

MIRNA FERNANDA DE OLIVEIRA

**PRESSUPOSTOS TEÓRICO-METODOLÓGICOS
PARA A ELABORAÇÃO DA BASE LEXICAL
DE UM THESAURUS ELETRÔNICO**

**Araraquara
2002**

MIRNA FERNANDA DE OLIVEIRA

**PRESSUPOSTOS TEÓRICO-METODOLÓGICOS
PARA A ELABORAÇÃO DA BASE LEXICAL
DE UM THESAURUS ELETRÔNICO**

Dissertação apresentada à Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, como parte dos requisitos para a obtenção do título de Mestre em Letras (Linguística e Língua Portuguesa).

Orientador: Prof. Dr. *Bento Carlos Dias da Silva*

**Araraquara
2002**

COMISSÃO JULGADORA

Orientador

Examinador (a) 1

Examinador (a) 2

*Aos meus pais Geraldo e Elza, e à
minha tia, Iria, meus primeiros
educadores, que com amor tomaram
minhas mãos e me conduziram ao
caminho do conhecimento. A eles,
minha eterna gratidão.*

AGRADECIMENTOS

A realização deste trabalho só foi possível graças à colaboração direta ou indireta de muitas pessoas. Manifesto minha gratidão a todas elas e de forma particular:

a todos os professores do Programa de Pós-Graduação em Lingüística e Língua Portuguesa, pelos valiosos ensinamentos;

aos funcionários da seção de Pós-Graduação em Lingüística e Língua Portuguesa, em especial à Rita e Gertz, pelo carinho e auxílio;

à CAPES, pelo apoio financeiro em forma de bolsa de estudos e ao MCT/FINEP, por apoiar projetos interdisciplinares, sem os quais esta dissertação não poderia ter sido desenvolvida;

aos funcionários da Biblioteca e do Pólo Computacional da Faculdade de Ciências e Letras, por sua colaboração;

aos colegas e professores do Núcleo Interinstitucional de Lingüística Computacional (NILC-São Carlos) pelo incentivo à pesquisa em Processamento Automático das Línguas Naturais em nosso país;

aos colegas que trabalharam no desenvolvimento do projeto do *thesaurus*, Helio, Daniela Amorim, Christie, pela amizade, companheirismo e troca de conhecimentos;

à Profª. Dra. Maria Helena Galvão Frem Dias da Silva, por seu carinho e incentivo;

ao Prof. Dr. Bento Carlos Dias da Silva, pela amizade, orientação, paciência e esclarecimento na partilha de conhecimentos e pelo privilégio em fazer parte de sua equipe;

a toda a minha família, pelo apoio, e em especial aos meus sobrinhos, fontes de alegrias eternas, sementes de futuros brilhantes;

a todos os meus amigos, sem exceção, e em especial à Soraya, Ludimila, Naila, Patricia, Daniela Munaretti, Iramaia e Helio, pelo carinho, alegria e amizade de todas as horas;

a Deus, por me permitir a dar mais um passo em minha vida.

A todos, meus profundos agradecimentos.

O modo científico de pensar é ao mesmo tempo imaginativo e disciplinado. Isso é fundamental para o seu sucesso. A ciência nos convida a acolher os fatos, mesmo quando eles não se ajustam às nossas concepções. Aconselha-nos a guardar hipóteses alternativas em nossas mentes, para ver qual se adapta melhor à realidade. Impõe-nos um equilíbrio delicado entre uma abertura sem barreiras para idéias novas, por mais heréticas que sejam, e o exame cético mais rigoroso de tudo – das novas idéias e do conhecimento estabelecido. Esse tipo de pensamento é também uma ferramenta essencial para a democracia numa era de mudanças.

Carl Sagan

RESUMO

Esta dissertação investiga os pressupostos teórico-metodológicos para a elaboração de uma base lexical para a compilação de um thesaurus eletrônico – um dicionário de sinônimos e antônimos armazenado na memória de um computador e acoplado a um processador de textos. Dada a inserção do trabalho no âmbito do Processamento Automático das Línguas Naturais (PLN), desse campo extraímos a metodologia básica de investigação, fundamentada em três fases distintas: a) fase lingüística, em que foram abordados o estudo do léxico, enquanto componente da gramática, enquanto constituinte do sistema cognitivo e enquanto módulo de um sistema de PLN, e o estudo das relações de sentido; b) fase das representações formais, em que foram formalizadas as relações de sinonímia e antonímia; c) fase de implementação, em que aplicou-se a metodologia desenvolvida ao longo do trabalho à compilação da base lexical do thesaurus eletrônico. Essa metodologia revelou-se consistente e eficiente, ao ser testada na montagem da base lexical de um thesaurus eletrônico para o Português do Brasil, contendo aproximadamente quarenta e quatro mil entradas organizadas em função das relações de sinonímia e antonímia.

Palavras-chave: semântica lexical; relações de sentido; léxico; thesaurus eletrônico; processamento automático das línguas naturais.

ABSTRACT

This thesis discusses the theoretic and methodological issues for the construction of a lexical database for the compilation of a thesaurus – a dictionary of synonyms and antonyms stored in a computer memory for use in word processing. Since this work belongs to the field of Natural Language Processing (NLP), from that field we extracted the basic methodology of investigation, divided into three different phases: a) linguistic phase, in which we studied both the lexicon - as part of the grammar of a language, as part of our cognitive system and as part of an NLP system -, and the sense relations within the lexicon; b) representational phase, in which we devised a formal representation for synonymy and antonymy relations; c) implementational phase, in which we applied the methodological framework we developed in the two preceding phases to the compilation of the thesaurus lexical database. The methodological principles proved to be efficient, for they were applied to the construction of an actual lexical database of a Brazilian Portuguese thesaurus, a lexical database currently containing nearly 44 thousand entries organized in terms of the synonymy and antonymy relations.

Keywords: lexical semantics; sense relations; lexicon; thesaurus; natural language processing.

SUMÁRIO

CAPÍTULO 1 - CONSTRUÇÃO DE UM <i>THESAURUS</i> ELETRÔNICO: EMPREENDIMENTO LINGÜÍSTICO E COMPUTACIONAL.....	1
1.1 INTRODUÇÃO.....	1
1.1.1 <i>Delimitação do objeto thesaurus eletrônico</i>	2
1.2 O PROCESSAMENTO AUTOMÁTICO DAS LÍNGUAS NATURAIS: ASPECTOS GERAIS	5
1.2.1 <i>Um breve panorama</i>	5
1.2.2 <i>A abordagem do PLN em três domínios</i>	8
1.2.3 <i>A base lexical e o sistema de PLN e o thesaurus eletrônico</i>	11
1.3 ESTRUTURA DA DISSERTAÇÃO	12
CAPÍTULO 2 - AS UNIDADES LEXICAIS E O LÉXICO	15
2.1 DUAS SEMÂNTICAS LEXICAIS	15
2.2 TRÊS ASPECTOS DO LÉXICO	16
2.2.1 <i>O léxico lingüístico</i>	17
2.2.1.1 A unidade lexical.....	17
2.2.1.2 O significado, sua composicionalidade e seus tipos	19
2.2.2 <i>O léxico mental</i>	25
2.2.3 <i>O léxico computacional</i>	34
2.3 PROPRIEDADES PARTILHADAS ENTRE UNIDADES LINGÜÍSTICAS E COMPUTACIONAIS	36
2.3.1 <i>Propriedades básicas</i>	36
2.3.1.1 Denotação ou categorização	36
2.3.1.2 Superordenação/subordinação	36
2.3.1.3 Atribuição	36
2.3.1.4 Demandas sobre os preenchedores de papéis de caso.....	37
2.3.1.5 Estruturas semânticas ou conceituais	37
2.3.2 <i>Propriedades derivadas ou de segunda ordem</i>	38
2.3.2.1 Implicação.....	38
2.3.2.2 Herança	38
2.3.3 <i>Propriedades psicológicas primárias</i>	38
2.3.3.1 Similaridade	38
2.3.3.2 Tipicidade ou representatividade.....	38
2.3.3.3 Hierarquias de nível básico	38
2.4 TEORIAS PARA A FORMALIZAÇÃO DE SIGNIFICADOS	39
2.4.1 <i>Traços semânticos</i>	39
2.4.2 <i>Redes Semânticas</i>	40
2.4.3 <i>Teoria dos protótipos</i>	42
CAPÍTULO 3 - ARQUITETURA DE LÉXICOS.....	45
3.1 <i>O modelo de Pustejovsky</i>	45
3.2 <i>O modelo de Jackendoff</i>	54
3.3 <i>Relacionando os dois pontos de vista: lexical e conceitual</i>	57
CAPÍTULO 4 - BASES DE DADOS LEXICAIS.....	60
4.1 MÉTODOS DE COMPILAÇÃO DE INFORMAÇÃO LÉXICO-SEMÂNTICA	60
4.2 IMPORTÂNCIA DOS DICIONÁRIO ENQUANTO FONTES DE INFORMAÇÃO LÉXICO-SEMÂNTICA	62
4.2.1 <i>O significado no dicionário: problemas</i>	68
4.3 MOTIVAÇÃO PARA A CONSTRUÇÃO DA BASE LEXICAL DO THESAURUS: A REDE WORDNET.....	75
4.3.1 <i>Fundamentos da rede Wordnet</i>	76
CAPÍTULO 5 - AS RELAÇÕES DE SENTIDO E A ESTRUTURAÇÃO DO LÉXICO	80
5.1 UM INVENTÁRIOS DAS RELAÇÕES DE SENTIDO.....	80
5.1.1 <i>Hierarquias</i>	81
5.1.1.1 Assimetria	81
5.1.1.2 Simetria.....	82
5.1.1.3 Transitividade.....	82
5.1.1.4 Não transitividade.....	82

5.1.2	<i>As relações de sentido fundamentais</i>	82
5.1.2.1	Identidade.....	82
5.1.2.2	Inclusão.....	83
5.1.2.3	Intersecção.....	83
5.1.2.4	Disjunção.....	83
5.1.3	<i>Uma tipologia de relações de sentido</i>	84
5.1.3.1	Sinonímia cognitiva.....	84
5.1.3.2	Hiponímia/hiperonímia.....	84
5.1.3.3	Compatibilidade.....	85
5.1.3.4	Incompatibilidade.....	85
5.1.3.5	Meronímia/holonímia.....	85
5.1.3.6	Troponímia.....	85
5.1.3.7	Relações parciais.....	86
5.1.3.8	Quase-relações.....	86
5.1.3.9	Pseudo-relações.....	86
5.1.3.10	Para-relações.....	87
5.1.3.11	Antonímia.....	88
5.2	AS RELAÇÕES DE SENTIDO DEFINIDORAS DO THESAURUS ELETRÔNICO.....	89
5.2.1	SINONÍMIA.....	89
5.2.1.1	<i>Sinonímia, contexto e o critério da possibilidade de substituição</i>	90
5.2.1.2	<i>Sinonímia, valor de verdade e implicação</i>	92
5.2.1.3	<i>Sinonímia e contexto</i>	93
5.2.1.4	<i>Sinonímia e traços componenciais</i>	93
5.2.1.5	<i>Sinonímia e conceitos</i>	95
5.2.1.6	<i>Sinonímia: uma tipologia</i>	98
5.2.1.6.1	Sinonímia absoluta.....	98
5.2.1.6.2	Sinonímia cognitiva.....	99
5.2.1.6.3	Sinonímia parcial.....	101
5.2.1.6.4	Sinonímia absoluta mas não-total.....	102
5.2.1.6.5	Sinonímia completa mas não-total.....	102
5.2.1.6.6	Sinonímia incompleta e não-total.....	102
5.2.1.6.7	Sinonímia conotativa.....	102
5.2.1.6.8	Sinonímia denotativa.....	102
5.2.1.6.9	Outros casos.....	102
5.2.2	ANTONÍMIA.....	103
5.2.2.1	<i>Contrastes binários ou dicotômicos</i>	105
5.2.2.1.1	Antonímia.....	105
5.2.2.1.2	Complementaridade.....	108
5.2.2.1.3	Reciprocidade.....	110
5.2.2.1.4	Oposição direcional.....	110
5.2.2.2	<i>Contrastes não-binários</i>	113
5.2.2.2.1	Conjuntos seriais.....	113
5.2.2.2.2	Conjuntos cíclicos.....	113
5.2.2.3	<i>Antônimos múltiplos ou lexemas com mais de um antônimo</i>	114
5.2.2.4	<i>Antonímia e conceitos</i>	114
5.3	AS RELAÇÕES DE SENTIDO E AS CLASSES DE PALAVRAS.....	115
5.3.1	<i>A organização semântica das classes de palavras consideradas</i>	116
5.3.1.1	Substantivos.....	116
5.3.1.2	Adjetivos e advérbios.....	119
5.3.1.3	Verbos.....	122
CAPÍTULO 6 - A BASE LEXICAL DO THESAURUS ELETRÔNICO.....		124
6.1	<i>Sinonímia e antonímia: estrutura de sua representação</i>	124
6.2	<i>Proposta de modelo para a interface de inserção de dados</i>	125
6.3	<i>Delimitação do corpus de referência</i>	129
6.4	<i>Parâmetros para filtragem de informações do corpus de referência</i>	133
6.5	<i>O algoritmo</i>	137
6.6	<i>O Editor do Thesaurus</i>	140
6.7	<i>O Editor e o Assistente de Edição: aspectos visuais e linhas gerais de funcionamento</i>	142
6.8	<i>A inserção de dados na interface: o procedimento do lingüista</i>	146
CAPÍTULO 7 - CONCLUSÕES, RESULTADOS E PERSPECTIVAS.....		150
REFERÊNCIAS BIBLIOGRÁFICAS.....		156

ANEXO 01	166
<i>O texto utilizado como ponto de partida para a análise de dois aplicativos do tipo “Dicionários de sinônimos” atualmente em uso no Brasil</i>	<i>166</i>
ANEXO 02	168
<i>Levantamento de sentidos e significados extraído dos aplicativos sob análise a partir das unidades textuais do anexo 01</i>	<i>168</i>
ANEXO 03	185
<i>As deficiências dos thesauri existentes</i>	<i>186</i>

LISTA DE FIGURAS

Fig. 1. A base lexical, ou base de dados lexicais, como um dos componentes essenciais de um sistema de PLN.	11
Fig. 2. Entradas lexicais fracionadas	27
Fig. 3. Modelo de representação do processamento lexical.	28
Fig. 4. Esquema sintético de processamento lexical.	28
Fig. 5. Tipos de entradas do léxico mental.	33
Fig. 6. Exemplo de uma rede semântica simples.	41
Fig. 7. Fragmento da base do LDOCE.	61
Fig. 8. A matriz lexical	79
Fig. 9. Classificação dos tipos de contraste.	105
Fig. 10. Comportamento dos pares de antônimos <i>devagar:rápido</i>	107
Fig. 11. Esquema de estruturação para a implementação do editor do <i>thesaurus</i>	126
Fig. 12. A interface de inserção de dados no <i>thesaurus</i>	143
Fig. 13. Em destaque, o Assistente de Edição da interface.	144
Fig. 14. Exemplo de armazenamento da entrada <i>abreviado</i> na base do <i>thesaurus</i>	145
Fig. 15. – A interface do dicionário de sinônimos v.2000, mostrando <i>viúvo</i> como sinônimo de <i>só</i> . ..	193
Fig. 16a e 16b – A interface de consulta do verbete <i>só</i> de ambas as versões dos aplicativos avaliados (v.97 e v.2000)	194

LISTA DE TABELAS

Tabela 1. Os seis tipos de objetos “ <i>thesaurus</i> ”.....	03
Tabela 2. Os sete tipos de significado.....	22
Tabela 3. Breve tipologia da sinonímia	134
Tabela 4. Breve tipologia das relações de contraste	134
Tabela 5. A base lexical do <i>thesaurus</i> , em números.....	152

CAPÍTULO 1 - CONSTRUÇÃO DE UM *THESAURUS* ELETRÔNICO: empreendimento lingüístico e computacional

1.1 Introdução

O grande distanciamento que separa os estudos da linguagem *per se* e o campo genericamente denominado processamento automático das línguas naturais (doravante PLN) é apontado por Dias-da-Silva (1996) como uma das causas da ineficiência do desempenho e da superficialidade no tratamento lingüístico de uma série de aplicativos - corretores ortográficos e gramaticais, tradutores automáticos (disponíveis na *internet*), dicionários eletrônicos, entre outros -, que a indústria de informática coloca, a cada ano, à disposição dos usuários de microcomputadores.

Sem perder de vista os procedimentos analíticos e descritivos da lingüística e utilizando-se também dos procedimentos formais e algorítmicos desenvolvidos no âmbito das ciências da computação, este trabalho visa à proposição de um arcabouço teórico-metodológico para o desenvolvimento de um tipo específico de aplicativo: um *thesaurus eletrônico*.

No desenvolvimento desse tipo de aplicativo, questões lingüísticas e computacionais devem ser necessariamente abordadas de modo solidário. Em outras palavras, o trabalho a ser desenvolvido pretende sistematizar um corpo específico de conhecimentos lingüísticos e computacionais necessários para a compilação de uma base de dados lexicais a partir da qual um *thesaurus eletrônico* possa ser automaticamente gerado.

Assim, este trabalho, ao se inserir no âmbito do PLN, campo interdisciplinar por excelência, na tarefa de discutir uma proposta de organização formal para a estrutura léxico-semântica subjacente a um *thesaurus eletrônico*, não só pretende apresentar um

modo de compilação de uma base de dados lexicais específica para esse tipo de aplicativo, como também visa contribuir para a aproximação entre os estudos lingüísticos e os estudos do PLN.

Como veremos mais adiante ainda neste capítulo, adotando-se a metodologia de desenvolvimento de projetos no campo do PLN proposta por Dias-da-Silva (1998), é possível fatorar o problema de construção de um *thesaurus eletrônico* em três domínios: lingüístico, das representações formais e da implementação computacional dessas representações.

1.1.1 Delimitação do objeto *thesaurus eletrônico*

A questão terminológica não pode ser descuidada, posto que o termo *thesaurus* é empregado por diferentes especialistas, para designar objetos bastante distintos. Não vamos fazer uma análise etimológica, mas simplesmente ler, nas entrelinhas, o que dizem algumas das principais obras de referência, sobretudo as de língua inglesa, uma vez que esse termo, na acepção pertinente para este trabalho, parece ter sido introduzido no português via informática e, conseqüentemente, via língua inglesa. Indiretamente, esse breve estudo ilustra também os problemas adicionais que os dicionários, no confronto de informações neles apresentadas, colocam para o consulente e para os pesquisadores.

A partir do levantamento realizado por Dias-da-Silva et al. (2000b, p. 03-06), na Tabela 1, distinguem-se seis tipos distintos de objetos denominados *thesaurus*, classificados de acordo com sua natureza.

Tabela 1. Os seis tipos de objeto “thesaurus”.

Objeto	Natureza	Definições
THESAURUS 1	Um tipo de inventário exaustivo que procura registrar o vocabulário de uma determinada língua, um “tesouro” vocabular.	<ul style="list-style-type: none"> • Thesaurus/ tesouro s.m. (lat.) Nome dado freqüentemente como título a léxicos de filologia e arqueologia, principalmente a dicionários exaustivos, compreendendo o vocabulário completo de uma língua. (<i>Grande Enciclopédia Larousse Cultural</i>, Nova Cultural, 1998) • tesouro/thesaurus/tesoiro [Do gr. <i>Thesaurós</i>, pelo lat. <i>thesauru.</i>] S. m. Coleção de palavras e/ou peculiaridades de uma língua, ou de determinado ramo do conhecimento, etc. (<i>Dicionário Aurélio Eletrônico – Século XXI – Versão 3.0</i>, Lexikon Informática Ltda., 1999)
THESAURUS 2	Um “dicionário organizado em função de conceitos lexicalizados” (Crystal, 1997, p. 158), ou seja, um dicionário onomasiológico, cujo precursor foi Roget ¹ .	<ul style="list-style-type: none"> • Tesouro [Do gr. <i>thesaurós</i>, pelo lat. <i>thesauru.</i>]/thesaurus [Do lat., pelo ingl., <i>thesauru.</i>] Dicionário analógico. (<i>Dicionário Aurélio Eletrônico – Século XXI – Versão 3.0</i>, Lexikon Informática Ltda., 1999) • tesouro/thesaurus sm (lat <i>thesauru</i>) Coleção de palavras agrupadas por conceitos e títulos, e não em ordem alfabética como num dicionário. <i>Exs: tesouro de sociologia, tesouro de economia política.</i> (<i>Michaelis Português – Moderno Dicionário da Língua Portuguesa - Versão 1.0</i>, DTS Software Brasil Ltda, 1998)
THESAURUS 3	Dicionário ancorado, de modo geral, nas relações de sinonímia e antonímia na organização dos lexemas	<ul style="list-style-type: none"> • thesaurus [1730–40; < L <i>thesaurus</i> < Gk <i>thesaurós</i> <i>treasure, treasury</i>] A dictionary of synonyms and antonyms. (<i>Random House Webster’s Unabridged Electronic Dictionary – Version 2.0</i>, Random House Inc, 1994) • thesaurus A book containing a store of words; specif., a book of synonyms and antonyms. (<i>Webster’s New World Dictionary & Thesaurus – Version 1.0</i>, Macmillan Publishers, 1997)
THESAURUS 4	Dicionário empregado no domínio da Informática e Documentação.	<ul style="list-style-type: none"> • Thesaurus [1730–40; < L <i>thesaurus</i> < Gk <i>thesaurós</i> <i>treasure, treasury</i>] <i>Computers</i> An index to information stored in a computer, consisting of a comprehensive list of subjects concerning which information may be retrieved by using the proper key terms. (<i>Random House Webster’s Unabridged Electronic Dictionary – Version 2.0</i>, Random House Inc, 1994) • Thesaurus <i>Computers</i> A categorized index of terms for use in information retrieval, as from a computer. (<i>Webster’s New World Dictionary & Thesaurus – Version 1.0</i>, Macmillan Publishers, 1997) • Tesouro/tesoiro sm (lat <i>thesauru</i>) (Não utiliza rubrica) Lista de termos autorizados para indexação e recuperação da informação, em um determinado sistema de documentação. (<i>Michaelis Português – Moderno Dicionário da Língua Portuguesa - Versão 1.0</i>, DTS Software Brasil Ltda, 1998) • Tesouro/thesaurus [Do lat., pelo ingl., <i>thesauru.</i>] S. m. <i>Docum.</i> Vocabulário controlado e dinâmico de descritores [palavra ou expressão utilizada em indexação e tesouro para representar, sem ambigüidade, um determinado conceito] relacionados semântica e genericamente, que cobre de forma extensiva um ramo específico de conhecimento. (<i>Dicionário Aurélio Eletrônico – Século XXI – Versão 3.0</i>, Lexikon Informática Ltda., 1999)
THESAURUS 5	Outro emprego,	<ul style="list-style-type: none"> • Tesouro/thesaurus (lat <i>thesauru</i>) sm <i>Inform</i>

¹ O *Thesaurus* de Roget levou 12 anos para ser concluído e foi publicado em 1852.

	também motivado pelo advento da Informática.	Arquivo contendo sinônimos que são exibidos como alternativas para uma palavra escrita de forma incorreta, durante uma verificação de ortografia. (<i>Michaelis Português – Moderno Dicionário da Língua Portuguesa – Versão 1.0</i> , DTS Software Brasil Ltda, 1998)
THESAURUS 6	Um tipo específico de ferramenta de auxílio à expressão lingüística, parte integrante vários processadores de textos	<ul style="list-style-type: none"> • Thesaurus [1730–40; < L <i>thesaurus</i> < Gk <i>thesaurós</i> <i>treasure, treasury</i>] <i>Computers</i> A dictionary of synonyms and antonyms stored in memory for use in word processing. (<i>Random House Webster's Unabridged Electronic Dictionary – Version 2.0</i> , Random House Inc, 1994)

Note-se que no português do Brasil a forma *thesaurus* convive hoje com os termos *tesouro* e *tesauro*. O primeiro é nitidamente evolução “natural” do termo latino e o segundo, adaptação do termo *thesaurus*, decorrência de crescente intrusão de termos ingleses. Apesar dessa variedade de objetos, um *thesaurus*, caracteriza-se por privilegiar tipos específicos de relação entre expressões lexicais. No sentido mais amplo, é um tipo particular de dicionário em que as entradas são estruturadas segundo um determinado critério como, por exemplo, relações conceptuais, relações léxico-semânticas, campos semântico-nocionais, sistemas terminológicos, entre outros.

Neste trabalho, o que denominamos *thesaurus eletrônico* é o **THESAURUS 6**, assim definido:

Um dicionário de sinônimos e antônimos armazenado na memória do computador e integrado a um processador de textos para consulta.

Como a construção desse objeto pressupõe metodologias desenvolvidas no âmbito do PLN, passamos, a seguir, a examinar o escopo de investigações desse campo com vistas à extração dos princípios norteadores para a compilação da base lexical a partir da qual um *thesaurus eletrônico* possa ser automaticamente gerado.

1.2 O Processamento Automático das Línguas Naturais: aspectos gerais

1.2.1 Um breve panorama

O advento dos computadores nos anos 40 e o crescente desenvolvimento de linguagens formais até os nossos dias têm possibilitado a cientistas de variadas áreas do conhecimento a realização de avanços tecnológicos significativos em diversos tipos de pesquisa.

A execução de tarefas pelo computador exigiu que o homem criasse uma linguagem própria para se comunicar com ele. Diante da impossibilidade da máquina “compreender” as línguas naturais, a saída foi a invenção de linguagens artificiais de programação (Basic, Fortran, Pascal, por exemplo), resolvendo-se em parte a questão da comunicação entre homem-máquina. No entanto, a evolução das linguagens de programação resultou em uma pluralidade de formalismos complexos e incompreensíveis para o usuário comum.

Para facilitar a difícil comunicação homem-máquina é que as pesquisas em PLN encontram seu desafio: desenvolver sistemas computacionais que sejam capazes de processar objetos de natureza lingüística (Dias-da-Silva, 1996, p. 12). Assim, a principal meta do PLN, e a mais ambiciosa, é desenvolver sistemas computacionais que permitam a comunicação entre o computador e o homem por meio de uma língua natural.

Essa idéia, abarcada com muito entusiasmo de início, revelou-se extremamente complexa com o decorrer das pesquisas. Dentre os entraves está a realização de trabalho cooperativo, abrangendo especialistas e domínios do conhecimento diversos, como é o caso, por exemplo, da lingüística e das ciências da computação. Mesmo diante da

absoluta especificidade dos objetos e métodos de estudo dessa duas disciplinas, o PLN pode ser concebido como uma área de investigação interdisciplinar, conforme defende Dias-da-Silva (1996): a busca de soluções para o PLN é uma tarefa que exige reflexões cuidadosas no campo da lingüística, sem perder de vista o trabalho de especialistas de outras áreas como ciência da computação, lógica, etc. Essas reflexões contribuirão para uma melhor acuidade na consecução de sistemas de PLN de qualquer natureza.

Hoje é possível ter acesso a muitos aplicativos computacionais que buscam dar conta de algum aspecto da linguagem humana. Entre os mais conhecidos estão os corretores ortográficos, os dicionários eletrônicos e programas ou “maquininhas” de bolso para realizar tradução automática. Um breve exame desses tipos de aplicativos demonstra sua ineficiência do ponto de vista lingüístico. As exigências comerciais e a demanda por aplicativos dessa natureza em um curto período de tempo acabam por produzir aplicativos que não dão conta da face lingüística do PLN, produzindo, por exemplo, “tradutores automáticos” incapazes de realizar uma tradução minimamente satisfatória.

Dentre as possibilidades de aplicação dos estudos do PLN, Cooper *et al*, 1996 citam:

- os corretores ortográficos e gramaticais – pacote de aplicativos que podem vir acoplados aos programas de processamento de texto (ao *Microsoft Word*, por exemplo); entre suas principais funções, destacam-se: checar ortografia e corrigir possíveis erros, além de solucionar problemas gramaticais como a concordância e a regência, por exemplo.
- os redatores de documentos estruturados – esses programas são uma evolução dos programas anteriores; esse tipo de programa auxilia o usuário a produzir textos de

documentos técnicos, documentos legais, etc., checando a consistência terminológica do documento, e tenta detectar novos termos.

- os programas de extração de informação – são programas que atuam geralmente em contextos livres de ambigüidade; eles extraem informações de relatórios ou textos jornalísticos, por exemplo, e transferem para um formulário específico, deixando lacunas nos lugares em que a informação foi retirada.
- programas de tradução interativa e tradução assistida por computador – são programas que simulam um diálogo entre a máquina e o tradutor humano; como todo procedimento de tradução, a tradução interativa requer decisões a respeito de ambigüidades e escolha apropriada de termos, que estão baseadas em fatores lexicais e/ou estruturais.
- programas de tradução “*offline*” – geralmente usados para realizar a tradução de grande quantidade de textos de modo absolutamente automático, sem a intervenção de tradutor humano, entre pares de línguas (por exemplo, do português para o japonês, e vice-versa).
- programas que constituem uma interface, capaz de “compreender” uma língua natural, entre o usuário e um sistema de computadores que armazenam grandes volumes de informação; são, portanto, programas que realizam buscas em bases de dados (numéricos, gráficos e/ou textuais), entre outros meios de armazenamento e estruturação de informação; esses sistemas são geralmente projetados para manipular a língua falada e sustentar diálogos e referências anafóricas.
- programas de geração de textos – são programas que transformam representações não lingüísticas em textos como, por exemplo, programas que geram relatórios financeiros e programas que geram resumos, entre outros gêneros.

- programas que transformam texto escrito em texto falado, e vice-versa – esses programas transformam grafemas em seqüências fônicas e vice versa.
- programas de manipulação de unidades lexicais – os mais variados tipos de dicionários e léxicos eletrônicos e computacionalmente tratáveis, incluindo aqui o *thesaurus eletrônico*.

1.2.2 A abordagem do PLN em três domínios

Uma das hipóteses de trabalho que exploramos aqui diz respeito à natureza da informação lingüística que deve ser investigada. Segundo essa hipótese, esse tipo de informação depende da arquitetura do sistema de PLN e dos objetivos a serem alcançados pelo sistema. No entanto, Dias-da-Silva (1996, p. 76) chama a atenção para o fato de que cada tipo de abordagem tem seus métodos próprios, que acabam por definir os contornos do objeto. Estudos em outras áreas do conhecimento podem trazer contribuições à lingüística, permitindo que determinado objeto lingüístico seja focalizado com outras lentes. Um exemplo é o estudo das linguagens formais iniciado por Chomsky, com apoio em modelos matemáticos; seu trabalho resultou em um revolucionário modelo formal de análise gramatical. Portanto, a arquitetura da metodologia para o desenvolvimento de um sistema em PLN tem duas faces: se a natureza dos objetivos do sistema computacional guia a busca de soluções no campo da lingüística, a teoria lingüística também possui seus próprios meios de abordar determinados fenômenos lingüísticos. É preciso buscar um equilíbrio entre esses dois caminhos.

Com o objetivo de propor uma metodologia de pesquisa para o desenvolvimento de aplicativos computacionais de PLN e procurar minimizar o problema da distância que separa os estudos lingüísticos e os estudos computacionais, Dias-da-Silva (1996,

1998a e 1998b) delineou um modelo baseado em três grandes domínios: domínio lingüístico, domínio representacional e domínio implementacional.

No domínio lingüístico realizam-se as reflexões que dizem respeito a objetos lingüísticos; é nesta fase que se investigam, por exemplo, questões acerca do que é linguagem, língua, discurso, léxico, gramática, semântica, pragmática, etc., ou seja, os grandes objetos de investigação da lingüística.

Em particular, vamos investigar concepções de léxico, de relações de sentido e de modelos de representação lexical e as técnicas de compilação de informação léxico-semântica.

No domínio **representacional** é necessário buscar meios de se representar formalmente os conhecimentos lingüísticos levantados no domínio anterior. É o momento de se procurar teorias e modelos que dêem conta da face conceitual dos fenômenos lingüísticos necessários para o desenvolvimento de um aplicativo específico; portanto, as representações lingüísticas propostas neste domínio precisam ser formais e computacionalmente tratáveis.

Ainda na fase representacional, Boguraev e Briscoe (1991, p.04) e Dias-da-Silva (1996, p.176) afirmam ser necessário abordar questões que abrangem os seguintes subdomínios: morfossintático – domínio da representação de gramáticas e de analisadores gramaticais, incluindo a representação das regras e das estruturas morfossintáticas e de léxicos enriquecidos com informações pragmático-discursivas; semântico – domínio da representação de estruturas semânticas, de domínios conceituais e estratégias computacionais de interpretação dessas representações; pragmático-discursivo – domínio da representação da estrutura do discurso e dos contextos pragmático-discursivo e situacional.

Briscoe e Boguraev (1991, p.04) afirmam ainda que, dependendo da natureza do aplicativo a ser desenvolvido, até a informação fonológica é relevante, pois dá conta do sistema de sons e da estrutura de palavras e enunciados de uma língua.

Neste trabalho procuraremos abordar diferentes tipos de formalização: formalização das relações léxico-semânticas de sinonímia e antonímia (relações que, como veremos, são constitutivas de um *thesaurus eletrônico*), formalização da estrutura da base de dados lexicais e formalização da estrutura do verbete do *thesaurus*.

Por fim, no domínio implementacional, colocam-se em relevo questões de como implementar todas as tarefas que um determinado sistema em PLN deve realizar, quais os módulos necessários para sua consecução, o papel de cada um desses módulos no sistema final e como essas partes devem ser organizadas para o seu funcionamento, além da especificação do fluxo de informações a ser gerenciado e a caracterização de um ambiente computacional para o seu desenvolvimento (Dias-da-Silva, 1996, p. 226).

Pretendemos, em particular, especificar o *editor do thesaurus*, através do qual as representações discutidas no domínio anterior são "materializadas" em uma base relacional de dados, no sentido computacional do termo, cujo conteúdo são unidades lexicais. Ainda neste domínio, são delimitados o *corpus de referência* (conjunto de dicionários de língua que constituem a fonte da qual deverá ser compilada a informação léxico-semântica), especificados os principais critérios para compilação e filtragem dessa informação e exemplificado o procedimento de montagem do verbete típico

Cabe ainda ressaltar que a proposta de trabalho em três fases busca conciliar reflexões de vários domínios do conhecimento, tanto do ponto de vista lingüístico quanto do ponto de vista computacional; a busca deste diálogo é fundamental para que se chegue a soluções viáveis e consistentes.

1.2.3 A base lexical e o sistema de PLN e o *thesaurus eletrônico*

Não se sabe ao certo quantos e quais são os componentes de um sistema de PLN, mas existem alguns que são imprescindíveis para sua construção. Dias-da-Silva (1996, p. 232) fornece o esquema de uma arquitetura genérica de um sistema de PLN, ilustrada na Figura 1.

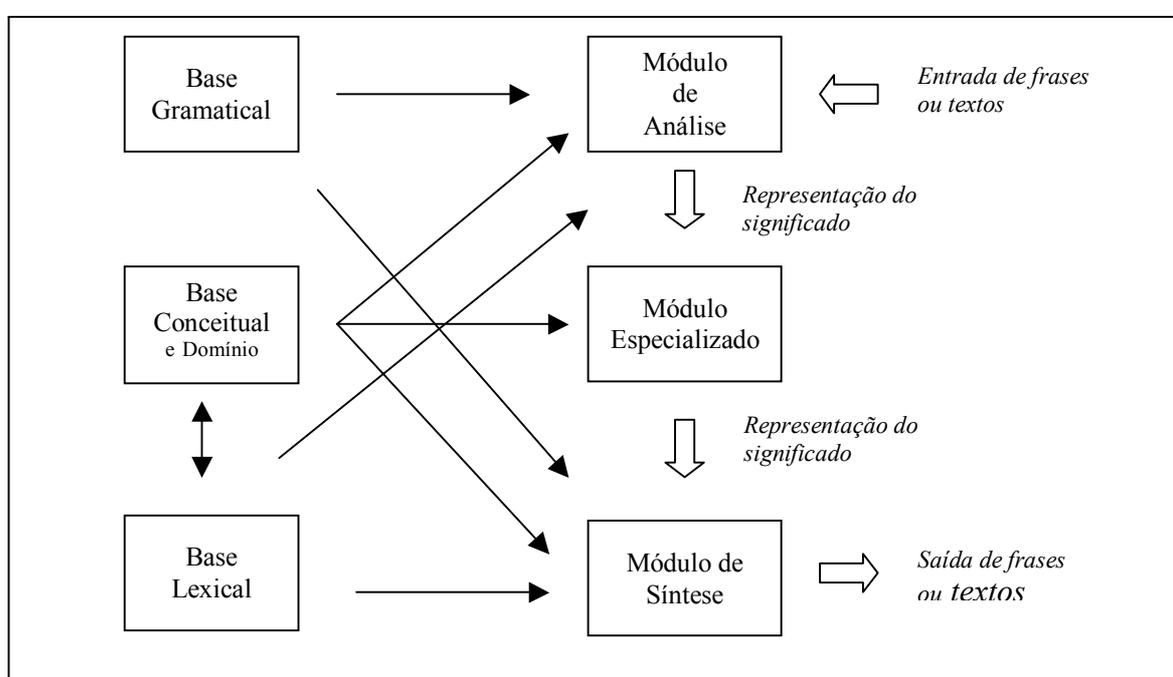


Fig. 1. A base lexical, ou base de dados lexicais, como um dos componentes essenciais de um sistema de PLN.

Dias-da-Silva (1996, p. 231) afirma que, com exceção do **módulo especializado (ME)**, que deve ser projetado em função dos objetivos previstos para o sistema, os demais módulos possuem um funcionamento padrão. Desta forma, para a proposição da metodologia de construção de um *thesaurus eletrônico*, aproveitamos deste esquema duas noções fundamentais: a base lexical e a base conceitual.

A **base lexical** é concebida como uma base de dados de natureza léxico-gramatical, contendo um conjunto de unidades lexicais que, de acordo com as

especificidades do sistema em desenvolvimento, poderá conter informações morfológicas, sintáticas, semânticas e pragmático-discursivas.

A **base conceitual** contém um modelo de mundo, tanto físico quanto conceitual, procurando descrever objetos, eventos, forças, propriedades, relações e atributos em termos de representações hierarquicamente estruturadas (Dias-da-Silva, 1996, p. 234). A base conceitual é importante porque contém o conjunto de categorias semânticas básicas, isto é, primitivos semânticos que simulam a “visão de mundo” do sistema, restringindo o universo discursivo a um determinado domínio.

Nesse contexto é que buscamos a motivação para propor um arcabouço teórico que dê sustentação para a elaboração de um *thesaurus eletrônico* para o português do Brasil; desta forma, utilizando-nos dos princípios da metodologia desenvolvida para a construção de sistemas de PLN, pretendemos extrair as diretrizes que nos levam à investigação de questões concernentes à lingüística, à representação formal de conceitos lingüísticos e à compilação de uma base de dados lexicais.

Definido o objeto (um dicionário eletrônico de sinônimos e antônimos da língua portuguesa) é preciso portanto constituir uma base de dados lexicais específica, computacionalmente estruturada em função de suas duas relações de sentido constitutivas: sinonímia e antonímia. Conseqüentemente, buscamos uma metodologia que norteie a construção dessa base e investigamos problemas em cada uma das fases (lingüística, representacional e implementacional), busca que constitui o corpo desta dissertação.

1.3 Estrutura da dissertação

Neste primeiro capítulo, apresentamos a contextualização do projeto de compilação da base de dados lexicais de um *thesaurus eletrônico*, um panorama do

campo de pesquisa em PLN e os princípios metodológicos que norteiam nossa pesquisa em PLN, a localização lógica da base lexical em um sistema de PLN genérico e a estrutura global desta dissertação.

No segundo capítulo prosseguimos com uma reflexão sobre os tipos de informação lingüística e não-lingüística relevantes para a construção de uma base lexical. Em primeiro lugar, apresentamos três concepções de léxico: componente da gramática (léxico lingüístico), componente do sistema cognitivo (léxico mental) e componente de um sistema de PLN (léxico computacional). Na seqüência, apresentamos as propriedades partilhadas entre unidades lingüísticas e computacionais e introduzimos as teorias para a formalização do significado.

No terceiro capítulo apresentamos reflexões sobre a arquitetura de léxicos, acompanhadas de um levantamento de modelos de representação lexical e conceitual.

No quarto capítulo, sobre as bases de dados lexicais, apresentamos os métodos de extração de informação léxico-semântica para a construção da base lexical do *thesaurus eletrônico*, refletimos sobre a importância dos dicionários como fonte de informação léxico-semântica e delineamos os fundamentos da rede *Wordnet*, cuja metodologia serviu de motivação para o nosso trabalho.

No quinto capítulo discutimos o papel das relações de sentido na estruturação do léxico e apresentamos um refinamento das relações de sinonímia e antonímia, constitutivas do *thesaurus eletrônico*.

No sexto capítulo esboçamos a representação formal da sinonímia e antonímia e partimos para a descrição dos passos referentes à implementação da base, dando destaque para a descrição do “*editor do thesaurus*” e de sua utilização pelo lingüista no processo de compilação da base lexical.

No sétimo capítulo resumimos as principais conclusões do trabalho, apresentamos a efetiva aplicação de parte da metodologia desenvolvida e apontamos para refinamentos necessários e trabalhos futuros.

Nos anexos, mostramos uma breve análise de duas versões do *thesaurus eletrônico* do português acoplado ao *Microsoft Word*: Anexo 1- o texto de onde extraímos as unidades lexicais para o teste; Anexo 2- quadro que apresenta um levantamento dos sentidos e significados das unidades textuais pesquisadas nos aplicativos sob análise.; Anexo 3- uma breve análise dos dados. Os resultados corroboram nossa iniciativa de propor a compilação de um *thesaurus eletrônico* linguisticamente motivado, com informações mais precisas e atualizadas.

CAPÍTULO 2 - AS UNIDADES LEXICAIS E O LÉXICO

2.1 Duas semânticas lexicais

A compilação de uma base de dados lexicais, ou simplesmente base lexical, componente de um sistema de PLN, demanda, no domínio lingüístico, o estudo do léxico. O ponto de partida está em duas disciplinas: a Semântica Lexical, oriunda das investigações no campo da Lingüística, e a Semântica Lexical Computacional, oriunda da Lingüística Computacional, ambas contribuindo para o desenvolvimento da disciplina do Processamento Automático de Línguas Naturais. A semântica computacional, embasada na vertente lógica dos estudos da linguagem, é fortemente motivada pela semântica formal de Montague e outros lógicos, lingüistas e filósofos da linguagem na década de 60 (Cann, 1993). A preocupação com a estruturação do significado levou esses estudiosos a investigar os mecanismos determinantes do significado tanto textual quanto sentencial. Chegaram à conclusão de que para obter resultados consistentes era preciso tratar os significados individuais das unidades lexicais com o rigor e precisão característicos dos métodos da lógica e da teoria matemática dos modelos, estabelecendo uma solidariedade entre as duas modalidades de investigação da **semântica lexical**.

Dentro do escopo dessas duas disciplinas, semântica lexical "lingüística" e semântica lexical "computacional", é central a questão da representação semântica; nesse sentido, os trabalhos de investigação essencialmente envolvem as seguintes questões (Cooper *et al*, 1995, p. 61):

- Como codificar a informação léxico-semântica com o menor grau de redundância (ambigüidade, subespecificações, estrutura *qualia*, etc.);

- Como relacionar a informação léxico-semântica tanto sintagmaticamente como paradigmaticamente (campos semânticos, “aglomerados semânticos”, colocações, etc.);
- Como explicitar o significado das palavras (decomposição, busca de primitivos semânticos, estruturas conceituais, etc.);
- Como relacionar argumentos sintáticos e argumentos semânticos (teorias de mapeamento lexical, teorias de ligação, estruturas conceituais, papéis temáticos, funções gramaticais, etc.);
- Como dar conta de inferências ativadas pelo léxico, (os postulados de significado, etc.);
- Como especificar formalismos e ferramentas que contribuam para a compreensão da estrutura do léxico (regras de redundância lexical, hierarquias de herança múltipla, especificações taxonômicas, valores padrão (*default values*), monotonicidade, etc.);
- Como compilar léxicos computacionais (e a informação semântica neles contidas) a partir de léxicos interpretáveis por computador (MRD’s – *machine readable dictionaries*), construção de léxicos em geral, trabalho de *corpus*, etc.

2.2 Três aspectos do léxico

Construir uma base de dados lexicais para fins de PLN pressupõe conhecer a natureza e características do Léxico de uma língua natural; assim sendo, neste segundo capítulo apresentaremos reflexões sobre aspectos do Léxico do ponto de vista das duas disciplinas matrizes acima mencionadas e do ponto de vista da Ciência Cognitiva. Cada

um desses domínios possibilita focalizarmos o léxico enquanto componente de três sistemas distintos: lingüístico, mental e computacional.

Neste ponto cabe lembrar que toda língua contém um vocabulário, ou léxico, que é complementar à gramática; esse vocabulário não apenas lista os lexemas da língua, mas associa a cada lexema toda informação que é requerida pelas regras da gramática. Ao conjunto de lexemas de uma dada língua chamamos de léxico; o léxico de uma língua pode estar armazenado tanto num léxico mental quanto sob a forma de dicionários (Handke, 1995, p. 49).

2.2.1 O léxico lingüístico

2.2.1.1 A unidade lexical

A língua pode ser concebida como um sistema de comunicação que emprega símbolos arbitrários chamados palavras. No entanto, devido à dificuldade de conceituação do termo, usamos o termo **item lexical** (ou "unidade lexical"), unidade abstrata que nos permite, por um lado, levantar um paradigma de formas como *correr*, *corre*, *correndo*, *correu*, etc. como realizações de um mesmo item lexical CORRER e, por outro, caracterizar a forma *fiou* em *Vovó fiou um novelo de lã* (= tecer) e *João não se fiou no novo empregado* (=depositar confiança em) como itens lexicais separados, FIAR₁ e FIAR₂, cada um encabeçando seu próprio verbete de dicionário. Unidades lexicais são complexos de forma-significado com propriedades semânticas discretas relativamente estáveis, que contraem relações de sentido entre si como, por exemplo, a antonímia (ex. *longo:curto*) e a hiponímia (ex: *cão:animal*), relações de que trataremos mais adiante.

Intercambiável com o termo unidade lexical encontramos também o termo **lexema**. Os lexemas podem ser considerados como unidades listadas no léxico, entendido como

“dicionário ideal” da língua (Cruse, 1995, p. 49); no campo da lexicografia, o lexema é considerado uma unidade de conteúdo léxico expresso no sistema lingüístico, como *hora*, *minuto*, *segundo*, etc, ou seja, lexemas são unidades léxicas pertencentes ao mesmo paradigma semântico ou unidades relacionadas sintagmaticamente por solidariedades do tipo afinidade, seleção e implicação (Bechara, 1999, p.387-8). É importante notar que no domínio da psicolingüística (especificamente no estudo do léxico mental) o termo lexema é empregado com outro sentido: aparece vinculado ao sistema articulador de produção de linguagem e é usado para denotar o elemento morfofonético responsável pela realização fonética de conceitos: é o “som” e a “forma” que se relaciona com determinado conceito (Bierwisch e Schreuder, 1992, p.42).

Já no campo da lexicografia, encontramos também os termos unidades lexicais ou léxicas, alfabeticamente organizadas em verbetes, tipicamente contendo informações referentes à morfologia, sintaxe, semântica e fonética da unidade; cada verbete é encabeçado pelo lema (*headword*), ou entrada, e constitui a forma canônica da palavra, a partir da qual se derivam outras formas com a adição de prefixos, sufixos e outros procedimentos tais como a flexão verbal.

Além de **lexemas simples**, também fazem parte do léxico os lexemas complexos e os lexemas compostos. O termo **lexema complexo** refere-se a dois processos de formação de palavras: a **derivação**, pelo qual há o acréscimo de algum tipo de afixo a uma raiz (ex.: *-mente*, que forma advérbios em português), resultando em formas diferentes, e a **flexão** que, a partir de uma raiz, produz todas as formas de um lexema que deverão ocorrer em ambientes sintáticos específicos (Lyons, 1977, p.521-2). Já o **lexema composto** é aquele cuja raiz é formada pela combinação de duas ou mais raízes (com ou sem modificação morfológica) (Lyons, 1977, p.534-5), processo chamado de **composição**; em português temos ocorrências do tipo substantivo + substantivo (*sofá-*

cama, peixe-espada), substantivo + adjetivo (*obra-prima, altar-mor*), verbo + substantivo (*guarda-roupa, mata-mosquito, porta-bandeira*). Em outras palavras,

A composição é um processo de formação de palavras que utiliza estruturas sintáticas para fins lexicais. Ou seja, mecanismos ou estruturas que são normalmente utilizados na formação de enunciados passam a ser utilizados na função de denominar e/ou caracterizar seres, que é uma função fundamental do léxico. (...) [é um processo que] permite a nomeação ou caracterização de seres pela junção de dois elementos semânticos, de existência independente no léxico, em apenas um elemento lexical. (Basílio, 1987, p. 30).

Destacam-se a nomeação descritiva (*papel-alumínio, navio-escola*), nomeação metafórica (*olho-de-sogra, louva-a-deus*), as combinações constantes (*guarda-roupa, guarda-chuva, guarda-costas; porta-aviões, porta-copos, porta-luvas; corre-corre, quebra-quebra*) e a composição por bases presas (ex.: a base *log-*, que aparece em palavras como *psicologia, ornitologia, patologia*, etc) (Basílio, 1987, p. 29-35).

2.2.1.2 O significado, sua composicionalidade e seus tipos

Ligada ao estudo do léxico, a Lingüística procura dar conta do estudo do **significado lexical**; quando da publicação de *Meaning of meaning*, em 1923 por Ogden e Richards, os autores levantaram uma série de hipóteses sobre a definição de significado, se seria uma propriedade intrínseca, palavras anexadas umas às outras no dicionário, a conotação de uma palavra, o lugar de qualquer coisa no sistema, ou símbolos aos quais os falantes recorrem para que seus referentes sejam reconhecidos, etc. (Leech, 1974, p. 01).

Posteriormente, Bloomfield considerou que o significado só poderia ser dado em termos de uma definição científica; Ogden, Richards e Bloomfield esbarravam no fato de que a grande dificuldade de se definir o significado está em sermos obrigados a usar a própria linguagem para fazê-lo, ou seja, no uso da metalinguagem: “não há como escapar da língua: uma equação como *cem=um cento* ou *sal=NaCl* não é a união de um

signo lingüístico com algo extralingüístico; é uma correspondência entre duas expressões lingüísticas que supostamente têm o mesmo significado” (Leech, 1974, p. 05, trad. nossa).

No entanto, não é possível considerar o significado apenas em termos de expressões lingüísticas; existe uma conexão entre língua e “mundo real”, conforme pode ser entendido através de exemplos. Em (1) *Meu tio sempre dorme sobre um de seus dedos do pé* e (2) *Meu tio sempre dorme acordado* temos certeza do absurdo expresso porque conhecemos o mundo em que vivemos e também pela contradição entre os dois significados de *dormir* e *acordar*; o problema provoca duas hipóteses: (1) se aquilo a que nos referimos pela palavra *significado* tem algum tipo de existência ou realidade ou (2) se tudo o que chamamos de *significado* tem um similar, ou idêntico, na natureza (Lyons, 1997, p.136). Portanto, as conceituações de significação variam entre o realismo extremo (que sobrepõe a língua ao mundo objetivo) e um relativismo que considera que é a língua que determina a capacidade do ser humano perceber o mundo (Ilari, 1985, p. 05); o fato é que o significado deve ser observado reconhecendo-se relações entre sentenças e entre seus elementos constituintes, o que pressupõe uma distinção entre conhecimento de língua e conhecimento de mundo (Leech, 1974, p.08-09).

Uma distinção significativa está entre o significado lexical e o significado sentencial: “o significado de uma sentença depende do significado de seus lexemas constituintes (incluindo lexemas frasais se existirem); e o significado de alguns lexemas, senão todos, depende do significado das sentenças em que ocorrem. Mas a estrutura gramatical das sentenças (...) também é relevante para a determinação de seu significado: portanto devemos reconhecer o significado gramatical como componente do significado da sentença” (Lyons, 1977, p.140).

Desta forma, o significado lexical está intimamente ligado ao significado em unidades maiores. Veja-se o caso das paráfrases: “além de descrever relações de sentido entre palavras, [essa relação serve para se] reconhecer relações de sentido entre construções gramaticais ou mesmo efeitos de sentido originados no contexto” (Ilari, 1985, 41). As unidades lexicais (sejam elas unidades independentes ou mesmo expressões) fazem parte do léxico, que não apenas lista os lexemas de uma língua, como também associa cada lexema com informações de dois tipos: sintática e morfológica. Tem-se assim constituída uma entrada lexical. Por exemplo, o verbo “ir” carrega dois tipos de informação: a) a de que pertence a uma ou mais subclasses de verbos intransitivos e (b) o paradigma de formas *ir, foi, ia, iria*, etc. (Lyons, 1981, p. 145). Conforme afirmado acima, muitos dos lexemas são compostos, como “couve-flor” e “sistema de produção de *clips* de papel” (Bierwish e Schereuder, 1992, p. 24); há ainda o caso de expressões das quais não há simplesmente como inferir um significado global a partir da soma das propriedades sintáticas e semânticas de seus constituintes. Um exemplo seria “chutar o balde”, que em português é sinônimo de “desistir”; quando uma expressão idiomática encontra correspondência com uma expressão não idiomática, diz-se que a última tem um significado literal, enquanto que a primeira, metafórica, tem um significado figurativo (Lyons, 1981, p.145). Portanto, o que chamamos de palavras constitui apenas uma parte do léxico de uma língua, e o significado lexical que aqui consideramos é o significado dos lexemas; é difícil distinguir o significado de lexemas compostos de clichês e colocações fixas, e, da mesma forma, embora o número de unidades do vocabulário de uma língua seja finito, é de tamanho indeterminado (Lyons, 1981, p.146).

Diante de tal diversidade tipológica, Leech (1975, p. 26) classificou sete tipos diferentes de significado, resumidos na Tabela 2:

Tabela 2. Os sete tipos de significado.

	1. Significado conceitual ou sentido	De conteúdo lógico, cognitivo ou denotativo	
	Significado associativo	2. Significado conotativo	O que é comunicado em virtude daquilo a que a língua se refere.
		3. Significado estilístico	O que é comunicado pelas circunstâncias sociais do uso da língua.
		4. Significado afetivo	O que é comunicado pelos sentimentos e atitudes do falante/escritor.
		5. Significado refletido	O que é comunicado pela associação com outro sentido da mesma expressão.
		6. Significado colocativo	O que é comunicado pela associação de palavras que tendem a ocorrer no ambiente de outra palavra.
	7. Significado temático	O que é comunicado pela maneira pela qual a mensagem é organizada em termos de ordem e ênfase.	

- **Significado conceitual** – também chamado de significado denotativo ou cognitivo; é formado a partir de dois princípios: contrastividade e estrutura de constituintes. Pelo princípio da contrastividade percebe-se a existência de traços contrastivos, como por exemplo +HUMANO, -MASCULINO, +ADULTO (para *mulher*) oposto a +HUMANO, +MASCULINO, -ADULTO (*menino*). Pelo princípio da estrutura de constituintes, entende-se que unidades lingüísticas maiores são constituídas por unidades menores (ex: uma sentença é constituída de sujeito e predicado, e assim por diante).
- **Significado conotativo** – ligado àquilo a que o significado “se refere”, em termos de seu valor comunicativo; desta forma, os traços componenciais de determinado objeto tornam-se um referente no mundo, o que portanto acaba por conferir-lhe mais ou menos traços de acordo com a época em que o referente esteja sendo utilizado (propriedades possíveis); o significado conotativo envolve cultura, período histórico e a experiência dos indivíduos. Portanto, ao contrário do significado conceitual, o conotativo é aberto, tão mutável quanto nossos sistemas de conhecimento e crenças.

- **Significado estilístico e afetivo** – ligado à situação em que uma expressão ocorre. Envolve aspectos dialetológicos, geográficos, sociais, grau de formalidade, época, se está em meio escrito ou falado, se é monólogo ou diálogo, se é língua de um domínio específico do conhecimento, modalidade e se identifica autores em particular.
- **Significado refletido e colocativo** – associações que o significado adquire com outros significados do mesmo ambiente de ocorrência. Ex: *bonito* e *belo*, referindo-se a uma pessoa de boa aparência, ou *Espírito Santo* e *aquele que conforta* para a terceira pessoa da trindade (Leech, 1974, p. 19-20).
- **Significado temático** – relacionado à maneira pela qual o falante organiza a mensagem que deseja veicular em termos de ordem, foco e ênfase. Ex: uma frase codificada no discurso direto e indireto; nela refletem-se as escolhas do usuário e suas intenções comunicativas. Ex: “*Sente-se aqui, já!*”, *disse Pedro* (discurso direto) e *Pedro pediu-lhe que se sentasse aqui* (discurso indireto).

Fora do escopo dos sete tipos de significado, Leech arrola mais dois tipos que considera relevantes, o significado pretendido (o que o falante pretende que seja entendido) e o significado interpretado (o que de fato o ouvinte entendeu), deixando clara a distinção entre o que se quer comunicar (intenção) e o que se consegue comunicar (efeito), visto que falante e ouvinte podem ter diferentes experiências cognitivas em relação a uma mensagem qualquer: a mente que emite uma mensagem não é igual àquela que a recebe.

Para a construção da base lexical do *thesaurus eletrônico*, a escolha deve recair na inserção de unidades lexicais que se referem aos significados denotativo e conotativo das unidades lexicais, não descuidando-se da questão da **polissemia** e **homonímia**.

Pela polissemia, distintos conteúdos são lexicalizados na mesma unidade lexical, ou dito de outra forma, são acepções diferentes para uma mesma entrada (Vilela, 1994, p. 26). Temos por exemplo *sorte = destino* e *sorte = fatalidade*. No dicionário, a polissemia é expressa por meio da indicação das várias acepções em um mesmo verbete.

Pela homonímia, dois ou mais significados diferentes podem associar-se a unidades lexicais distintas; um exemplo clássico do português é *bancol* (instituição financeira) e *banco2* (assento).

Segundo Lyons, a distinção entre homonímia e polissemia é indeterminada e arbitrária (1981, p. 431); no entanto, muitos teóricos buscaram a diferenciação aplicando critérios que permitissem uma distinção entre itens homônimos e polissêmicos, dos quais destacamos os critérios formal, semântico e etimológico.

Pelo **critério formal**, temos três aspectos a serem observados:

- Distribuição - a homonímia é vista sincronicamente e leva em conta a distribuição dos itens nas sentenças; portanto, se a distribuição na frase for diferente, estamos diante de um caso de homonímia. Ex: *um canto alegre / eu canto alegremente* – o primeiro *canto* é substantivo, o segundo é verbo; portanto, estando distribuídos de maneira diversa, trata-se de homonímia. A polissemia, por outro lado, envolve a identidade dessa distribuição, como em *o cabo do pelotão* e *o cabo da vassoura* (Almeida, 1990, p. 188); mas há problemas com esse critério, porque em “o banco da França” e “o banco da praça” a distribuição é idêntica e estamos diante de itens homônimos.
- Homofonia – são homônimos os lexemas com a mesma realização fonética. Ex: *nós* (terceira pessoa do plural) e *noz* (Lyons, 1981, p. 430);
- Homografia – são homônimos os lexemas com a mesma grafia. Ex: *sede* (secura) e *sede* (segunda pessoa do presente indicativo do verbo *ser*) (Lyons, 1981, p. 430).

Pelo **critério semântico**, a homonímia ocorre quando não há coincidência de nenhum traço componencial. Ex: *mangal* (fruta) e *manga2* (parte do vestuário). Já a polissemia ocorre se na oposição significativa das formas houver a intersecção de pelo menos um traço. Ex: *capa de chuva* e *capa de livro* – ambos possuem em comum o traço *cobrir* (Almeida, 1990, p. 188).

Finalmente, pelo **critério etimológico**, os casos de homonímia são identificados quando formas diferentes de origem dos lexemas convergem para uma mesma forma. Ex: *lat. SANUN:são* (saudável), *lat. SUNT: são* (terceira pessoa do verbo *ser*) (Vilela, 1994, P. 27).

A questão da polissemia e homonímia é fundamental na compilação de lexemas para a base lexical, como veremos no capítulo 6.

2.2.2 O léxico mental

Além de ser objeto de estudo da lingüística, o léxico também é alvo de investigações na psicolingüística, que fornece modelos de como o léxico se comporta no cérebro humano. O conhecimento lexical é parte essencial do conhecimento de uma língua, que deve ser concebido como um estado complexo, mais ou menos estável, do cérebro; a parte desse estado, cuja estrutura é formada pelo sistema lexical da língua, é denominada **léxico mental** (Bierwish e Schreuder, 1992, p. 26), tema desta seção.

Antes da produção de qualquer tipo de discurso que faz uso da linguagem humana, o falante realiza automaticamente uma seleção de itens lexicais adequados para a expressão de um pensamento: “a escolha dos itens lexicais em última instância reflete um grande leque de condições diferentes, mas de alguma forma relacionadas, que determinam a verdade, a propriedade e o efeito comunicativo das expressões lingüísticas” (Bierwish e Schreuder, 1992, p. 24, trad. nossa). Sendo assim, um falante

seleciona os itens lexicais de acordo com seu modelo mental interno e em função do ambiente externo no qual está imerso; desta forma, o falante incorpora os diferentes tipos de informação visual, auditiva, motora, inferencial, de conhecimento de mundo, etc, que acabam por moldar a mensagem que deseja verbalizar. Uma evidência disso é que uma mensagem pode ser expressa de diversas maneiras entre línguas diferentes ou até mesmo dentro de uma mesma língua; podemos dizer “Fique aqui!”, “Não saia daqui!”, ou “Não se mova!” como diferentes formas de expressar o mesmo conteúdo comunicativo.

A representação de um item no léxico mental é chamada de entrada lexical, que pode ser subdividida em unidade de acesso e especificação lexical correspondente, que contém a estrutura fonológica, morfológica, sintática e semântico-conceitual da entrada (Handke, 1995, p.113). Esta proposta adveio de Levelt, 1989 (*apud* Handke, 1995, p. 69) que, ao investigar a estrutura das entradas lexicais, sob o ponto de vista da produção de linguagem, concluiu que o processo de codificação gramatical, ou seja, o processo de gerar a estrutura de uma sentença, é independente da relação entre conteúdos e formas, sugerindo que a subdivisão básica de uma entrada lexical contenha aspectos relacionados à forma (neste contexto constituindo os “lexemas” ou “formas”) e aspectos relacionados ao conteúdo (neste contexto constituindo os “lemas”), ambos interligados por uma espécie de ponteiro lexical, isto é, um espécie de rótulo lexical, conforme ilustra a Figura 4 (Handke, 1995, p. 69).

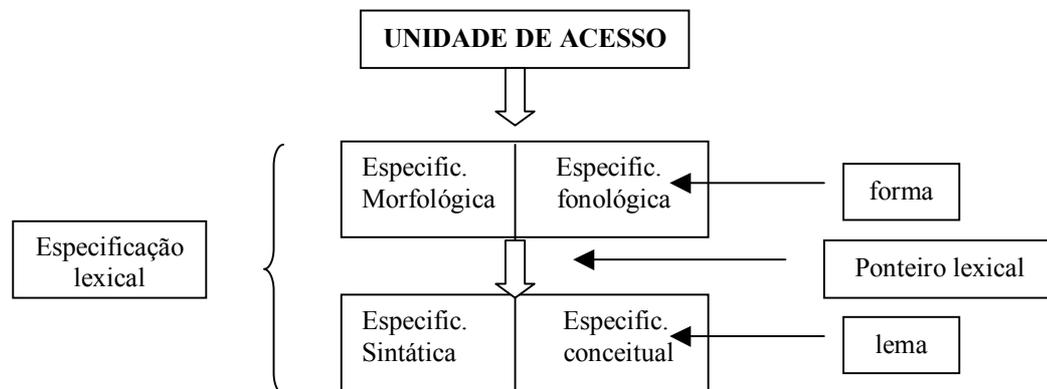


Fig. 2. Entradas lexicais fracionadas.

No estudo dos modelos de processamento lexical é possível refletir sobre a estrutura do léxico mental; vejamos, na Figura 3, a proposta de Handke (1995, p.35), baseada no esquema de representação das estruturas conceituais de Levelt (1989, 1992).

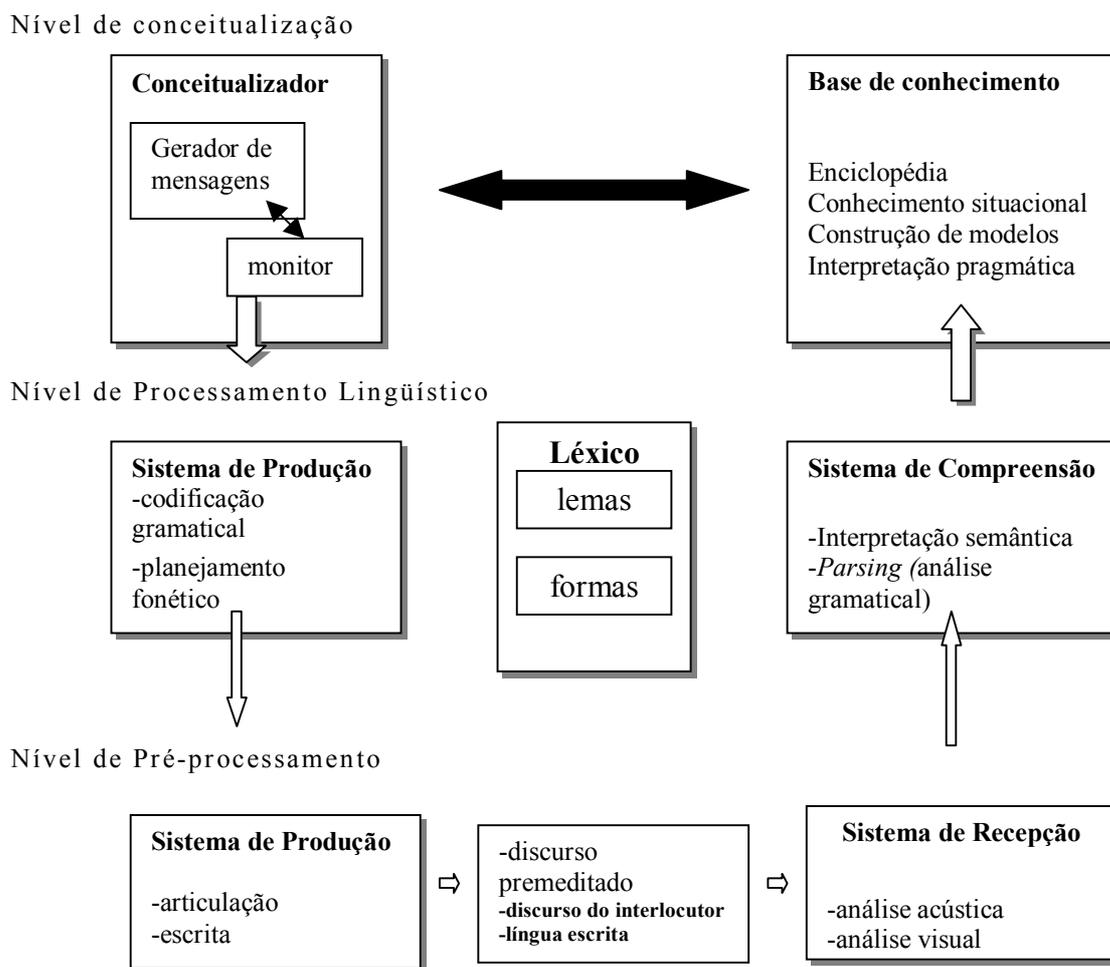


Fig. 3 – Modelo de representação do processamento lexical.

Proposta por Bierwish e Schereuder (1992, p. 25), apresentamos na Figura 4 a versão simplificada do mesmo modelo.

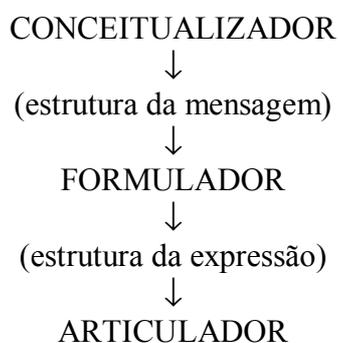


Fig. 4 – Esquema simplificado do processamento lexical.

No nível da conceitualização temos o **conceitualizador**, que interage com uma base de conhecimento que proporciona o conhecimento geral necessário para a geração e interpretação de uma mensagem; ele abrange todo tipo de informação perceptual, motora, emocional, conceitual e fornece a estrutura da mensagem que deverá então ser verbalizada e gramaticalmente organizada pelo **formulador**. O sistema representacional da estrutura da mensagem pode ser chamado de estrutura conceitual (EC); a EC é dependente da língua e está baseada em princípios gerais de organização, incluindo uma ontologia de conhecimentos comuns, conceitualização de espaço e tempo, condições gerais subjacentes ao conhecimento enciclopédico e sistemas de crenças.

No nível do processamento lingüístico temos o **formulador** que opera com elementos lingüísticos, de modo distinto conforme o sentido seja a produção ou a compreensão da linguagem. Na produção, a EC é transformada em uma estrutura lingüística, processo este subdividido em duas partes: (i) um primeiro estágio que lida com aspectos sintáticos e semânticos, convertendo a EC em uma estrutura sintática que descreve as principais propriedades de frases e (ii) um segundo estágio, em que a estrutura sintática é transformada em uma estrutura fonética. Já na compreensão, o processo é inverso: o sinal que chega é primeiro transformado em uma seqüência morfofonêmica e então esta é ampliada com informações sintáticas e semânticas e posteriormente transformada em uma representação conceitual.

Tal subdivisão do processamento lingüístico se dá devido à divisão de cada unidade lexical em duas partes: o **lema** (portadores de informações sintáticas e semânticas) e a **forma** (portadora das especificações morfológicas e fonológicas).

Um outro aspecto ligado ao formulador é que o significado lexical determina o escopo de configurações conceptuais às quais as unidades lexicais e suas combinações

devem corresponder; além do mais, os significados lexicais associam-se às condições invariantes de sua realização fonética. É através da informação dessa forma fonética (FF) que o **articulador** produz uma expressão lingüística.

O articulador presente no nível do processamento primário também engloba a produção e a compreensão da fala. Antes do articulador entrar em ação, o plano fonético é transformado em um programa fisiológico que serve de instrução para o sistema nervoso central e resulta ou na articulação da fala audível ou nas ações fisiológicas da produção da escrita.

No sentido da compreensão, o sinal da fala é acusticamente analisado por um componente perceptivo e, através de vários estágios intermediários, esse componente transforma a onda sonora em uma estrutura lingüisticamente interpretável. Se o sinal a ser analisado for gráfico, caso da escrita, postula-se um componente que opera no reconhecimento dos símbolos gráficos (Cf. Bierwish e Schereuder, 1992).

Central a todo o processamento está o **léxico**, que contém as unidades lexicais da língua e uma grande quantidade de informações associadas a cada um deles. A centralidade do léxico é atualmente reconhecida por lingüistas, psicolingüistas e lingüistas computacionais (Handke, 1995, p.39).

A forma semântica restringe a interpretação de uma mensagem no domínio conceitual, a exemplo do que ocorre com a forma fonética, que restringe o escopo de realizações articulatórias e faz ajustes de acordo com os padrões do sistema auditivo. É com essa relação entre a forma semântica e a estrutura conceitual que vamos lidar agora, a partir de três considerações principais (Bierwish e Schereuder, 1992, p. 31-32):

- 1) É preciso assumir que existem fatores lingüísticos e extra-lingüísticos na composição do significado; desta forma, a estrutura conceitual dá conta de todo o tipo de conhecimento factual ou crença que claramente não pertencem ao domínio

lingüístico; devem ser organizados e restringidos por condições que são independentes do conhecimento lingüístico. Temos, por exemplo, o caso de “João”. Temos um “João” enquanto entrada lexical e cada “João” que conhecemos, e isso não tem relação com o conhecimento de nossa língua, nossa gramática, etc. Saber mais sobre o João que mora em frente nossa casa afeta consideravelmente nosso conhecimento conceitual sobre ele, mas não afeta nosso conhecimento lexical sobre o item “João”;

2) Como conseqüência, podemos ter diferentes interpretações no caso do uso de certos itens lexicais em frases ambíguas; no caso, a ambigüidade só é resolvida graças ao conhecimento de mundo, que na Inteligência Artificial tem sido tratado quando se fala em *frames* e *scripts*;

3) Há que se considerar o conhecimento situacional que está envolvido na especificação da estrutura conceitual, e que não pode ser derivado do conhecimento lingüístico. A especificação das variáveis de local inicial ou final em “João foi” e “João vai” só pode ser determinada via informação contextual ou situacional.

Em suma, a estrutura conceitual em termos da qual a interpretação de expressões lingüísticas é especificada tem suas condições estabelecidas pela forma semântica, por meio de conhecimento enciclopédico, de informação contextual e de condições situacionais.

Diante de tais constatações, a estrutura do sistema lexical de uma língua L deve refletir várias facetas da linguagem, desde sua forma fonética até sua forma conceitual; assim sendo, uma entrada canônica E pode ser concebida como uma estrutura de dados, no sentido computacional desse termo, contendo vários tipos de informação agrupadas em 4 componentes (Bierwish e Schreuder, 1992, p. 26). Exemplo: o verbo *entrar*:

/entrar/ [+V, -N] $\lambda x \lambda y$ [y DO [MOVEy] : FIN[yLOC in x]]			
FF	TG	EA	FS

Em que:

FF – *Forma fonética* – especifica as condições segmentais e suprasegmentais que E impõe na realização fonética da expressão em que E ocorre.

TG (E) – *Traços gramaticais* - determinam as propriedades morfológicas e sintáticas de E e dos constituintes de hierarquia superior dos quais E é núcleo. Portanto, [+V –N] classifica *entrar* como um verbo que pode ser o núcleo de um sintagma verbal.

EA (E) – *Estrutura de argumentos* – seqüência de uma ou mais posições argumentais que especifica o número e tipo de complementos exigidos por E; no caso *entrar* exige um **locativo**.

FS (E) – *Forma semântica* – restringe o conteúdo proposicional da expressão contendo E.

Portanto, a partir do exame dos estudos acerca da constituição do léxico mental associados aos estudos do PLN, podemos verificar que o sistema lexical não é somente um conjunto fixo de elementos básicos organizados de acordo com os princípios do próprio sistema, mas também contém elementos que estendem as entradas lexicais básicas a um conjunto inclusivo de entradas lexicais que compreende as **entradas básicas** e as **complexas**, sendo que as entradas complexas, regulares e previsíveis com base em suas partes constituintes, não entram no conjunto de entradas “reais” mas constituem o que se pode chamar de entradas lexicais virtuais (Bierwisch e Schreuder, 1992, p.29); em português, por exemplo, um termo como “imexível” pode ser considerado virtual, pois a língua prevê mecanismos para sua existência mas a norma culta não o aprova (devido à existência de uma alternativa para o termo, “intocável”) deixando-o sob a condição de mera potencialidade. A Figura 5 esquematiza a proposta desses autores quanto à tipologia das entradas lexicais do léxico mental (Bierwisch e Schreuder, 1992, p. 30).

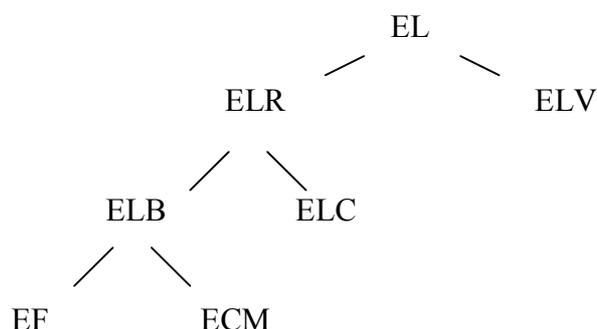


Fig. 5. Tipos de entradas do léxico mental.

Nesse esquema, o conjunto das entradas lexicais do léxico mental (EL) subdivide-se em dois tipos: entradas lexicais reais (ELR), que são as unidades que estão em uso (p. ex. *menino, comer, uva*) e entradas lexicais virtuais (ELV), que são unidades potenciais que passam a existir somente depois de verbalizadas e cuja permanência depende de aceitação da comunidade lingüística (ex. *imexível*). Dentro do conjunto das entradas lexicais reais, distinguem-se dois subtipos: as entradas lexicais complexas (ELC) e as entradas lexicais básicas (ELB). As primeiras são unidades lexicais resultantes de processos de derivação, flexão e outros processos morfológicos (p. ex. *couve-flor, porta-jóia, dir-se-ia*). Já as segundas abarcam as entradas de categorias maiores (ECM) e os elementos funcionais (EF). Aquelas correspondem às formas canônicas das categorias maiores (verbo, substantivo, adjetivo e advérbio) como, por exemplo, *casa, vela, canto*, e estas os elementos de natureza gramatical: *de, para, por*; além dos elementos flexionais: *-a, -ar, -s*, etc.

A classificação proposta acima para o léxico mental é um reflexo do léxico da língua, servindo, portanto, de sistematização do aparente caos que cerca o imenso conjunto de unidades lexicais de qualquer língua natural. Na tarefa de compilação de uma base lexical para o *thesaurus* é necessário ter em vista que tipo de unidades estão

presentes no sistema lexical para dele derivar uma parcela para a constituição do léxico a ser computacionalmente modelado; diante disso, fica justificada a exploração do conhecimento em torno do léxico mental em nosso trabalho.

2.2.3 O léxico computacional

No domínio do PLN, conforme mencionado no Capítulo 1, o termo **léxico** está diretamente associado à construção de bases lexicais voltadas para os mais variados propósitos de implementação; no entanto, cabe ressaltar uma importante distinção entre **léxico em uso** e **construção de léxico**. O léxico em uso refere-se à funcionalidade geral do módulo lexical dentro de um sistema de PLN, cujos aspectos remetem ao estudo do léxico mental; tal estudo permite delinear importantes semelhanças e diferenças entre o léxico armazenado na mente humana e o léxico a ser construído para fins computacionais, já que, pelo menos idealmente, a implementação de um léxico em meio digital e um poderoso algoritmo de manipulação de dados são motivados pelo que se sabe a respeito do processamento mental da linguagem humana e a simulação dos estágios desse processamento por um computador.

Desta forma, se a construção da base lexical de nosso dicionário de sinônimos e antônimos é elaborada a partir de princípios norteadores do PLN, é pertinente assinalar que são muitas as propostas de estruturação do léxico para fins computacionais; entre elas, podemos citar modelos criados por Boguraev e Briscoe (1989), Jackendoff (1991 e 1993), Handke (1995), Grefenstette (1994), etc. Todos esses modelos de construção de léxicos lidam com questões referentes à implementação, geração e administração de bases lexicais eletrônicas. A base lexical subjacente a um *thesaurus eletrônico* deverá

conter essencialmente a unidade lexical, um rótulo gramatical (substantivo, adjetivo, etc.) e estar estruturada em função das relações de sinonímia e antonímia.

A aparente distância entre as unidades do léxico mental e computacional é minimizada quando observamos que existem propriedades comuns entre as unidades semânticas e naturais. Uma gama variada de conceitos são indubitavelmente partilhados entre a psicolinguística e a linguística computacional. Sendo assim, vários níveis de descrição são aplicáveis às unidades léxico-semânticas: em um primeiro nível, o mais abstrato, as unidades semânticas são objetos físicos, o que leva a uma grande diferença entre as unidades naturais e as artificiais: um símbolo da máquina é um determinado estado dos componentes físicos do computador, enquanto um símbolo mental é um estado da rede neural integrante do cérebro; num segundo nível, considera-se a organização local dos símbolos, representada por um conjunto de *bits* e por uma determinada rede de neurônios e sinapses. No último nível, o cognitivo propriamente dito, as unidades lexicais são vistas na concepção saussureana de signo linguístico, possuindo duas partes (o significante e o significado), ou seja, a representação de aspectos morfológicos (conteúdo fonológico e ortográfico) e aspectos semântico/conceituais (seu significado) (Le Ny, 1995, p. 51). Desta forma, passam de unidades lexicais para o estatuto de unidades semânticas.

Assim sendo, as propriedades que se atribuem às unidades semânticas nas teorias de representação do conhecimento ou mesmo na linguística computacional encontram paralelo na psicologia cognitiva, sendo que todas são relacionais. A seguir mostramos as propriedades partilhadas e exploradas na semântica linguística e computacional. Nas fórmulas, L refere-se ao conteúdo semântico de um item lexical particular (Le Ny, 1995, p. 52-55). As propriedades são apresentadas a seguir.

2.3 Propriedades partilhadas entre unidades lingüísticas e computacionais

São as seguintes propriedades (cf. Le Ny, 1995):

2.3.1 Propriedades básicas

2.3.1.1 Denotação ou categorização

Pode ser representada pela fórmula $D(L,y)$, em que se lê “L se aplica a y”, “L toma y como uma de suas instâncias”, no qual y é um elemento de um determinado conjunto (extensão de um conceito ou significado associado a L); está presente em enunciados explícitos como “y é um L”, “y é uma instância de L”, ex: “Sócrates é um filósofo”, ou “o filósofo...” (tendo Sócrates sido mencionado antes). Tal propriedade também está envolvida em todas as atividades mentais que envolvem categorias ou conceitos: percepção, julgamento, predicação verbal, pensamento, etc. Em termos lógicos, esta propriedade é expressa pelos termos que possuem o papel funcional de predicadores naturais.

2.3.1.2 Superordenação/subordinação

Pode ser expressa por $Z(L,S)$, que corresponde a: “L é uma sub-categoria de S” ou “S é uma super-categoria de L”; geralmente “L é um tipo de S” ou “um L é um tipo de S”, construindo uma hierarquia que pode ser organizada em uma árvore conceitual. Ex: “um pardal é um pássaro, um pássaro é um animal...”.

2.3.1.3 Atribuição

Atribuição de pares de valores, propriedades, traços semânticos ou relações semânticas, em que uma categoria pode ser expressa por $A(L,p)$, sendo que o argumento p pode ser visto de várias formas, de acordo com o ponto de vista teórico; por exemplo:

- a) envolve a propriedade (L,p) , ou tem a propriedade (l, p) , ou envolve os traços (L, f) ou (l, f) , que é uma reformulação de p . Por exemplo, o “conceito de laranja” tem o traço “redonda”;
- b) expressões mais complexas podem ser elaboradas em termos de atributos/valores: $A(L,U,v)$, sendo o atributo v seu valor; por exemplo: o “conceito de laranja” envolve “forma” com o valor “redonda”.
- c) Relações ainda mais complexas envolvem:
 - Disjunção – ex: “uma maçã tem as cores verde, ou amarela, ou vermelha, etc.”
 - Frequências superficiais de valores – ex: “determinada qualidade de maçãs é geralmente cinza, às vezes amarela, raramente verde e nunca vermelha.”

2.3.1.4 Demandas sobre os preenchedores de papéis de caso

Refere-se aos preenchedores de papéis de caso (Fillmore, 1968) ou actanciação de verbos (Tesnière, 1959), $C(L,G,H,I)$; a relação diz respeito, por exemplo, a verbos que têm necessariamente dois papéis, em que G representa um agente, H um paciente, I vazio, como em “o menino empurra a menina”, e assim por diante.

2.3.1.5 Estruturas semânticas ou conceituais

Envolvem conjuntos de relações de várias ordens, e papéis, que correspondem a vários tipos de esquemas, cenários, *frames* e *scripts*, conceitos científicos, etc. Ex: o conceito de “restaurante” envolve os conceitos de “comida”, “comer”, “pagar”, etc.

2.3.2 Propriedades derivadas ou de segunda ordem

2.3.2.1 Implicação

$E(L,b)$ vincula ou implica (L,B) – ex: “se x é um pardal, então é um pássaro”.

2.3.2.2 Herança

$I(L,S,\{U\})$, ou “herda de $(L,S,\{U\})$ ”, onde S é uma unidade superordenada e $\{U\}$ um conjunto de atributos ou propriedades transferidos de S para L . Ex: “um pardal tem todas as propriedades de um pássaro”.

2.3.3 Propriedades psicológicas primárias

2.3.3.1 Similaridade

$S(L,B)$ ou “similar a (L,B) ”, ou $S(L,B,g)$, em que a similaridade inclui o atributo “ g ”; ex: “uma concha é similar a uma casa”. A similaridade é graduável: “ L é mais similar a B do que é a C ”, ou “ L é mais similar a B do que C ”.

2.3.3.2 Tipicalidade ou representatividade

Expressos por: $T(L,S)$, “típico de (L,S) ”, em que S é um superordenado de L , e L um hipônimo de S , mais típico do que outros, é um protótipo; ex: “pardais são pássaros típicos”.

2.3.3.3 Hierarquias de nível básico

$BL(L,S,I)$, em que S é a unidade maior, e I uma unidade no menor nível, definindo este nível. Ex: “cão” é o nível cognitivo básico na hierarquia de animais (de “animais” para “pastor alemão”).

Diante das semelhanças entre unidades lingüísticas e computacionais é possível buscar a proposição de formalismos para a representação do significado lexical.

Em primeiro lugar, temos os formalismos empregados para representar o significado lexical que são implementadas no âmbito da lingüística computacional. São eles os traços semânticos, as redes semânticas e os protótipos (Handke, 1995, p. 91; Dias-da-Silva e Oliveira, 2000). Em segundo lugar, temos os formalismos que procuram dar conta da estrutura da entrada lexical em termos das informações que contém e de suas relações com outras unidades do léxico; destes selecionamos dois: o léxico gerativo (Pustejovsky, 1991, 1994; 1996) e as estruturas conceituais (Jackendoff, 1993).

Começemos pelos formalismos propostos para representar o significado lexical.

2.4 Teorias para a formalização de significados

2.4.1 Traços semânticos

Os traços semânticos surgiram com os trabalhos de Katz/Fodor (1963) e Katz/Postal (1964), constituindo um dos primeiros formalismos para a representação formal do significado. Esse formalismo pressupõe uma técnica de segmentação dos conceitos lexicalizados nas unidades lexicais em partes menores. A essas unidades do significado dá-se o nome de traços semânticos, átomos de significado, primitivos semânticos ou componentes semânticos. Esses componentes não são lexemas, e sim parte de uma metalinguagem usada para a descrição dos componentes conceituais realizados nas línguas. Um exemplo de decomposição do item lexical por meio de pares de atributo valor – para cada atributo, um tipo de valor - vem a seguir:

[mosca: (ANIMAL), (INSETO), (2_ASAS), (6_PERNAS), (SEM_FERRÃO)]

Para fins de representação, é adequado que cada um dos atributos, como (ANIMAL), esteja associado a um valor específico, que pode ser binário ou numérico.

Desta forma, o resultado seria este:

[mosca: (ANIMAL+), (INSETO+), (ASAS:2), (PERNAS:6), (FERRÃO-)]

Existem atributos que não têm valor numérico ou binário, mas valor atômico.

Ex: (NOME:JOÃO).

Esse formalismo pode ser também empregado para representar um dos sentidos do verbo “voar”:

[voar: (TRANSFERIR+), (MOVER_ASAS-)]

Essa fórmula expressa: “alguém realiza o ato de voar sem usar asas”. É o caso de, por exemplo, “João voou para Londres”, como passageiro.

Esse formalismo possibilita ainda uma definição formal de relações de sentido. Por exemplo, um lexema A é um hipônimo de B se todos os traços de B estão contidos na especificação de traços de A. No caso da antonímia, pode-se dizer que é uma relação entre conjuntos de lexemas, se eles compartilham uma série de traços, mas diferem em um ou mais traços contrastantes.

Assim, concluímos que o formalismo dos traços é uma ferramenta poderosa para a representação do significado, especialmente para as relações de sentido. Mas ela também apresenta problemas, sendo o principal a seleção dos traços: quais e quantos traços postular? Questão que permanece sem resposta.

2.4.2 Redes Semânticas

É um formalismo que tem por base a psicologia e a ciência da computação, trazido à tona por especialistas como Collins (cientista da computação) e Quilian (psicólogo) na

década de 60. Pode ser visto como um modelo psicológico e computacional da memória humana.

Originalmente, as redes semânticas representavam conceitos nominais. Nos anos 70 e 80 o formalismo ganhou complexidade ao se adicionarem estudos de conhecimento predicativo, mecanismos para se lidar com conectivos lógicos, quantificação, causalidade, etc.

Numa rede semântica simples, os conceitos (objetos e eventos) são representados por nós e as inter-relações entre conceitos, por arcos ou *links*, conforme a figura 6 abaixo (Handke, 1995, p. 95).

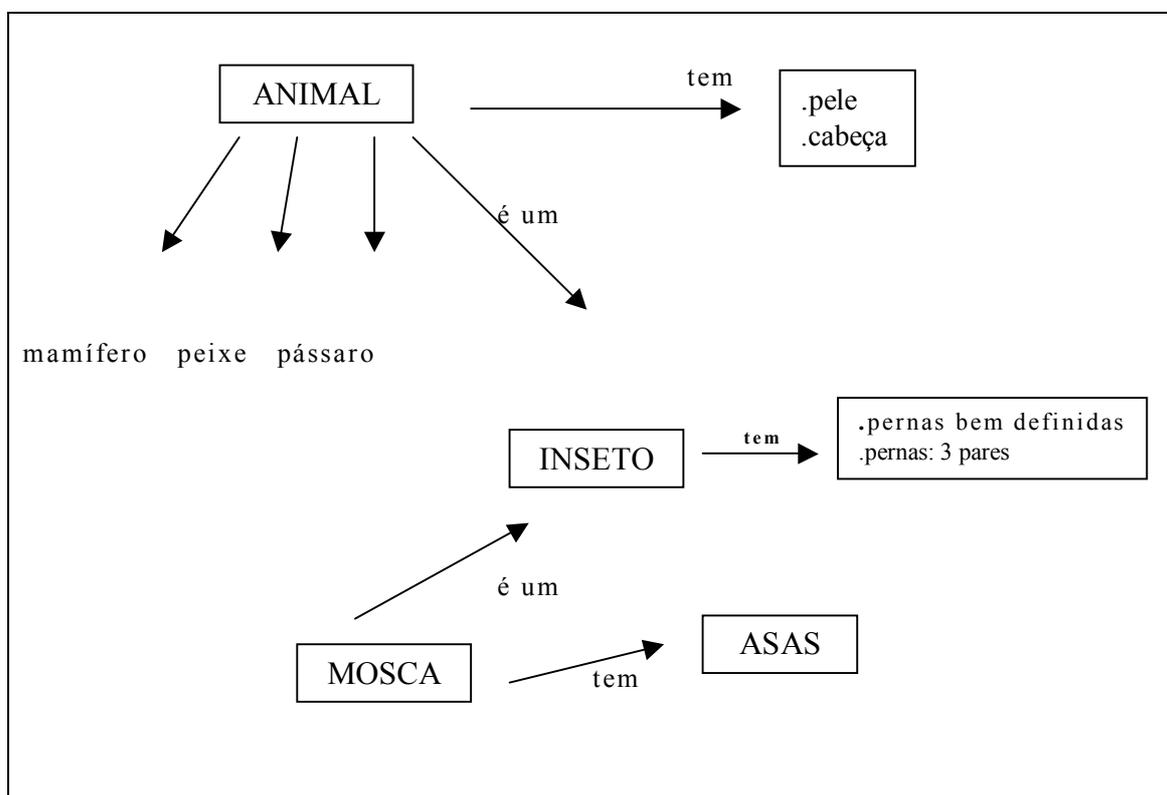


Fig. 6. Exemplo de uma rede semântica simples

As redes semânticas, pela forma elegante de representação de significados e pela facilidade de se inferir deduções, têm sido amplamente empregadas no âmbito da Inteligência Artificial, na ciência cognitiva e, conseqüentemente, no PLN. Por exemplo, para deduzir que a mosca tem cabeça, é só traçar a hierarquia É-UM, assumindo que os atributos associados com nós hierarquicamente superiores também são válidos para nós

hierarquicamente inferiores. É o que se chama de propriedade de herança, que permite inferir que MOSCA tem todos os atributos de INSETO, e que tanto INSETO como MOSCA têm todos os atributos de ANIMAL.

O formalismo é expressivamente forte porque permite formalizar inúmeras inter-relações através da rede. Ele ainda dá conta de instâncias conceituais, quando, por exemplo, permite captar os diferentes tipos de ASAS com seus atributos específicos (asa de inseto, pássaro, avião).

Em suma, o modelo é uma ferramenta poderosa para descrever e implementar hierarquias e associações de conhecimento e de conceitos, que através do mecanismo de herança, elimina informações redundantes, evitando a multiplicação desnecessária de traços semânticos para cada conceito.

Apesar de ainda serem usadas em aplicações computacionais, não são mais exploradas como modelos da memória humana. Atualmente, seu emprego restringe-se aos modelos conexionistas (Handke, 1995, p.97).

2.4.3 Teoria dos protótipos

Em meados dos anos 70, as teorias acerca da representação da linguagem na mente eram do tipo objetivista, segundo o qual o pensamento racional envolve a manipulação de símbolos abstratos arbitrários. Esses símbolos funcionam como representações internas da realidade externa, e as regras que as manipulam não têm acesso ao significado desses símbolos (Lakoff, 1988, p.119, *apud* Handke, 1995, p.97). Por exemplo, a sintaxe das línguas naturais pode ser vista como um conjunto de princípios algorítmicos que são aplicados a símbolos, palavras por exemplo, sem preocupação com o significado. Mais tarde esse posicionamento revelou-se inadequado, pois o

confinamento do pensamento a símbolos arbitrários cuja estrutura interna não pode ser utilizada pelos processos que operam sobre eles não pode ser mantido empiricamente.

Por outro lado, uma posição experimentalista revelou que estruturas conceituais significativas estão centradas na experiência social e em algum tipo de capacidade inata para estruturá-los.

Antes do surgimento dos termos cognição objetivista e experimentalista, este último era chamado de teoria dos protótipos. Nessa teoria, a categoria de pertença (*membership*) está centrada na representação do membro prototípico da classe que um certo item denota. Experiências baseadas nessa abordagem trouxeram dois tipos de resultados: prototípicos e de nível básico.

Os resultados prototípicos, por exemplo, demonstram que PARDAL é um membro mais representativo de PÁSSAROS do que AVESTRUZ ou PINGUIM. O membro mais representativo de uma categoria é chamado protótipo.

O protótipo em si não é o significado de uma palavra. Ele está localizado num espaço multidimensional com dimensões correspondentes aos atributos.

Os resultados de nível básico demonstram que certas categorias são mais básicas que outras, porque são reconhecidas mais rapidamente, aprendidas mais cedo, usadas mais frequentemente e processadas mais facilmente que outras categorias. Por exemplo, MOBÍLIA é o superordenado, CADEIRA é o nível básico e CADEIRA DE BALANÇO é uma categoria subordinada. A idéia por trás da postulação de categorias de nível básico é que elas:

- servem de base para a interação entre membros de uma categoria;
- servem como âncoras perceptuais;
- proporcionam imagens mentais que refletem a categoria toda.

Ambos os resultados são fundamentais para a categoria de associação, e portanto para a classificação do significado.

A teoria dos protótipos é potencialmente relevante para este trabalho pois, a partir da base lexical, poder-se-á realizar trabalho de verificação se um determinado conjunto de sinônimos aponta para um sentido prototípico ou não.

CAPÍTULO 3 - ARQUITETURA DE LÉXICOS

Sob um ponto de vista lingüístico, as unidades lexicais presentes num dicionário são definidas através do conhecimento lingüístico, por exemplo, através das relações de sentido como a sinonímia/antonímia, hiperonímia/hiponímia, etc. Associado ao conhecimento lingüístico está o conhecimento enciclopédico, que envolve conhecimento de mundo e relaciona uma entrada lexical a um esquema da memória humana; muitos sistemas desenvolvidos no âmbito da Inteligência Artificial utilizam-se de léxicos “enriquecidos” com informação conceitual; tais modelos fornecem uma estrutura do léxico mental e seu comportamento em relação ao nível lingüístico da linguagem, associando portanto, conceitos a itens lexicais da língua. Tal relação é explorada, de perspectivas diferentes, por Pustejovsky (1991,1994, 1996) e Jackendoff (1993).

3.1 O modelo de Pustejovsky

Pustejovsky (1991, 1994, 1996) defende que o uso de uma estrutura de representação do conhecimento oferece um vocabulário mais rico de informação lexical; sua teoria busca dar conta da criatividade da linguagem, criando mecanismos que possibilitariam a implementação de sistemas mais abertos, diferentes de algumas visões correntes que estudam o léxico de maneira estática. O autor lança mão da co-composicionalidade na derivação dos significados de um item lexical em diferentes contextos: “o contínuo refinamento e redefinição de que papéis um objeto desempenha em nosso ambiente, e como conceitualizamos aquele objeto como tendo diferentes propriedades em diferentes contextos é o processo de co-composição” (Pustejovsky, 1998, p.290, trad. nossa); tal afirmação é derivada da composicionalidade de Frege, que informalmente pode ser colocada da seguinte forma (Gazdar; Mellish, 1989, p.280) :

O significado do todo é função dos significados das partes.

A organização tradicional do léxico para fins de PLN assume que o significado de uma palavra pode ser exaustivamente definido por um certo número de sentidos. Uma desvantagem desse pressuposto é a necessidade de se especificar, no futuro, os novos contextos em que a palavra poderá ocorrer:

Sem uma apreciação da estrutura sintática de uma língua, o estudo da semântica lexical está destinado a falhar. Não há como divorciar o significado da estrutura que o carrega. (...) ; os significados das palavras deveriam de alguma forma refletir as mais profundas estruturas conceituais do sistema cognitivo e o domínio em que operam. (Pustejovsky, 1995, p. 5-6, trad. nossa)

Pensemos, por exemplo, num sistema de tradução automática, em que um programa tenha que escolher entre os sentidos da palavra aquele que mais se aproxima do contexto em questão. Mas... e se nenhum dos sentidos oferecidos se encaixar?

Segundo Pustejovsky, isso acontece porque as relações entre os sentidos dados num dicionário são muito rígidas e não refletem a inerente criatividade lingüística. Nem sempre é tarefa trivial selecionar o sentido adequado a um dado contexto, e conseqüentemente, resolver ambigüidades também não é tarefa fácil. Em seu modelo, ao invés de se especificar, para cada unidade lexical, um número fixo de sentidos, especificam-se aspectos representativos do significado lexical.

Para dar conta de certos fenômenos lexicais comuns, Pustejovsky e Boguraev (1993, p.203) postulam quatro níveis de representação:

- **Estrutura de argumentos** - relaciona a realização sintática de uma palavra com o número e tipo de argumentos que são identificados no nível da sintaxe e usados no nível da semântica.

- **Estrutura de eventos** - identifica o tipo de evento expresso pelos predicadores. Existem essencialmente três componentes para esta estrutura: o tipo de evento primitivo - estado (S), processo (P) ou transição (T): o foco do evento; e as regras para a composição de eventos.
- **Estrutura *qualia*** - define os atributos essenciais dos objetos, eventos e relações, associados com um item lexical. Ao postular componentes separados, o que é, em essência, uma estrutura de argumentos para nominais, os substantivos são elevados da posição de argumentos passivos para elementos ativos.
- **Estrutura de herança lexical** - determina as maneiras pelas quais uma palavra se relaciona com outras palavras no léxico. Além de oferecer informação sobre a organização de uma base de conhecimento lexical, esse nível de significados proporciona uma ligação explícita com o conhecimento de mundo (senso comum).

Os quatro níveis de estruturação caracterizam o modelo do **léxico gerativo**, motivado pela necessidade de se verificar que tipo de informação lexical contribui para a interpretação semântica das sentenças. O fenômeno de adaptação de expressões lingüísticas a novos contextos não é tão inexplicável; o fato é que dispositivos recursivos subjacentes à semântica de uma sentença dão margem a novos significados.

Esses quatro níveis de representação estão ligados entre si por dispositivos que permitem a interpretação de palavras em contextos diferentes; são eles: restrição de tipos, subseleção e co-composicionalidade.

A **estrutura de argumentos** de um item lexical pode ser vista como a micro especificação de sua semântica, sendo que os argumentos estão divididos em quatro tipos (Pustejovsky, 1994, p.63-64):

- a) argumentos verdadeiros (sintaticamente necessários) - parâmetros de um item lexical necessariamente realizados sintaticamente: p.ex. *John chegou tarde.*
- b) argumentos *default* (padrão ou semanticamente necessários)– parâmetros que fazem parte das expressões lógicas na estrutura *qualia*, mas não são necessariamente realizados sintaticamente: p.ex. *John construiu uma casa com tijolos.*
- c) argumentos *sombra* (argumentos incorporados)– parâmetros incorporados semanticamente no item lexical. Podem ser expressos por operações de subtipos ou de especificação discursiva: p.ex. *Mary e John dançaram uma valsa. (uma dança)*
- d) adjuntos verdadeiros (satélites espaço-temporais) – parâmetros necessariamente expressos na sintaxe que modificam a expressão lógica, mas são parte da interpretação situacional, e não estão ligados à representação de nenhum item lexical em particular; incluem expressões de modificação temporal e espacial: p.ex. *Mary foi para Nova Iorque na terça.*

Desta forma, a estrutura argumental de um dos sentidos do verbo *construir* teria a seguinte representação:

$$\left(\begin{array}{l} \mathbf{construir} \\ \text{estr.arg} \end{array} \left[\begin{array}{l} \text{arg.1 = indivíduo_animado} \\ \text{arg.2 = artefato} \\ \text{D-arg1 = material} \end{array} \right] \right)$$

A **estrutura de eventos** busca representar os eventos que denotam estados, processos e transições. A teoria proposta por Pustejovsky baseia-se no fato de que o significado é altamente estruturado, e não simplesmente uma coleção de traços semânticos. Sendo assim, no caso dos verbos, por exemplo, temos que cada um se

encaixa em um dos três tipos básicos de eventos: estado, processo ou transição (Bach, 1986; Dowty, 1979; Vendler, 1967, *apud* Pustejovsky, 1991, p.56). Verbos de estado denotam um só evento (*estar doente, amar, saber, etc*); verbos de processo expressam uma seqüência de eventos (*correr, empurrar, etc*) e nos verbos de transição, um evento identifica uma expressão semântica, avaliada em relação a suas oposições (*dar, abrir, construir, destruir, etc*) (Pustejovsky, 1991, p. 56).

A **estrutura *qualia*** é um sistema que caracteriza a semântica dos substantivos, muito parecida com a estrutura argumental de um verbo; ela determina o significado da mesma forma que os argumentos de um verbo determinam o seu. Inclui noções familiares como recipiente, espaço, superfície, figura, ou artefato; esses elementos da denotação de um objeto sempre foram cruciais para o nosso entendimento de como as coisas interagem com o mundo.

Resumidamente, a estrutura *qualia* de uma unidade lexical especifica quatro aspectos de seu significado, uma estrutura semântica mínima:

- a relação entre um objeto e suas partes constituintes (papel constitutivo)
- aquilo que o distingue de um domínio maior (papel formal)
- seu propósito e função (papel télico)
- fatores envolvidos em sua origem (papel agentivo)

No exemplo a seguir, *porta* e *lareira* podem ser usados para se referir a um objeto ou à uma abertura (Pustejovsky e Boguraev, 1993, p.205):

- (a) Eles passaram pela porta.
- (b) Ela vai pintar a porta de vermelho.
- (c) A fumaça negra preencheu a lareira.
- (d) A lareira está coberta de fuligem.

No caso de “porta”, a estrutura *qualia* especificaria o propósito ou uso do objeto através do papel télico:

$$\left[\begin{array}{l} \mathbf{porta} (x,y) \\ \text{const} = \mathbf{abertura} (y) \\ \text{formal} = \mathbf{objfísico} (x) \\ \text{telico} = \mathbf{caminhar.através} (P, w, y) \\ \text{agentivo} = \mathbf{artefato} (x) \end{array} \right]$$

A estrutura acima demonstra que uma porta é uma relação entre uma abertura de algum tipo, *y*, e o objeto físico em si, *x*. O papel télico se refere a um evento do tipo **processo** e um indivíduo *w* caminhando através da abertura da porta. Desta forma, como veremos mais abaixo, o sentido apropriado do substantivo é ativado com referência à estrutura *qualia*.

Vejamos um outro exemplo, trabalhado pelos mesmos Pustejovsky e Boguraev, (1993, p.198), no qual é considerado um conjunto de orações:

(1) *The island authorities sent out a fast little government boat, the Culpeper, to welcome us.* (As autoridades da ilha enviaram um pequeno barco rápido do governo, o Culpeper, para nos dar as boas vindas). - existe ambigüidade entre “um barco pilotado rapidamente” e “um barco que é naturalmente mais rápido” (lembrando que a ambigüidade está na língua inglesa).

(2) *Uma datilógrafa rápida* - uma pessoa que desempenha o ato de datilografar rapidamente.

(3) *O tênis é um jogo rápido* - os movimentos envolvidos no jogo são rápidos e imediatos.

(4) *Um livro rápido* - um que pode ser lido em pouco tempo.

(5) *Você pode decidir que um homem será capaz de tomar as decisões rápidas, difíceis.* - um processo que leva um curto período de tempo.

Temos portanto quatro sentidos de *rápido*:

rápido (1) - mover-se rapidamente

rápido (2) - realizar algum ato rapidamente

rápido (3) - fazer algo que requer pouco tempo

rápido (4) - algo que envolve movimentos rápidos

Podemos visualizar *rápido* como sempre predicando o papel télico de um substantivo, já que o papel télico é sempre considerado um evento. Para ilustrar, consideremos a estrutura *qualia* de um substantivo como *carro* (tradução nossa):

$$\left[\begin{array}{l} \mathbf{carro} (x) \\ \text{const} = \{\mathbf{corpo, motor, \dots}\} \\ \text{formal} = \mathbf{objfísico} (x) \\ \text{télico} = \mathbf{dirigir} (P, y, x) \\ \text{agentivo} = \mathbf{artefato} (x) \end{array} \right]$$

Note-se que a telicidade dá conta do propósito e função da do substantivo. Num sintagma como *um carro rápido*, é a relação de “dirigir”, vista como um evento, um processo P, que é modificado pela qualidade de ser rápido. Da mesma forma, para substantivos como *datilógrafa* ou *leitor*, é o papel télico que está sendo interpretado como *rápido*. No caso de *livro*, o papel télico especifica atividades comuns como ler ou escrever. Portanto, as interpretações de *rápido* nos exemplos acima podem ser derivados de um único sentido, e não há necessidade de enumerar os sentidos restantes.

Desta forma, a ambigüidade lexical parece ser resolvida através de regras de composição semântica, ao invés de considerar a seleção lexical como um processo em separado, divorciado do processo interpretativo. E uma consequência dessa visão é a diminuição do tamanho do léxico que se pode considerar.

A **estrutura de herança lexical**, explorada na lexicografia computacional na construção de léxicos específicos para fins de PLN, é uma propriedade pela qual a partir de uma entrada lexical no dicionário é possível montar uma hierarquia de itens relacionados entre si por relações de sentido. Por exemplo, a partir de uma definição do

item *árvore*, pode-se extrair que é uma planta grande, de madeira, perene e que tem um tronco distinto; assim sendo, é possível montar a seguinte hierarquia: *carvalho* @→*árvore*@→*organismo*, por exemplo, em que →@ representa a relação de hiponímia, transitiva e assimétrica, que pode ser lida como “é um tipo de”. Para os substantivos, por exemplo, tais hierarquias proporcionam um “esqueleto conceitual dos nomes” (Miller et al., 1993, p.12).

Na visão de Pustejovsky, um item lexical herda informação de acordo com a estrutura *qualia* que carrega; no entanto, tal representação ainda traz problemas, principalmente para sistema de desambiguação. O mérito está em associar um item lexical a um conceito. Vejamos um exemplo, adaptado de Pustejovsky e Boguraev, 1993, p.214, trad.nossa:

peça de teatro <i>é-um</i> livro		dicionário <i>é-um</i> livro
ler	ok	não
comprar	ok	ok
consultar	não	ok
começar	ok (?)	não

Na estrutura *qualia*, esses itens entrariam com as seguintes especificações:

Livro *é-formalmente* objeto físico
 Livro *é-telicamente* literatura
 Livro *é-agentivamente* literatura
 Dicionário *é-formalmente* livro
 Dicionário *é-telicamente* referência
 Dicionário *é-agentivamente* material compilado
 Peça *é-agentivamente* literatura
 Peça *é-telicamente* livro

Desta forma, ao se referir ao livro como um objeto físico, Pustejovsky associa informação conceitual ao item lexical, o que permite que se extraia uma estrutura *qualia* de herança (*Inh*) tipificada, evitando problemas na constituição de estruturas para línguas com múltipla herança. De acordo com Pustejovsky e Boguraev (1993, p. 216) a estrutura para o item “livro” seria:

[livro]

$$\lambda x [\text{livro}(x) \wedge \text{Formal}(x) = \text{Inh}(\text{obj_fis})$$

$$\wedge \text{Télico}(x) = \text{Inh}(\text{literatura})$$

$$\wedge \text{Agentivo}(x) = \text{Inh}(\text{literatura})].$$

Sendo assim, uma entrada lexical como *construir* poderia ser representada pela junção dos vários níveis descritos acima, assumindo-se que *construir* é um processo que se segue de um estado “de estar construído”; portanto, a entrada ficaria assim representada (Pustejovsky, 1995, p. 81; 1996, p.82):

$$\left(\begin{array}{l} \mathbf{construir} \\ \\ \text{ESTR_EVENT} = \left[\begin{array}{l} \mathbf{evento-transição} \\ \text{EVENTO1} = 1 \left[\begin{array}{l} \text{evento} \\ \text{TIPO} = \text{processo} \end{array} \right] \\ \\ \text{EVENTO2} = 2 \left[\begin{array}{l} \text{evento} \\ \text{TIPO} = \text{estado} \end{array} \right] \\ \\ \text{RESTRICÇÃO} = 1 \leq 2 \end{array} \right] \\ \\ \text{ESTR_ARG} = \left[\begin{array}{l} \mathbf{transitivo} \\ \text{ARG1} = 3 \left[\begin{array}{l} \text{humano} \\ \text{FORMA} = \text{criatura} \end{array} \right] \\ \\ \text{ARG2} = 4 \left[\begin{array}{l} \text{artefato_objeto} \\ \text{CONST} = 5 \\ \text{FORMA} = \text{obj. físico} \\ \text{AGENTE} = \text{artefato} \end{array} \right] \\ \\ \text{D-ARG1} = 6 \left[\begin{array}{l} \text{obj. físico} \\ \text{FORMA} = \text{massa} \end{array} \right] \end{array} \right] \\ \\ \text{QUALIA} = \left[\begin{array}{l} \mathbf{criar - RLS} \\ \text{FORMA} = \text{pred}(2, 4) \\ \text{AGENTE} = \text{ator}(1, 3, 5) \end{array} \right] \end{array} \right)$$

3.2 O modelo de Jackendoff

Enquanto a proposta de Pustejovsky procura dar conta da estrutura da entrada lexical num nível eminentemente lingüístico, Jackendoff (1975 e ss.) elaborou uma teoria para dar conta do nível conceitual da linguagem, buscando as relações entre as unidades lingüísticas e o conhecimento associado a elas, ou o domínio dos conceitos. Em seu modelo, busca captar as dependências conceituais de uma sentença; no exemplo *John colocou o carro na garagem*, propõe a seguinte representação:

[EVENTO	COLOCAR
(PESSOA	JOHN)
(COISA	CARRO)
(LUGAR	EM)
(COISA	GARAGEM)

Trata-se de trabalho semelhante ao de Pustejovsky, pois também descreve uma representação para estruturas conceituais, ou entidades que refletem o conhecimento humano. Fica a ressalva de que a proposta de Jackendoff é cognitivista enquanto a proposta anterior era eminentemente formalista.

Para Verspoor (1997), Jackendoff focaliza primariamente a semântica lexical de maneira sintaticamente relevante, ao mesmo tempo que vê as estruturas conceituais como entidades que carregam consigo conceitos ou pensamentos. Suas análises podem ser usadas para se discutir o problema das regularidades no léxico que têm influência direta na estrutura sintática; ele está preocupado com questões como a decomposição semântica, ontologia semântica, relações de papéis temáticos e funções que relacionam entidades.

Decomposição semântica

É de comum acordo entre vários especialistas (entre eles Pustejovsky) que para se chegar a generalizações sobre as relações entre forma sintática e o significado, é

necessária alguma forma de decomposição lexical. Também se concorda que não é possível decompor um significado em condições necessárias e suficientes para identificar as entidades com as quais uma palavra se relaciona; sempre haverá que se descobrir um elemento adicional do significado que é necessário para distinguir duas palavras que mantêm uma relação arbitrária entre si, por exemplo.

A ontologia semântica

Jackendoff propõe um inventário de conceitos primitivos (uma ontologia) que serve de base para a representação conceitual. Cada constituinte representacional corresponde a uma das categorias ontológicas abaixo (Verspoor, 1997):

Categoria ontológica	Heurística que permite detectar a categoria ontológica	
Coisa (<i>thing</i>)	<i>What did you buy?</i>	O que você comprou?
Evento (<i>event</i>)	<i>What happened next?</i>	O que aconteceu depois?
Estado (<i>state</i>)	<i>What happened was that Max was in Africa.</i>	O que aconteceu foi que Max estava na África.
Ação (<i>action</i>)	<i>What did you do?</i>	O que você fez?
Lugar (<i>place</i>)	<i>Where is my coat?</i>	Onde está meu casaco?
Caminho (<i>path</i>)	<i>Where did they go?</i>	Onde eles foram?
Propriedade (<i>property</i>)	<i>What was she like?</i>	Como ela era?
Modo (<i>manner</i>)	<i>How did you cook the eggs?</i>	De que maneira você preparou os ovos?
Quantidade (<i>amount</i>)	<i>How long was the fish?</i>	Qual era o tamanho do peixe?

Cada constituinte pode ser decomposto numa estrutura do tipo função-argumento. LUGAR, por exemplo, pode ser caracterizado como uma FUNÇÃO-LUGAR (COISA) e um CAMINHO como FUNÇÃO-CAMINHO (LUGAR). Jackendoff procura identificar funções que expliquem padrões gramaticais de combinação, usando as categorias conceituais ontológicas como um ponto de referência. A proposição de tal ontologia proporciona um ponto de partida para a identificação de generalizações sobre como as palavras podem se combinar para formar constituintes maiores. Graças à capacidade criativa da linguagem, fica claro que não se aprendem

simplesmente maneiras específicas de relacionar as palavras, existem padrões de combinação regulares e produtivos que se aplicam a classes de palavras e sentenças.

No entanto, a ontologia proposta por Jackendoff não prevê nenhum refinamento em termos de subcategorias cada vez mais específicas. Verspoor (1997) afirma que a proposta de Jackendoff poderia ser facilmente ampliada ao tomarmos as categorias levantadas por ele como início de uma estrutura hierárquica. Essa “nova” ontologia refletiria categorizações de entidades, eventos, etc. e as inter-relações.

Funções que relacionam entidades

Já foi dito que a identificação de estruturas de função-argumentos é necessária para se captar em relações semânticas entre entidades. Mas Jackendoff também está interessado em distinguir as relações semânticas entre verbos ou usos de verbos que diferem em sua realização sintática ou nos tipos de modificadores que podem aparecer com os verbos.

Verspoor (1997) afirma que, se forem examinados verbos veiculando informação espacial, verbos veiculando movimento de algo ao longo de um caminho e verbos especificando a localização de algo, pode-se encontrar uma distinção consistente entre eles. Ex:

- a) *A mosca voou ao redor da sala.* (CAMINHO)
- b) *O livro está dentro da sala.* (LUGAR)
- c) *A mosca voou dentro da sala* (LUGAR, mas com verbo de movimento)
- d) **O livro está ao redor da sala.* (CAMINHO, com verbo de localização, gerando incompatibilidade.) (Verspoor, 1997)

Diante disso, Jackendoff (1990) propõe dois tipos de representação:

-para verbos de movimento: [evento IR ([COISA];[CAMINHO])]

-para verbos de localização: [estado ESTAR ([COISA]; [LUGAR])]

O exemplo ilustra como a identificação de distinções na estrutura de função-argumento pode auxiliar no mapeamento sintático-semântico e na modelagem de interações apropriadas entre verbos e seus argumentos.

Um outro exemplo:

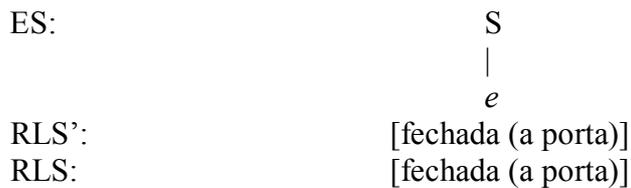
A boneca pertence a Beth. [SER poss. (a boneca; [para (Beth)])]	Beth recebeu a boneca. [IR poss. (a boneca; [para (Beth)])]
--	--

A partir da primeira representação, infere-se que a boneca é possuída por Beth; na segunda, que a boneca é transferida para Beth, que depois dessa transferência passa a possuí-la (é uma analogia com o processo de percorrer um caminho para chegar até Beth). Essas e outras extensões das funções espaciais sugerem a utilidade das funções identificadas por Jackendoff e refletem regularidades no uso dos componentes de uma língua através de domínios diferentes. Um ponto essencial a se observar aqui é que a teoria de extensibilidade de funções espaciais para outros domínios demonstra que muitos campos semânticos têm essencialmente a mesma estrutura, e que o domínio espacial define os termos em que muitos tipos de discursos serão estruturados. (Verspoor, 1997).

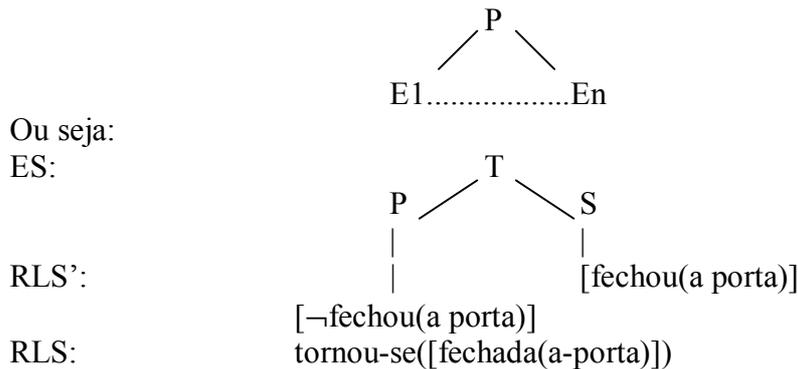
3.3 Relacionando os dois pontos de vista: lexical e conceitual

Pustejovsky e Boguraev (1993, p. 57-58) propõem uma comparação entre a sua própria proposta e a de Jackendoff. Para tanto, referem-se à proposta desse último como RLS (representação léxico-semântica) e à sua, RLS'. Sendo assim, as estruturas para ilustrar exemplos de estado, processo e transição ficariam da seguinte forma:

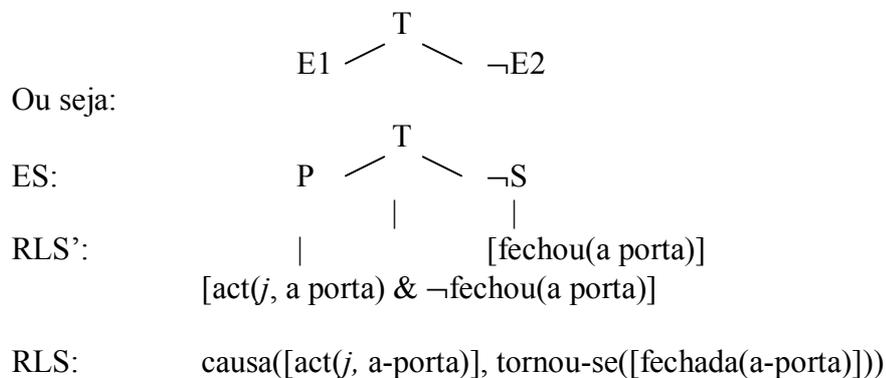
a) estado (S) – evento simples. Ex: *A porta está fechada.*



b) processos (P) – seqüência de eventos. Ex: *A porta fechou.*



c) transição (T) – um evento identificando uma expressão semântica, que é medido em relação à uma oposição (oposição entre dois eventos). Ex: *John fechou a porta.*



Ambas as estruturas sugerem que “fechar”, tanto em sua forma incoativa (*A porta fechou*) como na forma causativa (*John fechou a porta*) há uma transição de um estado para seu oposto. A representação RLS' difere apenas na especificação de uma ação sendo realizada para os casos causativos, ou seja, a expressão & indica simultaneidade de duas expressões dentro de um subevento. Embora o verbo seja ambíguo no nível RLS, os sentidos lógicos são capturados em ES ou RLS' (Pustejovsky, 1991, p. 58).

o modelo de representação de estruturas conceituais revela-se interessante para esse trabalho, pois a partir da base lexical do *thesaurus* pode ser feita uma indexação de conceitos a sentidos, conferindo à base lexical o estatus de uma ontologia de sentidos e conceitos para a língua portuguesa. Observe-se que esta ontologia seria mais completa na medida em que mais relações de sentido (hiperonímia, meronímia, troponímia, etc) fossem acrescentadas à base, dado que esta será construída apenas em função da sinonímia e antonímia.

CAPÍTULO 4 - BASES DE DADOS LEXICAIS

4.1 Métodos de compilação de informação léxico-semântica

O desenvolvimento de léxicos para fins computacionais teve origem na busca do desenvolvimento de técnicas que lidassem com a semântica envolvida na elaboração de ferramentas para tradução automática (IBM, 1959, *apud* Grefenstette, 1994, p.07); percebeu-se que era necessário que o significado de cada unidade e seu contexto de ocorrência fossem levados em conta. A história da busca pela classificação semântica passa pelo interesse dessa mesma comunidade de especialistas em buscar formas de se implementar versões digitais dos esquemas de classificação humanos.

Assim sendo, três métodos aparecem como alternativa para a extração de dados semânticos para fins computacionais:

- 1) construir um léxico a partir de *corpora*. Ex: sistema SEXTANT (Grefenstette, 1994). Tal sistema foi construído a partir de um *corpus* eletrônico, processado automaticamente; no entanto, pode-se construir um léxico manualmente, coletando ocorrências de *corpus* escrito um a um, anotando-os. O léxico do SEXTANT foi compilado com o objetivo de construir um dispositivo capaz de capturar as similaridades entre palavras de um *corpus*. A hipótese subjacente é de que unidades utilizadas de maneira semelhante em um *corpus* são semanticamente semelhantes. Após a análise gramatical (*parsing*) e extração de contexto sintático, tem-se para cada palavra do *corpus* um conjunto de “pistas” indicativas de significados. Por exemplo, sabe-se quais adjetivos modificam determinados nomes e verbos dos quais o substantivo é sujeito ou objeto. Considerando essas “pistas” como atributos das palavras, medidas de similaridade podem ser feitas; desta forma, o SEXTANT coleta esses atributos

e os usa para determinar se duas palavras estão sendo usadas de maneira similar no *corpus* (Grefenstette, 1994, p. 34).

- 2) extrair informação lexical de dicionários indexados/codificados automaticamente. Ex: o LDOCE² (Boguraev e Briscoe, 1989).

Todas as entradas do LDOCE (versão digital) estão codificadas e armazenadas sob a forma de registros, cada um em uma linha; cada linha começa com uma seqüência de dígitos, dos quais os seis primeiros referem-se a uma seqüência única e os outros dois codificam o identificador do registro. Isso é seguido de um número de campos específicos daquele registro; o caracter “<” é um separador de campos, conforme a figura 7 abaixo (Boguraev e Briscoe, 1989, p. 46).

```
RIVET2 v. 1 [T1; X9] to cause to fasten with rivets1...
28289801 < R0154300 < rivet
28289902 < 02 < <
28290005 < v <
28290107 < 0100 < T1; X9 < NAZV < HXS
28290208 < to cause to fasten with
28290318 < { *CA } RIVET { CB } { *46 } s { *44 } { *8A } :
```

Fig. 7. Fragmento da base do LDOCE.

Desta forma, 01 refere-se à entrada, que é então dividida em dois campos: número serial e entrada principal (aqui pode-se adicionar mais informação como silabação, variantes ortográficas e capitalização). Da mesma forma, 02 codifica a informação homográfica: número homográfico e informação adicional sobre a segmentação e padrões de acentuação em entradas compostas. O identificador para o código de definição “07” identifica um registro com quatro campos: número de sentidos, código gramatical, sujeito e código de entrada (Boguraev e Briscoe, 1989, p. 46). Assim, o léxico do LDOCE constitui-se um “Machine Readable Lexicon” ou léxico manipulável por computador (Handke, 1995, p. 223), adequado ao desenvolvimento de léxicos para fins de PLN.

- 3) extrair informação lexical manualmente, a partir de dicionários impressos da língua. Ex: o *Dicionário de Sinônimos* eletrônico do pacote MS Word-97, construído a partir da obra homônima de Antenor Nascentes, de 1981. A extração semântica via dicionários impressos tem origem no trabalho de Sparck Jones (1960, *apud* Grefenstette, 1994, p. 18), que reduziu cada sentido do verbete do dicionário aos principais nomes que aparecem na definição; desta forma, uma entrada como

Task3 a) a piece of work which has to be done; something that one has to do (usually involving labour or difficulty); a matter of difficulty. "He had taken upon himself a task beyond the ordinary strength of man."

é reduzida à seqüência *task – labour – work*, na qual os termos são considerados sinônimos. Semelhante princípio foi utilizado na construção da rede *WordNet* (Miller; Fellbaum, 1991).

Para a construção da base do *thesaurus eletrônico*, é necessário fazer uma opção. A primeira, de coleta a partir de *corpus*, deve ser avaliada, posto que exige muito tempo e equipe de trabalho numerosa; a segunda proposta, de extração de bases já digitalizadas, só é viável se houver bases disponíveis no formato digital para a língua alvo, o que não é o caso do português do Brasil. A última opção parece ser a mais operacional e factível, mesmo se deparamos com escassez de recursos materiais e humanos para o desenvolvimento da base lexical. Essa opção nos remete ao domínio da lexicografia.

4.2 Importância dos dicionário enquanto fontes de informação léxico-semântica

² *Longman Dictionary of Contemporary English*.

Não é possível estabelecer o número total das palavras em uso em um determinado idioma; além disso, qualquer tentativa de “contar” esse número é vã, pois o léxico é um sistema em aberto e em expansão (Biderman, 1978, p.151; Lyons, 1981, p. 146), com unidades entrando e outras saindo, num fluxo contínuo que impede qualquer tipo de delimitação.

No entanto, é possível delimitar o léxico de alguma forma; verifica-se que praticamente todas as línguas escritas possuem estoques de unidades lexicais: os dicionários. O dicionário é, portanto, uma fonte primária de informação lexical, uma vez que registra em seus verbetes a parcela da cultura de um povo estampada no léxico de sua língua; no verbete são registrados os mais variados tipos de descrição, de acordo com a natureza da obra de referência³ e os objetivos do consulente, que vão desde a tradução, busca de sinônimos e antônimos até a adequação estilística, etc.

Enquanto a ciência que estuda a construção dos dicionários, a **lexicografia**, conta com um conjunto de métodos e técnicas voltados para a descrição do estoque vocabular de um povo:

A lexicografia (e seu praticante, o lexicógrafo) está portanto preocupada com a tarefa de descrever todas ou algumas das palavras de uma ou mais línguas em termos de seus traços característicos, notadamente de seu significado (Hartmann, 1983, p. 4, trad. nossa).

A estrutura da entrada lexicográfica apresenta as seguintes características (Vilela, 1983, p. 08):

Entrada + informação (etimológica / ortográfica / fonética / gramatical) + definição (ou explicação) + exemplos (ou aplicação da entrada em contextos)

³Uma breve tipologia de dicionários pode ser arrolada: dicionários etimológico, de gírias, neologismos, pronúncia, sinônimos, dialetais, idiomáticos, de acrônimos, abreviações, bilíngues, monolíngues, dicionários no formato “duden” (dicionário bilíngue arranjado semanticamente, com ilustrações de página inteira), *crossword puzzle dictionaries* (arranjados pelo número de letras das palavras), livros de frequência e dicionários de rima (arranjados de acordo com o som da última sílaba), entre outros (Landau, 1996, p. 5-34).

A descrição do significado de uma entrada é tarefa árdua para o lexicógrafo, dependendo da natureza da obra de referência (se é uma obra bilíngüe, monolíngüe, etc.). É exatamente na definição da estrutura gramatical, que é parte da descrição de uma entrada, que um léxico difere muito de uma entrada de dicionário. Dependendo da descrição, um usuário pode compreendê-la com maior ou menor acuidade; já os modelos de formalização do significado dependem de primitivos que lhes permitam reconstruir um significado num sistema processador de língua (Handke, 1995, p. 65). Sendo assim, a maneira pela qual os significados devem ser representados por uma teoria lexical depende da teoria que lhe serve de suporte. Nesse domínio distinguem-se dois tipos: teorias construtivas e teorias diferenciais.

Numa teoria **construtiva**, a representação deve conter informação suficiente para dar conta de uma acurada construção do conceito (tanto pelo homem como por uma máquina). Tarefa difícil de ser executada; por exemplo, o modo pelo qual o conceito de “mesa” é especificado num dicionário (“peça de mobília que consiste em uma tábua fixa em quatro pernas”) só é acessível a quem já saiba o que é uma mesa.

Por outro lado, numa teoria **diferencial**, não se especificam os conceitos (Haensch, 1982, p. 278; Hartmann, 1983, p. 90; Miller; Fellbaum, 1991, p. 200) e postula-se que o significado pode ser representado por um símbolo capaz de ativar o conceito a ele associado na mente do falante (Miller; Fellbaum, 1991, p.200). As exigências de uma teoria diferencial são mais modestas, porém suficientes para a construção dos mapeamentos necessários. Se uma pessoa que lê uma definição precisa meramente identificar um conceito familiar, então um sinônimo (ou quase sinônimo) é geralmente suficiente:

Echando mano de la definición por sinónimos, se le puede dar al usuario una instrucción útil para el uso o la comprensión de una unidad léxica, siempre que quede asegurado – por otros factores dentro o fuera de la definición lexicográfica (éstos pueden ser de índole muy trivial) – que los sinónimos indicados sirvan sólo como

punto de partida para la delimitación del contenido de una unidad léxica y que no se pretenda que los sememas o contenidos de una unidad léxica se identifiquen exclusivamente mediante la indicación de sinónimos (Haensch, 1982, p. 279).

A **definição** é em si mesma um traço característico da linguagem, uma vez que a utilização da metalinguagem é um traço universal, já que toda língua possui uma maneira de definir o que é um “x”. O dicionário indica na sua definição o que distingue uma unidade ou expressão de outras unidades ou expressões; seu objetivo é solucionar problemas de compreensão e produção de textos ou na resolução de problemas de generalização (procura de um arquilexema ou hiperônimo), de especificação (hipônimo), de matização semântica (sinônimos) ou polarização sêmica (antônimos), ou ainda auxiliar nas tarefas de tradução de uma língua para outra. Comum a qualquer definição, é a noção aristotélica do “*definitio fit per genus proximum et differentiam specificam*” (“gênero próximo e diferença específica”), que equivale a dizer que a definição classifica dentro de uma determinada categoria o definido e o distingue dos outros membros da mesma categoria; é a chamada definição clássica. Greimas (1967, p. 72) coloca a definição como uma paráfrase ou expansão, no sentido de uma unidade de texto que engloba um enunciado, cujos elementos são os definidores e o definido. De qualquer forma, a definição deve seguir algumas regras básicas:

- a) deverá estabelecer uma relação entre o geral e o individual;
- b) não deverá ser formulada negativamente se puder ser formulada positivamente;
- c) não poderá ser circular.

Há que se observar ainda a categoria gramatical, a linguagem utilizada na definição, a equivalência e/ou substituição entre a entrada e a definição. Quanto aos elementos lingüísticos utilizados para definir, deverão ser mais freqüentes do que o termo definido (exigência que pode ser cumprida com o auxílio da lingüística de

corpus), e, no que diz respeito à capacidade de substituição entre definido-definição, isso só poderá ser feito a partir do pressuposto de que o leitor preencha as lacunas próprias da economia lexicográfica e que o lexicólogo forneça a chave dessa economia (Vilela, 1983, p. 08-11).

Para ilustrar a complexidade da informação contida na entrada do dicionário de língua portuguesa (em oposição às obras analógicas), tomemos alguns exemplos:

Entrada	Aurélio do séc. XXI	Michaelis
Formiga	<p>Formiga [Do lat. formica, por via popular.] S. f. 1. Zool. Designação comum a todos os insetos himenópteros, formicídeos, caracterizados por terem o hipopígio do macho em espinho voltado para cima. As fêmeas são dimórficas, as operárias ápteras, com suturas torácicas ausentes ou muito reduzidas, e as formas férteis com suturas presentes. Vivem em colônias. 2. Fig. Pessoa econômica e/ou trabalhadeira. 3. Fam. Pessoa que gosta muito de alimentos doces.</p> <p>* Como formiga. 1. Em grande quantidade: Havia no comício gente como formiga.</p>	<p>for.mi.ga s. f. Entom. Nome genérico dos insetos himenópteros que vivem em sociedade, debaixo da terra, em ninhos sobre árvores, no oco dos paus etc.</p>

Sumário da estrutura da entrada do lexema FORMIGA no dicionário (cf. Handke, 1995, p. 66):

Síntese da estrutura do lexema	Aurélio	Michaelis
Unidade de acesso	FORMIGA	FORMIGA
Gramática	s.f. (substantivo feminino)	s.f. (substantivo feminino)
História	[do lat. Formica, por via popular]	-
Significados	<p>1. Zool. Designação comum a todos os insetos himenópteros, formicídeos, caracterizados por terem o hipopígio do macho em espinho voltado para cima. As fêmeas são dimórficas, as operárias ápteras, com suturas torácicas ausentes ou muito reduzidas, e as formas férteis com suturas presentes. Vivem em colônias. 2. Fig. Pessoa econômica e/ou trabalhadeira. 3. Fam. Pessoa que gosta muito de alimentos doces.</p>	<p>Entom. Nome genérico dos insetos himenópteros que vivem em sociedade, debaixo da terra, em ninhos sobre árvores, no oco dos paus etc.</p>

Síntese da estrutura do verbete:

Estrutura da entrada lexicográfica	Aurélio	Michaelis
Entrada	<i>formiga</i>	<i>formiga</i>
Informação gramatical	s.f.	s.f.
Rótulos	Zoologia (zool.), sentido figurado (fig.), fam.	Entomologia (Entom.)
Especificação de significados	3 acepções numeradas	Uma acepção, não numerada
Expressões relacionadas	* Como formiga. 1. Em grande quantidade: Havia no comício gente como formiga.	-
Exemplos	“Havia no comício gente como formiga.”	-

Desta forma, a entrada do dicionário busca dar conta desde informações de cunho gramatical, passando pelo domínio semântico e em alguns casos, identificando seu uso (dimensão pragmática). As definições nem sempre obedecem à proposta aristotélica de definição de objetos, ou seja, nem sempre há uma especificação seguida de diferenciações. No exemplo da entrada *formiga*, por exemplo, o Dicionário Aurélio do Séc. XXI oferece na primeira acepção *1. Zool. Designação comum a todos os insetos himenópteros, formicídeos, caracterizados por terem o hipopígio do macho em espinho voltado para cima. As fêmeas são dimórficas, as operárias ápteras, com suturas torácicas ausentes ou muito reduzidas, e as formas férteis com suturas presentes. Vivem em colônias.* (grifo nosso). Percebe-se primeiro uma generalização (“todos os insetos himenópteros”), seguida das devidas especificações de cada um dos tipos do definido; temos portanto a definição clássica, aristotélica.

Agora observemos um outro exemplo, entrada *cara*:

Aurélio	Michaelis
[Do gr. Kára, ‘cabeça’, pelo lat. tard. cara, poss.] s.f. 1. a parte anterior da cabeça; rosto. 2. semblante, fisionomia. 3. a parte oposta à coroa, geralmente com uma efígie, em certas moedas.	Ca.ra s.f. 1. Parte anterior da cabeça; rosto, face. 2. Expressão do rosto; fisionomia, semblante. 3. Audácia, atrevimento, ousadia. 4. Face da moeda, oposta à coroa, geralmente com uma efígie. S.m. gír. Tratamento incivil que se dá a uma pessoa; indivíduo.

<p>4.fig. Aspecto, aparência, ar: O doente está de boa cara; O bife está com boa cara. 5.Ousadia, coragem: Não tem cara de vir até aqui. s.m. Bras. Gir. 6.pessoa que não se conhece. 7.indivíduo, sujeito: “já era de noite e eu estava no posto dois com esse cara chamado Fabinho” (Rubem Fonseca, A Coleira do Cão, p.168).</p>	
---	--

Neste caso, muitas das acepções da entrada *cara* não foram feitas aos moldes da definição clássica, mas ao contrário, utilizaram-se sinônimos para oferecer um sentido; por exemplo, a partir do dicionário Aurélio do Séc. XXI temos os sentidos:

cara, rosto;
cara, semblante, fisionomia;
cara, aspecto, aparência, ar;
cara, ousadia, coragem;
cara, indivíduo, sujeito.

Note-se que o sentido não foi dado por definições, e sim por sinônimos que constituem um átomo conceitual. Portanto, a estrutura das definições dos dicionários de língua portuguesa revela-se apropriada para os propósitos de construção da base lexical do *thesaurus*, em termos da terceira estratégia mencionada na página 85, posto que **oferece condições para a compilação de grande número de sinônimos e antônimos em termos de conjuntos.**

Sendo assim, há uma opção de montagem da base lexical do *thesaurus* de maneira semi-automática, elegendo-se dicionários do português como *corpus* de referência para extração de dados e posterior inserção na base lexical. Essas fontes serão arroladas no capítulo referente à implementação da base lexical do *thesaurus* (capítulo 4).

4.2.1 O significado no dicionário: problemas

Kilgarriff (1993), em busca dos tipos de **distinções** entre as acepções de um mesmo verbete, realizou um levantamento de como as distinções de significados são

feitas no *Longmans Dictionary of Contemporary English* (LDOCE), ou seja, ele buscava descobrir como e porque o lexicógrafo decidia se uma palavra teria um só significado ou dois, por exemplo, ou mesmo que parâmetros o lexicógrafo aplica na distinção entre os sentidos dentro de um mesmo verbete. A pesquisa foi realizada através de dois procedimentos. O primeiro desenvolveu um esquema de classificação para descrever os tipos de distinções que ocorrem mais comumente; o segundo examina a tarefa de cruzar os usos de uma palavra vindos de um *corpus* com os significados que o dicionário fornecia.

Para o primeiro estudo, esse autor procedeu a um levantamento das informações que um verbete de dicionário traz. Encontrou oito:

1. entradas diferentes (ex.: banco 1 e banco 2);
2. sentidos numerados dentro de uma entrada (acepções);
3. subdivisões das entradas numeradas, marcadas com a, b, c, etc.;
4. informações em parênteses (opcional);
5. distinção de uso figurado (“fig.”);
6. a definição principal contendo uma disjunção (ex.: melodia 1 canção ou cantiga....)
7. disjunção no código gramatical (se um nome é contável ou incontável, verbo transitivo ou intransitivo, etc.);
8. parênteses indicando uma disjunção (ex: meigo – de pessoa ou comportamento).

Ainda dentro do verbete, uma definição de sentido numerada pode ou não apresentar o seguinte:

1. código sintático
2. proposições subcategorizadas

3. texto de definição com:

- a) estrutura sintática
- b) conteúdo semântico
- c) restrição no escopo de aplicação
- d) especificação de um uso prototípico

4. exemplos com:

- a) estrutura sintática
- b) conteúdo semântico
- c) colocação (às vezes em negrito)

Ao analisar as definições, Kilgarriff isolou três tipos de problemas (cf. Dias-da-Silva e Moraes, 2000, p. 05):

1. **necessidade de traços**: se um traço é condição necessária para a palavra ser usada em determinado sentido, ou se é apenas fictício;

2. **consistência das informações do dicionário**: se a informação é consistente em diversos aspectos. Por exemplo, há uma entrada com dois sentidos diferentes, cada uma com seus traços característicos; mas ao se verificar os exemplos de suas ocorrências, percebe-se a existência de traços comuns entre os sentidos 1 e 2. Desta forma, a distinção do dicionário é inconsistente com as citações.

3. **Centralidade do sentido**: com a definição dada para a palavra usada naquele sentido, quanto de variação é aceitável para que tal palavra continue sendo classificada naquela acepção, ou seja, o quanto o significado pode se afastar do exemplo de ocorrência dado no verbete.

Conclui:

O experimentador identifica um sentido ao contrastar as especificações deste sentido com as especificações dos outros sentidos. (...) Na maioria dos casos, os contrastes podem começar a responder questões, mas é necessário evidência das citações para lembrar o experimentador das distinções que o lexicógrafo queria

captar, e determinar em que ponto os grupos de traços do lexicógrafo parecem não fazer sentido (Kilgarriff, 1993, p.372 – trad. nossa).

Para finalizar, Kilgarriff propôs cinco categorias para as distinções de sentido encontradas nos verbetes do LDOCE, as quais podem ser aplicadas na análise de outros dicionários. Exemplifiquemos cada uma das categorias através de verbetes extraídos do *Dicionário Aurélio do Século XXI*:

- Metáfora generalizante – é a distinção entre a palavra exata para uma determinada situação e um outro sentido, mais amplo, que pode ser aplicado a um número maior de situações. Em outras palavras, a metáfora generalizante é uma espécie de hiperônimo de outros sentidos, estando portanto presente em mais de uma acepção; a denotação de um sentido específico é um subconjunto de uma denotação de sentido geral.

chefe

1. O principal entre outros.
2. Aquele que exerce autoridade; que chefia, dirige: o chefe do governo.
3. O dirigente, o diretor, o patrão: chefe de uma firma.
4. Aquele que comanda ou governa: o chefe das tropas.

Todas as acepções, na realidade, carregam em si a acepção número 1; a distinção acaba sendo feita apenas através dos exemplos. Vejamos um outro caso:

maltratar

1. Tratar com violência; infligir maus-tratos a; bater em; espancar: Não se devem maltratar os animais.
2. Lesar fisicamente; mutilar.
3. Tratar com palavras rudes; tratar mal; receber mal.
4. Insultar, ultrajar, vexar.
5. Danificar, estragar, arruinar: As crianças maltratam qualquer objeto.
6. Bater, açoitar.
7. Causar dano(s) ou prejuízo(s) a.

Neste caso, a acepção número 7 se encaixa na idéia de metáfora generalizante, uma vez que está presente em todas as outras acepções, embora implicitamente (por exemplo, “tratar com violência” (01) causa danos (07), etc.

- “Must-be-there’s” (“devem estar lá”) – distinções em que, se a situação é tal que um sentido pode ser aplicado, então é uma consequência lógica que o outro também possa ser aplicado, embora a um outro aspecto da mesma realidade.

Exemplo:

casamento

1. Ato solene de união entre duas pessoas de sexos diferentes, capazes e habilitadas, com legitimação religiosa e/ou civil. [Sin.: matrimônio, enlace matrimonial, consórcio, (fam.) banho-de-igreja, e (pop.) casório.]
2. Cerimônia em que é celebrada essa união; núpcias, esponsais, boda(s), e (fam.) banho-de-igreja.
3. Fig. Aliança, união.
4. Fig. Combinação, harmonia.

“Casamento” é tanto um “ato” quanto uma “cerimônia”; um mesmo termo aplica-se a uma determinada situação e a outra, embora em um aspecto da mesma realidade.

- Mudança de domínio – nesse caso, os sentidos diferentes que uma palavra pode assumir estão tão distantes entre si que o lexicógrafo decidiu abrir dois sentidos distintos.

Exemplo:

Avião

[Do fr. avion.]

S. m.

1. Aer. Aeródino dotado de meios próprios de locomoção, e cuja sustentação se faz por meio de asas; aeroplano: O inventor do avião foi o brasileiro Santos Dumont.
2. Bras. Moça ou rapaz bonito.
3. Bras. Gír. Aquele que vai buscar droga para o usuário; vapor, vapozeiro.

No caso deste verbete, a distinção entre as acepções passa do denotativo para valor conotativo, no caso da relação entre 01 (denotativo) e 02 (conotativo, com rótulo de “brasileirismo”); o mesmo se dá entre 01 e 03. Em todo o caso, os sentidos estão bem distantes entre si, fazendo com o lexicógrafo optasse por abrir novas acepções.

- De tipo natural e social – nessa distinção, as entidades ou situações identificadas pelos diferentes sentidos possuem denotações distintas; embora as denotações tenham atributos em comum, sempre estarão em classes diferentes. Nesse caso, a definição do dicionário é mais um indicativo de classes de entidades.

Exemplo:

Crista

s.f.

1. Zool. Excrescência carnosa existente na cabeça dos galos e doutros galináceos.
2. Zool. A parte terminal do abdome dos pássaros, entre as coxas e a cauda.
3. Zool. Saliência no alto da cabeça de certos peixes e répteis.
4. Zool. Penacho, poupa.
5. Bot. Lígula da folha das palmeiras.
- (...)

No caso, a distinção foi dada através do uso de rótulos para as classes diferentes, zoologia (reino animal) e botânica (reino vegetal).

- Relações de tipos – relações entre as categorias das unidades. Por exemplo, num domínio de verbos, adjetivos e substantivos servem de distinção de tipo.

Exemplo:

cantar

v.t.d.

1. Dizer ou exprimir por meio do canto² (1): Cantava baladas para adormecer o filho. (...)
- (...)

(...)

s.m.

12. v. cantiga. (2)

O cantar do galo.

1. O amanhecer.

A distinção foi feita baseada na diferença de classes, verbo (*cantar=dizer*) e substantivo (*o cantar*).

O segundo estudo de Kilgarriff envolveu o uso de *corpus*, para cruzar seus usos com os sentidos dados no dicionário; para cada uma das citações selecionadas, um cruzamento foi feito para verificar se as ocorrências dos lexemas se encaixavam em um ou vários sentidos que o LDOCE dava para a palavra. Sem nos determos em maiores detalhes da realização do experimento, vejamos quais as possibilidades de resultado encontrados por Kilgarriff (1993, p.376):

1. exatamente um sentido se encaixou;
2. mais de um sentido estava próximo do sentido da citação;
3. dois ou mais sentidos não-exclusivos se aplicavam, trazendo contribuições diferentes para o significado da citação;
4. um uso foi indeterminado entre dois sentidos, especialmente os vocábulos não-referenciais;
5. um uso não se encaixou com nenhum dos sentidos, talvez porque era um uso figurativo incomum, ou raro, ou mesmo porque o lexicógrafo deixou de incluir algo em sua definição;
6. parece que a palavra estava sendo usada em um e apenas um sentido, mas não havia contexto suficiente para confirmar.

Em vista do que constatou, o autor fez algumas observações: primeiro que os dicionários não são feitos com vistas a encaixar todos os usos das palavras com os sentidos dados pelo próprio dicionário; e segundo, que esse fato não sirva de crítica negativa ao trabalho dos lexicógrafos, porque não há uma razão prática ou teórica que justifique porque isso seria possível (p. 378).

A conclusão a que se chega depois dos dois estudos é que revisar a tarefa já realizada pela tradição lexicográfica não é simples, embora se revele útil e produtiva.

Kilgarriff afirma:

A evidência tirada de ambos [os estudos] é que as distinções de sentido formam um conjunto altamente heterogêneo. (...) Às vezes dois sentidos de uma palavra são mutuamente exclusivos, mas geralmente não são, e para alguns usos, ambos os sentidos contribuem com elementos diferentes para o sentido. Existem usos em que uma distinção é irrelevante, e o contexto simplesmente não especifica nem um sentido nem o outro. Às vezes a chave para uma distinção está na semântica, em outras, na sintaxe, colocação ou pragmática. (...) Palavras diferentes, freqüentemente, não pertencem ao escopo de uma mesma dimensão, mas definem suas próprias dimensões, portanto o lexicógrafo precisa expressar a dimensão de variação antes de estabelecer uma palavra dentro de um escopo (Kilgarriff, 1993, p.381)

Em outras palavras, a tarefa de se extrair unidades dos dicionários é difícil, laboriosa, mas possível, desde que sejam observadas algumas condições, conforme veremos no capítulo quatro.

4.3 Motivação para a construção da base lexical do thesaurus: a rede WordNet

O centro dos estudos da **lexicografia computacional** está em verificar a viabilidade de se converter um conjunto de unidades lexicais de um determinado dicionário em módulos lexicais que possam ser utilizados de maneira automática por programas computacionais, estando portanto ao dispor das necessidades do sistema em desenvolvimento (Boguraev; Briscoe, 1991, p. 02); a pesquisa em tal campo tem obtido progressos significativos, convertendo os itens de um dicionário impresso em bases lexicais digitais, estruturadas, sistemáticas, contendo informações associadas a cada item lexical, especialmente sintáticas, em módulos para uso direto por sistemas em PLN (Kilgarriff, 1993, p. 367).

Já a **psicolexicologia** parte de pressupostos cognitivos em busca de modelos de representação da memória lexical, mas nem por isso deixa de se nortear por princípios da lexicografia tradicional. Segundo Miller e Fellbaum (1991, p. 198) os métodos lexicográficos são construtivos, pois buscam construir um produto, um dicionário; e um meio de adaptá-los aos métodos psicolinguísticos é utilizá-los na construção de modelos de memória lexical.

Assim sendo, em 1985 uma equipe de pesquisadores da Universidade de Princeton, EUA, começou a construir uma base lexical utilizando-se de várias hipóteses da psicolinguística, especialmente as que envolvem a representação do conhecimento. A base resultou em um *software*, uma representação *on-line* do léxico de língua inglesa chamada **WordNet**⁴, contendo mais de 54.000 diferentes entradas lexicais organizadas em mais de 48.000 conjuntos de sinônimos, construído com base nas noções de redes semânticas, conjuntos de sinônimos e de matriz lexical. Tal projeto foi viável porque os pesquisadores não descartaram o uso de métodos lexicográficos, posto que esses são construtivos, ou seja, pretendem resultar em dicionários. Uma maneira de adaptar tais métodos a hipóteses psicológicas de interesse aqui foi usá-los para construir um modelo de memória lexical, resultando assim em um produto que é fruto dos estudos da lexicografia computacional e psicolexicologia.

4.3.1 Fundamentos da rede *Wordnet*

Uma hipótese inicial para a consecução de um modelo como a *WordNet* foi considerar que diferentes categorias sintáticas de palavras podem ter diferentes tipos de mapeamento (Miller e Fellbaum, 1991, p.199).

⁴ <http://www.cogsci.princeton.edu/~wn/>

Uma pessoa que conheça uma palavra não percebe sua forma e seu significado de maneira separada, mas como dois aspectos de sua unidade fenomenológica. Uma maneira de representar essa distinção é através de um esquema de setas: de um lado, as formas; de outro, os significados, com setas indo e vindo entre elas:

<i>árvore</i>	→	<i>vegetal lenhoso cujo caule,</i>
	←	<i>chamado tronco, só se ramifica bem acima do nível do solo, ao contrário do arbusto, que exhibe ramos desde junto ao solo.</i>

Entretanto, tal esquema é muito abstrato para a construção de um modelo como a *WordNet*. O fato é que as associações entre formas e significados são do tipo *muitos:muitos* (leia-se “muitos para muitos”) - algumas formas têm vários significados diferentes, e alguns significados podem ser expressos de diversas formas. Sem esse mapeamento *muitos:muitos* seria impossível discutir as relações entre dois problemas essenciais: polissemia e sinonímia.

Polissemia e sinonímia são problemas que surgem no caminho de acesso ao léxico mental: um ouvinte ou leitor que reconhece uma forma tem que lidar com sua polissemia; um falante ou escritor que deseje expressar certo significado tem que decidir entre um conjunto de sinônimos. A polissemia é uma fonte de ambigüidades que pode conduzir à obtenção de informação irrelevante; a sinonímia é uma fonte de terminologias alternativas que pode resultar em falhas na busca de informação relevante.

Pela semântica lexical, assume-se que uma palavra é uma associação convencional entre um conceito lexicalizado e uma expressão que desempenha um papel sintático. Essa definição de palavra pode se referir tanto à expressão física quanto ao significado; portanto, o ponto de partida para a semântica lexical pode ser o mapeamento entre formas e significados (Miller *et al*, 1993, p. 04).

Para simular o mapeamento *muitos:muitos* proposto pela noção de matriz lexical, é preciso descobrir alguma forma de representar tanto forma como significado. Na falta de uma teoria psicológica adequada, os métodos desenvolvidos pelos lexicógrafos proporcionaram uma solução para o momento: as definições podem desempenhar o mesmo papel numa simulação que os significados desempenham na mente do usuário da língua. Sendo assim, a maneira pela qual os significados devem ser representados por uma teoria lexical depende do tipo da teoria: Conforme visto anteriormente, numa teoria **construtiva** há falhas nas definições: por exemplo, um usuário de dicionário somente compreenderá o verbete “*mesa*” se de fato já tiver um conhecimento do objeto. Por outro lado, numa teoria **diferencial**, assume-se que o usuário já tenha o conceito (Haensch, 1982, p. 278; Hartmann, 1983, p. 90; Miller; Fellbaum, 1991, p. 200) e que os significados podem ser representados por quaisquer símbolos que tornem possível distingui-los (Miller e Fellbaum, 1991, p.200).

Desse modo, a metodologia de análise semântica do lexicógrafo é basicamente **contrastiva (diferencial)**; a fim de isolar, é essencial que primeiro se compare, que se arranje uma série de itens lexicais que sejam sinônimos em um determinado espectro de traços semânticos, que estes se combinem em vários aspectos de modo que as áreas de contraste sejam colocadas em destaque (Hartmann, 1982, p. 90). Por exemplo, uma pessoa que saiba que *abalo* pode ser uma *sacudidela* ou significar *comoção*, o conjunto de sinônimos *{abalo, sacudidela}* e *{abalo, comoção}* servem como designadores não ambíguos dos dois significados de *abalo* (cf. Miller e Fellbaum, 1991, p.201). Esses conjuntos não explicam o conceito: **eles indicam que um dado conceito existe.**

Uma maneira de se confirmar a hipótese da teoria diferencial é a verificação de unidades e sentidos em termos de uma **matriz lexical**; portanto, tal estrutura pode ser representada para fins teóricos como um mapeamento entre palavras escritas e os

conjuntos de sinônimos aos quais elas pertençam, conforme a figura 8 abaixo (cf. Miller *et al*, 1993, p. 04):

SIGNIFICADOS	FORMAS				
	F1	F2	F3	...	Fn
S1	E1,1	E1,2			
S2		E2,2			
S3			E3,3		
...				...	
Sm					Em,n

Fig. 8. A matriz lexical.

S representa os significados; F, as formas. Uma entrada E na célula da matriz implica que a forma naquela coluna pode ser usada (em um contexto apropriado) para expressar o significado naquela linha. Portanto, a entrada E1,1 implica que a forma F1 pode ser usada para expressar o significado S1. Se existem duas entradas na mesma coluna, a forma é polissêmica; se existem duas entradas na mesma linha, as duas formas são sinônimas (Miller *et al*, 1993, p. 04).

Uma vez que um método de representação para significados lexicais é estabelecido, um problema central para a semântica lexical é caracterizar as relações entre significados, a fim de que sejam igualmente representadas. Cabe ressaltar que embora a *WordNet* sirva de motivação para a construção de uma base lexical para a compilação de um *thesaurus eletrônico*, seus objetivos são distintos: enquanto o primeiro é um modelo computacional que visa responder à questão “Qual é a natureza e organização dos conceitos lexicalizados que as palavras podem expressar?”, um *thesaurus eletrônico* pretende ser um dicionário de sinônimos e antônimos, armazenado na memória do computador para ser utilizado no processamento manual ou automático de textos (Dias-da-Silva *et al*, 2000, p. 04; 06).

CAPÍTULO 5 - AS RELAÇÕES DE SENTIDO E A ESTRUTURAÇÃO DO LÉXICO

5.1 Um inventários das relações de sentido

Embora a base lexical do *thesaurus* seja estruturada em função da sinonímia e antonímia, a literatura aponta para a existência de outras tantas relações de sentido no léxico, relevantes para o enriquecimento de um *thesaurus*; portanto, este capítulo inicia-se com um breve exame das muitas possibilidades de relações entre as unidades lexicais e culmina com o exame detalhado da sinonímia e antonímia.

Dentre as preocupações da semântica lexical, pura ou computacional, está o estudo de como os lexemas relacionam-se uns aos outros na expressão lingüística do significado, ou seja, o estudo das **relações de sentido**.

Sob o ponto de vista contextual, podemos afirmar que o significado de uma unidade lexical envolve também um grande número de relações sintagmáticas e paradigmáticas, tanto é que um falante é capaz de eliminar a ambigüidade de um item polissêmico de uma frase através do contexto. Entretanto, ao mesmo tempo que uma única unidade lexical contrai relações sintagmáticas e paradigmáticas, cada uma delas ainda se comporta como um todo, com características que as individualizam; uma prova disso é a de que o lexicógrafo pode “separar” os lexemas no dicionário numa tentativa de delimitar significados. A entrada do dicionário fornece um conjunto de informações relacionadas a um determinado item lexical utilizando-se de outros itens lexicais, pelo uso da metalinguagem; diante disso, podemos afirmar que cada lexema contrai relações com outros lexemas. Lyons confirma tal posição, afirmando que sob o ponto de vista estruturalista, defendido por Saussure e discípulos, cada elemento possui seu lugar num sistema e que sua função adquire função e valor através das relações que contrai com outras unidades do sistema; desta forma, o vocabulário de uma língua pode ser descrito

em função das relações de sentido tanto no plano paradigmático como sintagmático do sistema lingüístico (Lyons, 1981, p. 471).

Portanto, as relações de sentido são fundamentalmente de dois tipos: paradigmáticas e sintagmáticas. As relações paradigmáticas refletem como a realidade é apreendida em sua infinidade e controlada através de sua categorização, subcategorização e graduação ao longo de dimensões específicas de variação; elas representam as escolhas que um falante realiza quando está codificando sua mensagem. Já as relações sintagmáticas estão a serviço da coesão, seja adicionando informação redundante necessária a uma mensagem, seja controlando a contribuição semântica de cada elemento através da desambiguação, seja assinalando estratégias para interpretação (Cruse, 1995, p. 86).

As relações de sentido podem ser estabelecidas num esquema que inclua a noção de implicação, através da negação e afirmação e, para ser significativa, uma relação de sentido tem que ser ao menos sistemática, no sentido de que recorrem a um mínimo de pares de unidades lexicais relacionadas (unidade lexical aqui se refere à forma em conjunto com um significado) (p. 84).

5.1.1 Hierarquias

Uma hierarquia é um conjunto de elementos relacionados uns aos outros de maneira característica. Com relação à unidades lexicais, chamamos a atenção para algumas relações que as caracterizam, como as de dominância, com caráter direcional (Cruse, 1995, p. 112-114): assimetria, simetria, transitividade, não transitividade.

5.1.1.1 Assimetria

Um certo elemento A contrai uma relação R com um elemento B; se R é uma relação assimétrica, então B não contrai necessariamente a mesma relação R com A. Ex:

um canário é um pássaro. Um pássaro não é necessariamente um canário (pode ser um pardal, um colibri, etc.)

5.1.1.2 Simetria

A relação simétrica é válida para ambas as direções: se A é similar a B, então necessariamente B é similar a A. Ex: *garoto = menino*.

5.1.1.3 Transitividade

Se uma relação é transitiva entre um elemento A e B e também entre um terceiro elemento C, está garantida uma relação entre A e C. Ex: *se A é maior que B, e B é maior que C, então A é maior que C*. Logo, a relação “maior que” é transitiva.

5.1.1.4 Não transitividade

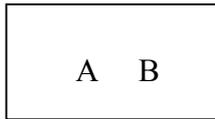
Se há uma relação entre A e B, e entre B e C, não ocorre necessariamente a mesma relação entre A e C. Ex: *se A é pai de B, e B pai de C, então A não poderia ser pai de C*. Portanto, a relação “pai de” é não transitiva.

5.1.2 As relações de sentido fundamentais

As relações entre as classes de palavras que estabelecem um grupo fundamental de relações de sentido são basicamente quatro (Cruse, 1995, p.87):

5.1.2.1 Identidade

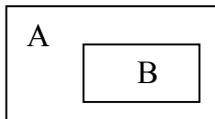
A classe A e classe B têm os mesmos membros.



Ex: garotos/meninos.

5.1.2.2 Inclusão

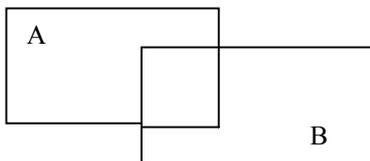
A classe B está inteiramente incluída na classe A.



Ex: pardal/pássaros.

5.1.2.3 Intersecção

A classe A e a classe B têm membros em comum mas cada uma tem membros que não se encontram na outra.



Ex: pardal/morcego.

5.1.2.4 Disjunção

A classe A e a classe B não têm membros em comum.



Ex: pardal/bola.

A seguir damos um panorama das relações de sentido em geral.

5.1.3 Uma tipologia de relações de sentido

5.1.3.1 Sinonímia cognitiva

Também considerada a sinonímia por excelência (Lyons, 1981, p. 478); é uma relação definida em termos de condições de verdade:

X é um sinônimo cognitivo de Y se (i) X e Y são sintaticamente idênticos, e (ii) qualquer sentença declarativa gramatical contendo X tem condições de verdade idênticas às de outra sentença S1, que é idêntica a S exceto se X for substituído por Y. (Cruse, 1995, p. 88, trad. nossa).

Ex: *Comprei uma moto.*

Comprei uma motocicleta.

A sinonímia, bem como a antonímia, serão amplamente discutidas adiante, neste mesmo capítulo.

5.1.3.2 Hiponímia/hiperonímia

Relação caracterizada pela inclusão de uma classe em outra. Aqui, X é hipônimo de Y (e conseqüentemente Y é um superordenado de X) se $A \text{ é } f(X)$ implicar mas não ser implicada por $A \text{ é } f(Y)$:

Isto é um cão. [necessariamente implica] *Isto é um animal.*

A relação **hiponímia/hiperonímia** é também chamada de relação de subordinação/superordenação; ocorre quando temos algo do tipo “um x é (um tipo de) y”. Ex: uma *laranjeira* é hipônimo de *árvore*; *árvore* é hipônimo de *planta*. Lyons (1977, p. 292) afirma que a hiponímia é transitiva e assimétrica e, desde que haja normalmente apenas um superordenado, esta relação semântica gera uma estrutura hierárquica largamente utilizada por cientistas da computação para construir sistemas de

recuperação de informações através da propriedade da herança lexical: o hipônimo herda todos os traços de um conceito mais genérico e adiciona pelo menos mais um traço que o distingue de seu superordenado e de quaisquer outros hipônimos daquele superordenado. Um exemplo: um *canário* é um *canário*, um *canário* é um *pássaro*, um *canário* é um *animal*. É uma relação muito utilizada nas definições de dicionário.

5.1.3.3 Compatibilidade

Relação lexical que corresponde à uma intersecção entre determinadas classes. Dois itens lexicais são ditos compatíveis quando há um superordenado comum. Ex: *cão* e *animal de estimação* são compatíveis; o superordenado é *animal*.

5.1.3.4 Incompatibilidade

Relação entre classes sem membros em comum. Dois itens X e Y são incompatíveis se uma sentença do tipo *A é f(X)* implica numa sentença do tipo *A não é f(Y)*. Exemplo: *gato* e *cão* são incompatíveis porque *é um gato* [implica] *não é um cão*.

5.1.3.5 Meronímia/holonímia

Também chamada de relação parte/todo. Ocorre quando temos algo do tipo “um x é parte de y”; é uma relação transitiva (com qualificações) e assimétrica (Cruse, 1995, p.158). Ex: *perna* é merônimo de *corpo humano*; *corpo humano* é holônimo de *perna*.

5.1.3.6 Troponímia⁵

⁵ Do grego *tropos*, maneira ou modo (Miller; Fellbaum, 1991, p. 216).

Relação que ocorre entre verbos; é caracterizada pelo elemento semântico “maneira”. Ex: *cochichar* é *falar* de certa maneira; portanto *cochichar* é tropônimo de *falar* (cf. Miller e Fellbaum, 1991, p. 215).

5.1.3.7 Relações parciais

Ocorrem entre itens lexicais que têm uma distribuição sintática que coincide parcialmente. Ex: *acabar* e *completar* – são sinônimos parciais por duas razões:

- a) *acabar* pode ocorrer sem um objeto direto (*Você acabou?*) e pode tomar um complemento preposicional (*Acabei de ler*);
- b) *completar* requer um objeto direto (*Você completou a tarefa?*) e não aceita complemento preposicional (**Completei de ler*).

Portanto, pode-se dizer que *completar* é um sinônimo cognitivo de *acabar* apenas num subconjunto de ocorrências gramaticais.

5.1.3.8 Quase-relações

Ocorrem quando “falta” um item lexical para ser o superordenado numa hierarquia. Ex: cores – *É vermelho, amarelo, verde* implica *É X*. *Colorido* não pode ser, pois em muitos contextos esse item exclui outras cores; portanto, uma fotografia colorida não pode ser simplesmente branca, preta e cinza; um lápis colorido não pode ser simplesmente preto, etc. Neste caso, *cor* serviria de um quase superordenado.

5.1.3.9 Pseudo-relações

Ocorrem quando dois itens lexicais “imitam” características contextuais um do outro em circunstâncias especiais. Ex:

Este triângulo tem três ângulos iguais.

Este triângulo tem três lados iguais.

Observe-se que nestas frases chama-se a atenção para as propriedades dos triângulos; entretanto, elas afirmam coisas diferentes, pois *ângulo* e *lado* possuem identidades semânticas distintas (Cruse, 1995, p. 98). Desta forma, as duas sentenças podem ser consideradas apenas como pseudo-sinônimas.

5.1.3.10 Para-relações

As sentenças de uma língua são consideradas bem formadas quando seus constituintes apresentam uma relação sintático-semântica apropriada. No entanto, a língua natural pode ser satisfeita com traços esperados; uma relação feita em termos de expectativas é chamada de para-relação. Ex:

.para-hiponímia – *cão* e *animal de estimação*

É um cão, mas é um animal de estimação. (paradoxo)

É um cão, mas não é de estimação. (normal)

.para-incompatibilidade – *estudante* e *gerente de banco*

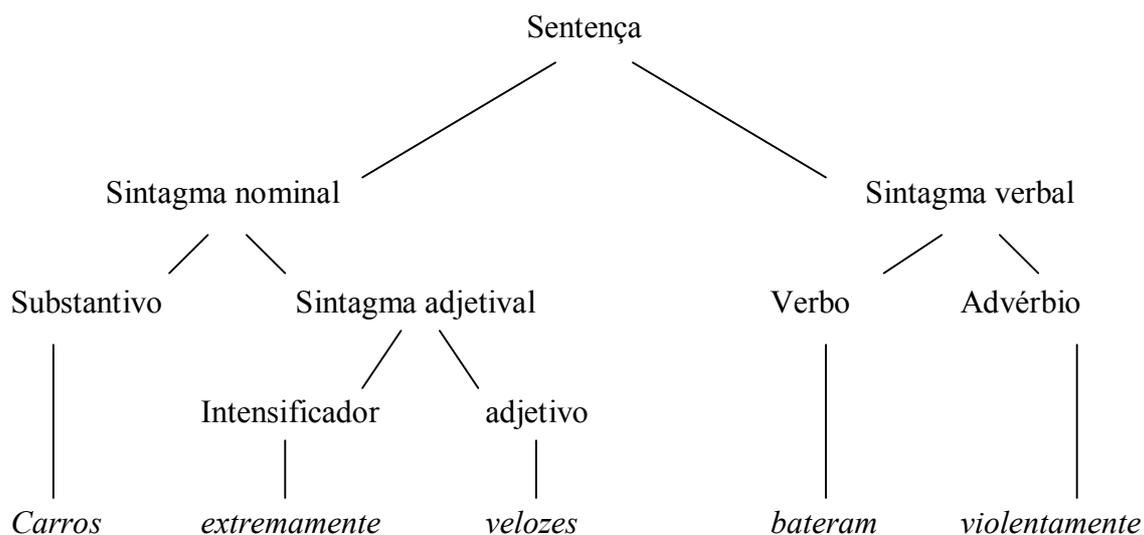
Ele é um estudante, mas é também um gerente de banco. (normal)

Ele é um estudante, mas não é um gerente de banco. (redundante)

- *relações sintagmáticas* – relações contraídas entre os elementos de uma sentença; estabelece-se uma interação entre os vários elementos regulada pela sintaxe da língua.

Ex: *Carros extremamente velozes bateram violentamente.*

No esquema de árvore:



Cada elemento da sentença contrai relações entre si e entre as construções; um substantivo está relacionado com o sintagma nominal, que por sua vez se relaciona com o sintagma verbal, que possui dois constituintes relacionados entre si, e assim por diante. As chamadas restrições semânticas ocorrem porque existem elementos específicos que preenchem o espaço sintático; temos assim um fenômeno semântico dependente da sintaxe.

5.1.3.11 Antonímia

Fenômeno da oposição de significados entre os lexemas (Lyons, 1977, p. 270). De maneira geral, este é o conceito do senso comum; no entanto, entre pares como *bom:mau*, *menor:maior*, *leste:oeste*, existem tipos diferentes de relação de contraste. A noção de contraste, por sua vez, parece determinar uma tendência humana a categorizar a experiência em termos de contrastes dicotômicos (Lyons, 1977, p. 277; 1981, p. 499).

Portanto, as relações semânticas representam associações que formam uma rede semântica complexa; saber em que ponto da rede uma palavra se situa é uma parte importante para se saber seu significado. É uma rede rígida e ao mesmo tempo aberta o

suficiente para admitir novos significados e novas associações entre os elementos constituintes.

Posto que um *thesaurus eletrônico* é um dicionário de sinônimos e antônimos, encerramos este capítulo com uma discussão aprofundada dessas duas relações. Dela pretendemos extrair os princípios que devem nortear a montagem da base lexical subjacente ao *thesaurus*.

5.2 As relações de sentido definidoras do thesaurus eletrônico

5.2.1 Sinonímia

A discussão acerca da sinonímia está presente nos estudos da linguagem desde a antigüidade, com as preocupações dos gregos acerca da identidade; Aristóteles, em seu “Organon”, utilizava-se do termo “sinônimo” para indicar a relação das “palavras às coisas”, e afirmou ser “homem” sinônimo de “boi”, no sentido de que ambos são “animais” (Rey-Debove, 1997, p. 91). Na realidade, o que Aristóteles chamou de sinonímia a lingüística hoje trata de hiperonímia; entretanto, fica com os gregos a origem do termo “sinônimo”, vinda do original “synónymon” (cf. Aurélio do séc. XXI).

Mais adiante, já no século XIX, Frege discutiu a questão da identidade na linguagem através de seu clássico exemplo:

(I) A estrela da manhã é a estrela da tarde.

Segundo Lyons (1977: p. 197-199), Frege afirma que as duas expressões possuem a mesma referência (*bedeutung*), já que ambas se referem a um mesmo planeta, Vênus; no entanto, não se pode afirmar que possuam o mesmo sentido (*sinn*), posto que teríamos uma tautologia:

(II) A estrela da manhã é a estrela da manhã.

Lyons chama a atenção para a questão da informatividade de ambas as frases. O exemplo I permanece mais informativo do que II, pois chama a atenção para uma informação que não poderia ser apreendida somente da leitura de II. Portanto, “estrela da manhã” e “estrela da tarde” não poderiam ser consideradas sinônimas, pois não têm o mesmo sentido. Porém, se as duas expressões fossem tratadas como nomes próprios, poderia-se afirmar que são desprovidas de sentido; mas se fossem tratadas de maneira óbvia a qualquer falante de uma língua em virtude de seu conhecimento da própria língua, não haveria problemas com:

(III) A estrela da manhã não é uma estrela (é um planeta).

Desta forma, não há contradição, e a carga informativa permanece. Através do exemplo de Frege, Lyons conclui: “expressões podem variar em sentido, mas ter a mesma referência; e ‘sinônimo’ significa ‘ter o mesmo sentido’, não ‘ter o mesmo referente” (Lyons, 1977 (1): p. 199, trad. nossa). Ou, conforme o esquema de Rossi (1997, p. 105):

A=A
Tautologia
Identidade *per se*

A=B
Sinonímia
Identidade entre dois termos distintos

A = B
x = x
identidade entre duas
maneiras de dizer, com base na
referência a um mesmo objeto

A = B
R1 = R2
semelhança entre duas representações

5.2.1.1 Sinonímia, contexto e o critério da possibilidade de substituição

Uma das premissas para a identificação da sinonímia é a da possibilidade de substituição dentro de um mesmo contexto [Lyons, (1981, p. 50-1), Geckeler (1976, p.285 *apud* Barbosa, 1994, p.151), Ullmann, 1963, p. 108-109; 1967, p. 159; Apresyan *et al*, 1969, p. 05)]. Assim sendo, dois termos serão considerados sinônimos se forem

intercambiáveis dentro de um mesmo contexto sem alterar o sentido das sentenças. No entanto, mesmo esta definição carece de refinamentos, visto que, por exemplo, pseudo-sinônimos são intercambiáveis em apenas alguns contextos, como veremos adiante.

Lyons (1977, p. 202) utiliza-se do exemplo das sentenças equativas, que trazem o verbo “ser” para expressar identidade (como no exemplo de Frege); essas mesmas sentenças encontram oposição nas declarativas que, mesmo utilizando-se do mesmo verbo, não são equivalentes:

“John é um bobo.”

No caso, “John” é uma expressão referente, mas “ser um bobo” é um predicativo. Lyons propõe que vejamos nas duas expressões dois sentidos diferentes: se no lugar de “John” colocarmos qualquer outra expressão equivalente (um pronome, um sintagma nominal, etc.) que identifique o mesmo indivíduo “John” no mesmo contexto de enunciação, então o significado descritivo da declaração não será afetado. Se fizéssemos o mesmo com “ser um bobo”, trocando por uma expressão que indicasse o mesmo predicativo, novamente o significado descritivo da sentença não seria afetado: “O critério da possibilidade de substituição na posição de sujeito nesta construção é de identidade referencial; o critério de substituição em posição predicativa é identidade de sentido” (Lyons, 1977 (1), p. 201). Portanto, um critério para distinção da igualdade/semelhança de sentidos é extraído ao mesmo tempo que dá mais uma definição de sinonímia:

Duas ou mais expressões serão definidas como tendo o mesmo sentido (i.e. serem sinônimas) sobre um certo escopo de enunciados se e somente se forem substituíveis nos enunciados sem afetar seu significado afetivo” (Lyons, 1977 (1), p. 202).

5.2.1.2 Sinonímia, valor de verdade e implicação

Para melhor compreender a questão do valor de verdade, Lyons (1977 (1), p. 202) sugere dois exemplos:

- (I) John é um bobo.
- (II) John é um lingüista.

Se considerarmos que o indivíduo John é bobo e lingüista, as proposições expressas pelas duas sentenças possuem o mesmo valor de verdade, mas não o mesmo significado descritivo, o que se conclui por intuição: I não significa o mesmo que II. Para confirmar a intuição do usuário da língua, Lyons (1977 (1), p. 202) sugere o seguinte: duas declarações serão descritivamente equivalentes (i.e. possuem o mesmo significado descritivo) se aquilo que uma delas acarreta, a outra também acarreta. Ou, na concepção de Quine (1960:5, *apud* Lyons, 1977 (1), p. 202), as sentenças são sinônimas se e somente se estiverem relacionadas pela relação de equivalência. Um exemplo de equivalência está na igualdade expressa entre os dois lados da expressão matemática “ $2 + 2 = 4$ ”; tais expressões são consideradas verdadeiras *a priori*. Outro exemplo de equivalência está nos postulados de Carnap, pelos quais se estabelece relações de vínculo e pressuposição:

- (x) $(Sx - \sim Cx) = \text{“Nenhum } X \text{ que seja solteiro é casado”}$
(cf. Lyons, 1977 (1), p. 203).

Desta forma, pode-se definir a sinonímia em função da implicação recíproca ou equivalência:

Se uma frase, F1, implica uma outra frase, F2, e se ocorre também o inverso, F1 e F2 são equivalentes: isto é, se $F1 \supset F2$ e se $F2 \supset F1$, então $F1 \equiv F2$ (em que \equiv significa “é equivalente a”). Se as duas frases equivalentes têm a mesma estrutura sintática e diferem uma da outra apenas pelo fato de uma ter a unidade lexical x, a outra tem y, então x e y são sinônimas. (Lyons, 1981, p. 478).

A sinonímia é geralmente tomada como uma relação entre palavras e não entre conceitos (Pustejovsky, 1994, p. 23). Uma definição de sinonímia que leva em conta a possibilidade de substituição em um mesmo contexto ou mesmo em outros é válida se não se alterar o valor de verdade de uma sentença; uma outra possibilidade é restringir a sinonímia apenas a contextos específicos, como no caso de [*abalo, balanço, sacudidela, solavanco*] e [*abalo, baque, choque, comoção, concussão, perturbação, sobressalto*] (cf. Miller et al, 1990).

5.2.1.3 Sinonímia e contexto

Lyons chama a atenção para o fato de que todas as relações de sentido são fortemente dependentes do contexto, mas a sinonímia **determinada** pelo contexto é de especial importância; por exemplo, em “a criança teve seu ovário retirado” fica claro que a criança é do sexo feminino. Ou, da mesma forma, dizer “vou ao supermercado buscar carne” ou “vou ao supermercado comprar carne” acaba de certa forma transformando “buscar” e “comprar” em sinônimos, demonstrando que a “não se pode estabelecer distinção clara entre a determinação ‘probabilista’ da sinonímia por meio de outras unidades lexicais no mesmo enunciado e a que se faz pelas características da situação em que ocorre o enunciado” (Lyons, 1981, p. 481).

5.2.1.4 Sinonímia e traços componenciais

Numa tentativa de dar um conceito para a **sinonímia**, D. A. Cruse parte de duas “intuições”: a de que certos pares ou grupos de itens lexicais têm uma semelhança muito grande uns com os outros e que alguns pares são mais sinônimos que outros. Sendo

assim, não haveria uma maneira absoluta de definir sinônimos senão analisando-os de duas formas:

-em termos de semelhanças necessárias e diferenças permitidas;

-contextualmente, por meio de *frames* (quadros) de diagnóstico.

Assim sendo, o autor faz uma primeira conceituação: “Sinônimos (...) são itens lexicais que são idênticos em relação a traços semânticos centrais, mas diferem em relação ao que podemos chamar de traços menores ou periféricos” (Cruse, 1995, p. 267).

Cruse identificou duas maneiras de coocorrência de sinônimos em frases:

- podem ocorrer caracteristicamente juntos em certos tipos de expressão, com a presença do marcador “ou seja” e “ou”:

Isto é uma onça, ou leopardo da neve.

-podem também ser usados contrastivamente, sendo que a diferença pode ser sinalizada por alguma expressão como “mais precisamente” ou “ou melhor”:

Ele foi assassinado, ou melhor, executado.

Essas duas intuições apontam para uma escala de sinonímia: “Não há uma motivação óbvia para a existência de sinônimos absolutos nas línguas, e poderia esperar-se que alguns deles caísse na obsolescência, ou que se desenvolveria alguma diferença semântica” (Cruse, 1995, p. 270).

Desta forma, sugere uma escala com os seguintes componentes:

sinonímia absoluta \Rightarrow sinonímia cognitiva \Rightarrow não sinonímia
--

Dentro dos limites de uma escala que identifica a sinonímia absoluta e a não sinonímia como pólos, muitos estudos buscaram identificar uma tipologia sinonímica; Ullmann (1963, p. 109) identifica o que chama de pseudo-sinônimos; Lyons (1997, p. 148) distingue sinônimos absolutos dos sinônimos completos; Lyons (1981, p.150)

caracteriza a sinonímia incompleta; temos ainda a sinonímia relativa, total parcial, total e absoluta, absoluta mas não total, sinonímia com restrições de denotação e comutabilidade, etc, (Vilela, 1994, p.28), o que torna a escala elaborada por Cruse passível de acréscimos e não só tendo a sinonímia cognitiva enquanto intermediária, como veremos adiante, lembrando que o que aparentemente é um grande número de tipos de sinonímia não passa de discordância quanto à nomenclatura conferida ao fenômeno.

5.2.1.5 Sinonímia e conceitos

As relações de sentido se aplicam não somente às unidades do léxico, mas também a frases e expressões fora do léxico; e uma das relações às quais o acesso lexical é sensível inclui a própria sinonímia. De acordo com Bierwisch e Schreuder (1992, p.36)⁶, duas expressões (lexicais ou frasais) A e B são sinônimas se A e B possuem o mesmo significado. De acordo com noções já expressas anteriormente, duas expressões com forma semântica idêntica deveriam ter a mesma interpretação conceitual em todos os contextos. Consideremos os seguintes exemplos:

(1) entrar no jardim

(2) ir para dentro do jardim

Dada a entrada lexical para *entrar* repetida em (3), a forma semântica de (1) apresenta-se como (4), em que a forma semântica de *o jardim* está abreviada como DEF GARDEN, substituindo a variável na posição de objeto:

(3) /entrar/ [+V, -N] $\lambda x \lambda y$ [y DO MOVE y]:[FIN [y LOC IN x]]

(4) /entrar no jardim/ VP

⁶ A noção da relação léxico-conceitos desses autores está delineada no capítulo três.

$$\lambda [y \text{ DO } [\text{MOVE } y]:[\text{FIN } [y \text{ LOC IN } [\text{DEF GARDEN}]]]]$$

Para demonstrar que a mesma forma semântica resulta para (2), considere-se as seguintes entradas para *ir* e *para dentro*:

(5) /ir/ [+V, -N] $\lambda P \lambda y [\lambda \text{ DO MOVE } y]:[P y]$

(6) /para dentro/ [-V, -N, +Dir] $\lambda x \lambda y [\text{FIN } [z \text{ LOC IN } x]]$

As preposições direcionais *para dentro* relacionam um tema *z* a um objetivo *x*, tal que o sintagma preposicional *para dentro do jardim* defina uma propriedade do tema *z*, a saber, a localização da parte final de alguma dimensão associada com *z* no interior do jardim:

(7) (a) [pp[p *dentro*] [np *o jardim*]]

(b) [$\lambda z [\text{FIN } [z \text{ LOC IN } [\text{DEF GARDEN}]]]]$]

Esse sintagma preposicional pode agora ser combinado com o verbo *ir*, em que sua forma semântica substitui a variável *P* relacionada às propriedades do sujeito do verbo:

(8) /ir para dentro do jardim/ VP

$$\lambda y [y \text{ DO } [\text{MOVE } y]: [\text{FIN } [y \text{ LOC IN } [\text{DEF GARDEN}]]]]$$

Pode-se tirar algumas conclusões destas formalizações. A forma semântica de (8) é de fato idêntica à de (4), o que prova que as duas expressões são opções para se

verbalizar o mesmo tipo de estruturas conceituais; portanto, uma escolha é feita no momento de se formular o enunciado. E mais, embora *entrar* de certa forma incorpore a forma semântica de *para dentro*, não há apenas um constituinte *para dentro* em (8), já que *dentro* é o núcleo de um sintagma preposicional que corresponde ao objeto de *entrar*. Em outras palavras, a escolha entre (4) e (8) implica uma diferença em se desmembrar a mensagem a ser codificada em partes constituintes que podem ser lexicalizadas e integradas de acordo com princípios gramaticais (Bierwisch e Schreuder, 1992, p. 38).

Um outro exemplo de sinonímia conceitual, agora no nível lexical, pode ser extraído de Jackendoff (1991, p. 253):

- a) comer
- b) devorar

Estruturas de representação:

$$\left[\begin{array}{l} \text{Comer} \\ \text{V} \\ [\text{CAUSA} ([\text{coisa }]^{\alpha} \text{A}, [\text{IR} ([\text{coisa }] <\text{A}>, \\ [\text{PARA} [\text{EM} [\text{BOCA-DE} [\alpha]]]])]) \end{array} \right]$$

$$\left[\begin{array}{l} \text{Devorar} \\ \text{V} \\ [\text{CAUSA} ([\text{coisa }]^{\alpha} \text{A}, [\text{IR} ([\text{coisa }] \text{A}, \\ [\text{PARA} [\text{EM} [\text{BOCA-DE} [\alpha]]]])]) \end{array} \right]$$

A única diferença é que *comer* traz o argumento A entre parênteses, indicando sua opcionalidade. O número de argumentos, portanto a transitividade do verbo, depende se A é escolhido ou não. Já *devorar* tem uma marca A obrigatória no segundo argumento, portanto é sempre transitivo (Jackendoff, 1991, p. 253).

5.2.1.6 Sinonímia: uma tipologia

Através do exame dos estudos acerca da sinonímia levantamos uma tipologia para o fenômeno, que apresentamos a seguir. Cabe ressaltar que cada teórico utilizou-se de certos parâmetros para a delimitação de seus tipos sinonímicos; ainda assim, é possível depreender equivalências e divergências a partir de suas definições, permitindo-nos relacionar os tipos uns com os outros e elaborar um quadro síntese dos tipos encontrados.

5.2.1.6.1 Sinonímia absoluta

É também chamada de **sinonímia completa e total** (Lyons, 1981, p. 476) ou **total e absoluta** (Martin, 1976, p.113 e ss., *apud* Vilela, 1994, p. 28). Sinônimos perfeitos ou puros são aqueles coextensivos e intercambiáveis em valor intelectual e afetivo (Ullmann, 1963, p. 109). Nestes incluem-se, por exemplo, os termos que se referem a doenças; exemplos do dicionário Aurélio do Séc. XXI:

malária – maleita – febre intermitente – impaludismo

gripe – influenza

tuberculose – tísica

Neste tipo de sinonímia, dois lexemas **a** e **b** são comutáveis em qualquer contexto, como *consoante surda/áfone* (Martin, 1976, p.113 e ss., *apud* Vilela, 1994, p. 28) sugere que a sinonímia total e absoluta ocorre apenas em contextos em que haja identidade denotativa e conotativa, e a linguagem técnica seria por excelência um campo para este tipo de sinonímia: *oftalmologista/médico dos olhos/oculista*.

Observe-se que a definição de Ullmann leva em conta as dimensões do significado cognitivo e afetivo. Segundo Lyons, essa distinção entre o significado

dependente do intelecto e daquele que se vale da imaginação e afetividade é pouco apropriada, pois o uso de uma unidade ou da escolha por se utilizar seus sinônimos depende muito da escolha do usuário da língua, seja por questões estilísticas, fonológicas, de origem, etc; desta forma, é preferível tratar esse tipo de sinonímia de cognitiva (1981, p. 478). Portanto, o que Ullmann chama de sinonímia absoluta, Lyons sugere que seja considerado também como sinonímia cognitiva; por questões didáticas, fiquemos com a distinção.

5.2.1.6.2 *Sinonímia cognitiva*

Também chamada de **sinonímia descritiva** ou **referencial** (Lyons, 1987, p. 145). Para que haja sinonímia cognitiva, um par de itens lexicais deve ter certas propriedades semânticas em comum; portanto, há propriedades em que os sinônimos cognitivos devem ser idênticos e outras em que devem diferir. Entre estas propriedades, destacamos o **modo semântico**, englobando o modo proposicional e o modo expressivo. Consideremos duas sentenças:

Acabo de sentir uma dor aguda. – modo proposicional

Ai! - modo expressivo

O significado proposicional depende da atitude expressa pela sentença em que opera (afirmação, pergunta, ordem, exclamação, etc.). Numa afirmação, é o significado apresentado que determina as condições de verdade com relação a determinado estado de coisas ou a outras afirmativas. Já o significado expressivo é válido apenas para o locutor, no local e momento da expressão: “Feche a *droga* da janela!” e veicula algum tipo de emoção ou atitude (dúvida, certeza, esperança, expectativa, surpresa, desapontamento, admiração, etc). Pares de itens lexicais que diferem somente nos traços expressivos inerentes são sinônimos cognitivos: *bebê* – *neném* ; *papai* – *pai*.

A relevância da observação do modo semântico para a sinonímia cognitiva está no fato de que “o significado herdado de um item lexical pode ser constituído por um ou ambos os tipos de significado; se dois itens lexicais são sinônimos cognitivos, então devem ser idênticos com relação a traços proposicionais mas podem diferir com respeito a traços expressivos” (Cruse, 1995, p. 273).

As **pressuposições** também são importantes na determinação da sinonímia cognitiva, uma vez que as restrições colocacionais estabelecem as condições de verdade de uma sentença. Na expressão “bater as botas” (= morrer), seu uso é possível apenas com sujeitos humanos, o que é uma restrição arbitrária. Tais restrições estão relacionadas à sinonímia cognitiva da seguinte forma: “definimos restrições de coocorrência colocacionais que são irrelevantes para as condições de verdade – ou seja, aquelas a respeito das quais os itens lexicais podem diferir e ainda serem sinônimos cognitivos” (Cruse, 1995, p. 289).

O **significado evocado** contribui para a coesão discursiva e tem um papel comunicativo secundário; não afeta o valor de verdade das sentenças em que ocorre e portanto proporciona uma fonte de variação entre sinônimos cognitivos. Está associado aos diferentes registros de uma língua, ou dialetos. Ex:

port. *menino – piá*

menino - curumim

Com relação ao registro, os sinônimos cognitivos apresentam três dimensões de variação: campo, modo e estilo.

Quando pertencem ao mesmo **campo**, os sinônimos cognitivos são aqueles itens que têm o mesmo referente, mas são lexicalizados de maneira distinta: *sal – cloreto de sódio*.

O **modo** está ligado ao meio de transmissão de uma mensagem, se escrita, falada, telegrafada, etc. Ex: *sobre / a respeito de*.

A dimensão do **estilo** está ligada às características da linguagem que marcam diferentes relações entre os participantes numa interação lingüística. Relaciona-se também com o uso formal ou informal da linguagem, sendo que os sinônimos cognitivos se multiplicam em áreas da experiência humana que possuem alta significância emotiva (morte, dinheiro, religião, etc.). Ex: *bater as botas, morrer, falecer*.

Lyons declara-se inclinado a considerar a sinonímia em si ao que os semanticistas chamam de sinonímia cognitiva (1981, p. 478).

5.2.1.6.3 *Sinonímia parcial*

Também chamada de **pseudo-sinonímia** (Ullmann, 1963, p. 109). Segundo este tipo de sinonímia, dois lexemas **a** e **b** são comutáveis em apenas alguns contextos. Ex: *chefe/patrão* (Vilela, 1994, p. 28).

Lyons, utilizando-se da distinção de Ullmann para o significado cognitivo e afetivo, distingue ainda a **sinonímia completa** para a equivalência dos sentidos cognitivo e afetivo, e **sinonímia total** para os sinônimos, “completos” ou não, que são intercambiáveis em todos os contextos; desta forma classifica quatro tipos de sinonímia: completa e total (que relacionamos no item 1 sob a denominação de sinonímia absoluta); absoluta mas não total; completa mas não total; incompleta e não total (Lyons, 1981, p. 476). Desta forma, os tipos sugeridos por Lyons acabam se encaixando na noção de sinonímia parcial .

5.2.1.6.4 *Sinonímia absoluta mas não-total*

A comutabilidade entre dois lexemas é restrita a alguns contextos: *principiante/debutante* (Vilela, 1994, p. 28; Lyons, 1981, p. 476).

5.2.1.6.5 *Sinonímia completa mas não-total*

Possuem equivalência de significado cognitivo e afetivo mas não se aplicam a todos os contextos. Ex: *professora/tia/mestre. / cara/face/rosto*

5.2.1.6.6 *Sinonímia incompleta e não-total*

Não possuem equivalência cognitiva e afetiva, tampouco se aplicam a todos os contextos. Ex: *roubar/assaltar*.

5.2.1.6.7 *Sinonímia conotativa*

Possuem a comutação restringida por seu valor conotativo: *chefe/patrão*.

5.2.1.6.8 *Sinonímia denotativa*

Sua comutação é restringida por seu valor denotativo: *autor/escritor, diretor/chefe, cara/face/rosto/feições (lavar a cara/face/rosto/feições?)*.

5.2.1.6.9 *Outros casos*

Outros casos de sinonímia apontados por Vilela, 1994, p. 29:

Domínios específicos: *porco/suíno/leitão*

Tabu: *banheiro/toalete/sanitário*

Diferenças de registro: *adiar/protelar, mãe/mamã, louro/loiro*

Diferenças diafásicas: *fazer/efetuar, usufruir de/usar*

Com restrição estilística: *encher/preencher; pesar/ponderar*

Sinônimos “novos”: *programa/software, comercialização/marketing*

5.2.2 *Antonímia*

Como mencionamos anteriormente, **antonímia** é o termo técnico utilizado para descrever o fenômeno da oposição de significados entre os lexemas (Lyons, 1977, p. 270). De maneira geral, este é o conceito do senso comum; no entanto, entre pares como *bom:mau, menor:maior, leste:oeste*, existem tipos de diferentes de relação de contraste. A noção de contraste, por sua vez, parece determinar uma tendência humana a categorizar a experiência em termos contrastes dicotômicos (Lyons, 1977, p. 277; 1981, p. 499).

Uma propriedade intrínseca aos contrários foi já identificada por Sapir, em 1944 (*apud* Lyons, 1977, p. 273), introduzindo a noção de **gradação**, pela qual existe algum tipo de comparação. Desta forma, Sapir traçou a distinção entre **opostos graduáveis** e **não-graduáveis**. **Os opostos graduáveis** seriam os lexemas que estariam em uma relação que permita que se use de comparação; ex: *quente:frio* – podem ser associados aos extremos de uma escala contínua cujos pontos intermediários correspondam à combinação de unidades quantificadoras como *pouco, um pouco, ligeiramente, bastante, muito*, etc (Campos & Xavier, 1991, p. 241). Por outro lado, os **opostos não-graduáveis** estão em relação de implicação, em que uma proposição nega a outra; ex: a proposição “X é fêmea” implica “X não é macho”, o que não acontece com os opostos graduáveis (dizer que “algo não está frio” não necessariamente quer dizer que “algo está quente”, pode “estar morno”), deixando claro que os opostos não-graduáveis podem estar implícitos; opostos explícitos caberiam em proposições do tipo *Minha casa é*

maior que a sua (Lyons, 1981, p. 492-495). Desta forma, a noção dos contrários aplicada às proposições pode ser estendida aos lexemas usados naquelas como predicativos (Lyons, 1977, p. 272).

Segundo Lyons (1977, p. 275), em algumas línguas a noção de contraste pode ser marcada morfologicamente, como podemos constatar no português: *casar:descasar*, *carregar:descarregar*, *formal:informal*.

Nesta breve introdução ao assunto, utilizamo-nos de termos os mais diferentes como contraste, oposição, opostos, etc. Lyons fez esta diferenciação, distinguindo os vários tipos de contraste existentes, deixando claras as noções de contraste e oposição; sua distinção nos servirá de base para a descrição do fenômeno da oposição de maneira geral.

Contraste é pois o termo mais geral, pois não implica um determinado número de elementos opositivos no conjunto respectivo e nem implica que a oposição seja paradigmática; **oposição** será o termo restrito aos contrastes dicotômicos ou binários (Lyons, 1977, p. 179). **Antonímia**, portanto, seria um tipo de oposição, que por sua vez é um tipo de contraste e estará restrita, conforme mostraremos, aos opostos graduáveis, pois estes manifestam a propriedade da polaridade mais explicitamente do que outros (Lyons, 1977, p. 279). Sendo assim, Lyons (1977, p. 270-290) classificou diferentes tipos de contraste, conforme podemos verificar na figura 9 abaixo.

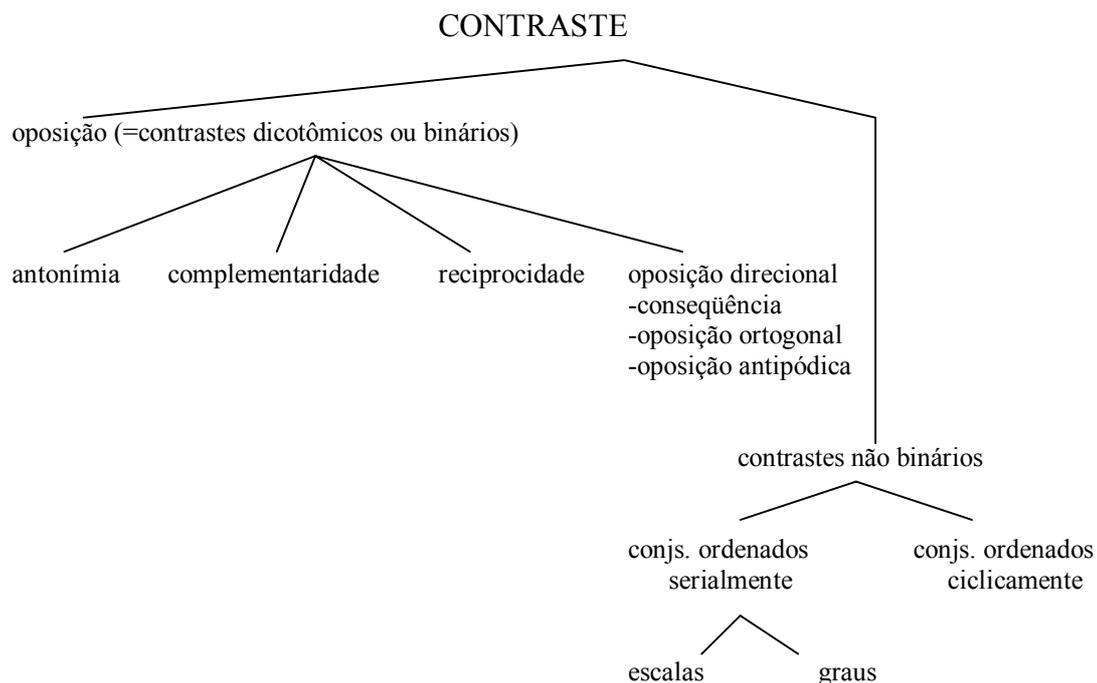


Fig. 9. A classificação dos tipos de contrastes.

A primeira subdivisão de **contraste** é feita entre **contrastos binários ou dicotômicos** e **contrastos não binários**; os primeiros subdividem-se em **antonímia**, **complementaridade**, **reciprocidade** e **oposição direcional**. Os conjuntos não binários subdividem-se em **conjuntos seriais** e **conjuntos cíclicos**, por sua vez subdivididos em **escalas** e **graus** (*ranks*).

5.2.2.1 Contrastos binários ou dicotômicos

5.2.2.1.1 Antonímia

Relação contrastiva que pode ser exemplificada com pares como *longo:curto*, *rápido:vagaroso*, *fácil:difícil*, etc; compartilham das seguintes características (Lyons, 1968, p. 463-4, *apud* Cruse, 1995, p. 204) :

- são totalmente graduáveis (a maioria são de adjetivos, há poucos verbos);

- denotam graus de alguma propriedade graduável como comprimento, velocidade, peso, acuidade, etc.
- quando intensificados, os membros do par se movem em direções opostas ao longo de uma escala, representando os graus da propriedade variável mais relevante. Ex: *muito pesado* / *muito leve*, que podem ser separados com maior acuidade numa escala de peso do que *pouco pesado* e *pouco leve*.
- os membros do par não esgotam um domínio: há valores de propriedade variável que se localizam entre eles. Como resultado, uma afirmação contendo um membro de um par de antônimos está em relação de contrariedade com a afirmação paralela do outro termo. Portanto, *Está comprido* e *Está curto* são contrários, e não contraditórios.

O que caracteriza os antônimos é essencialmente o fato de poderem ser graduáveis, sendo que a noção de “gradação” é emprestada de Sapir e se liga à operação de comparação de duas formas, explícitas ou implícitas: (i) comparação entre certa propriedade que está em maior grau e portanto caracteriza mais um dos membros de um par. Ex: *Nossa casa é maior que a sua*; (ii) comparação de dois estados em relação a uma certa propriedade. Ex: *Nossa casa está maior do que costumava ser*. Tais enunciados só podem ser interpretados se o outro termo da comparação puder ser recuperado no contexto (Lyons, 1981, p. 491-2).

O comportamento de um par de antônimos pode ser representado diagramaticamente, como na figura 10 abaixo (Cruse, 1995, p. 205).

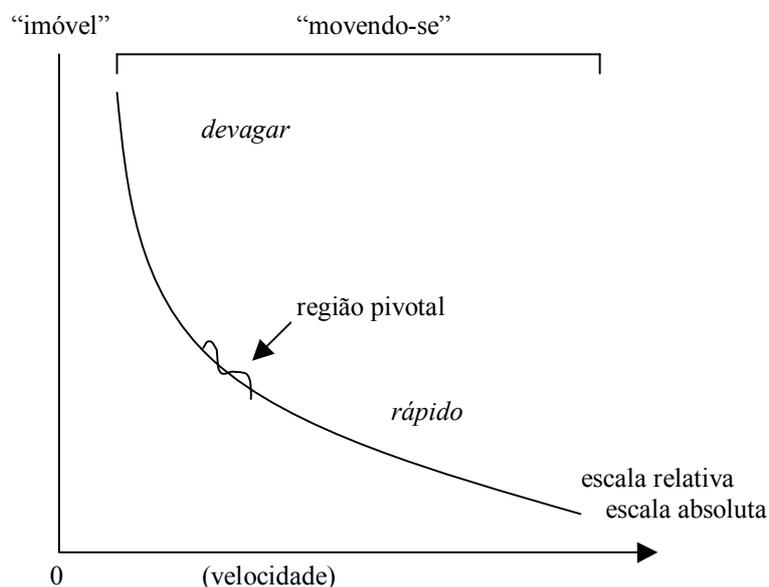


Fig. 10. Comportamento dos pares de antônimos *devagar:rápido*.

De um par de antônimos associados desta maneira, um tenderá a zero enquanto o outro tenderá à direção contrária; ambos estarão dispostos em torno de uma área neutra chamada **pivotal**, pois não pode ser referida por nenhum dos membros do par de antônimos (por exemplo, para o par *quente:frio*, a região pivotal poderia ser representada por *morno*). O valor de *devagar*, embora tenda a zero, nunca o alcança, apenas se aproxima; isso não é um fato físico, mas lingüístico, pois não se pode dizer *completamente devagar* quando se quer dizer *imóvel*. Portanto, o comportamento de *devagar* é típico dos membros orientados a zero dos pares de antônimos (Cruse, 1995, p. 205-6).

Assim sendo, é possível associar duas unidades antônimas aos extremos (ou pólos) de uma escala contínua cujos pontos intermediários correspondem à combinação de unidades lexicais com os quantificadores *pouco*, *um pouco*, *ligeiramente*, *bastante*, *muito*, *muitíssimo*, *mais/menos...que*, etc. Portanto, um par como *devagar:rápido* é de

antônimos graduáveis. Outros exemplos seriam *grande:pequeno*, *rico:pobre*, *bom:mau*, *longe:perto*, etc (Campos & Xavier, 1991, p. 241).

Cruse (1995, p. 208) sugere uma subdivisão da antonímia em três subgrupos:

- **antônimos polares** – ex: *longo:curto*. São tipicamente neutros e objetivamente descritivos; na maioria dos casos podem ser mensurados convencionalmente através de unidades de medida como polegadas, metros, gramas, etc.
- **antônimos sobrepostos** – ex: *bom:mau*, *limpo:sujo*. Os pares pertencentes a este grupo possuem valores fóricos como parte de seu significado: um termo é fórico (*bom*, *limpo*) e o outro é disfórico (*mau*, *sujo*). A identificação de tais pares não é clara, mas aparentemente referem-se a situações em que um estado desejável é menos a presença de alguma propriedade valiosa do que a ausência de uma propriedade indesejada, como a *sujeira*.
- **antônimos equípolentes** – ex: *quente:frio*, *feliz:triste*. Não são muitos, e referem-se a sensações ou emoções baseadas em reações subjetivas.

5.2.2.1.2 Complementaridade

É o tipo de oposição que se estabelece entre pares como *casado:solteiro*, *macho:fêmea*, *morto:vivo*, etc. Sendo assim, nos pares de opostos complementares, não há a possibilidade de inserção de um terceiro elemento entre eles (Cruse, 1995, p. 199). Do ponto de vista semântico, a expressão “o cão não estava vivo nem morto” é anômala. Também caracteriza tais pares o fato de que a negação de um implica a afirmação do outro, e vice-versa: $\sim x \supset y$ e $y \supset x$ (Lyons, 1981, p. 489); os opostos complementares constituem um subsistema em que são dois os valores possíveis, valores estes, por definição, incompatíveis; trata-se portanto de uma relação binária e complementar (Campos e Xavier, 1991, p. 237).

Entretanto, cabe aqui uma observação. A linguagem não foi “projetada” segundo os princípios da lógica, permitindo que usemos de expressões como “Estou *morto* de cansaço” (na realidade estou vivo, mas cansado), ou “Assista hoje a volta dos *mortos-vivos*”, ou ainda “Hoje estou *mais morto que vivo*”, algo que ocorre porque a complementaridade também é uma questão de grau, o que demanda um exame do contexto de aplicação de tais expressões antes que se aceite se a relação de vínculo é aceitável ou não (Cruse, 1995, p. 200).

Segundo Cruse (1995, p. 201-202) os opostos complementares podem ser subdivididos em quatro subtipos:

- a) **reversivos** – a oposição básica está entre “uma continuidade de um estado” e “mudança para um estado alternativo”. Ex: no conjunto *nascer:viver:morrer*, *nascer* e *morrer* estão em relação reversiva.
- b) **interativos** – os pares aqui envolvidos estão em uma relação de estímulo-resposta. Ex: num contexto de comando como *comandar:obedecer:desobedecer*, os dois últimos são complementares, enquanto que *comandar* e *obedecer* estão em relação interativa, representando estímulo e resposta.
- c) **satisfativos** – nos pares envolvidos, um termo representa uma tentativa de se realizar algo, enquanto que o outro representa uma performance completa e satisfatória. Ex: *procurar:encontrar*, *competir:vencer*, *mirar:atingir*.
- d) **contra-ativos** – neste caso, um termo representa uma ação agressiva e o outro, alguma espécie de meio para neutralizá-la. Ex: no conjunto *atacar:defender:render*, os dois últimos são contra-ativos pois representam uma resposta a uma ação original.

5.2.2.1.3 Reciprocidade

Também chamada de oposição conversa; é uma relação de oposição que ocorre entre pares do tipo *marido:esposa, comprar:vender*:

Em ambos os casos, a substituição léxica de um termo pelo antônimo correspondente ou recíproco se associa à transformação sintática que permuta o sintagma nominal, SN₁ e SN₂, e determina algumas mudanças “automáticas” na seleção da preposição apropriada (ou da flexão casual em outras línguas). Observe-se que esse traço “permutacional” é também característico da relação entre frases ativas e passivas correspondentes: SN₁ matou SN₂ implica e é implicado por SN₂ foi morto por SN₁. (Lyons 1981, p. 496-497)

Da mesma forma, esta relação só pode ocorrer com unidades lexicais usadas como predicados de dois lugares, caracterizando o chamado **predicado simétrico** (Campos e Xavier, 1991, p. 244):

João casou-se com Ana.

Ana casou-se com João.

Incluem-se nesta categoria formas de comparativo de antônimos graduáveis (*mais alto/baixo que*), forma ativa e passiva de verbos transitivos (*comer/ser comido*), locuções prepositivas de localização temporal ou espacial (*antes/depois de, à direita/à esquerda de*), etc (Campos e Xavier, 1991, p. 243).

5.2.2.1.4 Oposição direcional

Nesta relação, implica-se um movimento relativamente a uma direção entre duas direções opostas possíveis, tendo em conta um ponto determinado (Lyons, 1977, p. 281; Vilela, 1994, p. 171). São advérbios ou preposições: *norte:sul, acima:abaixo*, etc (Cruse, 1995, p. 223) ou ainda verbos como *vir:ir, chegar:partir, subir:descer*, etc (Vilela, 1994, p. 171). Podem ser subdivididos em conseqüência, ortogonal e antipódica (Lyons, 1977, p. 281-7); antipódicos, contrapartes, reversivos (independentes e

restitutivos), relacionais reversivos (diretos e indiretos) e pseudo-opostos (Cruse, 1995, p. 223-242); ambos os teóricos consideram a oposição antipódica.

- **Relação de conseqüência** – nesta relação, um elemento conseqüente pressupõe um elemento antecedente. Ex: *procurar:encontrar*.
- **Oposição ortogonal** – relação de conteúdo num campo composto de mais de dois elementos, sendo que determinados elementos do campo estão em oposição vertical (perpendicular). Ex: *norte* está em oposição ortogonal relativamente a *leste* e *oeste*; *oeste* está em relação ortogonal relativamente a *norte* e *sul* (Lyons, 1977, p. 282-3; Vilela, 1994, p. 171).
- **Oposição antipódica** – oposição diametralmente oposta entre certos elementos de um mesmo campo. Ex: *norte:sul*, *leste:oeste*. (Lyons, 1977, p. 282-3; Vilela, 1994, p. 171-2).
- **Contrapartes** – relação que ocorre entre pares que refletem algum tipo de desvio em uma superfície ou forma, com direções necessariamente opostas. Ex: *colina:vale*, *convexo:côncavo* (Cruse, 1995, p. 225).
- **Reversivos** – pares de verbos que denotam mudança de movimento em direções opostas. Ex: *ascender:descender*, *avançar:recuar*, *entrar:sair*, *embarcar:desembarcar*. Incluem os tipos (Cruse, 1995, p. 226-8):
 - **Restitutivos** – denotam a restituição de um estado anterior. Ex: *remover:restituir*, *matar:ressuscitar*, *destruir:reparar*. Note-se que há um

elemento dependente e um independente em termos de lógica, ao contrário do tipo seguinte.

- **Independentes** – embora não estejam vinculados da maneira dos restitutivos, nestes casos pode haver uma grande expectativa pragmática para a restauração de um estado anterior. Ex: *encher:esvaziar*. Uma fruta como o coco pode ser esvaziada sem antes ter sido preenchida.

- **Relacionais reversivos** – relação espacial entre opostos que especificam a direção de um em relação ao outro ao longo de alguma dimensão: se dois objetos A e B estão em locais diferentes, a direção de A em relação a B é exatamente oposta da direção de B em relação a A. Ex: *acima:abaixo, antes:depois* (Cruse, 1995, p. 231). Portanto, os opostos reversivos devem ser capazes de expressar relações assimétricas entre duas entidades. São de dois tipos:
 - **Diretos** – ambos os elementos do par envolvem um sintagma nominal que ocupa um “slot” (posição) valencial central, como em *preceder:seguir*. Ex: *A reunião foi seguida de um coquetel. / Um coquetel foi precedido da reunião.*

 - **Indiretos** – envolvem um objeto indireto que é periférico, em que um sintagma nominal central e um periférico podem ser intersubstituídos. Ex: *dar:receber* - *Miriam deu uma caixa para Arthur. / Arthur recebeu uma caixa de Miriam* (Cruse, 1995, p. 233-4).

- **Pseudo-opostos** – pares de opostos dependentes de contexto. Ex: *médico:paciente* ou *dentista:paciente, vítima:ladrão* ou *vítima:assassino*. Somente em contexto esses pares funcionam coerentemente (Cruse, 1995, p. 240-1).

5.2.2.2 Contrastes não-binários

Os opostos não binários são caracterizados pela relação de incompatibilidade (Lyons, 1977, p. 288); desta forma, apresentam-se em conjuntos de elementos lexicais cujos membros se caracterizam pela relação de conteúdos incompatíveis. Ex: conjunto de dias da semana: segunda-feira, terça-feira, quarta-feira, etc; Dizer “*Hoje é quarta-feira*” implica “*Hoje não é domingo, segunda, terça, etc.*” O importante a salientar é que a incompatibilidade é uma relação lexical, e como a oposição, está baseada no contraste dentro de um certo grau de semelhança (Lyons, 1977, p. 288).

Dividem-se em dois grupos: conjuntos seriais e conjuntos cíclicos.

5.2.2.2.1 Conjuntos seriais

Formados por conjuntos em que cada membro tem o seu lugar na série entre dois outros membros, em que há um membro inicial e um membro final (Vilela, 1994, p. 172). Subdividem-se em:

- **escalas** – conjuntos cujos membros são graduáveis, como {*quente, morno, frio, gelado*} (Lyons, 1977, p. 289);
- **graus** – conjuntos cujos membros não são graduáveis, como notas escolares, designações dos graus dos militares, números, etc. (Vilela, 1994, p. 172)

5.2.2.2.2 Conjuntos cíclicos

Conjuntos em que cada membro tem o seu lugar no ciclo entre dois outros membros em que há uma seriação. Ex: estações do ano, meses, etc. (Vilela, 1994, p. 172).

5.2.2.3 Antônimos múltiplos ou lexemas com mais de um antônimo

Existem casos em que um lexema apresenta mais de um antônimo. Assim sendo, são duas as hipóteses para tratamento de tais casos (Vilela, 1994, p. 173-4):

- a) lexema polissêmico – atribui-se um antônimo diferente a cada uma das variantes de conteúdo do lexema. Ex:

libertino	--	debochado	--	casto
libertino	--	irreligioso	--	religioso/crente

- b) interpretação homonímica – o antônimo é o selecionador do homonímico:

duro	--	duro1	--	mole (colchão)
	--	duro2	--	compassivo (coração)
	--	duro3	--	flácido (pele)
	--	duro4	--	flexível (pessoa)

.....

5.2.2.4 Antonímia e conceitos

Antônimos não possuem a mesma forma semântica, como demonstraremos a partir do par de adjetivos *alto* e *baixo* (Bierwish e Schereuder, 1992, p. 38):

/alto/ [+V+N] (λz) λx [QUANT MAX $x = v$ PLUS z]

/baixo/ [+V +N] (λz) λx [QUANT MAX $x = v$ MINUS z]

A forma semântica de *alto* representa a condição de que a quantidade da dimensão máxima de x é igual a algum valor v de comparação mais uma diferença de grau z . Tanto x como z estão ligados por abstrações que especificam posições de

argumentos, de modo que *alto* é um predicado de dois lugares que especifica uma relação entre um objeto e o grau de sua altura. Os parênteses em λz indicam a opcionalidade deste argumento, pois *alto* pode de fato aparecer como um predicado de um lugar com o grau de altura sendo computado a partir de informação contextual como em *John é alto* oposto a *John tem dois metros de altura*. A variável v não está ligada a nenhuma posição argumental. É um parâmetro livre a ser fixado em zero se possível, ou em outro caso, uma norma N dependente de contexto. A entrada de *baixo* difere apenas pela diferença negativa representada pelo funtor MINUS, que torna este antônimo o elemento marcado em relação a *alto*, com conseqüências características para uma possível escolha de valores para ambos z e v (Bierwish & Schreuder, 1992, 38-9).

5.3 As relações de sentido e as classes de palavras

A metodologia para a elaboração da base lexical do *thesaurus eletrônico*, como já mencionado neste trabalho, buscou subsídios na metodologia de construção da base lexical da rede *WordNet*.

Em primeiro lugar, a equipe de desenvolvimento decidiu que categorias lexicais deveriam ser incluídas na *WordNet*; foram escolhidos as classes maiores (substantivos, adjetivos, advérbios e verbos) e cada classe sintática recebeu um tratamento formal apropriado (Miller; Fellbaum, 1991, p. 199). Foi necessário verificar como se dá a organização das relações de sentido de cada uma dessas classes, pois essas relações se estabelecem entre as unidades de cada uma dessas classes de forma específica; assim, foi possível mapear os princípios básicos de organização das quatro classes selecionadas.

A relevância deste levantamento para a consecução da base lexical do *thesaurus eletrônico*, que difere da *WordNet* em termos de organização e objetivos, está no fato de

que saber como se organizam, do ponto de vista computacional, as diferentes classes consideradas, é fundamental na extração de dados dos dicionários impressos e sua inserção na base lexical, como demonstraremos logo a seguir, pois, a exemplo do que fez a equipe da *WordNet*, propomos também esta estratégia semelhante para a elaboração de um *thesaurus eletrônico*.

5.3.1 A organização semântica das classes de palavras consideradas

5.3.1.1 Substantivos

Muito já se especulou acerca da organização semântica dos substantivos. O que geralmente se concorda é que eles estão organizados hierarquicamente em níveis, do mais específico ao mais genérico, refletindo a propriedade denominada herança lexical. Desta forma, o *design* da base de substantivos da *WordNet* baseou-se nesta propriedade, basicamente por ser computacionalmente tratável, economizar espaço e por acabar com o problema da circularidade das informações encontradas num dicionário.

Um sistema de herança lexical leva em conta a relação de **hiperonímia**, como no exemplo:

carvalho @ → árvore @ → planta @ → organismo

em que @→ é a relação “é um tipo de”, transitiva e simétrica. Através da criação desta hierarquia, é possível relacionar outras unidades a um mesmo superordenado, como, por exemplo, *mogno*, *cedro*, *ipê*, etc, que teriam ponteiros que os conectassem com *planta* e *organismo*. Estes últimos estariam por sua vez associados a um conjunto de primitivos semânticos (uma espécie de ontologia básica, semelhante àquela proposta por Jackendoff), a partir dos quais estão organizados todos os substantivos na *WordNet* (Miller *et al*, 1993, p. 16):

{ato, ação, atividade}	{objeto natural}
{animal, fauna}	{fenômeno natural}
{artefato}	{pessoa, ser humano}
{atributo, propriedade}	{planta, flora}
{corpo, corpus}	{possessão, propriedade}
{cognição, concepção}	{processo}
{comunicação}	{quantidade, quantia}
{evento, acontecimento}	{relação}
{sentimento, emoção}	{forma}
{comida}	{sociedade}
{grupo, coleção}	{estado, condição}
{localidade, lugar}	{substância}
{motivo}	{tempo}

O número de níveis destas hierarquias raramente vai além de dez, e nesses casos ocorre o vocabulário técnico. Por exemplo, um *pônei Shetland* é um *pônei*, um *cavalo*, um *eqüideo*, um *perissodáctilo*, um *herbívoro*, um *mamífero*, um *vertebrado* e um *animal*: nove níveis, dos quais cinco são técnicos (Miller e Fellbaum, 1991, p. 205) . Sendo assim, associando unidades de níveis menores a unidades maiores na hierarquia, elimina-se a necessidade de informação redundante, reduzindo consideravelmente o tamanho da base de substantivos. As terminologias, ou “vocabulário técnico” passam a constituir um subsistema da *Wordnet*.

A herança lexical, no sentido de Pustejovsky (1995), é uma propriedade fundamental na organização dos nomes, da mesma forma que as relações de hiperonímia, meronímia e até certo ponto, a antonímia são igualmente relevantes.

Embora a maior parte da estrutura das hierarquias de substantivos seja gerada pela **hiponímia**, as distinções são dadas por traços que diferenciam um conceito de outro. Por exemplo, um *canário* é um *pássaro* que é *pequeno*, *amarelo*, que *canta* e *voa*, portanto todas essas características devem aparecer quando *canário* for ativado. E mais, *canário* herda de *pássaro* o fato de ter *bico* e *asas com penas*. A fim de refinar a informação da base lexical, seriam necessárias mais três tipos de informação:

- (1) Partes: bico, asas

(2) Atributos: pequeno, amarelo

(3) Funções: cantar, voar

Note-se que as partes são dadas por substantivos, atributos por adjetivos e funções por verbos. Sendo assim, associar a base de substantivos às outras bases daria à *WordNet* maior acuidade de informação; no momento, apenas as partes estão relacionadas dentro da base, pois se trata de uma relação entre substantivos (Miller *et al*, 1993, p. 18).

Bico e *asas* são merônimos de *pássaro*, portanto os merônimos podem servir como traços que os hipônimos podem herdar. Conseqüentemente, **meronímia** e **hiponímia** são intercambiáveis de maneira complexa.

Um problema freqüente ao se tentar estabelecer uma relação apropriada entre hiponímia e meronímia surge de uma tendência geral em se juntar traços muito altos na hierarquia. Por exemplo, se *roda* é considerada merônimo de *veículo*, então o *trenó* e a *nave espacial* herdarão rodas que não deveriam ter (Miller; Fellbaum, 1991, p. 206).

Quanto aos atributos, os adjetivos modificam substantivos, ou os substantivos servem de argumentos dos atributos: *tamanho (canário) = pequeno*. Para fins teóricos, o fato de que um *canário* é *pequeno* pode ser representado por um ponteiro, da mesma forma que *canário* → *pássaro*. Formalmente, a diferença é que não há ponteiro de retorno de *pequeno* em direção a *pássaro*. O ponteiro de *canário* para *pequeno* é interpretado com respeito ao superordenado imediato de *canário*, ou seja, de *pequeno* para *pássaro*, mas a conexão para um substantivo é perdida quando *pequeno* é acessado sozinho (Miller *et al*, 1993, p. 19).

Quanto às funções, no modelo da *WordNet*, “função” refere-se a descrição de algo que exemplifique o que o referente de um conceito prototipicamente faz, ou o que é normalmente feito com ou para eles. Parece funcionar melhor com *caneta (função:*

escrever). Dizer que a função de um *canário* é *voar* é um tanto quanto infeliz. O termo é usado na falta de um termo mais adequado. É algo mais aparente quando se trata de {ornamento, decoração}, por exemplo. Um ornamento pode ser de qualquer tamanho, forma ou composição; partes e atributos falham na busca do significado. Mas a função do ornamento é mais clara: tornar algo mais atraente; existem razões lingüísticas para assumir que a função de uma “coisa” é um traço de seu significado (Miller; Fellbaum, 1991, p.208).

No entanto, a “função” não é ainda muito consistente para a semântica lexical. Ficamos com a pergunta dos autores: afinal, qual é a função de “maçã” ou “gato”? (Miller *et al*, 193, p. 23). Essa questão, sem dúvida, é também um problema a ser resolvido no modelo de Pustejovsky (1995), posto que o *qualia* télico trata exatamente da especificação da “função” das entidades.

A antonímia e os substantivos

Embora não seja uma relação fundamental na organização dos substantivos, a **antonímia** merece menção. Está presente principalmente nas relações entre nomes deadjetivais, como *felicidade:infelicidade* e nas relações de parentesco, como *genro:nora, pai:mãe, filho:filha*, etc (Miller *et al*, 1993, p. 24).

Quando os três tipos de relações fundamentais – **hiponímia**, **meronímia** e **antonímia** - estiverem representadas na base de substantivos da *WordNet*, o resultado será uma rede altamente interconectada de itens lexicais, refinando o conhecimento acerca da organização da classe e contribuindo para os estudos do léxico.

5.3.1.2 Adjetivos e advérbios

Na consecução da *WordNet*, assumiu-se que os **advérbios** possuem a mesma organização semântica dos **adjetivos**. A maioria dos adjetivos pode ser usada atributivamente, mas muitos não o são; em dicionários, os adjetivos não-predicativos são geralmente definidos com algo do tipo “de ou pertencente a” algo.

A organização semântica dos adjetivos é bem diferente dos substantivos. A relação semântica básica entre eles é a **antonímia**. A importância da antonímia é compreensível quando se reconhece que a função dos adjetivos predicativos é expressar valores dos atributos, e que a maioria dos atributos é bipolar. Os antônimos são adjetivos que expressam valores em pólos opostos de um atributo.

No entanto, ainda ficam algumas questões:

1. se a antonímia é tão importante, por que há tantos adjetivos que parecem não ter antônimos?

2. quando dois adjetivos têm significados parecidos, por que eles não têm o mesmo antônimo?

O problema aqui é que a relação de antonímia entre as formas é diferente da relação de antonímia entre os significados. No caso do inglês, o que se percebe é que a antonímia é mais uma relação entre formas, posto que o prefixo *un-* se presta ao serviço de dar o antônimo de muitos adjetivos (ex: *able:unable*). Observa-se o mesmo para o português; vejamos o caso do adjetivo *fácil*; no dicionário Aurélio do Séc. XXI, a terceira acepção de *fácil* fornece o seguinte:

3. *Claro, simples, vulgar, natural: linguagem fácil; tem o estilo fácil.*

Verifica-se que *fácil* tem como sinônimos *claro, simples, vulgar e natural*. No entanto, verifica-se também que o dicionário fornece um só antônimo para cada uma das doze acepções de *fácil*, que é *difícil*. Percorrendo o caminho dentro dos sinônimos,

procuramos a acepção correspondente ao sentido de *fácil* na terceira acepção e verificamos se havia alguma menção ao antônimo *difícil*. Encontramos o seguinte:

Entrada	Acepção correspondente	Antônimo
Claro	16. Fácil de entender: letra clara; sentido claro.	Não há
Simples	4. Formado de poucos elementos, de fácil utilização ou compreensão; que não apresenta complexidade ou dificuldade; singelo: mecanismo simples, problema simples.	Não há
Vulgar	Não há acepção correspondente.	-
Natural	Não há acepção correspondente.	-

Ou seja, nenhuma das acepções encontradas apontou para o antônimo *difícil*, que permanece como antônimo somente de *fácil*, embora o dicionário mostre que *fácil* é sinônimo de *claro*, *simples*, *vulgar*, *natural*. O mesmo problema foi enfrentado pela equipe da *WordNet*, que daqui extraiu um dos princípios constituintes da base lexical: se *fácil* é similar a *claro*, *simples*, *vulgar*, *natural*, e *fácil* é antônimo de *difícil*, então o conjunto $\{fácil, claro, simples, vulgar, natural\}$ adquire, em relação a $\{difícil\}$, a condição de opostos conceptuais, mediados por *fácil* (cf. Miller e Fellbaum, 1991, p. 211). Desta forma, confirma-se a hipótese de que a antonímia é mais uma relação entre formas do que entre conceitos não é uma correspondência direta entre, por exemplo, *vulgar* e *difícil*: é uma relação de **oposição de conceitos**, por essa razão especificamos $\{difícil\}$ como antônimo de outro conjunto.

Por vezes, os pares antônimos são meramente frutos de “acordos”: aquele que conhece o inglês, por exemplo, reconhece em *weighty:weightless* (pesado/leve) um par de antônimos, mas ficarão confusos ao deparar com *heavy:weightless* (pesado/leve). Os significados são opostos, mas as formas não são familiares enquanto pares de antônimos (Miller; Fellbaum, 1991, p. 211). Podemos dizer o mesmo para o português, naquele caso de *fácil:difícil* e *simples:difícil*, sendo este último menos reconhecido que o primeiro.

Quanto às preferências de seleção, os **adjetivos** são seletivos em relação aos **substantivos** que modificam. A regra geral é que se o referente denotado por um substantivo não tem o atributo cujo valor é expresso pelo adjetivo, então a combinação nome-adjetivo requer uma interpretação figurativa ou idiomática. Ex: um prédio ou uma pessoa podem ser *altos*, pois possuem *altura* como atributo, mas *ruas* e *jardins* não têm altura, portanto *estradas altas* e *jardins altos* não admitem leitura ao pé da letra. Alguns adjetivos possuem aplicabilidade específica; portanto, um dos projetos para a *WordNet* é associar os adjetivos com suas preferências seletivas, adicionando um *link* entre eles e substantivos, algo que ainda não está implementado. Um exemplo seria associar o adjetivo relacional *paterno* ao substantivo *pai*.

5.3.1.3 Verbos

Os verbos possuem uma organização diferente da dos substantivos, adjetivos e advérbios. A principal relação lexical na organização verbal encontrada pela equipe da *WordNet* foi a princípio a **sinonímia**; depois desta, uma outra relação importante seria a hiponímia. No entanto, diante da dificuldade em se aceitar “*balbuciar* é um tipo de *falar*”, a solução foi recorrer ao componente semântico MODO e criar a relação chamada de **troponímia**, “x é y de certo modo”: então teríamos “*balbuciar* é um modo de *falar*”.

Uma terceira relação fundamental para os verbos é a própria **antonímia**, ocorrendo entre formas opostas morfologicamente (*acelerar:desacelerar*), verbos deadjetivos (*alongar:encurtar*), opostos diretos (*caminhar:correr*), opostos vinculados (*vencer:perder*, vinculando *jogar*), opostos reversivos (*comprar:vender*), indicativos de mudança de estado (*acordar:adormecer*), etc. (Miller; Fellbaum, 1991, p. 222-4).

Conhecer as relações de sentido prototípicas na organização das quatro classes sintáticas consideradas para a montagem da base lexical de um *thesaurus eletrônico* é de fundamental importância, pois, já que as únicas relações a serem representadas são a sinonímia e a antonímia.

CAPÍTULO 6 - A BASE LEXICAL DO *THESAURUS* ELETRÔNICO

Neste capítulo buscamos investigar maneiras de se representar as relações de sentido da sinonímia e antonímia com vistas à montagem da base lexical do *thesaurus*. Essencialmente, como já mencionado, a solução veio da adoção dos princípios utilizados na montagem da base lexical da *WordNet*, um modelo de representação do léxico de língua inglesa, de cuja metodologia utilizamo-nos de três noções essenciais (Dias-da-Silva *et al*, 2000, p. 06):

- a) adoção do **método diferencial**, que pressupõe a ativação de conceitos através de um conjunto de formas lexicais relacionadas pela sinonímia, eliminando a necessidade de haver um rótulo conceitual para cada acepção de uma entrada;
- b) a noção constitutiva básica de ***synset*** (conjunto de sinônimos);
- c) a noção de **matriz lexical**, que estabelece uma correspondência biunívoca entre sentido e *synset*.

6.1 Sinonímia e antonímia: estrutura de sua representação

De acordo com essa metodologia, considera-se que a sinonímia é uma relação que se estabelece entre unidades lexicais e não entre conceitos e que, se o critério para a validação da sinonímia é o da possibilidade de substituição, então é uma relação intracategorias, ou seja, entre substantivos e substantivos, verbos e verbos, e assim por diante (Pustejovsky, 1994, p. 23); ou seja, é absolutamente pertinente validar o critério da possibilidade de substituição para cada uma das classes sintáticas consideradas na montagem da base do *thesaurus eletrônico*.

A intuição de que é normal para um falante do inglês, por exemplo, considerar antônimos os pares *heavy:light* e *weighty:weightless*, mas não pares como

heavy:weightless ou *ponderous:airy*⁷, é captada na rede *WordNet* considerando-se que os adjetivos estruturam-se como aglomerados (*clusters*) formados por similaridade semântica e "gravitando" em torno de um adjetivo central (ou mais de um) que se opõe a um outro adjetivo, centro de outro aglomerado. Desta forma, *ponderous* é similar a *heavy* e *heavy* é antônimo de *light*. Portanto, uma **oposição conceitual** entre *ponderous/light* é mediada por *heavy* (Miller; Fellbaum, 1991, p. 211):

Conjunto de sinônimos {*ponderous, heavy*}

Conjunto de antônimos {*light*}

Portanto, a construção da base lexical do *thesaurus eletrônico* deve partir do princípio da similaridade/oposição conceitual. Diante disso, é necessário esboçar um esquema visual para a interface⁸ de inserção de dados na base lexical, acompanhado de testes de viabilidade dessa proposta para o português, a partir da montagem de conjuntos de sinônimos e antônimos que apontem para um átomo conceitual em comum.

6.2 Proposta de modelo para a interface de inserção de dados

A fim de se visualizar a estrutura proposta para a interface, em termos de conjuntos de sinônimos atrelados a conjuntos de antônimos, propomos um esquema semelhante ao da Figura 11.

⁷ Todos os pares referem-se à oposição *pesado:leve*. Em português, podemos citar o caso de *fácil:difícil* e *simples:difícil*, conforme capítulo três desta dissertação.

⁸ Em um sistema computacional, a interface gráfica é um conjunto de elementos de *hardware* e *software* destinados a possibilitar a interação com o usuário, (...) baseada no amplo emprego de imagens, (...) e que faz uso de um conjunto de ferramentas que inclui janelas, ícones, botões, e um meio de apontamento e seleção, como o mouse. (...) (*Dicionário Aurélio do Séc. XXI*).

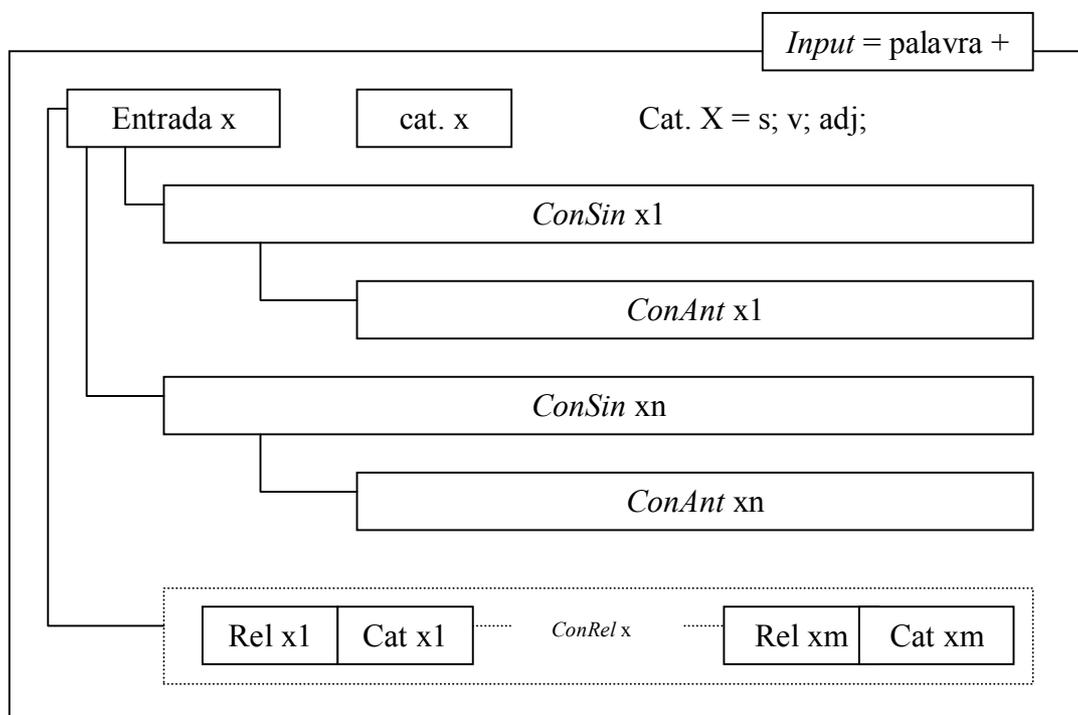


Fig. 11. Esquema de estruturação para a implementação do editor do *thesaurus*, prevendo a estrutura de conjuntos de sinônimos e antônimos (Dias-da-Silva *et al*, 2000).

Note-se que no esquema temos a *Entrada*, a especificação de sua categoria (verbo, advérbio, adjetivo ou substantivo) e, atrelado à entrada, um conjunto de sinônimos (*ConSin*), ao qual pode estar vinculado um conjunto de antônimos (*ConAnt*). Cada par *ConSin/ConAnt* constitui uma aceção da entrada, que pode ter várias aceções diferentes (*ConSinxn/ConAntxn*).

O *thesaurus eletrônico* que vislumbramos, além de possibilitar ao usuário opções de substituição em contextos afins, deve também atender a exigências formais, ou seja, demanda uma formalização dos significados que seja computacionalmente tratável, e que possibilite acesso rápido a todos os sinônimos e antônimos sistematicamente compilados. Portanto, com a adoção do modelo de representação proposto para a

WordNet, a tarefa de compilar a base lexical do *thesaurus* é consideravelmente reduzida. Em termos operacionais, reduz-se à montagem dos conjuntos de sinônimos e de antônimos. Desta forma, a estrutura básica da base deve conformar-se ao seguinte esquema (Dias-da-Silva e Moraes, 2000, p. 03):

Entrada n (categoria x)

Acepção n.1

Conjunto de sinônimos

Conjunto de antônimos

...

Acepção n.n

Conjunto de sinônimos

Conjunto de antônimos

Definida a forma de representação dos conceitos em termos de sinonímia e antonímia através das noções de *synsets* e da matriz lexical, é necessário refletir sobre a parte empírica da metodologia, ou seja, sobre como aplicá-la e de que fatores depende sua aplicação.

A opção por trabalho em equipe justifica-se pela escassez de tempo, se pensarmos em projetos financiáveis, da distribuição de computadores entre pesquisadores e do necessário revezamento de horários de trabalho. É preciso, portanto, compor uma equipe coordenada, que englobe lingüistas e cientistas da computação, dado o caráter de implementação computacional das relações semânticas da sinonímia e antonímia para a constituição da base lexical do *thesaurus eletrônico*.

Conforme visto no capítulo quatro, a montagem de bases lexicais torna-se viável a partir da extração de unidades lexicais dos dicionários de língua; portanto, é necessário constituir um *corpus de referência*, composto de dicionários de língua portuguesa. Além disso, é necessária a especificação de parâmetros de filtragem para a compilação de sinônimos e antônimos a partir dos dicionários que constituem esse *corpus*.

Do ponto de vista operacional, a metodologia que propomos neste trabalho, que foi efetivamente aplicada, conforme veremos no capítulo final desta dissertação, prevê a divisão do trabalho de compilação dos conjuntos de sinônimos e antônimos entre três lingüistas, sendo o primeiro responsável pela compilação da base de verbos, o segundo, pelos substantivos, e o terceiro, pelas bases de advérbios e adjetivos, munidos de uma lista de parâmetros elaborados com base na tipologia das relações de sentido discutidas no capítulo anterior e em critérios heurísticos, que apresentaremos oportunamente. Esses parâmetros norteiam a filtragem dos itens lexicais a partir do *corpus de referência*. Além disso, deve-se prever também um cientista da computação cuja tarefa é dedicar-se à implementação computacional das relações da sinonímia e antonímia, à programação da interface para o armazenamento e gerenciamento dos dados lexicais e, conseqüentemente, à elaboração de uma estratégia computacional que viabilize a concretização da metodologia proposta.

Assim sendo, para que todo o empreendimento tome corpo, é necessário um trabalho prévio ao desenvolvimento da interface computacional; trata-se de testes de viabilidade da estrutura proposta para o *thesaurus* em termos de conjuntos de sinônimos e antônimos, testes que visem verificar a validade da extração da informação lexical dos dicionários-fonte e ao mesmo tempo resultar em uma proposta de um algoritmo para o

desenvolvimento da arquitetura do sistema. Tal trabalho deve ser realizado em conjunto pela equipe de linguistas e de cientistas da computação.

A tarefa de compilação de dados para a base do *thesaurus* é de grande complexidade, demandando da equipe uma dedicação ao exame dos passos a serem tomados na extração de dados, viabilidade da proposta de estrutura de representação dos conjuntos de sinônimos e antônimos e futura implementação de uma interface. Vamos apresentar, na seqüência, os passos que a equipe deve seguir.

6.3 Delimitação do *corpus* de referência

Diante da complexidade de se explorar questões no âmbito do léxico, conforme fica evidenciada pelas discussões apresentadas nos três primeiros capítulos, das dificuldades de se encontrar um conceito de palavra, das diversas facetas do significado e, conforme lembra Dias-da-Silva et al (2000, p.05), da investigação do léxico, que consiste em enfrentar questões de natureza fonético-fonológica, passando pelas questões morfossintáticas, culminando com as complicadíssimas questões semânticas e pragmáticas, encontramos uma solução prática na extração de informações a partir de fontes já disponíveis. Tal decisão possibilita a agilização dos trabalhos, reduzindo a tarefa, conforme já foi dito, à extração e filtragem de informação das obras de referência escolhidas, resultantes de anos de dedicado trabalho lexicográfico e lingüístico. Além disso, ao adotarmos essas obras, eliminamos, em um primeiro momento, a necessidade de se montar cada verbete a partir do zero e pesquisar em *corpora*. Sugerimos como fonte de referência sete obras para a compilação da base lexical inicial do *thesaurus eletrônico*:

1 – NASCENTES, A. *Dicionário de sinônimos* – Apresenta as entradas sob a forma de conjuntos de sinônimos, seguidos de uma definição para cada um dos constituintes do

conjunto. Em alguns casos apresenta exemplos retirados de *corpora*, com citações e indicações bibliográficas; no prefácio dos editores, explica-se que as definições aqui encontradas são baseadas em trabalho anterior do próprio autor. A obra conta ainda com um índice remissivo.

2 – AZEVEDO, F. F. S. – *Dicionário analógico da língua portuguesa* – Possui como subtítulo “O primeiro *thesaurus* da língua portuguesa”, tendo sido elaborado aos moldes do *Thesaurus* de Roget (1852), conforme citação da própria “Apresentação” da obra: “(...) dada uma idéia, indica as palavras que podem expressar essa idéia ou que com ela têm analogia (...) 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”. Desta forma, a obra reúne 1000 conceitos classificados em seis grupos básicos (*relações abstratas, espaço, matéria, intelecto, vontade e afeições*) e se utiliza de rótulos gramaticais (adjetivo, advérbio, etc.) e estilísticos (brasileirismo, desusado, familiar, irônico, etc) na classificação de suas acepções.

3 – BORBA, F. S. (org.) *Dicionário gramatical de verbos do português contemporâneo do Brasil* – Obra exclusivamente dedicada à classe dos verbos, reunindo aproximadamente 6.000 deles sob a forma de entradas; fornece informações sobre a transitividade do verbo, relações entre predicado e argumentos (o que permite a classificação em quatro tipos de verbos – ação, processo, ação processo e estado), papéis dos argumentos (experimentador, beneficiário, afetado, paciente, efetuado, causativo, agente), especificação de mecanismos sintáticos individualizantes como a forma pronominal; apresenta ainda indicação de tempos verbais, topicalização, volição e intencionalidade, além da classificação do tipo de sujeito (agente, causativo, paciente ou inativo), quando apropriado. “Cada verbete segue uma ordem fixa. A descrição começa pelos valores que indicam ação-processo; seguem, na ordem, os que indicam processo,

ação, estado, auxiliaridade, modalização. No final, aparecem as expressões. Estabeleceu-se essa ordem porque os verbos de ação-processo são os mais numerosos da língua; as construções processivas são, em geral, derivadas das construções que indicam ação-processo” (pág. XVI). O dicionário ainda utiliza-se de abreviaturas ou rótulos como *automobilismo, coloquial, cinema, economia, etc.*

4 – FERNANDES, F. *Dicionário de sinônimos e antônimos da língua portuguesa* –

Obra que reúne sinônimos de 30.000 palavras, incluindo neologismos, brasileirismos e estrangeirismos, ainda que não explicitamente classificados desta forma; os únicos “rótulos” utilizados foram *antônimos, desusado, galicismo, pouco usado, pejorativo, plural, poético e sinônimo*. Na apresentação da primeira edição, de 1944, o autor afirma que o dicionário está organizado também lançando mão da analogia, pois os sinônimos muitas vezes se relacionam com outro “por simples vínculo analógico”. Em algumas entradas é possível encontrar exemplos retirados de *corpora* e formas pronominais; a remissão é feita dentro da própria entrada sob a forma “o mesmo que....” e não há índice remissivo.

5 – BARBOSA, O. *Grande dicionário de sinônimos e antônimos* –

O dicionário conta com entradas seguidas de seus sinônimos e antônimos, quando existirem; não há uso de rótulos além do indicativo de antonímia. Ao final apresenta-se um apêndice com “Antônimos de palavras sem sinônimos diretos”, como *alto-relevo/baixo relevo, eufonia/disfonia, etc.* Algumas entradas apresentam exemplos de ocorrência, sem citação da fonte de origem.

6 - MICHAELIS PORTUGUÊS – MODERNO DICIONÁRIO DA LÍNGUA

PORTUGUESA - Versão 1.0 – Reunindo mais de 200.000 verbetes e subverbetes, esta nova edição conta com um número maior de vocábulos novos, especialmente dos domínios das ciências e da tecnologia. Suas entradas apresentam categoria gramatical,

etimologia, registro, acepções, exemplos, subverbetes, expressões, formas irregulares e remissões. A edição impressa inclui um apêndice incluindo notas gramaticais, noções de lingüística e palavras latinas, além de utilizar-se de um grande número de rótulos (abreviaturas).

7 – NOVO DICIONÁRIO AURÉLIO ELETRÔNICO – SÉCULO XXI – VERSÃO

3.0 – Agora em versão eletrônica, o *Aurélio do séc. XXI* foi “ampliado em quase 25% e totalmente atualizado (...) com suas novas palavras e seus novos usos. (...) traz novas funções que facilitam ainda mais a consulta e multiplicam as possibilidades de acesso à informação: pesquisa reversa total, pesquisa de categorias gramaticais, pesquisas no âmbito de locuções, etimologias, exemplos e abonações e elementos de composição, além de uma nova interface, mais simples, mais ágil, mais fácil de usar.” (Prefácio). A obra conta com cerca de 345 mil verbetes, locuções e definições, dezenas de milhares de exemplos e abonações literárias (que abrangem 1.400 autores). A escolha da versão eletrônica é decorrente da própria metodologia deste trabalho, e encontrou eco nas seguintes inovações da interface de consulta: a) possibilidade de obter de listas de verbetes selecionados segundo critérios definidos pelo próprio usuário; b) navegação pelos verbetes, no conceito de hipertexto, podendo cada palavra que aparece na explicação dos significados levar ao verbete que a define; c) dicionário reverso, pelo qual a partir de palavras-chaves, contidas nos significados, obtêm-se os respectivos verbetes. “Também mereceu cuidado especial a apresentação, na tela, do significado dos verbetes. Aproveitaram-se os recursos dos monitores de vídeo, como a cor, para destacar as diferentes partes que compõem o texto (etimologia, categoria gramatical, rubricas, achegas, remissivas, exemplos e abonações).” (Apresentação).

A extração de sinônimos e antônimos a partir dos verbetes encontrados no *corpus de referência* é bastante produtiva; no entanto, é preciso levar em conta alguns problemas encontrados na própria constituição do significado no dicionário, conforme discutido do capítulo quatro. Portanto, para que se extraísse informação consistente do *corpus de referência*, elaboramos uma série de parâmetros para nortear o lingüista e padronizar sua tarefa de extração/inserção de dados na base lexical do *thesaurus*. Apresentamos brevemente os parâmetros adotados, a seguir.

6.4 Parâmetros para filtragem de informações do *corpus de referência*

Durante o processo de montagem dos conjuntos de sinônimos e antônimos da base do *thesaurus eletrônico*, o lingüista deve estar atento para dois aspectos: aqueles inerentes à linguagem e aqueles de ordem prática.

- 1) Critério motivado pela natureza da linguagem: observar a variedade do fenômeno da sinonímia e antonímia arrolada no capítulo cinco , conforme as tabelas 3 e 4 abaixo, que sintetizam brevemente tipologias e exemplos.

Tabela 3. Breve tipologia da sinonímia.

TIPO DE SINONÍMIA	EXEMPLOS
Absoluta	Malária – maleita – impaludismo
Cognitiva	Bebê – neném
Parcial	Chefe – patrão
Absoluta mas não total	Principiante – debutante
Completa mas não total	Cara – face – rosto
Incompleta e não total	Roubar – assaltar
Conotativa	Chefe – patrão
Denotativa	Autor – escritor

Tabela 4. Breve tipologia das relações de contraste.

TIPOS DE CONTRASTE	EXEMPLOS
Antonímia	Menor – maior
Complementaridade	Macho – fêmea
Reciprocidade	Marido – esposa
Oposição direcional	Chegar – partir
Conjuntos seriais	Quente – morno - frio – gelado
Conjuntos cíclicos	Janeiro – fevereiro – março – etc.

2) Critérios de ordem prática: observar os seguintes critérios heurísticos:

- Considerar somente as unidades lexicais dicionarizadas
- Na extração dos sinônimos e antônimos da definição de dicionário, observar:

- **Paráfrases** – por vezes será possível que a paráfrase seja substituída pelo verbo correspondente. Ex: *Tornar mais amplo = ampliar; colocar-se ou dispor-se em linha ou fileira = enfileirar*. Uma consequência de tal procedimento é a inclusão de verbos novos no corpo do verbete, enquanto sinônimos da entrada.
- **Adjuntos** – merecem atenção especial pois carregam traços como aspecto, modo, repetição, etc.; no entanto prevalece seu caráter acessório, portanto podem ser retirados da frase sem alterar o sentido da mesma. Geralmente não serão dispensados os adjuntos de modo e frequência; adjuntos de tempo e lugar podem ser dispensados.
- **Advérbios** – no caso da montagem da base dos verbos, dispensar os advérbios das definições. Ex: *Discorrer ou escrever amplamente sobre um assunto*. Ficamos com *discorrer e escrever*.
- **Regência** – tirar a especificação da regência e considerar o verbo. Ex: *dispor de = dispor, visar a = visar*.
- **Sintagmas preposicionais** – num primeiro momento, dispensar. Ex: *de cor bassa* (loc. Adjetiva). Considerar apenas as que se referem a advérbios: *à toa, à esmo* (loc. Adverbial.)
- **Uso metafórico** – deve ser considerado se estiver devidamente abonado pelas fontes. Ex: *viajar = divagar*.
- **Variantes** – considerar todas as acepções com a rubrica *variante*.
- **Formas pronominais** – a forma pronominal é especificada como sinônimo de um verbo não pronominal se a informação partir dos dicionários fonte. Em alguns casos, é necessário formar conjuntos apenas com os pronominais e um conjunto

semelhante, com os não pronominais, quando não for possível estabelecer uma relação de sinonímia entre uma forma pronominal e uma não pronominal.

- **Particípio** – no caso do particípio do verbo, se este não existir em nenhum dicionário em forma de entrada, deve-se consultar o verbo de origem e gerar uma entrada para ele.

- **Rubricas** – dispensar as entradas com rubricas de domínios específicos, com exceção de variantes e sentido figurado.

- **Homonímia** – é discriminada por um número apostro à entrada:

atentar 1 = {advertir, atender, notar, observar, reparar, ver}

{considerar, meditar, pensar, ponderar, refletir, reflexionar}

atentar 2 = {cometer, empreender, intentar}

atentar 3 = {tentar} ; {provocar, tentar}

- **Expressões e palavras relacionadas** – dispensar.

- **Relações de sentido** – entram somente sinonímia e antonímia. Portanto, ao verificar as definições de dicionário, dispensar informações do tipo:

✓ Troponímia: Cantar = dizer com certa entonação

✓ Hiponímia: Canário = pequena ave passeriforme, cantora, de plumagem amarela ou verde, bico curto e grosso

✓ Meronímia: Braço = parte do membro superior do homem, compreendida entre a espádua e o cotovelo, por oposição ao antebraço

6.5 O algoritmo

Para testar a proposta de representação em conjuntos de sinônimos e antônimos, elaboramos um algoritmo, ou seja, um conjunto de regras e operações bem definidas e ordenadas.

As etapas do algoritmo proposto para a montagem da base lexical foram baseadas em testes, de verificação da viabilidade da inserção das **entradas**, montagem dos conjuntos de **sinônimos** e **antônimos**, visando dois objetivos:

a) testar a hipótese de Miller; Fellbaum (1991, p.211), pela qual os conjuntos de sinônimos e antônimos seriam montados com base na semelhança/oposição conceitual (ver exemplo de *ponderous* e *heavy*, no capítulo anterior). Por esta proposta, todas as unidades semelhantes seriam colocadas em um mesmo conjunto, caso fosse detectado um antônimo comum a todas elas. Vejamos um exemplo a partir dos dados fornecidos pelo quadro abaixo, com o verbete *fácil*. Se a hipótese de Miller fosse aplicada, teríamos todas as acepções de *fácil* unidas por um antônimo em comum, *difícil*. Isso inviabilizaria a distinção de sentidos para cada uma das entradas do *thesaurus*, já que teríamos a união de todas elas em uma só, devido ao fato de serem relacionadas por um único antônimo.

Fácil

[Do lat. facile.]

Adj. 2g.

1. Que se faz ou se consegue sem custo ou esforço;
2. Que se aprende ou compreende sem custo;
3. Claro, simples, vulgar, natural;
4. Espontâneo, pronto;
5. Dócil, brando, suave;
6. Acessível, lhano, comunicativo;
7. Inclinado, tendente, propenso;
8. Crédulo, ingênuo, confiante;
9. Precipitado, irrefletido, imponderado;
10. Bras. Diz-se de indivíduo de honestidade duvidosa.

V. vida.

Adv.

11. Com facilidade; facilmente, sem demora ou esforço;
12. Sem oferecer dificuldade à compreensão de quem lê ou escuta.

[Ant. difícil. Pl. (do adj.): fáceis; superl. abs. sint. facilímo, facilíssimo.]

b) testar a viabilidade de se unir sinônimos em um mesmo conjunto baseando-se no reconhecimento da forma; se dois conjuntos apresentassem dois elementos em comum, ambos seriam unidos. Ex:

Conj. 1 = {**vaguear**, errar, vagar, **vagabundear**}

Conj. 2 = {**vaguear**, **vagabundear**, vagabundar, vadiar}

$1 \cup 2 = \{\text{vaguear, errar, vagar, vagabundear, vagabundar, vadiar}\}$

Isso, entretanto, não é possível. Vejamos os seguintes conjuntos, formados a partir das entradas **divagar, vaguear, devanear, vagar**⁹:

(3){divagar, vaguear}

(4){divagar, fantasiar, devanear}

(5){vaguear, errar, vagar, vagabundear}

(6){vaguear, vagabundear, vagabundar, vadiar}

(7){vaguear, devanear}

(8){devanear, fantasiar}

(9){devanear, imaginar, fantasiar, sonhar}

(10){vagar, errar, vagabundear, vaguear}

Pelo algoritmo proposto, teríamos a união dos conjuntos com dois elementos iguais, aqui demonstrada passo a passo:

3 {divagar, vaguear}

4 {divagar, fantasiar, devanear}

5, 6 e 10: {vaguear, errar, vagar, vagabundear, vagabundar, vadiar}

7 {vaguear, devanear}

8 e 9: {devanear, imaginar, fantasiar, sonhar}

A seguir, mais uma união:

4,8,9 = {divagar, fantasiar, devanear, imaginar, sonhar}

Teríamos então:

a) {vaguear, divagar}

b) {vaguear, errar, vagar, vagabundear, vagabundar, vadiar}

c) {vaguear, devanear}

d) {divagar, fantasiar, devanear, imaginar, sonhar}

Pelo dicionário de Barbosa (1996), temos os seguintes conjuntos:

{Divagar, vaguear, errar, vagabundear}

{Divagar, sonhar, fantasiar, devanear}

Desta forma, unindo a informação dos dois dicionários, teríamos:

{Vaguear, errar, vagar, vagabundear, vagabundar, vadiar, divagar}

{Divagar, fantasiar, devanear, imaginar, sonhar}

Ainda de acordo com Barbosa (1996) temos o seguinte conjunto:

{Vaguear, devanear, fantasiar}

Como resultado, temos o seguinte:

Vaguear, errar, vagar, vagabundear, vagabundar, vadiar, **divagar**

Divagar, fantasiar, devanear, imaginar, sonhar, **vaguear**

Portanto, com dois elementos em comum, há a união de todos os elementos em um só conjunto:

⁹ Cf. Dicionário *Aurélio do séc. XXI*.

{Vaguear, errar, vagar, vagabundear, vagabundar, vadiar,
divagar, fantasiar, devanear, imaginar, sonhar}

Desta forma, o algoritmo que inicialmente poderia agilizar o processo de montagem da base lexical não se revelou viável pois uniria acepções diferentes em um só conjunto. Evidentemente, no conjunto resultante (se o fôssemos deixar assim) teríamos *vagabundear* como sinônimo de *sonhar*. Logo, cabe ao lingüista examinar constantemente as informações do *corpus* de referência, estando sempre atento às informações cruzadas e cuidar para que o conjunto de sinônimos aponte para um só átomo de sentido, ou seja, sinônimos correspondentes a uma mesma acepção.

Sendo assim, a montagem dos conjuntos fica absolutamente sob a responsabilidade do lingüista, que, seguindo os critérios para filtragem de informações, deve compilar os sinônimos ou antônimos um a um.

Já o armazenamento e gerenciamento dos conjuntos de sinônimos e antônimos montados pelo lingüista são executados automaticamente, por meio de uma interface computacional gráfica, o Editor.

6.6 O Editor do *Thesaurus*

A fim de que os dados extraídos do *corpus* de referência sejam armazenados em forma de conjuntos de sinônimos e antônimos, é necessário programar uma interface computacional que, além de armazenar os conjuntos, permite ao lingüista editar os conjuntos já montados para acrescentar ou excluir elementos.

A partir dos esforços de pesquisadores em torno das questões representacionais e implementacionais, o editor do *thesaurus* (Dias-da-Silva et al, 2000) foi criado para atingir duas finalidades distintas:

- servir de interface computacional gráfica para a montagem da base do *thesaurus*;
- estruturar, armazenar e transformar os dados nela inseridos, função esta de um sistema de gerenciamento de banco de dados (SGBD) (Dias-da-Silva e Moraes, 2000, p. 06).

O SGBD serve à tarefa de unir conjuntos e verificar ambigüidades e inconsistências que possam ocorrer entre os conjuntos, uma vez que o editor não processa todos eles ao mesmo tempo em um mesmo computador. Isso se deve ao fato de que, por questões práticas, devemos prever que cada lingüista pode trabalhar com uma classe sintática diferente, em diferentes máquinas e até mesmo em diferentes horários.

Com a função básica de agilizar a entrada dos conjuntos, permitindo que se edite ou manipule cada verbete, o editor apresenta para o lingüista uma interface gráfica e ferramentas para a edição e gerenciamento das informações durante o processo de montagem da base do *thesaurus*: salvar informações, desfazer uma operação, editar campos, visualizar a base do *thesaurus*, listar entradas, imprimir partes da base do *thesaurus* e extrair dados estatísticos (número de entradas e de conjuntos, proporção entre o número de entradas e o número de conjuntos), entre outras.

Através dessa ferramenta, são criados e editados os conjuntos de sinônimos e antônimos que constituem a base do *thesaurus*, conferindo assim à sinonímia e antonímia uma “existência computacional” (Dias-da-Silva e Moraes, 2000, p. 06). Os arquivos gerados pelo editor armazenam os conjuntos no formato de texto. A **sinonímia** é especificada pela relação de “pertence a”, que se estabelece entre formas da língua e o

synset que as contém; já a **antonímia** é especificada como uma **relação entre pares de *synsets*** (Dias-da-Silva e Moraes, 2000, p. 06, grifo nosso).

Uma das funções do editor que merece destaque é a geração automática de verbetes, recurso pelo qual qualquer sinônimo digitado em um conjunto que representa uma determinada acepção de um verbete qualquer, é automaticamente transformado em entrada pelo editor, que se encarrega também de transportar para essa nova entrada tanto o conjunto em que esse sinônimo está inserido como o conjunto de antônimos, associado a esse conjunto, se houver, gerando um novo verbete (Dias-da-Silva e Moraes, 2000, p. 06).

6.7 O Editor e o Assistente de Edição: aspectos visuais e linhas gerais de funcionamento

A Figura 12 ilustra a interface de inserção de dados do *thesaurus*.

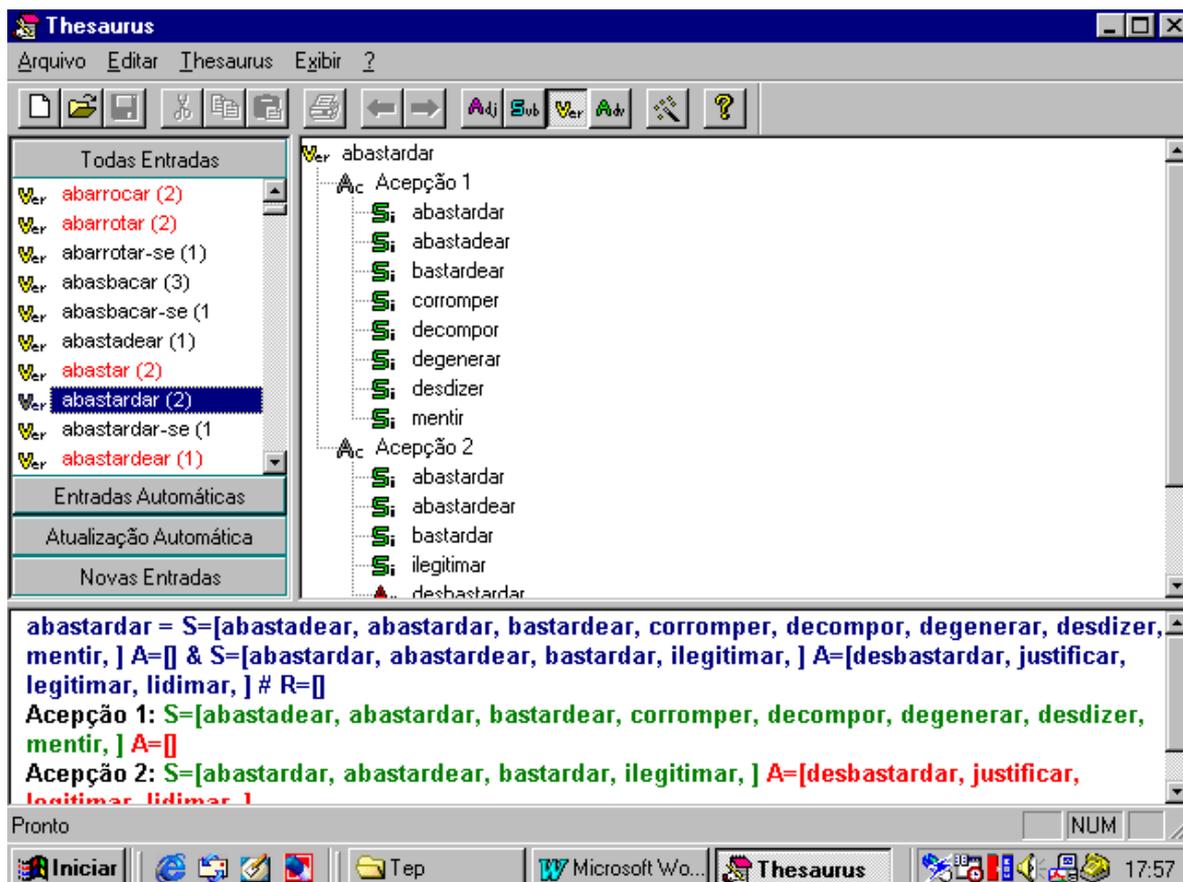


Fig. 12. A interface de inserção de dados no *thesaurus*. No detalhe, a entrada para o verbo *abastardar*.

Pode-se distinguir, além de uma barra de menus, três quadros redimensionáveis, dois superiores (ENTRADAS e VERBETES) e um inferior (CONJUNTOS). O quadro ENTRADAS, superior à esquerda, apresenta a lista, em ordem alfabética, das entradas inseridas na base. O quadro VERBETES, quadro superior à direita, exhibe o verbete completo, estruturado em forma de árvore, referente à entrada selecionada no quadro ENTRADAS. Com exceção do item ACEPÇÃO, que aparece na estrutura do verbete, todos os demais itens desse quadro podem ser renomeados. Ao darmos um duplo clique no item raiz ou nos itens Acepção X, a árvore pode ser expandida ou contraída; ao darmos um duplo clique nos itens terminais (palavras) é possível navegar pela base do *thesaurus* de forma semelhante aos navegadores para a *Internet*. O quadro CONJUNTOS, que ocupa toda a extensão inferior da janela da interface, exhibe todos os conjuntos que contêm uma ocorrência da entrada selecionada no quadro ENTRADAS.

A ferramenta essencial da interface é o **Assistente de Edição**, que auxilia o lingüista no processo de montagem da base do *thesaurus*. É ele que utiliza os filtros para verificar a consistência dos dados e dispara o algoritmo para atualizar automaticamente a estrutura de dados interna. Em termos gráficos, o **Assistente de Edição**, destacado na Figura 13, apresenta três campos para preenchimento e um campo de seleção do número da **Acepção**. Os campos para preenchimento são: **Entrada**, **Palavra ou Lista de Sinônimos** e **Palavra ou Lista de Antônimos**.

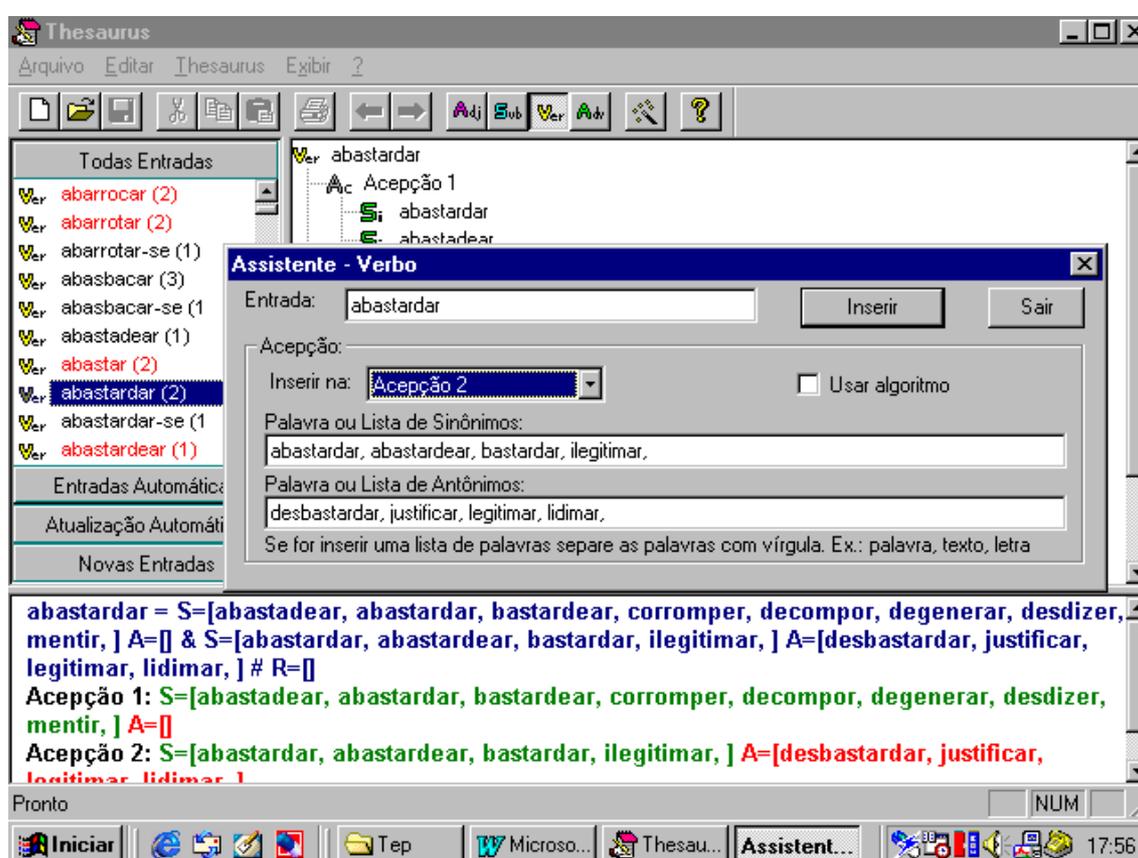


Fig. 13. Em destaque, o Assistente de Edição da interface. Note-se que ao selecionar a acepção (no caso, número 02), o assistente permite que se editem os conjuntos, efetuando acréscimos ou quaisquer outros tipos de alteração.

Após o lingüista digitar uma entrada no campo **Entrada**, o **Assistente de Edição** verifica se essa entrada já existe na base do *thesaurus*. Em caso afirmativo, os campos **Palavra ou Lista de Sinônimos** e/ou **Lista de Antônimos** são preenchidos com as informações recuperadas dessa base, permitindo que esses campos sejam

editados. Em caso negativo, ele cria novos campos. Ao terminar de preencher os campos, o lingüista confirma a entrada dos dados, clicando o botão **Inserir**, o que completa a montagem do verbete.

6.8 A estrutura subjacente da base lexical

Do ponto de vista lógico, a Figura 14 (Dias-da-Silva *et al*, 2000, p. 07) ilustra a forma subjacente da base lexical do *thesaurus*.

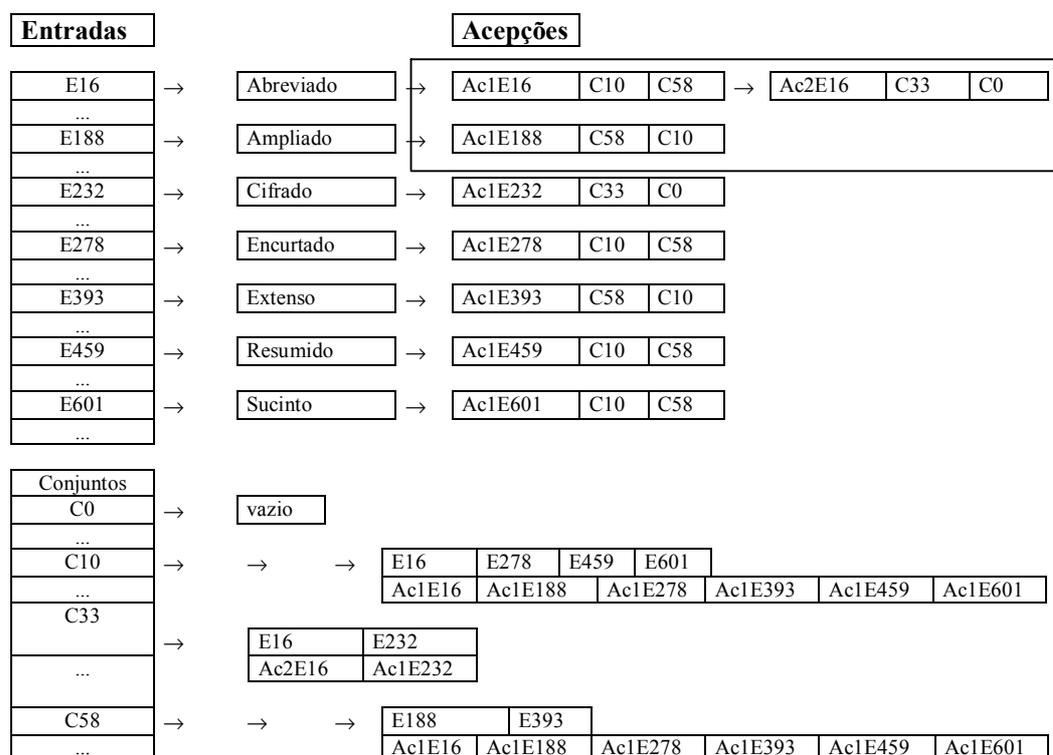


Fig. 14. Exemplo de armazenamento da entrada *abreviado* na base do *thesaurus*. Os índices Exx, Cxx e AcxxEyy são apenas para facilitar a ilustração, pois a estrutura interna utiliza ponteiros (endereços de memória) para indexar esses campos.

A estrutura apresenta duas listas principais: uma lista com todas as **Entradas (LE)**, dispostas em ordem alfabética, e uma lista de **Conjuntos (LC)**, os conjuntos de sinônimos (*synsets*), em que cada conjunto é formado a partir das entradas das acepções a que cada conjunto pertence. Cada entrada da LE, além de apresentar uma representação ortográfica do lexema, contém uma lista de acepções (LA); cada acepção é formada pela dupla de sinônimos e antônimos que apontam para seus respectivos

conjuntos na LC e por um campo que indica a qual entrada cada acepção pertence. Cada conjunto da LC é formado por uma LE e uma LA. A lista de entradas contém entradas que estão relacionadas entre si pela sinonímia e pode pertencer a várias acepções. Por poder pertencer a várias acepções, o conjunto formado pelas entradas precisa de uma lista de acepções para facilitar a pesquisa das entradas na detecção de relação, sinonímia ou antonímia, que cada conjunto contrai com as entradas que o contém e no gerenciamento desse conjunto pelo editor (Dias-da-Silva *et al*, 2000, p. 06-07).

6.8 A inserção de dados na interface: o procedimento do lingüista

Antes de dar início ao trabalho de extração/inserção de dados na base do *thesaurus*, a equipe de lingüistas deve seguir os passos apresentados a seguir.

Em primeiro lugar, é preciso que todos se inteirem do construto principal do *thesaurus*, ou seja, os conjuntos de sinônimos e antônimos. Em segundo lugar, por questões metodológicas, as classes sintáticas consideradas para a base lexical (verbos, advérbios, adjetivos e substantivos) devem ser distribuídas entre três lingüistas. Outra razão de distribuição de tarefas é o limitado número de computadores disponíveis. Os adjetivos e advérbios, por razões quantitativas, podem ser trabalhados pelo mesmo lingüista.

Distribuídas as classes sintáticas entre os membros da equipe, uma discussão deve ser feita de como os dados devem ser extraídos do *corpus* de referência para serem armazenados na base lexical; o procedimento é cotejar as informações das fontes selecionadas, para que uma confirme ou complete a informação de outra. Essa tarefa não é simples.

Extrair informação do *corpus* de referência para montar os conjuntos de sinônimos e antônimos envolve:

-aplicar os critérios de filtragem de informação lexical

-reflexão sobre a informação, não apenas um procedimento automático.

Sendo assim, é preciso atentar para vários tipos de problemas, como demonstram Dias-da-Silva e Moraes (2000); um deles é o da paráfrase. Observando as entradas dos dicionários que constituem o *corpus* de referência, verifica-se que é comum a ocorrência de tal fenômeno: por exemplo, em Weiszflog (op.cit.), na entrada de *prolongar*, a primeira acepção dá como definição *dar maior comprimento*, o que equivale a *encompridar*, cuja definição, no mesmo dicionário, é *tornar mais comprido*.

Outro tipo de problema refere-se à troponímia; por exemplo, *cochichar* é *falar em voz baixa*, pois *cochichar* é *falar de algum modo*; sendo assim, *cochichar* não é sinônimo de *falar*. No entanto, é preciso atenção para o fato de que nem sempre a troponímia está presente: “por exemplo, não há a relação aspectual de troponímia entre “labutar” e “trabalhar com intensidade”, porque “labutar” não é o mesmo que “trabalhar de um certo modo” (Dias-da-Silva e Moraes, 2000, p. 04).

Fica portanto atestada a importância de se observar os critérios de filtragem de informação léxico-semântica. Ilustremos o procedimento de extração/inserção de informação lexical através de um exemplo concreto, retirado de Dias-da-Silva e Moraes, 2000, a montagem do verbete *lembrar*. No dicionário de Weiszflog (op. cit.) temos a seguinte definição:

Lembrar

v.1.Tr.dir. Trazer à memória; recordar. 2. Tr. Ind. Vir à idéia, tornar-se recordado. 3. Pron. Recordar-se, ter lembrança de. 4. Tr. Dir. Fazer vir à memória por analogia ou semelhança. 5. Tr. Dir. Advertir, notar. 6. Tr. Dir. Sugerir. 7. Tr. Dir. Recomendar.

Após a filtragem da informação, é possível montar cinco conjuntos, representando cinco acepções diferentes de *lembrar*:

{lembrar, recordar}

{lembrar, advertir, notar}

{lembrar, sugerir}

{lembrar, recomendar}

{lembrar-se, recordar-se}¹⁰

A partir dessa primeira montagem dos conjuntos, é preciso verificar a consistência da informação extraída através da consulta aos verbetes *recordar*, *recordar-se*, *advertir*, *notar*, *sugerir* e *recomendar*. Temos, por exemplo, *recordar*:

Recordar

v.1. Tr. Dir. Trazer à memória. 2. Pron. Lembrar-se. 3. Tr. Dir. Fazer lembrar; ter analogia ou semelhança com; parecer. 4. Tr. Ind. Lembrar.

Esse verbete permite que se confirmem os dois conjuntos {lembrar, recordar} e {lembrar-se, recordar-se}, além de permitir que se construa um conjunto novo: {recordar, parecer}. Tal procedimento continua até que se esgotem todos os verbetes “atingíveis” a partir do verbete *lembrar*. Ao fim desse procedimento, retoma-se a ordem alfabética, como, por exemplo, consultando o verbete *esquecer* (Dias-da-Silva e Moraes, 2000, p. 5):

Esquecer

v.1. Tr. Dir. Deixar sair da memória; perder a memória de; tirar da lembrança, olvidar. 2. Pron. Perder a lembrança ou a memória; olvidar-se. 3. Tr. Dir. Não fazer caso de, pôr em esquecimento. 4. Tr. Ind. E intr. Escapar da memória, ficar em esquecimento: esqueceu-lhe o final do discurso. Seu prestígio foi momentâneo, passou e esqueceu. 5. Tr. Dir. Descurar-se de: Não esquecia as suas tarefas. 6. Pron. Perder a ciência ou a habilidade adquiridas: Já me esqueci do latim. 7. Pron. Descuidar-se: Meu secretário esqueceu-se de tudo. 8. Intr. Ficar dormente ou tolhido, perder a sensibilidade: naquela posição a perna esqueceu.

Após a filtragem das informações, temos os seguintes conjuntos:

{esquecer, olvidar}

{esquecer-se, olvidar-se}

{esquecer-se, descuidar-se, descurar-se}

Apesar de o verbete apresentar *descurar-se* e *descuidar-se* em acepções diferentes, a inserção desses dois itens em um mesmo conjunto justifica-se por dois motivos: esse mesmo dicionário apresenta *descurar-se* como sinônimo de *descuidar-se* no verbete *descurar*, e essa informação é confirmada em outras obras do *corpus* de referência (Dias-da-Silva e Moraes, 2000, p. 05).

Quanto aos antônimos, mesmo não tendo sido mencionados, é evidente a oposição *lembrar/esquecer* expressa através de paráfrases; por exemplo, no verbete *lembrar*, temos *trazer à memória* e, em *esquecer*, temos *deixar sair da memória*, *perder a memória de*; *tirar da lembrança*. Portanto, pode-se estabelecer a relação de antonímia entre os conjuntos {lembrar, recordar} e {esquecer, olvidar}.

Está, desta forma, ilustrado o procedimento de montagem de conjuntos de sinônimos e antônimos na base do *thesaurus*.

No capítulo final desta dissertação resumimos os principais passos apresentados neste trabalho e que devem nortear a compilação de um *thesaurus eletrônico*, ilustramos a aplicação de parte desses passos na montagem de uma base experimental, contendo hoje quarenta e quatro mil entradas, e apontamos os trabalhos futuros.

¹⁰ Esse conjunto pôde ser montado porque o sentido da forma pronominal *lembrar-se* é diferente de *lembrar*.

CAPÍTULO 7 - CONCLUSÕES, RESULTADOS E PERSPECTIVAS

Ao final da pesquisa, acreditamos ter alcançado nosso objetivo: o desenvolvimento de uma proposta metodológica para a compilação de uma base lexical particular para fins de PLN.

Seguindo a metodologia de desenvolvimento de projetos em PLN, dividida em três grandes fases (lingüística, representacional e implementacional), discutimos e equacionamos problemas referentes a cada uma das fases envolvidas.

Em primeiro lugar, realizamos um recorte terminológico em busca da delimitação do objeto *thesaurus* para o nosso trabalho, que foi definido como um dicionário de sinônimos e antônimos acoplado a um programa editor de textos, com a finalidade de auxiliar a tarefa da escrita do usuário de língua portuguesa.

Assim, partimos para o estudo da metodologia de trabalho no campo do PLN, a qual tem como uma de suas necessidades a construção de uma base lexical eletrônica, parte de qualquer sistema de PLN e essencial para o desenvolvimento de variados aplicativos computacionais.

Definido o objeto a ser construído, uma base lexical lingüisticamente adequada e destinada a representar computacionalmente as relações léxico-semânticas da sinonímia e antonímia, partimos para o exame de questões lingüísticas envolvidas no processo.

A fase lingüística envolveu o estudo de três formas de se considerar o léxico: como componente da gramática, do sistema cognitivo e de um sistema de PLN. Através do estudo de modelos de representação de entradas lexicais e de estruturas conceituais, buscamos as relações de sentido entre as unidades lexicais, levando-nos a concluir que o léxico é hierarquizado e passível de tratamento computacional. Dessa forma, partimos para a determinação de qual procedimento de construção de bases lexicais estaria de

acordo com nossas necessidades e disponibilidades; adotamos o método de extração de unidades lexicais a partir de fontes lexicográficas, ou seja, de um corpo de dicionários de língua portuguesa dos quais derivaríamos nossa base lexical. Através do exame dos métodos lexicográficos e da estrutura da definição nos dicionários, demonstramos a viabilidade da extração de itens lexicais do *corpus* de referência, dada a riqueza de sinônimos e antônimos fornecidos, desde que fossem observados critérios heurísticos para a construção dos conjuntos.

O estudo aprofundado das relações léxico-semânticas da sinonímia e antonímia foi fundamental para que observássemos o seu papel na constituição da estrutura do léxico e para partirmos para a investigação de uma forma de representação de tais relações que fosse computacionalmente tratável.

O resultado foi encontrado na adoção conjunta de três princípios: na utilização do método diferencial, na noção de matriz lexical (relacionando formas e significados) e na formalização das duas relações semânticas em termos de conjuntos de sinônimos e antônimos; assim, cada conjunto de unidades acabaria por corresponder a um átomo conceitual, eliminando a necessidade de um rótulo conceitual, apenas apontando para a existência do referido conceito.

Finalmente, abordamos questões referentes à implementação da base do *thesaurus*, como a necessidade de elaboração de um editor que fosse responsável pelo armazenamento e gerenciamento dos dados lexicais estruturados em termos das relações léxico-semânticas da sinonímia e antonímia e, por outro lado, do estabelecimento do procedimento de inserção de dados na base por parte do lingüista, que exigiu a especificação de parâmetros heurísticos para filtragem de informação lexicográfica a partir de um *corpus* de referência.

Há que se ressaltar que a metodologia aqui proposta foi aplicada concomitantemente à montagem da base por uma equipe envolvendo lingüistas e

cientistas da computação trabalhando em conjunto na difícil tarefa de conciliar os conhecimentos lingüístico e computacional. Assim, a tarefa de construir um léxico digital observando as relações de sentido entre as unidades do léxico foi concretizada através da construção de uma base lexical eletrônica, reunindo quatro grandes classes gramaticais (substantivos, verbos, adjetivos e advérbios), distribuídos conforme indica a Tabela 5 (Dias-da-Silva et al, 2001).

Tabela 5. A base lexical do *thesaurus eletrônico*, em números.

Classe gramatical	Nº. de entradas	Nº. de conjuntos
Substantivos	17388	8526
Verbos	11077	4129
Adjetivos	15073	6648
Advérbios	1139	567
Total	44677	19870

Os resultados obtidos são promissores, uma vez que a base lexical obtida através do emprego da metodologia aqui discutida revelou-se consistente e eficiente para a construção de bases eletrônicas semelhantes. Há que se ressaltar que antes mesmo de a equipe começar o trabalho de compilação da base lexical do *thesaurus eletrônico*, realizamos uma breve análise de dois aplicativos disponíveis para o português do Brasil, o *Dicionário de Sinônimos do Português* (acoplado ao pacote do Microsoft Word 97) e o *Dicionário de Sinônimos para a Língua Portuguesa* (acoplado ao pacote Microsoft Word 2000), aos quais chamamos respectivamente de v.97 e v.2000. A partir de um texto de Monteiro Lobato (ver anexo 01) extraímos 142 unidades lexicais e realizamos um levantamento de sentidos e significados em ambos os aplicativos sob análise (ver anexo 02). A partir dos dados desse levantamento, chegamos à algumas conclusões a respeito da qualidade de ambos os aplicativos, utilizando-nos inclusive de dados de avaliação da própria Microsoft (ver anexo 03). Essencialmente, os aplicativos mostraram deficiências como:

-pouca quantidade de itens lexicais reconhecidos – dos 142 itens analisados, a v.97 reconheceu 70 itens e a v.2000, 107;

- atribuição incorreta de categorias gramaticais – ex: *acolá* – *adjetivo* (v.97);
- desconhecimento de itens lexicais flexionados – ex: *captou* (v.97);
- sinônimos e acepções controversos em algumas entradas – ex: *mulher* como sinônimo de *costela* (v.97);
- interface gráfica não objetiva;
- escassez de antônimos – apenas a v.2000 apresenta antônimos.

Os resultados obtidos nessa breve análise deixam clara a necessidade de se construir um novo aplicativo que atenda a exigências de construção de interfaces computacionais para consulta de dados, procurando-se compilar uma maior quantidade de dados e conferir fidedignidade às informações nele contidas. Com uma base com mais de quarenta e quatro mil itens lexicais arranjados em conjuntos de sinônimos e antônimos, acreditamos ter viabilizado a futura implantação do *thesaurus eletrônico*, visto que a base oferece maior abrangência e qualidade de informações, pois foi lingüisticamente motivada.

Observamos que com esse trabalho é possível afirmar que podemos já pensar na confecção de uma rede semelhante à *WordNet* para o português do Brasil. Embora as duas únicas relações semânticas codificadas na base sejam a sinonímia e a antonímia, a base não deixa de ser uma rica fonte de unidades lexicais que permitem que se codifiquem outras relações de sentido entre formas e entre significados. Embora o empreendimento demande grande equipe de pesquisadores e seja tarefa laboriosa, ao menos é possível investigar uma parte dos problemas lingüísticos envolvidos no processo, abrindo caminho para a implantação, como dissemos, de uma possível *WordNet* para o português do Brasil.

Além desta, assinalamos perspectivas necessárias para que a base seja definitivamente implementada sob a forma de um *thesaurus eletrônico*:

- a) possibilidade de inserção dessa base na BDL (base de dados lexicais) do NILC¹¹ – essa base, em construção, pode ser enriquecida com o acréscimo dos dados desenvolvidos para o projeto do *thesaurus*; o desenvolvimento da BDL propõe, com base nos conjuntos de sinônimos que compõem o *thesaurus*, derivar uma ontologia para o português (Gregghi, 2001, p. 20);
- b) desenvolvimento de uma ontologia de conceitos compostos de unidades lexicais pertencentes a um mesmo núcleo de significado vinculadas a um “rótulo conceitual”, a exemplo da proposta de construção de ontologias de Jackendoff, apresentada neste trabalho;
- c) ampliação da própria base, com o acréscimo de outras informações tais como marcadores discursivos, estendendo a base para que também possa abrigar categorias pragmático-discursivas (cf. Dias-da-Silva & Oliveira, 2001);
- d) refinamento da base, com uma completa revisão dos conjuntos e sentidos construídos;
- e) servir de motivação para a construção de uma *WordNet* para o português do Brasil.

Merece destaque no empreendimento o trabalho conjunto realizado pela equipe interdisciplinar; com a cooperação de profissionais de variados ramos do conhecimento, o que torna possível a realização de trabalhos na área do PLN.

¹¹ A iniciativa de se construir o *Thesaurus* da Língua Portuguesa (*TeP*) surgiu no âmbito de pesquisas do Núcleo Interinstitucional de Linguística Computacional (NILC). O NILC, desde 1993, estrutura-se em função de pesquisadores da Universidade de São Paulo (USP), Campus de São Carlos, da Universidade Federal de São Carlos (UFSCar) e da Universidade Estadual Paulista (UNESP). Os especialistas pertencem a diversas áreas do conhecimento, como lingüística, ciência e engenharia da computação,

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, J. Ambigüidade lexical. **ALFA**. São Paulo: vol. 34, p. 187-193, 1990.
- ANDRADE, M. M. Organização da macroestrutura: problemas metodológicos. **I Encontro Nacional do GT de Lexicologia, Lexicografia e Terminologia da ANPOLL. Anais...** Rio de Janeiro: UFRJ, 1997.
- APRESYAN, YU. D; MEL'CUK, I.A.; ZOLKOVSKY, A.K. Semantics and Lexicography: towards a new type of unilingual dictionary. In: KIEFER, F. (Ed.) **Studies in Syntax and Semantics**. Boston: D. Reidel Publishing Company, 1969.
- AZEVEDO, F.F.S. **Dicionário analógico da língua portuguesa**. Brasília: Thesaurus, 1983.
- BARBOSA, O. **Grande dicionário de sinônimos e antônimos**. Rio de Janeiro: Ediouro, 1999.
- BARBOSA, M. A. Estrutura da obra lexicográfica: algumas questões semânticas. Assis: 1994. (MIMEO)
- _____. Lexicologia, lexicografia, terminologia, terminografia, identidade científica, objeto, métodos, campos de atuação. **II Simpósio Latino-Americano de Terminologia, I Encontro Brasileiro de Terminologia Técnico-Científica. Anais...** Brasília: CNPq/IBICT, 1990.
- _____. Relações de significação nas unidades lexicais. **I Encontro Nacional do GT de Lexicologia, Lexicografia e Terminologia da ANPOLL. Anais...** Rio de Janeiro: UFRJ, 1997.
- BASILIO, M. **Teoria Lexical**. São Paulo: Ática, 1987.

BECHARA, E. Estudo Estrutural do Léxico: a Lexemática. In: _____. **Moderna Gramática Portuguesa**. Rio de Janeiro: Lucerna, 1999.

BIDERMAN, M.T.C. **Teoria lingüística**. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

_____. O dicionário como norma na sociedade. **I Encontro Nacional do GT de Lexicologia, Lexicografia e Terminologia da ANPOLL. Anais...** Rio de Janeiro: UFRJ, 1997.

_____. As ciências do léxico. In: OLIVEIRA, A. M. P. P. ; ISQUERDO, A. N. (org.) **As ciências do léxico. Lexicologia, Lexicografia, Terminologia**. Campo Grande: UFMG, 1998.

_____. Os dicionários na contemporaneidade: arquitetura, métodos e técnicas. In: Idem. *Ibidem*.

BIERWISH, M.; SCHREUDER, R. From concepts to lexical items. **Cognition**, n.42, p. 23-60, 1992.

BOGURAEV, B.; BRISCOE, T. **Computational Lexicography for Natural Language Processing**. London/New York: Longman, 1989.

BORBA, F. S. **Pequeno vocabulário de lingüística moderna**. 2. ed. São Paulo: Cia. Editora Nacional, 1976.

BORBA, F.S. (coord.) **Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil**. São Paulo: Fundação Editora Unesp, 1990.

CAMPOS, M. H. C.; XAVIER, M. F. Estrutura semântica do léxico. In: _____. **Sintaxe e Semântica do Português**. Lisboa: Universidade Aberta, 1991.

CANN, R. Introduction. In: **Formal Semantics**. An introduction. Cambridge: Cambridge University Press, 1993.

CAVAZZA, M.; ZWEIGENBAUM, P. Lexical semantics: dictionary or encyclopedia? In: SAINT-DIZIER, P.; VIEGAS, E. (eds.) **Computational Lexical Semantics**. Cambridge: Cambridge University Press, 1995.

COOPER, R. et al. Lexical Semantics. In: _____. **Evaluating the state of the art**. The FraCas Consortium. Deliverable D10, January 3, 1995. Disponível em <http://www.stanford.edu/~dib/references.html> Acesso em: 10 mar. 2001.

_____. **A framework for Computational Semantics**. The Fracas Consortium. In: _____. Deliverable D16, Final Draft, January 1996. Disponível em <http://www.stanford.edu/~dib/references.html> Acesso em: 10 mar. 2001.

CRUSE, D. A. **Lexical Semantics**. Cambridge: Cambridge University Press, 1995.

CRYSTAL, D. **The Cambridge encyclopedia of the English language**. Cambridge: Cambridge University Press, 1997.

DIAS-DA-SILVA, B.C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. Araraquara, 1996. 272p. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara.

_____. Os Domínios Lingüístico e Tecnológico do Estudo do Processamento Automático das Línguas Naturais. **Estudos Lingüísticos**, Assis, vol. 26, p.612-617, 1998a.

_____. Bridging the gap between linguistic theory and natural language processing. **16th International Congress of Linguists. Proceedings...** Pergamon: Oxford/Elsevier, Paper 0425, 1998b.

DIAS-DA-SILVA, B. C. et al. A construção de um **thesaurus** eletrônico para o português do Brasil. **V PROPOR - Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada**. Anais... São Carlos: ICMSC, p. 01-10, 2000a.

_____. A construção de um **thesaurus** eletrônico para o português do Brasil. 2000b. (MIMEO).

_____. **Relatórios Técnicos FINEP**. 2001. (MIMEO)

DIAS-DA-SILVA, B. C.; MORAES, H. R. A questão da representação lingüístico-computacional da sinonímia e antonímia na compilação de um *thesaurus* eletrônico. In: XII CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNESP, 2000, São José do Rio Preto. **Resumos...** São José do Rio Preto, UNESP, 2000, p. 217.

_____. A questão da representação lingüístico-computacional da sinonímia e antonímia na compilação de um **thesaurus** eletrônico. 2000 (MIMEO)

_____. Filtragem de informação léxico-semântica na compilação de um *thesaurus* eletrônico. XIII Congresso de Iniciação Científica da UNESP. Bauru, 2001. (MIMEO).

DIAS-DA-SILVA, B. C.; OLIVEIRA, M. F. Estrutura do léxico: modelo lingüístico-computacional de representação das relações semânticas. **Estudos Lingüísticos XXX**. Marília: GEL, 2001. 1 CD.

_____. Inclusão de informação pragmático-discursiva na base lexical de um *thesaurus* eletrônico. Marília: GEL, 2001. (MIMEO)

DICIONÁRIO AURÉLIO ELETRÔNICO – SÉCULO XXI – VERSÃO 3.0, Lexikon Informática Ltda., 1999.

DICIONÁRIO DE SINÔNIMOS DO PORTUGUÊS, ferramenta do *Word 97*,

Microsoft, 1984-1997.

DUBOIS, J. et al. **Dicionário de lingüística**. Direção e coordenação geral da tradução Izidoro Blikstein. São Paulo: Cultrix, 1978.

FERNANDES, F. **Dicionário de sinônimos e antônimos da língua portuguesa**. São Paulo: Globo, 1997.

FODOR, J. A.; LEPORE, E. The Emptiness of the Lexicon: Reflexions on James Pustejovsky's **The Generative Lexicon**. **Linguistic Inquiry**. Boston: The MIT Press, vol. 29, n. 2, p. 269-288, 1998.

FRANKLIN, K. J. Lexicography considerations for Tok Pisin. **SIL Electronic Working Papers**, 198-002, 1998. Disponível em www.sil.org/silewp/1998/002/silewp1998-002.html Acesso em: 15 jul. 1999.

GAZDAR, G.; MELLISH, C. Features and the lexicon. In: _____. **Natural language processing in prolog: an introduction to computational linguistics**. New York: Addison Wesley, 1989.

GRANDE ENCICLOPÉDIA LAROUSSE CULTURAL. São Paulo: Nova Cultural, 1998.

GREGHI, J. G. **Uma base de dados lexicais para o português do Brasil**. Relatório de Qualificação. São Carlos: ICMSC/USP, 2001.

GUIMARÃES, E. **Os limites do sentido**. Campinas: Pontes, 1995.

GREFENSTETTE, G. **Explorations in automatic thesaurus discovery**. Boston/London/Dordrecht: Kluwer Academic Publishers, 1994.

GUIRAUD, P. **A semântica**. Tradução e adaptação Maria Elisa Mascarenhas. São Paulo: Difusão Européia do Livro, 1972.

HAENSCH, G. et al. **La Lexicografía**. Madrid: Editorial Gredos, 1982.

HANDKE, J. **The Structure of the lexicon**. Human versus Machine. Berlin/New York: Mouton de Gruyter, 1995.

HARTMANN, R.R.K. (Ed.) **Lexicography: Principles and Practice**. London: Academic Press Inc. (London) Ltd., 1983.

ILARI, R.; GERALDY, W. **Semântica**. 2 ed. São Paulo: Ática, 1985.

JACKENDOFF, R. S. **Semantic Structures**. Cambridge/Mass.: The MIT Press, 1990.

_____. **Idem**. 2.ed. Cambridge-Mass./London-England: The MIT Press, 1991.

_____. **Semantics and Cognition**. 3.ed. London/Cambridge: The MIT Press, 1993.

KATZ, J.J. Semantic Theory. In: STEINBERG, D. D.; JAKOBOVITZ, L. A. (ed.) **Semantics**. Cambridge: Cambridge University Press, 1975.

KILGARRIFF, A. Dictionary Word Sense Distinctions: An Inquiry Into Their Nature. **Computer and Humanities**, n.26, p.365-387, 1993.

LANDAU, S. I. **Dictionaries. The Art and Craft of Lexicography**. Cambridge: Cambridge University Press, 1996.

LEECH, G. **Semantics**. Middlessex: Penguin Books, 1974.

LEVIN, B.; PINKER, S. Lexical and conceptual semantics. **Cognition**, vol. 41 (1-3), p. 01-229, 1991.

LE NY, J. F. Mental lexicon and machine lexicon: Which properties are shared by machine and mental word representations? Which are not? In: SAINT-DIZIER, P.; VIEGAS, E. (org). **Computational lexical semantics**. Cambridge: Cambridge University Press, 1995.

LEVELT, W. J. M. **Speaking**. Cambridge/Mass.: The MIT Press, 1989.

_____. Accessing words in speech production: stages, processes and representations. **Cognition**, n. 42, 1992, p. 1-22.

LYONS, J. **Introduction to theoretical linguistics**. Cambridge: Cambridge University Press, 1968.

_____. Introduction: some basic terms and concepts. In: _____. **Semantics (I)**. Cambridge: Cambridge University Press, 1977.

_____. Reference, sense and denotation. Idem. Ibidem.

_____. The lexicon. In: _____. **Semantics. (2)** Cambridge: Cambridge University Press, 1977.

LYONS, J. **Introdução à lingüística geral**. Supervisão de tradução de Isaac Nicolau Salum. São Paulo: Editora Nacional/Edusp, 1979.

_____. **Language and linguistics**. An introduction. Cambridge: Cambridge University Press, 1981.

MICHAELIS PORTUGUÊS – MODERNO DICIONÁRIO DA LÍNGUA PORTUGUESA - Versão 1.0, DTS Software Brasil Ltda, 1998.

MILLER, G. A.; FELLBAUM, C. Semantic Networks of English. **Cognition**, n.41, p. 197-229, 1991.

MILLER, G. A. et al. **Five Papers on WordNet**. 1993. Disponível em <<http://www.cogsci.princeton.edu/~wn>> Acesso em: 30 mar. 1999.

NASCENTES, A. **Dicionário de sinônimos**. São Paulo: Nova Fronteira, 1981.

NUNES, M.G.V. et al. A Construção de um Léxico da Língua Portuguesa do Brasil para Suporte à Correção Automática de Textos, **Relatórios Técnicos do ICMSC**, n. 42, Agosto, 1996, 37 p.

OLIVEIRA, M. F. **O problema do significado: Da semântica lexical ao Processamento das Línguas Naturais**. Araraquara: Faculdade de Ciências e Letras (UNESP), 2000. (MIMEO)

PUSTEJOVSKY, J. The syntax of event structure. **Cognition**, vol. 41, p. 47-81, 1991.

_____. Linguistic constraints on type coercion. In: SAINT-DIZIER, P.; VIEGAS, E. (Ed.) **Computational Lexical Semantics**. Cambridge: Cambridge University Press, 1995.

_____. **The Generative Lexicon**. 2. ed. Cambridge (Mass.)/London: The MIT Press, 1996.

_____. Generativity and Explanation in Semantics: a reply to Fodor and Lepore. **Linguistic Inquiry**. Boston: The MIT Press, vol. 29, n. 02, p. 289-311, 1998.

PUSTEJOVSKY, J. ; BOGURAEV, B. Lexical knowledge representation and natural language processing. In: PEREIRA, F. C. N.; GROSZ, B. J. (Ed.) **Natural Language Processing**. Boston: The MIT Press, 1994.

PUSTEJOVSKY, J. Generativity and Explanation in Semantics: A reply to Fodor and Lepore. In: **Linguistic Inquiry**. Boston: The MIT Press, vol.29, n. 02, p. 289-311, 1998.

SAINT-DIZIER, P.; VIEGAS, E. An introduction to lexical semantics from a linguistic and a psycholinguistic perspective. In: _____. (org). **Computational lexical semantics**. Cambridge: Cambridge University Press, 1995.

RANDOM HOUSE WEBSTER'S UNABRIDGED ELECTRONIC DICTIONARY
– Version 2.0, Random House Inc, 1994.

REPORT ON THE BRAZILIAN THESAURUS LINGUISTIC EVALUATION.
Microsoft, 1998. (MIMEO)

REY-DEBOVE, J. La synonymie ou les échanges de signes comme fondement de la sémantique. In: BALIBAR-MRABTI, A. (Ed.) **La synonymie**. Paris: Larousse, 1997. (**Langages**, v. 128, p. 91-112).

ROGET, P.M. **Roget's Thesaurus**. Middlessex: Penguin Books, 1953. (ed. original 1852)

ROSSI, J. G. Considérations logico-philosophiques sur la synonymie. In: BALIBAR-MRABTI, A. (Ed.) **La synonymie**. Paris: Larousse, 1997. (**Langages**, v. 128, p. 105-126).

SUMMERS, A.(ed.) **Longman Dictionary of Contemporary English**. Essex: Longman, 1995.

ULLMANN, S. **The principles of semantics**. 3. ed. Oxford: Basil Blackwell, 1963.

_____. **Semántica**. Introducción a la ciencia del significado. 2. ed. Madrid: Aguilar, 1967.

VERSPoor, C. M. **Contextually-Dependent Lexical Semantics**. The University of Edimburg, 1997. Disponível em http://www.cogsci.ed.ac.uk/~kversp/ftp_html/thesis.html. Acesso em: 15 jul. 1999.

VILELA, M. **Estruturas léxicas do português**. Coimbra: Almedina, 1979.

_____. **Definição nos dicionários de português**. Rio Tinto: ASA, 1983.

WALKER, D.; IDE, N. Introduction: Common Methodologies in Humanities Computing and Computational Linguistics. **Computers and Humanities**, n. 26, p. 327-330, 1993.

WEBSTER'S NEW WORLD DICTIONARY & THESAURUS – Version 1.0, Macmillan Publishers, 1997.

WEISZFLOG, W. (ed.) **Michaelis português – moderno dicionário da língua portuguesa** (Versão 1.0). São Paulo: DTS Software Brasil Ltda, 1998.

YULE, G. Semantics. In: _____. **The study of Language**. 2. ed. Cambridge: Cambridge University Press, 1996.

ANEXO 01

**O texto utilizado como ponto de partida para a análise de dois aplicativos do tipo
“Dicionários de sinônimos” atualmente em uso no Brasil**

Texto utilizado como *corpus* para comparação entre duas versões do *Dicionário de Sinônimos* existente no pacote Office (97 e 2000).

O BURRO JUIZ

Monteiro Lobato

Disputava a gralha com o sabiá, afirmando que a sua voz valia a dele. Como as outras aves rissem daquela pretensão, a bulhenta matraca de penas, furiosa, disse:

— Nada de brincadeiras. Isto é uma questão muito séria, que deve ser decidida por um juiz. Canta o sabiá, canto eu, e a sentença do julgador decidirá quem é o melhor artista. Topam?

— Topamos! piaram as aves. Mas quem servirá de juiz?

Estavam a debater este ponto, quando zurrou um burro.

— Nem de encomenda! exclamou a gralha. Está lá um juiz de primeiríssima para julgamento de música, pois nenhum animal possui maiores orelhas. Convidê-mo-lo.

Aceitou o burro o juizado e veio postar-se no centro da roda.

— Vamos lá, comecem! ordenou ele.

O sabiá deu um pulinho, abriu o bico e cantou. Cantou como só cantam sabiás, garganteando os trinos mais melodiosos e límpidos. Uma pura maravilha, que deixou mergulhado em êxtase o auditório em peso.

— Agora eu! disse a gralha, dando um passo à frente.

E abrindo a bicanca matraqueou uma grita de romper os ouvidos aos próprios surdos.

Terminada a justa, o meritíssimo juiz deu a sentença:

— Dou ganho de causa à excelentíssima senhora dona Gralha, porque canta muito mais forte que mestre sabiá. (*)

Quem burro nasce, togado ou não, burro morre.

Fonte: www.releituras.com

ANEXO 02

Levantamento de sentidos e significados extraído dos aplicativos sob análise a partir das unidades textuais do anexo 01

TEXTO: O BURRO JUIZ (Monteiro Lobato)

Total de palavras no texto = 216
 Com exclusão de palavras repetidas – 162
 Com expressões – 142 unidades analisadas

TOTAL DE ITENS ANALISADOS - 142

Abreviaturas:

R = itens lexicais reconhecidos pelo aplicativo
NR = itens lexicais não reconhecidos pelo aplicativo
Acep. = aceção
Pal. rel. = palavras relacionadas
v.t. = ver também
sin = sinônimo
ant = antônimo

ITENS LEXICAIS	OFFICE 97		OFFICE 2000	
	R	NR	R	NR
a	x		x	
à				
à debater		x		x
à frente		x		x
abrindo		x	x	X

02 acep. ampère-s, para-outro (sem sinônimos)
 Sugere subst. por a, ha

30 acep. 01. acender-v. (sin. acender, abrir, ligar); fechar (antônimo), gás (veja também)
 02. afastar-v. (sin. afastar, abrir, apartar, arredar, espacar, separar); separação (veja também);
 03. alargar-v. (sin. alargar, abrir, dilatar, larguear); estrangular (ant.), estreitar (ant.), abertura (veja também), largo (veja também);
 04. ampliar-se-v (sin. ampliar-se, abrir, expandir); expansão (ver também);
 05. começar-v (sin. começar, abrir, dar início, principiar); fechar (ant.), reunião (veja também), sessão (veja também);
 06. começar-v (sin. começar, abrir, encabeçar, puxar fila); fechar (ant.) começo (veja também), lista (veja também), ponta (veja também), ponta (veja também), seqüência (veja também);
 07. cortar-v (sin. cortar, abrir, rasgar); fechar (ant.), envelope (veja também), pacote (veja também);

				<p>08. cortar-v (sin. cortar, abrir, incisar, tallar); fechar (ant.), suturar (ant.), corte (ver também);</p> <p>09. desabotoar-v (sin. desabotoar, abrir); abotoar (ant.), botão (v.t.)</p> <p>10. descerrar-v (sin. descerrar, abrir); fechar (ant.), abertura (v.t.), corte (v.t.), início (v.t.), separação (v.t.)</p> <p>11. desdobrar-v (sin. desdobrar, abrir, distender, estender); embolar (ant.), enrolar (ant.), massa (v.t.)</p> <p>12. fender-v (sin. fender, abrir, gretar, rachar); rachadura (v.t.)</p> <p>13. fender-v (sin. fender, abrir, furar); fechar (ant.), abertura (v.t.), furo (v.t.)</p> <p>14. vazár-v (sin. vazár, abrir); fechar (ant.), vazamento (v.t.), vazão (v.t.)</p> <p>15. cavado-adj. (sin. cavado, aberto, revolvido); fechado (ant.), cavar (v.t.)</p> <p>16. escancarado-adj. (sin. escancarado, aberto, fendido, hiante); fechado (ant.)</p> <p>17. acessível-adj. (sin. acessível, aberto, compreensível, inteligível); inacessível (ant.), abertura (v.t.), abrir (v.t.)</p> <p>18. acessível-adj. (sin. acessível, aberto, franco, franqueado, livre, patente); fechado (ant.)</p> <p>19. amplo-adj. (sin. amplo, aberto, dilatado, espaçoso, extenso, largo, vasto, fechado (ant.), abertura (v.t.), abrir (v.t.)</p> <p>20. campestre-adj. (sin. campestre, aberto, clara, claro)</p> <p>21. claro-adj. (sin. claro, aberto, intenso, vivo); escuro (ant.), abertura (v.t.), abrir (v.t.)</p> <p>22. cunhado-adj. (cunhado, aberto, gravado, impresso); abertura (v.t.), abrir (v.t.)</p> <p>23. dadivoso-adj. (sin. dadivoso, aberto, franco, generoso, largo, liberal, magnânimo, pródigo); avaro (ant.)</p> <p>24. desabotoado-adj. (sin. desabotoado, aberto, desapertado); abertura (v.t.), abrir (v.t.)</p> <p>25. descoberto-adj. (sin. descoberto, aberto, desatapado); coberto (ant.), abertura (v.t.), abrir (v.t.)</p> <p>26. desdobrado-adj. (sin. desdobrado, aberto, desenrolado, estendido); abertura (v.t.), abrir (v.t.)</p> <p>27. desfeito-adj. (sin. desfeito, aberto, desmanchado); abertura (v.t.), abrir (v.t.)</p> <p>28. evidente-adj. (evidente, aberto, manifesto); abertura (v.t.), abrir (v.t.)</p> <p>29. franco-adj. (sin. franco, aberto, leal, lhano, sincero); abertura (v.t.), abrir (v.t.)</p> <p>30. limpidó-adj. (sin. limpidó, aberto, nítido, embaciado (ant.), abertura (v.t.), abrir (v.t.)</p>
abriu	x	Sugere <i>abreu</i>		<p>30 acep. 01. acender-v (sin. acender, abrir, ligar, fechar (ant.), gás (veja também)</p> <p>02. afástar-v (sin. afástar, abrir, apartar, arredar, espacar, separar); separação (veja também);</p> <p>03. alargar-v (sin. alargar, abrir, dilatar, larguear); estrangular (ant.), estreitar (ant.), abertura (veja também), largo (veja também);</p> <p>04. ampliar-se-v (sin. ampliar-se, abrir, expandir); expansão (ver também);</p> <p>05. começar-v (sin. começar, abrir, dar início, principiar); fechar (ant.), reunião (veja também), sessão (veja também);</p> <p>06. começar-v (sin. começar, abrir, encabeçar, puxar fila); fechar (ant.) começo (veja também), lista (veja também), ponta (veja também), ponta (veja também), seqüência (veja também);</p> <p>07. cortar-v (sin. cortar, abrir, rasgar); fechar (ant.), envelope (veja também), pacote (veja também);</p> <p>08. cortar-v (sin. cortar, abrir, incisar, tallar); fechar (ant.), suturar (ant.), corte (ver também);</p> <p>09. desabotoar-v (sin. desabotoar, abrir); abotoar (ant.), botão (v.t.)</p> <p>10. descerrar-v (sin. descerrar, abrir); fechar (ant.), abertura (v.t.), corte (v.t.), início (v.t.), separação (v.t.)</p> <p>11. desdobrar-v (sin. desdobrar, abrir, distender, estender); embolar (ant.), enrolar (ant.), massa (v.t.)</p> <p>12. fender-v (sin. fender, abrir, gretar, rachar); rachadura (v.t.)</p> <p>13. fender-v (sin. fender, abrir, furar); fechar (ant.), abertura (v.t.), furo (v.t.)</p> <p>14. vazár-v (sin. vazár, abrir); fechar (ant.), vazamento (v.t.), vazão (v.t.)</p> <p>15. cavado-adj. (sin. cavado, aberto, revolvido); fechado (ant.), cavar (v.t.)</p> <p>16. escancarado-adj. (sin. escancarado, aberto, fendido, hiante); fechado (ant.)</p> <p>17. acessível-adj. (sin. acessível, aberto, compreensível, inteligível); inacessível (ant.), abertura (v.t.), abrir (v.t.)</p> <p>18. acessível-adj. (sin. acessível, aberto, franco, franqueado, livre, patente); fechado (ant.)</p> <p>19. amplo-adj. (sin. amplo, aberto, dilatado, espaçoso, extenso, largo, vasto); fechado (ant.), abertura (v.t.), abrir (v.t.)</p> <p>20. campestre-adj. (sin. campestre, aberto, clara, claro)</p>

							21. claro-adj. (sin. claro, aberto, intenso, vivo); escuro (ant.), abertura (v.t.), abrir (v.t.) 22. cunhado-adj. (cunhado, aberto, gravado, impresso); abertura (v.t.), abrir (v.t.) 23. dádioso-adj. (sin. dádioso, aberto, franco, generoso, largo, liberal, magnânimo, pródigo); avarento (ant.) 24. desabotoado-adj. (sin. desabotoado, aberto, desapertado); abertura (v.t.), abrir (v.t.) 25. descoberto-adj. (sin. descoberto, aberto, desapertado); coberto (ant.), abertura (v.t.), abrir (v.t.) 26. desdobrado-adj. (sin. desdobrado, aberto, desenrolado, estendido); abertura (v.t.), abrir (v.t.) 27. desfeito-adj. (sin. desfeito, aberto, desmanchado); abertura (v.t.), abrir (v.t.) 28. evidente-adj. (evidente, aberto, manifesto); abertura (v.t.), abrir (v.t.) 29. franco-adj. (sin. franco, aberto, leal, lhano, sincero); abertura (v.t.), abrir (v.t.) 30. limpido-adj. (sin. limpido, aberto, nítido); embaciado (ant.), abertura (v.t.), abrir (v.t.) 07 acep. 01. abrigar-v (abrigar, aceitar, acolher, asilar, receber); enjeitar (ant.), recusar (ant.), acolhida (v.t.) 02. acolher-v (sin. acolher, aceitar, admitir); enjeitar (ant.), recusar (ant.); acolhida (v.t.), aprovar (v.t.) 03. admitir-v (sin. admitir, aceitar, conhecer); conhecimento (v.t.) 04. admitir-v (sin. admitir, aceitar, consentir, deixar, deixar passar, fechar os olhos, permitir, tolerar) 05. adotar-v (sin. adotar, aceitar, tomar, vestir, vestimenta (v.t.), vestíário (v.t.), vestíbulo (v.t.) 06. receber-v (sin. receber, aceitar, tomar) 07. topar-v (sin. topar, aceitar)
aceitou	x		01 acep. aceitar-verb. (sin. <i>aceitar, receber, tomar</i>)	x			
afirmando	x		01 acep. afirmar-verb. (sin. <i>afirmar, assegurar, asseverar, atestar, certificar, garantir</i>)	x			05 acep. 01. Afiançar-v (sin. afiançar, afirmar, jurar) juramento (v.t.) 02. afiançar-v (sin. afiançar, afirmar, assegurar, asseverar, garantir), frança (v.t.) 03. alegar-v (sin. alegar, afirmar, declarar) 04. assegurar-v (sin. assegurar, a firmar, asseverar, certificar, garantir, protestar), protesto (v.t.) 05. assegurar-v (sin. assegurar, afirmar, certificar, garantir, prometer); promessa (v.t.) 03 acep. 01. Atualmente-outro (sin. atualmente, hoje, hoje em dia, presentemente); 02. cá-outro (sem sin.) 03. depressa-outro (sin. depressa, imediatamente, já, logo, prontamente)
agora	x		01 acep. agora-outro (sin. <i>agora, atualmente, hoje em dia, presentemente</i>)	x			
animal	x		01 acep. Animal-subst. (sin. <i>animal, bicho</i>)	x			02. animal-subst. (animal, besta, bicho) 03. animal-subst. (sin. animal, besta, bicho)
aos		x	Sugere substit. por <i>aios, amos, atos, azos</i>		X		
artista	x		01. acep. artêsão-subst. (sin. <i>artêsão, artífice, obreiro, operário, proletário, trabalhador</i>)	x			01. acep. artesão-subst. (sin. artesão, artífice)
as		x	Sugere substit. por <i>hás</i>	x			02 acep. ampere-subst. (sin. ampère, A); para-outro (sin. para, a), de (v.t.)
audifório		x					Sugere subst. por <i>audifória</i> .
aves	x		01 acep. ave-subst. (sin. <i>ave, pássaro</i>)	x			02 acep. bipede-subst. (sin. <i>bipede, ave, ornitóptero, passarinho, pássaro</i>); araponga (v.t.), araponguinha (v.t.), asa-de-telha (v.t.) avestruz (v.t.), azulão (v.t.), beija-flor (v.t.), bem-te-vi (v.t.), buraqueira (v.t.), codorna (v.t.), colhereiro (v.t.), corruira (v.t.), coró-coró (v.t.), cotovia (v.t.), cuitelão (v.t.), curiango (v.t.), cururuca (v.t.), gaviã (v.t.), gavião-tesoura (v.t.), graúna (v.t.), iratua (v.t.), jaburu (v.t.), jandaia (v.t.), jandaia (v.t.), volátil-subst. (sin. <i>volátil, ave</i>).
bicanca		x	Sugere substit. por <i>bichana</i>				
bico		x	02 acep., canto-subst. (sin. <i>bico, canto</i>) e biscate-subst. (sin. <i>biscate, galho</i>).	x			04 acep. acúleo-subst. (sin. acúleo, espigão, espinho, pico, ponta, puia); biscate-subst. (sin. biscate, galho, ganho, ponto, serviço avulso, trabalho avulso); canto-subst. canto, esquina, ponta, quina, vértice); pico-subst. (sin. pico,

			<i>gancho</i>			<i>ponta</i>
brincadeiras		x				08 acep. 01. Adivinha-subst. (sin. adivinha, adivinhação, brincadeira, charada, dificuldade, enigma, esfinge, parábola, problema, quebra-cabeça); adivinhar (v.t.), enigmático (v.t.) 02. brinco-subst. (brinco, brincadeira, brinquedo, jogo), brincar (v.t.) 03. caçoada-subst. (sin. caçoada, brincadeira, galhofa, zombaria); brincar (v.t.) 03. diversão-subst. (sin. diversão, brincadeira, folia); brincar (v.t.) 04. divertimento-subst. (sin. divertimento, brincadeira, entretenimento, passatempo), brincar (v.t.) 04. estúrdia-subst. (sin. estúrdia, brincadeira, paródia, pândega, troça) 05. festa-subst. (festa, brincadeira, festança), brincar (v.t.) 06. folgança-subst. (sin. folgança, brincadeira, folguedo)
bulhenta		x			x	
burro		x	01 acep. asno-subst. (sin. <i>asno, jegue, jericó, jumento</i>)	X		03 acep. 01. acéfalo-adj. (sin. acéfalo), acéfala (v.t.); 02. bate-orelha-adj. (sin. bate-orelha, bronco, estúpido, ignaro, ignorante); 03. asno-subst. (sin. asno, jegue, jericó, jumento)
canta		x	Sugere substit. por <i>canto</i>	x		03 acep. 01. cantiga-subst. (cantiga, cantar, canto, canção, cântico); 02. acenar-subst. (sin. acenar, aliciar, atrair, cantar, convidar, engodar, seduzir); repelir (ant.); 03. executar-v. (sin. executar, cantar); canto (v.t.), cantor (v.t.)
cantam		x	Sugere substit. por <i>catam e captam</i>	x		03 acep. 01. cantiga-subst. (cantiga, cantar, canto, canção, cântico); 02. acenar-subst. (sin. acenar, aliciar, atrair, cantar, convidar, engodar, seduzir); repelir (ant.); 03. executar-v. (sin. executar, cantar); canto (v.t.), cantor (v.t.)
canto	x		02 acep. bico-subst. (sin. <i>bico</i>) e canção-subst. (sin. <i>canção, cântico, cantiga</i>)	x		08 acep. 01. aresta-subst. (sin. aresta, esquina, quina); 02. balada-subst. (sin. balada, canção, hino, lied, serenata, ária), música (v.t.) 03. bico-subst. (sin. bico, esquina, ponta, quina, vértice); 04. cantiga-subst. (sin. cantiga, canlar, canção, cântico) 05. execução-subst. (sin. execução); executar (v.t.) 06. cantiga-subst. (sin. cantiga, cantar, canção, cântico) 07. acenar-v. (sin. acenar, aliciar, atrair, cantar, convidar, engodar, seduzir); repelir (ant.) 08. executar-v. (sin. executar, cantar); canto (v.t.), cantor (v.t.)
cantou		x	Sugere substit. por <i>canto, cantos, captou, catou, contou</i>	x		03 acep. 01. cantiga-subst. (cantiga, cantar, canto, canção, cântico); 02. acenar-subst. (sin. acenar, aliciar, atrair, cantar, convidar, engodar, seduzir); repelir (ant.); 03. executar-v. (sin. executar, cantar); canto (v.t.), cantor (v.t.)
centro	x		01 acep. centro -subst. (sin. <i>centro, meio</i>)	x		04 acep. 01. Cerne-subst. (sin. cerne, núcleo) 02. meio-subst. (sin. meio, ponto médio) 03. centralizar-v. (sin. centralizar, centrar); centro (v.t.) 04. localizado-adjet. (sin. localizado, centrado, sito, situado); centrar (v.t.)
com		x	Sugere substit. por <i>bon, coma, come, comi, como, como, cor, dom</i>		x	Sugere <i>coma</i> .
comecem	x		01 acep. começar -verb. (sin. <i>começar, encetar, estrear, iniciar, principiar</i>) e dá palavras relacionadas (<i>começo, princípio</i>).	x		05 acep. 01. Abanear-v. (sin. abanear, começar, pegar, por-se); banco (v.t.) 02. abrir-v. (sin. abrir, começar, dar início, principiar); fechar (ant.); reunião (v.t.), sessão (v.t.) 03. abrir-v. (sin. abrir, começar, encabeçar, puxar-fila); fechar (ant.); começo (v.t.), lista (v.t.), ponta (v.t.); sequência (v.t.) 04. dar início a – v. (sin. dar início a, começar, encetar, estrear, iniciar, principiar), encerrar (ant.); estréia (v.t.); início (v.t.); princípio (v.t.) 05. iniciar-v. (sin. iniciar, começar, nascer, principiar, ter início); morrer (ant.); terminar (ant.); nascimento (v.t.)
como	x		04 acep.: assim como-outro (sin. <i>assim</i>)	x		09 acep. 01. Aquêle-outro (sin. aquele, esse, porque, qual, que, quão)

				<i>como, bem como</i> ; como-outro (sin. <i>como, assim como, bem como</i>); porque-outro (sin. <i>porque, já que, pois, pois que, por isso que, porquanto que, uma vez que, visto como, visto que</i>); comer-verb. (sin. <i>comer, devorar, tragar</i>)				02. assim como-outro (sin. <i>assim como, bem como, porque, quando</i>) 03. já que –outro (sin. <i>já que, pois, pois que, por isso que, porquanto, porque, uma vez que, visto como, visto que</i>) 04. quanto-outro (sin. <i>quanto, quão</i>) 05. quanto-outro (sin. <i>quanto, quão grandemente, quão intensamente</i>) 06. abocanhar-v. (sin. <i>abocanhar, comer, devorar, engolir, tragar</i>); boca (v.t.) 07. alimentar-se-v (sin. <i>alimentar-se, comer, devorar, mandar</i>) 08. ingerido-adj. (sin. <i>ingerido, comido, mastigado</i>) 09. logrado-adj. (sin. <i>logrado, comido</i>) Sugere subst. por <i>conato</i> .
convidê-mo-lo		x					x	Sugere subst. por <i>conato</i> .
dá		x		Sugere substit. por <i>dá, daí, dar</i>				Dá uma lista de sugestões: <i>dadivoso, dado, dados</i>
dando		x		03 acp.: dar-verb. (sin. <i>dar, doar</i>); acomodar-se-verb. (sin. <i>acomodar-se, acostumar-se, adaptar-se, afazer-se, afeiçãoar-se, ajustar-se, amoldar-se, dar-se, habituar-se, identificar-se, modelar-se</i>); acontecer-verb. (sin. <i>acontecer, dar-se, ocorrer, passar-se, suceder</i>)				12 acp. 01. <i>Abonar-v</i> (sin. <i>abonar, dar, oferecer</i>); abono (v.t.); aprovação (v.t.) 02. <i>abrolhar-v</i> (abrolhar, dar, produzir); fruto (v.t.) 03. <i>apresentar-v</i> (apresentar, dar, deparar, oferecer, prestar, proporcionar); proporcional (v.t.); proporcção (v.t.) 04. <i>conceder-v</i> (conceder, conferir, dar, prestar); prestação (v.t.) 05. <i>doar-v</i> (doar, dar, presentear) 06. <i>emitir-v</i> (emitir, dar, rir); risada (v.t.) 07. <i>emprestar-v.</i> (emprestar, dar); receber (ant.); tomar (ant.); prestar (v.t.) 08. <i>fazer avançar – v</i> (sin. <i>fazer avançar, dar, promover</i>); promoção (v.t.) 09. <i>lançar-v</i> (lançar, dar, soltar); lançamento (v.t.) 10. <i>oferecer-v</i> (oferecer, dar, ofertar, trazer); receber (ant.) 11. <i>elemento-subst.</i> (sin. <i>elemento, dado, fenômeno</i>); fenomenal (v.t.) 12. <i>dados-subst.</i> (sin. <i>dados, dado, informação, números</i>); informação (v.t.)
daquela		x					X	
de		x		Sugere substit. por <i>dê, dei, der, deu</i>			x	01 acp. <i>desde-outro</i> (sin. <i>desde</i>); para (v.t.)
decidia		x		01 acp. decidir-verb. (sin. <i>decidir, deliberar, determinar, resolver</i>)			x	04 acp. 01. <i>Definir-v</i> (sin. <i>definir, decidir, deliberar, determinar, resolver</i>); titubear (ant.); decisão (v.t.), definição (v.t.); resolutivo (v.t.); 02. <i>resolver-v</i> (resolver, decidir, solucionar); solução (v.t.) 03. <i>ativo-adj.</i> (ativo, decidido, diligente, dinâmico, eficaz, enérgico, esforçado, executivo, vivo) 04. <i>deliberado-adj.</i> (deliberado, decidido, determinado, resolutivo, resolvido); indeciso (ant.); decidir (v.t.); decisão (v.t.)
decidirá		x		01 acp. decidir-verb. (sin. <i>decidir, deliberar, determinar, resolver</i>)			x	04 acp. 01. <i>Definir-v</i> (sin. <i>definir, decidir, deliberar, determinar, resolver</i>); titubear (ant.); decisão (v.t.), definição (v.t.); resolutivo (v.t.); 02. <i>resolver-v</i> (resolver, decidir, solucionar); solução (v.t.) 03. <i>ativo-adj.</i> (ativo, decidido, diligente, dinâmico, eficaz, enérgico, esforçado, executivo, vivo) 04. <i>deliberado-adj.</i> (deliberado, decidido, determinado, resolutivo, resolvido); indeciso (ant.); decidir (v.t.); decisão (v.t.)
deixou		x		Sugere substit. por <i>deitou</i>			x	07 acp. 01. <i>Abandonar-v</i> (sin. <i>abandonar, deixar, trocar</i>) 02. <i>aceitar-v</i> (aceitar, admitir, consentir, deixar, deixar passar, fechar os olhos, permitir, tolerar) 03. <i>afastar-se-v</i> (afastar-se, apartar-se, deixar) 04. <i>desaproveitar-v</i> (desaproveitar, deixar, desperdiçar, desprezar, perder), achar (ant.); ganhar (ant.); perdedor (v.t.); perdido (v.t.) 05. <i>herdar-v</i> (herdar, deixar, legar, passar, testar, transmitir); testado (v.t.); teste (v.t.) 06. <i>sair-v</i> (sair, deixar); permanecer (ant.); deixa (v.t.) 07. <i>abandonado-adj.</i> (abandonado, deixado, desamparado, largado)
dele		x		Sugere substit. por <i>gele, pele, sele</i>			x	Sugere <i>delatado</i> .

deu	x		03 acep.: dar-verb. (sin. <i>dar, doar</i>); acomodar-se-verb. (sin. <i>acomodar-se, acostumar-se, adaptar-se, afazer-se, afeiçãoar-se, ajustar-se, amoldar-se, dar-se, habituar-se, identificar-se, modelar-se</i>); acontecer-verb. (sin. <i>acontecer, dar-se, ocorrer, passar-se, suceder</i>).			12 acep. 01. Abonar-v (sin. abonar, dar, oferecer); abono (v.t.); aprovação (v.t.) 02. abrolhar-v (abrolhar, dar, produzir); fruto (v.t.) 03. apresentar-v (apresentar, dar, deparar, oferecer, prestar, proporcional); proporcional (v.t.); proporção (v.t.) 04. conceder-v (conceder, conferir, dar, prestar); prestação (v.t.) 05. doar-v (doar, dar, presentear) 06. emitir-v (emitir, dar, ir); risada (v.t.) 07. emprestar-v. (emprestar, dar); receber (ant.); tomar (ant.); prestar (v.t.) 08. fazer avançar – v (sin. fazer avançar, dar, promover); promoção (v.t.) 09. lançar-v (lançar, dar, soltar); lançamento (v.t.) 10. oferecer-v (oferecer, dar, ofertar, trazer); receber (ant.) 11. elemento-subst. (sin. elemento, dado, informação, número); informenon (v.t.) 12. dados-subst. (sin. dados, dado, informação, número); informação (v.t.) 02. carecer-01. Obrigação-v (sin. obrigação, dever) 02. carecer-v. (carecer, dever, necessitar, precisar); dever (v.t.); débito (v.t.)
deve	x		Remete a dever-subst. (sin. <i>dever, obrigação</i>); dá uma lista de sugestões a serem verificadas: <i>leve, neve, leve</i> .	x		06 acep. 01. alterar-v (sin. alterar, discutir acaloradamente, disputar, rezingar); 02. concorrer-v (concorrer, disputar, pleitear); 03. discutir-v (discutir, disputar, questionar) 04. esforçar-se-v (esforçar-se, disputar, lutar) 05. pleitear-v (pleitear, disputar); pleito (v.t.) 06. apostado-adj. (apostado, disputado, pleiteado)
disputava	x		Remete a <i>disputas e disputa</i> , que por sua vez remetem à mesma acep. alteração-subst. (sin. <i>alteração, contenda, contestação, debate, diferença, querela, rixa</i>)	x		05 acep. 01. Articular-v (articular, dizer, falar, proferir, pronunciar) 02. contar-v (contar, descrever, dizer, expor, narrar, proferir, referir, relatar) 03. significar-v (significar, dizer); significado (v.t.); significação (v.t.) 04. adágio-v (adágio, aforismo, anêxim, apotegma, axioma, brocardo, ditado, dito, máxima, parêmia, pensamento, proloquio, provérbio, rião, sentença) 05. celebração-subst. (celebração, dito, reza); celebrar (v.t.) Sugere subst. por <i>dó</i> .
disse	x		Remete a <i>dizer</i> apenas implicitamente; 02 acep. articular-verb. (sin. <i>articular, falar, proferir, pronunciar</i>) e contar-verb. (sin. <i>contar, narrar, referir, relatar</i>)	x		07 acep. 01. Cara metade-subst. (cara metade, consorte, costela, esposa, mulher, patroa, senhora); 02. dama-subst. (dama, madama, madame, matrona, mulher, senhora); cavalheiro (ant.) 03. D. –outro (D, Digno, Dom) 04. Da. – outro (sem sinônimos) 05. amo-subst. (amo, dono, empregador, patrão, senhor); senhoril (v.t.) 06. possuidor-subst. (possuidor, dono, senhor, titular); título (v.t.) 07. proprietário-subst. (proprietário, dono)
do	x		Sugere substit. por <i>dó, doa, doe, dom, dor, dou</i>	x		12 acep. 01. Abonar-v (sin. abonar, dar, oferecer); abono (v.t.); aprovação (v.t.) 02. abrolhar-v (abrolhar, dar, produzir); fruto (v.t.) 03. apresentar-v (apresentar, dar, deparar, oferecer, prestar, proporcional); proporcional (v.t.); proporção (v.t.) 04. conceder-v (conceder, conferir, dar, prestar); prestação (v.t.) 05. doar-v (doar, dar, presentear) 06. emitir-v (emitir, dar, ir); risada (v.t.) 07. emprestar-v. (emprestar, dar); receber (ant.); tomar (ant.); prestar (v.t.) 08. fazer avançar – v (sin. fazer avançar, dar, promover); promoção (v.t.) 09. lançar-v (lançar, dar, soltar); lançamento (v.t.)
dona	x		03 acep.: costela-subst. (sin. <i>costela, esposa, mulher, patroa, senhora, velha</i>), dama-subst. (sin. <i>dama, matrona, mulher</i>) e dono-subst. (sin. <i>dono, proprietário</i>). Note-se que <i>costela</i> refere-se à <i>mulher</i> .	x		
dou	x		03 acep.: dar-verb. (sin. <i>dar, doar</i>); acomodar-se-verb. (sin. <i>acomodar-se, acostumar-se, adaptar-se, afazer-se, afeiçãoar-se, ajustar-se, amoldar-se, dar-se, habituar-se, identificar-se, modelar-se</i>); acontecer-verb. (sin. <i>acontecer, dar-se, ocorrer, passar-se, suceder</i>).	x		

e									10. oferecer-v (oferecer, dar, ofertar, trazer), receber (ant.) 11. elemento-subst. (sin. elemento, dado, fenômeno); fenomenal (v.t.) 12. dados-subst. (sin. dados, dado, informação, números); informação (v.t.)
é	x					x			Sugere uma lista: ebonite, ebulição, eburneo, e-mail 01 acep. ser-verb. (sin. <i>ser, estar</i>) 02. ente-subst. (ente, indivíduo, ser) 03. consistir em -v (consistir em, equivaler a, ser) 04. constituir - v (constituir, ser, significar); significado (v.t.); significação (v.t.) 05. estar-v (estar, ficar, permanecer, ser); permanente (v.t.) 07. existir-v (existir, ser, viver) 08. ficar-v (ficar, ser, tornar-se) 09. querer dizer - v (querer dizer, ser, significar); significado (v.t.); significação (v.t.)
ele									Terceira pessoa singular masculino (sem sin.)
em									Sugere <i>in</i> .
em peso									Sugere <i>in</i> .
encomenda									04 acep. 01. Consignação-subst. (consignação, recomendação); consignar (v.t.) 02. encomendação-subst. (encomendação); encomendar (v.t.) 03. confiar-v (confiar, encomendar, recomendar-se) 04. encarregar-v (encarregar, encomendar, incluir, recomendar); esquecer (ant.); recomendação (v.t.); recomendável (v.t.)
está	x								08 acep. 01. Encontrar-se-v (encontrar-se, estar, fazer); jazigo (v.t.) 02. ficar-v (ficar, estar, permanecer, ser); permanente (v.t.) 03. circunstância-subst. (circunstância, conjuntura, estado, posição, situação) 04. condição-subst. (condição, estado, grau, nível, situação); nivelar (v.t.) 05. condição-v (condição, estado, pé, situação) 06. gente-subst. (gente, estado, nação, povo); país (v.t.) 07. império-subst. (império, estado, reich, reino) 08. nação-subst. nação, estado, pátria, terra, torrão); patriota (v.t.); patriotismo (v.t.)
estavam									08 acep. 01. Encontrar-se-v (encontrar-se, estar, fazer); jazigo (v.t.) 02. ficar-v (ficar, estar, permanecer, ser); permanente (v.t.) 03. circunstância-subst. (circunstância, conjuntura, estado, posição, situação) 04. condição-subst. (condição, estado, grau, nível, situação); nivelar (v.t.) 05. condição-v (condição, estado, pé, situação) 06. gente-subst. (gente, estado, nação, povo); país (v.t.) 07. império-subst. (império, estado, reich, reino) 08. nação-subst. nação, estado, pátria, terra, torrão); patriota (v.t.); patriotismo (v.t.)
este									01 acep. leste-subst. (leste, levante, nascente, oriente)
eu	x								Sugere <i>eucaristia</i> .
excelentíssima									11 acep. 01. Exma.-outro (Sem sinônimos); 02. Exmo - outro (Exmo., Excelentíssimo); 03. arrebatador-adj. (arrebatador, delirante, excelente, extraordinário, fantástico, magnífico, maravilhoso, primoroso, prodigioso); delírio (v.t.)

							04. belo-adj. (bela, excelente, magnífico, virtuoso); mesquinho (ant.); virtude (v.t.) 05. bonito-adj. (bonito, brilhante, excelente, magnífico); obscuro (ant.) 06. delicioso-adj. (delicioso, esquisito, excelente, requintado) 07. distinto-adj. (distinto, especial, excelente, seleta); comum (ant.); seletividade (v.t.); seletivo (v.t.); seleção (v.t.) 08. exímio-adj. (exímio, excelente, magistral, perfeito, superlativo) 09. fino-adj. (fino, excelente, finíssimo, inestimável, magnífico, precioso, rico, sumtuoso); reles (ant.); preciosidade (v.t.) 10. primo-adj. (primo, excelente, ótimo); péssimo (ant.) 11. grau-dez-subst. (grau-dez, excelente, inexcusável, perfeito)
exclamou		x					Sugere <i>exceção</i> .
êxtase		x			x		01. acep. arrebatamento-subst. (arrebatamento, deslumbramento, deslumbre, enlêvo, fascínio); deslumbrar (v.t.); ilusão (v.t.); mara vilhoso (v.t.)
forte	x				x		15. acep. 01. Agudo-adj. (agudo, grave, intenso, violento); brando (ant.) 02. alentado-adj. (alentado, possante, potente, pujante, reforçado, robusto, vigoroso); fraco (ant.); força (v.t.) 03. arrebatado-adj. (arrebatado, ativo, energético, frenético, impetuoso, intenso, veemente); inativo (ant.); intensificar (v.t.) 04. atlético-adj. (atlético, musculoso, vigoroso); fraco (ant.) 05. carregado-adj. (carregado, intenso); fraco (ant.) 06. duradouro-adj. (duradouro, duro, durável, resistente, seguro, sólido); resistir (v.t.); resistência (v.t.) 07. excessivo-adj. (excessivo, intenso, rude, violento); fraco (ant.); intensificar (v.t.) 08. fraco-adj. (fraco, robusto, saudável); débil (ant.); saúde (v.t.) 09. impetuoso-adj. (impetuoso, sobrepujante, violento); brutal (v.t.); feroz (v.t.) 10. intensivo-adj. (intensivo, intenso) 11. intenso-adj. (intenso, volumoso); fraco (ant.); volume (v.t.) 12. macho-adj. (macho, másculo, robusto, varonil, vigoroso); machista (v.t.) 13. robusto-adj. (robusto, sólido, vigoroso, válido); fraco (ant.); validade (v.t.); validar (v.t.) 14. castelo-subst. (castelo, fortaleza) 15. cidadela-subst. (cidadela, fortaleza, fortilim, praça forte)
furiosa	x				x		03. acep. 01. Acesso-adj. (acesso, furioso, irritado, raivoso); apagado (ant.); aplacado (ant.); acender (v.t.); excitar (v.t.) 02. assanhado-adj. (assanhado, colérico, encolorizado, enraivecido, furibundo, iracundo, irado, irroso, raivento, sanhudo, zangado); acalmado (ant.); irritado (v.t.); raiva (v.t.); zangar-se (v.t.) 03. delirante-adj. (delirante, desvaivado, frenético, furioso)
ganho de causa		x					Sugere que se verifique gado, galho, galo, ganchinho, ganho, gato.
garganteando		x					
gralha	x				x		02. acep. 01. canca-subst. (cancã, piom-piom); corvo-subst. (sem sin.)
grita		x					Sugere gaita, greta, grito, gruta.
isto		x					Sugere estio, estro, exato.
juiz	x				x		02. acep. julgador-subst. (julgador, jurado); jurar (v.t.); júri (v.t.); julgador-subst. (julgador, magistrado, meritíssimo; árbitro).
juizado		x					Sugere que se utilize <i>ajuizado</i> .
julgador	x				x		02. acep. juiz-subst. (juiz, jurado); jurar (v.t.); júri (v.t.); juiz-subst. (juiz, magistrado, meritíssimo, árbitro)

juízo		x	Sugere substituição por <i>juízo</i> .	x	07 acep. 01. Ajuizamento-subst. (ajuizamento, avaliação, ponderação); ajuizar (v.t.) 02. análise-subst. (análise, apreciação, crítica, exame); criticar (v.t.); crítico (v.t.) 03. apreciação-subst. (apreciação, conceito, opinião); apreciar. 04. apreciação-subst. (apreciação, exame); julgar (v.t.) 05. audiência-subst. (audiência, sessão solene); julgar (v.t.) 06. avaliação-subst. (avaliação, conceituação); conceituar (v.t.) 07. decisão-subst. (decisão, sentença); decidir-se (v.t.); decisivo (v.t.)
justa	x		02 acep. justa-subst. (sin. <i>justa, torneio</i>), equitativo-adj. (sin. <i>equitativo, justo</i>); palavras rel. (<i>equidade, justiça</i>); acep. rel. justiciero-adj. (sin. <i>justiciero, justicioso, justo</i>).	x	06 acep. 01. Briga-subst. (briga, combate, liça, luta, pelega, pugna, torneio); 02. torneio-subst. (sem sin.) 03. equitativo-adj. (equitativo, justiciero, justicioso, justo) 04. imparcial-adj. (imparcial, justo, reto); parcial (ant.); imparcialidade (v.t.) 05. lícito-adj. (lícito, justo, válido); ilícito (ant.); validade (v.t.); validar (v.t.) 06. batalhar-v (batalhar, brigar, combater, guerrear, justar, lutar, pelejar, pugnar)
lá	x		01 acep. acolá-adj. (sin. <i>acolá, alí</i>); no entanto, <i>acolá</i> está classificado como adjetivo.	x	02 acep. 01. Acolá-outro (acolá, ali, aí); aqui (v.t.) 02. acolá-outro (acolá, além, lá ao longe)
límpidos		x	Sugere substituição por <i>limpados</i> .	x	08 acep. 01. Aberto-adj. (aberto, límpido, nítido); embaciado (ant.); abertura (v.t.); abrir (v.t.) 02. brilhante-adj. (brilhante, claro, fulgente, limpo, luzido, límpido, nítido, polido); embaciado (ant.) 03. claro-adj. (claro, cristalino, diáfano, límpido, puro, transparente); cristalizar (v.t.) 04. claro-adj. (claro, lavado, límpido, puro); lavar (v.t.) 05. desanuviado-adj. (desanuviado, desneblado, límpido, sereno); embaciado (ant.) 06. diáfano-adj. (diáfano, límpido, translúcido, transparente); 07. franco-adj. (franco, ingênuo, límpido, simples, sincero) 08. puro-adj. (puro, límpido)
maiores	x		01 acep. antecedentes-subst. (sin. <i>antecedentes, antepassados, ascendentes, avoengos, avós, pais, progenitores</i>)	x	03 acep. 01. Pais-subst. (pais, progenitores); 02. antepassado-subst. (antepassado, ascendente, avó, maior); 03. espaço-subst. (espaço, extensão, maior, termo)
mais		x	De duas opções (<i>mas, mais</i>) sugere substituição por <i>mas</i> .	x	01 acep. além disso-outro (além disso, antes, ao mesmo tempo, igualmente, novamente, outra vez, também)
maravilha	x		02 acep. bonina-subst. (sin. <i>bonina, boas-noites</i>) e maravilha-subst. (sin. <i>maravilha, milagre, prodígio</i>)	x	10 acep. 01. Admiração-subst. (admiração, impressão, nota, surpresa) admirar (v.t.), deslumbramento (v.t.) milagre (v.t.); pasmo (v.t.); surpreendente (v.t.); surpreender (v.t.); surpreso (v.t.); susto (v.t.) 02. assombro-subst. (assombro, portento, prodígio) 03. boas-noites-subst. (boas-noites, bonina, flor-de-quatro-horas, maravilha-de-forquilha) 04. fenômeno-subst. (fenômeno, prodígio); fenomenal (v.t.) 05. abismar-v. (abismar, arrebatar, deslumbrar, enlevar, extasiar, fascinar, maravilhar); deslumbrado (v.t.); deslumbramento (v.t.) 06. arrebatar-se-v (arrebatar-se, cativar, encantar, maravilhar, seduzir); desiludir (ant.) 07. assombrar-v (assombrar, maravilhar, surpreender); surpresa (v.t.) 08. impressionar-v (impressionar, maravilhar, surpreender); admirar (v.t.); deslumbrar (v.t.); pasmar (v.t.); surpreendente (v.t.); surpresa (v.t.); surpreso (v.t.) 09. abismado-adj. (abismado, apaixonado, arrebato, deslumbrado, enlevado, estatelado, estático, exaltado, extasiado, extático, fascinado, maravilhado); deslumbramento (v.t.); deslumbrar (v.t.); iludido (v.t.); maravilhoso (v.t.) 10. admirado-adj. (admirado, impressionado, maravilhado, surpreso); admirar (v.t.); deslumbrado (v.t.); pasmaado (v.t.); surpreendente (v.t.); surpreender (v.t.); surpresa (v.t.); surpreso (v.t.)

mas				01 acep. mas -outro (sin. <i>mas, porém</i>)			01 acep. contido-outro (contido, entretanto, no entanto, porém, todavia)
matraca	x			01 acep. borralhar -subst. (sin. <i>borralhara, papa-ovo</i>)	x		01 acep. borralhara-subst. (borralhara, papa-ovo)
matraqueou		x				x	Sugere mata, mata-bicho, mata-borrão, matador, mata-mouros.
melhor		x		Sugere dois substituições: <i>melhora e melro.</i>	x		02 acep. 01. mais bem-adj. (mais bem, mais perfeito); pior (ant.) 02. mais bom-adj. (mais bom); pior (ant.)
melodiosos		x			x		03 acep. 01. Agradável-adj. (agradável, doce, melódico, suave); melodia (v.t.) 02. entoado-adj. (entoado, harmonioso, melódico, suave); modulação (v.t.) 03. harmonioso-adj. (harmonioso, melódico, sonoro, suave); sonorizar (v.t.); sonorização.
mergulhado	x			01 acep. imerso -adj. (sin. <i>imerso</i>); 01 acep. rel. imerso -verbo (sin. <i>imerso, mergulhar</i>)	x		03 acep. 01. Imerso-adj. (sem sin.); 02. afundar-v. (afundar, mergulhar); fundo (v.t.) 03. imergir-v. (imerso, mergulhar)
meritíssimo			x			x	02 acep. 01. Juiz-subst. (juiz, julgador, magistrado, árbitro); 02. MM. - outro (sem sin.)
mestre	x			02 acep. catedrático -subst. (sin. <i>catedrático, lente, professor</i>) e educador -subst. (sin. <i>educador, instrutor, mentor, professor</i>); pal. Rel. (<i>educar, instruir</i>).	x		03 acep. 01. Doutor-adj. (doutor, esculápio, licenciado, médico); doutorado (v.t.) 02. catedrático-subst. (catedrático, docente, doutor, educador, instrutor, lente, mentor, professor, titular, treinador); 03. conselheiro-subst. (conselheiro, guia, guru, orientador)
morre	x			01 acep. acabar -verbo. (sin. <i>acabar, expirar, extinguir-se, falecer, fenecer, finar-se, perecer</i>)	x		07 acep. 01. Acabar-v (acabar, expirar, extinguir-se, falecer, fenecer, finar-se, morrer, perecer); começar (ant.); nascera (ant.) 02. apitar-v (apitar, morrer); apito (v.t.) 03. expirar-v. (expirar, falecer, morrer, partir); partida (v.t.) 04. aprisionado-adj. (aprimado, caçado, morto); caçar (v.t.) 05. cadáver-adj. (cadáver, defunto, morto); cadavérico (v.t.) 06. defunto-adj. (defunto, falecido, finado, morto); vivo (ant.); atalide (v.t.); funeral (v.t.); fúnebre (v.t.) 07. inanimado-adj. (inanimado, inerte, morto)
muito		x		Sugere mutilo .		x	04 acep. 01. Abundantemente-outro (abundante, com excesso); pouco (ant.); 02. assaz-outro (assaz, bem, extremamente) 03. com intensidade-outro (com intensidade, em alto grau); pouco (ant.) 04. em abundância-outro (em abundância, em grande número); pouco (ant.)
música	x			01 acep. musical -adj. (sin. <i>musical</i>)	x		06 acep. 01. Bolero-subst. (bolero, canção) 02. composição-subst. (composição, harmonia, melodia, peça, tema, ária); cacofonia (ant.); silêncio (ant.); banda (v.t.); canção (v.t.); concerto (v.t.); musical (v.t.); músico (v.t.); músico (v.t.); orquestra (v.t.); sinfonia (v.t.); ópera (v.t.) 03. cânon-subst. (cânon, cânone) 04. mús. - subst. (sem sin.) 05. solfa-subst. (solfa); musical (v.t.); músico (v.t.); músico (v.t.) 06. musical-adj. (musical, músico)
nada				01 acep. cifra-subst. (cifra, zero)	x		20 acep. 01. Ausência-subst. (ausência, coisa nenhuma, coisa nula); tudo (ant.) 02. bagatela-subst. (bagatela, insignificância, ninharia); tudo (ant.) 03. coisa nenhuma-subst. (coisa nenhuma, migalha) 04. espaço-subst. (espaço, vazio, vácuo) 05. lufas-subst. (lufas, necas, nulidade, zero)

						06. abrolhar-v (abrolhar, aparecer, brotar, nascer, passar a existir, surgir, vir ao mundo); desaparecer (ant.); morrer (ant.); início (v.t.); nascimento (v.t.) 07. aparecer-v. (aparecer, apontar, despontar, nascer, raiar, surgir); sumir (ant.) 08. aparecer-v (aparecer, brotar, nascer, surgir); morrer (ant.); broto (v.t.) 09. aparecer-v. (aparecer, nascer, sair, surgir); morrer (ant.); 10. começar-v. (começar, iniciar, nascer, principiar, Ter início); terminar (ant.); nascimento (v.t.); 11. derivar-v (derivar, nascer, originar-se, proceder, prover); morrer (ant.); nascimento (v.t.); origem (v.t.) 12. despontar-v. (despontar, nascer, raiar, surgir); surgimento (v.t.) 13. encarnar-se-v (encarnar-se, humanar-se, nascer); morrer (ant.); nascimento (v.t.) 14. formar-se-v (formar-se, instituir-se, nascer); morrer (ant.); nascimento (v.t.) 15. gerar-se -v (gerar-se, nascer, surgir); morrer (ant.); nascimento (v.t.) 16. partir -v (partir, nascer, principiar); morrer (ant.); partida (v.t.) 17. vir ao mundo-v (vir ao mundo, nascer, vir à luz); morrer (ant.); nascimento (v.t.) 18. nado-adj (nado, nascido, nato); 19. nado-adj. (nado, nascido, nato); 20. boiar-v (boiar, flutuar, nadar, sobrenadar); a fundar (ant.)
não nasce	x x	Sugere nano. Sugere passee.	x x	x	x	Sugere <i>nazi</i> . Acep. 01. Abrolhar-v (abrolhar, aparecer, brotar, nascer, passar a existir, surgir, vir ao mundo); desaparecer (ant.); morrer (ant.); início (v.t.); nascimento (v.t.); 02. aparecer-v (aparecer, apontar, despontar, nascer, raiar, surgir); sumir (ant.) 03. aparecer-v (aparecer, brotar, nascer, surgir); morrer (ant.); broto (v.t.) 04. aparecer-v (aparecer, nascer, sair, surgir); morrer (ant.) 05. começar-v. (começar, iniciar, nascer, principiar, Ter início); morrer (ant.); terminar (ant.); nascimento (v.t.) 06. derivar-v (derivar, nascer, originar-se, proceder, prover); morrer (ant.); nascimento (v.t.); origem (v.t.) 07. despontar-v (despontar, nascer, raiar, surgir); surgimento (v.t.) 08. encarnar-se-v (encarnar-se, humanar-se, nascer); morrer (ant.); nascimento (v.t.) 09. formar-se -v (formar-se, instituir-se, nascer); morrer (ant.); nascimento (v.t.); 10. gerar-se -v (gerar-se, nascer, surgir); morrer (ant.); nascimento (v.t.) 11. partir-v (partir, nascer, principiar); morrer (ant.); partida (v.t.) 12. vir ao mundo-v (vir ao mundo, nascer, vir à luz); morrer (ant.); nascimento (v.t.) 13. nado-adj (nado, nascido, nato); 14. nado-adj. (nado, nascido, nato);
nem	x	Sugere bem, tem, unem, vem.	x		X	02 acep. 01. Nem um -outro (sem sin.) 02. qualquer -outro (sem sin.)
nenhum	X			x		Sugere nó, nu
no	X				x	
o	X				X	
ordenou	x	01 acep. classificar-verb. (classificar, coordenar, dispor, distribuir)		x		10 acep. 01. Classificar-v (classificar, coordenar, dispor, distribuir, ordenar) 02. colocar em ordem-v (colocar em ordem, ordenar); desordenar (ant.) 03. comandar-v (comandar, ordenar, preceituar, prescrever); comando (v.t.) 04. decretar-v (decretar, determinar, estabelecer, exigir, mandar, ordenar); comando (v.t.) 05. decretar-v (decretar, estabelecer, formular, legislar, ordenar) 06. elaborar-v (elaborar, ordenar) 07. composto-adj. (composto, corrigido, ordenado) 08. metódico -adj. (metódico, ordenado, sistemático); desordenado (ant.); sistema (v.t.)

								09. cachê-subst. (cachê, ordenado) 10. fêria-subst. (fêria, ordenado, salário)
orelhas	x		01 acep. orelha-subst. (orelha, ouvido)	x				02 acep. 01. Aurícula-subst. (aurícula, cavidade, orelha); auricular (v.t.) 02. ouvido-subst. (ouvido, orelha)
os		X				x		Sugere oscilação.
ou		X				X		
outras		x	Sugere outras.	x				04 acep. 01. Diferente-adj. (diferente, distante, distinto, diverso, outro); 02. imediato-adj. (imediato, outro, seguinte, ulterior); 03. mais um – outro (mais um, outro, um novo) 04. o restante – outro (o restante, o resto, outro)
ouvidos	x		01 acep. orelha-subst. (orelha)	x				02 acep. 01. Escutar-v. (escutar, ouvir); 02. orelha-subst. (orelha, ouvido)
para	x		02 acep. 01. A-outro (a); 02. Cessar-verb. (cessar, descontinuar, deter, interromper, parar, sobrestar, sustar).	x				01 acep. a-outro (a), de (v.t.)
passo	x		02 acep. 01. Braço-de-mar – subst. (braço-de-mar, estreito, mancha); 02. acontecer-verb. (acontecer, dar-se, ocorrer, passar-se, suceder)	x				11 acep. 01. braço-de-mar-subst. (braço-de-mar, estreito); baixa-mar (v.t.), mar (v.t.) 02. caminhar-v (caminhar, marcha) 03. comunicação-subst. (comunicação, ligação, passado, passagem); passar (v.t.) 04. etapa-subst. (sem sin.) 05. agüentar-v (agüentar, lidar, passar, sofrer, suportar) 06. atravessar-v (atravessar, cruzar, estender-se, passar, transpor) 07. decorrer-v (decorrer, mover, passar) 08. decorrer-v (decorrer, passar, surgir, transcorrer), surgimento (v.t.) 09. deixar-v (deixar, herdar, legar, passar, testar, transmitir); testado (v.t.); teste (v.t.) 10. influir-v. (influir, passar, transmitir) 11. sair-v (sair, passar); entrar (ant.)
penas			05 acep. aflição-subst. (aflição, agonia, amargura, angústia, ansiedade, consternação, desgosto, dor, incômodos, inquietação, mágoa, opressão, padecimento, pesar, sofrimento, suplício, tormento, tortura, trabalhos, transe, tribulação, tristeza). 02. aparato-subst. (aparato, pena de escrever) 03. castigo-subst. (castigo, punição) 04. clemência-subst. (clemência, comiseração, compaixão, dó, indulgência, misericórdia, piedade) 05. dor-subst. (dor, sentimento) Palavras relacionadas: sofrer	x				09 acep. 01. Aflição-subst. (aflição, cruz, infortúnio, pena, trabalhos); cruzar (v.t.) 02. aparato-subst. (aparato, caneta, esferográfica, pena) 03. castigo-subst. (castigo, corretivo, pena, penalidade, punição, correção (v.t.); corrigir (v.t.), penalizar (v.t.); 04. clemência-subst. (clemência, comiseração, compaixão, dó, indulgência, misericórdia, pena, piedade) 05. coima-subst. (coima, multa, pena); castigo (v.t.) 06. comiseração-subst. (compaixão, cálice, dor, dó, lástima, mágoa, pena, percepção, pesar, piedade, sentimento); lastimar (v.t.) 07. lástima-subst. (lástima, pecado, pena, tristeza); pecar (v.t.); 08. pluma-subst. (pluma, pena); plumagem (v.t.) 09. sanção-subst. (sanção, pena); sancionar (v.t.)
piaram	x		01 acep. piar-verb. (piar, pipiar, pipilar, pipitar)	x				01 acep. pipiar-v (pipiar, piar, pipilar, pipitar); ave (v.t.)
pois	x		01 acep. porque-outro (porque, como, já que, pois que, por isso que, porquanto, porque, uma vez que, visto como, visto que).	x				01 acep. como-outro (como, já que, pois que, por isso que, porquanto, porque, uma vez que, visto como, visto que).

ponto	x		que, uma vez que, visto como, visto que) 02 acep. 01. Cachoeira-subst. (cachoeira, cascata, catadupa, catarata, corredeira, corridas, queda, rápido, salto); 02. oceano-subst. (oceano, mar, pego, pélagos)	x	06 acep. 01. Alvo-subst. (alvo, mira, mosca, pontaria); 02. assunto-subst. (assunto, mote, questão, tema, tese) 03. bico-subst. (bico, biscate, galho, gancho, serviço avulso, trabalho avulso); 04. extremidade-subst. (extremidade, extremo, fim, parada, ponta, termo) 05. local-subst. (local, parte, zona); zoneamento (v.t.) 06. pontículo-subst. (pontículo, pontinho)
por		x	Sugere pôr, porá.	X	
porque	x		01 acep. porque-outro (porque, como, já que, pois, pois que, por isso que, porquanto, que, uma vez que, visto como, visto que)	x	04 acep. 01. Aquele-outro (aquele, como, esse, qual, que, quão); 02. assim como – outro (assim como, bem como, como, quando) 03. como-outro (como, já que, pois, pois que, por isso que, porquanto, uma vez que, visto como, visto que) 04. por isso que – outro (por isso que, porquanto, visto que)
possui	x		01 acep. possuir-verb. (possuir, ter)	x	01 acep. haver-v (haver, possuir, ser dono de, ter)
postar-se		x	Sugere postres.	X	
pretensão		x	Sugere pretendo, prevenção.	x	03 acep. 01. Ambição-subst. (ambição, anseio, aspiração, cobiça, vontade); pretender (v.t.); pretencioso (v.t.) 02. candidatura-subst. (candidatura); candidato (v.t.) 03. presunção-subst. (presunção); pretender (v.t.); pretencioso (v.t.)
primeiríssima		X		x	03 acep. 01. Inicial-adj. (inicial, primeiro, primitivo); derradeiro (ant.); último (ant.); início (v.t.); início (v.t.) 02. primo-adj. (primo, primeiro); último (ant.) 03. 1º-outro (1º, primeiro)
próprios	x		03 acep. 01. Adequado-adj. (adequado, azado, competente, conveniente, oportuno); 02. essencial-adj. (essencial, inerente, inseparável, intrínseco); 03. mesmo-adj. (mesmo)	x	07 acep. 01. Adequado-adj. (adequado, cômodo, favorável, próprio); inadequado (ant.) 02. adequado-adj. (adequado, azado, competente, conveniente, oportuno, próprio) 03. adequado-adj. (adequado, bom, próprio); impróprio (ant.); bondade (v.t.); bondoso (v.t.) 04. essencial-adj. (essencial, inerente, inseparável, intrínseco, próprio) 05. harmônico-adj. (harmônico, igual, proporcionado, proporcional, próprio, regular, simétrico); assimétrico (ant.); desproporcional (ant.); proporcionar (v.t.); proporção (v.t.); simetria (v.t.) 06. mensageiro-adj. (mensageiro, portador, próprio); impróprio (ant.) 07. mesmo-adj. (mesmo, próprio)
pulinho		x	Sugere sulino.	x	02 acep. 01. Batimento-subst. (batimento, estremecimento, latejo, palpitação, palpite, pancada, pulo, pulsação); 02. cabriola-subst. (cabriola, pincho, pinote, pirueta, pulo, salto, transposição)
pura	x		01 acep. castiço-adj. (castiço)	x	11 acep. 01. Bucólico-adj. (bucólico, ingênuo, puro); malicioso (ant.); bucolismo (v.t.) 02. castiço-adj. (castiço, legítimo, puro); 03. casto-adj. (casto, imaculado, inocente, modesto, puro, virgem, virginal); sensual (ant.); castidade (v.t.); virgindade (v.t.) 04. casto-adj. (casto, pudico, puro, virtuoso); virtude (v.t.) 05. claro-adj. (claro, cristalino, diáfano, límpido, puro, transparente); cristal (v.t.); cristalizar (v.t.) 06. claro-adj. (claro, lavado, límpido, puro); lavar (v.t.) 07. extremo-adj. (extreme, genuíno, puro, seleteo, simples, virgem) 08. ilibado-adj. (ilibado, imaculado, impoluto, incorrupto, intemerato, puro) 09. intacto-adj. (intacto, puro, virgem, virginal, virgineo) 10. intato-adj. (intato, intocado, puro, virgem); impuro (ant.); virgindade (v.t.) 11. límpido-adj. (límpido, puro)
quando		x	Sugere quando, quanto, quinado, suando.	x	01 acep. assim como – outro (assim como, bem como, como, porque)
que	x		02 acep. 01. Porque-outro (porque, como,	x	01 acep. aquele-outro (aquele, como, esse, porque, qual, quão)

ser			02 acep. ente-subst. (ente); ser-verb. (ser, estar)	x		08 acep. 01. Criatura-subst. (criatura, ente, homem, indivíduo, pessoa); 02. ente-subst. (ente, indivíduo) 03. consistir em -v (consistir em, equivaler a) 04. constituir -v (constituir, significar); significado (v.t.); significação (v.t.) 05. estar-v (estar, ficar, permanecer); permanente (v.t.) 06. existir-v (existir, viver) 07. ficar-v (ficar, tornar-se) 08. quero dizer-v (quero dizer, significar); significado (v.t.); significação (v.t.)
séria	x		01 acep. ajuizado-adj. (ajuizado, assisado, avisado, circunspeto, cordato, discreto, grave, judicioso, ponderado, prudente, sensato, sisudo, sério); imponderado (ant.); insensato (ant.); irresponsável (ant.); prudente (v.t.); sensatez (v.t.) 02. circunspeto-adj. (circunspeto, composto, sério) 03. firme-adj. (firme, sério, sólido); solidificar (v.t.)	x		03 acep. 01. Aconselhado-adj. (aconselhado, ajuizado, assisado, atinado, avisado, circunspeto, controlado, cordato, discreto, grave, judicioso, ponderado, prudente, sensato, sisudo, sério); imponderado (ant.); insensato (ant.); irresponsável (ant.); prudente (v.t.); sensatez (v.t.) 02. circunspeto-adj. (circunspeto, composto, sério) 03. firme-adj. (firme, sério, sólido); solidificar (v.t.) 01 acep. convir-v (convir, quadrar, servir, toar); toada (v.t.)
servirá	x		01 acep. empregar-verb. (empregar, servir-se, usar, utilizar)	x		02 acep. 01. Abandonado-adj. (abandonado, desamparado, viúvo); amparado (ant.); viuvez (v.t.) 02. apenas-outro (apenas, exclusivamente, somente)
só	x		03 acep. 01. Apenas-outro (apenas, exclusivamente, somente); 02. singular-adj. (singular, único); 03. só-adj. (só, sozinho)	x		
sua	x		01 acep. 01. Suar-verb. (suar, transpirar) Pal. rel. suor, transpiração	x		01 acep. transpirar-v. (transpirar, suar)
surdos		x	Sugere surdos, surtos	x		03 acep. esconso-adj. (esconso, oculto, secreto, surdo); surdez (v.t.); 02. impassível-adj. (impassível, indiferente, insensível, surdo); sensível (ant.); surdez (v.t.) 03. mouco-adj. (mouco, surdo); surdez (v.t.)
terminada	x		01 acep. concluir-verb. (concluir, acabar, aprontar, finalizar, findar, rematar, ultimar)	x		06 acep. 01. Acabar-v. (acabar, aprontar, completar, concluir, finalizar, findar, rematar, terminar, ultimar) 02. arrematar-v (arrematar, concluir, terminar) 03. concluir-v (concluir, encerrar, fechar-se, rematar, terminar); começar (ant.); encerramento (v.t.); término (v.t.) 04. concluir-v. (concluir, pôr fim, rematar, terminar, trancar); iniciar (ant.); tranca (v.t.); trancado (v.t.) 05. acabado-adj. (acabado, aprontado, complementado, completado, completo, concluído, finalizado, findo, pronto, rematado, terminado, ultimado); incompleto (ant.); completar (v.t.) 06. apagado-adj. (apagado, extinto, terminado)
togado		x	Sugere rogado, tocado, tomado		X	
topam		x	Sugere tocamos, toma, tomam.	x		08 acep. 01. Aceitar-v. (aceitar, topar); 02. achar-v. (achar, encontrar, topar) 03. acordar-v (acordar, concordar, topar) 04. bater-v. (bater, chocar-se, colidir, tocar, topar, trombar); trombada (v.t.) 05. dar com o pé - v (dar com o pé, topar) 06. deparar-v (deparar, topar) 07. embarrar-v (embarrar, esbarrar, topar) 08. jogar contra-v (jogar contra, topar)
topamos		x	Sugere tapamos, tocamos, tomamos.	x		08 acep. 01. Aceitar-v. (aceitar, topar); 02. achar-v. (achar, encontrar, topar) 03. acordar-v (acordar, concordar, topar) 04. bater-v. (bater, chocar-se, colidir, tocar, topar, trombar); trombada (v.t.) 05. dar com o pé - v (dar com o pé, topar)

								06. deparar-v (deparar, topiar) 07. embarrar-v (embarrar, esbarrar, topiar) 08. jogar contra-v (jogar contra, topiar)
trinos	x						x	Sugere triagem.
um	x						x	02 acep. 1 -outro (sem sin.) 02. algum -outro (algum, qualquer)
uma	x						X	02 acep. 1 -outro (sem sin.) 02. algum -outro (algum, qualquer)
valia	x						x	05 acep. 01. Valimento-subst. (valimento, valor); 02.. acudir-v (acudir, ajudar, amparar, auxiliar, defender, proteger, salvar, socorrer, valer) 03. custar-v (custar, importar, valer) 04. importar-v (importar, montar, representar, significar, valer); appear (ant.), demontar (ant.) 05. viger-v (viger, valer, vigorar)
vamos	x						x	03 acep. 01. Frequentar-v (frequentar, ir); freqüente (v.t.), freqüência (v.t.) 02. largar-v (largar, ir, partir, sair); voltar (ant.) 03. navegar-v (navegar, ir); navegação (v.t.)
veio	x						x	08 cep. 01. Depósito-subst. (depósito, filão, jazida, jazigo, mina, veiro); 02. Laivo-subst. (laivo, listra) 03. mina-subst. (mina, nascente); 04. riacho-subst. (riacho, ribeiro, veia); 05. chegar-v. (chegar, surgir, vir); surgimento (v.t.) 06. chegar-v (chegar, vir); partir (ant.) 07. decorrer-v (decorrer, derivar, dimanar, emanar, proceder, provir, vir) 08. nativo-adj. (nativo, natural, originário, oriundo, procedente, proveniente, vindo); im procedente (ant.); proceder (v.t.); procedência (v.t.)
voz	x						x	04 acep. 01. Boato-subst. (boato, rumor, ruído, zunzunzum); 02. brado-subst. (brado, clamor, vozear) 03. fala-subst. (sem sin.) 04. palavra-subst. (sem sin.)
zurrou	x						x	01 acep. ornear-v (ornear, ornejar, rebusnar, zurrar)

ANEXO 03

**Breve análise dos resultados obtidos
através do levantamento realizado no anexo 02**

As deficiências dos *thesauri* existentes

Os aplicativos do tipo “*thesaurus*” eletrônico disponíveis para uso no Brasil apresentam alguns problemas; uma evidência disso é o “Report on the Brazilian Thesaurus Linguistic Evaluation”, documento elaborado em 1998 por uma equipe da Microsoft, apresentando uma análise das deficiências do *thesaurus* acoplado ao pacote Office 9, portanto, incluído na versão 97 do redator de textos Microsoft Word. A partir da seleção de textos de autores como Lima Barreto, Millor Fernandes, Nelson Rodrigues, Oswald de Andrade, Luís Fernando Veríssimo e Machado de Assis, além de dois artigos de jornal, os examinadores procederam à aplicação do *thesaurus* a todos os textos, palavra por palavra. A partir dos resultados encontrados, procederam à sua avaliação, classificando os tipos de problemas em sete tipos, que arrolamos abaixo seguidos de exemplos e das sugestões de correção da equipe:

- **Qualidade de sinônimo *** - a lista de sinônimos apresenta um número considerável de entradas corretas, mas pode ser melhorada pela adição de mais sinônimos. Ex: *gordo*. “Apresenta poucos termos na lista. Eu sugeriria a inclusão de *volumoso* (no sentido de *grande* – *grande maço de notas* = *volumoso maço de notas*” (Op. cit. p. 06).
- **Qualidade de sinônimo **** - A lista de sinônimos mostra um número considerável de entradas corretas, mas não inclui as ocorrências mais comuns no português do Brasil, incluindo sinônimos que não são comuns, ou incluindo itens corretos e incorretos em relação à entrada principal. Ex: *visão*. “A lista não inclui o primeiro significado de *visão*, que é *ver*. Incluiríamos o sentido: *enxergar (verbo)*, com os sinônimos relevantes *observar, perceber* e então *ponto de vista (substantivo)*, com os sinônimos relevantes *percepção, compreensão*” (Op. cit. p. 11).

- **Qualidade de sinônimo ***** - Nenhum dos sinônimos apresentados na lista foram encontrados no material de referência. Ex: *humilhações*. “*Humilhação* e *humildade* não são sinônimos. Eu sugeriria *rebaixamento, vexame, afronta*.” (Op. cit. p.06).
- **Ordem imprópria** – sinônimos que não são apresentados de acordo com a ordem do sinônimo mais apropriado, mais comum vindo em primeiro lugar. Ex: *deformar*. “Eu colocaria *abastardar* em último lugar na lista” (Op. cit. p. 06).
- **Categorização gramatical incorreta** – significados com o código de categoria gramatical incorreto. Ex: *produto*. “*Fruto* é um substantivo, não um adjetivo; incluiríamos na lista de sentidos *mercadoria* e *receita*, e na lista de sinônimos, associado a *fruto, resultado*” (Op. cit. p.15).
- **Ausência de diferentes categorias gramaticais** – faltam sinônimos para categorias gramaticais diferentes. Ex: *veste*. “*Veste* está sendo considerado apenas como substantivo, mas também é uma flexão do verbo *vestir*, significando *trajar, envergar*” (Op. cit. p. 12).
- **Sinônimos controversos** – sinônimos que poderiam ser considerados ofensivos ou politicamente incorretos. Ex: *mulher*. “*Costela* e *velha*” (Op. cit. p.08).

Após esta análise, o relatório da Microsoft concluiu o seguinte:

- 1- apenas 17% das palavras analisadas foram encontradas no *thesaurus*.
- 2- A maior parte dos erros recai sobre a categoria qualidade de sinônimos.
- 3- A segunda maior fonte de erros acontece com o tipo categorias gramaticais diferentes, seguida de perto pelo problema da ordem imprópria de apresentação dos sinônimos.
- 4- O item sinônimos controversos merece cuidadosa revisão.
- 5- O não reconhecimento de itens poderia ser amenizado pela inclusão de um número maior de entradas, sem esquecer de vocabulário especializado e termos estrangeiros.

- 6- O aplicativo reconheceu a maioria das formas verbais flexionadas, mas desconheceu a flexão de substantivos (diminutivos), modificadores (advérbios) e pronomes. O mesmo se aplica a lexias compostas.
- 7- Nomes, sobrenomes, acrônimos, abreviações e numerais também não foram reconhecidos (Op. cit., 1998).

Basicamente, a equipe sugere que seja feita uma revisão do aplicativo para uma sensível melhora no seu desempenho.

Para o nosso trabalho, como fator de motivação para a construção da base lexical do *thesaurus*, também realizamos um breve exame de dois aplicativos, o *Dicionário de Sinônimos do Português*, acoplado ao pacote do *Microsoft Word 97*¹² (também objeto de revisão do relatório mencionado anteriormente) e o *Dicionário de Sinônimos para a Língua Portuguesa* acoplado ao pacote *Microsoft Word 2000*¹³. A partir de um texto de Monteiro Lobato¹⁴, que serviu como um *mini-corpus*, separamos todos os itens lexicais, excluimos os itens repetidos e conservamos algumas expressões do texto original; para análise, verificamos os itens originais, tais como se apresentam no texto, ou seja, flexionados; no total, foram verificadas 142 ocorrências com o auxílio dos aplicativos analisados e os resultados foram dispostos em listas, conforme mostra o anexo ao final deste trabalho. Feito esse breve exame, as conclusões revelam algumas qualidades e deficiências dos dicionários analisados¹⁵. Os problemas identificados serão arrolados a seguir, acompanhados de comentários que apontem

¹² *Dicionário de Sinônimos do Português*, 1993, por INSO Corporation. Adaptado do *Dicionário de Sinônimos de Antenor Nascentes*, 1981, Nova Fronteira. Todos os direitos reservados.

¹³ *Tesouro para a Língua Portuguesa*. Versão 7. Itautec Philco S. A. Todos os direitos reservados.

¹⁴ Ver anexo 01.

¹⁵ Para efeito de simplificação, chamaremos v.97 e v.2000 às respectivas versões dos aplicativos acoplados ao *Microsoft Word 97* e *Microsoft Word 2000*.

para a necessidade de evitar tais problemas na construção de um novo dicionário eletrônico de sinônimos e antônimos para a língua portuguesa:

- a) **quantidade de itens reconhecidos** – dos 142 itens verificados, ficou assim a proporção entre itens reconhecidos –R- e não reconhecidos –NR-:

Reconhecimento	v.97	v.2000
R	70	107
NR	72	35

- ✓ o reconhecimento de itens lexicais está diretamente relacionado com o tamanho da base de dados lexicais, ou seja, com o número de unidades disponíveis para consulta. Elaborando uma base maior, certamente aumentará o número de itens reconhecidos.

- b) **quantidade de sentidos oferecidos** – o número de acepções (sentidos) oferecidos pelos verbetes variou de um aplicativo para outro:

Sentidos oferecidos	Total	Média (sentidos/verbo)
v.97	112	0,78
v.2000	553	3,89

- ✓ Da mesma forma que a quantidade de dados disponíveis para consulta facilitará o reconhecimento de unidades pelo aplicativo, também poderão ser inseridos mais grupos de sinônimos, dando maior completude a cada um dos sentidos dos verbetes consultados.

- c) **categorias gramaticais diferentes** – por vezes, ocorre de uma determinada ocorrência ser rotulada de maneira incorreta. Ex:

- (v. 97) entrada *lá* – primeira acepção – *acolá* – adjetivo;
- falta de atribuição de certos rótulos, englobados no genérico “outro”: (v.97) *porque* – outro; (v. 2000) *como-outro*, *cá-outro*, *depressa-outro*, *com intensidade* – outro, etc.
- na v.2000, rótulos inadequados como *fazer avançar* – verbo, *consistir em* – verbo, etc.

- ✓ A atribuição atenta e correta de rótulos gramaticais contribuirá para aumentar a fidedignade das informações contidas na base de dados lexicais.

d) **não reconhecimento de itens flexionados** – Ex:

- (v.97) - não se reconhece os itens *cantam* e *cantou*, sugerindo substituições por itens como *canto*, *cantos*, *captou*, etc.
- (v.2000) – não se reconhece o item *exclamou*, sugerindo-se substituição pelo item *exceção*.

- ✓ Constata-se que os aplicativos não possuem um bom analisador morfológico, que remeteria o item flexionado à sua forma canônica, sem fazer sugestões de substituição não pertinentes.

e) **acepções controversas** – para o verbete *dona*, a v.97 oferece a acepção **costela**, o mesmo acontecendo na versão 2000, só que enquanto sinônimo da acepção **cara metade**. O dicionário *Aurélio do Séc. XXI* dá *costela* como *brasileirismo popular*, portanto questionamos sua sinonímia com *mulher*.

- ✓ Tal problema revela falta de precisão e aponta para a necessidade de se montar os grupos de sinônimos com maior acuidade, comparando informações com vários dicionários e até mesmo as questionando, com base em consulta a *corpus*, por exemplo. Lembramos que a v.97 foi construída com base na obra de Nascentes, 1981; lá, a informação se apresenta da seguinte forma: “Costela, dona, esposa, mulher, patroa, senhora, velha – Costela, dona,

patroa e velha são palavras familiares que lembram, respectivamente, o episódio bíblico da criação da mulher, a circunstância da posse, a subordinação, a idade” (...) p. 201. Portanto, o problema já se apresenta na própria fonte da informação do dicionário eletrônico, que foi transferido para a versão digital sem maiores questionamentos. A v.2000 apresenta o mesmo problema, mas não informa a fonte da informação lexical.

f) **acepções não relacionadas** – por vezes, algumas acepções arroladas como pertencentes a certas entradas podem ser contestáveis. Ex:

- (v. 2000) - entrada *dando*– 12. **Elemento-subst.** (sin. *Elemento, dado, fenômeno*); *fenomenal* (v.t.); 13. **Dados-subst.** (sin. *Dados, dado, informação, números*); *informação* (v.t.);
- (v. 2000) - entrada *nada*– 07. **Aparecer-v.** (*aparecer, apontar, despontar, nascer, raiar, surgir*); *sumir* (ant.); 08. **aparecer-v** (*aparecer, brotar, nascer, surgir*); *morrer* (ant.); *broto* (v.t.); 09. **aparecer-v.** (*aparecer, nascer, sair, surgir*); *morrer* (ant.); 10. **começar-v.** (*começar, iniciar, nascer, principiar, ter início*); *ter início* (ant.); *terminar* (ant.); *nascimento* (v.t.); 11. **derivar-v** (*derivar, nascer, originar-se, proceder, prover*); *morrer* (ant.); *nascimento* (v.t.); *origem* (v.t.); 12. **despontar-v.** (*despontar, nascer, raiar, surgir*); *surgimento* (v.t.); 13. **encarnar-se-v** (*encarnar-se, humanar-se, nascer*); *morrer* (ant.); *nascimento* (v.t.); 14. **formar-se-v** (*formar-se, instituir-se, nascer*); *morrer* (ant.); *nascimento* (v.t.); 15. **gerar-se –v** (*gerar-se, nascer, surgir*); *morrer* (ant.); *nascimento* (v.t.); 16. **partir –v** (*partir, nascer, principiar*); *morrer* (ant.); *partida* (v.t.); 17. **vir ao mundo-v** (*vir ao mundo, nascer, vir à luz*); *morrer* (ant.); *nascimento* (v.t.)

- (v. 2000) - entrada *está*– **06. Gente-subst.** (*gente, estado, nação, povo*); *país* (v.t.); **07. império-subst.** (*império, estado, reich, reino*); **08. Nação-subst.** (*nação, estado, pátria, terra, torrão*); *patriota* (v.t.), *patriotismo* (v.t.).

✓ Informações não relacionadas com uma entrada que está sendo consultada não deveriam ser arroladas.

g) **sinônimos controversos** – relacionada ao item e; desta vez, ao invés de acepções controversas, os próprios elementos das listas apresentam incoerências. Ex:

- (v. 97) – entrada *nada* – sin. *cifra*, *zero*;
- (v. 2000) entrada *música* – primeira acepção **bolero-subst.** (*bolero, canção*); *bolero* é um tipo de *canção*, portanto, hipônimo e não sinônimo;
- (v. 2000) entrada *morre* – 4^a acepção **aprisionado-adj.** (sin. *aprisionado, caçado, morto*); *caçar* (v.t.) – o Dicionário Aurélio do Séc. XXI não dá *morto* como sinônimo de *aprisionado*.
- (v.2000) entrada *só* – primeira acepção – **abandonado-adj.** (*abandonado, desamparado, viúvo*) – (ver Figura 15 abaixo); *amparado* (ant.); *viuvez* (v.t.);



Fig. 15. – A interface gráfica do dicionário de sinônimos v.2000, mostrando “viúvo” como sinônimo de “só”.

✓ Este problema reflete a falta de acuidade das informações contidas nos dois aplicativos; para contorná-lo, uma base de dados lexicais pode ser construída observando-se com maior rigor a fidedignidade das informações nela contidas.

h) **Interface não objetiva** – a própria interface de consulta dos verbetes não é clara ao apresentar sinônimos e sugerir substituições. Quando se quer buscar algum item pela interface, deve-se digitar a palavra no campo “inserir”, ao invés de “consultar”, em ambos os aplicativos, conforme podemos verificar nas figuras 16a e 16b abaixo.



Fig. 16a e 16b – As interfaces de consulta de ambas as versões dos aplicativos avaliados (respectivamente v.97 e v.2000); note-se que para consultar o verbete “só”, esse deve ser digitado no quadro intitulado “Inserir”.

- ✓ A interface de consulta final do dicionário eletrônico deve ser a mais amigável possível, evitando problemas de consulta do usuário.
- i) **Remissão implícita** – ao consultar um item flexionado, não há uma remissão direta à sua forma canônica. Ex:

- (v.97) - ao consultar o item *disse*, não há uma remissão explícita ao verbo *dizer*: **01. Articular-verb.** (*articular, falar, proferir, pronunciar*); **02. Contar-verb.** (*contar, narrar, referir, relatar*).
 - (v.97) – item *disputava* – remete a *disputas* e *disputa*, que por sua vez remetem ambas à mesma acepção **altercação-subst.** (*altercação, contenda, contestação, debate, diferença, querela, rixa*); não se remete ao verbo *disputar*.
- ✓ Talvez por incompletude da base de dados lexicais, não se remeta à forma canônica; mais uma vez, uma base mais completa minimizaria tal problema.
- j) **itens não reconhecidos**, mas para os quais o aplicativo sugere substituições incorretas – muitos itens não são reconhecidos pelas duas versões, mas mesmo assim são feitas sugestões de substituição. Ex:
- (v.97) – item *ele* – não reconhecido – sugestões de substituição – *gele, pele, sele*.
 - (v.2000) – item *e* – não reconhecido – sugestões de substituição: *ebonite, ebulição, ebúrneo, e-mail*.
- ✓ Mais uma vez, a questão da incompletude das bases, seguida de um problema que é a remissão a itens não relacionados com o item consultado.
- k) **antonímia** – antônimos foram encontrados apenas na v.2000, por acepção.
- ✓ Informação que nos permite concluir que a v.97 está realmente incompleta.

Os problemas apontam para uma série de aspectos que podem ser observados na construção de um *thesaurus* eletrônico para a língua portuguesa:

- a) necessidade de uma base de dados lexicais maior à disposição para consulta, o que conseqüentemente aumentaria o número de sinônimos e sentidos a serem reconhecidos pelo aplicativo;
- b) compilação rigorosa dos dados, a fim de evitar problemas de atribuição de rótulos incorreta, sentidos e sinônimos controversos, remissões implícitas e arrolagem de informação não relacionada com a entrada consultada;
- c) necessidade de inclusão de antônimos, o que não ocorre na v.97 do aplicativo analisado;
- d) necessidade de uma interface amigável de consulta para o usuário final.

Ainda assim, apontamos para a maior qualidade de uma das versões, a 2000, por apresentar maior número de entradas e sentidos além de apresentar antônimos ligados a algumas acepções.

Os resultados obtidos a partir do exame desses dois aplicativos revela a necessidade de um novo aplicativo que atenda a maiores exigências de construção de interfaces computacionais para consulta de dados, maior quantidade de dados e fidedignidade de informações nele contidas, o que confirma nossa motivação para a compilação de um *thesaurus eletrônico* para o português do Brasil.