. O *corpus* de referência é constituído pelos dicionários de sinônimos e dicionários analógicos elencados no Apêndice II como fontes de informação lexical. A escolha dessas obras não está isenta de incoerências, lacunas e imprecisões. No entanto, como veremos a seguir, são exatamente as idiossincrasias de cada fonte de informação lexical que

⁴³ Definições adaptadas de Aulete Digital in Dicionário Criativo (<http://www.dicionariocriativo.com.br/coracao>), acessado em 02 de setembro de 2012.

permite extrairmos um cálculo para a frequência de cada UL e, conseqüentemente, sua prototipicidade dentro de um CdSA.

A hipótese de que podemos extrair indícios de prototipicidade a partir da reutilização de dicionários existentes está fundamentada nas premissas i) de que essas obras, elaboradas por respeitados lexicógrafos ou corpo de lexicógrafos, seguem uma tradição lexicográfica fundamentada e que, no momento de sua feitura, já observaram os dados reais da língua e concluíram por incluir ou não uma determinada UL dentro de seus CSs; e ii) de que cada uma dessas obras, procurando atingir os mesmos objetivos (agrupar ULs pela relação de sinonímia), adotaram metodologias e critérios diferentes, o que resultou em CSs diferentes e, no entanto, equivalentes. As particularidades de cada obra fazem com que a equivalência entre os CSs não seja perfeita. Longe disso, eles se correspondem em uma espécie de *semelhanças de família*, em que nenhuma UL está necessariamente presente em todos os CSs, mas algumas poucas se repetem muitas vezes e outras tantas aparecem com menor frequência ou mesmo só em um dos CSs equivalentes. Calcular a frequência com que cada item lexical aparece em cada conjunto específico de sinônimos e grupo de analógicos nos dirá quais as ULs que melhor lexicalizam tal ou qual conceito ao mesmo tempo em que preserva uma longa cauda de ULs que, distantes do núcleo de sentido do CdSA, mantêm com ele alguma relação.

É como se estivéssemos perguntando aos lexicógrafos responsáveis por essas obras quais são as ULs que eles consideram indispensáveis para lexicalizar um dado conceito. Como os critérios são imperfeitos pela própria natureza da linguagem e as respostas de cada um – materializadas em suas respectivas obras – serão diferentes, podemos concluir que as que se repetem o maior número de vezes são mais relevantes do que as que se repetem com menor frequência. Se, por exemplo, para o conceito *HOMEM QUE AMA OU É AMADO*, o dicionário A nos diz que o melhor CS para instanciá-lo é {namorado, namorante}, o dicionário B diz que é {namorado, galã}, o dicionário C diz que é {namorado, pequeno} e o dicionário D diz que é {galã, pequeno}, então eu posso concluir, calculando a frequência de cada UL nos CSs equivalentes, que a UL que melhor representa o conjunto em questão é a UL *namorado*, seguida das ULs *galã* e *pequeno*. E a UL *namorante* seria a menos representativa, neste exemplo.



Acreditamos que os resultados obtidos deste modo serão equivalentes ou muito próximos aos resultados obtidos, por exemplo, por meio de questionários aplicados aos falantes de uma comunidade linguística relativamente homogênea. A inteligência emergente que nos falta devido ao fato de não estarmos lidando com centenas ou milhares de dados (questionários) é compensada pela qualidade das nossas fontes de informação, sabidamente produzidas ao longo de milhares de horas de trabalho técnico e especializado. Evidentemente, quanto mais dicionários de sinônimos e dicionários analógicos tivermos como fonte de informação, melhor. Trabalhos futuros poderão incorporar outros dicionários, *thesaurus* e recursos lexicais que não estejam contemplados no Apêndice II de forma a refinar os resultados. Poderão, também, incluir no algoritmo outras heurísticas como, por exemplo, atribuir pesos diferentes para cada fonte de informação lexical.

Como veremos na Seção 6.1.2.3, a reutilização de dicionários de sinônimos e dicionários analógicos como ponto de partida para a extração de informações de prototipicidade exigiu grande cuidado, principalmente com relação aos critérios para determinar se os CSs e CSAs supostamente equivalentes deveriam, de fato, ser agrupados como tal.

5.4 EQUACIONAMENTO LINGUÍSTICO-COMPUTACIONAL: TEORIA DOS CONJUNTOS DIFUSOS PARA MODELAR CATEGORIAS LÉXICO-CONCEITUAIS DIFUSAS

Para representar formalmente o conhecimento levantado na fase linguística, utilizaremos a Teoria dos Conjuntos Difusos (*Fuzzy Set Theory*) (ZADEH, 1965), visto que a Teoria dos Protótipos, no domínio linguístico, leva-nos a trabalhar com categorias que possuem limites vagos. Ou seja, se a Teoria Clássica de Conjuntos formaliza adequadamente o modelo clássico de categorização aristotélico, a Teoria dos Conjuntos

Difusos (doravante **TCD**) é o correspondente adequado na matemática moderna para a formalização de uma categoria léxico-conceitual de estrutura prototípica.

Os termos *conjunto* e *elemento* da Teoria Clássica de Conjuntos são considerados noções primitivas, não têm definição científica. *Conjuntos* são tentativas de organizar conhecimento sobre objetos agrupando-os conforme alguma propriedade ou critério, da mesma maneira que categorias léxico-conceituais agrupam conceitos lexicalizados por meio das relações léxico-semânticas e léxico-conceituais. Os objetos que formam o conjunto são os *elementos* do conjunto. Portanto, os elementos de categorias léxico-conceituais são conceitos lexicalizados.

Conjuntos são denotados, quase sempre, por letras maiúsculas e seus elementos por letras minúsculas. A letra U denota o conjunto universo, no qual está contido todos os possíveis elementos de um dado contexto. A noção de pertinência de elemento a conjunto é também primitiva. Para apontar que determinado objeto x é elemento de um conjunto A, escreve-se: $x \in A$. Para indicar que x não é elemento de A, escreve-se: $x \notin A$. Dado um conjunto universo U, há três métodos para a representação de conjuntos em U:

1. Para conjuntos finitos, podemos enumerar os seus elementos. Dessa forma, um conjunto A cujos elementos são a_1, a_2, \dots, a_n é escrito como $A = \{a_1, a_2, \dots, a_n\}$;
2. Pode-se também representá-lo por meio de uma proposição P, sendo comum o método de representação: $A = \{x | P(x)\}$. Ou seja, A é o conjunto de todos os elementos de U para os quais a proposição P(x) é verdadeira.
3. Ou ainda por meio de uma função característica, que discrimina quais elementos de U são elementos do conjunto e quais não são. A representação por função característica é escrita como:

$$\chi_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}$$

A função característica associa elementos do conjunto U a elementos do conjunto $\{0, 1\}$. Aqui, $\chi_A(x) = 1$ indica que x é elemento de A, enquanto $\chi_A(x) = 0$ indica que x não é elemento de A. Para obter a formalização matemática de um conjunto difuso (*fuzzy set*), Zadeh (1965) baseou-se no fato de que qualquer conjunto clássico (*crisp*) pode ser qualificado por uma função característica. Ou seja, a função característica de um conjunto *crisp* atribui o valor 1 ou 0 a cada elemento do conjunto U, discriminando, assim, os elementos que pertencem ao conjunto *crisp* dos que não pertencem. Como foi dito anteriormente, esta formalização matemática, de pertinência e não-pertinência,

satisfaz as necessidades de formalização do modelo clássico de categorização aristotélico, baseado numa estrutura de atributos necessários e suficientes: um ente pertence a uma categoria determinada se, e somente se, exibe todos os traços característicos que a definem. No entanto, como vimos em 5.2, a TP postula que as categorias conceituais não são estruturas homogêneas, mas se organizam em torno de um centro cognitivo exemplar, dentro do qual as entidades são ordenadas conforme *graus de pertinência*. Dessa forma, é impossível formalizar uma categoria conceitual prototipicamente estruturada por meio de um conjunto *crisp*.

Segundo Zadeh (1965), quando da definição de um conjunto A , sua função característica pode ser generalizada de modo a associar um valor a cada elemento do conjunto U , dentro de um determinado intervalo, que indica o grau de pertinência do elemento ao conjunto A . Essa função é chamada *função de pertinência* e o conjunto definido por ela é chamado conjunto *fuzzy*. Usualmente, o contra-domínio de funções de pertinência é o intervalo $[0,1]$. Assim, a função de pertinência associa elementos de um dado conjunto universo X a números reais do intervalo $[0,1]$, como mostra a Figura 5.3.

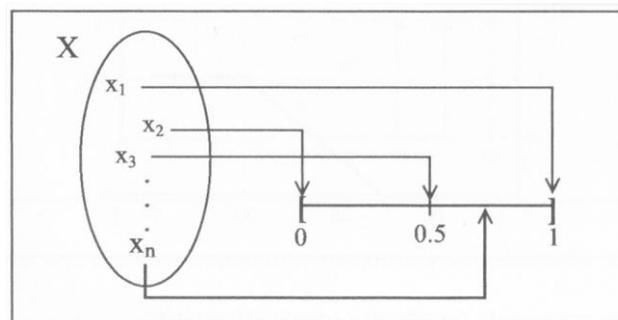


Figura 5.3. Representação Geral de uma função de pertinência (NICOLETI;CAMARGO, 2009, p. 25)

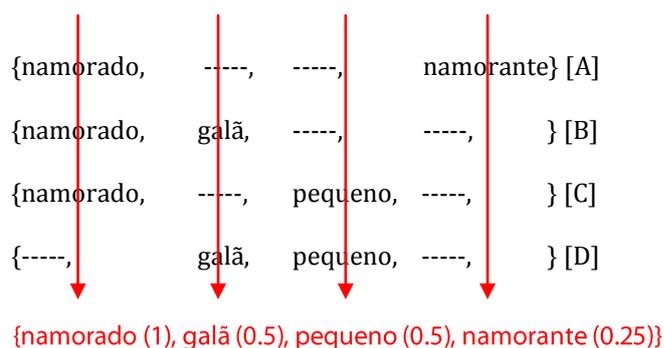
Dado um conjunto *fuzzy* A , sua função de pertinência é notada por μ_A ; ou seja:

$$\mu_A: X \rightarrow [0,1]$$

Em um conjunto de sinônimos *crisp* como, por exemplo, o CS_A {namorado, galã}, os elementos *namorado* e *galã* possuem o valor de pertinência 1; o que equivale a dizer que eles *pertencem* ao conjunto CS_A . E pertencem a ele com o mesmo grau de relevância. Neste caso, os elementos *pequeno* e *namorante* possuem o valor de pertinência 0; e isso é o

mesmo que dizer que eles *não pertencem* ao conjunto CS_A . Em nosso universo hipotético, o lexicógrafo que construiu esse CS não tinha outra opção além da de *incluir* ou *não incluir* as ULs em questão. Se ele tivesse dúvidas com relação à pertença de algumas ULs ou mesmo se possuísse o conhecimento de que algumas ULs são mais relevantes do que outras para o conjunto como um todo, não havia o que ser feito. Em um conjunto *crisp*, os elementos pertencem ou não pertencem ao conjunto. Não há meio termo.

Mas em um conjunto difuso, não funciona assim. Vamos utilizar a TCD para formalizar os efeitos prototípicos presentes nos CdSAs. Abaixo, reproduzimos o mesmo exemplo do final da Seção 5.3, agora com as ULs reunidas em um novo conjunto difuso e com a notação adequada, extraída da TCD.



O que queremos ilustrar com esse exemplo é que o achatamento de diferentes CSs resulta em um CS difuso, em que algumas ULs possuem um grau de pertença ao conjunto maior do que outras. A fusão (*merger*) dos CSs *crisp* CS_A , CS_B , CS_C e CS_D , por meio do cálculo da frequência das ULs, resulta em um novo conjunto de sinônimos difuso (CS difuso). Este é apenas um exemplo simplificado. Pois, além de conjuntos de sinônimos de vários tamanhos extraídos dos dicionários de sinônimos, nosso modelo utiliza-se também do que estamos chamando de conjuntos de sinônimos e analógicos (CSAs), retirados dos grupos de palavras dos dicionários analógicos. Como veremos, nos resultados apresentados na Seção 7, os CdSAs, fruto do achatamento entre CSs e CSAs *crisp*, tendem a ser muito maiores, contendo às vezes dezenas de ULs. O resultado obtido por meio do achatamento entre os CSs e CSAs equivalentes à categoria conceitual *HOMEM QUE AMA OU É AMADO*, ligada ao domínio léxico-conceitual *AMOR*, possui 95 ULs, cf. Quadro 5.2.

Quadro 5.2. CdSA para o conceito de HOMEM QUE AMA OU É AMADO do domínio AMOR

Domínio conceitual AMOR		
(S)	Glosa	Conjunto de Sinônimos e Analógicos difuso (CdSA)
(S7)	HOMEM QUE AMA OU É AMADO	{namorado (1), querido (0.815), bem (0.667), amor (0.63), amado (0.593), caro (0.556), predileto (0.556), amante (0.519), derricho (0.481), amigo (0.444), beijoqueiro (0.444), beijocador (0.407), apaixonado (0.37), dileto (0.333), esposo (0.333), estimado (0.296), enamorado (0.296), flerte (0.259), ídolo (0.259), namorado (0.259), frecheiro (0.259), noivo (0.259), zinho (0.259), chichisbéu (0.259), amásio (0.259), cujo (0.259), ficante (0.259), preferido (0.259), favorito (0.222), benzinho (0.222), estremeado (0.222), idolatrado (0.185), nubente (0.148), benzoca (0.148), coqueluche (0.148), pequeno (0.148), quindim (0.148), prezado (0.148), coisinha (0.148), amoreco (0.148), doce de coco (0.148), amorzinho (0.148), pretendido (0.111), tesouro (0.111), prometido (0.111), galã (0.111), beijador (0.111), de estimação (0.111), caído (0.074), adorado (0.074), queridinho (0.074), concubino (0.074), camarada (0.074), companheiro (0.074), dom-juan (0.074), mimoso (0.074), caricioso (0.074), futuro (0.074), precioso (0.074), menino (0.074), fã (0.074), conquistador (0.074), xodó (0.074), admirador (0.074), pretendente (0.074), camote (0.074), adorador (0.074), namoro (0.037), conversado (0.037), esposado (0.037), agradao (0.037), bem-amado (0.037), admirado (0.037), apreciado (0.037), bem-querer (0.037), recém-casado (0.037), desposado (0.037), néogamo (0.037), nóbio (0.037), doce (0.037), parceiro (0.037), ardente (0.037), babado (0.037), amoriscado (0.037), barregão (0.037), gamado (0.037), amancebado (0.037), embeicho (0.037), encambichado (0.037), rico (0.037), benquisto (0.037), estremeado (0.037), férvido (0.037), enrabiado (0.037), enxodoado (0.037)}

5.4.1 BANCO DE DADOS RELACIONAL FUZZY

Um banco de dados informatizado tem como objetivo manipular e recuperar informações armazenadas eletronicamente. Muitos modelos para representar estas informações têm sido propostos. Um deles é chamado de modelo relacional, que consiste em um conjunto multi-dimensional de relações manipuladas como tabelas.

Considerando um espaço amostral claramente definido, uma lista de ULs e uma lista de conceitos. Cada UL está relacionada a um conceito de acordo com um grau de pertinência, o qual expressa o relacionamento entre a UL e o conceito. O Quadro 5.3 ilustra o relacionamento entre as ULs e os conceitos (S), em que quanto maior a ligação entre a UL e o conceito, maior será o grau de pertinência desta UL ao conceito especificado. Assim, de acordo com os dados fornecidos no Quadro 5.3, pode-se

observar, por exemplo, que a UL1 possui melhor identificação com o conceito S1 do que com o conceito S2, com o qual não mantém relação nenhuma (0). No entanto, a UL1 possui uma ainda maior identificação com o conceito S3, se comparado com o conceito S1.

Quadro 5.3. Exemplo de banco de dados relacional *fuzzy*

ULs/Conceitos	S1	S2	S3	Sn
UL1	0.2	0	1	...
UL2	0.7	0.2	0.8	...
UL3	0.9	0.1	0.3	...
ULn

Não faz parte do escopo desta tese se aprofundar nas questões computacionais. No entanto, veremos um pouco mais sobre banco de dados *fuzzy* em 7.5, quando falaremos, resumidamente, do domínio computacional.

5.5 SÍNTESE DA SEÇÃO

A presente Seção discutiu e elaborou critérios de relevância para os grupos de palavras do DAD. Para tanto, foi necessário definir, de antemão, as noções de *categoria conceitual*, *categoria léxico-conceitual*, *domínio conceitual* e *domínio léxico-conceitual*. Esses termos e noções foram introduzidos para, em seguida, por meio de um breve panorama histórico das teorias da categorização, chegarmos até as modernas concepções da teoria cognitivista e apresentarmos as noções de *semelhança de família*, *Teoria dos Protótipos*, *Redes Radiais* e categorias de *nível de base*. Esses conceitos, que possuem destaque dentro da SC, constituem a base teórica fundamental para a nossa proposta de modelagem do DAD em termos de uma rede semântica difusa.

Como vimos, a SC postula que as categorias léxico-conceituais exibem uma estrutura prototípica, em que alguns dos seus elementos são mais representativos do que outros. Com os princípios de saliência semântica e da teoria dos protótipos definidos, a presente Seção discorreu sobre os procedimentos metodológicos a serem

aplicados para determinar a estrutura difusa de categorias léxico-conceituais. Chegamos à conclusão de que a melhor solução será utilizar as obras que compõe nosso *corpus* de referência como fonte de informação não apenas para a extração de ULs, CSs e CSAs, mas também para calcular a frequência de uso.

Por fim, procurando por uma maneira de representar formalmente o conhecimento levantado na fase linguística, a presente Seção discorreu sobre a TCD. Tendo em vista que a TP, no domínio linguístico, leva-nos a trabalhar com categorias que possuem limites vagos, a TCD demonstrou-se adequada à formalização de categorias léxico-conceituais de estrutura prototípica.

SEÇÃO 6 – MODELO DE REPRESENTAÇÃO E METODOLOGIA DE TRABALHO

As discussões empreendidas nas seções anteriores convergem nessa Seção para a proposição de um modelo de representação linguístico-computacional que satisfaça as necessidades de construção e implementação do DAD. Tal modelo de representação, batizado de *anaset*, será apresentado na Subseção 6.1, onde serão especificados sua definição, seus componentes e as etapas exigidas para a sua construção. Nosso propósito, neste ponto, é delinear um constructo representacional e uma metodologia de trabalho que, à luz das teorias apresentadas até aqui, faça jus ao objetivo principal desse trabalho, ou seja, a construção de um modelo linguístico-computacional para um dicionário analógico em ambiente digital. Esse desígnio exigiu, ao longo de todo o nosso trabalho, um encaminhamento da pesquisa para a proposição de uma representação linguística que possa ser codificada computacionalmente.

Em conjunto com a proposição do modelo de representação, são também apresentados nesta Seção os domínios léxico-conceituais que serão modelados como exemplo (6.2).

6.1 O CONSTRUCTO *ANASET*

O *anaset* é a estrutura por detrás de cada verbete do DAD e equivale aos grupos de palavras dos dicionários analógicos tradicionais. O DAD será organizado de forma que cada entrada corresponda a um domínio léxico-conceitual. Assim como um *synset* é um conjunto de unidades lexicais sinônimas que evocam um mesmo conceito, o *anaset* é um conjunto de categorias conceituais lexicalizadas e relacionadas que evocam um mesmo domínio conceitual. Portanto, cada domínio léxico-conceitual será modelado em um *anaset* diferente. Cada categoria conceitual será instanciada por um conjunto difuso de unidades lexicais sinônimas e análogas, fruto do achatamento (cf. 5.3) entre conjuntos de sinônimos (CSs) equivalentes extraídos de dicionários de sinônimos e conjuntos de sinônimos e analógicos (CSAs) extraídos de dicionários analógicos. Ou seja, as ULs não terão o mesmo *peso* dentro do conjunto.

Um *anaset* é, portanto, definido como uma estrutura léxico-conceitual que organiza categorias conceituais lexicalizadas no interior de um mesmo domínio conceitual valendo-se, para isso, de conjuntos difusos de sinônimos e analógicos.

O constructo *anaset* é construído tendo por base uma matriz léxico-conceitual (Quadro 6.1) onde as ULs extraídas do *corpus* e das obras de referência ocupam a coluna da extrema esquerda e as demais colunas correspondem às categorias conceituais pertinentes ao domínio léxico-conceitual em questão. No cruzamento, por exemplo, entre a linha **UL1** e a coluna **S1**, a célula será preenchida com os CSs e CSAs – extraídos dos dicionários de sinônimos e dicionários analógicos *X, Y, Z* –, que possuam a unidade lexical *UL1* e ao mesmo tempo instanciarem a categoria conceitual *S1*. Se, além disso, os CdSAs pertencerem também a um *frame* específico e, portanto, forem organizados sob a divisão ‘relações sintagmáticas’, então eles devem ser modelados conforme os elementos do *frame* (EF) ao qual pertencem. Se, por exemplo, *S5* e *S6* representam, respectivamente, os conceitos de *HOMEM QUE AMA OU É AMADO* e *MULHER QUE AMA OU É AMADA*, então eles pertencem ao EF *AMANTES*.

Quadro 6.1 – Componentes do *anaset* na matriz léxico-conceitual

Grande Categoria						
Anaset A	Relações Paradigmáticas			Relações Sintagmáticas		
	S1	S2	S3	EF1	EFn	
				S4	S5	Sn
UL1	{CS} [X] {CS} [Y] {CSA} [Z]					
UL2						
ULn						

Dessa forma, cf. Quadro 6.1, os componentes do *anaset* são:

- i. a Grande Categoria dentro da qual está inserido o domínio conceitual em questão (em verde escuro);
- ii. as unidades lexicais de entrada (UL) que lexicalizam as categorias conceituais do domínio (em azul);
- iii. as categorias conceituais (S) do domínio (em laranja);

- iv. a divisão das categorias conceituais entre os elementos do *frame* (EF) (em verde-água);
- v. a divisão das categorias conceituais entre relações paradigmáticas e sintagmáticas (em roxo);
- vi. os conjuntos de sinônimos (CS) e/ou analógicos (CSA) (em verde);
- vii. as marcações da fonte de informação lexical (em vermelho).

6.1.2 ETAPAS DE CONSTRUÇÃO DO *ANASET*

Nessa Seção iremos explicitar cada uma das quatro etapas necessárias para o preenchimento da matriz léxico-conceitual e a consolidação dos resultados para cada *anaset* que será modelado. A primeira etapa diz respeito à identificação e coleta das ULs que deverão ser extraídas do *corpus* e das obras de referência para ocuparem a coluna da extrema esquerda da matriz (6.1.2.1). A segunda etapa refere-se aos procedimentos de identificação e delimitação das categorias conceituais pertinentes ao domínio em questão, em que cada categoria conceitual corresponderá a uma coluna da matriz (6.1.2.2). A identificação, coleta e arranjo adequado dos conjuntos de sinônimos e/ou analógicos no interior da matriz satisfaz a terceira etapa do processo (6.1.2.3). Por fim, depois da matriz completamente preenchida, a quarta etapa refere-se à consolidação dos resultados (6.1.2.4).

Todos os dados obtidos ao longo das etapas de construção do *anaset* foram digitalizados e copiados para o programa *Microsoft Excel™*, local em que foram armazenados e, posteriormente, foram executados os cálculos necessários.

6.1.2.1 IDENTIFICAÇÃO E COLETA DAS ULs - ETAPA 1

O ponto de partida é gerar a lista de todas as ULs pertencentes ao domínio léxico-conceitual em questão que servirão como ULs de entrada para o *anaset* equivalente ao domínio. Essa lista de ULs é a que deverá ocupar a coluna da extrema esquerda na matriz léxico-conceitual do *anaset*. Tendo como base as ULs extraídas do *corpus* (cf. 6.2), procura-se nas fontes de informação léxico-conceitual (cf. Apêndice II) por ULs sinônimas e análogas que possam também pertencer ao mesmo domínio léxico-

conceitual. Para tanto, foram estabelecidos alguns critérios de modo a não se produzirem resultados distorcidos e/ou incompatíveis com os objetivos do DAD. São eles:

- (i) Não elencar as ULs estritamente técnicas e/ou de cunho essencialmente enciclopédico, que podem ser identificadas por não constarem em nenhuma das obras de referência de língua geral e/ou pela ausência ou baixa frequência em *corpus* equilibrados. Ex.: patafísica, peça-problema, memória emocional, crítica formalista etc.;
- (ii) Limitar a expansão da coleta das ULs pelos domínios léxico-conceituais adjacentes. Caso contrário, a coleta de ULs por meio de sinônimos e analógicos poderia progredir indefinidamente. No entanto, para proceder dessa forma, estamos levando em consideração os domínios léxico-conceituais consolidados no *Roget's Thesaurus* (1000 grupos de analógicos, cf. Seção 4). Por exemplo, a UL *namorado* nos leva, entre sinônimos e analógicos, às ULs *querido, caro, bem, benzinho, noivo* etc. A UL *noivo* nos leva às ULs *nubente, prometido, contraente, marido* etc. Todas poderiam entrar no domínio léxico-conceitual AMOR. No entanto, existindo o domínio léxico-conceitual CASAMENTO, ele prototipicamente acomoda as quatro últimas;
- (iii) Não elencar as ULs complexas, estrangeirismos e nomes próprios, pois não possuem entradas nos dicionários que compõe nosso *corpus* de referência. No entanto, poderão estar presentes no interior dos CSAs e serão mantidos para os cálculos das etapas seguintes. Ex.: amor materno, medula óssea, *commedia del'arte, fiancée, Afrodite, Tália*;
- (iv) As formas variantes foram agrupadas na mesma entrada e tiveram de ser alteradas para a forma preferencial dentro dos CSAs para que o algoritmo de contagem de frequência pudesse reconhecê-las adequadamente. Ex.: abdômen/abdome, curinga/coringa;

Esses critérios se mostraram necessários para que os objetivos inicialmente propostos fossem atingidos sem com que muitas outras heurísticas e etapas tivessem de ser criadas e muitas outras fontes de informação tivessem de ser incorporadas ao processo. Isso faria com que o trabalho como um todo tomasse uma proporção que não é compatível com o escopo dessa tese. Evidentemente, em trabalhos futuros, essas limitações deverão ser enfrentadas para que os resultados sejam ainda mais completos e refinados.

É importante frisar que a lista de ULs gerada para ocupar a coluna da esquerda da matriz não corresponde a todas as ULs que poderão estar presentes nos CSAs finais. Essa lista é o que vamos chamar de **ULs de entrada**. Ela serve para guiar o processo de

preenchimento da matriz e é limitada pelos critérios apontados anteriormente. No entanto, as ULs não incluídas como ULs de entrada ainda poderão aparecer nos resultados finais, dentro dos CdSAs.

Tomemos como exemplo a lista das ULs pertencentes ao domínio AMOR extraídas diretamente do *corpus*. Conforme 6.2, as ULs que aparecem nas letras das canções do álbum *Grande Circo Místico*, com composições de Chico Buarque e Edu Lobo, são *amante*, *beijo*, *bem*, *namorado*, *noiva* e *romance*. Consultando as fontes de informação lexical, vemos que a UL *amante* nos leva às ULs *amásio* e *amigo*, entre outras. A UL *amigo* pertence prototipicamente ao domínio léxico-conceitual AMIZADE e, portanto, conforme o critério (ii), não será incluída na lista de ULs de entrada do domínio AMOR. Já a UL *amásio* – que significa “indivíduo amancebado, amigado; amante” – deve ser incorporada à lista de ULs de entrada. Da mesma forma, a UL *noiva* nos leva às ULs *predileta* e *esposa*, entre outras. A UL *esposa* pertence prototipicamente ao domínio léxico-conceitual do CASAMENTO e deve ser descartada, ao passo que a UL *predileta* deverá compor a lista de ULs de entrada do domínio AMOR. Vejamos, no Quadro 6.2, um exemplo de resultado para a etapa 1. Uma tabela em que aparecem todas as ULs de entrada do *anaset* AMOR. Destacamos, em negrito, as ULs que pertencem à categoria conceitual MULHER QUE AMA OU É AMADA.

Quadro 6.2. ULs de entrada do *anaset* AMOR

Domínio léxico-conceitual AMOR (69 ULs)			
abraço	apego	dileção	namorisco
admiração	ardor	fervor	namoro
admirador	atração	flerte	noiva
adoração	beijo	frecheiro	noivo
adorador	beijocador	galã	paixão
afeição	beijoqueiro	galanice	predileção
aferro	bem	galanteador	predileta
afeto	benevolência	galanteio	predileto
agarramento	benquerença	idílio	pretendente
amada	calor	idolatria	pretendida
amado	carinho	inclinação	querença
amante	caro	intimidade	querida
amásia	conquistador	namorada	querido
amásio	cortejador	namoradeiro	simpatia
amor	dedicação	namorado	ternura
amorico	derretimento	namorador	
apaixonada	derricho	namoramento	
apaixonado	devoção	namorico	

Podemos ver que a UL *esposa* não é uma das ULs de entrada do domínio AMOR. No entanto, ela aparece nos resultados consolidados do CdSA referente à categoria conceitual MULHER QUE AMA OU É AMADA, do *frame* *Relação_amorosa*, do *anaset* AMOR (cf. 7.2.4). Porém, seu peso é de apenas 0.091, o que a faz ocupar a posição de número 57 dentro do conjunto, próxima à ULs como *amiga* (0.152), *mulher* (0.121), *recém-casada* (0.061) e *companheira* (0.03). No topo desse CdSA, com os maiores pesos, estão as ULs *namorada* (1), *amada* (0.788), *querida* (0.727) e *amante* (0.515), respectivamente. A UL *predileta*, com o peso de 0.455, aparece na sétima posição. Esse é apenas um exemplo de como ULs que não figuram como ULs de entrada podem aparecer nos CSAs consolidados, porém com uma relevância menor. Na Seção 7, ao longo da modelagem dos *anaset*s, veremos essas questões com mais clareza.

Uma vez identificadas, as ULs são inseridas, uma para cada linha, na primeira coluna de uma planilha do programa *Microsoft Excel™*. Cada planilha reproduz as linhas (ULs) e colunas (S) da matriz de um *anaset* específico. Com a primeira coluna preenchida com todas as ULs de entrada, a próxima etapa é identificar e delimitar as categorias conceituais pertinentes ao *anaset* que está sendo modelado.

6.1.2.2 IDENTIFICAÇÃO E DELIMITAÇÃO DAS CATEGORIAS CONCEITUAIS - ETAPA 2

A segunda etapa serve para construir a estrutura conceitual de cada *anaset* que será modelado. A estruturação conceitual dos *anaset*s começa pela definição da Grande Categoria à qual o domínio conceitual em questão está prototipicamente inserido. Isso se dará pela análise do domínio à luz do que foi discutido na Seção 4 e valendo-se dos sistemas conceituais citados em 4.4 como paradigmas.

Em seguida, tendo como base as fontes de informação léxico-conceitual (cf. Apêndice II.2), serão identificadas, para cada domínio conceitual, as categorias conceituais que deverão ser incluídas em cada *anaset*.

Tomemos como exemplo a modelagem conceitual do *anaset* TEATRO. Como veremos em 7.3.2, não é possível compreender o conceito TEATRO independentemente de um contexto social e cultural complexo. Veremos que o domínio conceitual TEATRO ancora-se, Dessa forma, na Grande Categoria MUNDO SOCIAL e, assim, a matriz do *anaset* TEATRO começa a ser preenchida, conforme o Quadro 6.3.

Quadro 6.3. Matriz preenchida com a Grande Categoria onde se inclui o *anaset* TEATRO: MUNDO SOCIAL

MUNDO SOCIAL						
TEATRO	Relações Paradigmáticas			Relações Sintagmáticas		
	S1	S2	S3	EF1	EF2	
				S4	S5	S6
UL 1	{CS} [X] {CS} [Y] {CSA} [Z]					

Em seguida, nosso objetivo passa a ser identificar quais são os conceitos ou categorias conceituais (S) que pertencem ao domínio TEATRO e que, portanto, estruturam conceitualmente o *anaset* TEATRO.

Por sua estrutura onomasiológica, os dicionários analógicos são as ferramentas apropriadas para a identificação e a delimitação dos conceitos. Ainda que, conforme 2.2, os dicionários analógicos não apresentem uma metodologia clara para a divisão dos conceitos, é possível identificar alguns agrupamentos mais ou menos homogêneos. No caso do domínio conceitual TEATRO, utilizamos ainda, em conjunto com os dicionários analógicos, alguns *thesaurus* e dicionários específicos do domínio como, por exemplo, o *Lexique International de Termes Techniques de Théâtre* (RAE, SOUTHERN, 1964) e o *Dicionário de Teatro* (PAVIS, 1999). Esse tipo de obra de referência específica de domínio foi utilizada complementarmente aos dicionários de língua geral e estão arroladas na Seção 6.3.2, item **iii**). Nelas, os termos são arrolados de acordo com uma estrutura conceitual própria do domínio. No *Dicionário de Teatro* (PAVIS, 1999), por exemplo, as ULs são organizadas sob oito categorias do índice temático: *dramaturgia, texto e discurso, ator e personagem, gênero e formas, encenação, princípios estruturais e questões de estética, recepção do espetáculo e semiologia*. Em conjunto com os dicionários analógicos, essas obras provêm o arcabouço conceitual adequado para a modelagem dos domínios que, pertencendo ao MUNDO SOCIAL, possuem uma estrutura interna típica dos domínios de especialidade. Como ponto de partida, utilizamos também os dicionários de língua geral, cujo percurso é semasiológico. Procurando pelo item lexical *teatro* nas entradas dos dicionários semasiológicos, encontramos uma correspondência entre as

diferentes acepções da palavra e alguns dos diferentes CSAs mais ou menos divisados no interior dos grupos de palavras dos dicionários analógicos.

Conforme o Quadro 6.4, as categorias conceituais serão instanciadas por uma pequena glosa e serão exemplificadas com CSs ou CSAs em português e em inglês, coletados nas fontes de informação, que instanciam o conceito ou parte da categoria conceitual em questão. Haverá, também, para cada categoria conceitual, uma frase-exemplo extraída do *corpus* do NILC para abonar pelo menos um de seus conceitos em situação de uso. Na sequência, as categorias conceituais serão separadas conforme relações paradigmáticas e sintagmáticas. Caso estejam dentro das relações sintagmáticas, as categorias conceituais serão também organizadas conforme os EFs aos quais correspondem. Por fim, as colunas da matriz (que correspondem às categorias conceituais) serão ordenadas de acordo com a sua prototipicidade e radialidade (cf. Seção 5).

Quadro 6.4. Exemplo de duas categorias conceituais do *anaset* TEATRO

Categorias conceituais do <i>anaset</i> TEATRO			
(S)	Glosa	CSs/CSAs	Exemplo
Relação Paradigmática			
(S1)	lugar ou edifício destinado à apresentação de obras dramáticas, óperas ou outros espetáculos públicos	{teatro, anfiteatro} [MIC] {theater, theatre, house} [WN]	“O centro abriga um teatro municipal com 320 lugares, biblioteca e o cine clube Sétima Arte.” [NILC]
Relação Sintagmática			
(S12)	partes do teatro	{palco, vistas, pano de boca, telão, tablado, prosênio, rampa, estrado, ribalta, gambiarra, bambolina, ..., trampolim, trapézio} [AZE] {stage, the boards, acting area, playing area, thrust stage, forestage, apron, passerelle, proscenium, ..., fire curtain, stage door} [RIT]	“É mais provável revê-la no palco como atriz.” [NILC]

O Quadro 6.4 exhibe dois exemplos de categorias conceituais pertencentes ao *anaset* TEATRO. Podemos ver que um dos conceitos prototípicos do domínio é o de TEATRO enquanto espaço físico, glosado por “lugar ou edifício destinado à apresentação de obras dramáticas, óperas ou outros espetáculos públicos”. Os CSs {teatro, anfiteatro} e {theater, theatre, house} extraídos, respectivamente, do dicionário de sinônimos Michaelis [MIC] e da WN.Pr [WN], servem como exemplo de instanciação do conceito nas

diferentes línguas. Por fim, na última coluna, como foi dito, uma frase-exemplo extraída do *corpus* do NILC abona o conceito em situação de uso.

No nosso modelo, o conceito S1 abarca ULs em relação paradigmática como, por exemplo, as ULs *teatro, anfiteatro, cinema, cinerama, teatro de arena, casa de ópera, sala de concertos, casa de espetáculos* etc. No entanto, o próprio conceito S1 possui uma estrutura interna de conceitos sintagmaticamente relacionados. A categoria conceitual PARTES DO TEATRO (S12) abarca esses conceitos e suas respectivas ULs como, por exemplo, as ULs *palco, cenário, tablado, proscênio, bastidores, pano de fundo, corrediça, camarim* etc.

Com as categorias conceituais identificadas e delimitadas, elas são inseridas na planilha do programa *Microsoft Excel™* – uma categoria conceitual para cada coluna. Neste ponto, temos a planilha preenchida com as ULs de entrada e com as categorias conceituais do *anaset*. Dessa forma, podemos iniciar o processo de preencher as células da planilha com os CSs e CSAs extraídos das fontes de informação.

6.1.2.3 IDENTIFICAÇÃO, COLETA E ARRANJO DOS CSs E CSAs - ETAPA 3

As células da matriz léxico-conceitual serão preenchidas com os CSs e CSAs identificados e coletados nas fontes de informação léxico-conceitual (Apêndice II). Para cada UL da matriz, buscam-se os conjuntos de sinônimos ou agrupamentos de analógicos que a contenham e que, ao mesmo tempo, instancie uma ou mais de uma categoria conceitual previamente definida, conforme 6.1.2.2.

Tomemos como exemplo a identificação, coleta e arranjo dos CSs e CSAs do domínio AMOR. Tendo como base a UL de entrada *afeição*, conforme Quadro 6.2, vamos às fontes de informação léxico-conceitual arroladas no Apêndice II e copiamos/digitalizamos todos os CSs e CSAs que possuem a UL *afeição* e que, aparentemente, pertencem ao domínio AMOR. Algumas vezes, os verbetes de entrada nos dicionários de sinônimos fazem remissão a outros verbetes. Nesses casos, seguimos a referência até encontrar o CS apropriado. Vejamos, por exemplo, o verbe **afeição** do *Dicionário Houaiss de Sinônimos e Antônimos* (HOUAISS, 2008):

- (i) **afeição** *s.f.* **1 afeto** (ver) **2 pendor:** inclinação (ver), vocação <*a. pelas artes*>

Seguindo a remissão da acepção 1, temos

- (ii) **afeto** *s.m.* adoração, afeição, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, (ben)querença, simpatia, ternura.

As demais acepções do verbete **afeto** são correspondentes à categoria gramatical dos adjetivos e, portanto, não nos interessa. Seguindo a remissão da acepção 2 do verbete **afeição**, no sentido especificado, temos

- (iii) **inclinação** *s.f.* 5 *fig.* **vocação:** aptidão, gosto, habilidade, jeito, pendor, propensão, tendência, veia <*tem i. artística*>

Evidentemente, esse sentido de *afeição* que aparece em (iii), como sinônimo de *inclinação, vocação, aptidão* etc., não pertence ao domínio AMOR, mas sim aos domínios léxico-conceituais TENDÊNCIA e/ou HABILIDADE. Já a acepção de **afeto** que aparece em (ii), referida pela acepção 1 do verbete **afeição** – conforme aparece em (i) – pertence, claramente, ao domínio léxico-conceitual AMOR. Dessa forma, para este caso, o único agrupamento de sinônimos identificado no *Dicionário Houaiss de Sinônimos e Antônimos* (HOUAISS, 2008) para a UL de entrada *afeição* é aquele elencado em (ii).

No entanto, as informações léxico-conceituais identificadas como válidas aos nossos propósitos devem ser coletadas de maneira padronizada, de forma que permaneçam simétricas e manipuláveis para futuras ações computacionais. Utilizaremos, para tanto, o mesmo padrão de formatação da WN.Pr. Ou seja, o CS é formatado como um conjunto de ULs separadas por vírgulas e fechadas entre chaves, tal como aparece abaixo.

$$\{UL_1, UL_2, UL_3, UL_n\}$$

Não pode haver nenhum outro tipo de caracter entre as chaves. O que exige adaptações em alguns poucos casos específicos, como em (ben)querença, que aparece em (ii). Na verdade, a formatação adotada em dicionários impressos serve para economizar espaço. Neste caso, (ben)querença está representando duas ULs diferentes: *benquerença* e *querença*. E é assim, separadas em palavras distintas, que iremos representar as ULs

dentro dos nossos CSs e CSAs. Da mesma forma, os verbetes e as entradas de cada acepção, marcadas em negrito pela formatação original do *Dicionário Houaiss de Sinônimos e Antônimos*, devem ser incluídas nos CSs e CSAs padronizados sem qualquer tipo de distinção gráfica. Assim, o CS resultante da coleta no *Dicionário Houaiss de Sinônimos e Antônimos* para a UL de entrada *afeição* deve ser:

{afeto, adoração, afeição, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU]

É dessa forma que os CSs e CSAs serão incluídos na matriz léxico-conceitual do *anaset* que lhe corresponde. No entanto, para organizar o texto da tese e facilitar a visualização dos CSs e CSAs coletados, eles serão antes exibidos em tabelas – uma para cada fonte de informação léxico-conceitual. O Quadro 6.5 mostra o exemplo de uma dessas tabelas, com alguns dos CSs extraídos do *Dicionário Houaiss de Sinônimos e Antônimos* para serem utilizados no *anaset* AMOR.

Quadro 6.5. Exemplo de CSs extraídos do *Dicionário Houaiss de Sinônimos e Antônimos* para o *anaset* AMOR

HOUAISS, A. <i>Dicionário Houaiss de Sinônimos e Antônimos</i> [HOU]
{adorador, fã},
{beijo, ósculo},
{caro, dileto},
{agarramento, agarra, agarração},
{agarramento, intimidade, ligação},
{favorito, predileto, preferido},
{idílio, romance, namoro},
{predileta, favorita, preferida},
{querida, admirada, apreciada},
{querido, admirado, apreciado},

Uma vez identificados, os CSs e CSAs são inseridos na planilha do programa *Microsoft Excel™*, sempre respeitando o seguinte procedimento: um CS/CSA deve ser disposto na célula i) cuja linha seja de uma UL de entrada presente no CS/CSA e ii) cuja a coluna seja de uma categoria conceitual cujo o CS/CSA instancie. Ou seja, no cruzamento, entre a linha **UL1** e a coluna **S1**, a célula será preenchida com os CSs e CSAs que possuem a unidade lexical *UL1* e ao mesmo tempo instanciem a categoria conceitual *S1*. Executar esse arranjo, no entanto, não é algo trivial. Exige muito cuidado e análise das fontes de informação para se chegar ao melhor arranjo dos CSs/CSAs. Isso porque nem sempre é claro perceber em qual categoria conceitual deve-se incluir um CS/CSA

específico, principalmente quando se trata de categorias conceituais, em relação paradigmática, pertencentes a um domínio do MUNDO PSÍQUICO. As categorias conceituais desse tipo parecem ser as que possuem fronteiras mais difusas. Às vezes, metade das ULs de um CS/CSA, em seu conjunto, instancia claramente uma determinada categoria conceitual, enquanto a outra metade parece instanciar melhor outra categoria conceitual. Evidentemente, tecnicamente falando, são todas as ULs de um CS/CSA que, juntas, instanciam uma categoria conceitual. Esse é o pressuposto do paradigma da WN.Pr e propostas congêneres. E, de fato, se olharmos primeiro para a categoria conceitual e depois para as ULs presentes no CS/CSA veremos que todas são candidatas àquela lexilização. Porque, no conjunto, percebemos que todas apontam na mesma direção. No entanto, alguns grupos menores de ULs dentro do mesmo CS/CSA lexicalizam melhor determinada categoria conceitual do que outros. E é exatamente esse um dos pontos-chave do nosso modelo. Como os CSs/CSAs coletados ainda são nivelados (ou seja, não difusos), todas as ULs possuem o mesmo *status* dentro do conjunto. Ao proceder, dessa maneira, com o arranjo dos CSs/CSAs na matriz tem-se a sensação de se estar diante do problema **figura-fundo**. Ao se olhar para um CS/CSA com a lente de uma determinada categoria conceitual, um grupo menor de ULs se torna saliente em relação às demais. Ao se trocar de lente, outro grupo de ULs no interior do mesmo CS/CSA é que se destaca. Sempre que isso acontecer, o CS/CSA deve ser incluído nas duas (ou mais) categorias conceituais, sem qualquer prejuízo ao modelo. O que parece ser um problema é na verdade uma das bases em que se assenta nossa proposta. Outros CSAs e, principalmente, CSs (menores) serão dispostos na mesma célula, outros na mesma coluna. Ao final do processo, depois de rodar o algoritmo, esses darão forma e saliência àqueles que sofrem de uma dupla instanciação.

O problema está, na verdade, quando há carência de CSs/CSAs para instanciar uma categoria conceitual que, claramente, deveria ser instanciada. Infelizmente, as fontes de informação possuem, às vezes, algumas lacunas de aceções. E o número reduzido de fontes para extração dos CSs/CSAs é uma fragilidade diante de problemas como esse. Se duas ou três fontes de informação apresentam lacunas, a consolidação do CdSA no interior dessa célula pode ser prejudicada. Felizmente, casos como esse são raros. E, na verdade, o resultado mais importante para nós é a consolidação das colunas e não das células, como veremos em 6.1.2.4.

6.1.2.4 CONSOLIDAÇÃO DOS RESULTADOS - ETAPA 4

Com a matriz completamente preenchida com i) as ULs de entrada, ii) as categorias conceituais (S) e iii) os CSs e CSAs correspondentes, os resultados podem ser consolidados de duas maneiras.

A primeira delas é consolidar os resultados no interior de cada célula, gerando CdSAs de motivação semasiológica. Ou seja, partindo de uma determinada UL, quais são os seus sinônimos e analógicos (dos mais fortemente relacionados para os menos relacionados) dentro daquele significado.

$$\text{célula UL1/S1} = \{\text{CS}\} [x] + \{\text{CS}\} [y] + \{\text{CSA}\} [z] = \{\text{CdSA_sem}\}_{\text{UL1/S1}}$$

O CdSA_{sem}_{UL1/S1} arrola os sinônimos e analógicos da UL1 no sentido S1 dos mais pertinentes para os menos pertinentes. Entre as chaves estão os CSs e CSAs que serão computados, cada qual marcado com o sua respectiva fonte de informação (entre colchetes).

As matrizes de cada *anaset* foram geradas no programa *Microsoft Excel™*, onde todos os dados foram armazenados e a partir do qual foram executados os cálculos necessários. Por conta do volume dos dados, não é possível replicar as matrizes em seu formato original aqui. Porém, no Quadro 6.6 podemos ver parte dos dados, considerando apenas cinco ULs de entrada (ULs) para apenas uma categoria conceitual (S). Neste exemplo, temos as primeiras ULs de entrada do *anaset* AMOR que possuem CSs e CSAs para a categoria conceitual S1, glosada por “afeição muito intensa de uma pessoa por outra”.

Quadro 6.6. Exemplo da matriz léxico-conceitual do *anaset* AMOR preenchida com CSs e CSAs

Anaset AMOR		
	ULs	Conjuntos de Sinônimos e/ou Analógicos (CSs/CSAs)
(S1) afeição muito intensa de uma pessoa por outra	afeição	{afeto, adoração, afeição , amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU] + {amor, afeição , afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, enfatuação} [IDS] + {dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição , dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência} [AZE] + { afeição , afeto, amor} [MIC] + { afeição , afeto, amizade, amor, benquerença, carinho, inclinação, propensão,

		simpatia} [FER] + {afeição, afeiçoamento, afetividade, afeto, amizade, amor, querença, querência} [TEP]
	afeto	{afeto, adoração, afeição, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU] + {amor, afeição, afeto , simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfatuação} [IDS] + {amor, carinho, idolatria, afeto , amorico, amorosidade, amorio, inclinação} [AZE] + {afeição, afeto , amor} [MIC] + {afeto, afeição, amizade, inclinação, simpatia, amor, paixão, emoção} [FER] + {afeição, afeiçoamento, afetividade, afeto , amizade, amor, querença, querência} [TEP]
	amor	{amor, afeição, afeto, amizade, apego, apreço, benquerença, benquerer, carinho, estima, simpatia, ternura} [HOU] + {amor, afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfatuação} [IDS] + {amor, carinho, idolatria, afeto, amorico, amorosidade, amorio, inclinação} [AZE] + {afeição, afeto, amor } [MIC] + {amor, afeto, afeição, amizade, dileção, inclinação, simpatia, paixão, bem-quer, benquerença, idolatria, adoração} [FER] + {afeição, afeiçoamento, afetividade, afeto, amizade, amor , querença, querência} [TEP]
	apego	{afeto, afeição, amizade, amor, apego , afinidade, devotamento, estima, encanto, fidelidade} [HOU] + {amor, afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego , gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfatuação} [IDS] + {dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição, dedicação, querença, admiração, apego , aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência} [AZE] + {apego, afeto, devoção, agarramento} [MIC] + {apego, apegamento, afeto, inclinação, amizade, amor, dedicação} [FER]

Tendo a matriz preenchida como no Quadro 6.6, rodamos o algoritmo de consolidação (*merger*) para cada célula, de forma a extrair como resultado um novo conjunto, em que a frequência de repetição das ULs entre os CSs e CSAs originais irá determinar o seu peso e, conseqüentemente, a sua relevância.

O algoritmo roda em cima de arquivos no formato .csv, que representa tabelas em formato de texto. Nesse tipo de arquivo, cada linha representa uma linha da tabela, e cada coluna é separada por ";". O algoritmo⁴⁴ consiste em percorrer o arquivo .csv utilizando-se de expressões regulares para identificar todas as ULs e seus respectivos CSs e CSAs. O Quadro 6.7 ilustra os resultados consolidados, após a aplicação dos cálculos do algoritmo, para as primeiras cinco ULs de entrada do *anaset* AMOR que lexicalizam a categoria conceitual S1.

⁴⁴ Para ver o algoritmo inteiro, abra o arquivo 'algorithm.txt' presente no CD em anexo.

Quadro 6.7. Exemplo da matriz léxico-conceitual do *anaset* AMOR consolidada com CSAs_difusos_sem

Anaset AMOR		
	ULs	CSAs_difusos_sem
(S1) afeição muito intensa de uma pessoa por outra	afeição	{afeição (1), amor (0.833), afeto (0.833), simpatia (0.667), benquerença (0.5), querença (0.5), ternura (0.5), carinho (0.5), amizade (0.5), apego (0.5), benevolência (0.333), dileção (0.333), dedicação (0.333), inclinação (0.333), estremecimento (0.167), estima (0.167), meiguice (0.167), derretimento (0.167), idílio (0.167), derricho (0.167), idiopatia (0.167), intimidade (0.167), querência (0.167), adoração (0.167), afetividade (0.167), afeioamento (0.167), conchego (0.167), propensão (0.167), constância (0.167), comunhão de sentimentos (0.167), zelo (0.167), desvelo (0.167), enfatuação (0.167), preferência (0.167), predileção (0.167), cuidado (0.167), chamego (0.167), fraternidade (0.167), caridade (0.167), boa vontade (0.167), aferro (0.167), admiração (0.167), gosto (0.167)}
	afeto	{afeto (1), amor (1), afeição (0.833), inclinação (0.5), carinho (0.5), simpatia (0.5), amizade (0.5), ternura (0.333), apego (0.333), querença (0.333), comunhão de sentimentos (0.167), caridade (0.167), meiguice (0.167), benquerença (0.167), adoração (0.167), dedicação (0.167), dileção (0.167), estima (0.167), fraternidade (0.167), boa vontade (0.167), amorio (0.167), amorosidade (0.167), paixão (0.167), emoção (0.167), querência (0.167), afetividade (0.167), afeioamento (0.167), amorico (0.167), idolatria (0.167), gosto (0.167), benevolência (0.167), chamego (0.167), cuidado (0.167), enfatuação (0.167), desvelo (0.167), zelo (0.167)}
	amor	{amor (1), afeto (1), afeição (0.833), inclinação (0.5), carinho (0.5), simpatia (0.5), amizade (0.5), benquerença (0.333), idolatria (0.333), ternura (0.333), apego (0.333), benevolência (0.167), boa vontade (0.167), fraternidade (0.167), comunhão de sentimentos (0.167), caridade (0.167), benquerer (0.167), apreço (0.167), estima (0.167), gosto (0.167), chamego (0.167), adoração (0.167), bem-quer (0.167), afeioamento (0.167), afetividade (0.167), querência (0.167), querença (0.167), paixão (0.167), dileção (0.167), zelo (0.167), cuidado (0.167), desvelo (0.167), enfatuação (0.167), amorio (0.167), amorosidade (0.167), amorico (0.167)}
	apego	{apego (1), afeto (0.8), amor (0.6), afeição (0.6), dedicação (0.4), benevolência (0.4), simpatia (0.4), inclinação (0.4), ternura (0.4), amizade (0.4), boa vontade (0.2), fidelidade (0.2), afinidade (0.2), gosto (0.2), chamego (0.2), fraternidade (0.2), caridade (0.2), encanto (0.2), comunhão de sentimentos (0.2), estima (0.2), devotamento (0.2), carinho (0.2), cuidado (0.2), zelo (0.2), derretimento (0.2), derricho (0.2), idílio (0.2), idiopatia (0.2), intimidade (0.2), apegamento (0.2), agarramento (0.2), devoção (0.2), conchego (0.2), constância (0.2), aferro (0.2), dileção (0.2), enfatuação (0.2), desvelo (0.2), predileção (0.2), preferência (0.2), admiração (0.2), querença (0.2), benquerença (0.2), estremecimento (0.2)}

Os resultados do Quadro 6.7 sugerem, por exemplo, que partindo da UL de entrada *afeição*, as ULs com relação de sinonímia mais forte são *amor*, *afeto* e *simpatia*, já

que são elas que possuem os maiores pesos dentro do conjunto⁴⁵. Outras ULs como *chamego*, *fraternidade*, *caridade* e *boa vontade* são bem menos relevantes dentro do conjunto, indicando que, no sentido S1, essas ULs possuem uma relação de sinonímia bem menos saliente com a UL *afeição*. Elas podem, inclusive, nem serem mais consideradas como instâncias da relação de sinonímia – podem ser apenas ULs analógicas. No entanto, se aplicarmos um corte em 0.5 para cada CdSA_sem⁴⁶, essa primeira maneira de consolidar os dados, de motivação semasiológica, produzirá resultados equivalentes aos encontrados em dicionários de sinônimos ou *thesaurus* eletrônicos como o TeP. Com a diferença de que, nesse caso, as ULs serão elencadas conforme sua relevância ao invés da ordem alfabética comumente utilizada. Vejamos, por exemplo, uma comparação entre (i) o CS bruto extraído do verbete **afeição** do *Dicionário Houaiss de Sinônimos e Antônimos* [HOU] e (ii) o CdSA_sem consolidado para a UL *afeição*, com um corte em 0.5.

- (i) {afeto, adoração, afeição, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU]
- (ii) {afeição (1), amor (0.833), afeto (0.833), simpatia (0.667), benquerença (0.5), querença (0.5), ternura (0.5), carinho (0.5), amizade (0.5), apego (0.5)} [CdSA_sem_0.5]_{afeição/S1}

No entanto, como o ambiente digital não apresenta limitações de espaço, não é necessário aplicar o corte em 0.5. A interface gráfica de um dicionário digital como esse pode apresentar primeiramente os resultados mais relevantes e permitir que o usuário role a tela (*scroll*) para baixo para ver mais resultados.

A verdade é que essa maneira de consolidar os dados da matriz pode gerar um produto complementar ao que estamos buscando, mas ainda não produz os resultados que necessitamos para o modelo do DAD. Como dissemos, esses resultados produzem um aperfeiçoamento em produtos similares aos dicionários de sinônimos e, provavelmente, serão utilizados para uma janela com essa função dentro do Dicionário Criativo. Mas para gerar os resultados para o DAD, devemos perpetrar outro tipo de cômputo.

A segunda maneira é consolidar os resultados de cada coluna, gerando conjuntos de sinônimos e analógicos difusos de motivação onomasiológica. Ou seja, partindo de um

⁴⁵ Evidentemente, a UL de entrada será sempre a UL com o maior peso = (1), visto que todo CS/CSA presente em sua célula deverá necessariamente contê-la.

⁴⁶ O que significa excluir do conjunto todas as ULs pontuadas com valores inferiores a 0.5.

significado S, explicitar quais são as ULs que melhor lexicalizam tal conceito. Se, como pretendemos, o primeiro algoritmo (cômputo nas células) já tiver sido rodado, podemos utilizar seus resultados como *input* para o novo algoritmo.

$$\text{coluna S1} = \{\text{CdSA_sem}\}_{\text{UL1/S1}} + \{\text{CdSA_sem}\}_{\text{UL2/S1}} + \{\text{CdSA_sem}\}_{\text{UL(\dots)/S1}} = \{\text{CdSA_ono}\}_{\text{S1}}$$

O CdSA_ono_{S1} arrola os sinônimos e analógicos mais relevantes para o conceito S1. Ou seja, tendo a matriz preenchida, rodamos o algoritmo de consolidação para cada coluna inteira, de forma a extrair um grande novo conjunto, em que a frequência de repetição das ULs entre os CSs e CSAs originais irá determinar o peso de cada uma no interior do CdSA_ono. Essa segunda maneira, de motivação onomasiológica, é que produz os resultados que interessam ser implementados no banco de dados difuso do DAD. Portanto, como temos feito desde o início da tese, sempre que estivermos utilizando a sigla CdSA, estaremos falando do constructo CdSA_ono. Para nos referirmos ao constructo de motivação semasiológica, utilizaremos a forma CdSA_sem.

Vejamos, como exemplo, no quadro 6.8, o consolidado onomasiológico para as duas primeiras categorias conceituais do *anaset* AMOR.

Quadro 6.8. Exemplo da matriz léxico-conceitual do *anaset* AMOR consolidada com CdSAs.

Anaset AMOR	
(S)	CdSAs
(S1) afeição muito intensa de uma pessoa por outra	{afeição (1), amor (0.958), afeto (0.896), simpatia (0.813), ternura (0.667), apego (0.604), carinho (0.563), inclinação (0.5), benquerença (0.458), benevolência (0.458), querença (0.438), dileção (0.396), amizade (0.396), dedicação (0.333), admiração (0.292), estima (0.292), predileção (0.25), preferência (0.229), derretimento (0.229), meiguice (0.229), idílio (0.229), aferro (0.208), constância (0.208), intimidade (0.208), derricho (0.208), conchego (0.208), estremecimento (0.208), idiopatia (0.208), fraternidade (0.208), chamego (0.188), boa vontade (0.188), caridade (0.188), ênfase (0.188), comunhão de sentimentos (0.188), gosto (0.188), desvelo (0.188), zelo (0.188), cuidado (0.188), adoração (0.125), idolatria (0.104), amorosidade (0.104), afeiçoamento (0.083), afetividade (0.083), querença (0.083), apreço (0.083), atração (0.083), amorico (0.083), amorio (0.083), doçura (0.063), brandura (0.063), afetuosidade (0.042), delicadeza (0.042), afinidade (0.042), paixão (0.042), bem-querer (0.042), propensão (0.042), carícia (0.042), amor suave (0.021), interesse (0.021), amor puro (0.021), empatia (0.021), queda (0.021), blandície (0.021), sensibilidade (0.021), blandícia (0.021), igualdade (0.021), confraternidade (0.021), afeição especial (0.021), identidade (0.021), calor (0.021), encantamento (0.021), encanto (0.021), devotamento (0.021), bem-quer (0.021), emoção (0.021), benquerer (0.021), fidelidade (0.021), devoção (0.021), afago (0.021), consideração (0.021), aconchego (0.021), afabilidade (0.021), agarramento (0.021), apegamento (0.021)}
(S2) atração (inclusive sexual) muito intensa entre duas pessoas	{paixão (1), ardor (0.684), fervor (0.526), chama (0.447), adoração (0.421), amor (0.395), atração (0.368), devoção (0.342), calor (0.316), êxtase (0.289), enlevo (0.289), flama (0.263), idolatria (0.237), arroubamento (0.211), paixonite (0.184), desejo (0.158), anelo (0.158), feitiçaria (0.158), enlevamento (0.158), arrebatamento (0.158), anseio (0.158), ansiedade (0.132), apaixonite (0.132), emoção (0.105), fogo (0.105), apego (0.105), entusiasmo (0.079), predileção (0.079), fascínio (0.079), inclinação (0.079), aferro (0.079), exaltação (0.079), atrativo (0.079), encanto (0.079), veneração (0.079), surto (0.053), admiração (0.053), explosão (0.053), agonia (0.053), paroxismo (0.053), afeto (0.053), afeição (0.053), simpatia (0.053), desespero (0.053), cena (0.053), tumulto (0.053), energia (0.053), paixoneta (0.053), acesso (0.053), paixa (0.053), febre (0.053), efervescência (0.053), tempestade (0.053), roubo (0.053), ímpeto (0.053), ebulição (0.053), sedução (0.053), querença (0.026), benquerença (0.026), dedicação (0.026), dileção (0.026), estremecimento (0.026), amor extremo (0.026), culto (0.026), tenção (0.026), preferência (0.026), latria (0.026), chamego (0.026), constância (0.026), feitiço (0.026), hipnotismo (0.026), fascinação (0.026), encantamento (0.026), interior (0.026), capricho (0.026), atratividade (0.026), magia (0.026), magnetismo (0.026), sentimento (0.026), rabicho (0.026), paixão cega (0.026), amor excessivo (0.026), tentação (0.026), amavio (0.026), loucura (0.026), impulso (0.026), parcialidade (0.026), intimidade (0.026), conchego (0.026), ternura (0.026), idiopatia (0.026), idílio (0.026), derricho (0.026), derretimento (0.026), benevolência (0.026), agarramento (0.026), favoritismo (0.026), propensão (0.026), gosto (0.026), cisma (0.026), pendor (0.026), agarração (0.026), interesse (0.026), fantasia (0.026)}

Os resultados do Quadro 6.8 sugerem, por exemplo, que as ULs que melhor lexicalizam o conceito S1, glosado por “afeição muito intensa de uma pessoa por outra”, são *afeição* (1), *amor* (0.95), *afeto* (0.89), *simpatia* (0.81), *ternura* (0.66), *apego* (0.60), *carinho* (0.56), *inclinação* (0.5) etc., nesta ordem. Podemos dizer, nesse caso, que a relação entre as ULs mais relevantes do conjunto é uma relação de sinonímia forte. São, portanto, intercambiáveis em contextos cujo significado pretendido é aquele de S1.

Conforme descemos pela cauda do conjunto, no entanto, as ULs com menores pesos vão deixando de ser boas candidatas para a função de lexicalizar o sentido central da categoria conceitual em questão e passam a funcionar bem como conexão para categorias conceituais relacionadas. Idealmente, quanto mais distante do topo do conjunto, menos a UL funcionará como *sinônima* das ULs melhor ranqueadas, passando a operar como uma unidade *analógica*.

Os CdSAs são normalmente conjuntos bem grandes, com dezenas de ULs, principalmente em *anaset*s cujo domínio inclui muitos conceitos abstratos, como no caso do domínio AMOR. No entanto, os CdSAs resultantes são agora internamente ordenados por relevância. Mais uma vez, poderíamos estabelecer um valor de corte para tornar os CdSAs mais enxutos caso alguma aplicação particular assim o exigisse. Porém, acreditamos que, para os fins a que se destina o DAD no Dicionário Criativo, os CdSAs devem se manter completos. A ideia é permitir que o usuário do DAD possa encontrar as ULs que melhor lexicalizem um conceito ou instanciem uma categoria conceitual. Conforme salientado na Seção 2.1, o problema para o qual um dicionário analógico deve fornecer a solução é o de que “para uma ideia que está dada, encontrar a palavra que a expresse mais adequadamente” (ROBERTSON, 1859, *introduction*). É exatamente o que o usuário do DAD poderá fazer consultando as ULs de maior relevância, exibidas sempre na ordem do valor de seus pesos, do maior para o menor. No entanto, não podemos perder de vista que um dos principais objetivos de um dicionário analógico é, também, permitir novas conexões analógicas entre ideias. Repetindo o que disse Roget (1911) em referência aos desígnios do seu *thesaurus*, “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”.

O usuário do DAD terá acesso às ULs de significado análogo àquele prototípico do conjunto ao descer pela cauda do CdSA. Ou seja, percorrendo as ULs com menor peso dentro do conjunto. Para a categoria conceitual S2, por exemplo, glosada por “atração (inclusive sexual) muito intensa entre duas pessoas”, as ULs de maior peso são *paixão* (1), *ardor* (0.684), *fervor* (0.526), *chama* (0.447). São elas as melhores candidatas para a lexicalização do conceito em questão. No entanto, se escorregarmos pela cauda do CdSA_{S2} até valores abaixo de 0.1, encontraremos as ULs *afeto* (0.053), *afeição* (0.053), *simpatia* (0.053) que, como vimos, são excelentes candidatas a lexicalização de outra

categoria conceitual, análoga a esta. No domínio AMOR, a categoria conceitual S2 é, em termos de análise de traços semânticos, a categoria S1 acrescida do traço de [intensidade]. E, em termos de uma experiência sócio-cultural estereotipada, S2 é um conceito fortemente relacionado com S1. É esperado, pelo menos em algumas fases, que o AMOR ROMÂNTICO (S1) venha acompanhado de uma “atração (inclusive sexual) muito intensa entre duas pessoas” (S2), com atestações como *seu coração estava em fogo, a febre do amor, o mais ardente calor do nosso amor etc.*

6.2 DOMÍNIOS LÉXICO-CONCEITUAIS MODELADOS: CORPO HUMANO, AMOR e TEATRO

Os domínios léxico-conceituais CORPO HUMANO, AMOR e TEATRO são os domínios que foram modelados como prova de conceito. Eles foram escolhidos por gerarem diferentes configurações no interior dos *anaset*s. O domínio CORPO HUMANO, como veremos em 7.2.2, pertence à Grande Categoria MUNDO FÍSICO, e talvez seja um dos exemplos mais emblemáticos de um domínio estruturado pela relação léxico-conceitual da meronímia. E, como é comum acontecer com domínios do MUNDO FÍSICO, ele é um domínio com categorias conceituais bem delimitadas por categorias adjacentes. Ao contrário, domínios do MUNDO PSÍQUICO, como o domínio léxico-conceitual AMOR, costumam possuir categorias conceituais com limites muito vagos, que compartilham muitas ULs em comum. O domínio AMOR é também um exemplo de domínio em que muitas categorias conceituais se relacionam paradigmaticamente. Já o domínio TEATRO, prototipicamente incluído na Grande Categoria MUNDO SOCIAL, como veremos em 7.3.2, é um exemplo de domínio em que a maioria das categorias conceituais se relacionam sintagmaticamente. Como no caso do domínio CORPO HUMANO, grande parte da estrutura conceitual do domínio TEATRO é percebida como um todo dividido em partes. Com a diferença de que essas partes são unidas por um tecido cultural e socialmente construído. Ou seja, não é possível apreender o domínio TEATRO sem ter como enquadramento uma determinada contextura social e cultural complexa.

Ao mesmo tempo, esses três domínios foram escolhidos por estarem entre os mais recorrentes na produção literária. O domínio AMOR é o mais trabalhado em poemas, canções e romances. O CORPO HUMANO também é amplamente utilizado, visto que quase qualquer tema evoca uma instância corporal – o tema amoroso, por exemplo,

constantemente alude a partes do corpo para instanciar um sentimento ou sensação abstrata, como o *olhos nos olhos*, o *beijo na boca*, o *coração palpitante*, a *face rubra* etc. Por fim, por possuir muitas vezes uma vocação metalinguística, a arte refere-se a si própria e o domínio ARTES (de onde recortamos o domínio TEATRO) é também bastante presente em canções, poemas e outras composições literárias. Um bom exemplo de vocação metalinguística, e desses três domínios sendo evocados com frequência, é o conjunto de canções composto para o espetáculo *Grande Circo Místico*, cujas letras constituem o *corpus* de pesquisa desta tese, cf. Apêndice I.

6.3 SÍNTESE DA SEÇÃO

Ao presente Seção teve por objetivo propor um modelo de representação e uma metodologia de trabalho que, à luz das teorias apresentadas nas seções anteriores, servirão, ao longo da próxima seção, para a construção do modelo linguístico-computacional do DAD. Dessa forma, o modelo de representação, batizado de *anaset*, foi definido como sendo uma estrutura léxico-conceitual que organiza categorias conceituais lexicalizadas no interior de um mesmo domínio conceitual valendo-se, para tanto, do constructo CdSA, anteriormente definido.

A metodologia de trabalho proposta ao longo da Seção sugeriu um percurso dividido em quatro etapas: i) identificação e coleta das ULs; ii) identificação e delimitação das categorias conceituais pertinentes ao domínio em questão; iii) identificação, coleta e arranjo adequado dos conjuntos de sinônimos e analógicos no interior da matriz; iv) consolidação dos resultados.

Em conjunto com a proposição do modelo de representação e da metodologia de trabalho, foram também explicitados, nessa seção, os domínios léxico-conceituais que serão modelados como exemplo: AMOR, CORPO HUMANO e TEATRO.

SEÇÃO 7 – CONSTRUÇÃO E IMPLEMENTAÇÃO DE UM MODELO PARA O DAD

Na Seção anterior, apresentamos o constructo *anaset* como um modelo de representação linguístico-computacional capaz de estruturar toda informação léxico-conceitual necessária para a construção do DAD. O objetivo da presente Seção é relatar um exercício de aplicação do modelo, aplicando cada passo do processo de construção e implementação do *anaset*, utilizando-se para isso de dados reais. Faz também parte do escopo dessa Seção final delinear como será a integração do DAD com o Dicionário Criativo e apresentar alguns esboços de interface gráfica e de usabilidade do sistema a serem implementados no produto final.

A Seção está dividida em cinco partes. As três primeiras relatam a aplicação do modelo para os três domínios léxico-conceituais citados em 6.2, começando pelo domínio AMOR (7.1), depois o domínio CORPO HUMANO (7.2) e, por fim, o domínio TEATRO (7.3). A quarta parte trata da integração do DAD com o Dicionário Criativo e as questões de interface e usabilidade (7.4). Por fim, a última parte discorre, resumidamente, sobre a etapa computacional que está sendo desenvolvida, paralelamente ao escopo desta tese, para implementar e disponibilizar o DAD em ambiente *Web* (7.5).

Todos os dados coletados foram digitalizados e copiados para o programa *Microsoft Excel™*, onde foram armazenados e executados os cálculos necessários. Ao longo dessa seção, parte dos dados foram copiados do *Excel™* e apresentados em tabelas apropriadas (do programa *Microsoft Word™*). Isso porque o volume dos dados brutos é muito extenso para ser integralmente replicado em tabelas do *Word™* e serem exibidos no meio do texto da tese. Dessa forma, para cada exemplar da tese, há um CD que contém todos os dados brutos armazenados em *Excel™* e um arquivo .txt com a descrição dos algoritmos utilizados para os cálculos e ordenamento dos pesos.

7.1 MODELAGEM DO ANASET AMOR

Essa Seção relata a aplicação do modelo no domínio léxico-conceitual AMOR. O percurso apresentado em 6.1.2 é seguido rigorosamente para se chegar à modelagem final do *anaset* AMOR. Para tanto, foram organizadas subseções para cada uma das etapas. Em 7.1.1 temos a identificação e coleta das ULs do domínio AMOR. Em 7.1.2

procedemos com a estruturação conceitual do *anaset* AMOR. Em 7.1.3 executamos a identificação, coleta e arranjo dos CSs e CSAs pertinentes ao *anaset* AMOR. Por fim, em 7.1.4 temos a consolidação dos resultados, em que apresentamos uma tabela que apresenta a modelagem final do *anaset* AMOR.

7.1.1 ETAPA 1 - IDENTIFICAÇÃO E COLETA DAS ULs DO DOMÍNIO AMOR

Conforme 6.1.2.1, o ponto de partida é gerar a lista com as ULs pertencentes ao domínio léxico-conceitual AMOR. Para gerar a lista das ULs pertencentes ao domínio AMOR partimos da lista de ULs extraídas do *corpus* (cf. Apêndice I). As seis ULs do domínio AMOR extraídas do *corpus* estão elencadas no Quadro 7.1.

Quadro 7.1. ULs do domínio AMOR extraídas do *corpus*

Domínio léxico-conceitual AMOR (6 ULs)
amante
beijo
bem
namorado
noiva
romance

Em seguida, foram consultadas as fontes de informação léxico-conceituais arroladas no Apêndice II.1 para identificar ULs sinônimas e análogas que possam também pertencer ao mesmo domínio léxico-conceitual. Dessa forma, procedeu-se à expansão do número de ULs que deverá ocupar a coluna da extrema esquerda na matriz léxico-conceitual do *anaset* correspondente ao domínio léxico-conceitual do AMOR.

Conforme os critérios apontados em 6.1.2.1, não foram incluídos na lista as ULs complexas, estrangeirismos e nomes próprios, pois estes não costumam possuir entradas próprias em dicionários e, de fato, não possuem em nenhuma das obras selecionadas como fonte de informação. Como exemplo das ULs que **não** foram incluídas na lista de ULs de entrada do domínio léxico-conceitual AMOR estão: *chama inédita*, *história de amor*, *laços de amor*, *fiancée*, *Afrodite*, *Vênus*, *Don-João*, *Casanova* etc. Na prática, todas as ULs que não possuem entrada em nenhuma das fontes de informação lexical foram automaticamente excluídas. É o caso também de alguns arcaísmos como *amorio* e *marrancho*. Porém, como alguns desses arcaísmos estão contemplados como

entradas em algumas das fontes de informação lexical – e considerando que para um dicionário analógico esta riqueza de ULs é importante –, apenas ser um arcaísmo não configura um critério para a exclusão da lista. O fato é que, pela metodologia e os procedimentos definidos para a construção do *anaset*, as ULs que não constam como entrada nas obras de referência devem ser automaticamente excluídas da lista. Em trabalhos futuros, caso se identifique a importância de se incluir essas ULs de volta na lista, deverá se encontrar heurísticas que suplantem as incompletudes do modelo aqui proposto.

Ainda com relação aos critérios apontados em 6.1.2.1, limitamos a expansão da coleta das ULs do domínio léxico-conceitual AMOR principalmente pelos domínios léxico-conceituais adjacentes: CASAMENTO, AMIZADE, DESEJO, CARÍCIAS e BENEVOLÊNCIA. Caso contrário, a coleta de ULs por meio de sinônimos e analógicos poderia progredir indefinidamente. Para sopesar o limite inicial de cada domínio conceitual utilizamos as mil categorias de Roget, replicadas no Dicionário Analógico do professor Azevedo (AZEVEDO, [1950] 2010). É por esse motivo que não foram incluídas na lista, por exemplo, as ULs *marido/ esposa* e *amigo/ amiga*, pois elas estão prototipicamente ligadas aos domínios léxico-conceituais CASAMENTO e AMIZADE, respectivamente.

No entanto, é importante frisar que as ULs não incluídas nessa lista inicial – seja por qualquer um dos critérios acima adotados – poderão aparecer, e muitas de fato aparecem, nos resultados finais, dentro dos CdSAs. Ou seja, elas não figuram como ULs de entrada no *anaset* AMOR, mas permanecem como ULs relacionadas, sem ocupar uma posição de alta relevância dentro dos conjuntos.

Seguindo a metodologia e os critérios estabelecidos em 6.1.2.1, chegamos a uma lista de 69 ULs de entrada para o *anaset* AMOR, conforme o quadro 7.2.

Quadro 7.2. ULs de entrada do *anaset* AMOR

Anaset AMOR: 69 ULs de entrada			
abraço	apego	dileção	namorisco
admiração	ardor	fervor	namoro
admirador	atração	flerte	noiva
adoração	beijo	frecheiro	noivo
adorador	beijocador	galã	paixão
afeição	beijoqueiro	galanice	predileção
aferro	bem	galanteador	predileta
afeto	benevolência	galanteio	predileto
agarramento	benquerença	idílio	pretendente
amada	calor	idolatria	pretendida
amado	carinho	inclinação	querença

amante	caro	intimidade	querida
amásia	conquistador	namorada	querido
amásio	corteador	namoradeiro	simpatia
amor	dedicação	namorado	ternura
amorico	derretimento	namorador	
apaixonada	derricho	namoramento	
apaixonado	devoção	namorico	

7.1.2 ETAPA 2 - ESTRUTURAÇÃO CONCEITUAL DO ANASET AMOR

A estruturação conceitual do *anaset* AMOR começa por definir em qual das grandes categorias o domínio se insere. E, para tanto, vamos recorrer ao que foi discutido na Seção 4, voltando a comparar os sistemas conceituais lá citados, só que dessa vez com o foco no domínio AMOR. A comparação não tem o objetivo de especular sobre quais dos sistemas conceituais é o mais filosoficamente correto, o que seria virtude de um trabalho do campo da filosofia dar cumprimento. Todos possuem méritos dentro do contexto de aplicação de cada um. O objetivo é demonstrar uma convergência entre eles e apenas optar pelo modelo que nos parece mais adequado para o contexto de um dicionário analógico digital que integra parte fundamental de um dicionário voltado para a criatividade.

No quadro sinóptico de categorias elaborado por Roget o domínio AMOR se encontra na ponta da seguinte ramificação: **classe VI. Afeições> divisão III. Afeições Simpáticas> 1º) Sociais> 897. Amor**. Exatamente a mesma estrutura conceitual que encontramos no *Dictionnaire Idéologique* (ROBERTSON, 1859) e no Dicionário Analógico da língua portuguesa (AZEVEDO, [1950] 2010). Na WN.Pr, o conceito prototípico AMOR (LOVE⁴⁷), definido pela glosa “uma forte emoção positiva de respeito e carinho”, é hipônimo direto do conceito EMOÇÃO (EMOTION⁴⁸) e tem perto do topo da sua hierarquia conceitual o conceito ABSTRAÇÃO (ABSTRACTION⁴⁹). Na EWN, o conceito AMOR é categorizado também como uma emoção (EMOTION) que, por sua vez, é entendida como uma experiência mental (EXPERIENCE/MENTAL). No caso da WN.Pr, acima da categoria conceitual ABSTRAÇÃO, têm-se ainda, no topo da hierarquia, a categoria

⁴⁷ {07558676} <noun.feeling>S: (n) love (a strong positive emotion of regard and affection)

⁴⁸ {07495208} <noun.feeling>S: (n) emotion (any strong feeling)

⁴⁹ {00002137} <noun.Tops>S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples).

conceitual ENTIDADE (ENTITY). E, no caso da EWN, o topo da hierarquia do conceito AMOR termina na divisão ENTIDADE DE 2ª ORDEM (2ND ORDER ENTITY).

No HTOED, o conceito AMOR (LOVE) é hipônimo direto do conceito EMOÇÃO (EMOTION) que, por sua vez, alinha-se diretamente abaixo da Grande Categoria MUNDO MENTAL (THE MENTAL WORLD), como se vê ilustrado abaixo:

02 The mental world

02.02 Emotion

02.02.22 Love

Podemos perceber que a hierarquia do HTOED é uma estrutura conceitual mais simplificada. E, como dissemos ao longo da Seção 4, é com esse tipo de estrutura simplificada que o nosso modelo melhor se ajusta. Uma estrutura conceitual simples possui a vantagem de tornar o modelo mais fluido e adaptável e a desvantagem de torná-lo menos preciso para alguns tipos de aplicação computacional. Sendo assim, para os fins da construção de um dicionário analógico digital que estimule a criatividade, entendemos que a fluidez e a adaptabilidade são mais benéficos para o modelo do que um sistema conceitual rigidamente hierarquizado.

É difícil comparar sistemas conceituais diferentes, tanto em seus propósitos como na estrutura em si. No caso dos dicionários analógicos e do HTOED, temos uma arquitetura conceitual que serve apenas de ancoragem para as entradas que, na maioria das vezes, constituem domínios conceituais completos. No modelo das *wordnets*, não podemos falar em domínios conceituais. O que temos são conceitos instanciados em *synsets* que se conectam por meio de algumas poucas relações semânticas. Nas *wordnets* não há um domínio conceitual AMOR no qual podemos, por exemplo, incluir o conceito de NAMORADO (BOYFRIEND). Por esse motivo, utilizamos para fins de comparação o conceito prototípico AMOR. Ainda assim a comparação é válida, pois os domínios conceituais são construímos sempre em torno de um protótipo (cf. 5.2). Para um dicionário de analogias não é relevante o fato de que o conceito NAMORADO não seja uma entidade abstrata, por exemplo. O fato de que o conceito seja indispensável para pelo menos um tipo de RELAÇÃO AMOROSA e de que RELAÇÃO AMOROSA seja um conceito diretamente relacionado com o conceito prototípico de AMOR é o que se prova relevante. Evidente que poderá haver no modelo, futuramente, uma entrada em que o conceito NAMORADO apareça em relação paradigmática a outros papéis sociais desempenhados por sujeitos e, dessa forma, tal entrada estará incluída na Grande

Categoria MUNDO SOCIAL. Esse é um dos motivos do porque nossa discussão na Seção 3 sobre relações paradigmáticas e sintagmáticas é tão importante. No entanto, como dissemos, estamos categorizando o domínio conceitual AMOR a partir de seu conceito prototípico: “sentimento de afeição muito intensa de uma pessoa por outra”. Tomando como referência as estruturas conceituais antes mencionadas, principalmente a do HTOED, identificamos que o domínio conceitual AMOR deve ser representado dentro da Grande Categoria MUNDO PSÍQUICO, conforme Quadro 7.3.

Quadro 7.3. Matriz preenchida com a Grande Categoria onde se inclui o *anaset* AMOR: MUNDO PSÍQUICO

MUNDO PSÍQUICO						
AMOR	Relações Paradigmáticas			Relações Sintagmáticas		
	S1	S2	S3	EF1	EF2	
				S4	S5	S6
UL 1	{cs1} [X] {cs2} [Y] {cs3} [Z]					

Em seguida, nosso objetivo passa a ser identificar quais são as categorias conceituais (S) que estruturam o domínio AMOR. Por sua estrutura onomasiológica, os dicionários analógicos seriam ferramentas apropriadas para a delimitação dos conceitos. No entanto, conforme 2.2, os dicionários analógicos não apresentam uma metodologia clara para a divisão dos conceitos no interior dos grupos de palavras. E, quando o fazem, não explicitam por meio de uma glosa qual é o conceito que está ali sendo lexicalizado em sinônimos e analógicos. No entanto, procurando pelo item lexical *amor* nas entradas dos dicionários semasiológicos, encontramos uma correspondência entre as diferentes acepções da palavra e os diferentes conjuntos de sinônimos e analógicos mais ou menos divisados no interior dos grupos de palavras dos dicionários analógicos.

Ao consultar por *amor* nas fontes de informação léxico-conceitual arroladas no Apêndice II.2, especificamente os itens a-d (Dicionários monolíngues da Língua Portuguesa), identificamos dez diferentes sentidos. Vejamos:

[AUR-S] **1** Sentimento que predispõe alguém a desejar o bem de outrem. **2** Sentimento de dedicação absoluta de um ser a outro, ou a uma coisa. **3** Inclinação ditada por laços de família. **4** Inclinação sexual forte por outra pessoa. **5** Afeição, amizade, simpatia. **6** Objeto do amor

[AUL-S] **1** Sentimento que faz alguém querer o bem de outrem ou de alguma coisa. **2** Afeto profundo, devoção de uma pessoa a outra. **3** Sentimento terno e caloroso de uma pessoa por outra, inclusive de natureza física e sexual. **4** Relação amorosa. **5** O ato sexual. **6** Inclinação, apego ao que desperta prazer ou empatia. **7** Sentimento de devoção a Deus; VENERAÇÃO. **8** O ente objeto do amor. **9** Cuidado, zelo, dedicação.

[MIC-S] **1** Sentimento que impele as pessoas para o que se lhes afigura belo, digno ou grandioso. **2** Grande afeição de uma a outra pessoa de sexo contrário. **3** Afeição, grande amizade, ligação espiritual. **4** Objeto dessa afeição. **5** Benevolência, carinho, simpatia. **6** Tendência ou instinto que aproxima os animais para a reprodução. **7** Desejo sexual. **8** Ambição, cobiça. **9** Culto, veneração. **10** Caridade.

[HOU-S] **1** forte afeição por outra pessoa, nascida de laços de consanguinidade ou de relações sociais. **2** atração baseada no desejo sexual. **3** (*por extensão de sentido*) relação amorosa; caso, namoro. **4** (*por extensão de sentido*) atração sexual natural entre espécies animais. **5** afeição baseada em admiração, benevolência ou interesses comuns; forte amizade. **6** (*derivação por metonímia*) a pessoa ou a coisa amada. **7** devoção, adoração. **8** (*sentido figurado*) devoção de uma pessoa ou um grupo de pessoas por um ideal concreto ou abstrato. **9** (*derivação por metonímia*) o objeto de tal interesse ou veneração. **10** demonstração de zelo, de dedicação.

Podemos aplicar a filtragem abaixo para chegarmos a seis sentidos de ordem paradigmática e quatro sentidos de ordem sintagmática:

S1: [AUR-S] (1) = [AUL-S] (1+2) = [MIC-S] (2) = [HOU-S] (1) = afeição muito intensa de uma pessoa por outra

S2: [AUR-S] (4) = [AUL-S] (3) = [MIC-S] (6+7) = [HOU-S] (2) = atração (inclusive sexual) muito intensa entre duas pessoas

S3: [AUR-S] (5) = [AUL-S] (2) = [MIC-S] (3) = [HOU-S] (5) = amizade intensa entre duas pessoas

S4: [AUR-S] (2) = [AUL-S] (7) = [MIC-S] (1) = [HOU-S] (7+8) = devoção a um ideal concreto ou abstrato

S5: [AUL-S] (9) = [MIC-S] (9) = [HOU-S] (10) = demonstração de zelo

S6: [AUL-S] (4) = [MIC-S] (3) = relação amorosa

S7: [AUR-S] (6) = [AUL-S] (8) = [MIC-S] (4) = [HOU-S] (10) = homem que ama ou é amado

*S8: [AUR-S] (6) = [AUL-S] (8) = [MIC-S] (4) = [HOU-S] (10) = mulher que ama ou é amada

S9: [AUL-S] (4) = [MIC-S] (3) = homem que corteja o amor de uma mulher

S10: [AUL-S] (5) = manifestação física do amor

O conceito S1 é o sentido prototípico do domínio. É esse sentido que ajuda a explicar, nos moldes de uma rede radial (cf. 5.2.3), as interpretações ligadas aos outros

sentidos. O conceito é lexicalizado por ULs como *amor*, *afeto* e *carinho*. Os conceitos de S2 e S3 são sentidos bifurcados em relação paradigmática com o protótipo. É uma relação de especialização em que S2 denota um sentimento mais intenso e potencialmente de interesse sexual. Ao contrário de S3, que denota um afeto um pouco mais brando, entre pessoas que se admiram e que não almejam expressar esse amor fisicamente. *Paixão*, *ardor* e *fervor* são ULs frequentemente associadas ao conceito S2, ao passo que S3 é mais comumente lexicalizado pelas ULs *simpatia*, *admiração*, *amizade* e *afeição*. Os conceitos S4 e S5 parecem derivar respectivamente dos conceitos S2 e S3, com a diferença de que o sentimento nesse caso se dirige a objetos, eventos ou ideais abstratos. O conceito S4 está fortemente ligado, como veremos adiante, às ULs *devoção*, *adoração*, *veneração*, onde há uma paixão intensa e virtualmente cega a um ideal concreto ou abstrato. Exemplos de uso poderiam ser os sintagmas nominais “amor a Deus” ou “amor à Pátria”. Já o sentido denotado por S5 evoca um sentimento de zelo ou cuidado para com algo que seja importante como, por exemplo, o “amor ao trabalho”. Tanto em S4 quanto em S5, o contexto evocado é do sujeito que direciona o seu amor para um objeto inanimado, o que os afasta do conceito prototípico de AMOR, que evoca o *frame* *Relação_amorosa*, em que os elementos centrais são sujeitos que se relacionam por meio do sentimento amoroso.

O *frame* *Relação_amorosa* estrutura um fato social complexo conceitualmente expresso por S6 e que é lexicalizado por ULs como *namoro*, *romance*, *flerte* e *caso*. Ou seja, o conceito S6 está em relação paradigmática ao sentido S1, mas é exatamente ele que possui uma estruturação interna, na forma de *frame*. A partir desse ponto, estamos no terreno dos conceitos que se relacionam sintagmaticamente dentro do domínio conceitual AMOR, enquadrados pelo *frame* *Relação_amorosa*. Os conceitos S7 e S8 denotam, respectivamente, HOMEM QUE AMA OU É AMADO e MULHER QUE AMA OU É AMADA. Em outros domínios conceituais, a divisão de uma mesma categoria conceitual pela questão do gênero não seria necessária ou mesmo recomendada. No entanto, a estruturação conceitual do domínio AMOR e o seu reflexo na língua exigiram do nosso modelo essa subdivisão. Diferentemente de outros domínios, a questão do gênero é relevante para o domínio conceitual AMOR e do *frame* prototipicamente evocado por ele. A língua portuguesa, assim como diversas outras línguas, possuem ULs específicas para os dois gêneros dos sujeitos de uma relação amorosa. Algumas ULs como *amante* e *xodó*, por exemplo, servem para os dois gêneros e outras como *namorado (a)* e *amásio (a)*

resolvem a questão do gênero no nível sintático. Porém, ULs como *dulcineia*, *concupina* e *janeleira*, por exemplo, lexicalizam apenas o conceito MULHER QUE AMA OU É AMADA, e nunca o seu equivalente masculino.

O conceito S9 também é marcado pelo gênero, com a diferença de não possuir um equivalente do gênero oposto. O sentido denotado por S9 é o de HOMEM QUE CORTEJA UMA MULHER e, portanto, é um conceito específico do gênero masculino. O fato de não existir um equivalente feminino para o PB, é um aspecto interessante de como a cultura e os padrões sociais se refletem na estrutura léxico-conceitual de uma língua (cf. 4.6). Apesar de não ser inconcebível a imagem de uma mulher cortejando um homem, esse comportamento não é esperado ao ponto de ser lexicalizado. Apenas o conceito do sujeito que possui um comportamento estereotipado dentro do *frame* social *Relação_amorosa* é que possui ULs para instanciá-lo, tais como *corteador*, *galanteador* e *mulherengo*.

Para construir o *frame* *Relação_amorosa* tomamos como inspiração o *frame* *Personal_relationship* da FrameNet, já que esta não possui um *frame* específico *Loving_relationship*. Os EFs centrais do *frame* *Personal_relationship* são exatamente os parceiros (*partners*) que se relacionam entre si. No entanto, como foi dito em 3.3, cada *frame* especifica os EFs que dele participam. Dessa forma, assim como os lexemas *vender* e *comprar*, no *frame* *Comércio_de_mercadorias*, têm como seus EFs centrais, respectivamente, as funções ‘vendedor’ e ‘comprador’, os lexemas que instanciam S7 e S8 do nosso específico *frame* *Relação_amorosa* terão como EF central a função ‘amante’ (EF1). Já o conceito S9 ancora-se no EF não central ‘pretendente’ (EF2).

Como dissemos em 3.3, o que nos interessa utilizar do modelo da FrameNet é a sua coletânea de estruturas conceituais culturalmente definidas, que atua na organização de segmentos da realidade e que, portanto, nos auxilia na estruturação de categorias conceituais que se relacionam por meio de enquadramentos.

Apenas o dicionário AUL-S tem uma acepção para o conceito do “ato sexual”, apesar do sintagma verbal “fazer amor” ser frequente nas atestações em *corpus*. Por outro lado, estava faltando uma categoria conceitual para acomodar ULs como *transa*, *sexo*, *cópula*, *beijo*, *abraço*, *carícia* etc. que aparecem arroladas em conjuntos claramente definidos nos grupos de palavras dos dicionários analógicos no interior do domínio AMOR. Olhando para o *frame* *Personal_relationship*, encontramos o EF MANNER,

que sugere a maneira pela qual os parceiros se relacionam. No caso do nosso *frame* *Relação_amorosa*, uma das maneiras pelas quais os amantes se relacionam é por meio da MANIFESTAÇÃO FÍSICA DO AMOR (S10), conceito lexicalizado por ULs como *sexo*, *carinho*, *cafuné* e *carícia*, cujo EF pode ser rotulado como ‘maneira de amar’ (EF3).

O Quadro 7.4 apresenta a matriz léxico-conceitual do *anaset* AMOR preenchida com a estrutura conceitual discutida ao longo dessa seção.

Quadro 7.4. Matriz do *anaset* AMOR preenchida com a estrutura conceitual

MUNDO PSÍQUICO										
AMOR	Relações Paradigmáticas						Relações Sintagmáticas			
	S1	S2	S3	S4	S5	S6	EF1		EF2	EF3
							S7	S8	S9	S10
UL 1										

O Quadro 7.4 complementa-se com as informações presentes no Quadro 7.5, em que as categorias conceituais estão instanciadas por uma pequena glosa e por CSs/CSAs representativos. Além disso, a última coluna apresenta abonações extraídas do *corpus* do NILC ou diretamente retiradas das fontes de informação lexical.

Quadro 7.5. Estrutura conceitual do *anaset* AMOR

Estrutura conceitual do <i>anaset</i> AMOR			
(S)	Glosa	CSs/CSAs	Exemplo
Relação Paradigmática			
(S1)	Afeição muito intensa de uma pessoa por outra	{amor, carinho, idolatria, afeto, amorico, amorosidade, amorio, inclinação} [AZE] {love} [WN]	“Ligado ao amor , constitui-se numa das mais gratificantes formas de inter-relacionamento pessoal, embora o amor se baseie mais em fatores mentais e emocionais do que em pura sexualidade” [NILC]
(S2)	Atração (inclusive sexual) muito intensa entre duas pessoas	{amor, atração, desejo, fascínio, interesse, paixão} [HOU] {love, sexual love, erotic love} [WN]	“Nesse ciberlocal, o objetivo é ajudar os apaixonados a manter a chama do amor acesa” [NILC]
(S3)	Amizade intensa entre duas pessoas	{amor, afeição, afeiçoamento, afetividade, afeto, amizade, querença, querência} [TEP] {closeness, love, regard, affection, intimacy, fondness, companionship, comradeship} [RIT]	“ amor pelos antigos colegas” [HOU-S]
(S4)	Veneração de, adoração de, devoção a	{amor, adoração, culto, devoção, tenção, veneração} [TEP]	“Com este espírito, está o professor Darcy Ribeiro, senador

	um ideal concreto ou abstrato	{adoration, idolization, idolisation} [WN]	pelo PDT do Rio de Janeiro, brasileiro adorável que tem pela educação um amor desbragado” [NILC]
(S5)	Demonstração de zelo	{amor, aplicação, atenção, carinho, cuidado, dedicação, desvelo, diligência, vigilância, vigília, zelo} [TEP]	“ amor ao trabalho” [HOU-S]
(S6)	Relação amorosa	{romance, caso, namoro} [TEP] {love affair, romance} [WN]	“A nossa relação era um caso de amor ” [NILC]
Relação Sintagmática			
(S7)	Homem que ama ou é amado	{namorado, camote, derriço, pequeno, xodó} [TEP] {beloved, dear, dearest, honey, love} [WN]	“Outras têm vontade de trazer o marido, o namorado ” [NILC]
(S8)	Mulher que ama ou é amada	{namorada, amante, amor, conversada, derriço, namoro, pequena} [HOU] {beloved, dear, dearest, honey, love} [WN]	“Dalila foi o amor de Sansão” [AUL-S]
(S9)	Homem que corteja uma mulher	{namorado, enamorado, galã, conquistador, admirador, apaixonado, pretendente, noivo, adorador, fã, amante} [IDS] {lover, admirer, adorer, amorist, infatuate, paramour, suitor, wooer, pursuer, follower} [RIT]	“As novelas exaltam o conquistador , o mulherengo” [NILC]
(S10)	Manifestação física do amor	{carinho, abafo, acalentamento, acalento, acarinhamento, afago, agrado, cafuné, carícia, mimo, quindim, quitute, tagaté} [TEP] {caress, fondle} [WN]	“Estou precisando de um carinho feminino” [NILC]

7.1.3 ETAPA 3 - IDENTIFICAÇÃO, COLETA E ARRANJO DOS CSS E CSAS DO ANASET AMOR

As células da matriz léxico-conceitual serão preenchidas com CSs e CSAs identificados e coletados nas fontes de informação léxico-conceitual apresentadas no Apêndice II.1, itens **a-f**. Para cada UL de entrada da matriz, busca-se os CSs e CSAs que a contenham e que, ao mesmo tempo, instanciem uma, ou mais de uma, das categorias conceituais previamente definidas e separadas na etapa anterior (Etapa 2).

Os CSs e CSAs coletados são apresentados em seis quadros, um para cada fonte de informação diferente. O cabeçalho de cada quadro identifica a fonte de informação de

onde os CSs e CSAs foram extraídos, seguindo a mesma ordem em que elas são apresentadas no Apêndice II.1.

a) HOUAISS, A. <i>Dicionário Houaiss de Sinônimos e Antônimos</i> [HOU]
{adorador, fã},
{beijo, ósculo},
{caro, dileto},
{agarramento, agarra, agarração},
{agarramento, intimidade, ligação},
{favorito, predileto, preferido},
{idílio, romance, namoro},
{predileta, favorita, preferida},
{querida, admirada, apreciada},
{querido, admirado, apreciado},
{abraço, amplexo, abraçamento, enlaçamento},
{amada, amante, namorada, querida},
{amor, amizade, afinidade, fraternidade},
{apaixonado, amante, enamorado, namorado},
{bem, querida, amada, amor},
{bem, querido, amado, amor},
{derretimento, amor, atração, encantamento},
{derriço, namorada, parceira, pequena},
{derriço, namorado, parceiro, pequeno},
{dileção, afeto, predileção, preferência},
{idolatria, devoção, fervor, loucura},
{pretendente, candidato, aspirante, concorrente},
{amada, amante, amor, bem-amada, bem-querer},
{amor, dedicação, cuidado, fidelidade, zelo},
{carinho, cuidado, capricho, desvelo, esmero},
{derretimento, requebro, dengo, denguiço, trejeito},
{derriço, namoro, corte, galanteio, rabicho},
{devoção, adoração, afeição, dedicação, veneração},
{esposo, companheiro, côjuge, consorte, marido},
{galanteador, galante, namorador, cortejador, requestador},
{galã, namorador, cortejador, galanteador, requestador},
{adoração, amor, idolatria, latria, paixão, veneração},
{amor, atração, desejo, fascínio, interesse, paixão},
{amor, devoção, adoração, entusiasmo, respeito, veneração},
{amor, relacionamento, aventura, caso, namoro, romance},
{ardor, paixão, amor, desejo, exaltação, fervor},
{benevolência, afeto, afabilidade, benquerença, estima, simpatia},
{galanteio, namoro, amor, chamego, embeijamento, flerte},
{inclinação, atração, apego, interesse, queda, simpatia},
{predileção, escolha, inclinação, opção, preferência, propensão},

{namorada, amante, amor, conversada, derriço, namoro, pequena},
{admirador, aficionado, amante, apaixonado, entusiasta, fanático, louco, maníaco},
{adoração, louvor, culto, idolatria, latria, respeito, reverência, veneração},
{amante, apaixonado, enamorado, namorado, amásio, camarada, companheiro, concubino},
{atração, afinidade, afeição, afeto, amor, inclinação, simpatia, vínculo},
{namorador, cativador, cortejador, femeeiro, galanteador, mulherengo, namorado, requestador},
{noiva, desposada, esposada, esposa, nubente, pretendida, prometida, recém-casada},
{dedicação, apreço, adoração, apego, atenção, consideração, desvelo, paixão, zelo},
{galantaria, galanteria, galanice, delicadeza, donaire, elegância, garbo, graça, primor},
{afeto, afeição, amizade, amor, apego, afinidade, devotamento, estima, encanto, fidelidade},
{amorzinho, amorico, amorio, aventura, caso, linhada, namorico, namorisco, namorilho, romance},
{atração, sedução, atratividade, encanto, encantamento, fascinação, feitiço, hipnotismo, magia, magnetismo},
{concubina, amante, amásia, barregã, caso, dama, fêmea, manceba, mulher, súcuba},
{noivo, desposado, esposado, esposo, néogamo, nóbio, nubente, pretendido, prometido, recém-casado},
{ternura, carícia, afago, agrado, blandícia, carinho, festa, festinha, meiguice, mimo},
{arroubo, arroubamento, abarretamento, embevecimento, encantamento, encanto, enleio, enlevo, enlevação, entusiasmo, êxtase},

b) POLITO, AG. <i>Michaelis Dicionário de Sinônimos e Antônimos</i> ; [MIC]	
{abraço, amplexo},	
{afeição, amizade},	
{beijo, ósculo},	
{beijoqueiro, beijador},	
{derriço, namorada},	
{derriço, namorado},	
{querença, querer},	
{adorador, admirador, fã},	
{afeição, afeto, amor},	
{agarramento, agarra, agarração},	
{amante, amásio, amigo},	
{conquistador, mulherengo, sedutor},	
{derriço, namoro, xodó},	
{galã, namorado, galanteador},	
{galã, namorador, galanteador},	
{idílio, amor, romance},	
{noivo, nubente, esposo},	
{querida, cara, estimada},	
{admiração, estima, afeição, apreço},	
{adoração, paixão, amor, admiração},	
{amada, namorada, bem, tesouro},	
{amada, querida, cara, estimada},	
{amante, amásia, concubina, amiga},	
{amor, atenção, diligência, cuidado},	
{amor, estima, simpatia, amizade},	
{amor, paixão, atração, desejo},	

{apaixonada, enamorada, caída, gamada},
{apaixonado, enamorado, caído, gamado},
{apego, afeto, devoção, agarramento},
{ardor, paixão, entusiasmo, arrebatamento},
{ardor, paixão, entusiasmo, energia},
{atração, atrativo, fascínio, encanto},
{atração, simpatia, afeição, admiração},
{bem, amada, querida, amor},
{bem, amado, querido, amor},
{benevolência, estima, afeto, simpatia},
{carinho, cuidado, atenção, desvelo},
{caro, querido, estimado, prezado},
{culto, idolatria, veneração, adoração},
{derretimento, dengo, afetação, requebro},
{devoção, dedicação, veneração, adoração},
{favorito, preferido, predileto, querido},
{fervor, devoção, piedade, religiosidade},
{fervor, entusiasmo, energia, paixão},
{flerte, namorico, caso, aventura},
{galanteio, namoro, xodó, chamego},
{galanteio, corte, galanteria, lisonja},
{idolatria, adoração, veneração, paixão},

c) FERNANDES, F. *Dicionário de Sinônimos e Antônimos da Língua Portuguesa*; [FER]

{amor, idílio},
{carinho, cuidado},
{derriço, namorada},
{derriço, namorado},
{namorado, namorante},
{abraço, abraçamento, amplexo},
{admirador, apreciador, fã},
{admirador, namorado, admirante},
{adoração, amor extremo, paixão},
{agarramento, apego, ligação},
{amada, amante, amásia},
{amor, rabicho, namoro},
{amorico, namorico, namorisco},
{apaixonado, amante, namorado},
{ardor, paixão, amor},
{atração, simpatia, afeição},
{carinho, ternura, amor},
{derriço, namoro, rabicho},
{esposo, marido, consorte},
{flerte, galanteio, namoro ligeiro},

{frecheiro, namorado, galanteador},
{galã, namorado, galanteador},
{idolatria, amor excessivo, paixão cega},
{idílio, amor suave, amor puro},
{namorado, requestado, galanteado},
{preferência, primazia, predileção},
{admiração, respeito, afeição, simpatia},
{adorador, admirador, fã, namorado},
{adoração, culto, veneração, latria},
{atração, atrativo, fascínio, encanto},
{benevolência, estima, afeto, simpatia},
{chama, flama, fogo, amor},
{coração, amor, afeto, afeição},
{corteador, cumprimenteiro, mesureiro, galanteador},
{dileção, estima, afeto, afeição especial},
{inclinação, amor, afeição, simpatia},
{namorada, conversada, apaixonada, pequena},
{namorado, agradado, enamorado, apaixonado},
{preferência, predileção, simpatia, pendor},
{pretendida, namorada, requestada, noiva},
{querença, afeto, simpatia, afeição},
{adorante, adorador, idólatra, cultor, venerador},
{afeição, cortesia, afabilidade, benevolência, urbanidade},
{aferro, apego, predileção, paixão, inclinação},
{amigo, caro, dileto, querido, amigável},
{amor, ardor, patriotismo, entusiasmo, dedicação},
{amor, caridade, fraternidade, filantropia, altruísmo},
{beijeiro, caricioso, mimoso, menineiro, beijador},
{carinho, carícia, afago, mimo, meiguice},
{caro, querido, estimado, prezado, precioso},
{derretimento, dengue, requebro, desvanecimento, afetação},
{dom-joão, sedutor, mulherengo, femeeiro, lovelace},
{favorito, favorecido, mimoso, predileto, preferido},
{galanice, donaire, garbo, galantaria, gentileza},
{galanteador, amável, galante, corteador, namorado},
{intimidade, amizade, familiaridade, camaradagem, trato íntimo},
{namorado, galanteador, namorado, namorado, namorado},
{noiva, prometida, futura, pretendida, nubente},
{noivo, prometido, futuro, pretendido, nubente},
{predileta, preferida, estremecida, amada, dileta},

d) TEP 2.0 (<http://www.nilc.icmc.usp.br/tep2/>); [TEP]

{amor, paixão},

{beijo, ósculo},

{noiva, prometida},
{noivo, prometido},
{querença, querência},
{admirador, adorador, fã},
{atração, inclinação, simpatia},
{beijeiro, beijador, beijocador},
{carinho, meiguice, ternura},
{cortejador, doneador, galanteador},
{dileção, afeição, predileção},
{esposo, homem, marido},
{favorito, predileto, preferido},
{namoro, caso, romance},
{predileta, favorita, preferida},
{abraço, abarcamento, abraçamento, amplexo},
{afeição, bem-querer, benquerença, estima},
{amada, cara, dileta, querida},
{amado, caro, dileto, querido},
{ardor, chama, flama, paixão},
{dileção, estima, afeto, afeição especial},
{inclinação, amor, afeição, simpatia},
{intimidade, afinidade, colácia, familiaridade},
{namorico, flerte, namorilho, namorisco},
{predileção, escolha, favoritismo, preferência},
{pretendente, aspirante, candidato, pretendedor},
{admiração, afeição, afeto, inclinação, simpatia},
{agarramento, aferro, agarração, apego, chamego},
{ardor, calor, emoção, fervor, paixão},
{arroubamento, arrebatamento, encantamento, enlevamento, vôo},
{conquistador, dom-joão, dom-juan, gavião, sedutor},
{derriço, camote, namorado, pequeno, xodó},
{galanice, cortejo, galantaria, galanteio, galanteria},
{paixão, cegueira, fanatismo, obcecação, obsessão},
{devoção, adoração, amor, culto, tenção, veneração},
{fervor, devoção, diligência, piedade, religião, religiosidade},
{galanteio, chamego, namoro, paquera, rabicho, xodó},
{querido, apreçado, apreciado, caro, estimado, prezado},
{galanteador, namoradeiro, namorado, namorador, renteador, requebrador, requestador},
{afeição, afeioamento, afetividade, afeto, amizade, amor, querença, querência},
{carinho, amor, aplicação, atenção, cuidado, dedicação, desvelo, diligência, vigilância, vigília, zelo},
{apaixonada, amoriscada, ardente, babada, caída, embeçada, enamorada, encambichada, enrabichada, enxodozada, férvida, namorada},
{apaixonado, amoriscado, ardente, babado, caído, embeçado, enamorado, encambichado, enrabichado, enxodozado, férvido, namorado},
{carinho, abafo, acalentamento, acalento, acarinhamento, afago, agrado, cafuné, carícia, mimo, quindim, quitute, tagaté}

e) AZEVEDO, F.F.S. – <i>Dicionário Analógico da Língua Portuguesa</i> ; [AZE]
{noiva, pretendida, futura, nubente},
{noivo, futuro, nubente, pretendido},
{agarramento, entranha, aspiração, galanteio, galanice},
{namoro, namorico, namorisco, flerte, amizade colorida, namoramento, xaveco},
{amor, carinho, idolatria, afeto, amorico, amorosidade, amorio, inclinação},
{paixão, adoração, ardor, fervor, chamas, calor, devoção, atração, êxtase, enlevamento, arroubamento, enlevo, feitiçaria},
{chichisbéu, caro, esposo, bem, amigo, querido, derriço, predileto, cujo, zinho, frecheiro, beijocador, beijoqueiro, flerte, namorido, ficante},
{tesão, sensualidade, amor carnal, lascívia, erotismo, luxúria, volúpia, voluptuosidade, transa, sexo, cópula, sexualidade, beijo, abraço, carícia, orgasmo, gozo},
{dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência},
{namorada, arrojada, cupida, apaixonada, querida, amorzinho, amada, dulcineia, derriço, benzinho, predileta, anjo, querubim, serafim, ídolo, deusa, inclinação, objeto da simpatia, cotó, namoradeira, janeleira, frança, pau de cabeleira},
{amante, proco, namorado, namorador, namoradeiro, pretensor, pretendente, admirador, vegete, apaixonado, adorador, galanteador, cortejador, adorante, galã, amoroso, jacaré, marrancho, derriçador, quebra-esquinas, babão, bandoleiro, conquistador, dom-juan, Casanova}

f) FLORENZANO, E. <i>Dicionário de Ideias Semelhantes</i> . [IDS]
{intimidade, afinidade, colacia, familiaridade},
{atração, tentação, sedução, atrativo, amavio},
{licenciosidade, libertinagem, beijoca, beijo, ósculo, abraço},
{adoração, culto, devoção, voto, homenagem, serviço, aspiração, genuflexão, prostração},
{namorada, apaixonada, pequena, garota, anjo, ídolo, menina, deusa, diva, noiva, amante},
{querido, caro, idolatrado, estremado, estremecido, doce, rico, precioso, estimado, prezado, benquistado},
{namorado, enamorado, galã, conquistador, admirador, apaixonado, pretendente, noivo, adorador, fã, dom-juan, amante},
{casado, esposo, amo, senhor, dono, chefe, patrão, homem, macho, consorte, cônjuge, companheiro, velho},
{amado, idolatrado, estremecido, estimado, quindim, doce de coco, amor, amorzinho, amoreco, coisinha, benzinho, benzoca, namorado, ídolo, coqueluche},
{namorico, carinho, carícia, meiguice, mostra de afeto, estima, mimo, gentileza, afago, flerte, festinhas, galanteios, galanterias, brincadeiras, brinquedo},
{anseio, ansiedade, anelo, paixão, chama, flama, paixonite, apaixonite, arrebatamento, enlevo, êxtase, idolatria, adoração, devoção, fervor, ardor},
{amistosidade, relações amistosas, harmonia, concórdia, afeição, benevolência, amabilidade, afeto, simpatia, amigança, cordialidade, compreensão mútua, camaradagem, boa vontade, parcialidade, favoritismo, reciprocidade de sentimentos},
{desejo, anseio, anelo, fantasia, inclinação, pendor, cisma, predileção, favoritismo, propensão, gosto, amor, apego, parcialidade, impulso, interior, capricho},
{amor, afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfase},
{querida, preferida, predileta, dileta, de estimação, amada, idolatrada, estremecida, estimada, quindim, doce de coco, amor, amorzinho, amoreco, coisinha, benzinho, benzoca, namorada, ídolo, coqueluche},
{derriço, namoro, flerte, namorice, namorico, namorilho, namoricho, galanteio, azeite, camote, cera, grude, mormaço, paleio, pé-de-alferes, prosa, sumbaré, suruba, tribofe, xodó, agarramento},
{favorito, querido, preferido, predileto, dileto, de estimação, amado, idolatrado, estremecido, estimado, quindim, doce de coco, amor, amorzinho, amoreco, coisinha, benzinho, benzoca, namorado, ídolo, coqueluche},

{predileção, afeições, qualidades, disposição, afetividade, afeto, natureza, espírito, índole, temperamento, têmpera, pendor, inclinação, predisposição, tendência, idiosincrasia, capricho, veia, veneta, humor, simpatia, amor, caráter},
{paixão, paixa, paixoneta, paixonite, calor, febre, ardor, fogo, chama, flama, tumulto, efervescência, ebulição, ímpeto, tempestade, arroubo, arroubamento, acesso, paroxismo, explosão, surto, cena, agonia, desespero},
{caridade, virtude, benevolência, graça de Deus, boa vontade, humanidade, humanitarismo, ausência de egoísmo, abnegação, desprendimento, desinteresse, dedicação, gentileza, bondade, amabilidade, amor, fraternidade, cordialidade, coração bem-nascido, ternura, amizade, tolerância, consideração, mercê, piedade, misericórdia, clemência}
{intimidade, afinidade, colácia, familiaridade},
{atração, tentação, sedução, atrativo, amavio},
{licenciosidade, libertinagem, beijoca, beijo, ósculo, abraço},
{adoração, culto, devoção, voto, homenagem, serviço, aspiração, genuflexão, prostração},
{namorada, apaixonada, pequena, garota, anjo, ídolo, menina, deusa, diva, noiva, amante},

Não é possível reproduzir aqui a matriz completamente preenchida com todos os CSs e CSAs em suas respectivas células. Dessa forma, para ver os arranjos dos CSs e CSAs dentro da matriz léxico-conceitual do *anaset* AMOR, é necessário o arquivo *anaset_AMOR.csv* – presente no CD em anexo – no programa *Microsoft Excel™*. No entanto, o Quadro 7.6 exemplifica o arranjo dos CSs e CSAs para as ULs *afeição*, *amor* e *ardor* e para as categorias conceituais S1, glosado por “afeição muito intensa de uma pessoa por outra”, e S2, glosado por “atração (inclusive sexual) muito intensa”.

Quadro 7.6. Exemplo do *anaset* AMOR preenchido com CSs e CSAs

MUNDO PSÍQUICO		
AMOR	Relações Paradigmáticas	
	S1	S2
Afeição	{afeto, adoração, afeição, amizade, amor, apego, carinho, dedicação, dileção, estima, meiguice, benquerença, querença, simpatia, ternura} [HOU] {amor, afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, enfatuação} [IDS] {dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência} [AZE] {afeição, afeto, amor} [MIC] {afeição, afeto, amizade, amor, benquerença, carinho, inclinação, propensão, simpatia} [FER] {afeição, afeioamento, afetividade, afeto, amizade, amor, querença, querença} [TEP]	

<i>Amor</i>	{amor, afeição, afeto, amizade, apego, apreço, benquerença, benquerer, carinho, estima, simpatia, ternura} [HOU] {amor, afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfase} [IDS] {amor, carinho, idolatria, afeto, amorico, amorosidade, amorio, inclinação} [AZE] {afeição, afeto, amor} [MIC] {amor, afeto, afeição, amizade, dileção, inclinação, simpatia, paixão, bem-quer, benquerença, idolatria, adoração} [FER] {afeição, afeiçoamento, afetividade, afeto, amizade, amor, querença, querença} [TEP]	{amor, atração, desejo, fascínio, interesse, paixão} [HOU] {amor, paixão, atração, desejo} [MIC] {chama, flama, fogo, amor} [FER] {desejo, anseio, anelo, fantasia, inclinação, pendor, cisma, predileção, favoritismo, propensão, gosto, amor, apego, parcialidade, impulso, interior, capricho} [IDS] {amor, paixão} [TEP]
<i>Ardor</i>		{ardor, paixão, amor, desejo, exaltação, fervor} [HOU] {anseio, ansiedade, anelo, paixão, chama, flama, paixonite, apaixonite, arrebatamento, enlevo, êxtase, idolatria, adoração, devoção, fervor, ardor} [IDS] {paixão, adoração, ardor, fervor, chama, calor, devoção, atração, êxtase, enlevamento, arroubamento, enlevo, feitiçaria} [AZE] {ardor, paixão, entusiasmo, energia} [MIC] {ardor, paixão, amor} [FER] {ardor, chama, flama, paixão} [TEP] {ardor, calor, emoção, fervor, paixão} [TEP]

7.1.4 ETAPA 4 - CONSOLIDAÇÃO DOS RESULTADOS DO *ANASET AMOR*

Com a matriz completamente preenchida com as ULs de entrada (7.1.1), as categorias conceituais (7.1.2) e os CSs e CSAs correspondentes (7.1.3), os resultados podem ser consolidados seguindo-se os procedimentos apresentados em 6.1.2.4. No entanto, ainda que os cálculos para a consolidação dos resultados tenham sido efetuados para as duas maneiras referidas em 6.1.2.4, só os resultados para a consolidação de ordem onomasiológica é que serão exibidos nessa seção⁵⁰. Isto porque são esses os resultados que serão implementados no banco de dados relacional difuso do DAD.

O Quadro 7.7 reproduz, portanto, apenas os resultados consolidados por meio do algoritmo de ordem onomasiológica. Na primeira coluna à esquerda temos as dez categorias conceituais do *anaset AMOR*, seguidas, na coluna adjacente, por suas respectivas glosas. Na última coluna são apresentados os CdSAs correspondentes⁵¹.

Quadro 7.7. Resultados consolidados para o *anaset AMOR*

⁵⁰ Para ver a consolidação dos resultados de ordem semasiológica, é necessário abrir o arquivo CdSA_sem_AMOR.csv, presente no CD em anexo.

⁵¹ Utilizamos um corte em 0.1 para exibir os resultados sem ocupar tanto espaço no meio do texto da tese. Para ver os resultados completos, é necessário abrir o arquivo anaset_AMOR.csv no programa *Microsoft Excel*TM.

Resultados consolidados para o <i>anaset</i> AMOR		
(S)	Glosa	CdSAs
Relação Paradigmática		
(S1)	Afeição muito intensa de uma pessoa por outra	{afeição (1), amor (0.958), afeto (0.896), simpatia (0.813), ternura (0.667), apego (0.604), carinho (0.563), inclinação (0.5), benquerença (0.458), benevolência (0.458), querença (0.438), dileção (0.396), amizade (0.396), dedicação (0.333), admiração (0.292), estima (0.292), predileção (0.25), preferência (0.229), derretimento (0.229), meiguice (0.229), idílio (0.229), aferro (0.208), constância (0.208), intimidade (0.208), derriço (0.208), conchego (0.208), estremecimento (0.208), idiopatia (0.208), fraternidade (0.208), chamego (0.188), boa vontade (0.188), caridade (0.188), ênfatuação (0.188), comunhão de sentimentos (0.188), gosto (0.188), desvelo (0.188), zelo (0.188), cuidado (0.188), adoração (0.125), idolatria (0.104), amorosidade (0.104), [...]}
(S2)	Atração (inclusive sexual) muito intensa entre duas pessoas	{paixão (1), ardor (0.684), fervor (0.526), chama (0.447), adoração (0.421), amor (0.395), atração (0.368), devoção (0.342), calor (0.316), êxtase (0.289), enlevo (0.289), flama (0.263), idolatria (0.237), arroubamento (0.211), paixonite (0.184), desejo (0.158), anelo (0.158), feitiçaria (0.158), enlevamento (0.158), arrebatamento (0.158), anseio (0.158), ansiedade (0.132), apaixonite (0.132), emoção (0.105), fogo (0.105), apego (0.105), [...]}
(S3)	Amizade intensa entre duas pessoas	{simpatia (1), afeição (1), afeto (0.848), amor (0.609), amizade (0.565), ternura (0.478), querença (0.478), dileção (0.457), apego (0.457), benquerença (0.457), benevolência (0.435), dedicação (0.391), admiração (0.391), estima (0.37), predileção (0.37), intimidade (0.326), carinho (0.304), preferência (0.304), inclinação (0.261), derretimento (0.217), aferro (0.217), estremecimento (0.217), constância (0.217), idiopatia (0.217), derriço (0.217), idílio (0.217), conchego (0.217), afinidade (0.152), boa vontade (0.152), fraternidade (0.152), meiguice (0.152), adoração (0.13), amabilidade (0.109), favoritismo (0.109), camaradagem (0.109), cordialidade (0.109), afetividade (0.109), [...]}
(S4)	Veneração de, adoração de, devoção a um ideal concreto ou abstrato	{adoração (1), devoção (0.92), veneração (0.64), paixão (0.48), fervor (0.44), idolatria (0.4), amor (0.4), culto (0.36), dedicação (0.36), ardor (0.36), êxtase (0.28), chama (0.28), enlevo (0.28), piedade (0.16), latria (0.16), arrebatamento (0.16), afeição (0.16), ansiedade (0.16), apaixonite (0.16), anelo (0.16), flama (0.16), anseio (0.16), paixonite (0.16), feitiçaria (0.12), arroubamento (0.12), enlevamento (0.12), tenção (0.12), atração (0.12), respeito (0.12), calor (0.12), [...]}
(S5)	Demonstração de zelo	{cuidado (1), amor (0.9), desvelo (0.9), carinho (0.8), dedicação (0.8), zelo (0.7), atenção (0.6), apego (0.4), diligência (0.4), vigilância (0.3), aplicação (0.3), vigília (0.3), afeto (0.3), afeição (0.3), boa vontade (0.2), benevolência (0.2), simpatia (0.2), fraternidade (0.2), ternura (0.2), gosto (0.2), comunhão de sentimentos (0.2), caridade (0.2), inclinação (0.2), adoração (0.2), ênfatuação (0.2), devoção (0.2), chamego (0.2), consideração (0.1), fidelidade (0.1), paixão (0.1), devotamento extremo (0.1), apreço (0.1), veneração (0.1), ardor (0.1), devotamento (0.1), patriotismo (0.1), entusiasmo (0.1), esmero (0.1), capricho (0.1), [...]}
(S6)	Relação amorosa	{namoro (1), namorico (0.833), flerte (0.8), galanteio (0.633), xodó (0.533), agarramento (0.533), namorilho (0.533), namorisco (0.5), derriço (0.433), namorice (0.367), namoramento (0.367), paleio (0.333), romance (0.333), grude (0.333), chamego (0.333), cera (0.333), pé-de-alferes (0.333), prosa (0.333), camote (0.267), suruba (0.267), mormaço

		(0.267), tribofe (0.267), azeite (0.267), sumbaré (0.267), caso (0.267), namoricho (0.267), amor (0.233), rabicho (0.233), aventura (0.233), amorico (0.2), corte (0.167), xaveco (0.167), amizade colorida (0.167), linhada (0.133), amorio (0.133), amorinho (0.133), apego (0.133), ligação (0.1), embeijamento (0.1), agarração (0.1), idílio (0.1), [...]
Relação Sintagmática		
(S7)	Homem que ama ou é amado	{namorado (1), querido (0.815), bem (0.667), amor (0.63), amado (0.593), caro (0.556), predileto (0.556), amante (0.519), derriço (0.481), amigo (0.444), beijoqueiro (0.444), beijocador (0.407), apaixonado (0.37), dileto (0.333), esposo (0.333), estimado (0.296), enamorado (0.296), flerte (0.259), ídolo (0.259), namorado (0.259), frecheiro (0.259), noivo (0.259), zinho (0.259), chichisbéu (0.259), amásio (0.259), cujo (0.259), ficante (0.259), preferido (0.259), favorito (0.222), benzinho (0.222), estremecido (0.222), idolatrado (0.185), nubente (0.148), benzoca (0.148), coqueluche (0.148), pequeno (0.148), quindim (0.148), prezado (0.148), coisinha (0.148), amoreco (0.148), doce de coco (0.148), amorzinho (0.148), pretendido (0.111), tesouro (0.111), prometido (0.111), galã (0.111), beijador (0.111), de estimação (0.111), [...] }
(S8)	Mulher que ama ou é amada	{namorada (1), amada (0.788), querida (0.727), amante (0.515), ídolo (0.485), amor (0.485), predileta (0.455), apaixonada (0.455), deusa (0.424), noiva (0.394), benzinho (0.333), amorzinho (0.333), dulcineia (0.333), anjo (0.303), pequena (0.303), derriço (0.303), bem (0.273), preferida (0.273), amásia (0.273), dileta (0.242), inclinação (0.212), nubente (0.212), pretendida (0.212), concubina (0.212), objeto da simpatia (0.182), cotó (0.182), pau de cabeloira (0.182), namoradeira (0.182), garota (0.182), tesouro (0.182), frança (0.182), janeleira (0.182), cupida (0.182), serafim (0.182), estimada (0.182), estremecida (0.182), prometida (0.182), querubim (0.182), arrojada (0.182), amoreco (0.152), idolatrada (0.152), de estimação (0.152), coisinha (0.152), benzoca (0.152), doce de coco (0.152), coqueluche (0.152), quindim (0.152), barregã (0.152), amiga (0.152), menina do coração (0.121), menina dos olhos (0.121), menina do peito (0.121), menina (0.121), diva (0.121), mulher (0.121), [...] }
(S9)	Homem que corteja uma mulher	{galanteador (1), namorado (0.903), admirador (0.903), adorador (0.871), galã (0.839), namorador (0.806), amante (0.774), pretendente (0.774), conquistador (0.742), apaixonado (0.71), cortejador (0.677), dom-juan (0.677), namoradeiro (0.581), fã (0.484), adorante (0.419), pretensor (0.419), vegete (0.387), quebra-esquinas (0.387), babão (0.387), bandoleiro (0.387), Casanova (0.387), derriçador (0.387), marrancho (0.387), proco (0.387), jacaré (0.387), amoroso (0.387), noivo (0.258), enamorado (0.258), requestador (0.194), mulherengo (0.194), femeeiro (0.161), candidato (0.129), cativador (0.129), aspirante (0.129), sedutor (0.129), [...] }
(S10)	Manifestação física do amor	{carícia (1), carinho (1), abraço (0.889), beijo (0.889), galanteio (0.667), galanteria (0.667), galanice (0.667), ternura (0.667), ósculo (0.667), galantaria (0.556), agarramento (0.556), meiguice (0.556), amplexo (0.444), mimo (0.444), afago (0.444), cortejo (0.333), gentileza (0.333), derretimento (0.333), abraçamento (0.333), afeto (0.333), requebro (0.333), agarração (0.333), beijoca (0.333), corte (0.222), volúpia (0.222), agarra (0.222), luxúria (0.222), lisonja (0.222), sexualidade (0.222), voluptuosidade (0.222), orgasmo (0.222), licenciosidade (0.222), gozo (0.222), cópula (0.222), lascívia (0.222), transa (0.222), sexo (0.222), erotismo (0.222), denguiço (0.222), amor carnal (0.222), entranha (0.222), donaire (0.222), garbo (0.222), libertinagem (0.222), delicadeza

(0.222), aspiração (0.222), sensualidade (0.222), afetação (0.222), tesão (0.222), agrado (0.222), denego (0.222), apego (0.111), galanteio (0.111), aferro (0.111), abarcamento (0.111), enlaçamento (0.111), brandura (0.111), doçura (0.111), festinha (0.111), festa (0.111), blandícia (0.111), xeta (0.111), chamego (0.111), acarinhamento (0.111), acalento (0.111), acalentamento (0.111), brinquedo (0.111), abafo (0.111), cafuné (0.111), [...]

7.2 MODELAGEM DO ANASET CORPO HUMANO

Essa Seção relata a aplicação do modelo no domínio léxico-conceitual CORPO HUMANO. Para tanto, foram organizadas subseções para cada uma das etapas. Em 7.2.1 temos a identificação e coleta das ULs do domínio CORPO HUMANO. Em 7.2.2 procedemos com a estruturação conceitual do *anaset* CORPO HUMANO. Em 7.2.3 executamos a identificação, coleta e arranjo dos CSs e CSAs pertinentes ao *anaset* CORPO HUMANO. Por fim, em 7.2.4, temos a consolidação dos resultados, em que apresentamos uma tabela que representa a modelagem final do *anaset* CORPO HUMANO.

7.2.1 ETAPA 1 - IDENTIFICAÇÃO E COLETA DAS ULs DO DOMÍNIO CORPO HUMANO

Para gerar a lista das ULs pertencentes ao domínio léxico-conceitual CORPO HUMANO partimos da lista de ULs extraídas do *corpus*. As dez ULs do domínio CORPO HUMANO extraídas do *corpus* aparecem elencadas no Quadro 7.8.

Quadro 7.8. ULs do domínio CORPO HUMANO extraídas do *corpus*

Domínio léxico-conceitual CORPO HUMANO (10 ULs)	
	rosto
	corpo
	coração
	pele
	boca
	perna
	orelha
	bigode
	olho
	pé

Em seguida, foram consultadas as fontes de informação léxico-conceituais e para identificadas as ULs sinônimas e análogas que pertencem ao mesmo domínio conceitual. Dessa forma, procedeu-se à expansão do número de ULs de entrada que deverá ocupar a coluna da extrema esquerda na matriz léxico-conceitual do *anaset* CORPO HUMANO.

Conforme os critérios apontados em 6.1.2.1, não foram incluídas na lista as ULs complexas. Como exemplo das ULs que **não** foram incluídas como ULs de entrada no

anaset CORPO HUMANO estão: *aparelho fonador, caixa torácica, batata da perna, céu da boca, medula óssea, coluna vertebral, espinha dorsal, nervo óptico, monte de Vênus.*

Ao contrário do domínio léxico-conceitual AMOR, o domínio CORPO HUMANO não possui outros domínios conceituais adjacentes com os quais compartilhe fronteiras difusas. Como é comum acontecer com domínios do MUNDO FÍSICO, o domínio CORPO HUMANO é bem delimitado. Ou seja, os sinônimos e analógicos de uma UL do domínio CORPO HUMANO nos levarão sempre para ULs desse mesmo domínio, com exceção dos casos em que claramente se trata de uma conexão metafórica.

Seguindo a metodologia e os critérios estabelecidos em 6.1.2.1, chegamos a uma lista de 148 ULs de entrada⁵² para o *anaset* CORPO HUMANO, conforme o Quadro 7.9.

Quadro 7.9. ULs de entrada do *anaset* CORPO HUMANO

Domínio léxico-conceitual CORPO HUMANO (148 ULs)			
<u>abdômen</u> /abdome	cutícula	medula	pulso
<u>amígdala</u> /amídala	cútis	mindinho	punho
<u>anca</u> /ancas	dedo	miolo	<u>quadril</u> /quadrís
anular	dedo médio	moleira	quartos
ânus	dente	munheca	rabo
artelho	derme	<u>nádega</u> /nádegas	retina
artéria	diafragma	narina	reto
axila	duodeno	nariz	<u>rim</u> /rins
bacia	encéfalo	nuca	rosto
barriga	entranha	<u>olho</u> /olhos	<u>seio</u> /seios
beço	epiderme	ombro	septo
bexiga	esclerótica	orelha	<u>sobrancelha</u> /sobrancelhas
boca	esôfago	ouvido	sovaco
bochecha	estômago	ovário	supercílio
braço	face	pai-de-todos	testa
bumbum	faringe	palato	testículo
bunda	fígado	palma	tímpano
busto	flanco	pálpebra	tórax
cabeça	fronte	pança	tornozelo
cabeleira	fuça	pâncreas	torso
cabelo	garganta	panturrilha	traqueia
cachola	gengiva	papo	traseiro
<u>cadeira</u> /cadeiras	genitália	pé	trompa
calcanhar	glânde	<u>peito</u> /peitos	tronco

⁵² Algumas células do quadro apresentam também as formas variantes encontradas nas diferentes fontes de informação. A forma sublinhada é a que foi escolhida como padrão.

cambito	goela	pele	umbigo
cangote	gogó	pelo	unha
cara	ilharga	pênis	uretra
cérebro	indicador	perna	útero
cílio	intestino	pescoço	úvula
cocuruto	íris	pinto	vagina
colo	lábio	pituitária	<u>venta/ventas</u>
<u>cólon/colo</u> ²	laringe	pleura	ventre
coração	língua	polegar	vesícula
costas	lombo	pomo-de-adão	víscera
cotovelo	mamilo	prepúcio	vulva
crânio	mão	próstata	
cuca	mata-piolho	pulmão	

7.2.2 ETAPA 2 – ESTRUTURAÇÃO CONCEITUAL DO ANASET CORPO HUMANO

No quadro sinóptico de categorias elaborado por Roget o domínio CORPO HUMANO se encontra no extremo da seguinte ramificação: **classe III. Matéria> divisão III. Matéria Orgânica> 2º) Sensação> II. Em especial> 6º) Luz> 440e. Partes do Corpo Humano**. Na WN.Pr, os conceitos relacionados com partes do corpo humano, como CABEÇA (HEAD⁵³), PÉ (FOOT⁵⁴) e CORAÇÃO (HEART⁵⁵), organizam-se, em diferentes níveis, abaixo do conceito PARTES DO CORPO (BODY PARTS⁵⁶). Os dois primeiros são categorizados como PARTES EXTERNAS DO CORPO (EXTERNAL BODY PART⁵⁷), sendo que PÉ se encontra ainda abaixo da ramificação EXTREMITY > VERTEBRATE FOOT > FOOT. Já o conceito de CORAÇÃO encontra-se na ponta da seguinte ramificação:

⁵³ {05546258} <noun.body>S: (n) head, caput (the upper part of the human body or the front part of the body in animals; contains the face and brains)

⁵⁴ {05570899} <noun.body>S: (n) foot, human foot, pes (the part of the leg of a human being below the ankle joint)

⁵⁵ {05396148} <noun.body>S: (n) heart, pump, ticker (the hollow muscular organ located behind the sternum and between the lungs; its rhythmic contractions move the blood through the body)

⁵⁶ {05227735} <noun.body>S: (n) body part (any part of an organism such as an organ or extremity)

⁵⁷ {05232383} <noun.body>S: (n) external body part (any body part visible externally)

<noun.body>S: (n) **heart**, pump, ticker (the hollow muscular organ located behind the sternum and between the lungs; its rhythmic contractions move the blood through the body)

<noun.body>S: (n) **internal organ**, viscus (a main organ that is situated inside the body)

<noun.body>S: (n) **organ** (a fully differentiated structural and functional unit in an animal that is specialized for some particular function)

<noun.body>S: (n) **body part** (any part of an organism such as an organ or extremity)

Na EWN, os conceitos relacionados com partes do corpo humano estão categorizados como PART e LIVING, do topo da ontologia. No caso da WN.Pr, acima do conceito PARTES DO CORPO, têm-se ainda a ramificação dos conceitos PART > THING > PHYSICAL ENTITY > ENTITY. E, no caso da EWN, o topo da hierarquia dos conceitos relacionados com partes do corpo humano termina na divisão ENTIDADE DE 1ª ORDEM (1ST ORDER ENTITY), que abrange exatamente as entidades concretas, perceptíveis pelos sentidos e localizadas em algum ponto no tempo e em um espaço tridimensional (Cf. 4.2).

No HTOED, a categoria conceitual CORPO HUMANO (HUMAN BODY) é hipônimo direto de VIDA (LIFE) que, por sua vez, alinha-se diretamente abaixo da Grande Categoria MUNDO EXTERNO (THE EXTERNAL WORLD), como se vê ilustrado abaixo:

01 The external world

01.02 Life

02.02.05 The Human Body

Tomando como referência as estruturas conceituais antes mencionadas, principalmente a do HTOED, identificamos que o domínio conceitual CORPO HUMANO deve ser representado dentro da Grande Categoria MUNDO FÍSICO.

Quadro 7.9. Matriz preenchida com a Grande Categoria onde se inclui o *anaset* CORPO HUMANO: MUNDO FÍSICO

MUNDO FÍSICO						
CORPO HUMANO	Relações Sintagmáticas					
	S1	S2	S3	S4	S5	S6
UL 1	{cs1} [X] {cs2} [Y] {cs3} [Z]					

Diferentemente do domínio conceitual AMOR, em que a maioria dos conceitos são organizados por meio de relações paradigmáticas, o domínio CORPO HUMANO é talvez um dos exemplos mais emblemáticos de um domínio estruturado pela relação léxico-conceitual da meronímia (relação sintagmática). O corpo humano é um todo percebido conceitualmente como dividido em partes. O conceito BRAÇO não é um gênero do conceito CORPO HUMANO, mas uma parte dele (Cf. 3.3). No domínio conceitual AMOR, mais da metade dos conceitos estão em relação paradigmática e um deles, o conceito de RELAÇÃO AMOROSA, evoca um *frame* social, que possui uma estrutura interna, em que identificamos conceitos em relação sintagmática. O domínio CORPO HUMANO, entendido como PARTES DO CORPO HUMANO, se estrutura, todo ele, pela relação sintagmática de parte-todo (*part-whole*). O domínio conceitual CORPO HUMANO é um *frame* físico e, como tal, mantém relações de contiguidade entre os seus conceitos. É sintomático que, diferentemente do domínio AMOR, o domínio conceitual CORPO HUMANO possua um *frame* correspondente dentro da FrameNet – ainda que limitado às partes externas do corpo humano –, o *frame* `Observable_body_parts`. Como foi dito em 3.3, as relações paradigmáticas são melhor representadas dentro de estruturas como as das *wordnets* ao passo em que as relações sintagmáticas são melhor representadas em esquemas como os das *framenets*. No entanto, o *frame* `Observable_body_parts` da FrameNet de Berkeley não estrutura o domínio CORPO HUMANO internamente, ou seja, não responde à pergunta: quantos e quais são os conceitos para as partes do corpo humano em uma granularidade que seja relevante para a estruturação do domínio? Pois, apesar de ser uma estrutura que melhor representa as relações sintagmáticas, não é vocação do projeto da FrameNet descrever *frames* físicos, mas sim *frames* sociais. Ou seja, no projeto da FrameNet, um *frame* é entendido como uma estrutura conceitual

culturalmente definida que atua na organização de segmentos da realidade social. Por este motivo, a FrameNet mostra-se útil como uma referência para modelar o *frame* de *Relação_Amorosa* (no domínio AMOR) e mais ainda do *frame* *Arte_Performática* (no domínio TEATRO), mas não para auxiliar na modelagem e descrição de entidades físicas como as que se apresentam no caso do domínio CORPO HUMANO.

Conseqüentemente, para delimitar e estruturar os conceitos internos ao domínio CORPO HUMANO, não utilizamos como ponto de partida nem o modelo da FrameNet e nem os dicionários semasiológicos tal como o procedimento efetuado para o domínio AMOR. A UL *corpo humano* não constitui entrada em dicionários deste tipo e, mesmo que houvesse um verbete **corpo humano**, suas partes (cabeça, pé, coração) não seriam, e nem poderiam ser, arroladas como sendo acepções de *corpo humano*. Tivemos de recorrer, portanto, a outros modelos e encontramos nos dicionários visuais a referência adequada para a estruturação conceitual do domínio.

O nosso objetivo, seguindo a tradição dos dicionários analógicos, é estruturar o domínio CORPO HUMANO em campos conceituais valendo-se de uma granularidade que seja útil para o consulente de um dicionário criativo e não para o consulente de um dicionário técnico de anatomia. Como entidades do mundo físico possuem um número muito reduzido de sinônimos, emparelhar uma UL como *pé* com o seu estrito conceito PÉ e outra UL como *tornozelo* com seu estrito conceito TORNOZELO, faria com que a estruturação conceitual do domínio tivesse quase o mesmo número de conceitos quanto tem de ULs, o que apresenta muito pouca utilidade prática. A solução de estruturar o domínio a partir de categorias conceituais um pouco mais extensas permite agrupar conceitos e ULs que, inclusive, podem ser intercambiáveis em alguns contextos, como no par de frases “ele torceu o *tornozelo*” e “ele torceu o *pé*”.

Cada categoria conceitual do domínio CORPO HUMANO, que equivale a cada uma das colunas da matriz léxico-conceitual do *anaset* correspondente, foi definida segundo as divisões propostas pelo *Dicionário Visual do Corpo Humano SBS* (cf. Apêndice II.2, item f). Na figura 7.1, temos um exemplo para a categoria conceitual BOCA.

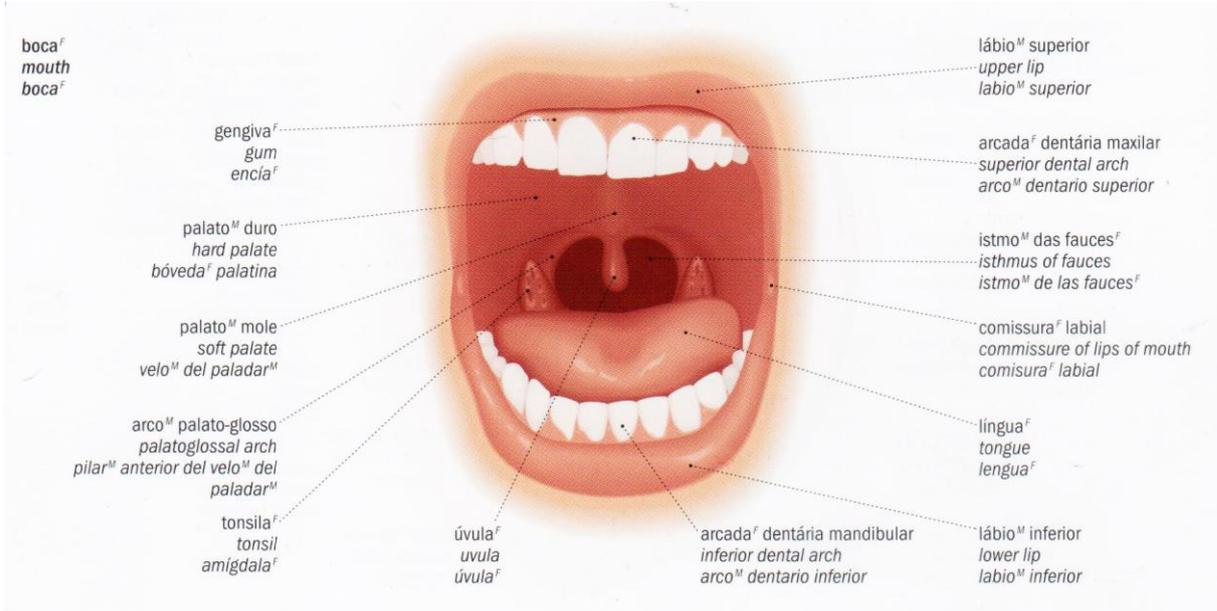


Figura 7.1. Exemplo da categoria conceitual BOCA extraído do *Dicionário Visual do Corpo Humano SBS*

Como veremos, na Etapa 3, o conjunto de ULs que aparece em cada categoria conceitual do *Dicionário Visual do Corpo Humano SBS* será transformado, mediante algumas adaptações, em um CSA e será aproveitado no preenchimento das células da matriz do *anaset* CORPO HUMANO.

O Quadro 7.10 apresenta a matriz léxico-conceitual do *anaset* CORPO HUMANO preenchida com a estrutura conceitual discutida ao longo dessa seção.

Quadro 7.10. Matriz do *anaset* CORPO HUMANO preenchida com a estrutura conceitual

MUNDO FÍSICO										
CORPO HUMANO	Relações Sintagmáticas									
	S1	S2	S3	S4	S5	S6	S7	S8	S(...)	S24
UL 1										

O Quadro 7.10 complementa-se com as informações presentes no Quadro 7.11, em que aparecem todas as conceituais pertencentes ao *anaset* CORPO HUMANO. Elas aparecem instanciadas por uma pequena glosa e por CSs/CSAs ilustrativos. Além disso, a última coluna apresenta abonações extraídas do *corpus* do NILC.

Quadro 7.11. Estrutura conceitual do *anaset* CORPO HUMANO

Estrutura conceitual do <i>anaset</i> CORPO HUMANO			
(S)	Glosa	CSs/CSAs	Exemplo
Relação Sintagmática			
(S1)	Parte superior do corpo humano	{cabeça, cachola, cachimônia, coco, bola, caco, cuia, chocolateira, bestunto, pinha, quengo, tola, touda, topete, sinagoga} [FER] {head, caput} [WN]	“Esse sistema compreende o encéfalo , localizado na cabeça , no interior do crânio [...]” [NILC]
(S2)	Parte anterior da cabeça	{cara, cariz, face, queixo, rosto, semblante, vulto} [TEP] {countenance, physiognomy, phiz, visage, kisser, smiler, mug} [WN]	“[...] além de quebrar o nariz e sofrer diversas escoriações no rosto .” [NILC]
(S3)	Conjunto do pelos que crescem na cabeça	{cabelo, cabeleira, cabelama, cabeladura, cabelame, cabelugem, coma, encabeladura, juba} [HOU] {hair} [WN]	“[...] tem cabelo castanho claro pintado e aparenta ter 32 anos.” [NILC]
(S4)	Órgão da visão	{olho, pálpebra, cílio, pupila, íris, esclera, retina, córnea, nervo óptico, cristalino, conjuntiva, esclerótica} [SBS] {eye, oculus, optic, orb, peeper, visual organ, retina, lens, cornea, sclera, optic nerve, ..., aqueous humor, vitreous humos} [RIT]	“A córnea transparente do olho é curva e funciona como uma lente.” [NILC]
(S5)	Órgão da audição	{orelha, aurícula, nambi, pavilhão auricular} [HOU] {ear} [WN]	“Por isso, também a orelha é formada de tecido cartilaginoso.” [NILC]
(S6)	Órgão do olfato, parte inicial das vias respiratórias	{nariz, septo, narina, asa} [SBS] {nose, nasal organ, snout, smeller, proboscis, beak, schnoz, muzzle, ..., nasal cavity, olfactory nerve} [RIT]	“O nariz , por exemplo, para ter a sua forma, precisa de uma armação — o septo nasal .” [NILC]
(S7)	Cavidade situada na cabeça, delimitada externamente pelos lábios e internamente pela faringe	{boca, gengiva, palato, úvula, arcada dentária, dente, língua, lábio} [SBS] {mouth, oral cavity, lips, tongue, taste buds, mandible, jaw, maw, gums, ..., uvula, teeth} [RIT]	“resultando um som que, depois de passar pela boca , sendo trabalhado pela língua, dentes e lábios, constituirá a voz da pessoa.” [NILC]
(S8)	Parte do corpo entre o tronco e a cabeça	{garganta, goela, fauce, gasganete, gorja, pescoço} [HOU] {neck, cervix} [WN]	“A tireóide localiza-se no pescoço , à frente da traquéia ” [NILC]
(S9)	Sistema circulatório	{coração, artéria, veia, aorta, miocárdio, endocárdio, sangue, vaso, vaso sanguíneo, ..., glóbulo, plaqueta} [IDS] {heart, ticker, pump, endocardium, atria, ventricles} [RIT]	“O coração , por exemplo, é o órgão central do sistema circulatório .” [NILC]
(S10)	Parte do tronco que vai do pescoço ao abdome	{peito, tórax, torso} [TEP] {thorax, chest, pectus} [WN]	“Dor no tórax nem sempre é infarto.” [NILC]
(S11)	Sistema respiratório	{boca, nariz, epiglote, faringe, laringe, corda vocal, esôfago, traqueia, pulmão, brônquio,	“Abertura da traqueia para a colocação de um tubo para melhorar a entrada de ar nos pulmões .” [NILC]

		diafragma} [SBS] {lungs, bellow, lights, diaphragm, windpipe, trachea, weasand, wizen, bronchus, bronchi, bronchial tube, epiglottis} [RIT]	
(S12)	Parte do corpo humano entre o tórax e a pelve	{barriga, ventre, abdômen, abdome, pança, bandulho, pandulho} [FER] {abdomen, venter, stomach, belly} [WN]	“É o tecido gorduroso, que fica abaixo da pele, em todo o corpo, mas principalmente no abdome (barriga) .” [NILC]
(S13)	Sistema digestivo	{boca, língua, glândula salivar, faringe, esôfago, estômago, fígado, pâncreas, ..., reto, esfíncter, ânus} [SBS] {mouth, maw, salivary glands, gullet, crop, craw, throat, pharynx, esophagus, gorge, weasand, ..., rectum, anus} [RIT]	“Membrana formada por epitélio estratificado (reveste a cavidade bucal , a faringe e o esôfago) ou epitélio simples (reveste o estômago e o intestino).” [NILC]
(S14)	Sistema urinário	{rim, bexiga, ureter, uretra, glândula suprarrenal} [SBS]	“Os médicos removeram então o outro rim e o ureter , tubo que liga o rim à bexiga .” [NILC]
(S15)	Parte correspondente ao peito no corpo feminino	{peito, seio, mama, teta} [MIC] {breast, bosom, knocker, boob, tit, titty} [WN]	“O decote deixa à mostra o bico de seu seio direito e revela parte do esquerdo.” [NILC]
(S16)	Membros superiores	{ombro, axila, braço, cotovelo, antebraço, pulso, mão} [SBS] {arm, forearm, wrist, elbow, upper arm, biceps} [RIT]	“O outro é negro, alto, usa bigode e estava com um braço enfaixado.” [NILC]
(S17)	Extremidade do membro superior, articulada com o antebraço pelo punho e terminada pelos dedos	{mão, palma, dorso, unha, dedo, dedo mínimo, dedo médio, dedo anular, dedo indicador, dedo polegar, pulso, lúnula} [SBS] {hand, paw, finger} [RIT]	“Ela poderia ser medida também de outras maneiras, como, por exemplo, com a mão espalmada.” [NILC]
(S18)	Membros inferiores	{perna, coxa, panturrilha, joelho, pé, dedo do pé, tornozelo, dorso do pé, calcanhar} [SBS] {leg, limb, shank, gam, pin, legs, wheels, shin, cnemis, ankle, tarsus, calf, knee, thigh, ham, popliteal space} [RIT]	“Ele sofreu fratura no perônio direito e «lesões extensas» na mesma perna , com perda de musculatura da panturrilha (batata da perna) .” [NILC]
(S19)	Extremidade do membro inferior abaixo da articulação do tornozelo e terminada pelos artelhos	{patela, pé, chanca, canastra, pesunho, prancha, patola, toesa, metatarso, artelho, tornozelo, maléolo} [AZE] {foot, dog, puppy, toe} [RIT]	“Acho que torci o pé direito, contou.” [NILC]
(S20)	Parte posterior do tronco	{região lombar, lombo, rim, isquiu, costão, ilharga, quadril, anca, cadeira, hipocôndrio, costas} [AZE]	“Talvez uma injeção para dor nas costas .” [NILC]
(S21)	Parte do corpo humano da cintura à articulação da coxa	{anca, cadeira, quadril, costaneira, lombo, nádega, culatra, pousadeiro, pousadouro, bunda, bumbum, ..., rabicho, rabadela} [AZE]	“A calçinha ficou muito pequena e não passou pelo quadril .” [NILC]

		{back end, backside, behind, bottom, bum, derrière, fanny, fundament, gluteus maximus, haunches, posterior, ..., seat, tush} [RIT]	
(S22)	Sistema reprodutor feminino	{lábio maior, lábio menor, clitóris, vagina, colo do útero, útero, ovário, tuba uterina, vulva} [SBS] {vulva, yoni, cunt, vagina, clitoris, glans clitoridis, pudenda, labia, labia majora, labia minora, lips, nymphae, ovary, ..., uterus, womb} [RIT]	“A vagina é um tubo que liga o útero ao meio externo.” [NILC]
(S23)	Sistema reprodutor masculino	{pênis, saco escrotal, glande, prepúcio, meato urinário, uretra, testículo, epidídimo, corpo cavernoso, próstata} [SBS] {penis, phallus, lingam, glans penis, gonads, testes, testicles, balls, nuts, rocks, ballocks, nads, ..., scrotum, bag, basket} [RIT]	“O esperma é liberado através da uretra no pênis .” [NILC]
(S24)	Membrana mais ou menos espessa que reveste o corpo humano	{tegumento, tegmento, córion, pele, derma, indúvia, tez, epiderme, cútis, carão, cutícula, película, tona} [AZE] {skin, tegument, cutis, dermis, derma, dermat, corium, true skin, cuticle, epithelium, ..., endoderm, entoderm} [RIT]	“Esse conjunto é envolto externamente pela pele .” [NILC]

7.2.3 ETAPA 3 - IDENTIFICAÇÃO, COLETA E ARRANJO DOS CSS E CSAS DO ANASET CORPO HUMANO

As células da matriz léxico-conceitual serão preenchidas com os CSS e CSAs coletados nas fontes de informação léxico-conceitual, listadas no Apêndice II.1, itens **a-e** e **g**. Houve a necessidade, no caso particular do domínio CORPO HUMANO, diferentemente dos dois outros domínios, de se fazer a substituição do *Dicionário de Ideias Semelhantes* [IDS] pelo *Dicionário Visual SBS do Corpo Humano* [SBS] como uma das seis fontes de informação lexical que estamos utilizando como padrão para a extração dos CSS e CSAs. Isso se deve ao fato de que o [IDS] possui um número muito pequeno de CSAs relativos ao domínio CORPO HUMANO. Mantê-lo como fonte de informação para este domínio provocaria distorções indesejadas nos resultados. Sendo o domínio CORPO HUMANO concernente à grande categoria MUNDO FÍSICO e, pelo que foi explanado em 3.3 e 4.7 sobre os conceitos pertencentes a esta categoria serem relacionados metonimicamente – em especial em relações semânticas de parte-todo –, foi natural a seleção do dicionário visual [SBS] para ocupar a lacuna deixada pelo dicionário analógico [IDS]. Os dicionários

visuais são ideias para extrair relações de parte-todo, visto que a totalidade do léxico presente neles denota objetos físicos discretos e suas partes.

Os CSs e CSAs coletados são apresentados em seis quadros, um para cada fonte de informação diferente. O cabeçalho de cada quadro identifica a fonte de informação de onde os CSs e CSAs foram extraídos, seguindo a mesma ordem em que elas são apresentadas no Apêndice II.1.

a) HOUAISS, A. <i>Dicionário Houaiss de Sinônimos e Antônimos</i> [HOU]
{cérebro, encéfalo},
{cólon, colo},
{fronte, cabeça},
{glânde, bálano},
{intestino, víscera},
{lábio, boca},
{nariz, narina},
{ouvido, orelha},
{peito, tórax},
{pele, odre},
{pescoço, cerviz},
{pescoço, colo},
{pulso, mão},
{punho, mão},
{seio, mama},
{testa, cabeça},
{testa, frente},
{testículo, gônada masculina},
{ventre, intestino},
{vulva, vagina},
{amígdala, agalha, tonsila},
{calcanhar, talão, coice},
{coração, peito, tórax},
{cílio, celha, pestana},
{flanco, ilharga, ilhal},
{lombo, nádega, bunda},
{mindinho, auricular, mínimo},
{miolo, medula, tutano},
{munheca, unha, pulso},
{olho, olhar, vista},
{panturrilha, barriga da perna, sura},
{partes pudendas, genitália, genitais},
{pé, chispe, pezunho},
{quartos, anca, quadril},
{seio, útero, ventre},
{sovaco, axila, sovaqueira},
{tímpano, membrana do ouvido, tambor},
{unha, casco, úngula},
{venta, narícula, narina},
{ventre, entranha, útero},
{ventre, estômago, barriga},
{víscera, entranha, órgão interno},

{abdômên, abdome, barriga, ventre},
{artelho, dedo do pé, pedartículo, pododáctilo},
{bacia, pelve, ilhargá, pélvis},
{bunda, cadeira, nádega, nalga},
{busto, seio, mama, peito},
{cachola, crânio, cabeça, testa},
{cara, rosto, face, frente},
{colo, peito, busto, torso},
{colo, pescoço, cérvix, cerviz},
{costas, dorso, costado, lombo},
{cútis, epiderme, pele, tez},
{dedo, dígito, gadanho, garra},
{dedo, mão, garra, unha},
{entranha, víscera, órgão interno, órgão},
{face, expressão, fisionomia, semblante},
{fuça, rosto, cara, face},
{gogó, maçã-de-adão, nó-de-adão, pomo-de-adão},
{madre, matriz, ventre, útero},
{membro, pênis, falo, fálus},
{orelha, aurícula, nambi, pavilhão auricular},
{palato, abóboda palatina, céu da boca, paladar},
{pinto, pênis, falo, fálus},
{pulso, punho, carpo, munheca},
{pálpebra, capela, párpado, prega},
{rabo, bunda, cadeira, nádega},
{rosto, fisionomia, aparência, feições},
{rosto, frente, dianteira, frente},
{sobaco, sovaco, axila, sovaqueira},
{torso, busto, talhe, tronco},
{venta, narícula, narina, nariz},
{víscera, míudos, frescura, tripa},
{anca, quadril, cadeira, ilhargá, nádega},
{barriga, abdômen, estômago, panturra, ventre},
{beicho, lábio, beicoca, beicola, beicorra},
{cocuruto, crânio, coco, moleira, testa},
{cuca, crânio, cabeça, caso, testa},
{moleira, bregma, cocuruto, fontanela, sincipúcio},
{papo, bócio, estruma, papeira, tireomegalia},
{sobrancelha, celha, sobrecílio, sobrolho, supercílio},
{artéria, vaso, vaso sanguíneo, canal, conduto, duto},
{barriga, abdômen, abdome, estômago, panturra, ventre},
{busto, tórax, peito, talhe, torso, tronco},
{cabeça, crânio, coco, cocuruto, moleira, testa},
{crânio, cérebro, cabeça, casco, inteligência, pensamento},
{cuca, cérebro, cabeça, intelecto, inteligência, raciocínio},
{garganta, goela, fauce, gasganete, gorja, pescoço},
{ombro, omoplata, costas, escápula, espádua, lombo},
{pança, abdômên, abdome, barriga, panturra, ventre},
{cara, expressão, aparência, feições, fisionomia, semblante, traço},
{cangote, nuca, cerviz, cachaço, toitiço, toutiço, cogote, congote},
{cabelo, cabeleira, cabelama, cabeladura, cabelame, cabelugem, coma, encabeladura, juba},
{cangote, cogote, cachaço, cérvix, cerviz, congote, gogote, nuca, pescoço}

b) POLITO, AG. <i>Michaelis Dicionário de Sinônimos e Antônimos</i> ; [MIC]
{amígdala, tonsila},
{artelho, dedo do pé},
{barriga da perna, batata da perna},
{calcanhar, talão},
{colo, regaço},
{coração, peito},
{céu da boca, palato},
{cílio, pestana},
{cólon, colo},
{estômago, bucho},
{fronte, testa},
{lábio, beiço},
{lábio, boca},
{mamilo, bico do peito},
{moleira, abóboda craniana},
{moleira, fontanela},
{munheca, pulso},
{nádega, nalga},
{orelha, ouvido},
{ouvido, audição},
{palma, palma da mão},
{papo, estômago},
{pescoço, colo},
{polegar, póllice},
{pulmão, bofe},
{pé, pata},
{testículo, bago},
{tornozelo, artelho},
{trompa, trompa de falópio},
{tronco, corpo},
{tímpano, ouvido},
{unha, úngula},
{ventre, útero},
{ânus, cu},
{úvula, campainha},
{axila, sovaco, sobaco},
{bacia, pelve, pélvis},
{busto, peito, seio},
{cabeça, crânio, coco},
{cadeira, quadril, anca},
{cara, aparência, aspecto},
{cara, fisionomia, semblante},
{cara, rosto, face},
{cuca, cabeça, inteligência},

{cútis, tez, pele},
{dente, canino, presa},
{esclerótica, branco do olho, clara},
{garganta, goela, fauce},
{glânde, extremidade do pênis, cabeça},
{gogó, pomo-de-adão, adão},
{gônada, testículo, ovário},
{ilharga, lado, ilhal},
{indicador, dedo indicador, fura-bolo},
{intestino, víscera, entranha},
{medula, tutano, miolo},
{mindinho, mínimo, dedo mínimo},
{miolo, cérebro, massa encefálica},
{mão, garra, unha},
{narina, fossa nasal, venta},
{narina, nariz, venta},
{nariz, narinas, venta},
{ombro, espádua, espalda},
{pai-de-todos, médio, dedo médio},
{papo, bócio, papeira},
{peito, tórax, torso},
{polegar, dedo polegar, mata-piolho},
{seu-vizinho, anular, dedo anular},
{supercílio, sobancelha, sobrolho},
{vagina, vulva, boceta},
{ventre, abdômen, abdome},
{ventre, barriga, pança},
{anca, quadril, cadeira, nádega},
{artéria, vaso, duto, conduto},
{bunda, nádega, bumbum, traseiro},
{busto, torso, tórax, peito},
{cabelo, cabeleira, coma, guedelha},
{cambito, perna fina, gambito, caniço},
{costas, dorso, costado, lombo},
{cutícula, epiderme, película, cútis},
{derme, derma, pele, couro},
{entranha, víscera, intestino, órgão interno},
{nariz, narigão, penca, bicanca},
{olho, olhar, vista, visão},
{peito, seio, mama, teta},
{pele, epiderme, película, cutícula},
{pinto, pênis, membro, cacete},
{polegar, dedo polegar, mata-piolho, pólce},
{rabo, nádega, traseiro, bunda},
{rosto, fisionomia, semblante, feições},

{útero, madre, matriz, ventre},
{barriga, abdômen, abdome, ventre, pança},
{bunda, nádega, bumbum, traseiro, rabo},
{epiderme, pele, cutícula, cúits, película},
{nuca, cangote, cachaço, toutiço, cogote},
{pênis, falo, membro, pinto, cacete},
{ventre, barriga, pança, abdomên, abdome},
{cangote, nuca, cachaço, toutiço, cogote, congote},
{ventre, barriga, pança, abdomêm, abdome, bucho}

c) FERNANDES, F. <i>Dicionário de Sinônimos e Antônimos da Língua Portuguesa</i> ; [FER]
{amígdala, tonsila},
{anular, seu-vizinho},
{axila, sovaco},
{bacia, pelve},
{barriga da perna, panturrilha},
{boca, lábio},
{busto, peito},
{cocuruto, cocuruta},
{colo, peito},
{colo, regaço},
{cólon, colo},
{esôfago, goelha},
{fronte, cabeça},
{fronte, testa},
{goela, traqueia},
{nádega, nalga},
{ombro, espádua},
{ouvido, orelha},
{peito, coração},
{peito, tórax},
{pele, couro},
{pele, epiderme},
{pele, odre},
{pescoço, cachaço},
{punho, pulso},
{quaril, alcatra},
{seio, colo},
{seio, peito},
{unha, úngula},
{venta, bochecha},
{artelho, tornozelo, maléolo},
{bunda, nádega, sesso},
{busto, tórax, arcabouço},

{cambito, pernil, perna fina},
{colo, pescoço, ombro},
{costas, dorso, lombo},
{cílio, celha, pestana},
{dedo médio, maior-de-todos, pai-de-todos},
{entranha, víscera, intestino},
{flanco, seio, ventre},
{goela, garganta, tragadeiro},
{gogó, pomo-de-adão, garganta},
{miolo, cérebro, massa encefálica},
{miolo, medula, tutano},
{munheca, pulso, mão},
{palato, céu da boca, paladar},
{papo, papeira, bócio},
{pescoço, colo, garganta},
{pulmão, bofe, voz},
{pulso, punho, mão},
{pé, pata, chispe},
{quadril, cadeira, anca},
{rosto, frente, fronte},
{torso, busto, tórax},
{tímpano, timbale, ouvido},
{tórax, torso, arcabouço},
{venta, olfato, nariz},
{ventre, útero, entranha},
{útero, madre, matriz},
{beiço, lábio, beiçola, beiçorra},
{cachola, cabeça, cachimônia, bestunto},
{calcanhar, tacão, coice, talão},
{cambito, perna, perna fina, perna magra},
{cangote, cachaço, cerviz, cogote},
{crânio, caveira, cérebro, inteligência},
{dente, presa, defesa, colmilho},
{epiderme, pele, tez, cútis},
{fuça, cara, focinho, venta},
{garganta, colo, pescoço, gasganete},
{garganta, goela, laringe, gorja},
{narina, venta, fossa nasal, narícula},
{nuca, cogote, toutiço, cachaço},
{olho, olhar, vista, percepção visual},
{papo, pança, bucho, estômago},
{vaso, caminho, duto, canal},
{vesícula, bolha, empola, bexiga},
{ânus, fiofó, viegas, ano},
{bucho, barriga, ventre, pança, bandulho},

{cabelo, coma, crina, gadelha, grenha},
{cara, rosto, face, aspecto, semblante},
{cútis, pele, epiderme, tez, cute},
{derme, derma, pele, couro, córion},
{flanco, vazio, ilharga, lado, ilhal},
{glãnde, bolota, bálano, lande, lãndea},
{indicador, índex, índice, mostrador, fura-bolos},
{mínimo, mindinho, minguinho, meiminho, auricular},
{nariz, penca, bitácula, tromba, venta},
{rosto, aparência, aspecto, feições, presença},
{sobrancelha, supercílio, sobrolho, sobreolho, sobrecenho},
{ventre, barriga, abdômen, adbome, pança},
{dedo, polegar, poléx, pólice, mata-piolho, cata-piolho},
{dedo polegar, polegar, poléx, pólice, mata-piolho, cata-piolho},
{nágega, cadeira, quadril, anca, coxa, ilharga},
{pomo-de-adão, nó na garganta, nó na goela, gogó, nó-de-adão, maçã-de-adão},
{barriga, ventre, abdômen, abdome, pança, bandulho, pandulho},
{traseiro, ânus, nádega, pódice, pousadeiro, rabo, sesso},
{rosto, cara, face, focinho, fuça, lata, visagem, semblante, fisionomia},
{cara, rosto, face, semblante, fisionomia, fachada, facha, frontispício, lata, tacho},
{cabeça, cachola, cachimônia, coco, bola, caco, cuia, chocolateira, bestunto, pinha, quengo, tola, touta, topete, sinagoga}

d) TEP 2.0 (http://www.nilc.icmc.usp.br/tep2/); [TEP]
{beicho, lábio},
{boca, lábio},
{busto, seio},
{busto, torso},
{calcanhar, talão},
{colo, pescoço},
{colo, regaço},
{crânio, caveira},
{cutícula, película},
{garganta, goela},
{mindinho, mínimo},
{orelha, aba},
{orelha, ouvido},
{panturrilha, sura},
{pele, membrana},
{pele, odre},
{polegar, pólex},
{unha, úngula},
{axila, sovaco, sobaco},
{cútis, pele, tez},
{derme, derma, pele},

{fronte, frente, testa},
{lombo, dorso, lombada},
{munheca, pulso, punho},
{peito, tórax, torso},
{pele, coiro, couro},
{pestanda, celha, cílio},
{sobrancelha, sobrolho, supercílio},
{tímpano, tambor, atambor},
{anca, cadeira, quadril, quartos},
{barriga, abdôm, abdome, ventre},
{barriga, abdômen, abdome, ventre},
{barriga, bojo, saliência, ventre},
{cocuruto, cocuruta, crista, grimpa},
{entranha, intestino, tripa, víscera},
{estômago, ânimo, disposição, fígado},
{fronte, face, frente, rosto},
{narina, aleta, narícula, venta},
{nuca, cachaço, toutiço, toitiço},
{unha, gadanho, gafa, garra},
{costas, cacunda, canastra, dorso, lombo},
{gogó, maçã-de-adão, nó-da-garganta, nó-de-adão, pomo-de-adão},
{peito, mama, poma, pomo, seio},
{rosto, aspecto, fisionomia, semblante, vulto},
{seio, madre, matriz, útero, ventre},
{cara, cariz, face, queixo, rosto, semblante, vulto}

e) AZEVEDO, F.F.S. – *Dicionário Analógico da Língua Portuguesa*; [AZE]

{pâncreas, baço},
{cílio, pestana, celha},
{sobrancelha, supercílio, sobrolho},
{útero, madre, matriz},
{rim, ureter, bexiga, uretra},
{tórax, petrina, torso, busto},
{papo, papeira, trasorelho, bócio, caxumba},
{garganta, laringe, fauce, goela, tragadeira, traqueia},
{nariz, beque, batata, bicanca, penca, focinho},
{barriga, abdômen, abdome, ventre, bandulho, pança, panturra},
{coração, artéria, veia, vaso capilar, sangue, linfa, plasma},
{moleira, cabeça, cabeça, cocuruto, sincipúcio, carrapito, copa},
{seio, teta, mama, peito, glândula mamal, mamilo, poma},
{víscera, miudezas, entranha, debulho, miúdos, intestino, tripa},
{coração, artéria, veia, vaso, vaso capilar, sangue, linfa, plasma},
{dente canino, dente do siso, dente molar, laníar, cabeiro, cartucheira, colmilho, presa, sobredente},
{encéfalo, cérebro, cerebelo, bulbo raquiano, bulbo raquidiano, ponte, corpo caloso, tálamo, hipotálamo},

{partes pudendas, genitália, sistema reprodutor, órgãos genitais, vagina, pito, clitóris, vulva, monte de vênus},
{gasnete, gasnate, gasganete, garganta, pescoço, cachaço, nuca, queixo, mento, pomo-de-adão},
{partes pudendas, genitália, sistema reprodutor, órgãos genitais, pênis, pipi, períneo, testículo, ovo, glândula do escroto},
{polegar, indicador, fura-bolo, dedo médio, maior-de-todos, dedo mínimo, mindinho, seu-vizinho, mata-piolhos, anular},
{polegar, indicador, fura-bolos, dedo médio, maior-de-todos, dedo mínimo, mindinho, seu-vizinho, mata-piolho, dedo anular},
{tegumento, tegmento, córion, pele, derma, indúvia, tez, epiderme, cútis, carão},
{boca, bocaça, bocarra, língua, céu da boca, palato, úvula, campainha, dente, lábio, beiço},
{gâmbia, perna, sanco, perônio, fíbula, tíbia, fêmur, joelho, panturrilha, barriga da perna, rótula},

g) SBS - <i>Dicionário Visual SBS do Corpo Humano</i> ; [SBS]
{mama, mamilo, aréola, glândula mamária},
{nariz, septo, narina, asa},
{derme, epiderme, pele, poro, pelo},
{rim, bexiga, uréter, uretra, glândula suprarrenal},
{cérebro, cerebelo, coluna vertebral, medula espinhal, filamento terminal, dura-máter},
{ombro, axila, braço, cotovelo, antebraço, pulso, mão}
{boca, gengiva, palato, úvula, arcada dentária, dente, língua, lábio},
{perna, coxa, panturrilha, joelho, pé, dedo do pé, tornozelo, dorso do pé, calcanhar}
{lábio maior, lábio menor, clitóris, vagina, colo do útero, útero, ovário, tuba uterina, vulva},
{orelha, aurícula, lóbulo, concha, hélice, anti-hélice, bigorna, martelo, vestibulo, cóclea},
{pênis, saco escrotal, glande, prepúcio, meato urinário, uretra, testículo, epidídimo, corpo cavernoso, próstata},
{boca, nariz, epiglote, faringe, laringe, corda vocal, esôfago, traqueia, pulmão, brônquio, diafragma},
{mão, palma, dorso, unha, dedo, dedo mínimo, dedo médio, dedo anular, dedo indicador, dedo polegar, pulso, lúnula},
{olho, pálpebra, cílio, pupila, íris, esclera, retina, córnea, nervo óptico, cristalino, conjuntiva, esclerótica},
{coração, artéria, veia, aorta, miocárdio, endocárdio, sangue, vaso, vaso sanguíneo, plasma, aurícula, ventrículo, glóbulo, plaqueta},
{boca, língua, glândula salivar, faringe, esôfago, estômago, fígado, pâncreas, vesícula biliar, duodeno, intestino grosso, intestino delgado, reto, esfíncter, ânus, cólon}

O Quadro 7.12 exemplifica o arranjo dos CSs e CSAs para as ULs *cabeça*, *face* e *fron*te e para as categorias conceituais S1, glosado por “parte superior do corpo humano”, e S2, glosado por “parte anterior da cabeça”. Para ver os arranjos completos dos CSs e CSAs dentro da matriz léxico-conceitual do *anaset* CORPO HUMANO, abra o arquivo *anaset_CORPO.csv* no programa *Microsoft Excel™*.

Quadro 7.12. Exemplo do *anaset* CORPO HUMANO preenchido com CSs e CSAs

MUNDO FÍSICO		
CORPO HUMANO	Relações Sintagmáticas	
	S1	S2
<i>cabeça</i>	{cérebro, cabeça, crânio, sede do pensamento} [IDS] {cabeça, cachola, cachimônia, coco, bola, caco, cuia, chocolateira, bestunto, pinha, quengo, tola, touta, topete, sinagoga} [FER] {cabeça, crânio, coco, cocuruto, moleira, testa} [HOU] {cabeça, crânio, coco} [MIC] {cabeça, sincipúcio, crânio, pericrânio, sensório, sensório comum, miolo, cérebro, tinote, cachimônia, bestunto, cabeçorra, tonta, bola, tola, mieleira, testo, cachola, toutiço, caco, carola, caveira, cerebelo} [AZE]	
<i>face</i>		{fronte, face, frente, rosto} [TEP] {cara, cariz, face, queixo, rosto, semblante, vulto} [TEP] {cara, rosto, face, aspecto, semblante} [FER] {face, expressão, fisionomia, semblante} [HOU] {fronte, rosto, cara, face} [HOU] {cara, rosto, face, semblante} [MIC] {face, semblante, rosto, cara, carão, caraça, verônica, frente, testa, testaça, presença, facha, focinho, lata, fuça, fisionomia} [AZE]
<i>fronte</i>	{fronte,cabeça} [FER] {fronte,cabeça} [HOU]	{fronte, frente, testa} [TEP] {fronte, face, frente, rosto} [TEP] {fronte, testa} [FER] {cara, rosto, face, frente} [HOU] {fronte, testa} [MIC] {testa, frente} [HOU] {face, semblante, rosto, cara, carão, caraça, verônica, frente, testa, testaça, presença, facha, focinho, lata, fuça, fisionomia} [AZE]

7.2.4 ETAPA 4 - CONSOLIDAÇÃO DOS RESULTADOS DO ANASET CORPO HUMANO

Com a matriz completamente preenchida com as ULs de entrada (7.2.1), as categorias conceituais (7.2.2) e os CSs e CSAs correspondentes (7.2.3), os resultados podem ser consolidados seguindo-se os procedimentos apresentados em 6.1.2.4. Lembrando que apenas os resultados para a consolidação de ordem onomasiológica é que serão exibidos aqui⁵⁸.

O Quadro 7.13 reproduz os resultados consolidados por meio do algoritmo de ordem onomasiológica. Na primeira coluna à esquerda temos os vinte e três campos

⁵⁸ Para ver a consolidação dos resultados de ordem semasiológica, é necessário abrir o arquivo CdSA _sem_CORPO.csv, presente no CD em anexo.

conceituais que estruturam o *anaset* CORPO HUMANO, seguidas na coluna adjacente por suas respectivas glosas. Na última coluna são apresentados os CdSAs correspondentes.

Quadro 7.13. Resultados consolidados para o *anaset* CORPO HUMANO

Resultados consolidados para o <i>anaset</i> CORPO HUMANO		
(S)	Glosa	CdSAs
Relação Sintagmática		
(S1)	Parte superior do corpo humano	{cabeça (1), cérebro (0.87), crânio (0.739), miolo (0.478), cerebelo (0.348), cachola (0.348), sincipúcio (0.304), caveira (0.304), testa (0.304), moleira (0.304), cocuruto (0.304), bestunto (0.261), cachimônia (0.261), bola (0.217), coco (0.217), caco (0.217), encéfalo (0.217), tola (0.217), inteligência (0.217), testo (0.174), toutiço (0.174), sensório (0.174), mioleira (0.174), pericrânio (0.174), sensório comum (0.174), tinote (0.174), tonta (0.174), cabeçorra (0.174), carola (0.174), [...]}
(S2)	Parte anterior da cabeça	{rosto (1), cara (0.897), face (0.828), fronte (0.793), semblante (0.621), fisionomia (0.483), testa (0.483), focinho (0.379), fuça (0.345), lata (0.276), frente (0.241), facha (0.241), presença (0.241), carão (0.207), caraça (0.207), testança (0.207), verônica (0.207), venta (0.172), aspecto (0.138), vulto (0.138), aparência (0.138), feições (0.138), queixo (0.103), cariz (0.103), [...]}
(S3)	Conjunto do pelos que crescem na cabeça	{cabelo (1), cabeleira (0.857), coma (0.714), cabeladura (0.571), pelo (0.571), cabelame (0.429), encabeladura (0.429), cabelama (0.286), guedelha (0.286), cabelugem (0.286), juba (0.286), penugem (0.143), grenha (0.143), gadelha (0.143), crina (0.143)}
(S4)	Órgão da visão	{olho (1), cílio (1), sobrancelha (0.909), sobrolho (0.909), supercílio (0.909), esclerótica (0.818), nervo óptico (0.727), conjuntiva (0.727), córnea (0.727), retina (0.727), íris (0.727), pupila (0.727), pálpebra (0.636), cristalino (0.636), esclera (0.545), celha (0.545), pestana (0.455), olhar (0.273), vista (0.273), lumes (0.182), menina do olho (0.182), humor aquoso (0.182), sobreolho (0.182), hialoide (0.182), úvea (0.182), corpo vítreo (0.182), sobrecílio (0.182), mácula (0.182), coroide (0.182), sobrecenho (0.182), [...]}
(S5)	Órgão da audição	{orelha (1), ouvido (1), tímpano (0.538), aurícula (0.385), vestíbulo (0.308), bigorna (0.308), martelo (0.308), cóclea (0.308), canal auditivo (0.231), aurículo (0.231), aparelho auditivo (0.231), conca (0.231), concha auditiva (0.231), pavilhão auricular (0.231), pavilhão (0.231), órgãos acústicos (0.231), labirinto (0.231), estribo (0.231), trágus (0.231), tambor (0.154), [...]}
(S6)	Órgão do olfato, parte inicial das vias respiratórias	{nariz (1), narina (0.923), venta (0.923), narícula (0.462), septo (0.231), asa (0.231), penca (0.231), bicanca (0.154), narinas (0.154), aleta (0.154), fossa nasal (0.154), [...]}
(S7)	Cavidade situada na cabeça, delimitada externamente pelos lábios e internamente pela faringe	{lábio (1), boca (0.704), palato (0.63), dente (0.593), beíço (0.556), úvula (0.556), língua (0.519), céu da boca (0.37), campainha (0.296), bocaça (0.259), bocarra (0.259), arcada dentária (0.259), gengiva (0.259), beijorra (0.148), beijola (0.148), beijoca (0.111), presa (0.111), [...]}
(S8)	Parte do corpo entre o tronco e a cabeça	{garganta (1), pescoço (1), goela (0.7), cachaço (0.55), colo (0.55), pomo-de-adão (0.5), nuca (0.45), gasganete (0.4), fauce (0.4), gogó (0.4), cogote (0.35), gorja (0.3), toutiço (0.3), cerviz (0.3), nó-de-adão (0.25), maçã-de-adão (0.25), laringe (0.25), cangote (0.25), papeira (0.2), bócio (0.2), papo

		(0.2), traqueia (0.15), gasnete (0.15), amígdala (0.15), congote (0.15), tonsila (0.15), gasnate (0.1), queixo (0.1), adão (0.1), cérvix (0.1), mento (0.1), toitiço (0.1), nó-da-garganta (0.1), tragadeira (0.1), [...]
(S9)	Sistema circulatório	{coração (1), artéria (0.833), vaso (0.833), sangue (0.5), peito (0.5), veia (0.5), plasma (0.5), duto (0.5), vaso sanguíneo (0.333), canal (0.333), conduto (0.333), linfa (0.333), tórax (0.167), aurícula (0.167), ventrículo (0.167), glóbulo (0.167), plaqueta (0.167), endocárdio (0.167), caminho (0.167), vaso capilar (0.167), vaso capilar (0.167), aorta (0.167), miocárdio (0.167)}
(S10)	Parte do tronco que vai do pescoço ao abdome	{peito (1), torso (0.95), tórax (0.95), busto (0.75), colo (0.25), tronco (0.2), petrina (0.2), coração (0.15), arcabouço (0.15), talhe (0.15), regaço (0.15), seio (0.1), [...]}
(S11)	Sistema respiratório	{pulmão (1), traqueia (1), esôfago (0.875), laringe (0.875), brônquio (0.75), boca (0.75), corda vocal (0.75), diafragma (0.75), epiglote (0.75), nariz (0.75), faringe (0.75), bofe (0.25), vias respiratórias (0.125), tragadeira (0.125), goela (0.125), fauce (0.125), garganta (0.125), voz (0.125)}
(S12)	Parte do corpo humano entre o tórax e a pelve	{abdome (1), barriga (1), ventre (1), pança (0.727), estômago (0.455), panturra (0.318), bandulho (0.273), bucho (0.227), papo (0.182), [...]}
(S13)	Sistema digestivo	{estômago (1), cólon (0.864), esôfago (0.818), faringe (0.773), intestino (0.773), ânus (0.727), reto (0.727), víscera (0.682), pâncreas (0.591), entranha (0.591), fígado (0.591), intestino delgado (0.545), língua (0.545), vesícula biliar (0.545), duodeno (0.545), glândula salivar (0.545), boca (0.545), esfíncter (0.545), intestino grosso (0.545), tripa (0.5), golelha (0.227), canal intestinal (0.182), piloro (0.182), jejuno (0.182), ceco (0.182), lenço (0.182), peritônio (0.182), miúdos (0.136), barriga (0.136), ventre (0.136), debulho (0.136), órgão interno (0.136), colo (0.136), miudezas (0.136), [...]}
(S14)	Sistema urinário	{bexiga (1), ureter (0.857), rim (0.857), uretra (0.857), glândula suprarrenal (0.429), bolha (0.143), vesícula (0.143), empola (0.143)}
(S15)	Parte correspondente ao peito no corpo feminino	{seio (1), peito (0.941), mama (0.706), busto (0.353), colo (0.353), poma (0.353), teta (0.353), mamilo (0.294), regaço (0.235), glândula mamal (0.176), pomo (0.118), [...]}
(S16)	Membros superiores	{axila (1), sovaco (0.75), cotovelo (0.5), ombro (0.5), antebraço (0.5), braço (0.5), sobaco (0.313), articulação (0.25), cúbito (0.25), ulna (0.25), úmero (0.25), rádio (0.25), charneira (0.25), gínglimo (0.25), mão (0.25), pulso (0.25), espádua (0.188), sovaqueira (0.125), [...] }
(S17)	Extremidade do membro superior, articulada com o antebraço pelo punho e terminada pelos dedos	{mão (1), pulso (1), unha (0.75), dedo médio (0.714), dedo (0.643), dedo mínimo (0.607), munheca (0.571), polegar (0.536), mindinho (0.464), dedo anular (0.464), palma (0.429), dedo polegar (0.429), indicador (0.393), punho (0.393), seu-vizinho (0.393), mata-piolho (0.357), dedo indicador (0.357), maior-de-todos (0.357), dorso (0.286), lúnula (0.286), anular (0.286), carpo (0.25), fura-bolo (0.214), palma da mão (0.214), pólice (0.214), mínimo (0.214), pai-de-todos (0.214), raia (0.179), sabugo (0.179), metacarpo (0.179), fura-bolos (0.179), úngula (0.143), garra (0.143), médio (0.143), mata-piolhos (0.107), auricular (0.107), cata-piolho (0.107), poléx (0.107), falange (0.107), [...]}
(S18)	Membros inferiores	{panturrilha (1), perna (1), barriga da perna (0.625), joelho (0.5), perna fina (0.5), cambito (0.5), sura (0.375), caniço (0.25), sanco (0.25), gambito (0.25), gâmbia (0.25), perônio (0.25), pé (0.25), fíbula (0.25), tornozelo (0.25), dorso do pé (0.25), calcanhar (0.25), dedo do pé (0.25), coxa

		{(0.25), tibia (0.25), fêmur (0.25), rótula (0.25), pernil (0.125), batata da perna (0.125), perna magra (0.125)}
(S19)	Extremidade do membro inferior abaixo da articulação do tornozelo e terminada pelos artelhos	{tornozelo (1), pé (0.889), artelho (0.889), calcanhar (0.778), dedo do pé (0.556), maléolo (0.444), talão (0.444), perna (0.333), panturrilha (0.333), dorso do pé (0.333), coxa (0.333), joelho (0.333), coice (0.222), pata (0.222), prancha (0.222), patola (0.222), metatarso (0.222), toesa (0.222), pesunho (0.222), chispe (0.222), patela (0.222), chanca (0.222), canastra (0.222), pezunho (0.111), tacão (0.111), pododáctilo (0.111), pedartículo (0.111)}
(S20)	Parte posterior do tronco	{lombo (1), costas (0.929), ilharga (0.786), dorso (0.714), cadeira (0.643), anca (0.643), quadril (0.643), ilhal (0.429), costado (0.357), flanco (0.357), região lombar (0.286), lado (0.286), rim (0.286), hipocôndrio (0.286), ísquio (0.286), costão (0.286), quartos (0.143), canastra (0.143), nádega (0.143), vazio (0.143), cacunda (0.143), [...]}
(S21)	Parte do corpo humano da cintura à articulação da coxa	{nádega (1), cadeira (0.875), quadril (0.813), anca (0.781), bunda (0.688), rabo (0.563), nalga (0.5), ânus (0.469), pódice (0.469), sesso (0.438), bumbum (0.438), pousadeiro (0.406), cu (0.344), lombo (0.344), alcatra (0.344), traseira (0.313), rabiosque (0.313), assento (0.313), culatra (0.313), costaneira (0.313), pousadouro (0.313), reiras (0.313), rabiote (0.313), rabioste (0.313), reto (0.313), fundilho (0.313), tarso (0.313), rabeira (0.313), rabicho (0.313), ás de copas (0.313), rabadela (0.313), rabadilha (0.313), rabisteco (0.313), posterioridade (0.313), rabada (0.313), cola (0.313), cauda (0.313), traseiro (0.281), quartos (0.188), ilharga (0.125), [...]}
(S22)	Sistema reprodutor feminino	{útero (1), ventre (0.769), vulva (0.692), vagina (0.692), matriz (0.538), madre (0.538), clitóris (0.385), seio (0.385), partes pudendas (0.308), genitália (0.308), monte de vênus (0.231), sistema reprodutor (0.231), pito (0.231), órgãos genitais (0.231), ovário (0.231), lábio maior (0.154), colo do útero (0.154), boceta (0.154), tuba uterina (0.154), lábio menor (0.154), entranha (0.154), [...]}
(S23)	Sistema reprodutor masculino	{pênis (1), testículo (0.833), glande (0.667), epidídimo (0.417), uretra (0.417), meato urinário (0.417), corpo cavernoso (0.417), saco escrotal (0.417), prepúcio (0.417), próstata (0.417), genitália (0.333), partes pudendas (0.333), ovo (0.25), pipi (0.25), glândula do escroto (0.25), órgãos genitais (0.25), sistema reprodutor (0.25), falo (0.25), membro (0.25), pinto (0.25), períneo (0.25), bálano (0.167), cacete (0.167), fálus (0.167), [...]}
(S24)	Membrana mais ou menos espessa que reveste o corpo humano	{pele (1), epiderme (0.531), cútis (0.5), tez (0.438), derme (0.344), derma (0.313), córion (0.188), cutícula (0.188), película (0.188), indúvia (0.156), carão (0.156), tegmento (0.156), couro (0.156), tegumento (0.156), [...]}

7.3 MODELAGEM DO ANASET TEATRO

Essa Seção relata a aplicação do modelo no domínio léxico-conceitual TEATRO. Para tanto, foram organizadas subseções para cada uma das etapas de construção do *anaset* equivalente. Em 7.3.1 temos a identificação e coleta das ULs do domínio TEATRO. Em 7.3.2 procedemos com a estruturação conceitual do *anaset* TEATRO. Em 7.3.3 executamos a identificação, coleta e arranjo dos CSs e CSAs pertinentes ao *anaset* TEATRO. Por fim, em 7.3.4 temos a consolidação dos resultados, em que apresentamos uma tabela que representa a modelagem final do *anaset* TEATRO.

7.3.1 ETAPA 1 - IDENTIFICAÇÃO E COLETA DAS ULs DO DOMÍNIO TEATRO

Essa Seção tem como objetivo arrolar as ULs de entrada para o *anaset* TEATRO. Para tanto, partimos da lista de ULs extraídas do *corpus* (cf. 6.2). As vinte e uma ULs do domínio TEATRO extraídas do *corpus* estão elencadas no Quadro 7.14.

Quadro 7.14. ULs do domínio TEATRO extraídas do *corpus*

Domínio léxico-conceitual TEATRO (21 ULs)
pintura
atriz
dança
papel
cenário
comédia
pagantes
bis
artista
palmas
palco
cortina
coxia
palhaço
corista
arquibancada
maestro
star
cinema
dançarina
bailarina

Em seguida, foram consultadas as fontes de informação léxico-conceituais arroladas em 6.3.1 para identificar ULs sinônimas e análogas que possam também

pertencer ao mesmo domínio. Dessa forma, procedeu-se à expansão do número de ULs que deverá ocupar a coluna da extrema esquerda na matriz léxico-conceitual do *anaset* correspondente ao domínio léxico-conceitual TEATRO.

Conforme os critérios apontados em 6.1.2.1, não foram incluídas na lista as ULs complexas, estrangeirismos e nomes próprios. Tampouco foram incluídas na lista as ULs estritamente técnicas e/ou de cunho essencialmente enciclopédico. Como exemplo das ULs que **não** foram incluídas na lista de ULs de entrada do *anaset* TEATRO estão: *artista dramático, cena lírica, peça teatral, star, jeune premier, avant-première, costumier, happening, gag, clown, Melpômene, Tália, Tépsis, patafísica, peça-problema, memória emocional, crítica formalista etc.*

Ainda com relação os critérios apontados em 6.1.2.1, limitamos a expansão da coleta das ULs do domínio léxico-conceitual TEATRO principalmente pelos domínios léxico-conceituais adjacentes: REPRESENTAÇÃO, ARTES, ARTISTA, INTÉRPRETE, ARENA e DIVERTIMENTO.

Seguindo a metodologia e os critérios estabelecidos em 6.1.2.1, chegamos a uma lista de 107 ULs de entrada para o *anaset* TEATRO, conforme o Quadro 7.15.

Quadro 7.15. ULs de entrada do *anaset* TEATRO

Domínio léxico-conceitual TEATRO (107 ULs)			
acrobata	comediante	extra	montagem
arlequim	comediógrafo	fala	ópera
artista	comparsa	fantoches	palco
ato	contrarregra	farsante	palhaço
ator	coro	farsista	palmas
atriz	cortina	farsola	pantomima
audiência	dançarino	figurante	papel
autômato	dançariz	figurinista	peça
bailador	debutante	figurino	pelotiqueiro
bailarim	desenlace	funâmbulo	personagem
bailarina	desfecho	galã	pierrô
bailarino	diretor	gerente	pintura
bastidores	drama	ginasta	plateia
bis	dramatologia	herói	polichinelo
boneco	dramaturgia	histrião	prólogo
bonifrate	dramaturgo	interlúdio	protagonista
bufão	elenco	intermédio	público
bufo	empresário	intérprete	representação
cabotino	encenação	intervalo	saltatriz
camarim	entreato	jogral	saltimbanco

camarote	entremezista	malabarista	teatro
cantor	epílogo	mamulengo	texto
cena	equilibrista	maquiador	títere
cenógrafo	espectador	maquiagem	tragédia
chocarreiro	espetáculo	maquinista	truão
clímax	estreia	marionete	
colombina	estrela	mímica	
comédia	estrelato	monólogo	

7.3.2 ETAPA 2 - ESTRUTURAÇÃO CONCEITUAL DO ANASET TEATRO

No Dicionário Analógico da língua portuguesa (AZEVEDO, [1950] 2010) o equivalente ao nosso domínio TEATRO é a entrada **599. Drama**. No quadro sinóptico de categorias a entrada se encontra na ponta da seguinte ramificação: **classe IV. Entendimento> divisão II. Comunicação das Ideias> 3º) Meios> II. Meios convencionais> c. Linguagem escrita> 599. DRAMA**. Na WN.Pr, o conceito prototípico TEATRO (THEATRE⁵⁹), definido pela glosa “a arte de escrever e produzir peças”, é hipônimo direto do conceito COMUNICAÇÃO (COMMUNICATION⁶⁰) e tem perto do topo da sua hierarquia, assim como o conceito prototípico de AMOR, o conceito de ABSTRAÇÃO. Na EWN, o conceito de TEATRO é também um hipônimo do conceito COMUNICAÇÃO (COMMUNICATION) que, por sua vez, é entendido como uma experiência mental (EXPERIENCE/MENTAL) e, portanto, é categorizado no topo da ontologia da EWN como uma ENTIDADE DE 2ª ORDEM (2ND ORDER ENTITY).

No HTOED, o conceito TEATRO (THEATRE) não é categorizado como hipônimo do conceito COMUNICAÇÃO, mas sim dentro da categoria ARTE PERFORMÁTICA (PERFORMANCE ART) que, por sua vez, alinha-se diretamente abaixo da categoria LAZER/ARTE (LEISURE/THE ARTS). As duas categorias – COMUNICAÇÃO e LAZER/ARTE – estão no mesmo nível dentro da hierarquia do HTOED e pertencem à Grande Categoria MUNDO SOCIAL (THE SOCIAL WORLD). O esquema abaixo mostra a hierarquia conceitual do conceito TEATRO dentro da estrutura do HTOED:

⁵⁹ {07019235} <noun.communication>S: (n) dramaturgy, dramatic art, dramatics, theater, theatre (the art of writing and producing plays)

⁶⁰ {06262268} <noun.communication>S: (n) communication, communicating (the activity of communicating; the activity of conveying information)

03 The social world

03.11 Leisure/The Arts

03.11.03 Performance Arts

03.11.03.04 Theatre

Diferentemente das demais estruturas conceituais anteriormente citadas, o HTOED não organiza o conceito TEATRO como uma entidade abstrata e puramente mental, ou seja, como uma faculdade intelectual do ser humano. De fato, não é possível compreender o domínio TEATRO independentemente de um contexto social e cultural complexo. Portanto, dentro do nosso modelo, o domínio conceitual TEATRO ancora-se na Grande Categoria MUNDO SOCIAL, conforme o Quadro 7.16.

Quadro 7.16. Matriz preenchida com a Grande Categoria onde se inclui o *anaset* TEATRO: MUNDO SOCIAL

MUNDO SOCIAL						
TEATRO	Relações Paradigmáticas			Relações Sintagmáticas		
	S1	S2	S3	EF1	EF2	
				S4	S5	S6
UL 1	{cs1} [X]					
	{cs2} [Y]					
	{cs3} [Z]					

Ao consultar por *teatro* nas obras de referência semasiológicas, identificamos três núcleos de sentido (um sentido prototípico e dois sentidos por extensão metonímica) e outros três sentidos por extensão metafórica dos três primeiros. Esses últimos, não serão modelados nesse trabalho, na medida em que pertencem a outros domínios conceituais como veremos adiante. Por isso, para não interferir no padrão de representação das categorias conceituais que serão modeladas (marcadas pela letra 'S'), os conceitos metafóricos que apontam para outros domínios conceituais serão marcados pela letra 'M' (de **M**etafóricos). Vejamos:

[CAM] **1** Edifício onde se encenam peças teatrais e espetáculos afins; **2** Arte de representar; **3** Conjunto das obras dramáticas de uma época (teatro medieval), de um país (teatro brasileiro), de um autor (teatro de Martins Penna), de uma corrente estética (teatro realista).

[AUL] **1** *Teat.* Local com palco próprio para encenação de peças, óperas, recitais etc.; **2** *Teat.* Grande anfiteatro onde eram realizados jogos e espetáculos públicos na antiga Roma; ANFITEATRO; CIRCO; **3** *Teat.* A arte de representar; **4** Profissão da pessoa que representa peças teatrais; **5** *Liter.* A literatura escrita para ser encenada; coleção das obras dramáticas de um autor, ou dos atores de um país; DRAMATURGIA; **6** *Fig.* Obra, documento ou tipo de manual escrito para instruir sobre certos princípios; exemplo, modelo; regra; **7** *Fig.* Lugar onde se passa um acontecimento importante, notável; CENÁRIO; PALCO; **8** *Fig.* Aparência vã, miragem, ilusão; **9** Atitude falsa; FINGIMENTO

[MIC] **1** Casa ou lugar destinado à representação de obras dramáticas, óperas ou outros espetáculos públicos. **2** Circo, anfiteatro. **3** Conjunto das obras dramáticas de um autor. **4** Coletânea das obras dramáticas de uma nação. **5** Literatura ou arte dramática. **6** A arte de compor obras dramáticas ou de representá-las. **7** A profissão de ator ou de atriz. **8** Lugar onde se verifica qualquer acontecimento notável. **9** Aparência vã, miragem, ilusão. **10** Obra escrita para instruir sobre certos princípios; exemplo, modelo, regra.

[HOU] **1** Lugar ou edifício destinado à apresentação de obras dramáticas, óperas ou outros espetáculos públicos; **2** (*derivação por metonímia*) O ofício ou a arte teatral; **3** conjunto das obras dramáticas de um autor, de uma época, de um país; **4** literatura ou arte dramática; **5** (*derivação por metáfora*) Local onde se passa algum acontecimento notável; palco; **6** (*derivação por metáfora*) falsa realidade, aparência vã; ilusão, miragem; **7** fingimento, hipocrisia que se exterioriza com dramaticidade

Podemos aplicar a filtragem abaixo para chegarmos a apenas seis sentidos:

S1: [CAM] (1) = [AUL] (1+2) = [MIC] (1+2) = [HOU] (1) = teatro (edifício, local)

S2: [CAM] (2) = [AUL] (3+4) = [MIC] (5+6+7) = [HOU] (2) = teatro (arte, técnica)

S3: [CAM] (3) = [AUL] (5) = [MIC] (3+4+5) = [HOU] (3+4) = teatro (obra, produto)

M1: [AUL] (7) = [MIC] (8) = [HOU] (5) = ext. metafórica de S1 = lugar de acontecimentos notáveis

M2: [AUL] (8+9) = [MIC] (9) = [HOU] (6+7) = ext. metafórica de S2 = falsidade, fingimento

M3: [AUL] (6) = [MIC] (10) = ext. metafórica de S3 = exemplo, modelo, regra

A palavra *teatro* no sentido de M1 (lugar onde se desenrola um acontecimento importante) pode fazer parte de um *synset* como {palco, cena, cenário, teatro}, com a atestação em *corpus* das seguintes expressões: “palco de grandes acontecimentos”, “cenário de muitas lutas” e “foi teatro de muitos e variados acontecimentos”. No entanto, a frequência em *corpus* da UL *teatro* no sentido de M1 mostra-se infinitamente menor do que a frequência dos demais sentidos, exceto o sentido M3, para o qual não se achou atestação. Os sentidos de M1, M2 e M3 – extensões metafóricas de S1, S2 e S3, respectivamente – redirecionam (apontam) para outros domínios conceituais. O sentido

M2, por exemplo, deve apontar para o domínio conceitual FALSIDADE. Dessa forma, os sentidos nucleares S1, S2 e S3 são as categorias conceituais em relação paradigmática que serão modeladas no interior do domínio TEATRO.

A categoria conceitual S1, glosada por “edifício destinado à apresentação de obras dramáticas e outros espetáculos públicos”, é uma categoria com estrutura interna. Ela se estrutura internamente como o domínio CORPO HUMANO, ou seja, um todo percebido conceitualmente como dividido em partes. O conceito PALCO não é um gênero do conceito TEATRO (S1), mas uma parte dele (Cf. 3.3). No entanto, ao contrário da estrutura conceitual CORPO HUMANO, TEATRO (S1) não é uma classe natural, e sim um artefato socialmente produzido, modificado ao longo de séculos de história humana, sendo, a cada momento, concebido segundo certas condições culturais, materiais e econômicas. Segundo o *Dicionário de Teatro* (VASCONCELLOS, 2001), por exemplo, a arquitetura do Teatro Grego era estruturada em dois segmentos separados, o ‘Théatron’ e a ‘Skené’, ligados por um grande espaço reservado à orquestra (chamado ‘Orchéstra’), e a capacidade do auditório era de cerca de 20 mil pessoas. Já a arquitetura do Teatro Elisabetano proporciona um espaço bem menor para a orquestra, inova no formato do palco e tinha uma capacidade que variava de 1.500 a 2 mil espectadores. Hoje, a maioria dos teatros não possui mais o fosso da orquestra e a lotação pode variar de algumas dezenas para algumas centenas de lugares.

A categoria conceitual TEATRO (S1) é estruturada como um *frame* físico-social, ou seja, como partes físicas que, interligadas, compõe um todo socialmente construído. Pela granularidade adotada, o *frame* físico-social de TEATRO (S1) abrange apenas duas categorias conceituais: PARTES DO TEATRO (S4) e PLATEIA (S5). A categoria S4 compreende ULs como, por exemplo, *palco, cenário, prosscênio, cortina, bastidores e ribalta*. A categoria S5 compreende ULs como *plateia, auditório, camarote, poltronas, galeria* etc. As categorias conceituais S4 e S5 correspondem, respectivamente, aos EF1 e EF2 do *frame* Teatro_prédio.

A categoria conceitual S3, glosada por “literatura ou arte dramática”, é também uma categoria com estrutura interna. O *frame* Teatro_texto, abrange quatro categorias conceituais: AUTOR DRAMÁTICO (S6), GÊNEROS DE OBRA DRAMÁTICA (S7), PARTES DA OBRA DRAMÁTICA (S8) e FIGURA DRAMÁTICA (S9). A categoria S6 compreende ULs como, por exemplo, *dramaturgo, comediógrafo e entremezista*. A categoria S7 abarca ULs como *tragédia, comédia, pantomima, ópera* etc. A categoria S8

emgloba ULs como, por exemplo, *prólogo*, *epílogo*, *entreato* e *clímax*. E, por fim, a categoria S9 contém ULs como *personagem*, *herói*, *galã* e *prima-dona*, por exemplo. As categorias conceituais S6, S7, S8 e S9 correspondem, respectivamente, aos EF1, EF2, EF3 e EF4 do *frame* *Teatro_texto*.

Para modelar a categoria conceitual S2, glosada por “o ofício ou a arte teatral”, tomamos como inspiração o *frame* *Performing_arts* da FrameNet, já que esta não possui um *frame* específico *Theatrical_performing*. Dessa forma, fizemos uma adaptação do cabeçalho (*Definition*) do *frame* *Performing_arts* para o nosso *frame* específico *Espectáculo_teatral* e identificamos os respectivos EFs, conforme (i). Ao lado de cada EF aparece a sua categoria conceitual correspondente. Os *frames* anteriormente citados (*Teatro_prédio* e *Teatro_texto*), também aparecem citados no *frame* *Espectáculo_teatral*, já que estão intimamente vinculados a ele. A categoria conceitual S2 é o protótipo do domínio TEATRO e seu *frame* correspondente é o que estrutura todas as relações léxico-conceituais do domínio.

- (i) Os artistas [EF1/S10], juntamente com pessoal por trás das cenas [EF2/S11], executam um espetáculo teatral [S2] valendo-se de determinado espaço físico [*Teatro_prédio*] e de acessórios e/ou objetos [EF3/S12] adequados para tal. Tal espetáculo se desenrola de acordo com um script [*Teatro_texto*]. O objetivo da performance é criar uma experiência para um público que então irá julgar os seus méritos [EF4/S13].

O Quadro 7.17 apresenta a matriz léxico-conceitual do *anaset* TEATRO preenchida com a estrutura conceitual discutida ao longo dessa seção.

Quadro 7.17. Matriz do *anaset* AMOR preenchida com a estrutura conceitual

MUNDO SOCIAL													
TEATRO	Relações Paradigmáticas			Relações Sintagmáticas									
				Teatro_prédio		Teatro_texto				Espectáculo_teatral			
	S1	S2	S3	EF1	EF2	EF1	EF2	EF3	EF4	EF1	EF2	EF3	EF4
				S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
UL 1													

O Quadro 7.17 complementa-se com as informações presentes no Quadro 7.18, em que as categorias conceituais estão instanciadas por uma pequena glosa e por

CSs/CSAs representativos. Além disso, a última coluna apresenta abonações extraídas do *corpus* do NILC ou diretamente retiradas das fontes de informação lexical.

Quadro 7.18. Estrutura conceitual do *anaset* TEATRO

Domínio conceitual TEATRO			
(S)	Glosa	CSs/CSAs	Exemplo
Relação Paradigmática			
(S1)	edifício destinado à apresentação de obras dramáticas e outros espetáculos públicos	{teatro, anfiteatro} [MIC] {theater, theatre, house} [WN]	“O centro abriga um teatro municipal com 320 lugares, biblioteca e o cine clube Sétima Arte.” [NILC]
(S2)	o ofício ou a arte teatral	{teatro, drama} [TEP] {dramaturgy, dramatic art, dramatics, theater, theatre} [WN]	“O teatro é, no mínimo, tão importante quanto o cinema.” [NILC]
(S3)	literatura ou arte dramática	{peça, peça teatral, texto} [MIC] {drama} [WN]	“Mais tarde, desenvolveram-se outros gêneros: o romance, o teatro , a crônica.” [NILC]
Relação Sintagmática			
(S4)	partes do teatro	{palco, vistas, pano de boca, telão, tablado, prosênio, rampa, estrado, ribalta, gambiarra, bambolina, ..., trampolim, trapézio} [AZE] {stage, the boards, acting area, playing area, thrust stage, forestage, apron, passerelle, proscenium, ..., fire curtain, stage door} [RIT]	“É mais provável revê-la no palco como atriz.” [NILC]
(S5)	público	{plateia, bancada, arquibancada, torrinha, cadeiras, poltronas, frisas, camarote, poleiro, galeria, público, auditório} [IDS] {auditorium, seating, parquet, orchestra, pit, box, box seat, loge, ..., balcony, mezzanine} [RIT]	“Na época de Shakespeare, a ralé ficava na plateia , os burgueses nos balcões e os nobres nos camarotes , como um desfile de escola de samba.” [NILC]
(S6)	autor de peças dramáticas	{dramaturgo, autor dramático, comediógrafo, entremezista, mimógrafo} [AZE] {dramatist, playwright} [WN]	“Descobriu níveis desconhecidos no maior dramaturgo brasileiro, Nelson Rodrigues.” [NILC]
(S7)	gêneros de obra dramática	{drama, tragédia, coturno, comédia, ópera, grande ópera, ópera séria, opereta, zarzuela, cena cômica, cena lírica, ..., folhetim, épico} [AZE] {tragedy, tragic drama, melodrama, tragic flaw, buskin, ..., cothurnus, tragic muse} [RIT]	“É uma comédia e um drama ao mesmo tempo.” [NILC]
(S8)	partes de obra dramática	{êxodo, cenário, quadro, introdução, entrecho, epílogo, prólogo, libreto} [AZE] {act, scene, number, turn, bit, shtick, routine, introduction, expository scene, monologue, prologue,	“O motor da ação dramática, anuncia-o o protagonista no prólogo [...]” [NILC]

		epilogue, ..., encore, ovation} [RIT]	
(S9)	figura dramática	{primeiro trágico, prima-dona, herói, protagonista, personagem, debutante, estreante, galã, amoroso, ingênua, pai nobre, ator, atriz} [AZE] {role, part, piece, cue, lines, side, cast, character, person, personage, lead, lady, hero, heroine, ..., stock part, actor} [RIT]	“Não é como construir um personagem de Shakespeare”
(S10)	o conjunto de artistas ligados a um espetáculo ou a uma companhia	{ator, artista, astro, estrela, personagem, figurante, corista, extra, cantor, dançarino, ..., comediante, humorista} [IDS] {actor, actress, player, stage player, performer, playactor, histrion, histrio, thespian, ..., matinee idol, romantic lead} [RIT]	“Mas como ator , se você está realmente transmitindo emoções, você nunca é ridículo.” [NILC]
(S11)	equipe de um espetáculo teatral	{fornecedor, maquinista, ponto, apontador, figurinista, maquiador, diretor, ..., cenógrafo, fotógrafo} [AZE] {theater man, theatrician, showman, director, auteur, stage director, stage manager, costume designer, ..., barker, ballyhoo man} [RIT]	“No palco, cenógrafo , dramaturgo, coreógrafos e o próprio diretor têm papéis secundários.” [NILC]
(S12)	acessórios ou instrumentos cênicos necessários à produção de um filme, uma peça teatral	{make-up, maquiagem, maquilagem, pintura} [TEP] {property, prop, practical piece, handprop, costume, theatrical makeup, makeup, ..., clown white, spirit gum} [RIT]	“Dois outros aspectos importantes: maquiagem e figurino .” [NILC]
(S13)	aclamação	{palmas, aplausos, aclamação, saudação, ovação, vivas} [HOU] {applause, plaudit, éclat, acclaim, acclamation, clap, handclap, ..., ovation, standing ovation} [RIT]	“Os aplausos foram entusiasmados.” [NILC]

7.3.3 ETAPA 3 - IDENTIFICAÇÃO, COLETA E ARRANJO DOS CSS E CSAs DO ANASET TEATRO

As células da matriz léxico-conceitual serão preenchidas com os CSs e CSAs identificados e coletados nas fontes de informação léxico-conceitual.

Os CSs e CSAs coletados são apresentadas em seis quadros, um para cada fonte de informação diferente. O cabeçalho de cada quadro identifica a fonte de informação de onde os CSs e CSAs foram extraídos, seguindo a mesma ordem em que elas são apresentadas no Apêndice II.1.

a) HOUAISS, A. <i>Dicionário Houaiss de Sinônimos e Antônimos</i> [HOU]
{artista, intérprete},
{bastidores, coxia},
{bis, repetição},
{cenógrafo, cenarista},
{coro, coral},
{drama, tragédia},
{epílogo, paralipômenos},
{equilibrista, malabarista},
{fala, texto},
{funâmbulo, acrobata},
{ginasta, acrobata},
{herói, protagonista},
{mamulengo, teatro de bonecos},
{palco, caixa de cena},
{papel, personagem},
{prólogo, prefácio},
{truão, bobo},
{ato, parte, divisão},
{cena, palco, teatro},
{comediante, cômico, humorista},
{comediante, pelotiqueiro, saltimbanco},
{comparsa, figurante, extra},
{estreia, inauguração, abertura},
{estrela, astro, galã},
{monólogo, monodrama, solilóquio},
{mímica, pantomima, representação},
{palco, teatro, arte teatral},
{palhaço, fantoche, títere},
{personagem, intérprete, papel},
{plateia, auditório, sala},
{audiência, espectadores, ouvintes, público},
{cabotino, comediante, palhaço, saltimbanco},
{cena, cenário, decoração, palco},
{drama, melodrama, dramalhão, novela},
{drama, peça, obra, texto},
{estreia, debute, começo, início},
{figurino, vestuário, indumentária, traje},
{protagonista, herói, ator, intérprete},
{arlequim, bufão, farsante, palhaço, truão},
{bufão, arlequim, bobo, bufo, palhaço},
{encenação, montagem, apresentação, dramatização, representação},
{entreato, entrecena, interlúdio, intermédio, intervalo},

{funâmbulo, marionete, bonifrate, fantoche, títere},
{palco, estrado, tablado, jirau, plataforma},
{peça, peça teatral, montagem, encenação, representação},
{títere, marionete, bonifrate, fantoche, presepe},
{circo, anfiteatro, arena, coliseu, estádio, pavilhão},
{espetáculo, encenação, cena, montagem, peça, representação},
{farsante, brincalhão, galhofeiro, gracejador, palhaço, trocista},
{histrião, brincalhão, farsante, farsista, gracejador, palhaço},
{mamulengo, fantoche, boneco de engoço, marionete, presepe, títere},
{palmas, aplausos, aclamação, saudação, ovação, vivas},
{pantomima, palhaçada, arlequinada, bufonaria, momice, truanice},
{plateia, público, assistência, audiência, auditório, espectadores},
{robô, títere, autômato, boneco, fantoche, marionete},
{cortina, cortinado, bambinela, corrediça, empanada, estore, reposteiro},
{dançarino, bailadeiro, bailador, bailão, bailarino, dançadeiro, dançante},
{desfecho, fim, final, finalização, epílogo, arremate, remate},
{epílogo, conclusão, desfecho, fim, final, arremate, remate},
{pintura, maquiagem, maquilagem, cosméticos, make-up, arrebique, rebique},
{desenlace, epílogo, desfecho, fecho, fim, final, arremate, remate},
{palhaço, bufão, alerquim, bobo, bufo, cômico, farsante, gracioso, histrião, mimo, momo, polichinelo, saltimbanco},
{jogral, bobo, alerquim, bufão, bufo, chocarreiro, farsante, farsista, histrião, maninelo, palhaço, saltimbanco, truanaz, truão},
{acrobata, acróbata, anemóbata, aramista, burlatim, cremnóbata, equilibrista, funambulista, funâmbulo, ginasta, petaurista, petauristário, volantim, volatim, volteador},
{clímax, apogeu, acme, ápice, assomada, auge, coroamento, coronal, culminância, cume, cúmulo, cúspide, esplendor, fastígio, magnificência, máximo, pináculo, píncaro, remate, zênite, zina}

b) POLITICO, AG. *Michaelis Dicionário de Sinônimos e Antônimos*; [MIC]

{artista, ator},
{ato, parte},
{camarote, frisa},
{cena, cenário},
{cena, palco},
{cenógrafo, cenarista},
{cinema, cine},
{circo, coliseu},
{dançarino, bailarino},
{dançatriz, dançarina},
{drama, peça},
{elenco, artistas},
{estrela, astro},
{estrela, celebridade},
{monólogo, solilóquio},
{mímica, pantomina},

{palco, tablado},
{palmas, aplausos},
{papel, parte},
{protagonista, herói},
{teatro, anfiteatro},
{androide, robô, autômato},
{ator, comediante, cômico},
{bailarina, dançarina, bailadeira},
{bis, repetição, reiteração},
{boneco, fantoche, títere},
{bufão, palhaço, saltimbanco},
{cortina, cortinado, véu},
{entreato, intervalo, pausa},
{espectador, assistente, ouvinte},
{espetáculo, peça, representação},
{extra, figurante, comparsa},
{histrião, cômico, comediante},
{peça teatral, peça, texto},
{pintura, maquilagem, maquiagem},
{protagonista, herói, personagem},
{protagonista, personagem, principal},
{teatro, arte de representar, palco},
{teatro, palco, cenário},
{truão, bobo, bobo da corte},
{acrobata, equilibrista, funâmbulo, acróbata},
{audiência, plateia, público, assistência},
{bufo, bufão, palhaço, saltimbanco},
{desenlace, desfecho, final, epílogo},
{entreato, intervalo, interlúdio, intermédio},
{estreia, início, princípio, começo},
{farsante, palhaço, alerquim, bufão},
{farsola, brincalhão, zombateiro, galhofeiro},
{figurante, figura, ator, intérprete},
{figurino, traje, roupa, vestimenta},
{jogral, bufão, palhaço, saltimbanco},
{palhaço, arlequim, bufão, farsante},
{público, plateia, assistência, auditório},
{teatro, fingimento, encenação, fita},
{truão, bufão, palhaço, saltimbanco},
{títere, marionete, fantoche, bonifrate},
{clímax, apogeu, auge, máximo, ápice},
{plateia, pessoas, público, assistência, audiência},
{prólogo, introdução, prefácio, preâmbulo, prelúdio},
{saltimbanco, bufão, palhaço, pelotiqueiro, truão},
{comediante, humorista, bufão, cômico, histrião, palhaço}

c) FERNANDES, F. <i>Dicionário de Sinônimos e Antônimos da Língua Portuguesa</i> ; [FER]
{autômato, robô},
{camarote, frisa},
{cantor, cantador},
{comediante, cômico},
{comparsa, figurante},
{debutante, estreante},
{dramaturgo, autor de dramas},
{elenco, conjunto de artistas},
{equilibrista, funâmbulo},
{herói, protagonista},
{intervalo, entreato},
{monólogo, solilóquio},
{arlequim, amante, amante cínico},
{bailador, bailadeiro, dançarino},
{cena, arte dramática, teatro},
{circo, pavilhão, afiteatro},
{dançarino, bailarino, dançador},
{farsante, gracejador, pantomimeiro},
{intermédio, intervenção, entreato},
{mímica, gesticulação, gestos},
{personagem, figura dramática, figura de romance},
{público, auditório, assistência},
{saltatriz, dançarina, bailarina},
{teatro, anfiteatro, circo},
{teatro, encenação, fita},
{ator, artista, comediante, cômico},
{bailarino, bailarim, bailador, dançarino},
{entreato, intervalo, entrecena, entremez},
{espectador, testemunha, assistente, observador},
{fantoche, autômato, bonifrate, títere},
{funâmbulo, acrobata, equilibrista, dançarino},
{ginasta, acrobata, acróbata, atleta},
{palco, palanque, tablado, estrado},
{palmas, aplausos, aclamação, ovação},
{pantomima, parlapatice, pantomina, pantomimice},
{pelotiqueiro, saltimbanco, prestidigitador, malabarista},
{polichinelo, bobo, bufão, truão},
{protagonista, herói, personagem, figurante},
{atriz, artista, estrela, estrela de teatro, estrela de cinema},
{bonifrate, fantoche, boneco, títere, boneco de engoço},
{cortina, estore, bambolina, bambinela, corrediça},
{epílogo, conclusão, remate, fecho, fim},

{espetáculo, exibição, divertimento, cena, representação},
{estreia, inauguração, começo, início, princípio},
{cena, decoração teatral, cenário, adornos, palco, espetáculo},
{clímax, gradação, apogeu, ápice, auge, ponto culminante},
{desenlace, desfecho, desenredo, epílogo, final, remate},
{espetáculo, contemplação, exibição, divertimento, cena, representação},
{farsola, fanfarrão, jactancioso, chocarreiro, galhofeiro, farsista},
{acrobata, dançarino, palhaço, saltimbanco, equilibrista, funâmbulo, malabarista},
{bonifrate, fanteche, boneco, boneco de engonço, automato, títere, marionete},
{bufo, bufão, bobo, truão, jogral, palhaço, arlequim},
{chocarreiro, bobo, bufão, truão, jogral, farsante, farsista},
{histrião, palhaço, bobo, farsista, saltimbanco, pelotiqueiro, bufão},
{palhaço, saltimbanco, arlequim, bobo, histrião, truão, clown},
{pelotiqueiro, saltimbanco, farsante, histrião, burlantim, volantim, polichinelo},
{jogral, truão, bobo, bufão, farsista, histrião, palhaço, chocarreiro},
{truão, bobo, palhaço, bufão, chocarreiro, pelotiqueiro, saltimbanco, truanaz},
{bufão, bobo, truão, palhaço, arlequim, farsista, farsante, histrião, saltimbanco, volatim},
{desfecho, conclusão, termo, remate, epílogo, resultado, desenlace, fim, desenredo, desentrecho},
{prólogo, prefácio, introdução, proêmio, prolusão, preâmbulo, prelúdio, preliminar, antelóquio, prefação}

d) TEP 2.0 (http://www.nilc.icmc.usp.br/tep2/); [TEP]
{ator, artista},
{autômato, robô},
{bailador, bailante},
{cinema, cine},
{coro, coral},
{encenação, dramatização},
{espetáculo, show},
{estreia, debute},
{estreia, première},
{farsista, goliardesco},
{intermédio, entremeio},
{monólogo, solilóquio},
{palmas, aplausos},
{pelotiqueiro, malabarista},
{plateia, auditório},
{teatro, drama},
{dançarino, bailarim, bailarino},
{mímica, gesticulação, pantomima},
{palco, estrado, tablado},
{papel, finalidade, função},
{teatro, anfiteatro, palco},
{make-up, maquiagem, maquilagem, pintura},
{truão, bobo, maninelo, truanaz},

{truão, pelotiqueiro, saltimbanco, truanaz},
{títere, bonifrate, fantoche, marionete},
{bailarina, bailadeira, dançadeira, dançarina, saltatriz},
{epílogo, conclusão, fecho, posfácio, remate},
{clímax, acme, ápice, apogeu, auge, zina},
{teatro, encenação, fantochada, farsa, fingimento, fita},
{acrobata, acróbata, aramista, equilibrista, funâmbulo, volantim, volatim},
{cena, cenário, painel, palco, pano de fundo, panorama, proscênio},
{intervalo, entrepausa, intercadência, intermissão, intermitência, interrupção, parada},
{títere, autômato, boneco, fantoche, marionete, paspalho, robô},
{desenlace, desenlaçamento, desenleio, desenredo, deslindamento, deslinde, destrinça, elucidação},
{prólogo, antelóquio, encabeçamento, exórdio, preâmbulo, prefácio, proêmio, prolusão},
{desenlace, conclusão, desenredo, desfecho, fim, final, solução, terminação, termo},
{truão, arlequim, bobo, bufão, farsante, histrião, palhaço, polichinelo, truanaz}

e) AZEVEDO, F.F.S. – <i>Dicionário Analógico da Língua Portuguesa</i>; [AZE]
{anteato, entreato, intermédio, entremez, embrechado},
{dramaturgo, autor dramático, comediógrafo, entremezista, mimógrafo},
{máscara, mascarado, fantasiado, dominó, saltimbanco},
{bastidores, coulisse, camarim, episcênio, palanque, borlista},
{peça teatral, representação teatral, espetáculo, dramaturgia, arte histriônica, peça, composição dramática, drama},
{êxodo, cenário, quadro, introdução, entredo, ato, cena, epílogo, prólogo, libreto},
{teatro, teatro de arena, cinema, cinerama, casa de ópera, politeama, anfiteatro, circo, hipódromo},
{execução, representação, récita, desempenho, mise-en-scène, encenação, montagem, estreia, jogo de cena, jogo, mímica},
{fornecedor, maquinista, ponto, apontador, gerente, contrarregra, empresário, figurinista, maquiador, continuísta, diretor, cenógrafo, fotógrafo},
{auditorio, público, claque, casa, cadeiras, poltronas, plateia, cávea, galeria, balcão, balcão nobre, varanda, camarote, frisa, torrinha, galinheiro},
{primeiro trágico, prima-dona, herói, protagonista, personagem, debutante, estreante, galã, amoroso, ingênua, pai nobre, ator, atriz},
{palco, vistas, pano de boca, telão, tablado, proscênio, rampa, estrado, ribalta, gambiarra, tímele, chaspulho, mezzanino, orquestra, boca de cena, bambolina, maroma, trampolim, trapézio},
{títere, franca-tripa, fantoche, boneco de engoço, autômato, polichinelo, mamulengo, marionete, manequim, jagodes, monha, modelo, padrão, figura de cera, figurilha, figurino, escorço, boneco, boneca, figura},
{charlatão de circo, charlatão de feira, pelotiqueiro, acrobata, voador, barrista, anemobata, funâmbulo, barlatim, volteador, equilibrista, malabarista, argolista, ginasta, saltador, saltatriz, dançarino, dançartriz, bailador, bailarino, bailarina, bailarim, cantor, cantora, cantatriz, transformista, arara},
{drama, tragédia, coturno, comédia, ópera, grande ópera, ópera séria, opereta, zarzuela, cena cômica, cena lírica, revista, vaudeville, comedieta, autos, soap opera, lever de rideau, comédie larmoyante, farsa, mimo, farsalhão, burlata, divertimento, comédia bufa, ópera bufa, dramalhão, arlequianada, palhaçada, besteirol, cena muda, momo, mímica, pantomima, mourisca, mimodrama, baile, bailado, bailete, bailarico, fandango, melodrama, tragicomédia, monodrama, monólogo, duólogo, trilogia, mistério, vista de teatro, filme, novela, folhetim, épico}

{papel, caracterização, companhia, elenco, corpo de baile, repertório, ator, atriz, artista, artista dramático, intérprete, personagem, protagonista, estrela, figura, trágico, trágica, cômico, galã, comediante, apinário, pantomimeiro, pantomimo, pantomineiro, mimo, machatim, diteriade, titeriteiro, palhaço, clown, arlequim, bufo, bufão, histrião, jogral, chocarreiro, truão, farsante, farsola, farsista, cabotino, columbina, polichinelo, boneco, mamulengo, bonifrate, títere, franca-tripa, androide, figurante, figuranta, comparsa, supranumerário, entremezista, comparsaria, prestigiador, mágico, ilusionista}

f) FLORENZANO, E. <i>Dicionário de Ideias Semelhantes</i> . [IDS]
{mímica, pantomima},
{peça, peça teatral, texto},
{palhaço, bufão, arlequim, colombina, pierrô},
{títere, bonifrate, fantoche, marionete, boneco de engonço},
{cinema, teatro, circo, sala de concertos, casa de espetáculos, concha acústica, café-concerto},
{teatro, palco, cenário, cena, local de ação, proscênio, coliseu, hipódromo, circo},
{peça, peça teatral, cena, quadro, ato, cortina, esquete, cenário, representação, encenação},
{drama, teatro, cena, peça, arte teatral, arte histriônica, melodrama, tragédia, palco, comédia, ópera},
{peça, peça teatral, introdução, encenação, prelúdio, prólogo, exposição, desenvolvimento, enredo, epílogo, clímax, apoteose},
{plateia, bancada, arquibancada, torrinha, cadeiras, poltronas, frisas, camarote, poleiro, galeria, público, auditório},
{elenco, companhia, corpo de baile, bailado, personagem principal, figura secundária, ponta, mocinho, prima-dona, cantora principal, papel, repertório, programação},
{ator, artista, astro, estrela, personagem, figurante, corista, extra, cantor, dançarino, trágico, cômico, comediante, humorista},
{humorista, palhaço, comediante, cômico, bufão, bobo, truão, polichinelo, tônico, arlequim, caricaturista, imitador, burlão, burlantim},
{representação, peça, cena, quadro, ato, encenação, prelúdio, prólogo, exposição, desenvolvimento, enredo, epílogo, clímax, apoteose, texto},
{bico, cauda, rabo, apêndice, remate, epílogo, cerrar da cortina, cair do pano, desenlace, limite, cobro, peroração, palavra final, última deixa, paralisação, expiração, parada},
{charlatão, quacre, tartufo, hipócrita, impostor, pelotiqueiro, jogral, malabarista, prestidigitador, mágico, mago, feiticeiro, macumbeiro, pai-de-santo, nigromante, cartomante, astrólogo},
{melodrama, tragédia, comédia, ópera, ópera bufa, farsa, burleta, teatro de revista, folias, auto, interlúdio, intervalo, pantomima, bailado, espetáculo, mascarada, chanchada, mistério, arlequinada}

Como não é possível reproduzir aqui a matriz preenchida com todos os CSs e CSAs, é necessário abrir o arquivo *anaset_TEATRO.csv* – presente no CD em anexo – no programa *Microsoft Excel™*, para ver os arranjos dos CSs e CSAs na matriz léxico-conceitual do *anaset* TEATRO. No entanto, o Quadro 7.19 exemplifica o arranjo dos CSs e CSAs para as ULs *drama*, *teatro* e *representação* e para as categorias conceituais S1, glosado por “edifício destinado à apresentação de obras dramáticas e outros espetáculos públicos”, e S2, glosado por “o ofício ou a arte teatral”.

Quadro 7.19. Exemplo do *anaset* TEATRO preenchido com CSs e CSAs

MUNDO SOCIAL		
TEATRO	Relações Paradigmáticas	
	S1	S2
<i>drama</i>		{drama, teatro, cena, peça, arte teatral, arte histriônica, melodrama, tragédia, palco, comédia, ópera} [IDS] {drama, peça} [MIC] {drama, peça, obra, texto} [HOU] {peça teatral, peça, representação teatral, espetáculo, dramatologia, dramaturgia, arte histriônica, drama} [AZE] {drama, teatro} [TEP]
<i>representação</i>		{representação, encenação, apresentação, dramatização, montagem, peça teatral, peça} [HOU] {execução, representação, récita, desempenho, mise-en-scène, encenação, montagem, estreia, jogo de cena, jogo, mímica} [AZE] {espetáculo, contemplação, exibição, divertimento, cena, representação} [FER] {espetáculo, peça, representação} [MIC] {peça, peça teatral, cena, quadro, ato, cortina, esquete, cenário, representação, encenação} [IDS]
<i>teatro</i>	{teatro, teatro de arena, cinema, cinerama, casa de ópera, politeama, anfiteatro, circo, hipódromo} [AZE] {cinema, teatro, circo, sala de concertos, casa de espetáculos, concha acústica, café-concerto} [IDS] {teatro, anfiteatro, circo} [FER] {teatro, anfiteatro} [MIC] {teatro, anfiteatro, palco} [TEP]	{drama, teatro, cena, peça, arte teatral, arte histriônica, melodrama, tragédia, palco, comédia, ópera} [IDS] {teatro, encenação, fita} [FER] {palco, teatro, arte teatral} [HOU] {teatro, drama} [TEP] {teatro, encenação, fantochada, farsa, fingimento, fita} [TEP]

7.3.4 ETAPA 4 - CONSOLIDAÇÃO DOS RESULTADOS DO *ANASET* TEATRO

Com a matriz completamente preenchida com as ULs de entrada (7.3.1), as categorias conceituais (7.3.2) e os CSs e CSAs correspondentes (7.3.3), os resultados podem ser consolidados seguindo-se os procedimentos apresentados em 6.1.2.4⁶¹.

O Quadro 7.20 reproduz os resultados consolidados por meio do algoritmo de ordem onomasiológica. Na primeira coluna à esquerda temos as treze categorias conceituais que estruturam conceitualmente o *anaset* TEATRO, seguidas na coluna seguinte de suas respectivas glosas. Na última coluna, são apresentados os CdSAs correspondentes ao domínio.

⁶¹ Para ver a consolidação dos resultados de ordem semasiológica, é necessário abrir o arquivo CdSA_sem_TEATRO.csv, presente no CD em anexo.

Quadro 7.20. Resultados consolidados para o *anaset* TEATRO

Domínio conceitual TEATRO		
(S)	Glosa	CdSAs
Relação Paradigmática		
(S1)	lugar ou edifício destinado à apresentação de obras dramáticas e outros espetáculos públicos	{teatro (1),anfiteatro (0.714),circo (0.429),palco (0.429),cinema (0.286),cinerama (0.143),teatro de arena (0.143),casa de ópera (0.143),politeama (0.143),hipódromo (0.143),sala de concertos (0.143),local de ação (0.143),proscênio (0.143),cena (0.143),cenário (0.143),casa de espetáculos (0.143),concha acústica (0.143),café-concerto (0.143)}
(S2)	o ofício ou a arte teatral	{peça (1),encenação (0.818),representação (0.727),teatro (0.682),espetáculo (0.591),drama (0.591),cena (0.591),peça teatral (0.5),arte histriônica (0.409),montagem (0.409),palco (0.318),melodrama (0.273),arte teatral (0.273),dramatologia (0.227),representação teatral (0.227),fita (0.227),ópera (0.227),tragédia (0.227),comédia (0.227),dramaturgia (0.227),dramatização (0.182),ato (0.182),texto (0.182),farsa (0.182),récita (0.136),mise-en-scène (0.136),desempenho (0.136),estreia (0.136),jogo de cena (0.136),mímica (0.136),jogo (0.136),divertimento (0.136),quadro (0.136),execução (0.136),fingimento (0.136),apresentação (0.136), [...]}
(S3)	literatura ou arte dramática	{texto (1),peça (0.667),peça teatral (0.667),fala (0.333)}
Relação Sintagmática		
(S4)	partes do teatro	{palco (1),cena (0.545),cenário (0.455),tablado (0.364),proscênio (0.364),palanque (0.273),cortina (0.273),bastidores (0.273),estrado (0.273),painel (0.182),pano de fundo (0.182),corrediça (0.182),panorama (0.182),borlista (0.182),camarim (0.182),estore (0.182),bambinela (0.182),episcênio (0.182),cortinado (0.182),bambolina (0.182),coulisse (0.182), [...]}
(S5)	público	{plateia (1),público (1),auditório (0.733),camarote (0.533),assistência (0.533),audiência (0.467),poltronas (0.4),galeria (0.4),torrinha (0.4),cadeiras (0.4),frisa (0.333),cávea (0.2),balcão (0.2),espectadores (0.2),casa (0.2),claque (0.2),auditorio (0.2),balcão nobre (0.2),frisas (0.2),varanda (0.2),poleiro (0.2),arquibancada (0.2),galinheiro (0.2),bancada (0.2),assistente (0.133),espectador (0.133), [...]}
(S6)	autor de peças dramáticas	{dramaturgo (1),autor dramático (0.75),comediógrafo (0.75),entremezista (0.75),mimógrafo (0.75),autor de dramas (0.25)}
(S7)	gêneros de obra dramática	{tragédia (1), comédia (0.938),pantomima (0.938),mímica (0.813),drama (0.813),melodrama (0.813),ópera (0.75),monólogo (0.688),farsa (0.688),cena (0.625),mistério (0.563),burleta (0.563),bailado (0.563),ópera bufa (0.563),novela (0.5),monodrama (0.5),dramalhão (0.5),cena muda (0.438),ópera séria (0.438),besteirol (0.438),grande ópera (0.438),momo (0.438),baile (0.438),mimodrama (0.438),coturno (0.438),mourisca (0.438),palhaçada (0.438),cena lírica (0.438),mimo (0.438),cena cômica (0.438),autos (0.438),revista (0.438),farsalhão (0.438),zarzuela (0.438),arlequianada (0.438),opereta (0.438),comédia bufa (0.438),divertimento (0.438),bailete (0.438),ato (0.438),vista de teatro (0.438),filme (0.438),folhetim (0.438),bailarico (0.438),épico (0.438),trilogia (0.438),fandango (0.438),tragicomédia (0.438),duólogo (0.438),solilóquio (0.25),peça (0.188),arte histriônica (0.188),auto (0.188),teatro (0.188),arte teatral (0.188), [...]}
(S8)	partes em que se divide uma peça de teatro	{epílogo (1),entreato (0.722),intermédio (0.667),prólogo (0.611),intervalo (0.611),desfecho (0.556),final (0.5),clímax (0.5),ato

		(0.5),desenlace (0.444),interlúdio (0.444),remate (0.444),cena (0.389),quadro (0.389),prelúdio (0.389),fim (0.389),conclusão (0.333),peça (0.278),encenação (0.278),exposição (0.278),enredo (0.278),representação (0.278),desenvolvimento (0.278),entrecena (0.278),texto (0.278),estreia (0.278),apoteose (0.278),auge (0.222),ápice (0.222),desenredo (0.222),apogeu (0.222),introdução (0.222),começo (0.167),fecho (0.167),início (0.167),arremate (0.167),preâmbulo (0.167),termo (0.167),prefácio (0.167),entremez (0.167),terminação (0.111),debute (0.111),solução (0.111),libreto (0.111),antelóquio (0.111),embrechado (0.111),anteato (0.111),entrecho (0.111),zina (0.111),máximo (0.111),cenário (0.111),proêmio (0.111),acme (0.111),êxodo (0.111),prolusão (0.111),princípio (0.111),parte (0.111),[...]
(S9)	figura dramática	{protagonista (1),herói (0.909),personagem (0.818),papel (0.455),ator (0.455),atriz (0.364),galã (0.364),ingênuo (0.273),amoroso (0.273),pai nobre (0.273),primeiro trágico (0.273),intérprete (0.273),estreador (0.273),prima-dona (0.273),debutante (0.273),coral (0.182),coro (0.182),figurante (0.182), [...]}
(S10)	o conjunto de artistas ligados a um espetáculo ou a uma companhia	{artista (1),ator (0.966),figurante (0.931),estrela (0.828),cômico (0.828),comediante (0.828),personagem (0.655),intérprete (0.621),comparsa (0.621),trágico (0.586),protagonista (0.517),galã (0.483),atriz (0.483),figura (0.448),elenco (0.448),histrião (0.414),palhaço (0.414),bufão (0.414),clown (0.379),farsante (0.379),truão (0.379),titeriteiro (0.379),bufo (0.379),jogral (0.379),chocarreiro (0.379),arlequim (0.379),pantomimeiro (0.379),diteriade (0.379),repertório (0.379),caracterização (0.379),papel (0.379),companhia (0.379),corpo de baile (0.379),artista dramático (0.379),trágica (0.379),machatim (0.379),mimo (0.379),pantomineiro (0.379),apinário (0.379),pantomimo (0.379),farsista (0.379),figurante (0.379),entremezista (0.379),farsola (0.379),comparsaria (0.379),prestigiador (0.379),extra (0.379),ilusionista (0.379),mágico (0.379),supranumerário (0.379),boneco (0.379),androide (0.379),columbina (0.379),cabotino (0.379),polichinelo (0.379),mamulengo (0.379),franca-tripa (0.379),títere (0.379),bonifrate (0.379),astro (0.31),humorista (0.241),cantor (0.207),dançarino (0.207),corista (0.207),herói (0.138),estrela de teatro (0.103),estrela de cinema (0.103), [...]}
(S11)	equipe de um espetáculo teatral	{cenógrafo (1),contrarregra (0.8),apontador (0.8),ponto (0.8),fornecedor (0.8),maquinista (0.8),gerente (0.8),empresário (0.8),fotógrafo (0.8),diretor (0.8),continuista (0.8),figurinista (0.8),maquiador (0.8),cenarista (0.2)}
(S12)	acessórios ou instrumentos cênicos necessários à produção de um filme, uma peça teatral	{pintura (1),maquilagem (1),maquiagem (1),make-up (0.667),cosméticos (0.333),arrebique (0.333),rebique (0.333),traje (0.333),figurino (0.333),vestimenta (0.167),roupa (0.167),indumentária (0.167),vestuário (0.167)}
(S13)	aclamação	{palmas (1),aplausos (1),ovação (0.5),aclamação (0.5),bis (0.5),repetição (0.5),saudação (0.25),vivas (0.25),reiteração (0.25)}

7.4 INTEGRAÇÃO COM O DICIONÁRIO CRIATIVO E PROPOSTAS DE INTERFACE DO DAD

O *website* **Dicionário Criativo** (www.dicionariocriativo.com.br) está no ar, em versão *beta*, desde junho de 2012. A versão *beta* já disponibiliza todos os módulos citados em 1.1. Uma parceria com a editora **Lexikon**, a maior editora de obras de referência do Brasil, permite-nos utilizar os conteúdos do Aulete Digital para os módulos 1 e 3, e o conteúdo do *Dicionário Analógico da Língua Portuguesa* do professor Azevedo, revisto e atualizado, no módulo 2 (conforme Figura 1.1). Os módulos 4 e 5 utilizam um banco de dados próprio, com cerca de 2 mil provérbios e 6 mil citações, respectivamente. Atualmente, esses são os dois únicos módulos nos quais os usuários podem inserir suas sugestões e contribuir para o crescimento da base. Portanto, o número de provérbios e citações aumenta diariamente. O módulo 6 faz uso da API⁶² oficial da Wikipedia para retornar trechos dos verbetes da enciclopédia virtual. E, por fim, o módulo 7 é fruto de uma parceria com o banco de imagens **GettyImages**.

Nessa versão *beta* do Dicionário Criativo, as consultas por palavra-chave retornam resultados, exclusivamente, (i) para a palavra buscada ou (ii) para a palavra buscada mais sua flexão sintática.

(i) palavra-chave **namorado**: resultados para *namorado*;

(ii) palavra-chave **namorado**: resultados para *namorado, namorada, namorados, namoradas*.

Conforme o objetivo da aplicação, esse conjunto de resultado estrito é satisfatório e até desejável. No entanto, para muitas aplicações, entre elas a do Dicionário Criativo, o resultado da busca pode ser ampliado por meio de sinônimos e conexões analógicas.

A integração do DAD com o Dicionário Criativo tem, portanto, dois objetivos. O primeiro, como dissemos em 1.1, é substituir o atual dicionário analógico, do professor Azevedo, no módulo 2. Ou seja, o DAD deverá prover todo o conteúdo necessário para o módulo 2. O segundo objetivo é servir de recurso intermediário entre a palavra-chave de *input* no campo de pesquisa e os resultados dos demais módulos. Sua estrutura de conjuntos difusos de sinônimos e analógicos permitirá que ele extenda os resultados dos outros módulos, por exemplo, trazendo provérbios e citações que contenham não

⁶² *Application Programming Interface* (Interface de Programação de Aplicativos, em português) é um conjunto de rotinas e padrões de programação para acesso a uma aplicativo de software baseado na Web.

apenas a palavra-chave de *input* e suas flexões sintáticas, mas também provérbios e citações que contenham ULs sinônimas e análogas. Hoje, uma pesquisa no Dicionário Criativo pela palavra *amor*, por exemplo, retorna, no módulo 5 (citações), resultados tais como os que aparecem na Figura 7.2.

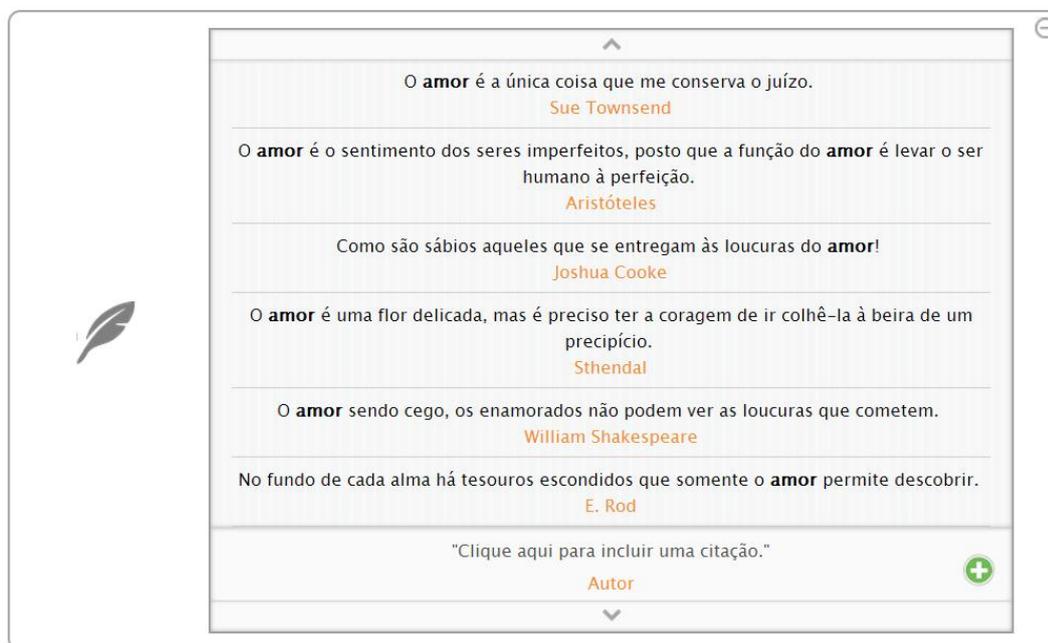


Figura 7.2. Resultados do Dicionário Criativo, no módulo de citações, para a busca pela UL *amor*.

A integração do DAD com o Dicionário Criativo possibilitará que os resultados se ampliem, nesse caso, para citações que contenham a UL *afeto*, por exemplo. Ao rolar os resultados para baixo, poderemos encontrar a citação "A prova de um **afeto** puro é uma lágrima", atribuída a George (Lord) Byron (cf. NEVES DA SILVA, 2012). Ou, por exemplo, a citação "Um **beijo** é um segredo que se diz na boca e não no ouvido" – atribuída a Jean Rostand –, por ela incluir a UL análoga *beijo*.

O DAD ocupando o módulo 2, devidamente integrado ao Dicionário Criativo, funcionará como um painel de controle dos outros módulos. Navegar pelos domínios, categorias conceituais e ULs do DAD, permitirá diferentes configurações de resultados. Uma experiência que vai além dos resultados obtidos exclusivamente por meio de buscas por palavras-chave.

Outro exemplo de como essa integração deverá funcionar como inteligência semântica para todo o sistema do Dicionário Criativo é a base do DAD funcionando como camada intermediária para recuperar informação de outras bases de dados com

estrutura onomasiológica. Tomemos como exemplo o módulo de locuções e expressões idiomáticas (EI). Da maneira como está na versão *beta*, uma pesquisa por *amor* no Dicionário Criativo exhibe nesse módulo apenas as locuções e EI's que possuem a palavra *amor* na sua composição, conforme (i)⁶³.

- (i) *Amor platônico*: O que é isento de desejo sexual.
De mil amores: Com o maior prazer, com todo o gosto.
Fazer amor: Ter relações sexuais.
Pelo amor de Deus: Por favor, por caridade.
Por amor à arte: De modo gratuito, desinteressado
Por amor de: Por causa de; em atenção a.
Um amor: Pessoa ou coisa muito bonita, graciosa; um encanto; um doce.

Nosso intuito, no entanto, é que o módulo retorne também resultados de ordem onomasiológica. Ou seja, as locuções e EI's que, mesmo não contendo a palavra de *input* e/ou seus sinônimos e analógicos, expressem conceitos do domínio AMOR. Valendo-se da base de dados do próprio *Dicionário Analógico da Língua Portuguesa* (AZEVEDO, 1950 [2010]), podemos exhibir resultados como *cair na graça de, serem dois corações em um só, possuir o coração de, reinar no coração de, ser a tampa da panela de alguém* e outros, já que elas aparecem organizadas dentro do verbete **897. Amor**, equivalente ao *anaset* AMOR no DAD. No entanto, o dicionário do professor Azevedo tem limitações e deve ser complementado, principalmente no que diz respeito às EI's, com outras fontes. Uma delas é o *Dicionário onomasiológico de expressões idiomáticas do Brasil* (DOEIB)(RIVA, 2009), com 1562 EI's separadas em 400 categorias conceituais. Nele, as categorias não foram lexicalizadas de modo a corresponderem aos 1000 verbetes do Azevedo. Elas precisam ser recuperadas, portanto, por meio dos sinônimos e analógicos dos CdSAs do DAD, seguindo o peso das ULs para apresentar os resultados dos mais relevantes para os menos relevantes. O Quadro 7.21 apresenta as EI's recuperadas ao se buscar por todas as ULs sinônimas e análogas (com peso acima de 0.1) do *anaset* AMOR do DAD na coluna de categorias conceituais (S) do DOEIB.

⁶³ <http://www.dicionariocriativo.com.br/amor>, acessado em 10 de janeiro de 2013.

Quadro 7.21. Pequena amostra de EI's do DOEIB, separadas por temas e com seus respectivos significados.

S	EI's	Significado
ADMIRAÇÃO	<i>de (se) tirar o chapéu</i>	admirável [alusão ao cumprimento, não tão comum hoje em dia, de se tirar o chapéu para reverenciar outra pessoa]
AFINIDADE	<i>alma gêmea</i>	pessoa com quem se tem profundas afinidades, geralmente em relacionamento amoroso.
AMOR	<i>dar o céu</i>	querer fazer até o impossível por quem se ama.
DESEJO	<i>estar de quatro por</i>	sentir-se muito atraído por alguém.
EMOÇÃO	<i>falar ao coração</i>	emocionar [orig. sup.: alusão ao órgão do corpo humano metaforicamente retomado como o centro das emoções]
EMOÇÃO	<i>falar direto ao coração</i>	emocionar vivamente [orig. sup.: alusão ao órgão do corpo humano metaforicamente retomado como o centro das emoções]
EMOÇÃO	<i>mudar de cor</i>	ficar pálido ou ruborizado devido a uma forte emoção [orig.: alusão à ruborização da pele diante de determinadas circunstâncias]
EMOÇÃO	<i>ter um nó na garganta</i>	não conseguir expressar-se; estar triste com algo.
EMPATIA	<i>ir com a cara de</i>	ter simpatia por [orig. sup.: alusão ao estabelecimento de vínculo positivo com algo ou alguém por conta da expressão facial]
SEDUÇÃO	<i>cair matando</i>	paquerar.
SEDUÇÃO	<i>dar bola</i>	dar confiança a alguém [orig.: esporte; alusão aos parceiros trocam jogadas em momentos decisivos]
SEDUÇÃO	<i>dar em cima</i>	paquerar com insinuações freqüentes.
SEDUÇÃO	<i>mulher fatal</i>	mulher muito atraente e sedutora [orig.: cinema; designação comum para atrizes ou mulheres famosas]
SENTIMENTO	<i>voz do coração</i>	os sentimentos mais íntimos [orig.: alusão à parte do corpo humano considerada o centro das emoções]
TENTAÇÃO	<i>fruto proibido</i>	qualquer coisa que, por ser proibida, se mostra mais cobiçada e tentadora [orig.: bíblica; Eva cede à tentação do diabo, disfarçado de serpente, e divide o fruto proibido, "árvore da ciência do bem e do mal", com Adão]

Não há dúvidas de que as EI's que aparecem no Quadro 7.21 devam aparecer como resultados dentro do *anaset* AMOR. Evidentemente, os resultados serão reconfigurados e refinados conforme se navegue pelas categorias conceituais internas ao *anaset*. O usuário poderá percorrer as categorias do DAD por meio de interface especialmente desenvolvida para o módulo 2.

Imaginemos a futura implementação de um dicionário de rimas em um novo módulo do Dicionário Criativo. Poderemos utilizar a base do DAD para apresentar palavras que rimam e que, ao mesmo tempo, instanciam a mesma categoria conceitual. Para uma busca por *afeição*, no *anaset* AMOR, na categoria conceitual AFEIÇÃO (S1), o módulo de rimas poderá exibir, por exemplo, as ULS *afeição* (1), *inclinação* (0.5), *dileção* (0.396), *dedicação* (0.333), *admiração* (0.292), *adoração* (0.125), *atração* (0.083), *paixão* (0.042), *propensão* (0.042), *emoção* (0.021), *devoção* (0.021), *consideração* (0.021), antes

de exibir outras ULs terminadas em *ão* de outras categorias e domínios. No entanto, se o usuário clicar em PAIXÃO (S2), o módulo de rimas exibirá as ULs *paixão* (1), *adoração* (0.421), *atração* (0.368), *devoção* (0.342), *emoção* (0.105), *exaltação* (0.079), *veneração* (0.079), *admiração* (0.053), *explosão* (0.053), *afeição* (0.053), *ebulição* (0.053), *sedução* (0.053), *dedicação* (0.026), *dileção* (0.026), *tenção* (0.026), *fascinação* (0.026), *tentação* (0.026), *propensão* (0.026), *agarração* (0.026), nessa ordem, antes de exibir outras ULs com a mesma terminação.

A Figura 7.3 exibe um exemplo de interface do módulo 2 para uma busca pelo item léxico *amor*⁶⁴. Por padrão, é assim que o módulo correspondente ao DAD será exibido. A partir do *input amor*, o sistema recupera o CdSA em que o *input* aparece melhor ranqueado, identifica o *anaset* correspondente e exibe todo o *anaset* separando os CdSAs por coluna. A palavra de *input* aparece destacada em laranja. Para acessar os demais CdSAs, o usuário deverá utilizar as setas situadas na parte inferior da caixa ou o botão de *scroll* do mouse. As colunas então se movimentam para a esquerda abrindo espaço para as demais colunas. Para ter acesso a todas as ULs de um determinado CdSA, o usuário deverá clicar sobre a UL que encabeça o respectivo CdSA, grafada em capitular e em negrito. Dessa forma, as demais colunas abrem espaço para que todas as ULs do CdSA selecionado sejam exibidas.

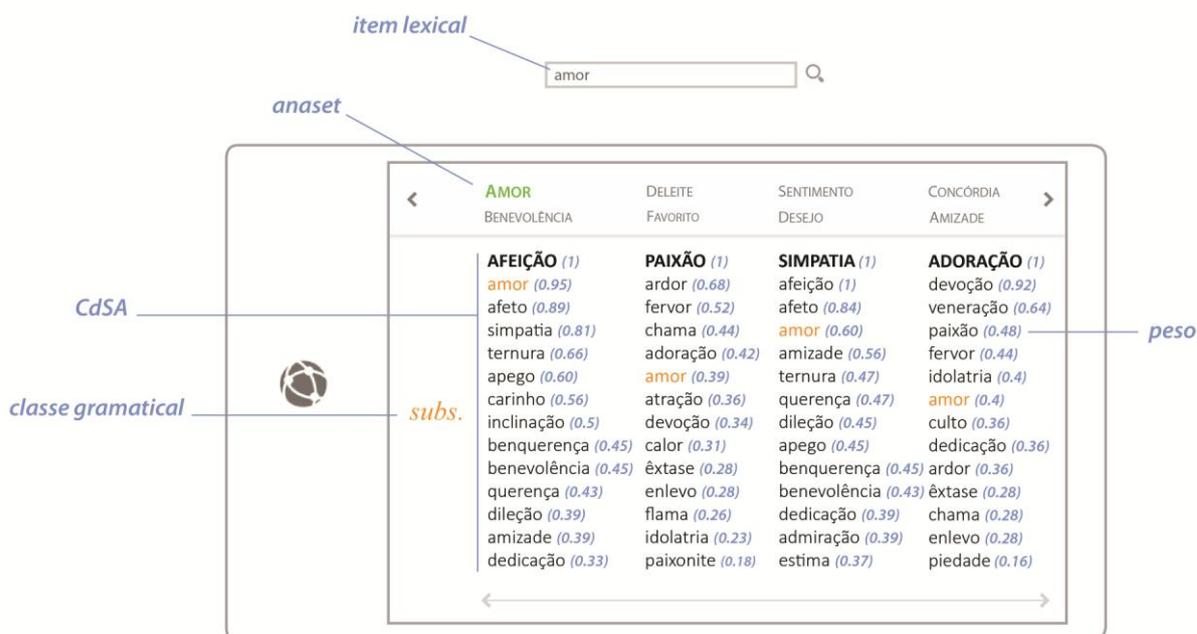


Figura 7.3. Exemplo de interface do DAD para uma busca por *amor*.

⁶⁴ As informações em azul são informações de background e, portanto, não serão exibidas para o usuário.

O cabeçalho do módulo exibe todos os *anaset*s em que haja CdSAs que contenham, por sua vez, o item léxico de *input*. O *anaset* que estiver em exibição no momento aparece destacado em verde. Caso existam muitos *anaset*s, as setas no cabeçalho permitirão que o usuário navegue para a esquerda e a direita para ter acesso a todos. Caso o usuário selecione o CdSA de uma categoria conceitual que possua estrutura interna, como no caso da categoria encabeçada por NAMORO, então o sistema exibirá, além de todas as ULs de NAMORO, também o *frame* *Relação_amorosa*. Assim, o usuário terá acesso também às ULs dos CdSAs correspondentes às categorias conceituais NAMORADO, NAMORADA, GALANTEADOR e CARÍCIA.

Nosso objetivo com o design da interface de cada módulo do Dicionário Criativo é que as informações estejam disponíveis ao usuário sob uma estrutura que propicie uma fácil visualização e que estimule a criatividade. O sistema deverá permitir, em poucos cliques, que o usuário i) encontre a melhor maneira de lexicalizar uma ideia e/ou ii) encontre outras ideias complementares, desviantes, metonímicas e metafóricas, enfim, ideias análogas. Estamos falando, portanto, de uma interface gráfica limpa, iconográfica e com hiperlinks inteligentes. Implementado em HTML 5 e valendo-se da tecnologia de interface adaptativa, o Dicionário Criativo poderá ser acessado, via internet, em qualquer dispositivo computacional: computadores, *tables* e *mobiles*.

7.5. ASPECTOS COMPUTACIONAIS

Apesar do domínio computacional não fazer parte do escopo da presente tese, faremos, nesta seção, uma breve explanação da etapa computacional que está sendo desenvolvida, graças a recursos do CNPq⁶⁵ e da empresa Dicionário Criativo, para implementar e disponibilizar o DAD em ambiente *Web*. Para tanto, estão sendo utilizadas as seguintes tecnologias: i) a linguagem de programação PHP; ii) o servidor de banco de dados relacionais MySQL, para armazenamento das informações léxico-conceituais do DAD; iii) e as linguagens de marcação HTML5 e CSS3 para geração da interface gráfica para o usuário.

⁶⁵ CNPq – Programa RHAe Pesquisador na Empresa – Processo 456448/2012-8.

É importante ressaltar que todas as tecnologias supracitadas são recursos gratuitos e bastante difundidos, utilizados em inúmeros projetos de aplicações *Web*. Com isso, além do menor custo financeiro, a resolução de problemas técnicos é facilitada, visto que há diversos profissionais que trabalham com estas tecnologias e fóruns direcionados.

Formalmente, o desenvolvimento de uma aplicação *Web* é dividido em três partes principais: i) a geração da base de dados; ii) os mecanismos de funcionamento interno do *website*, conhecido também como *back-end*; e iii) os mecanismos de apresentação do conteúdo para o usuário, conhecido também como *front-end*. A seguir, é apresentado o desenvolvimento de cada uma dessas etapas computacionais.

7.5.1 BANCO DE DADOS

As ULs, os CdSAs e os *anaset*s foram representados por meio de quatro tabelas e as relações entre elas no bando de dados. Esse modelo é visualmente representado pela Figura 7.4.

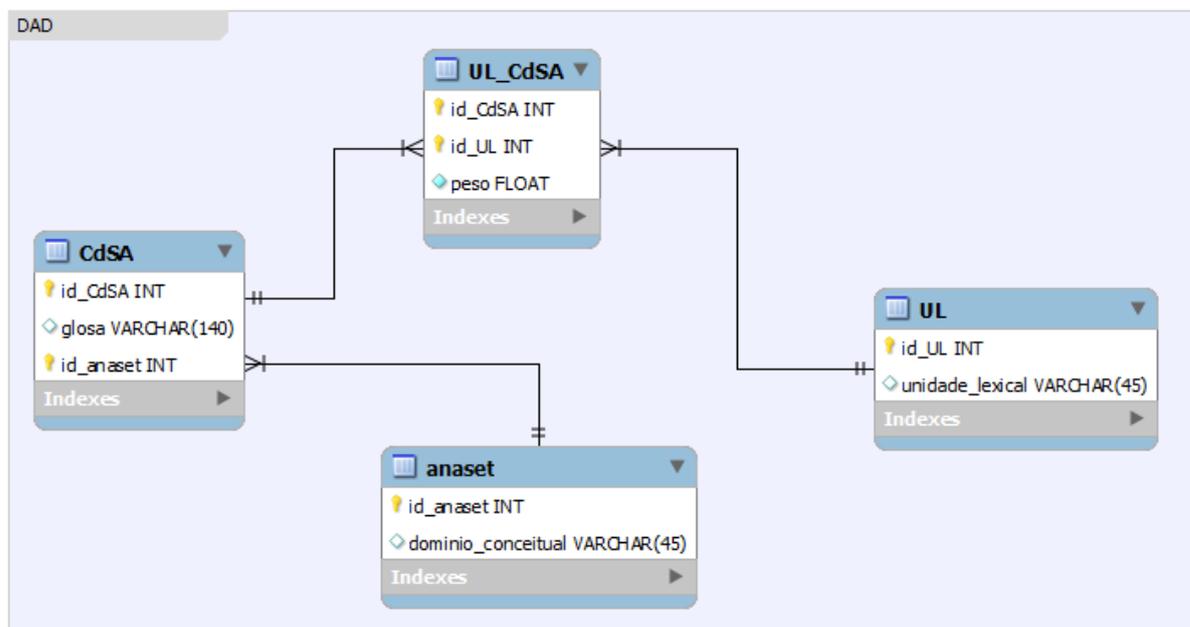


Figura 7.4 Modelo de representação dos *anaset*s, CdSAs e ULs no banco de dados

Nessa forma de representação, têm-se os seguintes componentes:

i) **Tabelas**, representadas pelos retângulos e cujos componentes são: a) **título da tabela**, disposto na parte superior e em negrito; b) **campos**, que representam as informações armazenadas em cada tabela. À direita de cada campo, em capitulares, é identificada a maneira pela qual as informações serão representadas (VACHAR = sequência de caracteres; INT = número inteiro; FLOAT = número real)

ii) **Relações**, identificadas por linhas que conectam duas tabelas, que representam relações lógicas entre estas.

A tabela **UL** armazena todas as ULs presentes em, no mínimo, um CdSA. Toda UL armazenada no banco de dados possui um número único, denominado **id_UL**, que a identifica e permite relacioná-la com outras tabelas.

Cada CdSA do DAD é registrado na tabela **CdSA**, que é composta por i) um identificador numérico único (**id_CdSA**), ii) uma **glosa** (que representa uma descrição informal de seu significado) e iii) um número, que aponta para um *anaset* cadastrado na tabela **anaset**. Esse último campo é utilizado para registrar o *anaset* em que o CdSA se encontra.

A tabela **UL_CdSA** relaciona as tabelas **UL** e **CdSA** por meio de seus identificadores (**id_UL** e **id_CdSA**, respectivamente). Assim, é possível representar que um CdSA possui uma ou mais ULs em seu interior e registrar o peso da relação de uma UL no CdSA correspondente (campo **peso**).

Se um CdSA **S**, com identificador **id_S**, é composto pelas unidades lexicais **UL1**, **UL2** e **UL3** registradas na tabela **UL** e, cujos identificadores únicos serão **id_UL1**, **id_UL2** e **id_UL3**, teremos, por exemplo, a tabela **UL_CdSA** como representada abaixo:

Quadro 7.22. Exemplo de preenchimento da tabela **UL_CSA_difuso**

id_CdSA	id_UL	peso
id_S	id_UL1	FLOAT
id_S	id_UL2	FLOAT
id_S	id_UL3	FLOAT

Dessa forma, é possível representar um modelo de banco de dados *fuzzy*, como exemplificado no Quadro 5.3 (Seção 5.4.1), onde diferentes ULs possuem relações com determinados conceitos **S** por meio de um grau de pertinência.

Para popular o banco de dados são utilizados os arquivos disponibilizados no formato .csv – presentes no CD anexo à tese. Para executar essa transferência dos dados, foi elaborado um simples *script* em linguagem de programação Python, de forma a gerar um arquivo a ser importado para a base de dados. Dessa forma, a tabela 7.23 apresenta

um exemplo das primeiras linhas da tabela UL_CdSA do banco de dados, preenchida com as informações extraídas do arquivo `anaset_AMOR.csv` para o CdSA S1, glosado por “afeição muito intensa de uma pessoa por outra” (conforme Quadro 7.7).

Quadro 7.23. Exemplo da tabela UL_CSA_difuso preenchida com os dados

id_CSA_difuso	id_UL	peso
id_S1	id_afeição	1
id_S1	id_amor	0,958
id_S1	id_afeto	0,896
id_S1	id_simpatia	0,813
id_S1	id_ternura	0,667
id_S1	id_apego	0,604

7.5.2 BACK-END

Para a realização do mecanismo de funcionamento interno do *website* são necessários diversos algoritmos computacionais como, por exemplo, o algoritmo de consulta na base de dados. Para a consulta na base de dados foi utilizada a linguagem de consulta SQL, que é amplamente reconhecida como a linguagem padrão de consulta em bases de dados relacionais. Como exemplo, o trecho a seguir apresenta uma linha de instrução em linguagem SQL que retorna todos os CdSAs em que uma determinada UL ocorre.

```
SELECT CdSA.id_CdSA, CdSA.glosa
FROM CdSA, UL, UL_CdSA
WHERE UL_CdSA.id_CdSA = CdSA.id_CdSA
AND UL_CdSA.id_UL = UL.id_UL
AND UL.unidade_lexical = "<palavra procurada>";
```

7.5.3 FRONT-END

O mecanismo de transfência de dados para a interface, ou *front-end*, foi elaborado utilizando-se a linguagem PHP. Para fazer a apresentação dos dados obtidos do *back-end* utilizou-se o mecanismo de *template* SMARTY#. A função de um mecanismo de *template* é transferir os dados gerados pelo *back-end* para o modelo de uma *webpage*. Nessa *webpage*, campos especificados por uma linguagem própria do *template* são então

alterados e os dados inseridos da maneira correta. Dessa forma, utilizando-se este mecanismo, a lógica na qual os dados são manipulados independe do modo de apresentação destes. Assim, o *website* se torna mais flexível a mudanças e os problemas existentes são facilmente identificados.

Além do mecanismo de *template* em PHP, também foi utilizada a linguagem *JavaScript*. Nesse caso, para fazer o mecanismo de carregamento da tabela do DAD dentro do Dicionário Criativo. Assim, com uma técnica conhecida como AJAX, a tabela do DAD consegue ser carregada paralelamente ao restante do *website*, oferecendo ao usuário a impressão de velocidade na apresentação da *webpage*.

7.6 SÍNTESE DA SEÇÃO

A presente Seção relatou, etapa por etapa, um exercício de aplicação do modelo, desenvolvido ao longo de toda a tese e sistematizado na Seção 6, utilizando-se de dados reais de três domínios léxico-conceituais: AMOR, CORPO HUMANO e TEATRO.

O exercício de aplicação do modelo proposto demonstrou trazer consigo inúmeros desafios. Desafios de ordem lexicográfica e computacional. Mesmo trabalhando com apenas três domínios léxico-conceituais, o volume de dados foi considerável. Principalmente, no que diz respeito a encontrar o pareamento entre os diferentes CSs e CSAs potencialmente equivalentes. Isso porque as diferentes fontes de informação lexical trabalham com diferentes pressupostos e não possuem uma regularidade, por exemplo, no número de acepções para cada UL de entrada. Será um grande desafio estender a aplicação do modelo para os demais domínios léxico-conceituais e as demais classes gramaticais de cada um deles. No entanto, estamos confiantes que os resultados foram satisfatórios e, com alguns ajustes e a sua futura expansão para todo o léxico, deverão servir para os propósitos inicialmente estabelecidos. Ou seja, servir como conteúdo para um dicionário analógico digital, integrado ao *website* Dicionário Criativo, e, ao mesmo tempo, como inteligência semântica para diversas aplicações que necessitem de uma rede semântica difusa.

Fez parte do escopo dessa seção, também, descrever a maneira pela qual se dará a integração do DAD com o *website* Dicionário Criativo. Para tanto, foram dados alguns exemplos de integração, considerando o funcionamento do *website* que já está no ar em

versão *beta*, tanto na parte de *back-end* quanto na parte de *front-end*. Foram apresentados ainda alguns esboços de interface gráfica e de usabilidade a serem implementados no produto final, de forma a demonstrar como os dados resultantes desse trabalho serão utilizados pelo sistema do *website* Dicionário Criativo e, também, como serão exibidos ao usuário final.

Por fim, a presente Seção discorreu sobre a etapa computacional que está sendo desenvolvida, paralelamente ao escopo dessa tese, para implementar e disponibilizar o DAD em ambiente *Web*. Como se trata de uma terceira etapa – a etapa computacional –, que dá continuidade ao que foi proposto como finalidade para a presente tese, trata-se de um trabalho ainda em andamento.

SEÇÃO 8 – CONCLUSÃO

O objetivo geral desta tese foi propor um modelo da parte léxico-conceitual de uma aplicação linguístico-computacional correspondente a um *Dicionário Analógico Digital*. Tal modelo deverá, em trabalhos futuros, ser ampliado para toda a língua e, devidamente integrado à base de dados e contribuindo para a lógica de funcionamento do *website* Dicionário Criativo, deverá servir como ferramenta de auxílio à escrita. Dessa forma, espera-se que seja utilizado por escritores, jornalistas, publicitários e demais produtores de conteúdo, contribuindo para a produtividade desses profissionais e para a qualidade de seus trabalhos.

A tese foi desenvolvida seguindo uma metodologia de pesquisa que integra três domínios mutuamente complementares: o linguístico, o linguístico-computacional e o computacional. O domínio computacional foi apenas sugerido, como parte de um trabalho em andamento, na medida em que não fez parte do escopo desta tese. Os domínios linguístico e linguístico-computacional estruturaram as seções intermediárias da tese (Seções 3, 4 e 5), em que determinadas questões, levantadas na Seção 2, exigiam uma hipótese a ser investigada e solucionada dentro destes dois domínios. A Seção 6 propôs o modelo de representação do DAD e a Seção 7 teve como objetivo implementar o modelo para três domínios léxico-conceituais, de forma a validar a proposta. A seguir, apresentamos nossas conclusões a respeito dos objetivos, hipóteses e resultados de cada seção.

A Seção 2 serviu para localizar a nossa proposta dentro da tradição lexicográfica dos *thesaurus* e dicionários analógicos. A partir da exposição e da reflexão sobre o estado da arte desse tipo de obra lexicográfica, pudemos sugerir algumas inovações e construir algumas hipóteses que serviram para orientar o restante do trabalho. Afinal, não é possível propor novidades para um determinado domínio sem antes mapeá-lo com certo grau de profundidade. Definimos que os dicionários analógicos são obras em que o percurso se faz do nível conceitual para o nível lexical. Tal direção, de ordem onomasiológica, satisfaz uma função codificadora, como vimos, na medida em que auxilia os consulentes a encontrarem a palavra ou expressão adequada para uma ideia em mente. O que nos levou à hipótese de que a estrutura deste tipo de obra, desde que adaptada para o ambiente digital e renovada a partir de alguns pressupostos, deveria vir

a funcionar como o 'cérebro' do *website* Dicionário Criativo, cujo objetivo é exatamente servir como uma ferramenta de auxílio à escrita disponível na internet. Alguns comentários dos usuários da versão *beta* do Dicionário Criativo mostram que o objetivo da ferramenta está sendo cumprido e acreditamos que o será ainda mais depois da integração com o DAD.

A pesquisa perpetrada na Seção 2 constatou que não há, ainda, para o PB, um dicionário analógico digital que seja resultado de pesquisa acadêmica ou que respeite critérios de tradição lexicográfica. Tal constatação reforçou nossa justificativa de se construir um modelo para um dicionário analógico em suporte digital que, no futuro, pudesse orientar a implementação de uma ferramenta deste tipo para o PB, em consonância com ferramentas semelhantes já desenvolvidas para outros idiomas.

Diante do diagnóstico efetuado ao longo da Seção 2, foram identificados três principais limitações das obras em uso corrente e foram sugeridas soluções para cada uma delas, à luz dos avanços nas pesquisas em Semântica Lexical e, dentro do contexto de uma migração do suporte impresso para o suporte digital.

Primeiramente, identificamos que as relações semânticas paradigmáticas e as relações semânticas sintagmáticas não são explicitadas no interior dos grupos de analógicos em nenhuma das obras analisadas. Isto posto, sugerimos a necessidade de se explicitar essa diferença. Esse desafio foi trabalhado ao longo da Seção 3, valendo-se do estudo das redes *wordnets* como modelo de estrutura fundada em relações semânticas paradigmáticas e do estudo das redes *framenets* como parâmetro de uma arquitetura pautada por relações sintagmáticas. A investigação de tais modelos, de seus constructos e das teorias por detrás de cada um foram imprescindíveis para a construção do nosso próprio modelo na Seção 6 e, posteriormente, para a modelagem dos dados ao longo da Seção 7. Nosso modelo, nossos constructos e nossa metodologia foram todos construídos em analogia ao modelo, aos constructos e às metodologias das *wordnets* e das *framenets*. Utilizar os dois paradigmas em concomitância foi o que nos permitiu chegar aos resultados que chegamos, ou seja, uma estrutura léxico-conceitual abrangente e coerente com os nossos propósitos. Quando representamos, por exemplo, conforme os resultados da Seção 7.1, a UL *namorado* como elemento de um conjunto de sinônimos que instancia a categoria conceitual NAMORADO que, por sua vez, é elemento central do *frame* *Relação_amorosa* dentro do domínio léxico-conceitual AMOR, estamos, direta ou indiretamente, nos utilizando de seus constructos e do arcabouço teórico que

lhes dá sustentação. Com relação às investigações acerca do equacionamento linguístico-computacional a Seção 3 retrocedeu a noções essenciais da Teoria das Redes para só depois chegar ao nosso interesse específico em redes léxico-semânticas. Esse retorno à teoria mais fundamental foi importante para tomar conhecimento dos diferentes tipos de redes existentes e de suas topologias características. Foi dessa perspectiva mais ampla que surgiu o *insight* de que, quando formalizadas em grafos, as relações sintagmáticas formam redes randômicas, enquanto que relações paradigmáticas geram redes livres de escala. Se, futuramente, como é nosso propósito, exibirmos os resultados consolidados em uma visualização de rede, como em um *Visual Thesaurus*, veremos, por exemplo, que do domínio CORPO HUMANO, em que predominam relações sintagmáticas entre as categorias conceituais, surgirá uma rede randômica. Do domínio AMOR, ao contrário, em que predominam conceitos abstratos e a relação preponderante entre as categorias conceituais é a relação paradigmática, veremos emergir uma rede livre de escala, com apenas alguns aglomerados randômicos – no caso, entre as categorias conceituais do *frame* *Relação_amorosa*. Queremos ressaltar que, apesar de nossos resultados não serem apresentados visualmente em forma de rede, com links e nós, sua lógica e formalismo funcionam como uma rede, como na WN.Pr, com a diferença de que os nossos resultados constituem rede difusa.

Outra questão levantada ao longo da Seção 2 referiu-se à estrutura hierárquica dos dicionários analógicos e, principalmente, com relação às Grandes Categorias no topo da hierarquia. Constatou-se que os planos de classificação das ideias e os quadros sinópticos de categorias são idiossincráticos, o que dificulta a consulta por parte do usuário final. Tal constatação motivou a nossa proposta de simplificação do sistema hierárquico conceitual do nosso modelo em relação àquele consagrado na tradição do *Roget's Thesaurus*. Tendo como paradigma o modelo do *Historical Thesaurus of the Oxford English Dictionary* e como referencial teórico os preceitos da Linguística Cognitiva, a Seção 4 defendeu e propôs uma estrutura hierárquica encabeçada por três Grandes Categorias: MUNDO FÍSICO, MUNDO PSÍQUICO e MUNDO SOCIAL. Além, ainda, da intersecção entre cada duas dessas três Grandes Categorias: PSÍQUICO-FÍSICO, PSÍQUICO-SOCIAL e FÍSICO-SOCIAL.

A estrutura hierárquica foi simplificada também no sentido vertical, ou seja, com relação ao número de níveis de uma categoria conceitual até uma Grande Categoria. O modelo proposto possui apenas três níveis conceituais. Uma determinada UL pertence a

i) um CdSA (categoria conceitual), que pertence a ii) um *anaset* (domínio léxico-conceitual) que, por sua vez, integra iii) uma das Grandes Categorias. A única exceção a esse percurso é quando um CdSA faz parte de um *frame* e, portanto, essa dimensão conceitual é explicitada, como no exemplo da UL *namorado* dado anteriormente. Ou seja, em consequência do que foi diagnosticado na Seção 2 e do que foi discutido e defendido na Seção 4, o modelo aqui proposto não mantém nenhuma das classes, divisões e subdivisões do sistema conceitual consagrado pelo *Roget's Thesaurus* e replicado no *Dicionário Analógico da Língua Portuguesa* do professor Azevedo. No entanto, foi mantida uma correlação, em termos de granularidade, com os mil verbetes da tradição rogetiana. Os *anaset*s AMOR, CORPO HUMANO e TEATRO, correspondem, respectivamente, aos verbetes **897. Love**, **440e. Body Parts** e **599. Drama** do *Roget's Thesaurus* original (ROGET, [1852] 1911). Para o trabalho futuro de expandir o modelo aqui desenvolvido para toda a língua, estamos considerando, pelo menos inicialmente, os mil verbetes do *Roget's Thesaurus* como referência.

É importante ressaltar que um *anaset*, e seus CdSAs, não se limitam a elencar conceitos relacionados exclusivamente a uma determinada Grande Categoria. No entanto, o *anaset* deve ser prototipicamente incluído em uma delas, sempre considerando os CdSAs centrais do domínio como referência para a inclusão nesta ou naquela Grande Categoria. Como dissemos, as Grandes Categorias são dimensões entrecruzadas e a dispersão de um mesmo conceito lexicalizado entre elas é resultado, como era o desejado, da estrutura rizomática e difusa dos *anaset*s. Conforme o modelo proposto evolua de um simples protótipo para as mil categorias de Roget, veremos surgir conexões analógicas impensáveis no suporte impresso, o que deverá estimular o trabalho criativo, conforme os objetivos inicialmente aventados. Em trabalhos futuros, com o DAD disponibilizado na *Web* via Dicionário Criativo, os próprios usuários poderão incluir ou fazer surgir novas relações analógicas mediante i) um sistema de monitoramento dos cliques em *links* que ajudará a criar trilhas ou mesmo clareiras na topologia da rede e ii) ferramentas de inclusão ou exclusão de nós e arcos respaldados na inteligência coletiva de milhares de usuários.

Por fim, ao longo da Seção 2, constatamos que os grupos de palavras dos dicionários analógicos carecem de critérios explícitos para a sua organização interna. Não é clara a distinção entre as categorias conceituais de um mesmo domínio conceitual. Como vimos, tal constatação é inclusive declarada no tópico “como usar este dicionário”,

do Dicionário Analógico (AZEVEDO, 2010, p. xi), em que diz que “os grupos não têm uma estrutura lógica, embora as palavras estejam, geralmente, agrupadas por proximidade semântica”. Se, por um lado, o grande número de ULs sob um mesmo domínio léxico-conceitual é o grande diferencial desse tipo de obra em relação aos dicionários de sinônimos, a ausência de critérios para agrupar e dispor a informação léxico-semântica dentro dos grupos de palavras faz com que esse volume se torne um problema para o usuário. Dessa forma, foi necessário a produção de critérios que organizassem a informação léxico-semântica no interior dos grupos de palavras. A solução encontrada, detalhada ao longo da Seção 5, foi utilizar conjuntos de sinônimos para organizar os grupos de analógicos. Ou seja, recortar os volumosos grupos de palavras analógicas em grupos menores, utilizando como medida os conjuntos de sinônimos (CSs) extraídos de diferentes dicionários de sinônimos. Como cada conjunto de sinônimos evoca, idealmente, uma categoria conceitual, os grupos de palavras passaram, dentro do nosso modelo, a ser internamente bem estruturados. No entanto, a comparação entre dicionários analógicos e dicionários de sinônimos, levada a cabo na Seção 2, reforçou a nossa hipótese de que a distinção entre sinônimos e analógicos, em certo sentido, é nebulosa, com a evidente excessão de quando se trata de uma relação sintagmática e que, portanto, só pode se dar entre categorias conceituais explicitamente distintas. Mas, categorias conceituais em relação paradigmática, dentro de um mesmo domínio léxico-conceitual, possuem entre si, conforme os resultados apresentados na Seção 7, uma grande permissividade. Seus limites, com relação às ULs que lhes instanciam, são verdadeiramente difusos. Fica evidente, agora, conforme se olhe para os dados extraídos das diferentes fontes de informação léxico-conceitual na Seção 7, que os conjuntos de sinônimos (que instanciam categorias conceituais) são construídos de maneiras diferentes a partir de metodologias e concepções diferentes. Alguns dicionários de sinônimos, como o [MIC], por exemplo, tendem a construir os seus CSs mantendo uma relação de sinonímia bastante restritiva e, por isso, seus CSs contêm, normalmente, poucas ULs. Outros dicionários, como o [HOU], por exemplo, são mais tolerantes com relação aos critérios de inclusão de uma determinada UL como instanciadora de uma determinada categoria conceitual e, portanto, seus CSs são mais ricos em número de ULs. Dessa forma, decidimos não trabalhar com um constructo representacional para o agrupamento de ULs que possuísse limites bem definidos e que pretendesse instanciar um conceito unívoco. Ao contrário, percebemos que poderíamos utilizar as

idiosincrasias dos dicionários de sinônimos e dicionário analógicos para calcular a frequência que cada UL possui entre CSs e CSAs equivalentes e, conseqüentemente, inferir sua prototipicidade dentro de um novo conjunto, dessa vez difuso. Como vimos, as particularidades de cada obra fazem com que os CSs e CSAs se correspondam em uma espécie de *semelhança de família*. Calcular a frequência com que cada UL aparece em CSs e CSAs correspondentes, permitiu que descobrissemos, conforme os resultados apresentados na Seção 7, quais as ULs que melhor lexicalizam determinada categoria conceitual, ao mesmo tempo em que preserva uma longa cauda de ULs que, mesmo distantes do núcleo de sentido da categoria, mantêm com ela alguma relação. E é essa cauda longa que interliga os diferentes sentidos dentro de um mesmo domínio ou mesmo para fora dele, preservando as conexões analógicas. Trabalhos futuros poderão calcular o grau de relação entre diferentes categorias conceituais por meio do cômputo das ULs (e heurísticas com relação aos seus pesos) que elas compartilham e, possivelmente, inferir inclusive o tipo de relação semântica.

Como dissemos em 5.3, é provável que os resultados obtidos segundo tal metodologia sejam próximos aos resultados obtidos por centenas ou milhares de dados coletados diretamente a partir de questionários aplicados aos falantes de uma comunidade linguística homogênea. No entanto, apenas outra pesquisa, voltada para este propósito, poderá indicar se isso de fato é verdade. Futuramente, pensamos em aplicar uma pesquisa como esta, em meio digital, valendo-se da própria plataforma do Dicionário Criativo e os seus milhares de usuários. E, dessa forma, comparar os resultados obtidos mediante as duas metodologias.

O Dicionário Criativo foi o vencedor do Prêmio Santander de Empreendedorismo 2011 na categoria de Tecnologia da Informação e Comunicação (TIC). O prêmio é destinado aos alunos de graduação e pós-graduação e tem como objetivo reconhecer e apoiar o desenvolvimento de projetos e tecnologias inovadoras. O aluno⁶⁶ e o professor orientador⁶⁷ desta tese receberam como prêmio um curso de empreendedorismo no *Babson College* (Boston – MA) com todos os custos de transporte, estadia e alimentação pagos. E o projeto recebeu um apoio financeiro no valor de R\$50 mil. O projeto também foi um dos selecionados para participar do programa de aprimoramento tecnológico do SEBRAE, chamado SEBRAETEC, tendo recebido um apoio financeiro no valor de R\$24

⁶⁶ Felipe Iszlaji de Albuquerque. Lattes: <http://lattes.cnpq.br/8967679794366528>

⁶⁷ Bento Carlos Dias da Silva. Lattes: <http://lattes.cnpq.br/2202787473720818>

mil. Esses recursos permitiram a implementação computacional da versão *beta* do Dicionário Criativo e o seu respectivo aprimoramento. Ainda em 2011, o projeto foi um dos sete selecionados para participar do programa de formação chamado Empreendedores Criativos. Conforme o *website* do programa⁶⁸, o “Empreendedores Criativos é um programa que reúne pesquisa, informação, qualificação profissional e uma rede nacional de empresários e gestores em busca de inovação, novas referências e ferramentas adequadas ao desenvolvimento de negócios criativos”. Seis meses após o lançamento da versão *beta*, em junho de 2012, o *website* conquistou mais de 22 mil usuários mensais e mais de 200 mil pesquisas efetuadas por mês.

Desde o início de 2012, o projeto do Dicionário Criativo foi formalizado e se tornou microempresa, incubada no Centro de Inovação, Empreendedorismo e Tecnologia (CIETEC), com sede no campus da USP/ Butantã. Em novembro do mesmo ano, fomos contemplados no edital do programa RHAЕ Pesquisador na Empresa do CNPq⁶⁹ e, desde de o início de 2013, estamos trabalhando com dois pesquisadores⁷⁰ vinculados ao NILC com o objetivo de dar continuidade à presente pesquisa e replicar o modelo aqui proposto para todos os mil verbetes do *Roget’s Thesaurus* e, também, expandir para as demais categorias gramaticais: verbo, adjetivo e advérbio. No entanto, para dar conta desse volume de trabalho em apenas dois anos, serão feitos alguns aperfeiçoamentos na metodologia de construção dos *anaset*s como, por exemplo, a proposição de heurísticas e a respectiva construção de algoritmos que tornem os procedimentos de algumas etapas em processos semi-automáticos. Esse trabalho deverá também ampliar o número de dicionários, *thesaurus* e recursos lexicais utilizados como fontes de informação lexical com o objetivo de refinar os resultados. Deverá, ainda, incluir no algoritmo outras heurísticas como, por exemplo, atribuir pesos diferentes para cada fonte de informação lexical. Em parceria com o CIETEC e com o NILC, o projeto (no. do processo 456448/2012) tem como pesquisadores co-executores, especialistas em PLN, os professores doutores Thiago Alexandre Salgueiro Pardo⁷¹ e Ariani Di Felippo⁷². Além de gerar um diferencial competitivo para a microempresa Dicionário Criativo, a proposta apresentada tem o potencial de influir fortemente sobre as

⁶⁸ <http://www.empreendedorescriativos.com.br>, acessado em março de 2013.

⁶⁹ Chamada Pública MCTI/SETEC/CNPq Nº 17/2012 RHAЕ Pesquisador na Empresa.

⁷⁰ Bolsistas vinculados ao projeto em março/2012: i) Pedro Paulo Balage (<http://lattes.cnpq.br/1018152686077947>) e ii) Fernando Antônio Azevedo Nóbrega (<http://lattes.cnpq.br/0167275230724673>)

⁷¹ <http://lattes.cnpq.br/7078737101024368>

⁷² <http://lattes.cnpq.br/8648412103197455>

Tecnologias da Informação, da Cultura e da Educação, na medida em que o projeto do Dicionário Criativo se insere na intersecção dessas três grandes áreas. O Dicionário Criativo é uma plataforma que procura se utilizar de todo o potencial das tecnologias digitais e da internet para preservar e ao mesmo tempo propagar a língua portuguesa de maneira inovadora, escalável e de alto impacto.

Trabalhos futuros deverão, também, reunir e colocar em comparação outras propostas de redes semânticas difusas, que estão sendo desenvolvidas em paralelo, em outras partes do mundo. Como exemplo, podemos citar o projeto da Onto.PT⁷³, que encontra-se em desenvolvimento, no âmbito do doutoramento de Hugo Gonçalo Oliveira, dentro da Universidade de Coimbra, sob a orientação do prof. Dr. Paulo Gomes. Dessa forma, futuras pesquisas poderão propor o alinhamento de redes semânticas difusas de diferentes países.

⁷³ <http://ontopt.dei.uc.pt>, acessado em março de 2013.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLAN, K. *Natural language semantics*. Malden, MA: Blackwell Publishers, 2001.
- ALLEN, J. *Natural Language Understanding*. Redwood City, CA: Benjamin-Cummings Publishing Co., Inc, 1995.
- ANGIONI, L. *Introdução à Teoria da Predicação em Aristóteles*. Campinas: Unicamp, 2006.
- ATKINS, B. T. S.; RUNDELL, M. *The Oxford guide to practical lexicography*. New York: Oxford University Press, 2008.
- AZEVEDO, F. F. S. *Dicionário analógico da língua portuguesa: ideias afins*. Brasília: Thesaurus, 1983.
- BABINI, M. *Do conceito à palavra: os dicionários onomasiológicos*. *Ciência e Cultura*, abr./jun. 2006, v. 58 (2), p.38-41.
- BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The Berkeley FrameNet Project. In: *Proceedings of COLING-ACL'98*. Montreal, Quebec: ACL/ Morgan Kaufmann Publishers, 1998.
- BARABÁSI, A-L. *Linked: the new science of network*. Perseus, Cambridge, 2002.
- BARABÁSI, A-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, 509-512, 1999.
- BARBOSA, O. *Grande dicionário de sinônimos e antônimos*, 16^a ed., Rio de Janeiro: Ediouro, 2000.
- BERBER SARDINHA, A. P. *Linguística de corpus*. São Paulo, Barueri: Editora Manole, 2004.
- BERLIN, B.; KAY, P. *Basic Colors Terms*. Berkeley and Los Angeles: University of California Press, 1969.
- BIDERMAN, M. T. C. *Teoria Lingüística*. São Paulo: Martins Fontes, 1978 (2a ed. 2001).
- BOISSIERE, P. *Dictionnaire analogique de la langue française: répertoire complet des mots par les idées, des idées par les mots*. Paris: Aug. Boyer M (s.d.)
- BORBA, F. da S. *Organização de dicionários: uma introdução à lexicografia*. São Paulo: Editora UNESP, 2003.
- BRISCOE E. J., BOGURAEV, B. (eds.) *Computational Lexicography for Natural Language Processing*. London: Longman, 1989.
- CASARES, J. *Diccionario ideológico de la lengua española*. Barcelona: G. Gili, 1959.
- COLLINS, A. M., QUILLIAN, M. R. Retrieval time from semantic memory. *Journal of verbal behavior and verbal learning*, v. 8, 240-247, 1969.
- CORREIA, M. *Os Dicionários Portugueses*. Lisboa: Caminho, 2009.
- CROFT, W., CRUSE, A. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.
- CRUSE, D. A. *Lexical semantics*. New York: Cambridge University Press, 1986.

____. *A Glossary of Semantics and Pragmatics*. Edinburgh: Edinburgh University Press, 2006.

CRYSTAL, D. *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge Univ. Press, 1997.

DELBECQUE, N. *Linguística cognitiva: compreender como funciona a linguagem*. Lisboa: Instituto Piaget, 2008.

DELEUZE G, GUATTARI F. *O anti-edipo: capitalismo e esquizofrenia*. Rio de Janeiro: Imago, 1976.

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, 612-7, 1998.

____. O estudo lingüístico-computacional da linguagem. *Letras de Hoje*, Porto Alegre, v. 41, n. 2, p. 103-138, 2006.

____, MORAES, H. R. A construção de thesaurus eletrônico para o português do Brasil. *Alfa*, São Paulo; Editora da UNESP, v. 47(2), p. 101-115, 2003.

DIAS-DA-SILVA, BC; FELIPPO, A. Di. Uma introdução à engenharia do conhecimento lingüístico. *Revista de Letras*, v. 1, 57-72, 2008.

DIAS-DA-SILVA, B.C.; DI FELIPPO, A., NUNES, M.G.V. The automatic mapping of Princeton WordNet lexical-conceptual relations onto the Brazilian Portuguese WordNet database. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION – LREC, 6, 2008. Marrakech, Morocco. *Proceedings...* Marrakech, 2008.

DIAS-DA-SILVA, B.C.; FERREIRA, M. C.; MARCELLINO, E. R. Léxicos, ontologias e a web semântica In: *Estudos linguísticos da FCLAr: desafios na pós-graduação*. São Paulo: Cultura Acadêmica, 59-78, 2011.

DIRVEN, R.; PÖRINGS, R. (eds.) *Metaphor and Metonymy in Comparison and Contrast*. Berlin, New York: Mouton de Gruyter, 2003.

ECO, U. *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press, 1984.

Erdős, P.; Rényi, A. On random graphs. In: *Publicationes Mathematicae Debrecen*, v. 6, 290-297, 1959.

DIESSEL, R. *Graph Theory*, v. 173, Graduate Texts in Mathematics. Springer, 2^a ed., 2000. Disponível em: <http://www.math.uni-hamburg.de/home/diessel/books/graph.theory> acessado em 07/03/2012.

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

FELTES, H. P. M. *Semântica cognitiva: ilhas, pontes e teias*. Porto Alegre: Edipucrs, 2007.

FERNANDES, F. *Dicionário de sinônimos e antônimos da língua portuguesa*. São Paulo: Globo, 1997.

FILLMORE, C.J. The case for case reopened. In: P. COLE; J. SADDOCK (eds.), *Grammatical relations*. New York: Academic Press, 59-81, 1977.

- _____. Frame semantics. In: THE LINGUISTICS SOCIETY OF KOREA, *Linguistics in the morning calm*. Seoul: Hashin, p. 111-137, 1982.
- FILLMORE, C.J.; PETRUCK, M.R.L.; RUPPENHOFER, J; WRIGHT, A. FrameNet in action: the case of Attaching. *International Journal of Lexicography*, 16(1):297-332, 2003.
- FLORENZANO, E. *Dicionário de Ideias Semelhantes*. Rio de Janeiro: Ediouro, 1982.
- FRANCIS, W. N. A standard corpus of edited present-day American English for computer use. *Literary Data Processing Conference Proceedings*, 9-11 September 1964. In: BESSINGER, J.B.; PARRISH, S.M.; ARADER, H.F. (eds.), 79-89. Armonk, N.Y.: IBM Corporation, 1965.
- GEERAERTS, D. Cognitive Linguistics. In: VERSCHUEREN, J.; ÖSTMAN J.-O.; BLOMMAERT, J. (eds.) *Handbook of Pragmatics*. Amsterdam: John Benjamins, 111-116, 1995.
- GENTNER, D. Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155-70, 1983.
- GLOCK, H.-J. *Dicionário Wittgenstein*. Oxford: B. Blackwell, 1996.
- GRISHMAN, R. *Computational linguistics*. Cambridge: Cambridge University Press, 1986.
- GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: *International Journal Human-Computer Studies* (43):5-6, 1995, p.907-928.
- HAIMAN, J. The iconicity of grammar: isomorphism and motivation. *Language*, 56(3): 515-40, 1980.
- HANDKE, J. *The structure of the lexicon: human versus machine*. Berlin: Mouton de Gruyter, 1995.
- HALLIG, R.; WARTBURG, W. Von. *Begriffssystem als Grundlage für die Lexikographie*. Berlin: Akademie-Verlag, 1952.
- HARPER-COLLINS. *Collins Thesaurus of the English Language*. Glasgow: Harper-Collins Publishers, 1995 [2002].
- HANNEMAN, R.; RIDDLE, M. *Introduction to social network methods*. Riverside: University of California, 2005.
- HOLLANDA, C. B. *Tantas Palavras*. São Paulo: Cia. das Letras, 2006.
- HOUAISS, A., VILLAR, M. de S. *Dicionário eletrônico Houaiss da língua portuguesa*. (versão 1.0). Rio de Janeiro: Editora Objetiva, 2001. 1 CD-ROM.
- HOUAISS, A. *Dicionário Houaiss: sinônimos e antônimos*. Instituto Antônio Houaiss, 2 ed. São Paulo: Publifolha, 2008.
- HOVY, E. H. Combining and standardizing large-scale, practical ontologies for machine translation and other uses. In: *Proceedings of the first international conference on language resources and evaluation (LREC)*, 535-42, Granada, 1998.
- 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.
- JOHNSON, S. *De onde vêm as boas ideias*. Rio de Janeiro: Zahar, 2011.

- KATZ, J. J. *Language and other abstract objects*. Totowa, N. J.: Rowman & Littlefield, 1981.
- KAY, C. et al. (eds.). *Historical Thesaurus of the Oxford English Dictionary*. Oxford: Oxford University Press, 2009.
- KEMPTON, W. *The folk classification of ceramics: a study of cognitive prototypes*. New York: Academic Press, 1981.
- KIPFER, B. A. (ed.). *Roget's Internacional Thesaurus*. New York: Harper-Collins Reference, 70^a ed., 2010.
- KÖVECSES, Z. *The language of love: the semantics of passion in conversational English*. Lewisburg: Bucknell University Press, 1988.
- LABOV W. The boundaries of words and their meanings. In: BAILEY, C.-JN; SHUY, R. W. (eds.). *New ways of analyzing variation in English*. Washington: Georgetown University Press, 1973.
- LACERDA, R. C.; LACERDA, H. R. C.; ABREU, E. S. – *Dicionário de Provérbios: francês, português, inglês*. São Paulo: Editora UNESP, 2004.
- LAKOFF, G.; JOHNSON, M. *Metaphors We Live By*. Chicago: The Un. of Chicago Press, 1980.
- LAKOFF, G.; JOHNSON, M. *Philosophy in the Flesh*. New York: Basic Books, 1999.
- LAKOFF, G. *Women, fire and dangerous things*. Chicago: University of Chicago Press, 1987.
- LANGACKER, R. W. *Foundations of Cognitive Grammar*. Vol I e II. Theoretical Prerequisites. Stanford. California: Standford University Press, 1987 e 1991.
- LEHRER, A. *Semantic fields and lexical structure*. Amsterdam and New York: North Holland, 1974.
- LYONS, J. *Structural semantics*. Oxford, Basil Blackwell, 1963.
- 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.
- MEDNICK, S. A. The associative basis of the creative process. *Psychol Review*, 69:220–232, 1962.
- MILLER, G.; BECKWITH, R.; FELLBAUM, C.; GROSS; D.; MILLER, K. Introduction to WordNet: An on-line lexical database. In: *International Journal of Lexicography* (special issue), 3(4):235-312, 1990.
- MILLER, G. A., FELLBAUM, C. Semantic networks of English. *Cognition*, 41, n. 1-3, p. 197-229, 1991.
- MITKOV, R. (Ed.). *The Oxford handbook of computational linguistics*. Oxford, New York: Oxford University Express, 2004.
- OLD, J. L. *The semantic structure of Roget's, a whole-language Thesaurus*. PhD Dissertation. Indiana University, 2003.
- OLIVEIRA, M.F. *Pressupostos teórico-metodológicos para a elaboração da base lexical de um thesaurus eletrônico*. Araraquara, 2002, 209f. Dissertação (Mestrado em Letras) – Faculdade de Ciências e Letras, UNESP, Araraquara, 2002.

- PALMER, M. Multilingual resources, multilingual information management: current levels and future abilities. *Linguistica Computazionale*, Piza, v.14-15, p.1-33, 2001.
- PAVIS, P. *Dicionário de teatro*. São Paulo: Perspectiva, 1999.
- PEIRCE, C. S. *Semiótica*. São Paulo: Perspectiva, 2010.
- PINKER, S. *Do que é feito o pensamento: a língua como janela para a natureza humana*. São Paulo: Companhia das Letras, 2008.
- POLITO, AG. *Michaelis dicionário de sinônimos e antônimos*. São Paulo: Melhoramentos, 2009.
- PRÉVOT, L.; BORGO, S.; OLTRAMARI, A. Interfacing Ontologies and Lexical Resources. In: OntoLex 2005, Jeju Island. *Proceedings...* Jeju Island: [s.n.], p. 91-102, 2005.
- PUSTEJOVSKY, J. *The generative lexicon*. Cambridge, Mass.: The MIT Press, 1995.
- QUILLIAN, M. R. Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12, p. 410-430, 1967.
- QUILLIAN, M. R. Semantic Memory. In: M. Minsky. *Semantic information processing*. Cambridge, Mass.: MIT Press. p. 227-70, 1968.
- RAE, K.; SOUTHERN, R. *Lexique International de termes techniques the théâtre*. Bruxellas: L'institut International du Théâtre, 1964.
- RIVA, H. C. *Dicionário onomasiológico de expressões idiomáticas usuais na língua portuguesa do Brasil*. São José do Rio Preto, 2008, 315 f. Tese (Doutorado em Letras) - Instituto de Biociências, Letras e Ciências Exatas, UNESP, São José do Rio Preto, 2008.
- 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] 1911.
- ROHRER, T. Embodiment and experientialism. In: GEERAERTS, D.; CUYCKENS, H. (eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford Univ. Press, 25-47, 2007.
- ROSCH, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology, General*, 104, 192-233, 1975.
- ROSCH, E.; LLOYD, B. B. *Cognition and Categorization*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.
- RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M.; JOHNSON, C.; SCHEFFCZYK. *FrameNet II: Extended theory and practice*. Disponível em: <http://framenet.icsi.berkeley.edu/> acessado em: 21/09/2009.
- SALOMÃO, M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*, v. 7, n. 3, p. 171-182, 2009.
- SAUSSURE, F. *Curso de Linguística Geral*. São Paulo: Cultrix, 1916 (2006).
- SCHANK, R. C.; ABELSON, R. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum, 1977.

- SCHANK, R. C.; KASS, A. Knowledge representation in people and machines. In: ECO, U.; SANTAMBROGIO, M.; VIOLI, P. (eds.). *Meaning and Mental Representations*. Indianapolis: Indiana University Press, 181-200, 1988.
- SILVA, P. N. *Dicionário de Citações*. Lisboa: Âncora Editora, 2012.
- SPITZER, C. S. J. *Dicionário Analógico da Língua Portuguesa*. 2. ed. Porto Alegre: Livraria do Globo, 1952.
- TAYLOR, J. R. *Linguistic Categorization: Prototypes in linguistic Theory*. Oxford: Clarendon Press, 1989.
- ULLMANN, S. *Semântica: uma introdução à ciência do significado*. Trad. de J.A. Osório Mateus. Lisboa: Fundação Calouste Gulbenkian, 1964.
- NIRENBURG, S.; BEALE, S.; MAHESH, K.; ONYSHKEVYCH, B.; RASKIN, V.; VIEGAS, E.; WILKS, Y.; ZAJAC, R. Lexicons in the Mikrokosmos Project. In: *Proceedings of the Artificial Intelligence and Simulated Behavior Workshop on Multilinguality in the Lexicon*. Brighton, 1996.
- VISUAL THESAURUS. Version 3.0. Thinkmap Inc., 2005.
- VOSSSEN P. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Dordrecht: Kluwe, 1998.
- VOSSSEN, P. Ontologies. In: MITKOV, R. (ed.). *The Handbook of Computational Linguistics*. Oxford: Oxford University Press, 464-482, 2003.
- ZADEH, L. A. Fuzzy Sets. In: *Information and Control*. v.8, p.338-353, 1965.
- ZUGMAN, F. *O mito da Criatividade: Desconstruindo verdades e mitos*. Rio de Janeiro: Elsevier, 2008.
- WIERZBICKA, A. *Semantic primitives*. Frankfurt am Main, Athenaum, 1972.
- WITTGENSTEIN, L. *Philosophical Investigations*. Oxford: Basil Blachwell, 1953.

APÊNDICE I – CORPUS DA PESQUISA

O *corpus* da pesquisa é constituído das letras das canções do álbum intitulado *Grande Circo Místico*, com composições de Chico Buarque e Edu Lobo (ver Anexo I). As canções foram especialmente compostas como trilha sonora para o espetáculo homônimo de dança e teatro montado na década de 1980 pelo Balé Teatro Guaíra, de Curitiba. As letras e as canções ultrapassaram os limites do espaço teatral e foram imortalizadas no álbum de 1982 que reuniu grandes intérpretes da Música Popular Brasileira: Gal Costa, Gilberto Gil, Jane Duboc, Tim Maia, Simone e Zizi Possi. As canções versam sobre essa atmosfera de circo, dança e teatro, com personagens como palhaços, dançarinas, atores e atrizes, além de temas românticos. Selecionamos essa obra como *corpus* porque, conforme 2.4, Chico Buarque é declaradamente usuário do Dicionário Analógico do professor Azevedo e é reconhecido por escrever letras de grande elaboração poética.

O álbum possui dez faixas, sendo nove delas com letras e uma instrumental. As letras possuem, proporcionalmente ao tamanho do álbum, uma grande quantidade de ULs dos domínios léxico-conceituais CORPO HUMANO, AMOR e ARTES (de onde extraímos as ULs do domínio TEATRO). No Quadro I.1, elencamos o número da faixa no disco, o título e o ano da canção e, na última coluna, as ULs que foram extraídas e que compõe o nosso *corpus* inicial.

Quadro I.1 – ULs extraídas por faixa do álbum *Grande Circo Místico*.

Grande Circo Místico (1982) – Chico Buarque e Edu Lobo		
faixa	título/ ano	ULs
1	Beatriz (1982)	<i>pintura, rosto, atriz, dança, papel, cenário, comédia, pagantes, bis</i>
2	A Bela e a Fera (1982)	<i>corpo, coração</i>
3	Na Carreira (1982)	<i>amante, artista, coração, boca, perna, pele, palco</i>
4	Ciranda da Bailarina (1982)	<i>bailarina, namorado, orelha, bigode</i>
5	O Circo Místico (1982)	<i>artista, palco, dançarina, espectador, elenco</i>
6	A História de Lily Braun (1982)	<i>romance, olhos, cinema, show, star, beijo</i>

7	Opereta do Casamento (1982)	<i>noiva, palhaço, corista, dançarina, maestro, cortina, pé</i>
8	Valsa dos Clowns (1982)	<i>palhaço, boca, coração, coxia, arquibancada</i>
9	Sobre Todas as Coisas (1982)	<i>(instrumental)</i>
10	Meu Namorado (1982)	<i>namorado, bem</i>

Foram extraídos apenas os substantivos dos domínios AMOR (i), CORPO HUMANO (ii) e TEATRO (iii), totalizando 38 ULs, divididas conforme abaixo:

- (i) Domínio AMOR (6 ULs): amante; romance; bem; namorado; beijo; noiva.
- (ii) Domínio CORPO HUMANO (10 ULs): rosto; corpo; coração; pele; boca; perna; orelha; bigode; olho(s); pé.
- (iii) Domínio TEATRO (22 ULs): pintura; atriz; dança; papel; cenário; comédia; pagante(s); bis; artista; palmas; palco; cortina; coxia; palhaço; show; corista; arquibancada; maestro; *star*; cinema; dançarina; bailarina.

Cada uma dessas listas é o ponto de partida para se gerar as listas de todas as ULs pertencentes a cada um dos domínios léxico-conceituais modelados. Tendo como base as ULs extraídas do *corpus*, procura-se nas fontes de informação léxico-conceitual (Apêndice II.1) as ULs sinônimas e análogas que possam também pertencer ao mesmo domínio, de forma a expandir o número de ULs de entrada de cada *anaset*.

APÊNDICE II – FONTES DE INFORMAÇÃO LÉXICO-CONCEITUAL

Faremos aqui, a análise de obras de referência disponíveis, publicadas em papel ou em meio digital (do português e inglês), com vistas à utilização dessas obras como fontes de conhecimento léxico-conceitual. Elas estão divididas em dois grupos, conforme o uso que faremos delas: (i) fontes de informação léxico-conceitual para extração de ULs e CSs/CSAs (II.1); (ii) fontes de informação léxico-conceitual para definição dos sentidos das ULs, delimitação das categorias conceituais e extração das glosas (II.2).

Diante da carência de dicionários do PB computacionalmente tratáveis, procedeu-se a seleção de fontes de informação léxico-conceitual que, apesar de suas limitações, já são amplamente utilizadas para muitas questões lexicográficas da língua portuguesa e mostraram-se também apropriadas para os propósitos desta tese.

II.1. FONTES DE INFORMAÇÃO LÉXICO-CONCEITUAIS PARA COLETA DE ULs E DE CSs/CSAs

Os dicionários de sinônimos e dicionários analógicos listados a seguir foram selecionados como fonte de informação léxico-conceitual para a coleta das ULs e extração dos CSs/CSAs que servirão para o preenchimento da matriz dos *anaset*s. Todos os dicionários de sinônimos selecionados se utilizam da noção de sinonímia contextual. Além disso, são os mais representativos em língua portuguesa, atestação que se faz por serem os mais referenciados e utilizados em outras pesquisas da área.

A marcação entre chaves à direita dos nomes das fontes é a sigla que ao longo da análise dos dados será utilizada para identificar a procedência das informações.

i) Dicionários de Sinônimos

- a) HOUAISS, A. *Dicionário Houaiss de Sinônimos e Antônimos*; [HOU]
- b) POLITO, AG. *Michaelis Dicionário de Sinônimos e Antônimos*; [MIC]
- c) FERNANDES, F. *Dicionário de Sinônimos e Antônimos da Língua Portuguesa*; [FER]
- d) TEP 2.0 (<http://www.nilc.icmc.usp.br/tep2/>); [TEP]

ii) Dicionários Analógicos/ Onomasiológicos

- e) AZEVEDO, F.F.S. – *Dicionário Analógico da Língua Portuguesa*; [AZE]
- f) FLORENZANO, E. *Dicionário de Ideias Semelhantes*. [IDS]

g) SBS - *Dicionário Visual do Corpo Humano*; [SBS]

Quando falamos em CSs, estamos nos referindo ao constructo resultante da extração padronizada (cf. 6.1.2.3) do conhecimento léxico-conceitual sobre ULs sinônimas compilado nas obras elencadas em (i). Quando falamos em CSAs, referimo-nos ao constructo resultante da extração padronizada do conhecimento léxico-conceitual compilado nas obras do tipo (ii), ou seja, dicionários com estrutura onomasiológica, que agrupam ULs sinônimas e análogas por campos semânticos e domínios léxico-conceituais. É possível perceber em (ii) que o *Dicionário Analógico da Língua Portuguesa* do padre Carlos Spitzer (1952) não foi incluído como fonte de extração de CSAs. Isso porque o seu conteúdo é muito parecido com o do *Dicionário Analógico da Língua Portuguesa* do professor Azevedo ([1950] 2010). Na escolha entre os dois, pesou a importância e maior reconhecimento deste último, sendo ao mesmo tempo o primeiro dicionário do tipo em língua portuguesa e o único com edição atualizada e revista.

Segue abaixo a especificação técnica e quantitativa de cada uma das obras selecionadas:

a) HOUAISS, A. *Dicionário Houaiss de Sinônimos e Antônimos*; [HOU]

Lançado em 2008 (2ª edição) pela editora PubliFolha, o *Dicionário Houaiss de Sinônimos e Antônimos* é, das obras comerciais aqui selecionadas, a que possui melhor fundamentação teórica e melhor estruturação interna dos verbetes. Na introdução da obra (p. x-xiv), Mauro de Salles Vilar, membro da Academia Brasileira de Filologia e diretor do projeto, resgata a problemática da sinonímia desde a antiguidade até fins do século XX, citando, por exemplo, autores como LYONS (1968), ULLMANN (1977) e CRUSE (1986).

Conforme os dados apresentados na introdução (p. xiii), o *Dicionário Houaiss de Sinônimos e Antônimos* registra 196 mil ULs agrupadas em um número de CSs não indicado e distribuídas em 20.158 entradas. Cada CS ligado a uma acepção da UL é introduzido por uma palavra-chave em negrito que “categoriza a identidade semântica da sinonímia que apresenta para aquela acepção” (p. xii). Além disso, cada acepção é introduzida por um número e termina com um exemplo em itálico e entre chaves. Exemplo do verbete *amor* (omitindo-se os antônimos):

amor /ô/ - s.m. **1 afeição**: afeto, amizade, apego, apreço, benquerença, benquerer, carinho, estima, simpatia, ternura <a. a país, amigos, animais> **2 ambição**: apego, apetite, cobiça, fome, ganância, interesse, sede, sofreguidão <a. ao poder> **3 amizade**: afinidade, fraternidade <a. pelos colegas> **4 atração**: desejo, fascínio, interesse, paixão <a. ao perigo> **5 coito**: cópula <fazer a.> **6 dedicação**: cuidado, fidelidade, zelo <a. ao trabalho> **7 fig. devoção**: adoração, entusiasmo, respeito, veneração <a. à pátria, a Deus> **8 relacionamento**: aventura, caso, namoro, romance <sei de todos os seus a.>

b) POLITO, AG. *Michaelis Dicionário de Sinônimos e Antônimos*; [MIC]

Lançado em 2009 (2ª edição) pela editora Melhoramentos, o *Michaelis Dicionário de Sinônimos e Antônimos* possui 18 mil verbetes. No entanto, nem a obra nem o catálogo oficial da editora informam o número de ULs e de CSs correspondentes.

Os verbetes do *Michaelis Dicionário de Sinônimos e Antônimos* também explicitam cada acepção por meio de um número. No entanto, raramente são apresentados exemplos dentro dos verbetes. Exemplo do verbo *amor* (omitindo-se os antônimos):

amor *sm* **1** afeição, afeto. **2** estima, simpatia, amizade. **3** adoração, devoção, culto. **4** paixão, atração, desejo. **5** atenção, diligência, cuidado.

c) FERNANDES, F. *Dicionário de Sinônimos e Antônimos da Língua Portuguesa*; [FER]

A primeira edição da obra de Francisco Fernandes foi lançada em 1944. No entanto, para este trabalho, foi utilizada a 43ª edição, lançada em 2002 pela editora Globo, com revisão, atualização e ampliação do consagrado gramático e etimólogo Celso Pedro Luft. O catálogo oficial da editora informa apenas que a atual versão consigna sinônimos de 30.000 palavras, incluindo neologismos e brasileirismos.

A diagramação do *Dicionário de Sinônimos e Antônimos da Língua Portuguesa* utiliza apenas o sinal gráfico do ponto final (.) para separar cada acepção no interior dos verbetes. Por vezes, algumas acepções são acompanhadas de trechos de autores consagrados para exemplificar o contexto de uso. Exemplo do verbo *amor* (omitindo-se os antônimos):

AMOR *Sin.* Afeto, afeição, amizade, dileção, inclinação, simpatia, paixão, bem-querer, benquerença, idolatria, adoração. Ardor, patriotismo, entusiasmo, dedicação: *Vereis um novo exemplo de amor dos pátrios feitos* (Camões). Cupido, Eros. Chama, flama, fogo. Caridade, fraternidade, filantropia, altruísmo. Idílio. Rabicho, namoro.

d) TEP 2.0 (<http://www.nilc.icmc.usp.br/tep2/>); [TEP]

Conforme a Subseção 2.5.1. o TEP 2.0 é a nova versão do TEP, um dicionário eletrônico de sinônimos e antônimos para o português do Brasil desenvolvido por pesquisadores brasileiros associados ao NILC. Atualmente, o TeP 2.0 contém 19.888 CSs e 44.678 ULs, tendo a média de 2,5 unidades por CSs.

Os verbetes do TEP 2.0 separam em linhas diferentes cada acepção e as introduz por meio de um número. Exemplo do verbo *amor* (omitindo-se os antônimos):

amor (Substantivo)

1. amor, paixão

2. amor, adoração, culto, devoção, tenção, veneração

3. amor, aplicação, atenção, carinho, cuidado, dedicação, desvelo, diligência, vigilância, vigília, zelo

4. amor, afeição, afeiçoamento, afetividade, afeto, amizade, querença, querência

e) AZEVEDO, F.F.S. – *Dicionário Analógico da Língua Portuguesa*; [AZE]

A primeira edição da obra de Francisco Ferreira dos Santos Azevedo foi lançada em 1950, oito anos após o falecimento do autor. No entanto, para este trabalho, foi utilizada a 2ª edição, lançada em 2010 pela editora Lexikon, com revisão, atualização e ampliação do lexicógrafo Paulo Geiger. Na Seção **como usar este dicionário** (p. x), a obra é apresentada como contendo quase

100 mil itens lexicais distribuídos em mais de mil grupos de analógicos (os grupos estão numerados de 1 a 1.000, mas há grupos intermediários). Conforme Seção 2.2, esses mil grupos estão organizados em três níveis de ramificação: i) 6 Classes; ii) 24 Divisões; iii) 93 Subdivisões.

Diferentemente dos dicionários de sinônimos anteriormente apresentados, os padrões gráficos dos verbetes do *Dicionário Analógico da Língua Portuguesa* são mais complexos. A diagramação do *Dicionário de Sinônimos e Antônimos da Língua Portuguesa* utiliza apenas o sinal gráfico do ponto final (.) para separar cada acepção no interior dos verbetes. Por vezes, algumas acepções são acompanhadas de trechos de autores consagrados para exemplificar o contexto de uso. Exemplo do verbete *amor* (apenas substantivos):

897. Amor, carinho, idolatria, afeto, amoricos, amorosidade, amorio (ant.), inclinação (desejo) 865; dileção, predileção, preferência, simpatia, estremecimento, benquerença, afeição, dedicação, querença, admiração, apego, aferro, constância, idílio, derriço, derretimento, idiopatia, ternura, intimidade, conchego, benevolência 906; agarramento, entranha, aspiração, galanteio, galanice; namoro, namorico, namorisco, flerte, amizade colorida, paixão, adoração, ardor, fervor, chamas, calor, devoção, atração, êxtase, enlevamento, arroubamento, enlevo, feitiçaria, namoramento, xaveco (gír.);

chama inédita; chama de cintilações desconhecidas/ de efeitos maravilhosos; amatividade, Cupido, Afrodite, Vênus, Ondim ou Ondina;

mirto, murta, setas do amor;

história/ laços/ caso/ negócio de amor; dedo de Cupido (namoro) 902; olhar amoroso, transportes amorosos, filoginia, erotídeas; amor inextinguível/ verdadeiro/ irremediável/ inabalável/ acendrado/ sincero/ ardente/ profundo/ indestrutível/ imaculado/ platônico/ paterno/ materno/ filial, conjugal/ incondicional/ inabalável/ sem fim/ infinito; piedade filial, favorito 899; popularidade, prestígio, influência; amante, proco (desus.), namorado, namorador, namoradeiro, pretensor, pretendente, admirador, vegete = amante velho, apaixonado, adorador, galanteador, cortejador, adorante, galã, amoroso, jacaré (pop.), marrancho (pop.), derriçador, quebra-esquinas, babão, Lotário, bandoleiro, conquistador, D. João (libertino) 962, Casanova;

chichisbéu, caro, esposo, bem, amigo, querido, derriço, predileto, cujo, zinho (bras.), frecheiro, beijocador, beijoqueiro, flerte, namorado (pop.), ficante (bras. gír.);

namorada, arrojada (ant.), cupida, apaixonada, querida, amorzinho, amada, Dulcineia, derriço, benzinho, predileta, anjo, querubim, serafim, ídolo, deusa, inclinação, objeto da simpatia, cotó, namoradeira, janeleira, frança, pau de cabeleira;

noivo, noiva, fiancée, pretendida, futura, nubente; casal de pombinhos, morada de amor, ninho, dois corações num só, Romeu e Julieta, Abelardo e Heloísa;

tesão, sensualidade, amor carnal, lascívia, erotismo, luxúria, volúpia, voluptuosidade, transa, sexo, cópula, sexualidade, beijo, abraço, carícia, orgasmo, gozo.

f) FLORENZANO, E. *Dicionário de Ideias Semelhantes*. [IDS]

A primeira edição do *Dicionário de Ideias Semelhantes* data de 1961. No entanto, para este trabalho, foi utilizada a edição de 1993, lançada pela editora Ediouro. Na **nota introdutória** do editor, é feita a referência ao *Thesaurus* de Roget, cuja metodologia foi “adotada em parte” (p. 5). Trata-se, de fato, de um léxico analógico, em que “as palavras se apresentam por afinidade, partindo-se do sentido para o vocábulo” (p. 5). Sua estrutura, no entanto, apresenta algumas particularidades. Diferentemente do *Roget's Thesaurus* e do *Dicionário Analógico da Língua Portuguesa*, seu plano de classificação da ideias se divide em nove categorias (aqui chamadas de NOÇÕES), iniciadas por letras maiúsculas: A) NOÇÕES ABSTRATAS; B) NOÇÕES AFETIVAS; C) NOÇÕES DINÂMICAS; D) NOÇÕES ESTÉTICAS; E) NOÇÕES FÍSICAS; F) NOÇÕES INTELLECTUAIS; G) NOÇÕES MORAIS; H) NOÇÕES RELIGIOSAS; I) NOÇÕES SOCIAIS. Cada verbete (grupo de analógico) é iniciado por uma letra (que lhe indica a categoria) e um número (que lho distingue dos demais dentro da mesma categoria). No total são 945 grupos de analógicos. O total de CSs e de ULs não é informado.

Outra particularidade é a explicitação de diferentes conjuntos de sinônimos no interior dos grupos de analógicos. Em oposição à obra do professor Azevedo, aqui as ULs que evocam um

mesmo conceito estão, na maioria dos casos, claramente agrupadas e delimitadas. Os agrupamentos se iniciam por uma palavra-chave em capitular e negrito e terminam com um ponto final. Exemplo do verbete *amor* (apenas substantivos):

B-79 **AMOR**: afeição, afeto, simpatia, comunhão de sentimentos, ternura, carinho, caridade, fraternidade, boa vontade, benevolência, apego, gosto, inclinação, chamego, cuidado, zelo, desvelo, ênfase. **ANSEIO**: ansiedade, anelo, paixão, chama, flama, paixonite, apaixonite, arrebatamento, enlevo, êxtase, idolatria, adoração, devoção, fervor, ardor. **ATRATIVO**: amavios, encantamento, encanto, popularidade, simpatia, “it”. **AMOR MATERNO**: ternura, maternal, inclinação natural, afeição de mãe. **CUPIDO**: Eros, Vênus, Afrodite; cinto, rosa, espuma do mar, cabra, mirto, esmeralda. **NAMORADO**: enamorado, galã, conquistador, Don Juan, admirador, apaixonado, pretendente, noivo, adorador, fã; amante. **NAMORADA**: apaixonada, pequena, garota, anjo, ídolo, menina, deusa, diva, noiva; amante. **NAMORO**: flerte, namorice, namorico, namorilho, namoricho, galanteio, azeite, camote, cera, derriço, grude, mormaço, paleio, pé-de-alferes, prosa; sumbaré, suruba, tribofe, xodó, agarramento. **AMADO**: querido, caro, idolatrado, estremado, estremecido, doce, rico, precioso, estimado, prezado, benquistado.

g) SBS - *Dicionário Visual SBS do Corpo Humano*; [SBS]

Utilizamos a primeira edição do *Dicionário Visual SBS do Corpo Humano*, lançada em 2009, para essa pesquisa. O *Dicionário Visual SBS do Corpo Humano* contém ilustrações para as principais partes do corpo humano e aponta a terminologia, em seis idiomas, das características e dos mecanismos dos diferentes sistemas anatômicos. A edição possui 192 páginas e foi elaborada originalmente pela editora canadense *QA International*.

Os dicionários visuais são constituídos por imagens que representam um objeto com indicação do nome de cada uma de suas partes, ou um campo léxico-conceitual com suas entidades e, eventualmente, com alguma forma de estruturação interna. Por este motivo, assim como os dicionários analógicos, nesse tipo de dicionário o percurso proposto é também de caráter onomasiológico (CORREIA, 2009, p. 46). Portanto, não são elencados, neste caso, conjuntos de sinônimos. As ULs são agrupadas por instanciarem conceitos que se relacionam por meronímia, ou seja, relações de parte-todo.

O *Dicionário Visual SBS do Corpo Humano* aparece citado como fonte de extração de CSAs porque, no caso particular do domínio CORPO HUMANO, diferentemente dos dois outros domínios, houve a necessidade de utilizá-lo no lugar do *Dicionário de Ideias Semelhantes* [IDS] como uma das seis fontes de informação lexical que estamos utilizando como padrão para a extração dos CSs e CSAs. Isso se deve ao fato de que o [IDS] possui um número restrito de CSAs relativos ao domínio CORPO HUMANO. Mantê-lo como fonte de informação para este domínio provocaria distorções nos resultados. Ao mesmo tempo, os dicionários visuais são ideais para extrair relações de parte-todo e o domínio CORPO HUMANO, como vimos, é um exemplo emblemático de domínio estruturado por relações desse tipo. Também por esse motivo, o *Dicionário Visual SBS do Corpo Humano* está arrolado não só como fonte de extração de CSAs nessa seção, como também como fonte de informação para a delimitação das categorias conceituais (do domínio CORPO HUMANO) na Seção seguinte.

II.2 FONTES DE INFORMAÇÃO LÉXICO-CONCEITUAIS PARA DEFINIÇÃO DOS SENTIDOS DAS ULS, DELIMITAÇÃO DAS CATEGORIAS CONCEITUAIS E EXTRAÇÃO DAS GLOSAS

Os dicionários monolíngues da língua portuguesa e os dicionários de especialidades listados a seguir foram selecionados como fonte de informação léxico-conceitual para a definição dos sentidos das ULS e para a identificação e delimitação de categorias conceituais, além de servirem como fonte para a extração das glosas. Os quatro dicionários monolíngues listados em (i) são considerados os mais representativos em língua portuguesa e são tradicionalmente empregados em pesquisas por conta da reconhecida qualidade e pela facilidade de acesso e manuseio. Igualmente, os dicionários de especialidades listados em (ii) possuem qualidade reconhecida e servem como complemento aos dicionários monolíngues para a definição dos sentidos das ULS e a identificação e delimitação das categorias conceituais dos domínios de especialidade. Em (iii) aparece listado o *Roget's International Thesaurus* (2010), última edição revista e atualizada do *Roget's Thesaurus* original. Essa última edição, de responsabilidade da linguista e lexicógrafa Barbara Ann Kipfer, resolve o problema de falta de estruturação interna dos grupos de palavras da versão original. Nele, os grupos de palavras são organizados internamente por categorias conceituais, separadas em conjuntos de ULS e identificadas por um numeral. Dessa forma, ela é utilizada como referência, em língua inglesa, para a estruturação conceitual dos domínios. A WN.Pr, listada em (iv), também é utilizada como referência para a identificação de conceitos equivalentes codificados em língua inglesa. Já a FrameNet de Berkeley, também listada em (iv), é empregada como referência para a estruturação conceitual das categorias que se relacionam sintagmaticamente, em forma de *frames*.

i) Dicionários monolíngues da Língua Portuguesa

- a) HOUAISS, A. e VILLAR, M. S. – *Dicionário Houaiss da língua portuguesa*. [HOU-S]
- b) *Michaelis Português – Moderno Dicionário da Língua Portuguesa* – V. 1.0; [MIC-S]
- c) *Novo Dicionário Aurélio Eletrônico – Século XXI* – V. 3.0; [AUR-S]
- d) *Caldas Aulete Digital* – V. 1.0; [AUL-S]

ii) Dicionários de especialidades

- e) SBS - *Dicionário Visual do Corpo Humano*; [SBS]
- f) VASCONCELLOS, L. P. – *Dicionário de Teatro* [VAS]

- g) PAVIS, P. – *Dicionário de Teatro*; [PAV]
- h) CAMPOS, G. – *Glossário de Termos Técnicos do Espetáculo*; [CAM]
- i) RAE, K. e SOUTHERN, R. – *Lexique International de Termes Techniques de Théâtre*; [LIT]

iii) Dicionário Analógicos e Thesaurus

- j) ROGET, P. M. – *Roget's International Thesaurus* [RIT]

iv) Recursos léxico-computacionais

- k) WordNet Search 3.0 (<http://wordnetweb.princeton.edu>)
- l) The Berkeley FrameNet Project (<http://framenet.icsi.berkeley.edu>)

ANEXO I - CANÇÕES DE CHICO BUARQUE E EDU LOBO PARA O BALÉ *GRANDE CIRCO MÍSTICO* (1982).

BEATRIZ

Olha
 Será que ela é moça
 Será que ela é triste
 Será que é o contrário
 Será que é pintura
 O rosto da atriz
 Se ela dança no sétimo céu
 Se ela acredita que é outro país
 E se ela só decora o seu papel
 E se pudesse entrar na sua vida
 Olha
 Será que é de louça
 Será que é de éter
 Será que é loucura
 Será que é cenário
 A casa da atriz
 Se ela mora num arranha-céu
 E se as paredes são feitas de giz
 E se ela chora num quarto de hotel
 E se eu pudesse entrar na sua vida
 Sim, me leva para sempre, Beatriz
 Me ensina a não andar com os pés no chão
 Para sempre é sempre um triz
 Aí, diz quantos desastres tem na minha mão
 Diz se é perigoso a gente ser feliz
 Olha
 Será que é uma estrela
 Será que é mentira
 Será que é comédia
 Será que é divina
 A vida da atriz
 Se ela um dia despencar do céu
 E se os pagantes exigirem bis
 E se um arcanjo passar o chapéu
 E se eu pudesse entrar na sua vida.

A BELA E A FERA

Ouve a declaração, oh bela
 De um sonhador titã
 Um que dá nó em paralela
 E almoça rolimã
 O homem mais forte do planeta
 Tórax de Superman
 Tórax de Superman
 E coração de poeta

Não brilharia a estrela, oh bela
 Sem noite por detrás
 Tua beleza de gazela
 Sob o meu corpo é mais

Uma centelha num graveto
 Queima canaviais
 Queima canaviais
 Quase que eu fiz um soneto

Mais que na lua ou no cometa
 Ou na constelação
 O sangue impresso na gazeta
 Tem mais inspiração
 No bucho do analfabeto
 Letras de macarrão
 Letras de macarrão
 Fazem poema concreto

Oh bela, gera a primavera
 Aciona o teu condão
 Oh bela, faz da besta fera
 Um príncipe cristão
 Recebe o teu poeta, oh bela
 Abre teu coração
 Abre teu coração
 Ou eu arrombo a janela

NA CARREIRA

Pintar, vestir
 Virar uma aguardente
 Para a próxima função
 Rezar, cuspir
 Surgir repentinamente
 Na frente do telão
 Mais um dia, mais uma cidade
 Pra se apaixonar
 Querer casar
 Pedir a mão

Saltar, sair
 Partir pé ante pé
 Antes do povo despertar
 Pular, zunir
 Como um furtivo amante
 Antes do dia clarear
 Apagar as pistas de que um dia
 Ali já foi feliz
 Criar raiz
 E se arrancar

Hora de ir embora
 Quando o corpo quer ficar
 Toda alma de artista quer partir
 Arte de deixar algum lugar
 Quando não se tem pra onde ir

Chegar, sorrir
 Mentir feito um mascate
 Quando desce na estação
 Parar, ouvir

Sentir que tatibitati
 Que bate o coração
 Mais um dia, mais uma cidade
 Para enlouquecer
 O bem-querer
 O turbilhão

Bocas, quantas bocas
 A cidade vai abrir
 Pruma alma de artista se entregar
 Palmas pro artista confundir
 Pernas pro artista tropeçar

Voar, fugir
 Como o rei dos ciganos
 Quando junta os cobres seus
 Chorar, ganir
 Como o mais pobre dos pobres
 Dos pobres dos plebeus
 Ir deixando a pele em cada palco
 E não olhar pra trás
 E nem jamais
 Jamais dizer
 Adeus

CIRANDA DA BAILARINA

Procurando bem
 Todo mundo tem pereba
 Marca de bexiga ou vacina
 E tem piriri, tem lombriga, tem ameba
 Só a bailarina que não tem
 E não tem coceira
 Berruga nem frieira
 Nem falta de maneira
 Ela não tem

Futucando bem
 Todo mundo tem piolho
 Ou tem cheiro de creolina
 Todo mundo tem um irmão meio zarolho
 Só a bailarina que não tem
 Nem unha encardida
 Nem dente com comida
 Nem casca de ferida
 Ela não tem

Não livra ninguém
 Todo mundo tem remela
 Quando acorda às seis da matina
 Teve escarlatina
 Ou tem febre amarela
 Só a bailarina que não tem
 Medo de subir, gente
 Medo de cair, gente
 Medo de vertigem

Quem não tem

Confessando bem
 Todo mundo faz pecado
 Logo assim que a missa termina
 Todo mundo tem um primeiro namorado
 Só a bailarina que não tem
 Sujo atrás da orelha
 Bigode de groselha
 Calcinha um pouco velha
 Ela não tem

O padre também
 Pode até ficar vermelho
 Se o vento levanta a batina
 Reparando bem, todo mundo tem pentelho
 Só a bailarina que não tem
 Sala sem mobília
 Goteira na vasilha
 Problema na família
 Quem não tem

Procurando bem
 Todo mundo tem...

O CIRCO MÍSTICO

Não
 Não sei se é um truque banal
 Se um invisível cordão
 Sustenta a vida real

Cordas de uma orquestra
 Sombras de um artista
 Palcos de um planeta
 E as dançarinas no grande final

Chove tanta flor
 Que, sem refletir
 Um ardoroso espectador
 Vira colibri

Qual
 Não sei se é nova ilusão
 Se após o salto mortal
 Existe outra encarnação

Membros de um elenco
 Malas de um destino
 Partes de uma orquestra
 Duas meninas no imenso vagão

Negro refletor
 Flores de organdi
 E o grito do homem voador
 Ao cair em si

Não sei se é vida real

Um invisível cordão
Após o salto mortal

A HISTÓRIA DE LILY BRAUN

Como num romance
O homem dos meus sonhos
Me apareceu no dancing
Era mais um
Só que num relance
Os seus olhos me chuparam
Feito um zoom

Ele me comia
Com aqueles olhos
De comer fotografia
Eu disse cheese
E de close em close
Fui perdendo a pose
E até sorri, feliz

E voltou
Me ofereceu um drinque
Me chamou de anjo azul
Minha visão
Foi desde então ficando flou

Como no cinema
Me mandava às vezes
Uma rosa e um poema
Foco de luz
Eu, feito uma gema
Me desmilinguindo toda
Ao som do blues

Abusou do scotch
Disse que meu corpo
Era só dele aquela noite
Eu disse please
Xale no decote
Disparei com as faces
Rubras e febris

E voltou
No derradeiro show
Com dez poemas e um buquê
Eu disse adeus
Já vou com os meus
Numa turnê

Como amar esposa
Disse ele que agora
Só me amava como esposa
Não como star
Me amassou as rosas
Me queimou as fotos

Me beijou no altar

Nunca mais romance
 Nunca mais cinema
 Nunca mais drinque no dancing
 Nunca mais cheese
 Nunca uma espelunca
 Uma rosa nunca
 Nunca mais feliz

OPERETA DO CASAMENTO

Nem assaz alhures e antanho
 Era um evento tamanho
 A sagração nupcial
 Vinha a noiva de gargantilha
 Caçoleta e rendilha
 Diadema e torçal

Mas se houvesse algum embaraço
 Dera a moça um mau passo
 Quanto horror e desdém
 Ela ia parar no convento
 Ia dormir ao relento
 Ou deitar nos trilhos do trem

Do pudor da noiva a bandeira
 Após a noite primeira
 Desfraldava-se ao sol
 A sua virtude escarlate
 Igual brasão de tomate
 Enobrecendo o lençol

Mas se não houvesse tal mancha
 É que outra mancha mais ancha
 Se ocultava por trás
 E o rapaz pagava o malogro
 Com a vendeta do sogro
 Ou com a malícia dos mortais

"Oh meu pai, oh meu pai, por favor
 Condenai o nosso amor
 De langor e luxúria!
 Mas poupai, oh meu pai
 Nosso filho
 Da fúria do Senhor!"

O guri nasceu apressado
 Nem um mês de casado
 Tinha quem o gerou
 Quando o pai caiu nos infernos
 Foi nos braços maternos
 Que ele se pendurou

Quando a mãe caiu na sarjeta
 Foi seguindo a opereta

Na garupa do avô
 Quando o avô caiu do cavalo
 Foi chorar no intervalo
 E mais um ato começou

Palhaço, corista
 Trapézio, dançarina
 Maestro, cortina
 É fé na flauta e pé na pista

VALSA DOS CLOWNS

Em toda canção
 O palhaço é um charlatão
 Esparrama tanta gargalhada
 Da boca pra fora
 Dizem que seu coração pintado
 Toda tarde de domingo chora

Abra o coração
 Do palhaço da canção
 Eis que salta outro farrapo humano
 E morre na coxia
 Dentro do seu coração de pano
 Um palhaço alegre se anuncia

A nova atração
 Tem um jovem coração
 Que apertado por estreito laço
 Amanhece partido
 Dentro dele sai mais um palhaço
 Que é um palhaço com um olhar caído

E esse charlatão
 Vai cantar sua canção
 Que comove toda a arquibancada
 Com tanta agonia
 Dentro dele um coração folgado
 Cantarola uma outra melodia

Em toda canção
 O palhaço é um charlatão
 E esse charlatão
 Vai cantar uma canção

SOBRE TODAS AS COISAS

Pelo amor de Deus
 Não vê que isso é pecado, desprezar quem lhe quer bem
 Não vê que Deus até fica zangado vendo alguém
 Abandonado pelo amor de Deus

Ao Nosso Senhor
 Pergunte se Ele produziu nas trevas o esplendor

Se tudo foi criado - o macho, a fêmea, o bicho, a flor
Criado pra adorar o Criador

E se o Criador
Inventou a criatura por favor
Se do barro fez alguém com tanto amor
Para amar Nosso Senhor

Não, Nosso Senhor
Não há de ter lançado em movimento terra e céu
Estrelas percorrendo o firmamento em carrossel
Pra circular em torno ao Criador

Ou será que o deus
Que criou nosso desejo é tão cruel
Mostra os vales onde jorra o leite e o mel
E esses vales são de Deus
Pelo amor de Deus
Não vê que isso é pecado, desprezar quem lhe quer bem
Não vê que Deus até fica zangado vendo alguém
Abandonado pelo amor de Deus

MEU NAMORADO

Ele vai me possuindo
Não me possuindo
Num canto qualquer
É como as águas fluindo
Fluindo até o fim
É bem assim que ele me quer
Meu namorado
Meu namorado
Minha morada
É onde for morar você

Ele vai me iluminando
Não iluminando
Um atalho sequer
Sei que ele vai me guiando
Guiando de mansinho
Pro caminho que eu quiser
Meu namorado
Meu namorado
Minha morada é onde for morar você

Vejo meu bem com seus olhos
E é com meus olhos
Que o meu bem me vê

ANEXO II - LISTA DE ARQUIVOS DO CD EM ANEXO

- a) algorithm.txt
- b) anaset_AMOR.csv
- c) anaset_CORPO.csv
- d) anaset_TEATRO.csv
- e) CdSA_sem_AMOR.csv
- f) CdSA_sem_CORPO.csv
- g) CdSA_sem_TEATRO.csv