Ariani Di Felippo

Representação Lingüístico-computacional dos Adjetivos Valenciais do Português

Araraquara
2004

Ariani Di Felippo

Representação Lingüístico-computacional dos Adjetivos Valenciais do Português

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

Orientador: Prof. Dr. Bento Carlos Dias da Silva

Araraquara
2004

Di Felippo, Ariani

Representação lingüístico-computacional dos adjetivos valenciais do português / Ariani Di Felippo. – Araraquara, 2004.
120 f.: 30 cm

Dissertação (Mestrado – Universidade Estadual Paulista, Faculdade de Ciências e Letras)
Orientador: Bento Carlos Dias da Silva

1. Lingüística 2. Língua Portuguesa 3. Léxico 4. Adjetivo

Ariani Di Felippo

Representação Lingüístico-computacional dos Adjetivos Valenciais do Português

Comissão Julgadora

Presidente e Orientador:

1º Examinador (a):

2º Examinador (a):

Aos meus pais

Agradecimentos

A **UNESP/Ar**, pelo apoio institucional.

A **CAPES**, pelo apoio financeiro.

Ao professor e orientador **Bento Carlos Dias da Silva**,
pela paciência, estímulo e orientação.

Aos **colegas e coordenadores do NILC,**
pelo apoio pessoal e profissional.

A **minha família**, pelo carinho de sempre.

Resumo

Tendo como principal motivação a necessidade cada vez mais urgente da construção de léxicos lingüisticamente motivados para aplicações do Processamento Automático das Línguas Naturais, apresenta-se, neste trabalho, uma proposta de representação lingüístico-computacional para os *adjetivos valenciais* do português do Brasil. Para tanto, buscou-se identificar essa classe de adjetivos e suas principais propriedades e investigar alguns dos principais formalismos de representação de informações léxico-gramaticais. Dessa forma, propôs-se um modelo de entrada lexical para os adjetivos valenciais baseado no construto formal das *estruturas de traços* ou *matrizes de atributo-valor*. Como contribuição prática, apresenta-se também uma proposta de inclusão de novas informações para os adjetivos da base da rede Wordnet.Br.

Abstract

Considering urgent the demands on the construction of linguistically-motivated lexicon for Natural Language Processing applications, this thesis presents a computational-linguistic representation for the Brazilian Portuguese adjectives which project a *valence* or *argument structure*. Accordingly, we identified their valence elements and main properties and investigated some outstanding lexical-grammatical representation formalisms. Then we proposed a lexical entry template for these adjectives based on *features structures* or *attribute-value matrix* formalism. As a practical contribution, we also show how part of such template can contribute to enrich the Wordnet.Br adjectival lexical database.

Lista de figuras

Figura 1: O PLN e suas disciplinas matrizes 2
Figura 2: Equacionamento metodológico dos problemas do PLN: os três domínios de Dias-da-Silva (1996, 1998) 8
Figura 3: Arquitetura de um sistema de PLN genérico de Dias-da-Silva (1996) 12
Figura 4: Arquitetura do processamento mental da linguagem de Handke (1995) e Lowie (1998). 19
Figura 5: Modelo de entrada lexical bipartida de Handke (1995) e Lowie (1998). 22
Figura 6: Representação em notação MAV, descrita com base na HPSG. 74
Figura 7: Modelo de entrada lexical para os adjetivos valenciais 79
Figura 8: Proposta de representação do adjetivo *pálido*1 83
Figura 9: Exemplificação do traço CONC 84
Figura 10: Representação do adjetivo *pálido*2 85
Figura 11: Representação da entrada do adjetivo *descendente* 86
Figura 12: Representação da entrada lexical do adjetivo *doador* 87
Figura 13: Representação da entrada lexical do adjetivo V2 *temeroso* 88
Figura 14: Representação do adjetivo V4 *transferível* 89

Lista de quadros

Quadro 1: Recursos teórico-metodológicos fornecidos pelas disciplinas matrizes do PLN 3
Quadro 2: Modelo de entrada canônica decomposta em quatro componentes..... 25
Quadro 3: Os componentes da entrada lexical e as informações léxico-gramaticais correspondentes 26
Quadro 4: Classificação distribucional dos adjetivos..... 49
Quadro 5: Padrões de subcategorização dos adjetivos valenciais 63

Sumário

Dedicatória v
Agradecimentos vi
Resumo vii
Abstract viii
Lista de figuras ix
Lista de quadros x
Sumário xi

Capítulo 1 Introdução 1
1.1. Contexto, motivação e justificativa 1
1.2. Objetivo 6
1.3. Metodologia 8
1.4. Estrutura da dissertação 10

Capítulo 2 Dos olhares sobre o "léxico" e do esquema de representação da informação lexical 11
2.1. Do olhar lingüístico-computacional sobre o léxico 11
2.2. Do léxico nos estudos lingüísticos: a macroestrutura 15
2.3. Do léxico sob o olhar da Psicolingüística 18
2.4. Da sistematização da informação lexical: esquema geral de representação 25
2.4.1. Do aprofundamento teórico do modelo e de sua utilização neste trabalho 26
2.4.1.1. Do componente FG 26
2.4.1.2. Do componente TG 27
2.4.1.3. Do componente EA 28
2.4.1.4. Do componente FS 30

Capítulo 3 Dos adjetivos do português: propriedades e classificações 32
3.1. Da categorização do léxico: considerações sobre os adjetivos 32
3.2. Da "vertente grega" do adjetivo 34
3.3. Da caracterização geral dos adjetivos do português 37
3.4. Do ponto de vista morfossintático: as marcas de gênero, de número e de grau 38
3.4.1. As desinências de *número* e de *gênero* 38
3.4.2. Do grau 40

3.5. Das subclasses dos adjetivos: os qualificadores e os classificadores 41
3.5.1. Dos *qualificadores* e a macroestrutura do léxico 41
3.5.2. Dos *classificadores* 44
3.6. Das funções sintáticas dos qualificadores e classificadores 46
3.7. Das propriedades sintático-semânticas dos adjetivos 50
3.7.1. Dos adjetivos em Padn: a função de *modificadores* 50
3.7.1.1. Da semântica dos qualificadores em Padn 50
3.7.1.2. Da semântica dos classificadores 52
3.7.2. Dos adjetivos em Ppred: a função de *predicadores* 55
3.7.3. Dos *qualificadores* e a microestrutura do léxico 56
3.7.4. Da valência adjetival 56
3.7.4.1. Da valência lógico-semântica 56
3.7.4.2 Da valência sintática 57
3.7.4.3. Da valência semântica ou *estrutura de argumentos* 66
3.8. Da delimitação dos adjetivos valenciais e de suas propriedades: uma síntese 69

Capítulo 4 Da representação lingüístico-computacional dos adjetivos valenciais 71
4.1. Dos formalismos 71
4.1.1. Da *estrutura de traços* ou *matriz de atributo-valor* 72
4.1.2. Da lógica de predicados 75
4.1.3. Dos *frames* 76
4.2. Do formalismo para a representação dos adjetivos valenciais 78
4.3. Da proposta de representação lingüístico-computacional 78
4.3.1. Das *estruturas de traços* que compõem o modelo de entrada lexical 80
4.3.1.1. Da estrutura de traço simples FG 80
4.3.1.2. Da estrutura de traços complexa DE 80
4.4. Da aplicação do modelo: alguns exemplos de entradas lexicais 82

Capítulo 5 Da proposta de edição de informações na base da rede Wordnet.Br 91
5.1. Das características da base da Wordnet.Br 91
5.2. Da proposta de inserção de novos dados 92
5.3. Da estrutura atual da base da Wordnet.Br 93
5.4. Da proposta de extensão da informação-base da rede Wordnet.Br 94
5.5. Da proposta de extensão da informação periférica da rede Wordnet.Br 96

Capítulo 6 Das considerações finais 97

Referências bibliográficas 99

Anexo 1 Textos consultados: siglas e obras 108

Capítulo 1 Introdução

1.1. Contexto, motivação e justificativa

Desde que os computadores foram introduzidos em nossa cultura, na década de 40, fazê-los "entender" instruções necessárias para a execução de tarefas tem sido um desafio para os cientistas.

A primeira solução encontrada para esse problema foi a criação das *linguagens de programação* como: COBOL, FORTRAN, PASCAL, C, PROLOG, etc. Ainda assim, a comunicação entre o homem e o computador era extremamente difícil e rígida, do ponto de vista humano, e, conseqüentemente, outra saída foi encontrada: a construção das *interfaces gráficas,* ou seja, programas que transformam a informação em objetos gráficos (ícones, menus, figuras, etc.).

A preocupação com a comunicação homem-máquina, entretanto, foi muito além. Desde os primórdios da computação, inúmeros pesquisadores se aventuram na concretização de uma outra solução para o problema: a criação de programas capazes de interpretar mensagens codificadas em *língua natural*.

Apesar do entusiasmo inicial pelo desenvolvimento de sistemas computacionais que possibilitassem que a comunicação homem-máquina fosse feita por meio de uma língua natural, as realizações nesse sentido foram mais modestas. Dada a complexidade da tarefa de criar programas computacionais capazes de "compreender" as línguas, a maioria dos pesquisadores passou a focalizar o desenvolvimento de programas que realizam tarefas bastante específicas, como *correção ortográfica, análise sintática, tradução,* entre outras. A área em que esses programas têm sido desenvolvidos recebeu o nome de Processamento Automático dos Línguas Naturais (PLN) (DIAS-DA-SILVA, 1998). Atualmente, essa área

tem recebido denominações diversas como a de Tecnologia ou Engenharia da Linguagem Humana (TLH) (CALZOLARI, ZAMPOLLI, LENCI, 2002; ATKINS, 2002). Neste trabalho, em específico, optou-se pela denominação PLN, por esta ressaltar a natureza mais científica e menos tecnológica do campo.

O PLN, segundo Dias-da-Silva (1996), constitui um domínio complexo e multifacetado, cujo objetivo é *a projeção e implementação de sistemas computacionais que processam língua natural*. Aliás, desenvolver projetos de pesquisa dessa natureza significa percorrer várias disciplinas matrizes. Com base em Dias-da-Silva (1996), tais disciplinas são descritas na Figura 1.

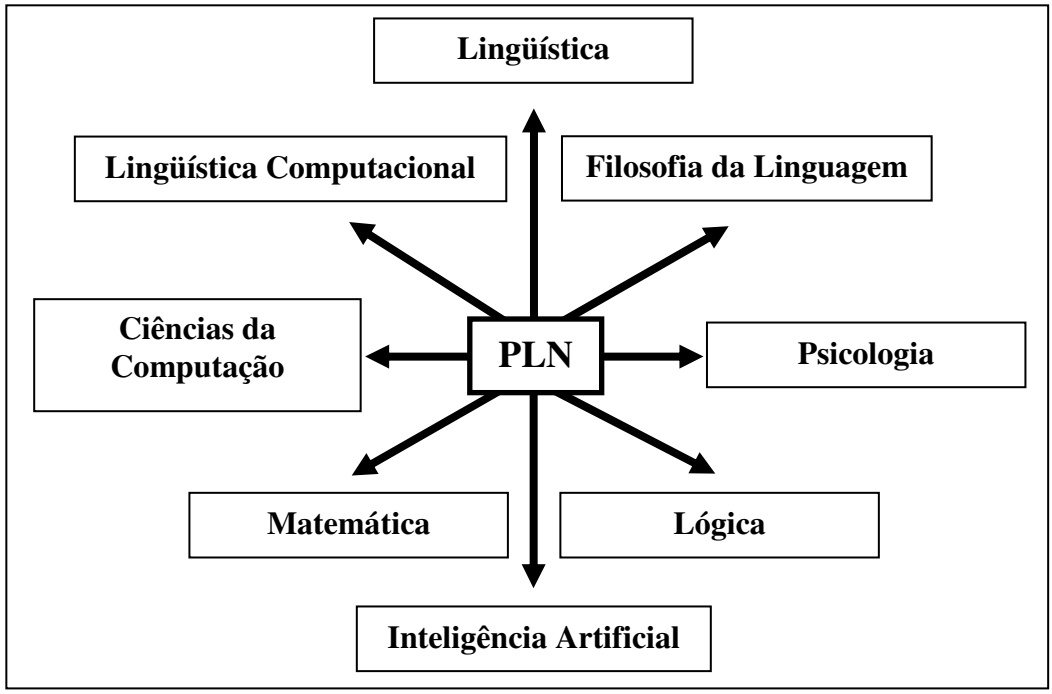


Figura 1: O PLN e suas disciplinas matrizes

As disciplinas descritas na Figura 1 fornecem os recursos teóricos e metodológicos para o PLN. Dentre eles, o Quadro 1 destaca os recursos fornecidos pela Lingüística (DIAS-DA-SILVA, 1996):

Filosofia da Linguagem: *teoria da referência, atos de fala,* etc.;

Psicologia: *teorias* e *modelos de organização da memória*, etc.;

Lógica: *lógica de predicados, lógica proposicional*, etc.

Inteligência Artificial: *representação do conhecimento*, *estratégias de inferências*, etc.;

Matemática: *relações* e *funções, teoria de modelos, teoria de conjuntos*, etc.;

Ciência da Computação: *algoritmos*, *autômatos, redes de transição*, etc.;

Lingüística Computacional: *linguagens de programação*, etc.;

Lingüística: *análise gramatical, análise pragmática, teoria do texto*, etc.

Quadro 1: Recursos teórico-metodológicos fornecidos pelas disciplinas matrizes do PLN

Tais recursos teóricos e metodológicos são essenciais para o avanço dos estudos do processamento automático das línguas naturais. Os recursos fornecidos pela Lingüística, em específico, vêm contribuindo para a construção de sistemas de PLN lingüisticamente motivados e capazes de simular aspectos da competência e do desempenho lingüísticos humanos.

Ressalta-se também que a contribuição dessas disciplinas, sobretudo da Lingüística, não é unidirecional, como pode sugerir a Figura 1. As pesquisas em PLN também têm permitido a proposição de modelos lingüísticos mais completos[1], explícitos[2] e, conseqüentemente, mais operacionais[3] (DIAS-DA-SILVA, 1996).

Um bom exemplo da contribuição do trabalho colaborativo entre lingüistas e pesquisadores do PLN ocorreu, por exemplo, na proposição e no desenvolvimento da

[1] Modelos gramaticais que prevêem, além do desenvolvimento de cada um dos componentes da gramática, modos de inter-relacionar um modelo da competência com um modelo do desempenho.
[2] Modelos gramaticais cujas informações são especificadas em termos de linguagens formais (p. ex: lógica de predicados, estruturas de atributo-valor, entre outras).
[3] Modelos gramaticais que podem ser "interpretados" pelos sistemas de PLN devido à representação formal de suas informações.

Gramática Funcional ("*Functional Grammar*" - FG)[4] de Dik (1997). A implementação da FG permitiu que se realizassem testes e que se comprovasse a necessidade de efetuar modificações no conteúdo lingüístico do modelo, com o objetivo de torná-lo mais completo (SIEWIERSKA, 1991, p.19).

Atualmente, após cinqüenta anos de pesquisas acumuladas, o PLN ainda é um domínio em plena ebulição, no qual sistemas computacionais são projetados, analisados, implementados e mesmo comercializados, haja vista o acelerado crescimento da indústria de informática. A seguir, estão descritos alguns dos tipos de sistemas de PLN (aplicações/produtos) mais desenvolvidos (COLE, 1995; COOPER et al. 1996; MANI, 2000):

- **Sistemas de correção gramatical e ortográfica:** sistemas cujas principais funções são (i) identificar os erros de ortografia e gramática e (ii) sugerir possíveis alternativas a cada erro identificado.
- **Sistemas de tradução:** sistemas que visam à tradução de grande quantidade de textos de modo automático, ou seja, sem a intervenção do tradutor humano.
- **Sistemas de sumarização de textos**: sistemas que visam à produção de sumários (resumos) a partir de textos-fonte.
- **Sistemas de armazenamento e manipulação de unidades lexicais:** sistemas que gerenciam o fluxo de informação lexical dos mais diversos tipos de sistemas de PLN e que podem também dar origem a dicionários e léxicos eletrônicos, p.ex.: um *thesaurus* eletrônico e bases de dados lexicais.
- **Sistemas de processamento de fala (do inglês, "*speech recognition systems*"):** sistemas que transformam textos escritos em textos falados, ou seja, transformam seqüências de *grafemas* em seqüência fônicas, ou vice-versa.

[4] Neste trabalho, optou-se por fazer referência a essa e a outras gramáticas por meio de suas abreviaturas.

Todos esses tipos de sistemas de PLN, diferentes dos demais tipos de sistemas computacionais, são criados para manipular o código lingüístico. Todos eles pressupõem um tipo de "arquivo" em que são armazenadas as unidades lexicais (palavras e expressões) que são manipuladas pelo sistema durante os procedimentos de interpretação e/ou produção de língua natural. Nesse "arquivo", concebido como uma base de dados léxico-gramatical, são especificadas, para cada unidade nele contida, informações de natureza lexical, morfológica, sintática, semântica e, até mesmo, pragmático-discursiva, dependendo das especificidades do sistema de PLN para o qual o arquivo foi desenvolvido (PALMER, 2001; DIAS-DA-SILVA, OLIVEIRA, 2002). Do ponto de vista do PLN, esse mega arquivo é a base de dados lexicais do sistema, isto é, o seu "léxico".

Atualmente, em função das aplicações reais para as quais os sistemas de PLN são escritos, é premente, na construção de um tradutor automático, por exemplo, a compilação de léxicos (monolíngües e/ou multilíngües) que sejam (HANDKE, 1995; VIEGAS, RASKIN, 1998):

(i) *manipuláveis* pelo sistema do qual fazem parte, isto é, léxicos cujas informações sejam explicitamente especificadas por meio de um esquema de representação formal (ou formalismo);

(ii) *lingüisticamente motivados*, tanto do ponto de vista da robustez (isto é, léxicos que contenham uma quantidade de unidades compatível com o léxico de uma língua natural) quanto da qualidade das informações associadas às entradas lexicais.

Dessa forma, a construção de léxicos para fins do PLN requer, sobretudo, a investigação dos diferentes formalismos de representação da informação lexical e das propriedades dos itens lexicais.

1.2. Objetivo

Assim, propõe-se, neste trabalho, a representação lingüístico-computacional para uma das classes lexicais de palavras da língua portuguesa, a saber, a dos *adjetivos*[5]. A razão para a escolha dessa classe de palavras pautou-se, basicamente, em dois fatores.

O primeiro diz respeito ao fato de que as classes dos verbos e dos substantivos têm merecido uma atenção maior, por parte dos investigadores, que classes como a dos adjetivos e advérbios (PUSTEJOVSKY, 1996). Reagindo, assim, à tendência apontada, privilegiou-se a classe dos adjetivos, haja vista que, segundo Vilela (1995), essa é, depois das classes dos verbos e dos substantivos, a terceira classe mais representada na língua portuguesa.

O segundo fator diz respeito ao fato de que os adjetivos ocupam lugar de destaque na exteriorização de uma *visão de mundo*[6], pois parece ser por meio dos adjetivos que se manifesta com bastante clareza a subjetividade do uso lingüístico (BORBA, 1996). Segundo Borba (1996), a tradução da realidade corre por conta da seleção lexical, momento em que o falante ou escolhe uma palavra portadora de conteúdo objetivo/ subjetivo ou dá ao item um peso objetivo/ subjetivo de acordo com o contexto e situação.

Em outras palavras, quando se diz que é por meio do adjetivo que se manifesta a subjetividade do uso lingüístico, faz-se referência às opiniões do locutor (relacionadas à crença, valores, afetividades e registro do que ocorre no mundo exterior) transpostas na língua (FAULSTICH, 1990; BAKHTIN, 1986). Naturalmente, qualquer palavra lexical se destina a esse papel, mas o adjetivo ocupa lugar de destaque.

---

[5] A representação proposta poderá ser empregada, por exemplo, na compilação dos léxicos monolíngües para sistemas de tradução automática.

[6] Entende-se *visão de mundo* como o conjunto de crenças, opiniões, seleção e registro do que ocorre no mundo objetivo (BORBA, 1996).

Dessa forma, os sistemas de PLN não podem ignorar a "interpretação" dos adjetivos nas tarefas de análise e/ou síntese de língua (textos); os sistemas de PLN necessitam "entender" ou, no mínimo, "emular" as propriedades da classe adjetival (RASKIN, NIRENBURG, 1995).

Na verdade, propõe-se a representação lingüístico-computacional dos adjetivos *valenciais*, os quais apresentam *estrutura de argumentos* (ou valência). Ressalta-se que essa estrutura tem sido amplamente empregada para representar parte das informações léxico-gramaticais das unidades do léxico (LEVIN, PINKER, 1991; SAINT-DIZIER, VIEGAS, 1995; PUSTEJOVSKY, 1996).

Na busca de equacionamento para o trabalho, que se enquadra no âmbito do PLN, tomou-se por base Dias-da-Silva (1996; 1998), que fornece os subsídios metodológicos essenciais para o desenvolvimento de projetos multidisciplinares na área do PLN. A metodologia proposta por Dias-da-Silva fundamenta-se nas estratégias desenvolvidas para o campo denominado "engenharia do conhecimento". Nesse campo, a construção de "sistemas de conhecimento" (em inglês, "*knowledge system*")[7] pressupõe a especificação dos tipos de conhecimento que os especialistas possuem e de como esse conhecimento é armazenado, manipulado, aplicado e adquirido (DIAS-DA-SILVA, 1996). Assim, concebendo o estudo do PLN como um tipo particular de "sistema de conhecimento", isto é, como um sistema de conhecimentos de natureza lingüística, Dias-da-Silva propõe uma metodologia, descrita na próxima seção, que distribui as atividades de pesquisa em três domínios ou fases.

Neste ponto do trabalho, cabe salientar também que essa metodologia busca unir reflexões dos vários domínios do conhecimento envolvidos no processamento das línguas naturais (lingüística, psicolingüística, ciência da computação). Essa conciliação é essencial

[7] Sistemas computacionais construídos para representar complexos de conhecimentos e para serem aplicados automaticamente no processo de *resolução de problemas* (AMAREL, 1990).

para que soluções viáveis e consistentes sejam encontradas para os problemas postos pelo processamento automático de processamento das línguas.

Ressalta-se ainda que, como objetivo secundário, propõe-se também a inserção de novas informações sintático-semânticas referentes aos adjetivos na base da rede Wordnet.Br (DIAS-DA-SILVA, OLIVEIRA, MORAES, 2003).

1.3. Metodologia

A proposta metodológica de Dias-da-Silva estabelece que os estudos na área do PLN distribuem-se em três atividades complementares, agrupadas, segundo sua natureza, em três domínios, conforme ilustrado na Figura 2.

Figura 2: Equacionamento metodológico dos problemas do PLN: os três domínios de Dias-da-Silva (1996, 1998)

No **domínio lingüístico**, as atividades concentram-se na reflexão sobre os fatos da língua natural e do seu uso. Desnecessário sinalizar que a realização das atividades pertencentes ao *domínio lingüístico* é condição necessária para a realização das atividades previstas nos outros dois domínios – o *representacional* e o *implementacional*. Neste trabalho, em específico, foram desenvolvidas atividades relacionadas aos domínios *lingüístico* e *representacional*.

Partindo-se do objetivo geral, ou seja, o de propor a representação lingüístico-computacional dos adjetivos valenciais, as atividades do domínio lingüístico concentraram-se

nas seguintes tarefas: (i) sob o ponto de vista da Psicolingüística: investigação teórica sobre como a informação lexical (ou conhecimento lexical) é armazenada, manipulada e representada na mente do falante; (ii) sob o ponto de vista da Lingüística: investigação teórica sobre a categoria adjetival; identificação dos adjetivos valenciais do português e de suas principais propriedades.

Com base na investigação psicolingüística sobre como a informação lexical (ou conhecimento lexical) é armazenada, manipulada e representada na mente do falante, propôs-se um esquema geral de representação da informação lexical no qual estão especificados os tipos de informação (lexical) lingüisticamente relevantes para o processamento automático das línguas naturais.

Da investigação da classe dos adjetivos do português, foi possível identificar os adjetivos valenciais e, principalmente, as suas propriedades sintáticas e semânticas. Tal investigação, em específico, fundamentou-se nas informações fornecidas pela literatura, sendo que, para a exemplificação das propriedades desses adjetivos, foram utilizadas ocorrências retiradas dos córpus do Laboratório de Lexicografia (doravante, LL) da Faculdade de Ciência e Letras – UNESP/Araraquara e do Núcleo Interinstitucional de Lingüística Computacional (NILC) – USP/São Carlos. O córpus do LL abriga textos escritos de literaturas romanesca, técnica, oratória, jornalística e dramática, totalizando aproximadamente 190 milhões de ocorrências em palavras de linguagem escrita desde 1950. O córpus do NILC[8] reúne textos escritos de gêneros literário, jurídico, técnico e científico, jornalístico, didático, entre outros, totalizando aproximadamente 35 milhões de ocorrências[9].

[8] O córpus do NILC está subdividido em três diretórios: *córpus corrigido*, *córpus semicorrigido* e *córpus não-corrigido*. Os usos dos adjetivos foram extraídos apenas do *córpus corrigido* (aproximadamente 30 milhões de ocorrências). Assim, neste trabalho, quando utilizado o termo "córpus do NILC", entende-se *corpus corrigido*. <http://acdc.linguateca.pt/acesso/>.

[9] A todo uso extraído do córpus do LL é associada uma sigla que indica a referência do texto do qual o uso foi extraído, p.ex.: AF (*A festa*. ANGELO, I. 1978) (BORBA, 2002). A todo uso extraído do córpus do NILC, por sua vez, é associado um conjunto de informações que especifica o gênero e o tipo de texto. Para mais informações sobre a notação empregada como referência do córpus do NILC, conferir Pinheiro e

No **domínio representacional**, estudam-se modelos formais de representação para os conhecimentos reunidos no domínio lingüístico que sejam computacionalmente tratáveis. No caso específico deste trabalho, foram investigados os seguintes modelos de representação: *estrutura de traços* ou *matrizes de atributo-valor, lógica de predicados* e *frames*. Foi exatamente com base na investigação desses formalismos que se propôs a representação lingüístico-computacional dos adjetivos valenciais.

1.4. Estrutura da dissertação

Neste **primeiro** capítulo, apresentou-se a contextualização do trabalho, mais especificamente, o campo de pesquisa em que a investigação se insere, seu objetivo e seu equacionamento metodológico.

No **segundo** capítulo, investiga-se a concepção de "léxico" sob o ponto de vista computacional e (psico)lingüístico e apresenta-se um esquema geral de representação da informação lexical elaborado com base nessa investigação.

No **terceiro** capítulo, faz-se, ainda que de modo resumido, uma apresentação da classe dos adjetivos do português, enfatizando suas subclasses e funções. Por meio dessa revisão, identificam-se os adjetivos valenciais e suas principais propriedades sintáticas e semânticas.

No **quarto** capítulo, discutem-se alguns dos principais modelos formais para a representação de informações lexicais e, a partir deles, propõe-se a representação das propriedades dos adjetivos valenciais identificadas no capítulo anterior.

No **quinto** capítulo, apresenta-se a proposta de extensão da base da rede Wordnet.Br para que informações sintático-semânticas dos adjetivos valenciais possam ser inseridas.

Por fim, no **sexto** capítulo, são feitas as considerações finais sobre esta dissertação.

Aluisio (2003). No Anexo 1, estão descritas as siglas empregadas neste trabalho e suas respectivas definições.

Capítulo 2 Dos olhares sobre o "léxico" e do esquema de representação da informação lexical

Como mencionado, o como objetivo deste trabalho é propor a representação lingüístico-computacional dos adjetivos valenciais. Para tanto, apresenta-se, em primeiro lugar, o esquema geral de representação das informações lexicais que serviu de base para a representação proposta para os adjetivos valenciais do português.

Dessa forma, dividiu-se esta seção em duas partes. Na primeira, são apresentadas algumas reflexões sobre o "léxico" do ponto de vista do PLN, da Lingüística e da Psicolingüística. Com essas reflexões, foi possível (i) delimitar o objeto denominado "léxico" no domínio do PLN, (ii) identificar as características e funções desse objeto e, por fim, (iii) delimitar os pressupostos (psico)lingüísticos para a proposição do esquema formal de representação da informação lexical. Esse esquema, inclusive, é apresentado na segunda parte desta seção.

2.1. Do olhar lingüístico-computacional sobre o léxico

Teoricamente, as arquiteturas propostas para sistemas de PLN acabam por espelhar a arquitetura proposta para o sistema lingüístico (ALLEN, 1987; FRAZIER, 1989). Como decorrência, um sistema de PLN deve possuir módulos autômatos, que realizam tarefas específicas e especializadas, e módulos que armazenam um modelo de conhecimento proposicional, que visa a criar simulacros de parcelas de mundo que lhe servem de referencial para interpretar os enunciados lingüísticos. Apesar da arquitetura de um sistema de PLN variar de acordo com as especificidades da aplicação, dois grupos de componentes são imprescindíveis para a implementação de qualquer sistema desse tipo: as *bases de*

conhecimento e os *módulos de processamento* que atuam sobre essas bases (DIAS-DA-SILVA, 1996). A Figura 3 ilustra esses dois grupos de componentes.

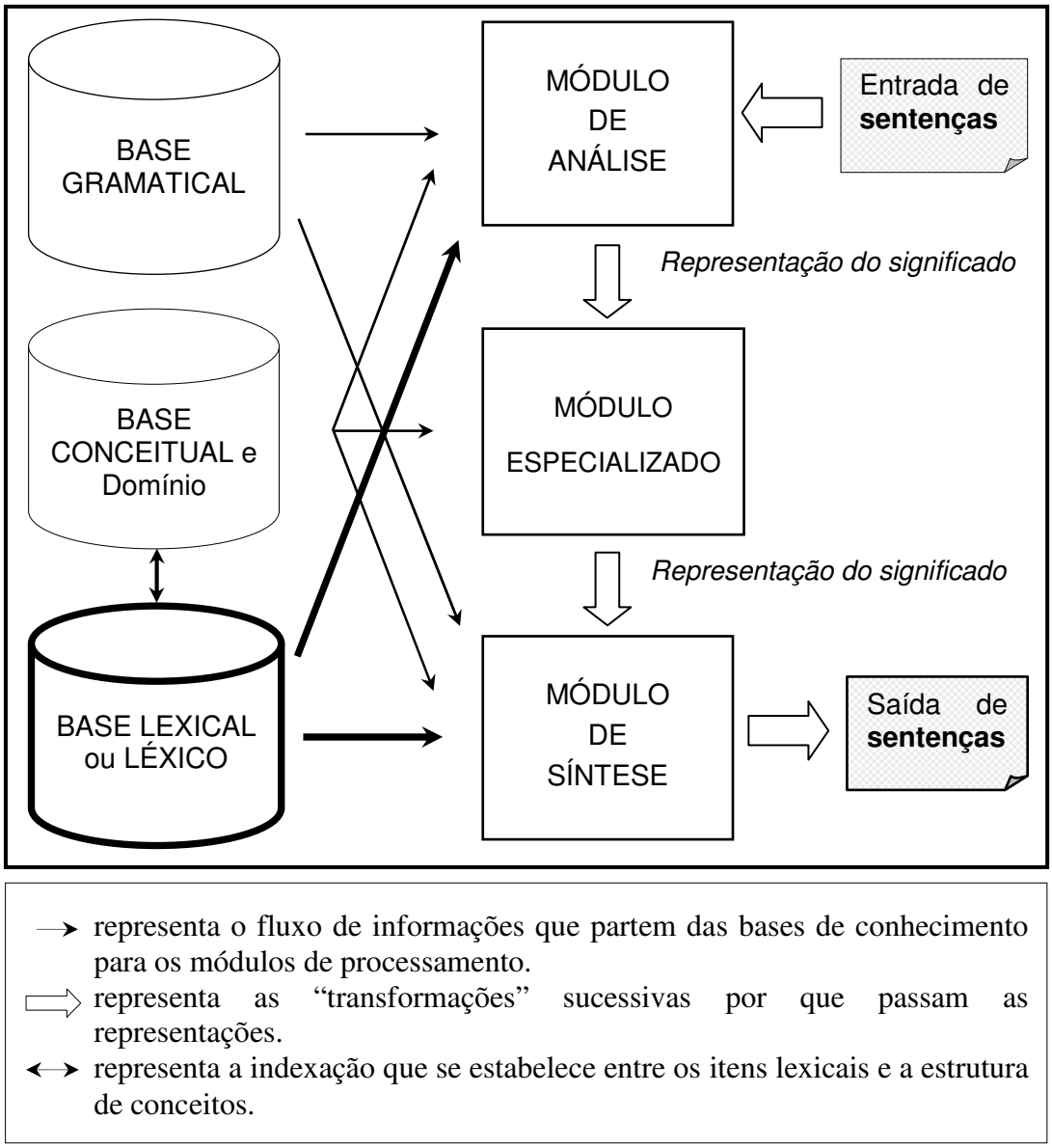


Figura 3: Arquitetura de um sistema de PLN genérico de Dias-da-Silva (1996)

Os módulos de conhecimento podem ser divididos em três módulos: o de *análise,* o *especializado* e o de *síntese*. As bases de conhecimento podem ser dividas em três bases: *gramatical, conceitual* e *lexical*. Com exceção do *módulo especializado*, os demais módulos de processamento e as bases de conhecimentos, embora os conteúdos possam variar em

função da especificidade do sistema, possuem estrutura e funcionamento semelhantes. Toda a especificação dos módulos descrita a seguir foi extraída de Dias-da-Silva (1996).

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

O **módulo de síntese** (MS), por sua vez, transforma a representação abstrata gerada pelo MA em uma seqüência de "frases contextualizadas". Ao realizar a tarefa de construção de uma representação semântica, por exemplo, o MA utiliza-se, dependendo da sofisticação do sistema de que é parte, das bases gramatical, conceitual e lexical para executar todas ou parte das análises: morfológica, sintática, semântica e, até mesmo, pragmática. Assim, cada base de conhecimento, por sua vez, fornece ao MA informações de natureza diferente (cf. também HUTCHINS, SOMERS, 1997).

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

A **base conceitual** fornece um modelo do mundo físico e conceitual, descrevendo tipos básicos de objetos, eventos, propriedades, relações e atributos em termos de representações hierarquicamente estruturadas, isto é, a sua estrutura consiste em uma rede de unidades conceituais interligadas em termos de relações de hiponímia/ hiperonímia, entre outras. Essa base também pode fornecer conceitos mais específicos, ou seja, conceitos referentes a domínios particulares do conhecimento ou conceitos relacionadas a tarefas específicas para a qual o módulo esteja sendo projetado.

Em particular, à **base lexical**, fica a tarefa de fornecer, aos MA e MS, a coleção de unidades lexicais, para as quais se faz necessária a especificação de conjuntos de traços morfológicos, sintáticos, semânticos e pragmático-discursivos (cf. também BOGURAEV, BRISCOE; 1989; BRISCOE, 1991; SANFILIPPO, 1995; PALMER, 2001). Esse tipo de base de dados, no domínio do PLN, é definido como o léxico do sistema e recebe a denominação de **léxico tratável por máquina** ("*machine tractable dictionary*") (WILKS, 1988).

Todos os sistemas de PLN desenvolvidos para serem aplicações reais necessitam de léxicos que sejam *lingüisticamente motivados*, tanto do ponto de vista da (i) robustez (isto é, léxicos que contenham uma quantidade de unidades compatível com o léxico de uma língua natural) quanto da (ii) pertinência ("qualidade") das informações associadas às entradas. Isso acontece porque o desempenho de um sistema de PLN depende diretamente do número de entradas do léxico e da qualidade das informações associadas a essas entradas (DORR, 1993; SAINT-DIZIER, VIEGAS, 1995; PALMER, 2001). Mas quais são, por exemplo, os itens lexicais de uma língua que devem estar em um léxico "computacional"? Ou ainda, quais os tipos de informação lexical lingüisticamente relevantes para o processamento automático das línguas naturais (HANDKE, 1995; PALMER, 2001)? Para tentar responder a essas e a outras questões, pesquisadores do PLN têm buscado, nos domínios lingüístico e psicolingüístico, os subsídios para a construção de léxicos lingüístico-computacionais.

No caso específico deste trabalho, é importante identificar quais são os tipos de informações lexicais lingüisticamente relevantes para o processamento das línguas naturais. Para tanto, passa-se a investigar a natureza do léxico sob o ponto de vista dos estudos (psico)lingüísticos. Nesse enfoque, são investigadas as seguintes questões: (i) a concepção de léxico do ponto de vista da Lingüística, focalizando a *macroestrutura* do léxico; (ii) a concepção de *léxico mental*, ou seja, o papel do léxico no processamento cognitivo da linguagem; (iii) o acesso (entendido aqui como sinônimo de "fazer uso do léxico" ou

"preparar o léxico para ser usado") às unidades lexicais na mente do falante; (iv) a organização da microestrutura do léxico mental.

Ressalta-se que a opção por investigar tais questões sob o ponto de vista psicolingüístico justifica-se pelo fato de que os estudos realizados nesse domínio têm contribuído consideravelmente para a construção de léxicos lingüístico-computacionais. A rede **WordNet**[10], inicialmente proposta para a língua inglesa, pode ser vista como um exemplo paradigmático de como os estudos psicolingüísticos têm contribuído com o desenvolvimento de léxicos lingüístico-computacionais. Essa rede consiste em uma base relacional de unidades lexicais (palavras e expressões) da língua inglesa organizadas à semelhança de um *thesaurus* (FELLBAUM, 1999). O que, de fato, diferencia a rede WordNet de um *thesaurus* "tradicional" é a sua organização, baseada em propriedades vinculadas ao *léxico mental* (cf. MILLER; FELLBAUM, 1991; MILLER et al. 1990, 1993).

2.2. Do léxico nos estudos lingüísticos: a *macroestrutura*

Na teoria lingüística, o conceito de *léxico mental* surge com a redefinição da natureza do léxico. De acordo com Basílio (1999), em abordagens gerativas, o léxico deixa de ser o vocabulário da língua como realidade externa; o objeto de estudo do lingüista é o *léxico mental*. Essa redefinição traz diversas implicações para a teoria lingüística, conforme relatado por Basílio (1999). Resumindo os aspectos mais relevantes, a autora destaca algumas questões relativas ao estudo do léxico em abordagens gerativas, dentre elas, estão: (i) a delimitação das fronteiras entre o conhecimento lingüístico e não lingüístico no e do léxico; (ii) a interação

[10] O nome dessa base lexical está grafado com "N" maiúsculo para diferenciá-la das demais, caracterizando-a como "a mãe de todas as wordnets". A elaboração da rede WordNet foi idealizada (e realizada) na Universidade de Princeton, EUA, e engloba, naturalmente, apenas a língua inglesa. Atualmente, várias comunidades de PLN já possuem seus aplicativos no formato *Wordnet*. Dentre esses aplicativos, citam-se, por exemplo, as redes para o espanhol, italiano, francês, tcheco, entre outras, e, ainda em fase de construção, as redes Wordnet-PT e a Wordnet.Br.

entre o léxico e os diferentes componentes da gramática e (iii) a pertinência ou não de objetos ao léxico.

Ressalta-se que não há um consenso em Lingüística sobre a natureza do léxico mental e sobre a sua função dentro do sistema lingüístico. Mas existem, no entanto, algumas hipóteses a respeito, em específico, da organização global do léxico (ou macroestrutura).

De acordo com Anderson (1992), especificamente, o léxico não é entendido somente como uma base de dados, cuja única função é a de armazenamento dos itens lexicais. O autor se coloca a favor da idéia de que o léxico é um componente do conhecimento lingüístico e que, como tal, deve ser entendido como "o conhecimento que um falante tem sobre as palavras da língua" (ANDERSON, 1992). Partindo dessas premissas, o autor explica que, ao se dizer que uma determinada palavra está dentro do léxico, diz-se que ela é reconhecida por esse componente da gramática, e não necessariamente que está dentro de uma lista finita de itens. Segundo Perini (1999), "o léxico deve ser entendido em termos de redes de correspondência em vários níveis e conceitua o item léxico como uma trilha de propriedades fonológicas, morfológicas sintáticas e semânticas" (PERINI, 1999). Percebe-se que Perini também considera que o léxico pressupõe uma organização diferente de uma lista e um tipo de conhecimento peculiar. Dado o número elevado dos elementos do léxico mental e da complexidade combinatória resultante desse número, pressupõe-se que os itens estejam organizados de maneira funcional, para que o falante possa recuperar rapidamente não só o significado de um item, mas também todas as suas características gramaticais e usos. Em outras palavras, isso quer dizer que o léxico mental, do ponto de vista da macroestrutura, apresenta uma intrincada rede de relações que se estabelecem entre seus constituintes (MEL'ČUK, 1988). Essas relações, consideradas *intrínsecas*, são responsáveis pela macroestrutura do léxico e podem ser **diretas,** ou seja, representadas no interior das entradas

lexicais (LEVELT, 1993)[11]. Segundo Biderman (1981) e Lyons (1981), pode-se postular que as associações estabelecidas entre os itens lexicais são de duas formas: (i) paradigmáticas; (ii) sintagmáticas.

(1) Relações paradigmáticas

Diz-se que as unidades lexicais pertencem ao mesmo paradigma quando uma puder ser substituída pela outra em um mesmo ponto do sintagma (BORBA, 1991). Tais relações podem ser de natureza:

(a) morfossemântica: isto é, uma família de itens lexicais que tem uma mesma raiz, p.ex.: embalar, embalado, embalador (JACKENDOFF, 1975; BASÍLIO, 1987, 1994);

(b) lógico-conceitual: diz-se lógico-conceitual porque esse tipo de relação se estabelece entre conceitos (MURPHY, 2002); dentre elas, citam-se as relações de *hiponímia* (p.ex.: laranjeira é hipônimo de árvore) e *hiperonímia* (p.ex.: árvore é hiperônimo de laranjeira);

(c) léxico-semântica: diz-se léxico-semântica (também denominada de *relação de sentido*) porque esse tipo de relação se estabelece entre unidades lexicais e não entre conceitos (GROSS, FISCHER, MILLER, 1989); são elas: as relações de *sinonímia* (p.ex.: asfixiado, sufocado) e de *antonímia* (p.ex.: grande – pequeno)[12].

(2) Relações sintagmáticas

Nas relações sintagmáticas, as unidades de associam em seqüências porque são compatíveis. Em outras palavras, pode-se dizer que tais relações são resultantes da combinatória freqüente entre itens lexicais. A principal relação sintagmática de macro-estrutura lexical é a **colocação** (inglês: *collocation*). O termo colocação pode ser usado para se

[11] Tais relações não foram previstas no esquema de representação lexical e não foram incluídas na representação dos adjetivos valenciais.

[12] No caso específico dos adjetivos, mais especificamente dos adjetivos *qualificadores*, ressalta-se que a relação paradigmática léxico-semântica típica dessa classe é a *antonímia* (GROSS, FISCHER, MILLER, 1989).

referir a seqüências de unidades lexicais que coocorrem habitualmente como, por exemplo, *feliz aniversário, dias de sol* (CRUSE, 1986; STUBBS, 2001).

Após essas breves considerações a respeito da concepção de léxico (mental) do ponto de vista dos estudos lingüísticos, são feitas, a seguir, alguns observações, do ponto de vista da Psicolingüística, sobre: (i) a concepção de léxico no processamento cognitivo da linguagem; (ii) o acesso às unidades lexicais na mente do falante; (iii) a organização da microestrutura do léxico.

2.3. Do léxico sob o olhar da Psicolingüística

Unindo pressupostos da Lingüística e da Psicologia, a Psicolingüística estuda a existência e o funcionamento de mecanismos mentais envolvidos no processamento da linguagem humana (SAINT-DIZIER, VIEGAS, 1995).

Com o intuito de compreender, entre outras questões, como ocorre o armazenamento de o acesso aos itens lexicais de uma determinada língua, os psicolingüistas postulam a existência de um **léxico mental** (LM), definido como a parte do conhecimento lexical do indivíduo delimitada por sua língua (BIERWISCH, SCHREUDER, 1992; LEVELT, 1992). De acordo com Bock (1982), Bierwisch e Schreuder (1992) e Levelt (1992), o léxico mental ocupa lugar de centralidade no processamento cognitivo da linguagem, o qual envolve três tipos de processos: (i) **conceitualização** (especificação de conceitos); (ii) **formulação** (seleção de palavras e construção de representações sintáticas e fonéticas); (iii) **articulação** (produção da fala).

Esses três processos – conceitualização, formulação e articulação – e o papel desempenhado pelo LM estão ilustrados na Figura 4.

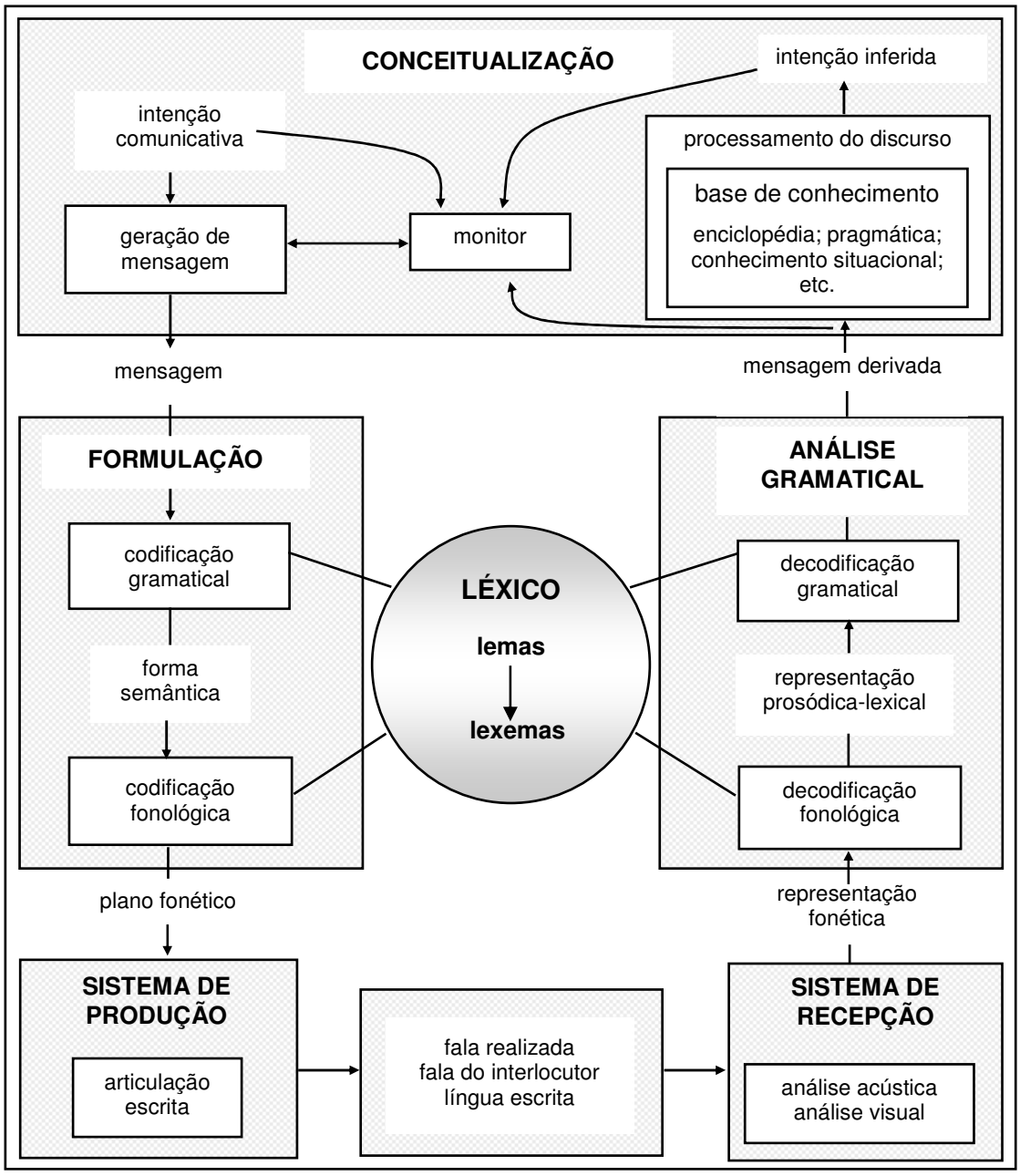


Figura 4: Arquitetura do processamento mental da linguagem de Handke (1995) e Lowie (1998).

Para explicar o funcionamento desses três processos e o papel central desempenhado pelo LM no processamento da linguagem, é descrito, com mais detalhes, o sistema ou processo de **produção** de enunciados (lado esquerdo da Figura 4).

Nesse processo, a nomeação de um objeto perceptível envolve: (i) a identificação do objeto (conceitualização); (ii) a seleção de uma representação sintático-semântica do objeto,

assim como a codificação dessa representação em termos fonológicos (formulação); (iii) a transformação da representação fonológica em realização fonética, que constitui o nome do objeto (articulação).

Mais especificamente, o processo de identificação do objeto ou **conceitualização** ativa uma robusta *base de conhecimento* que contém informações extralingüísticas provenientes de diversas fontes (visual, auditiva, motora, emotiva, conceitual, entre outras), além de princípios gerais de organização conceitual (ontologia do senso comum, conceitualizações do espaço e tempo, condições gerais subjacentes ao conhecimento enciclopédico ou a sistemas de crença, etc). O processo de conceitualização gera uma *estrutura conceitual* (EC) (pré-lingüística), que é a mensagem a ser verbalizada e organizada gramaticalmente pela *formulação*, no caso, essa mensagem será o *nome* do objeto.

Já a **formulação** é responsável por transformar essa EC em um enunciado lingüístico. Essa transformação é mediada pelo LM, que, como já se disse, é a parte do conhecimento lexical delimitada pela língua do falante. Assim, a língua do indivíduo delimita o LM, que, por sua vez, media a transformação da *estrutura conceitual* em um *enunciado lingüístico*. Em outras palavras, pode-se dizer que os estímulos recebidos por um indivíduo (= estrutura conceitual) são traduzidos em itens lexicais de acordo com regras e princípios de cada língua. A essa hipótese, Glanzer e Clark (*apud* BIDERMAN, 1981) deram a denominação de **elo verbal** (do inglês, "*verbal-loop hypothesis*").

De acordo com Bierwisch e Schreuder (1992), a conversão ou transformação de uma estrutura conceitual em um enunciado lingüístico é feita em dois estágios: *ativação de lemas* e *ativação de lexemas*. Postula-se, assim, a existência do LM no nível lingüístico. O LM é central a todo o processamento da linguagem e contém todas as informações sobre os itens

lexicais da língua, isto é, seus *lemas* e *lexemas*[13]. Dessa forma, o primeiro estágio da formulação é responsável pela seleção da representação sintático-semântica do objeto. Para tanto, é ativado, no LM, o **lema** do objeto, ao qual estão associadas informações sintáticas e semânticas que determinam, por exemplo, sua forma semântica, categoria sintática e estrutura de argumentos. De acordo com Bierwisch e Schreuder (1992), o resultado do primeiro estágio, a *formulação*, é uma *forma semântica* (FS)[14]. No segundo estágio, essa FS é transformada na *forma fonológica* (FF). Para que essa transformação seja possível, é ativado, no LM, o **lexema** do objeto, ao qual estão associadas informações fonológicas e morfológicas.

Por fim, a **articulação** opera sobre a FF, ativando programas articulatórios que produzem a realização fonética do objeto.

Logo, partindo-se do modelo de Bierwisch e Schreuder (1992), a **seleção do lema** consiste na ativação e seleção de um *lema* a partir da estrutura conceitual (EC). Já a **codificação da forma** consiste na construção de um programa articulatório que envolve a seleção de morfemas e de segmentos da forma da palavra e na ligação desses elementos às suas respectivas posições em uma estrutura denominada *esqueleto da forma da palavra* (Esq). Por exemplo, o Esq da estrutura silábica de *gato*, [σ' σ], é preenchida pelos segmentos [/g/ /a/$_{\mu}]_{\sigma'}$ e [/t/ /u/$_{\mu}]_{\sigma.}$[15]

[13] Cabe ressaltar que os termos *lema* e *lexema* não estão sendo empregados no sentido típico do campo da Lexicografia, isto é, uma representação canônica das entradas de um dicionário (lema) ou de uma unidade lexical virtual que compõe o léxico (lexema) (BIDERMAN, 1999). Para Bierwisch e Schreuder (1992), *lema* é a representação das propriedades sintático-semânticas de um item lexical e *lexema* é a representação das estruturas morfológica e fonológica de um item lexical.

[14] Vale ressaltar que há divergências quanto à postulação dos níveis EC e FS. A Bierwisch e Schreuder (1992), que defendem essa proposta, opõe-se, por exemplo, Jackendoff (1991, 1997), que propõe um nível único denominado nível da *estrutura léxico-conceitual* (ELC). Para Jackendoff, a FS não é concebida como uma entidade distinta da *estrutura conceitual*, mas parte dela.

[15] Os símbolos **σ** (sigma) e **μ** (mi) são usados para representar, em fonologia métrica, a estrutura de uma palavra em termos dos elementos abstratos: **sílaba** (**σ**) e **mora** (**μ**). O número de moras de uma sílaba caracteriza-a como *pesada* ou *leve*, p.ex.: sílabas que contêm vogal longa ou consoante final apresentam duas moras e são consideradas sílabas *pesadas*; sílabas que contêm vogais curtas (como as de *gato*) apresentam uma mora e são consideradas sílabas *leves* (LEVELT, 1993).

Tendo em vista que o acesso aos itens lexicais realiza-se nas etapas de *seleção do lema* e *codificação da forma* da palavra, a representação da microestrutura do LM, ou seja, da estrutura interna das entradas lexicais, subdivide-se em *unidade de acesso* e *especificação lexical*. A Figura 5 ilustra essa bipartição das entradas.

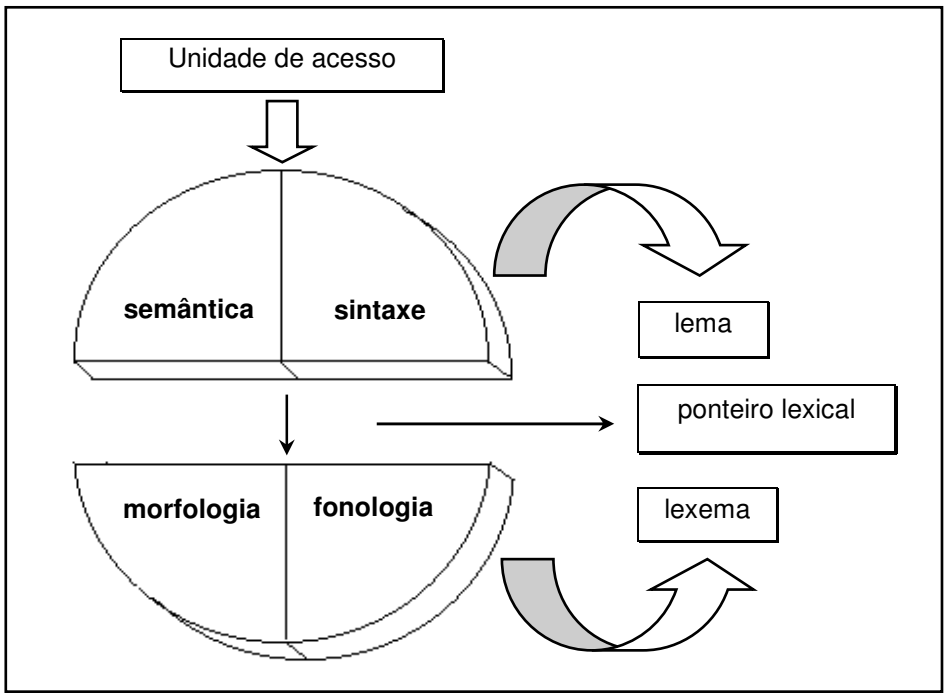


Figura 5: Modelo de entrada lexical bipartida de Handke (1995) e Lowie (1998).

(1) Da unidade de acesso

A **unidade de acesso** é o item lexical propriamente dito. Como bem salienta Langacker (1972) e Basílio (1999), a questão da delimitação das unidades que devem ser consideradas como pertencentes ao léxico é antiga e tem sido discutida sob diferentes perspectivas.

Com relação à forma dessas unidades, os especialistas divergem quanto a se a representação lexical delas se faz por *palavras* ou por *morfemas* (radicais ou raízes), seja na Lingüística, na Psicolingüística ou no PLN. De um modo geral, pode-se dizer que há três concepções de unidade ou item lexical. Na primeira, listam-se, no léxico, as formas que

servem de base para a formação de outras formas, isto é, as raízes e radicais, os afixos e as palavras funcionais; a esse tipo de léxico é dado o nome de "*root lexicon*" ("léxico de raízes" ou "léxico de morfemas"). Na segunda, são listadas todas as formas possíveis da língua, inclusive as flexionadas; a esse tipo de léxico é dado o nome de "*full-form lexicon*" ("léxico de formas plenas", em oposição, portanto, à "léxico de morfemas") (BUTTERWORTH, 1983). Na terceira concepção, que é lingüístico-computacional, listam-se a raiz e outras formas-base idiossincrásicas, isto é, que são empregadas em processos morfológicos não regulares (HANDKE, 1995). Por exemplo, na entrada do verbo *agir*, seriam listadas as formas AG-, base para a maioria das formas do verbo *agir*, e AJ-, para as formas seguidas das letras *a* e *o* (ajo, aja, ajas, ajamos, ajais, ajam).

Além dessas três concepções, uma quarta pode ainda ser identificada. Nessa concepção as formas pertencentes ao paradigma flexional são marcadas como realizações discursivas (p.ex.: formas *embalar, embalou, embalando*) de um *item lexical canônico* (EMBALAR). Por outro lado, as formas pertencentes ao paradigma derivacional (p.ex.: *embalar, embaladeira, embalado*) são marcados como itens lexicais distintos e, conseqüentemente, com entradas lexicais também distintas (LYONS, 1979). Nessa concepção, observa-se que o termo *item lexical* refere-se ao sistema, isto é, à língua, entidade abstrata e supra-individual (*langue*), opondo-se, portanto, à *palavra*, o discurso realizado (*parole*). Vale ressaltar que há outras denominações que são comumente aplicadas à unidade da *langue* e à unidade da *parole*. Por exemplo, Muller (1964) emprega a denominação *vocable* para a unidade do léxico e *mot* para a unidade ocorrente no texto; Biderman (1999), por sua vez, emprega o termo *lexema* para designar a unidade virtual do léxico e o termo *lexia* para designar as realizações discursivas dos lexemas. Neste trabalho, optou-se por empregar o termo *item lexical* para designar a unidade do léxico. Cabe ressaltar ainda que, nessa concepção, os itens pertencentes às classes lexicais são comumente expressos por meio de

suas *formas canônicas*. No caso do adjetivo, a forma canônica equivale ao masculino singular (p.ex.: alto, bonito, cansado, embalado, etc.).

Ainda sobre a questão da delimitação das unidades pertencentes ao léxico, ressalta-se que, neste trabalho, foram tomados como base os pressupostos do projeto WordNet (FELLBAUM, 1999). De acordo com esses pressupostos, é possível estabelecer que *embalado* (=ninado), em (1) O bebê dormiu embalado por um barulhinho bom, *embalado* (=embrulhado), em (2) Pacote embalado a vácuo, e *embalado* (= animado), em (3) O Real Madrid é o time mais embalado do campeonato, são itens lexicais distintos, isto é, *embalado*1, *embalado*2 *e* *embalado*3.

O critério empregado para a identificação de itens distintos é o da substituição por sinônimos. De fato, *embalado*1, *embalado*2 e *embalado*3 podem ser substituídos, sem que haja alteração substancial do significado das sentenças em que ocorrem acima, por seus respectivos sinônimos: p.ex.: *ninado, embrulhado* e *animado*, respectivamente.

(2) Da especificação lexical

A **especificação lexical** de um item é a representação do seu *lema* e do seu *lexema*, os quais estão interligados por um ponteiro lexical (isto é, cada *lema* "aponta" para um *lexema* correspondente) (cf. Figura 5). Como já se disse, o **lema** é a representação das propriedades semânticas e sintáticas do item lexical; especifica as condições conceituais que garantem o uso apropriado do item, indicando, entre outras coisas, sua classe gramatical e seus argumentos; o **lexema**, por sua vez, é a representação das estruturas morfológica e fonológica de um item lexical.

Após a apresentação do léxico sob o ponto de vista das pesquisas psicolingüísticas, fica evidente a centralidade do componente lexical no sistema de processamento cognitivo da

linguagem (BOCK, 1982; LEVELT, 1992; 1993; BIERWISCH, SCHREUDER, 1992; LOWIE, 1998). A seguir, propõe-se o esquema (psico)lingüístico de representação da informação lexical

2.4. Da sistematização da informação lexical: *esquema geral de representação*

Partindo-se, então, da investigação sobre as questões relativas ao léxico sob o ponto de vista (psico)lingüístico, elabora-se, a seguir, um esquema de representação em que estão especificados os tipos de informação lexical lingüisticamente relevantes para o processamento das línguas naturais, o qual servirá base para a proposição da representação dos adjetivos valenciais. Pode-se, então, conceber uma entrada lexical canônica como uma estrutura de dados, no sentido computacional desse termo, contendo quatro *componentes* lingüísticos, os quais estão especificados no Quadro 2.

		COMPONENTES
DIMENSÕES	**Dimensão Lexemática**	**FG**
	Dimensão Lemática	**TG**
		EA
		FS

Quadro 2: Modelo de entrada canônica decomposta em quatro componentes

Cada componente do Quadro 2 especifica tipos distintos de informação lexical, responsáveis pela microestrutura do léxico lingüístico-computacional. Tais tipos estão sintetizados no Quadro 3.

Dimensão Lexemática	**FG** (E) (Leia-se: *Forma gráfica* de E) Esse componente especifica a forma gráfica e a estrutura morfológica de E.
Dimensão Lemática	**TG** (E) (Leia-se: *Traços gramaticais* de E) Esse componente determina as propriedades sintáticas de E e dos constituintes de hierarquia superior dos quais E é núcleo.
	EA (E) (Leia-se: *Estrutura de argumentos* de E) Esse componente especifica a seqüência de uma ou mais posições argumentais que corresponde aos argumentos exigidas por E.
	FS (E) (Leia-se: *Forma semântica* de E) Esse componente especifica o conteúdo proposicional da expressão contendo E, restringindo-a.

Quadro 3: Os componentes da entrada lexical e as informações léxico-gramaticais correspondentes

A seguir, são feitas algumas observações teóricas e práticas sobre os quatro componentes do esquema de representação lexical do Quadro 3.

2.4.1. Do aprofundamento teórico do modelo e de sua utilização neste trabalho

2.4.1.1. Do componente FG

Como mencionado, o componente FG especifica a forma *gráfica* (ortográfica) e a estrutura *morfológica* das entradas lexicais. As características fonético-fonológicas das entradas, que também podem ser especificadas nesse componente, não foram previstas no esquema porque este trabalho concentra-se exclusivamente no tratamento computacional da modalidade escrita[16].

[16] O destaque dado à modalidade escrita pauta-se no fato de que o tratamento dos aspectos ligados à fonética, fonologia e prosódia, essencial para o desenvolvimento de sistemas de fala, necessita de investigação à parte (DIAS-DA-SILVA, 1996; HUTCHINS, SOMERS, 1997).

Do componente FG, a representação dos adjetivos valenciais contemplou apenas a forma *gráfica*, isto é, a forma ortográfica dos itens lexicais: a seqüência de *grafemas* que constitui um item lexical[17]. Os *grafemas* são, na verdade, os símbolos gráficos unos, constituídos por traços gráficos distintivos, que permitem ao falante perceber visualmente os itens da língua, da mesma forma que os *fonemas*[18] permitem ao falante perceber esses itens auditivamente na língua oral. Na língua portuguesa, há símbolos gráficos (p.ex.: <c>, <s>) que podem representar em certos contextos um mesmo fonema (p.ex.: os símbolos gráficos <c>, <s> podem representar o mesmo fonema /s/), mas como grafemas podem distinguir, na língua escrita, os homófonos da língua oral (p.ex.: *cela* = tipo de aposento/ *sela* = arreio de cavalgadura).

2.4.1.2. Do componente TG

No componente TG da entrada de um item lexical, especificam-se as propriedades sintáticas desse item. Presume-se que as propriedades sintáticas realizam-se por meio das categorias gramaticais primárias e das categorias gramaticais secundárias[19]. A especificação da categoria primária do item engloba, além da classe, a descrição dos traços subcategoriais (isto é, esquema de subcategorização).

Do ponto de vista do PLN, a especificação de informações sobre as categorias primárias e secundárias dos itens lexicais é essencial para que os analisadores morfológico (ou

[17] A especificação da estrutura morfológica dos itens lexicais, no entanto, não foi projetada na representação dos adjetivos valenciais.

[18] Por se tratar de um trabalho multidisciplinar, julga-se importante e necessário sinalizar que o "*Fonema* é a menor unidade destituída de sentido, passível de delimitação na cadeia da fala. Cada língua apresenta, em seu código, um número limitado e restrito de fonemas [...] que se combinam sucessivamente, ao longo da cadeia da fala, para constituir os significantes das mensagens, e se opõem, segmentalmente, em diferentes pontos da cadeia da fala, para distinguir as mensagens umas das outras. Sendo esta sua função essencial, o fonema é seguidamente definido como a unidade distintiva mínima" (DUBOIS, 1973, p. 280).

[19] Por categoria gramatical primária, entende-se a classe gramatical a que o item pertence, por exemplo, s*ubstantivo, verbo* e *adjetivo*. Por categorias gramaticais secundárias, entende-se as categorias de *gênero, número, modo, caso, aspecto, voz*, entre outras (LYONS, 1979).

léxico) e sintático possam atribuir categorias gramaticais a uma forma ***x*** e verificar a validade do relacionamento sintático da mesma com as demais formas da sentença, construindo, assim, uma estrutura abstrata da sentença que contenha a forma ***x***.

2.4.1.3. Do componente EA

A todo **predicador** (PR) está associada uma **estrutura de argumentos** (estrutura-a) ou *valência*[20]. Os predicadores são itens lexicais semanticamente incompletos que, por isso, precisam necessariamente ligar-se a outros elementos – seus argumentos (As) – para adquirir um valor semântico completo (DIK, 1997; NEVES, 2000).

As relações semânticas estabelecidas entre um predicador e seus argumentos são representadas por *papéis temáticos* (ou funções semânticas) (GRUBER, 1965; FILLMORE, 1968; PALMER, 1994; DAVIS, 2001).

Há dez anos, o construto denominado de estrutura-a foi equacionado como o número de argumentos (A) requerido ou projetado por um PR (GRIMSHAW, 1992). De acordo com autores como Williams (1981) e Marantz (1984), a estrutura-a é um conjunto de As marcados como **internos** (subcategorizados) ou **externos** (não subcategorizados)[21]. Entretanto, com a crescente importância de princípios como o Princípio da Projeção[22] e Critério-θ[23] na Teoria da Regência e Ligação ("*Government-Binding Theory*" – GB; cf. CHOMSKY, 1981) e com o

[20] Por predicador, entende-se todo elemento que atribui uma determinada propriedade a um certo termo ou estabelece uma relação entre termos, ou seja, uma *predicação* (MIRA MATEUS, et al., 1994; NEVES, 1997).
[21] Williams (1981) e Marantz (1984) definem os *papéis temáticos* de acordo com a concepção clássica de Gruber (1965) e Fillmore (1968), ou seja, como um conjunto de rótulos conceituais que definem a participação dos argumentos na "cena" projetada pelo predicador.
[22] Segundo esse princípio, as representações em cada nível sintático (estrutura profunda, estrutura superficial e forma lógica) são projetadas do léxico, isto é, observam as propriedades temáticas e de subcategorização dos itens lexicais (RAPOSO, 1992).
[23] O Critério-θ tem como finalidade assegurar que as posições projetadas segundo o Princípio de Projeção sejam devidamente preenchidas por argumentos (RAPOSO, 1992).

desenvolvimento das teorias *lexicalistas*[24], um novo ponto de vista sobre a estrutura-a emergiu, segundo o qual ela constitui uma interface entre a semântica e a sintaxe (GRIMSHAW, 1992; LEVIN, PINKER, 1991; BRESNAN, 1981, 1982, 2000; SAG, WASOW, 1999; SAG, 1997; SELLS, 1985).

Além dos *papéis temáticos*, também são empregadas as chamadas *restrições selecionais* na descrição da estrutura de argumentos de um predicador. Tais restrições são traços de natureza semântica caracterizadores dos argumentos selecionados pelo predicador. O princípio básico do emprego dessas restrições é associar a cada argumento do predicador uma lista de *traços* (F_i) que restringem o conteúdo semântico dos argumentos. Essa lista de restrições pode ter diferentes formatos (SAINT-DIZIER, VIEGAS, 1995):

(i) $[F_i]$: traço semântico único, p.ex.: *humano, animado*, etc.;

(ii) $[F_1 \wedge F_2 \ldots \wedge F_n]$: uma conjunção de traços que expressa um conjunto de restrições que devem ser satisfeitas;

(iii) $[F_1 \vee F_2 \ldots F_m]$: uma disjunção de traços: uma das restrições deve ser satisfeita;

(iv) uma combinação de conjunções e disjunções.

O verbo *comer,* por exemplo, projeta dois argumentos, A1 e A2, cujos papéis temáticos são: Agente e Objetivo (ou Tema), respectivamente. O argumento Agente requer o traço [animado] e o argumento Objetivo, por sua vez, requer no mínimo o traço [concreto]. Em outras palavras, os traços *animado* e *concreto* restringem o conteúdo semântico do A1 e A2 do verbo *beber*, respectivamente. O exemplo em (4), elaborado com base na Gramática Funcional - FG (DIK, 1997), ilustra o uso de *papéis temáticos* e de *restrições selecionais* na representação da estrutura de argumentos:

[24] A LFG (KAPLAN, BRESNAN, 1982), a GPSG (GAZDAR et al. 1985) e a HPSG (POLLARD, SAG, 1987; 1994) são exemplos paradigmáticos de modelos lexicalistas. Nesses modelos, o léxico "projeta" a sintaxe (WASOW, 1985).

(4) **comer** [V] (x_1: **<***anim***>** (x_1))$_{Ag}$ (x_2: **<***conc***>** (x_2))$_{Objetivo}$

Em (4), a **estrutura de predicado**[25] especifica a forma ortográfica (*comer*), a categoria gramatical (V) e a valência ou estrutura de argumentos de *comer*. Esta, por sua vez, consiste em duas posições, indicadas pelas variáveis x_1 e x_2, cujas funções semânticas ou papéis temáticos são, respectivamente, Agente (Ag) e Objetivo. Os argumentos indicados pelas variáveis x_1 e x_2 apresentam as restrições selecionais <animado> e <concreto>, respectivamente.

2.4.1.4. Do componente FS

Como mencionado, especifica-se, neste componente, a *forma semântica* (FS) de um item lexical. Segundo Bierwisch e Schreuder (1992), a FS de um item representa a contribuição desse item para o significado das expressões que o contêm. A natureza dessa contribuição é assunto controverso nos estudos lingüísticos. Apesar da controvérsia, uma hipótese geral parece ser a de que a FS de um item restringe o conteúdo proposicional das expressões que o contêm.

Na Teoria do Léxico Gerativo de Pustejovsky (1996), por exemplo, a *forma semântica* de um predicador é especificada em termos de quatro estruturas ou níveis de representação sobre as quais operam mecanismos gerativos. São eles:

(a) a estrutura de argumentos: responsável pela relação entre o léxico e a sintaxe; especifica o número e o tipo de argumentos lógicos, além de especificar o modo como esses argumentos são realizados sintaticamente;

[25] Na FG, os predicados da língua estão armazenados no léxico em *estruturas de predicados*, que especificam um predicado juntamente com um "esqueleto" das estruturas nas quais ele pode aparecer.

(b) a estrutura qualia: responsável pela especificação dos "modos de significação", apresenta os atributos e valores de um objeto em função dos *quale*: FORMAL (de que ***x*** feito); CONSTITUTIVO (as partes de ***x***); TÉLICO (a função ou finalidade de ***x***) e AGENTIVO (como ***x*** origina-se);

(c) a estrutura de eventos: responsável pela descrição dos eventos, estados e transições, fornece os elementos para a representação semântica dos predicados;

(d) a estrutura de herança: responsável, do ponto de vista das categorias léxico-conceituais, pela hierarquização dos itens lexicais, em termos de relações de semelhança, oposição ou inclusão dos itens, impõe ao léxico uma organização global.

No caso deste trabalho, salienta-se que as propriedades relativas à FS dos adjetivos valenciais não foram investigadas e, conseqüentemente, não foram projetadas no modelo de representação lingüístico-computacional (cf. Capítulo 4). Na verdade, tomando o Léxico Gerativo (PUSTEJOVSKY, 1996) como referência, considera-se que a *estrutura de argumento* representa "parte" da informação referente à FS dos adjetivos valenciais.

Capítulo 3 Dos adjetivos do português: propriedades e classificações

3.1. Da categorização do léxico: considerações sobre os adjetivos

Com relação à categorização gramatical do léxico, muito daquilo que ainda hoje é ensinado, tanto nas gramáticas tradicionais (em sua maioria, normativas) como nos manuais de lingüística, fundamente-se na tradição greco-latina. A tradição latina mais antiga distinguia oito "partes da oração": *nome, pronome, verbo, advérbio, particípio, conjunções, preposições* e *interjeições* (BIDERMAN, 2001). Na Idade Média, os gramáticos latinos incluíram nessa lista três novas classes: *nome substantivo, nome adjetivo* e *numeral*. Observa-se que, na obra do gramático alexandrino Dionísio de Trácia (de 170 a 90 a.C.), considerada a primeira descrição gramatical ampla e sistemática publicada no Ocidente, praticamente todas essas categorias já eram objeto de especulação: *nome, pronome, verbo, advérbio, particípio, conjunção, artigo, preposição* e *artigo* (NEVES, 1987).

Esses dados revelam que a tradição gramatical das línguas ocidentais, que se consolida na Renascença, apenas adaptou as classes identificadas no grego e no latim[26]. Foi exatamente isso o que aconteceu com línguas como o português, o espanhol, o alemão, entre outras. E assim tem-se nomeadas as seguintes classes de palavras: *substantivo, adjetivo, pronome, artigo, verbo, advérbio, preposição, conjunção, interjeição* e *numeral*. Quanto a esse esquema classificatório, vários gramáticos e lingüistas vêm fazendo repetidas críticas, evidenciando deficiências e propondo modelos alternativos. Mesmo sem descrever com mais detalhes tais modelos, afirma-se, com base em Biderman (2001), que nenhuma dessas propostas mostrou-se totalmente ideal para a classificação dos elementos do léxico. Contudo, ao contrário do que

[26] Dionísio de Trácia separa o particípio da classe dos verbos, o que não se conservou na tradição gramatical (NEVES, 1987).

o impasse epistemológico revelado por essas críticas possa suscitar, existem certamente esquemas de classificação que operam no léxico das línguas, pois qualquer falante de uma determinada língua utiliza, ainda que de modo intuitivo, um processo classificatório dos itens lexicais que lhes permite distribuir esses itens nas suas realizações discursivas.

Uma questão que aqui se coloca é: haveria classificações universais operando sobre os itens de toda e qualquer língua? No sentido de contribuir com a hipótese de que a estrutura de base das línguas é universal, Lakoff (1970), Lyons (1979) e outros gerativistas procuraram demonstrar que adjetivos e verbos não são constituintes distintos e sim pertencentes à mesma classe, cuja etiqueta seria *verbo*. A argumentação desses teóricos para a reunião de adjetivos e verbos em uma mesma classe pauta-se no fato de que os adjetivos, assim como os verbos, têm função predicativa, como *bonita* em Maria é bonita. Assim, itens como *bonita* seriam marcados, no léxico, com os traços /+ verbo/ e /+ adjetivo/, enquanto que itens como *é* seriam marcados com os traços /+ verbo/ e /– adjetivo/. Lyons destaca, em favor da hipótese de que adjetivos e verbos pertencem a classe *verbo*, o fato de que há línguas, como o chinês, em que o que se entende por adjetivo em nossa tradição gramatical corresponde a um grupo de verbos *estativos*[27] (p.ex.: Tāmen hěn *gǎo* (verbo estativo)) (em português, Eles são muito altos.) (BIDERMAN, 2001).

Na verdade, línguas como o chinês evidenciam o *relativismo* dos processos classificatórios. No português, inclusive, o adjetivo exerce, assim como o verbo, a função predicativa, como em O homem é *alto*. No entanto, esse paralelismo de comportamento gramatical não justifica colocar adjetivos e verbos em uma mesma classe, pois o fato de muitos adjetivos ocorrem no português em função predicativa, comportando-se, portanto, como verbos, não significa que o mesmo ocorra com todos os adjetivos e seus empregos (p.ex.: reforma tributária/ *a reforma é tributária). Além disso, os adjetivos não

[27] Os *estativos* são verbos que não ocorrem no aspecto progressivo, pois se referem ao estado de coisas e não à ação, ao acontecimento ou processo, p.ex.: os verbos como *saber, odiar*. A duração já está implícita nesses verbos (LYONS, 1979).

compartilham com os verbos a propriedade de expressar, por meio de flexões, categorias como tempo, modo e aspecto, por exemplo. Como já mencionara Sapir (1949), há, então, um esquema ou modelo de classificação peculiar a cada língua. A função do lingüista é identificar o esquema subjacente ao sistema lingüístico sob análise.

No que diz respeito ao objeto deste trabalho, é válido afirmar que o *adjetivo* no português é uma classe independente, que apresenta propriedades típicas (BORBA, 1996). Nas seções seguintes, relata-se a breve revisão feita a respeito dos adjetivos e identificam-se as principais características dessa *classe*. Vale salientar que a revisão que se segue não pretende ser exaustiva, pelo contrário, os tópicos revisados dão apenas uma idéia parcial do que tem sido abordado pelos lingüistas a respeito da classe adjetival. No entanto, por meio dessa revisão, foi possível identificar os principais tipos e propriedades dos adjetivos do português do Brasil e, mais especificamente, dos valenciais.

3.2. Da "vertente grega" do adjetivo

Nossa tradição gramatical herdou, via Portugal, o esquema de classificação greco-latino. Segundo Neves (1987), também inspiração do título desta subseção, a primeira classificação das palavras de que se tem registro foi estabelecida pelos gregos, especificamente por Platão. Este tomava o **lógos** (proposição lógica, expressão do juízo) como unidade maior e, a partir dessa unidade, destacava suas partes constitutivas, ou seja, o **ónoma** (nome) e o **rhêma** (verbo), de modo que a constituição das sentenças faz-se pela onomástica e pela retórica. Já aí tem sido vista a identificação de *ónoma* com sujeito e *rhêma* com predicado; e já se isolam os verbos como *rhêma*. Assim, o verbo é definido como o que exprime a ação e o nome, o sinal que se aplica ao sujeito dessa ação. Aristóteles mantém a dicotomia nome/verbo:

> O nome se define como uma voz que tem significação [forma através das quais se dizem coisas] convencional, sem referência de tempo, e que não tem nenhuma parte que, tomada separadamente, seja significativa. O verbo se define como o que acrescenta à sua significação a do tempo; que também não tem nenhuma parte que, tomada separadamente, tenha significação; e que indica sempre alguma coisa afirmada de alguma outra coisa (NEVES, 1987, p.133).

Quanto à classificação proposta por Platão e mantida por Aristóteles, algumas observações são importantes. Primeira, as classes *nome* (ônoma) e *verbo* (rhêma) não compreendem exatamente as classes que foram assim rotuladas nos sistemas posteriores de análise sobre os quais de baseiam as gramáticas escolares (LYONS, 1979). Em outras palavras, *nome* era o termo que funcionava como sujeito de um predicado e o *verbo* era o termo que expressava a ação ou afirmava a qualidade. Segunda, os verbos e os adjetivos, assim como são definidos atualmente, pertenciam a uma mesma classe, *verbo*. A justificativa para esse agrupamento é dada por Lyons:

> Platão e Aristóteles consideravam como função mais típica do adjetivo e do verbo a predicação [atribuição de propriedades às coisas], ao passo que a função mais característica do substantivo era a de denominar o sujeito da predicação. (LYONS, 1979, p.340).

Na própria definição de *verbo* já existe uma menção à função típica de predicação: "ele indica sempre alguma coisa afirmada de alguma coisa". O *verbo* é, portanto, e elemento predicador, independentemente de classe gramatical (NEVES, 1987, p.134).

Posteriormente a Platão e Aristóteles, mais especificamente na obra de Dionísio da Trácia, o *nome* era (i) definido como "parte do discurso flexionável em casos e que designa uma pessoa ou coisa" e (ii) classificado basicamente pelos critérios morfológico e semântico (LYONS, 1979, p.335). Quanto à morfologia, o nome era classificado em dois tipos: *primitivos* e *derivados* (definidos exatamente do modo como se faz atualmente). Do ponto de

vista semântico ou nocional, o *nome* se distribuía em vinte e quatro espécies nocionais bastante diversas, as quais correspondem, nas classificações atuais, a substantivos, adjetivos, pronomes e numerais (NEVES, 1987). Uma dessas espécies era o **epíteto** (ou adjetivo), isto é, "o que é atribuído homonimicamente a nomes próprios e comuns e que indica elogio ou censura, p.ex.: sábio (que é "da alma"); rápido (que é "do corpo"); rico (que é "do extrínseco")" (NEVES, 1987, p.145-6). Com Apolônio Díscolo, o *epíteto* passou também a indicar "grandeza, quantidade ou disposição de alma, ou algo semelhante" e não apenas "elogio ou censura", como em Dionísio de Trácia (NEVES, 1987, p.160).

Observa-se, então, que os gramáticos Dionísio da Trácia e Apolônio Díscolo abandonaram a classificação de Platão (e Aristóteles) e a substituíram por uma outra, que reunia na mesma classe - a dos *nomes* - o que hoje se denomina de *substantivo* e *adjetivo*. Salienta-se que essa reunião era decorrente das flexões causais de gênero e de número que "substantivos" e "adjetivos" manifestavam, tanto em grego quanto em latim.

Segundo Borba (1991) e Bechara (2001), somente na Idade Média, com os gramáticos escolásticos, fez-se a distinção entre *nome substantivo* e *nome adjetivo*. O *nome adjetivo* subdividia-se em **qualificativos** e **determinativos**. Os qualificativos são aqueles que acrescentam "qualidades" às compreensões[28] dos substantivos que modificam (p.ex.: alto, inteligente, entre outros), sendo que a "qualidade" expressa por ele pode ser *acidental* ou *inerente* ao substantivo: quando *acidental*, o adjetivo qualificativo é **restritivo** e, quando *inerente*, **explicativo**[29]. Por exemplo, em homem racional, a palavra *racional* é um adjetivo explicativo porque indica uma qualidade essencial (ou inerente a) de *homem*, isto é, qualidade que está incluída na idéia expressa pela palavra *homem*; em homem razoável, a palavra *razoável* é um adjetivo restritivo, pois acrescenta uma propriedade acidental às

[28] Por *compreensão* (ou *intensão*) de um substantivo, entende-se o conjunto de propriedades que fazem um indivíduo ser designado por ele (BORGES NETO, 1991, p.12).
[29] Abandonou-se a distinção entre adjetivos qualificativos *restritivos* e *explicativos* nas classificações posteriores.

essenciais expressas pela palavra *homem*. Os **determinativos**, por sua vez, são aqueles que delimitam áreas na extensão[30] do substantivo que modificam (p.ex.: *estes*, *aqueles*, *dois*, *vários*, *certo*, *mesmo*, entre outros) (PEREIRA, 1918). Em outras palavras, operam sobre o conjunto de entidades denotadas pelo núcleo nominal.

Com o advento da Nomenclatura Gramatical Brasileira (NGB)[31], em 1959, dez classes de palavras ou gramaticais foram rotuladas: *substantivo, adjetivo, pronome, artigo, numeral, verbo, advérbio, preposição conjunção* e *interjeição*. Os elementos do português rotulados "adjetivo", em específico, ficaram restritos aos *qualificativos*, sendo os *determinativos* distribuídos nas categorias *artigo, numeral* e *pronome* (BORGES NETO, 1991).

3.3. Da caracterização geral dos adjetivos do português

Para a caracterização geral da classe, decidiu-se adotar a conceituação proposta por Borba (1996). Segundo essa conceituação, pressupõe-se que:

(i) a adjetivação é um processo mental de diferenciação, discriminação e seleção;

(ii) o adjetivo é palavra de natureza abstrata;

(iii) a classe adjetival constitui uma categoria não autônoma, tendo como uma de suas características não incidir sobre si próprio.

Cabe salientar, aliás, que a propriedade (iii) pode ser inferida da própria etimologia da palavra "adjetivo". O *ad-*, prefixo latino da palavra, indica idéia de "acompanhamento", ou seja, o adjetivo está sempre "preso" a um outro elemento. Além disso, observa-se que, etimologicamente, "adjetivo" é uma palavra latina que pertence à mesma família de "adjacente", "jacente" e tantas outras palavras. No verbete *jact-*, elemento de composição

[30] A *extensão* de um substantivo, em oposição a *intensão*, é o conjunto de indivíduos que esse substantivo denota (BORGES NETO, 1991, p.12).

[31] A NGB foi instituída tendo em vista as inúmeras terminologias distintas empregas no ensino de língua portuguesa.

latino, o Dicionário Houaiss (HOUAISS, 2001) fornece uma lista extensa de palavras dessa família. Isso quer dizer que o adjetivo é sempre colocado próximo ou ao lado de um substantivo. Dessa forma, é próprio dessa classe não incidir sobre si mesmo, mas aplicar-se a um suporte sobre o qual o discurso faz incidir momentaneamente (VILELA, 1995; BORBA, 1996).

A seguir, sobretudo porque este trabalho destina-se também a leitores de outras áreas do conhecimento que não necessariamente estejam familiarizados com os estudos sobre as classes de palavras, considera-se oportuno sumarizar cada uma das características da classe dos "adjetivos" apontadas na literatura estudada. Para tanto, recorre-se a diferentes perspectivas teóricas e leva-se em conta o entrelaçamento de propriedades existente principalmente entre a classe dos adjetivos e a dos verbos.

3.4. Do ponto de vista morfossintático: as marcas de gênero, de número e de grau

3.4.1. As desinências de *número* e de *gênero*

Salienta-se que a própria definição de *epíteto*, elaborada por de Dionísio da Trácia, ou seja, "o que é atribuído homonimicamente a nomes próprios e comuns", faz alusão ao fato de que os adjetivos apresentam os mesmos acidentes flexionais que os substantivos.

(1) Da "categoria gramatical secundária" de número

A manifestação mais comum da categoria de *número* é a distinção entre *plural* e *singular* (p.ex.: casa/ casas), existente em muitas línguas do mundo. Tal distinção relaciona-se com o reconhecimento de que pessoas, coisas e animais podem ser enumerados e referidos, individual e coletivamente, por meio de substantivos (LYONS, 1979). Dessa forma, a

categoria de *número* é típica dos substantivos. Em português, especificamente, o adjetivo concorda com o número do substantivo (p.ex.: bicicleta *azul*; bicicletas *azuis*). A flexão numérica do adjetivo é, portanto, derivada de regras de concordância com o substantivo que ele modifica. O adjetivo no português conhece, então, os dois números, *singular* e *plural*, tal como os substantivos, sendo que o {-s} é a marca própria do plural, oposta ao vazio do singular. A regra geral de formação do *plural* é o acréscimo ao adjetivo singular do sufixo flexional {–s} (p.ex.: grande + -s = grandes) (cf. CUNHA, CINTRA, 1985). Quanto aos adjetivos compostos, vale lembrar que normalmente só o último varia (p.ex.: luso-brasileiras), com algumas exceções (p.ex.: surdos-mudos).

(2) Da "categoria gramatical secundária" de gênero

O substantivo sempre tem um gênero, o que não acontece com o adjetivo, que assume o gênero do substantivo. Em algumas línguas, como o inglês, os adjetivos têm uma única forma, invariável, quer o substantivo seja *masculino* (p.ex. *beautiful* car) (em português, "carro bonito") ou *feminino* (p.ex.: *beautiful* house) (em português, "casa bonita"). Em português, por outro lado, o adjetivo flexiona-se de acordo com o gênero do substantivo (p.ex.: carro *bonito*; casa *bonita*), podendo ser, então, *masculino* (p.ex.: bonito) ou *feminino* (p.ex.: bonita).

Com relação à formação do feminino, os adjetivos do português são classificados como *uniformes* ou *biformes*. Os uniformes são aqueles que apresentam uma só forma para acompanhar nomes masculinos e femininos (p.ex.: *breve* exame; *breve* prova)[32]; são também denominados adjetivos *de dois gêneros* (2 g). A formação dos biformes (isto é, adjetivos que têm uma forma para acompanhar o substantivo masculino e outra para o feminino) segue de perto as mesmas regras que se aplicam aos substantivos, sendo que a regra geral é o acréscimo, para o feminino, do sufixo flexional {–a} (átono final) com a supressão

[32] Cunha e Cintra (1985) apresentam sete regras para os adjetivos *uniformes*.

da vogal temática, quando ela existe no singular (p.ex.: bel(o) + -a = bela)[33]. Aliás, essa identidade de forma, isto é, ter os mesmos acidentes flexionais (gênero e número), também foi responsável por levar gramáticos normativos a colocar adjetivos e substantivos em uma mesma classe, a dos *nomes*.

Dessa forma, vê-se que, do ponto de vista morfológico, o adjetivo tem o poder, na maioria das vezes, de apresentar duas terminações de gênero, sem que, com isso, seja uma palavra de gênero determinado e sem que o conceito por ele designado corresponda a um gênero real (BHAT, 1994).

3.4.2. Do grau

Muitos gramáticos normativos afirmam que os adjetivos podem flexionar-se em *número*, *gênero* e *grau*, indicando, em seguida, as diversas regras de flexão (p.ex.: CUNHA, CINTRA, 1985; BECHARA, 1992). Rocha Lima (1972) e Bechara (2001), entretanto, reconhecendo que o *grau* não é um processo de flexão no português, não o consideram como tal. Nesse sentido, Bechara (2001) observa que o *grau* só figura na sessão "Flexões do adjetivo" por ter sido contemplado como tal pela NGB.

Mais especificamente, a maioria dos adjetivos em português apresenta grau *absoluto* (p.ex.: fácil, triste). Entretanto, alguns adjetivos apresentam comumente uma possibilidade de indicar, por meio de um morfema adicional, o grau superlativo da propriedade que expressam (MATTOSO CÂMARA, 1970; LYONS, 1979). Temos, por exemplo, *facílimo,* para *fácil*, *lindíssimo,* para *lindo,* etc. Daí o nome: *grau superlativo* (*absoluto sintético*). Pode-se formar também esse grau com o a acréscimo de um prefixo ou de um pseudoprefixo, como *arqui-* (p.ex.: arquimilionário), *hiper-* (p.ex.: hipersensível), *super-* (p.ex.: superexaltado), *ultra-*

[33] Para a obtenção de mais detalhes sobre as demais regras de formação dos adjetivos *biformes* do português, sugere-se conferir Cunha e Cintra (1985).

(p.ex.: extrafino), etc. (CUNHA; CINTRA, 1985). Segundo Mattoso Câmara (1970), não há uma obrigatoriedade no emprego do adjetivo com esse sufixo de superlativo ou grau intenso. Na realidade, o que se tem com os superlativos é uma *derivação* possível em muitos adjetivos. Em outras palavras, a expressão de grau não é um processo de flexão em português porque não faz parte do mecanismo de concordância da língua, ou seja, o grau acontece para o adjetivo sem que necessariamente ocorra para o substantivo e vice-versa (MATTOSO CÂMARA, 1970). Por exemplo, em *homem lindíssimo*, o substantivo se manteve sem a marcação de grau.

Como delineado, o componente TG das entradas lexicais engloba informações sobre a especificação das categorias gramaticais primárias e secundárias. As categorias gramaticais secundárias, em específico, dizem respeito ao conjunto de flexões que caracterizam os itens lexicais. Tendo em vista que o *grau* não é um processo de flexão em português, ficou restrito ao componente TG dos adjetivos valenciais a função de especificar as informações relativas às categorias gramaticais secundárias de *gênero* e de *número*, que são resultantes de regras de concordância.

3.5. Das subclasses dos adjetivos: os *qualificadores* e os *classificadores*

3.5.1. Dos *qualificadores* e a macroestrutura do léxico

Neste trabalho, parte-se do princípio de que os adjetivos do português podem ser classificados em duas grandes classes semânticas: a dos *qualificadores* (QLs) e a dos *classificadores* (CLs) (ROCHA, 1994; BORBA, 1996; NEVES, 2000).

Um adjetivo é QL quando expressa o *valor* de um atributo do substantivo que não necessariamente compõe o feixe das propriedades que o definem. Mais especificamente, dizer

"*x* é QL" ou "*x* QL" pressupõe postular que há um atributo A, tal que A(x) = adj, isto é, o adjetivo QL é o valor da função A(x). Por exemplo, dizer o muro é *alto* ou muro *alto* significa postular a função ALTURA(muro) = *alto*. Ou ainda, dizer a flor é *branca* ou flor *branca* pressupõe postular o atributo do tipo COR, tal que COR(flor) = branca. Em outras palavras, pode-se dizer que esses adjetivos qualificam o substantivo, o que pode implicar uma característica mais ou menos subjetiva, mas sempre revestida de uma certa vaguidade (NEVES, 2000).

Assim definidos, são qualificadores (i) os adjetivos formados com prefixos negativos (p.ex.: desagradável, imaturo); (ii) os formados por prefixos intensificadores (p.ex.: hipervazio, super-reativa), (iii) os deverbais (p.ex.: delicioso, respeitado, gostoso), (iv) os que admitem sufixo superlativo ou sufixo diminutivo com valor de intensificação (p.ex.: fraquíssimo; pequenininho) (NEVES, 2000).

O termo "atributo" empregado na definição do adjetivo QL faz referência a um construto que pode ser igualmente denominado "qualidade", "traço" ou "dimensão". O teste comumente empregado para a identificação dos QLs é a nominalização do atributo expresso pelo adjetivo, p.ex.: o muro é *alto*/ muro *alto* > a altura do muro. Devido ao fato de aceitarem a nominalização do atributo, Quirk et al. (1991) classificam os QLs como sendo tipicamente **inerentes**, contrapondo-se aos **não-inerentes**, que não aceitam a nominalização da propriedade que expressam, p.ex.: crise ministerial > *a *ministerialidade* da crise.

Certos adjetivos são, em princípio, QLs, mas, combinados com determinados substantivos, podem exercer a função semântica de classificadores. A seguir, exemplifica-se essa permeabilidade entre as classes:

(5) a. Era o vestido *branco* da filha, os sapatos *brancos*, o véu *branco*, as flores de laranjeira. (CG) (= qualificador)

b. Finalmente, o homem *branco* se apresenta aos índios. (AVL) (classificador)

Segundo Neves (2000), os QLs podem passar a CLs especificamente em construções cristalizadas, como em (6).

(6) a. Água *doce*, o mar e o solo úmido. (GAN)

b. O mar fica a trinta léguas de distância mas diz o povo que escuta o estrondo da estrela cadente quando se afoga na água *salgada*. (BP)

Quanto às relações da macroestrutura do léxico, salienta-se, como já se disse, que a relação léxico-semântica básica dos QLs é a **antonímia** ou "oposição de sentido" (DEESE, 1964; JUSTESON, KATZ, 1989). Para Lyons (1979), a antonímia, há muito, é reconhecida como uma das mais importantes relações semânticas.

Para o equacionamento lingüístico-computacional dessa relação léxico-semântica no âmbito do projeto WordNet (FELLBAUM, 1999), pressupõe-se que: (i) o QL expressa o *valor* de um "atributo" do substantivo; (ii) os atributos são *bipolares*, isto é, adjetivos antônimos como grande é pequeno, por exemplo, expressam os valores dos pólos do atributo TAMANHO; (iii) os atributos podem ser *graduáveis* (contínuos) ou *não-graduáveis* (dicotômicos), ou seja, o atributo graduável ALTURA, por exemplo, varia em um contínuo de "alturas" entre os valores alto e baixo, ao quais, na verdade, são os valores dos pólos do atributo ALTURA; já o atributo não-graduável SEXO, por exemplo, apresenta apenas dois valores: macho e fêmea; e, por isso, são denominados dicotômicos.

Cabe ressaltar ainda que a antonímia é uma relação *léxico-semântica*, ou seja, é um tipo de relação que se estabelece entre unidades da língua e não entre conceitos. Um fato lingüístico que corrobora com essa definição é a existência de antônimos formados pelo acréscimo de um sufixo derivacional (p.ex.: *in-*/*im-*, *anti-*, *a-*, etc.) que muda a polaridade do significado do adjetivo qualificador, p.ex.: abalável/ inabalável (FELLBAUM, 1999).

3.5.2. Dos *classificadores*

De acordo com Borba (1996), Demonte (1999b) e Neves (2000), os adjetivos CLs colocam o substantivo com o qual ocorrem em uma determinada subclasse, nomeando-a (p.ex.: reforma *previdenciária*); daí o rótulo *classificador*. Borba (1996) salienta que a relação semântica que se estabelece entre reforma e previdenciária, em reforma *previdenciária*, é de natureza externa porque o adjetivo apenas coloca o substantivo em uma determinada classe.

Em outras palavras, pode-se dizer que a função semântica desses adjetivos é a de combinar ou relacionar propriedades denotativas de um substantivo com as propriedades denotativas de outro substantivo (LOBATO, 1993; BASÍLIO, GAMARSKI, 2002). Por exemplo: em reforma *previdenciária* > reforma da previdência, observa-se que a paráfrase "da previdência" sinaliza que previdenciária liga a entidade reforma à entidade previdência, ou seja, o adjetivo classificador previdenciária relaciona propriedades denotativas do substantivo reforma e do substantivo previdência. Em outras palavras, pode-se dizer que o adjetivo QL é um modo de conceber e o CL é um modo de relacionar entidades (BORBA, 1996). Por isso se diz que os CLs são **relacionais** (GROSS, MILLER, 1990; BOSQUE, PICALLO, 1996; DEMONTE, 1999b).

Vale ressaltar que, os adjetivos CL são geralmente denominais e correspondem à locução adjetival "*de* + substantivo". Quando o adjetivo equivale a um sintagma preposicional desse tipo, convém verificar, em dependência do substantivo com o qual ocorrem, se a noção é qualificadora ou classificadora (NEVES, 2001). Por exemplo, o adjetivo criminal, em ato *criminal,* é um QL, pois pode ser relacionado à propriedade que indica, no caso, a criminalidade do ato; já em advogado *criminal*, o adjetivo criminal é CL (p.ex.:

*criminalidade do advogado.). Muitas vezes, quando um adjetivos tipicamente CL passa a QL, ele comumente está em uso metafórico, com possibilidade de anteposição, p.ex.:

(7) a. Desconhecido olhava a cena tomado dum *subterrâneo* temor. (N)

b. A *polivalente* personalidade de César Salgado impõe, à crítica, o dever de partir de um determinado critério, para situá-lo no panorama da cultura brasileira. (FI)

Enquanto os adjetivos QLs expressam "qualidades" ou "valores de atributos" dos substantivos, os CLs são comumente definidos, em obras lexicográficas, por meio de paráfrases como "de ou pertencente/ relativo a X" (MILLER, FELLBAUM, 1991). A seguir, ilustra-se esse tipo de definição com os dois verbetes (8) e (9):

(8) **criminal** (FERREIRA, 1999)

[Do lat. criminale.]

Adj. 2 g.

1. Relativo ou pertencente a crime; criminoso, crime.

(9) **campestre** (WEISZFLOG, 1998)

adj m+f (lat campestre)

Relativo ao campo; rural rústico.

Concluindo: tendo em vista que a classificação das palavras, independentemente do rótulo a elas atribuído, tem como principal objetivo agrupar aquelas que têm comportamento gramatical semelhante, parece pertinente a classificação dos adjetivos em **qualificadores** e **classificadores**.

3.6. Das funções sintáticas dos *qualificadores* e *classificadores*

As duas funções sintáticas básicas desempenhadas pelos adjetivos relacionam-se com sua posição: posição adnominal (Padn) e posição predicativa (Ppred) (BORBA, 1996). Em Padn, os adjetivos desempenham a função denominada **adjunto adnominal** e, em Ppred, a de **predicativo** (NEVES, 2000).

Em função adnominal, o adjetivo é periférico no sintagma nominal; diz-se que o adjetivo associa-se ao substantivo de forma direta. Em função predicativa, o adjetivo é núcleo no sintagma verbal, portanto, núcleo do predicado (BORBA, 1996). Mais especificamente, quando em Ppred, o adjetivo associa-se ao substantivo de forma indireta, ou seja, liga-se ao sujeito da oração (o funcionário) por meio de um verbo-suporte ou copulativo, isto é, verbo semanticamente vazio que funciona como suporte do *tempo*, *número* e *pessoa* (VILELA, 1995, NEVES, 2000; BIDERMAN, 2001). O verbo copulativo, entretanto, pode não estar expresso na sentença (NEVES, 2000), como em (10):

(10) Apesar de *amáveis*, era evidente que também os Barros estavam constrangidos. (A)
(= apesar de serem amáveis)

Quando ocorrem com verbos copulativos, pode-se dizer que os adjetivos predicativos do português "controlam" a seleção dos verbos copulativos *ser* e *estar*. Com o verbo *ser*, o adjetivo relaciona-se a qualidades permanentes, p.ex.: O funcionário é *alto*; com o verbo *estar*, relaciona-se a qualidades temporárias, p.ex.: O funcionário está *alto* (com aquele sapato).

Vale ressaltar que a classificação em qualidades *permanentes* (ou essenciais) e *temporárias* (ou acidentais) originou-se na escola aristotélica (LYONS, 1979; HACKING,

2001). Outras duas terminologias são encontradas para a mesma classificação. A primeira, que distingue adjetivos do tipo *individual level* (que se relacionam a qualidades permanentes, essenciais) e adjetivos do tipo *stage level* (que se relacionam a qualidades temporárias, acidentais), é atribuída a Carlson (1977) e retomada por Kratzer (1989), Pustejovsky (1996) e Pesetsky e Torrego (2002). A segunda, que distingue adjetivos *individuais* (ou estáveis) e adjetivos *episódicos* (ou de estado), é empregada por Demonte (1999b).

Em (11a), estão listados alguns adjetivos qualificadores e, em (11b), alguns classificadores. A seguir, observa-se o comportamento desses elementos em relação às funções sintáticas de adjunto adnominal e de predicativo.

(11) a. alegre; bonito; calmo; frio. b. campestre; citadino; municipal; rural.

Tomando *calmo* e *alegre* como exemplos do grupo em (11a), observa-se, por meio das ocorrências em (12) e (13), que tais elementos podem desempenhar as duas funções sintáticas básicas.

(12) a. O marido, exato burocrata, era cordato, de temperamento *calmo*, ponderado e seco que nem uma crise ministerial. (LI-LT) (Padn)

b. [...] sentia-se em torno dela o doce e *calmo* calor das paixões [...]. (LI-LT) (Padn)

c. André Santos é *calmo* e sabe sair jogando com muita facilidade. (JO-FSP) (Ppred)

(13) a. Nunca vou esquecer da fisionomia *alegre* daquela primeira professora. (VEJ) (Padn)

b. Nele a erudição foi transformada em *alegre* licor de jenipapo. (CAR-O) (Padn)

c. O garoto era comunicativo e *alegre*. (JO-FSP) (Ppred)

Assim, os itens *calmo* e *alegre* e, conseqüentemente, todos os demais elementos pertencentes ao conjunto em (11a), podem ocorrer em função adnominal e em função predicativa. Tomando agora os itens *rural e municipal* como exemplo dos elementos do conjunto em (11b), buscou-se encontrar nos córpus de referência deste trabalho, ou seja, no córpus do LL e no córpus do NILC, ocorrências desses adjetivos em função adnominal e predicativa. Nesses córpus, no entanto, os adjetivos *rural* e *municipal* aparecem apenas em função adnominal, como ilustram os exemplos em (14).

(14) a. Desanimado, estende sua dor à pobre população *rural* ... (LI-LT)

b. Palmira morreu e foi sepultada no cemitério *municipal*. (LI-LT)

Quanto à função predicativa, ressalta-se que esta não é exclusiva, mas sim característica dos adjetivos do grupo em (11a). Há certos adjetivos do grupo em (11b) que admitem a função predicativa quando em condições contextuais específicas, como (i) em construções contrastivas (p.ex.: Estas viaturas são municipais; aquelas, não.); (ii) com repetição do núcleo do sintagma nominal (p.ex.: Esta estrada é uma *estrada* vicinal.) (CASTELEIRO, 1981; BORBA, 1996).

Dessa forma, vê-se que os adjetivos qualificadores listados em (11a) não exercem as mesmas funções, ou melhor, não ocorrem nos mesmos contextos sintático que os adjetivos classificadores listados em (11b). Observa-se que os adjetivos em função adnominal são comumente denominados **atributivos** e os adjetivos em função predicativa são denominados **predicativos** (BOLINGER, 1967; LEVI, 1978, QUIRK et al., 1991; TEYSSIER, 1968).

Salienta-se também que os adjetivos qualificadores e os classificadores diferem não só quanto às funções adnominal e predicativa, mas também quanto à *gradação* e *comparação* (BORBA, 1996). Enquanto os adjetivos qualificadores aceitam *gradação* e *comparação*,

como ilustrado nos exemplos em (15), o mesmo não acontece com os classificadores (p.ex.: *A casa é muito *rural/ *Uma casa é mais *rural* do que a outra). A gradação, inclusive, parece ser o traço sintático típico do adjetivo e, por isso, alguns autores como Postal (1969) e Nique (1974) consideram os QLs **verdadeiros adjetivos**. E mais, os adjetivos qualificadores e classificadores não podem ser coordenados, por exemplo: *engenheiro *civil* e *rude* (LEVI, 1978).

(15) a. André Santos é muito *calmo* e sabe sair jogando com muita facilidade. (LI-LT)

b. Carlos é mais clamo do que Luis.

A correspondência entre as propriedades sintáticas e as subclasses é representada no Quadro 4.

Adjetivos	**Propriedades distribucionais**				**Classificações**
	(a)	(b)	(c)	(d)	
alegre	+	+	+	+	
bonito	+	+	+	+	qualificadores
calmo	+	+	+	+	
frio	+	+	+	+	
campestre	+	?	-	-	
citadino	+	?	-	-	classificadores
municipal	+	?	-	-	
rural	+	?	-	-	

(a) = ocorrer em função adnominal;
(b) = ocorrer em função predicativa;
(c) = gradação;
(d) = comparação.
(?) = reduzida aceitabilidade.

Quadro 4: Classificação distribucional dos adjetivos

3.7. Das propriedades sintático-semânticas dos adjetivos

3.7.1. Dos adjetivos em Padn: a função de *modificadores*

Quando em posição adnominal, diz-se que os adjetivos exercem a função de *modificação*. De acordo com Perini (1998), o processo de **modificação** seria um misto de semântica e sintaxe. Semanticamente, esse processo significa que um adjetivo teria seu significado amalgamado ao de outro elemento, formando um todo semanticamente integrado; assim, casa exprime uma entidade, e casa azul exprime a mesma entidade, acrescida de algum ingrediente de significado. Sintaticamente, a noção de modificação parece referir-se à ocorrência conjunta dentro de um constituinte. Desse modo, casa azul forma um constituinte (ou sintagma).

3.7.1.1. Da semântica dos qualificadores em Padn

Como mencionado, o adjetivo QL expressa o que seria o *valor* de um *atributo* do substantivo ao qual se liga. Assim, para os QL, o processo de modificação, anteriormente definido, pode ser entendido semanticamente da seguinte maneira: em Padn, o adjetivo QL expressa o valor de atributo *preexiste* ao julgamento do falante. De acordo com essa visão, pressupõe-se que a categoria do substantivo seja um conjunto de propriedades ou atributos e que a função do adjetivo QL em Padn ou em função modificadora é a de preencher o valor de um desses atributos. Por exemplo, pode-se entender que o substantivo *casa* seja previamente composto pelos atributos TAMANHO, COR e CÔMODOS; em casa azul, o QL especifica o valor do atributo COR, preexistente ao julgamento do falante.

Em decorrência da posição ocupada pelo adjetivo QL no sintagma nominal e da natureza semântica do substantivo com o qual está ligado, ocorrem diferenças nos resultados semânticos (NEVES, 2000). Pode-se dizer que, em alguns casos, a posição do adjetivo no sintagma nominal e o tipo semântico do substantivo determinam se o atributo, cujo valor é expresso pelo QL, é o do **referente**[34] ou o da **referencia**[35]. Por exemplo:

(16) a. Quem me contou foi um homem *velho* que esteve lá. (B)

b. Apresento-te um *velho* amigo, companheiro de colégio. (AV)

De acordo com Chierchia e McConnell-Ginet (1990), o adjetivo velho, em (16a), tem interpretação **intersectiva** e, por isso, (17a) implica (17b) e (17c):

(17) a. um homem *velho*

b. → o indivíduo é *homem*

c. → o indivíduo é *velho*

Nesse caso, diz-se que o *atributo*, cujo valor (velho) é expresso pelo adjetivo, é do *referente*, sendo o QL, portanto, uma *propriedade de primeira ordem* (LYONS, 1977). Quando indica *propriedade de primeira ordem*, o adjetivo QL expressa o valor de um atributo do *referente* (= da entidade), no caso, "ser humano do sexo masculino". Já em (17b), o adjetivo velho tem interpretação **não-intersectiva**; nesse caso, (18a) implica (18b), mas não necessariamente (18c) (CHIERCHIA; MC-CONNELL-GINET, 1990):

[34] O termo *referente* está sendo empregado para designar as entidades (do mundo real ou fictício) a que as expressões lingüísticas fazem referência (DUBOIS et al., 1973, p. 512).

[35] O termo *referência* está sendo empregado para designar o modo lingüístico pelo qual se refere às entidades. (DUBOIS et al., 1973, p. 511). Por exemplo: o indivíduo "Ronaldo" pode ser objeto de diferentes referências, p.ex.: "o melhor jogador de futebol do ano", "o (ex-)marido de Milene Rodrigues", "o pai de Ronald", etc.

(18) a. este *velho* amigo

b. → este indivíduo é *amigo*

c. /→ este indivíduo é *velho*[36]

Dessa forma, o adjetivo velho, em (16b), pode expressar o valor de um atributo do *referente* ou da *referência*. Quando expressa o valor de um atributo da *referência*, diz-se que o QL é uma **propriedade de segunda ordem**. (LYONS, 1977), pois expressa o valor do atributo de uma outra propriedade, no caso, "ser amigo".

Para Bouillon (1999), o adjetivo velho, em (16b), quando especifica o valor de um atributo da *referência,* funciona como um *modificador de evento*. Para esse autor, o adjetivo QL expressa o valor do atributo do evento derivado do substantivo amigo, ou seja, "a relação de amizade".

Tomando como ilustração o modelo de Léxico Gerativo (LG) de Pustejovsky (1996), em que os substantivos são representados semanticamente pelos seus "modos de significação", mais especificamente, pelos *qualia* FORMAL (de que ***x*** é feito), CONSTITUTIVO (as partes de ***x***), TÉLICO (a função/ finalidade de ***x***) e AGENTIVO (como *x* origina-se), pode-se dizer que velho, em (16b), pode incidir tanto no valor *indivíduo* como no valor *evento* do quale FORMAL de substantivos como do tipo de amigo (AMARO, 2002).

3.7.1.2. Da semântica dos classificadores

Como mencionado, a função semântica do CL é colocar o substantivo com o qual ocorre em uma determinada classe. Quando em função adnominal, os adjetivos *classificadores* ocorrem preferencialmente pospostos ao núcleo do sintagma, p.ex.:

[36] O símbolo (/) indica padrão de inferência não válido.

(19) a. Encostado à primeira bancada, um rapaz lia um folheto; ao longo da mesa *presidencial*, na frente, atrás, dos lados, havia um vaivém continuado. (LI-LT)

b. Não estaria a moça representando uma cena da comédia *matrimonial* que a divertia? (LI-LT)

c. A geradora, em fase de licenciamento para operar, quer enquadrar-se às normas da legislação *ambiental*. (JO-FSP)

Cabe ressaltar que pode haver, no entanto, construções cristalizadas em que o adjetivo classificador vem anteposto, como em (20) (NEVES, 2000):

(20) O *Pátrio poder* era exercido pelo homem, com a ajuda da mulher, até 1997, quando saiu a lei do divórcio. (VEJ)

Em (19), os adjetivos são, na verdade, argumentos dos substantivos, ou seja, eles expressam o que seria o complemento dos substantivos (p.ex.: mesa *presidencial* > mesa do presidente; comédia *matrimonial* > comédia do matrimônio; legislação *ambiental* > legislação *do ambiente*). No plano morfológico, os adjetivos classificadores são geralmente denominais e correspondem à locução adjetival "*de* + substantivo", como em mesa *presidencial* < do presidente ou em comédia *matrimonial* > do matrimônio; por essa razão, teóricos como Coates (1971) denominam esses adjetivos de *denominais*. De acordo com Neves (2000), tais adjetivos têm, portanto, a mesma distribuição, no texto, que essas locuções, e freqüentemente se coordenam com elas, como ilustrado em (21).

(21) Entende-se, assim, o aparecimento dos sistemas *digestivo* [de digestão]*, respiratório* [de respiração]*, de transporte, excretor* [de excreção]. (FIA)

Salienta-se ainda que a impossibilidade de anteposição desses adjetivos no sintagma nominal pode estar ligada ao fato de que o português é uma língua que lineariza seus complementos à direita do predicador (no caso, o substantivo) e, sendo esses adjetivos complementos dos substantivos, não poderiam ser colocados à esquerda do substantivo (MENUZZI, 1992). O fato desses adjetivos não aceitarem a posição predicativa, por outro lado, parece se relacionar com a impossibilidade de associá-los ao nome da propriedade correspondente (nominalização), já que eles não expressam propriedades, p. ex: a crise ministerial > *ministerialidade da crise (LEVI, 1978; BORBA, 1996).

No entanto, há alguns adjetivos *classificadores* que não desempenham função de complemento do substantivo; esse é o caso, por exemplo, de profissional, habitual e pluvial, em (22).

(22) a. O jogador *profissional* [= *x* joga por profissão] só pode passar à classe de amador após ter cumprido o contrato que tenha celebrado com um clube [...] (JO-FSP)

b. Lílian finalizou a carta dizendo que é ouvinte *regular* [= *x* lê regularmente] de nossos programas [...] (JO-FSP)

c. Poderosos Senhores que faziam uso da navegação *fluvial* [= *x* navega por rio] para transportar suas riquezas e mercadorias [...] (JO-FSP)

Os adjetivos descritos em (22) equivalem a sintagmas preposicionais de valor adverbial; em outras palavras, diz-se que esses adjetivos são *circunstanciais*. Segundo Borba (1996) e Demonte (1999a), tais adjetivos podem expressar vários valores semânticos, por exemplo: *local* (p.ex.: traumatismo *craniano* > que ocorreu no crânio), *tempo* (p.ex.: publicação *mensal* > feita por mês), entre outros.

3.7.2. Dos adjetivos em Ppred: a função de *predicadores*

Se, em Padn, o processo desencadeado pelos adjetivos é a *modificação*, em Ppred, como o próprio nome indica, o processo é a **predicação**. Em outras palavras, entende-se que, quando modificadores, os QLs expressam que o *valor* de um *atributo* "preexiste" ao julgamento do falante e, em função predicativa, os QLs instauram um processo de predicação, ou seja, de "atribuição" de propriedade. Sendo, portanto, **predicadores** (PRs), os QLs selecionam (ou projetam) os argumentos com os quais ocorrem (ROCHA, 1994; BORBA, 1996; NEVES, 2000). Por exemplo:

(23) a. Michaelowsky (A1) está *desgostoso* com a cor governista do jornal (A2). (LI-LT)

b. Camargo (A1) era pouco *simpático* à primeira vista. (LI-LT)

O adjetivo *desgostoso*, em (23a), é de valência 2, ou seja, projeta (ou exige) dois argumentos no nível lógico-semântico: argumento 1 (A1) e argumento 2 (A2). No nível sintático, o A1 realiza-se na forma de um SN (Michaelowsky); o A2, que corresponde ao complemento (Comp), realiza-se na forma de um sintagma preposicional (SPrep) *com*-SN (com a cor governista do jornal). Em outras palavras, diz-se que o adjetivo *desgostoso* subcategoriza dois constituintes, p.ex.: *desgostoso* [SN, SPrep]. Esses dois constituintes formam o *esquema de subcategorização* de *desgostoso*. Os argumentos A1 e A2 projetados por esse adjetivo associam-se a papéis temáticos que especificam a *estrutura de argumentos* ou *valência semântica* desse adjetivo. No caso, o A1 (Michaelowsky) associa-se ao papel temático Experimentador (Ex) e o A2 (com a cor governista do jornal), ao Causativo (Ca).

O adjetivo *simpático*, em (23b), por sua vez, projeta (ou requer) apenas um argumento no nível lógico-semântico, A1, que se realiza sintaticamente na de SN (Camargo). Assim, no quadro de subcategorização do adjetivo *simpático*, consta apenas um constituinte, cuja forma é SN. O argumento A1 projetado por *simpático* associa-se ao papel temático Objetivo (Ob). No caso da ocorrência em (23b), a expressão à primeira vista é apenas um *circunstancial* e não faz parte da valência do adjetivo.

3.7.3. Dos *qualificadores* e a microestrutura do léxico

Sendo, então, predicadores ou valenciais, como em (23), salienta-se que os adjetivos podem ser descritos e representados, quanto à *microestrutura* do léxico, como os verbos. Isso quer dizer que a *entrada lexical* desses adjetivos deve conter as informações relativas à *estrutura de argumentos* e ao *esquema de subcategorização*.

3.7.4. Da valência adjetival

Partindo-se das premissas da Gramática de Valência (Borba, 1996), afirma-se que a relação estabelecida entre um predicador e seus argumentos ocorre em três níveis: *lógico-semântico, sintático ou morfossintático* e *semântico*.

3.7.4.1. Da valência lógico-semântica

No nível lógico-semântico, caracteriza-se o número de argumentos projetados ou requeridos por um predicador. De acordo com Busse e Vilela (1986), há duas interpretações possíveis para a valência adjetival. Na primeira, consideram-se apenas os constituintes

dependentes dos adjetivos; na segunda, considera-se também o constituinte em função de sujeito. Neste trabalho, optou-se pela segunda interpretação e, conseqüentemente, os adjetivos valenciais do português podem ter (BORBA, 1996):

- **Valência 1**, quando projetam apenas um argumento, p.ex.:

(24) Meu pai (A1) era *alto*, loiro e de olhos azuis. (JO-FSP)

- **Valência 2**, quando podem projetar dois argumentos, p.ex.:

(25) [..] Renunciou à convicção porque ela (A1) não era *útil* a seus propósitos (A2). (JO-FSP)

- **Valência 3**¸ quando podem projetar três argumentos, p.ex.:

(26) O réu (A2) era *condenável* à pena de morte (A3) pelo juiz. (A1)

- **Valência 4**, quando podem projetar quatro argumentos, p.ex.:

(27) A carga (A2) era *transportável* do estaleiro (A3) para o navio (A4) por guindastes. (A1)

3.7.4.2 Da valência sintática

Todo predicador, além de determinar o número de argumentos lógico-semânticos, determina também (i) a categoria sintagmática dos argumentos e (ii) a relação semântica que se estabelece entre ele e seus argumentos.

No nível sintático, estabelecem-se as categorias sintagmáticas dos argumentos do predicador. Em outras palavras, pode-se dizer que, entendidos como predicadores, os adjetivos selecionam a categoria sintagmática dos constituintes com as quais podem ou não podem ocorrer. A esse fenômeno, pode ser dada a denominação **esquema** ou **quadro de subcategorização** (CHOMSKY, 1970; CHOMSKY, LASNIK, 1977; RAPOSO, 1992; MIRA MATEUS et al., 1994) ou valência morfossintática (ou sintática) (ROCHA, 1994; BORBA, 1996).

(1) Opcionalidade vs obrigatoriedade

Em decorrência de se considerar o constituinte na função de sujeito como argumento, observa-se que os adjetivos valenciais sempre terão um argumento subcategorizado *obrigatório*, o sujeito (BRESNAN, 1982; GAZDAR et al., 1985; POLLARD, SAG, 1994). Com exceção do argumento em função de sujeito, os demais argumentos dos adjetivos são *opcionais ou facultativos*, ou seja, podem ou não ocorrer sintaticamente. Por exemplo, em garoto *obediente* (a alguém), é fácil perceber o sintagma preposicional *a-SN* omitido. Quando os argumentos são omitidos, verifica-se que a significação intrínseca do adjetivo é expressa e os argumentos ausentes são interpretados de modo "subentendido". A seguir, estão mais alguns exemplos.

(28) a. [...] o seqüestrado (A1) estava *pálido* quando chegou na Polícia [...] (JO-FSP)

b. Os Edomitas (A1) eram *descendentes* de Esaú (A2) (JO-FSP)

c. [...] esse fator (A2) era *acusável* de deformação eleitoral (A3) [por X] (A1) [...]. (JO-FSP)

d. A energia espiritual (A2) das flores era *transferível* [de X] (A3) para a água (A4) [por Y] (A1) (JO-FSP)

Nos exemplos em (28c,d), vê-se que nem todos os argumentos realizam-se sintaticamente. No entanto, os argumentos elípticos podem ser recuperados contextualmente. Por exemplo, em (28c), o adjetivo *acusável* projeta três argumentos lógico-semânticos, mas somente estão sintaticamente realizados: esse fator e de deformação eleitoral. Em (28d), o adjetivo *transportável*, por sua vez, projeta quatro argumentos lógico-semânticos, mas somente dois ocorrem, a energia espiritual e para a água. Os outros dois, no entanto,

podem ser recuperados pelo contexto, p.ex.: a energia espiritual das flores era *transferível* [de X] para a água [por Y].

(2) Das relações sintáticas

Observa-se também pelos exemplos em (28) que a relação sintática que se estabelece entre os adjetivos e seus respectivos argumentos podem ser de três tipos: *subjetiva, objetiva* e *preposicional* (BORBA, 1996). Os argumentos que estão em relação subjetiva desempenham a função de *sujeito* (Suj) e os que estão em relação objetiva ou preposicional desempenham a função de *complemento* (C). Por exemplo, em (28b), o constituinte os Edomitas está em relação subjetiva com *descendente*, desempenhando, conseqüentemente, a função de *sujeito*. O constituinte de Esaú, por sua vez, está em relação preposicional, desempenhando a função sintática de *complemento*.

Para se identificar tais relações, recorre-se ao paralelismo existente entre o esquema de subcategorização projetado pelo adjetivo e o projetado pelo verbo ou substantivo morfologicamente associado. Por exemplo, (28c) relaciona-se a X acusa esse fator de deformação eleitoral. Assim, o sintagma recuperado pelo contexto, por X, está em relação subjetiva com *acusável*, o sintagma esse fator está em relação objetiva e de deformação eleitoral, em relação preposicional.

(3) Da realização sintática dos argumentos

Os adjetivos de valência 1, do ponto de vista sintático, podem ser denominados **intransitivos**, já que não apresentam argumento em função de complemento (p.ex.: O muro é *alto*). Os demais, de valência 2, 3 e 4, são denominados **transitivos**, pois apresentam um argumento em função de complemento, independentemente de subcategorizarem ou não outro(s) complemento(s) (p.ex.: esse fator era *acusável* de deformação [por X])

(MALING, 1983; PUSTEJOVSKY, 1996; SPENCER, 1997; NEWMEYER, 1998). A seguir, são feitas algumas observações sobre a realização sintática dos argumentos em função dos tipos valenciais.

(1) Adjetivos de Valência 1 (V1): em decorrência de se considerar o constituinte na função de sujeito como argumento dos adjetivos, observa-se que os adjetivos V1 sempre partirão de um *índice 1*. Esse A1 é o argumento que está em relação subjetiva com o adjetivo. Considerando-se que os valenciais são os adjetivos em posição predicativa, os adjetivos monovalentes podem apresentar um único argumento (A1), na forma de sintagma nominal (29a) ou oracional (29b), p.ex.:

(29) a. Matemática (A1 = SN) é *difícil*? (JO-FSP)

b. É *difícil* ser liberal em Portugal. (A1 = O) (JO-FSP)

(2) Adjetivos de Valência 2 (V2): esses adjetivos podem apresentar até dois argumentos lógico-semânticos realizados na sintaxe, o A1 e o A2. Em posição predicativa, o A1, ou seja, o *sujeito*, pode ocorrer na forma de sintagma oracional (30a), preposicional (30b) ou nominal (30c), p.ex.:

(30) a. Numa capital rica e próspera como a cidade do Rio de Janeiro, era *indispensável* [para X] que se fundasse um grêmio (A1 = O) [...]. (JO-FSP)

b. O inconsciente (A2) é *incontrolável* pelo consciente. (A1 = SPrep) (JO-FSP)

c. O jovem (A1 = SN) estava *receoso* da descoberta (A2). (JO-FSP)

O adjetivo *indispensável* projeta dois argumentos lógico-semânticos; no entanto, na sentença em (30a), apenas um argumento ocorre, o sujeito, que está sob a forma de sintagma oracional, no caso, que se fundasse um grêmio. O A2 elíptico pode ser recuperado pelo contexto, p.ex.: [...] era *indispensável* [para X] que se fundasse um grêmio. Em (30b), o argumento que está em relação subjetiva com o adjetivo *incontrolável* realiza-se na forma de sintagma preposicional: *pelo consciente*. Já em (30c), o A1 está sob a forma de um sintagma nominal, no caso, *o jovem*. Quanto aos demais argumentos, observa-se que, em (30b), o A2 (o inconsciente) está em relação objetiva com *incontrolável*, pois a sentença em (30b) relaciona-se a "o consciente não controla o inconsciente"; nesse caso, diz-se que o A2 é o complemento (C) do adjetivo. Em (30c), o A2 (C), realizado sintaticamente na forma de sintagma preposicional (da descoberta), está em relação objetiva com o adjetivo *receoso*, desempenhando função de complemento.

Dessa forma, observa-se que em relação objetiva, o argumento pode realizar-se como sintagma nominal, como em (30b), ou preposicional, como em (30c).

(3) **Adjetivos de Valência 3 (V3):** esse tipo de adjetivo pode projetar o argumento externo sob o forma de sintagma nominal e sintagma preposicional, p.ex.:

(31) a. O rapaz (A1 = SN) era *doador* de órgãos (A2) para transplante (A3). (JO-FSP)

b. O réu (A2) é *condenável* à pena de morte (A3) pelo juiz (A1 = SPrep). (LI-LT)

Em (31a), o A1 realiza-se sintaticamente sob a forma de sintagma nominal (o rapaz); em (31b), sob a forma de sintagma preposicional (pelo juiz). Em (31a), o A2 ou C1 do adjetivo *doador* (de órgãos) está em relação objetiva com o adjetivo em questão. Já o A3 ou C2 está em relação preposicional com *doador*. Em (31b), o A2 (C1) (o réu) está em relação

subjetiva com o adjetivo *condenável,* pois a sentença em (31b) relaciona-se a "o juiz condena o réu à pena de morte"; e o A3 (C2) (à pena de morte) está em relação preposicional.

(4) Adjetivos de Valência 4 (V4): com relação aos adjetivos que projetam quatro argumentos lógico-semânticos, observa-se que o argumento na função sintática de sujeito, o A1, realiza-se sintaticamente na forma de sintagma preposicional (Sprep = [por-SN]). A seguir, estão dois exemplos extraídos de Rocha (1994)[37].

(32) a. O texto (A2) *traduzido* do inglês (A3) para o português (A4) por Gina (A1 = Sprep) apresentava alguns erros.

b. A funcionária (A2) *transferida* do almoxarifado (A3) para a chefia (A4) pelo Sr. Antunes (A1 = Sprep) obteve êxito.

Observa-se que, em (32a), o A2 está em relação objetiva com o adjetivo, pois a sentença em (32a) equivale a Gina traduziu o texto do inglês para o português. Já o A3 e o A4 estão em relação preposicional. O mesmo pode ser observado para o exemplo em (32b).

Com base nos dados sobre a realização sintática da valência adjetival, foi possível estabelecer alguns padrões de subcategorização para os adjetivos valenciais, também conhecidos por *frames sintáticos*, os quais estão descritos no Quadro 5:

[37] Tendo em vista que os adjetivos valenciais foram definidos como sendo aqueles em função sintática predicativa, cabe ressaltar que se considera a existência de uma predicação implícita nos exemplos em (32) e (33b).

Tipo	Padrões de subcategorização e exemplos
Adjetivo V_1	(a) (A1: SN) + adj P.ex.: *O rapaz é bonito*. (b) (A1: O) + adj P.ex.: *Cantar é maravilhoso*.
Adjetivo V_2	(a) (A1: SN) + adj + (A2: SPrep [com-SN]) P.ex.: *A velha estava desgostosa com a vida*. (b) (A1: SN) + adj + (A2: SPrep [de-SN]) P.ex.: *O rapaz está temeroso da descoberta*. (c) (A1: SN) + adj + (A2: SO[38] [de-O]) P.ex.: *O rapaz está desejoso de que você volte*. (d) (A1: SO) + adj + (A2: SPrep [para-SN]) P.ex.: *Cantar era maravilhoso para fulano*. (e) (A2: SN) + adj + (A1: SPrep [por-SN]) P.ex.: *O rapaz era incontrolável pelo juiz.* (f) (A1: SN) + adj + (A2: SPrep [sobre-SN]) P.ex.: *O professor não estava bem informado sobre a vida dos gatos.*
Adjetivo V_3	(a) (A1: SN) + adj + (A2: de-SN) + (A3: para/a-SN) P.ex.: *Fulano era doador de órgãos para transplantes*. (b) (A2: SN) + adj + (A3: a-SN) + (A1: por-SN) P.ex.: *O réu é condenável à pena de morte pelo juiz.*
Adjetivo V_4	(a) (A2: SN) + adj + (A3: de-SN) + (A4: para-SN) + (A1: por-SN) P.ex.: *A funcionária transferida do almoxarifado para a chefia pelo Sr. Antunes* [...].

Quadro 5: Padrões de subcategorização dos adjetivos valenciais

Analisando a complementação sintática dos adjetivos, observa-se que os complementos adjetivais são introduzidos por preposições. Uma possibilidade de explicação para tal fato relaciona-se a Teria do Caso, mais especificamente, ao Filtro de Caso[39] proposto pela Teoria da Regência e Ligação (*Government-Binding Theory* – GB) (CHOMSKY, 1981).

[38] SO = sintagma oracional.

[39] O Filtro de Caso pode ser concebido como uma condição de licenciamento morfofonológica sobre os sintagmas determinantes foneticamente realizados. De acordo com esse filtro, um *sintagma determinante* foneticamente realizado manifesta um Caso abstrato, sendo o qual excluído pela gramática. A violação do Filtro de Caso é suficiente para marcar uma frase como não gramatical (RAPOSO, 1992, p. 350-2). Os

Esse Filtro de Caso estabelece que um sintagma determinante (SD) foneticamente realizado, mas sem a marca de Caso é agramatical. Dessa forma, os sintagmas determinantes devem ser marcados com Caso (p.ex.: nominativo, se sujeito; acusativo, se objeto direto; etc.). A atribuição de Caso, no entanto, é feita por classes de expressões específicas, por exemplo, os verbos e as preposições. O Caso nominativo é atribuído pela categoria *concordância* (*agreement* em inglês) Infl/[+Agr] (RAPOSO, 1992). Por exemplo, na sentença Os deputados saír*am*, o DP *os deputados* recebe Caso nominativo da flexão -am (Infl/[+Agr]). Já em sentenças do tipo Leonardo está apaixonado Maria, *Maria* não tem como receber Caso porque *apaixonado* não atribui Caso. Assim, a sentença só pode ser "salva" pela introdução de um elemento atribuidor de Caso: a preposição "por", que atribui um caso obliquo ao SD *Maria* (CHIERCHIA, 2003). Por essa razão é que os complementos dos adjetivos são sempre introduzidos por preposições.

Chomsky (1986), no entanto, propôs revisões na Teoria do Caso. O autor sugere que os nomes (substantivos) e os adjetivos também são atribuidores Casuais, ambos, inclusive, atribuindo Caso genitivo, que é realizado superficialmente por meio da preposição *de*. Nessa visão da Teoria do Caso, Chomsky faz uma distinção entre dois tipos de Casos; por um lado, o Caso genitivo atribuído pelas categorias N (nome) e A (adjetvo) e o Caso oblíquo atribuído pela categoria P; por outro lado, o Caso nominativo atribuído por Infl/[+Agr] e o Caso acusativo atribuído por V. Ao primeiro tipo, o autor dá a denominação de Caso *inerente*, ao segundo, de Caso *estrutural*.

Segundo Raposo (1992), a principal diferença entre esses dois tipos de Casos diz respeito ao nível em que o Caso é atribuído: enquanto o Caso estrutural é atribuído em estrutura superficial, o Caso inerente é atribuído em estrutura profunda. Essa distinção tem reflexos na estrutura temática da categoria que atribui o Caso: enquanto um Caso inerente

---

sintagmas nominais são, na verdade, *grupos de determinantes* (DPs, do inglês, "determiner phrase"), ou seja, são projeções da categoria D e não da categoria N (substantivo) (STOWEL, 1981).

encontra-se ligado a uma função-θ particular na estrutura temática da categoria atribuidora e é atribuído em estrutura profunda juntamente com essa função-θ, o Caso estrutural, por outro lado, não está necessariamente associado a uma função-θ, podendo, inclusive, ser atribuído por categorias que não atribuem função-θ (p.ex.: a categoria Infl/[+Agr]).

Ainda segundo Raposo (1992), a Caso inerente particularmente atribuído por N e A, o Caso genitivo, é realizado em estrutura profunda por meio da preposição *de*. Dessa forma, é possível compreender o que acontece com sentenças do tipo Estou orgulhoso de você ou O valor deduzido da conta bancária [pelo correntista] era irrisório. Nesses exemplos, a preposição introduz o complemento do adjetivo para satisfazer o Filtro de Caso, atribuindo, conseqüentemente, o Caso inerente genitivo e uma função-θ ao complemento. O adjetivo *deduzido*, por exemplo, atribui ao seu complemento a função-θ de Origem. No entanto, o Caso inerente atribuído pelos As também pode ser marcado por preposições específicas, que juntamente com os As (já que elas parecem ser selecionadas pelos As para a introdução dos complementos), atribuem determinado papel temático (ou função-θ) ao SD. Por exemplo: na sentença *A velha estava desgostosa com a vida*, a preposição *com*, juntamente com *desgostosa*, atribuem o papel temático Ca(usativo) ao SD a vida.

Um outro modo de explicar a característica dos adjetivos de terem seus complementos introduzidos por preposições também se relaciona com o Filtro de Caso. Nessa outra hipótese, o Caso é considerado como um traço da categoria T(empo) (PESETSKY; TORREGO, 2002).

Pesetsky e Torrego (2002) partem da hipótese de que o núcleo (funcional) de uma sentença é a categoria T(empo) e que a diferença de complementação entre os predicadores verbais e adjetivais reflete distinções quanto à categoria T. Para esses autores, a frase realiza duas instâncias da categoria T, denominadas de T_s e T_o. O T_s seria o tempo do *sujeito* e o T_o seria o tempo do objeto. A predicação verbal instancia esses dois aspectos da categoria T porque é característico desse predicador poder expressar dois subeventos temporalmente

distintos. Por exemplo, o verbo *ler* expressa dois subeventos distintos. O primeiro subevento é um *processo*, mais especificamente, um predicado com um argumento Agente, e o segundo subevento é a complementação do processo, ou seja, um predicado com um elemento adicional, o objeto lido, o evento télico. O tempo do primeiro subevento seria o T_s e o tempo do segundo subevento seria o T_o. Segundo esses autores, os predicadores adjetivais não expressam subeventos temporalmente distintos (PESETSKY; TORREGO, 2002). Por conseguinte, os predicadores adjetivais se caracterizam por não instanciarem tempo algum, já que o T_s é determinado pelo verbo copulativo. Por isso que a preposição é necessária como introdutora do complemento para expressar o T_o. De fato, segundo Pesetsky e Torrego (2002), a preposição deveria ser categorizada como um dos atributos da categoria T.

3.7.4.3. Da valência semântica **ou** *estrutura de argumentos*

Como mencionado, o adjetivo valencial estabelece relações semânticas com seus argumentos, ou seja, o sentido ou valor semântico de um adjetivo valencial é determinado pela combinatória (semântica) estabelecida entre ele e seus argumentos (PUSTEJOVSKY, 1996). Para efeitos práticos, a literatura tem demonstrado serem suficientes, para a descrição da estrutura de argumentos ou valência, os seguintes papéis temáticos: *agente* (Ag)*, experimentador* (Ex)*, beneficiário* (B)*, objetivo* (Ob)*, locativo* (L)*, instrumental* (I)*, causativo* (Ca)*, meta* (M)*, origem* (Or)*, resultativo* (R)*, temporal* (Tp) *e comitativo* (Co) (FILLMORE, 1968; BORBA, 1996).

No caso dos adjetivos monovalentes, o significado adjetival é determinado pela relação semântica estabelecida entre o adjetivo e seu único argumento. Por exemplo: na sentença O garoto é *esperto*, em que *esperto* tem sentido de "astuto", pode-se associar o papel temático Objetivo (indica a entidade à qual se verifica uma situação) ao argumento

realizado sintaticamente por meio do sintagma nominal o garoto. Nesse caso, no entanto, o argumento que recebe o papel temático Objetivo necessariamente possui o traço +*humano*; esse traço, inclusive, pode ser visto como a *restrição selecional* imposta pelo adjetivo ao argumento. Na sentença A água estava *esperta*, em que *esperto* tem sentido de "estimulante", associa-se o mesmo papel temático ao argumento (Objetivo); nesse caso, no entanto, o argumento que recebe o papel temático Objetivo possui o traço -*animado*. Resumindo, o adjetivo esperto, quando em combinação com um argumento de traço +humano, tem sentido de astuto, finório, matreiro (p.ex.: Este garoto (A) é *esperto*); quando em combinação com argumento de traço –animado, *esperto* tem o sentido de vibrante, estimulante (p.ex.:. A água (A) estava *esperta*).

No caso dos adjetivos de valência 2, 3 e 4, o estatuto semântico (= sentido) é circunscrito na relação (semântica) estabelecida entre os adjetivos e os seus vários argumentos. A seguir, estão alguns exemplos.

(33) a. O som alto (A1) era *desagradável* a todos (A2). (JO-FSP)

b. O funcionário (A2) *transferido* do almoxarifado (A3) para a chefia (A4) pelo Sr. Antunes (A1) obteve êxito. (ROCHA, 1994)

Em (33a), o argumento A1, realizado sintaticamente sob a forma do sintagma nominal o som alto, recebe o papel temático Causativo, pois a sentença em (33a) relaciona-se a "o som alto causa desagrado a todos". Já o sintagma preposicional a todos recebe o papel temático Experienciador, pois, segundo Borba (1996), esse papel temático traduz uma experiência ou uma disposição mental. Em (33b), o A1, introduzido pela preposição *por,* recebe o papel temático Agente, pois a sentença em (33b) equivale a "Sr. Antunes transferiu o funcionário do almoxarifado para a chefia". O A2, que está em relação

objetiva com *transferido*, recebe o papel temático Objetivo, que pode ser entendido como a entidade a que se verifica uma situação. Já os argumentos realizados sintaticamente sob a forma dos sintagmas preposicionais do almoxarifado e para a chefia recebem, respectivamente, os papeis temáticos Origem (ponto de partida) e Meta (ponto de chegada).

(1) Da complementação semântica

Como mencionado, a preposição introdutora de complemento colabora com o adjetivo predicador na atribuição de um papel temático ao argumento por ela introduzido (PESETSKY; TORREGO, 2002; CASTILHO, 2003).

É esse o caso, por exemplo, da preposição *a* em (33a), que se combina com o adjetivo *desagradável* para atribuir o papel temático Experienciador ao SN todos, e das preposições *de*, *para* e *por* em (33b), que se combinam com o adjetivo *transferido* para atribuir os papéis temáticos Origem, Meta e Agente aos SNs almoxarifado, chefia e Sr. Antunes, respectivamente. Observa-se, agora, a preposição *de* em (34), quando introduz o complemento de adjetivos que expressam o sentido "subtração" (NEVES, 2000).

(34) a. O valor da doação *deduzido* do imposto de renda [...] (JO-FSP)

b. O valor de R$ 90,00 *abatido* do salário bruto [...] (JO-JB)

Ao introduzir complemento de adjetivos que expressam "subtração", a preposição *de* combina-se com o adjetivo (no caso, *deduzido* e *abatido*) para atribuir o papel temático Origem (Or) ao argumento por ela introduzido. No caso do adjetivo *deduzido*, cabe ressaltar que as ocorrências retiradas dos córpus revelaram que esse item apresenta, além do sentido "subtração", um outro sentido, o qual está ilustrado em (35).

(35) a. Estimativa – representa um valor de uma característica do universo, *deduzido* da amostra. (NP)

Em (35), o complemento adjetival também recebe o papel temático Or, no entanto, o sentido de *deduzido* é de "inferência". A distinção entre os sentidos de "subtração" e "inferência" pode ser traduzida em termos das **restrições de seleção** que o adjetivo valencial impõe aos seus complementos. Em (34), imposto (de renda) conforma-se ao traço semântico [+(origem da) subtração] e, em (35), amostra conforma-se ao traço semântico [+(origem da) inferência]. Assim, *deduzido* e *abatido*, em (34), têm sentido de "subtração", e *deduzido*, em (35), de "inferência". Para explicitar, na estrutura de argumentos ou valência semântica, os dois sentidos de *deduzido*, recorre-se ao expediente formal das **restrições de seleção** (36):

(36) **deduzido$_1$**: [(x1)$_{\text{Objetivo}}$ (x2: <*subtração*>)$_{\text{Origem}}$] (= subtraído)

deduzido$_2$: [(x1)$_{\text{Objetivo}}$ (x2: <*inferência*>)$_{\text{Origem}}$] (= inferido)

3.8. Da delimitação dos adjetivos valenciais e de suas propriedades: uma síntese

Da exposição sobre os adjetivos do português, foi possível delimitar os *valenciais* e identificar suas principais propriedades. Com relação à delimitação dos valenciais, cabe ressaltar que, neste trabalho, considera-se valencial o adjetivo em *função predicativa*, entendida aqui como um misto de sintaxe e semântica. Sintaticamente, a noção de predicação refere-se à ocorrência do adjetivo como núcleo do sintagma verbal. Semanticamente, "predicação" significa que o adjetivo atribui uma propriedade ao substantivo.

No que diz respeito às propriedades dos valenciais, vê-se que, do ponto de vista morfológico, o adjetivo tem o poder, na maioria das vezes, de apresentar duas terminações de *gênero* e *número*, resultantes de regras de concordância. Do ponto de vista sintático, os adjetivos são *predicativos* e selecionam as categorias sintagmáticas de seus constituintes, ou seja, dos argumentos subcategorizados. E, do ponto de vista (sintático-)semântico, os adjetivos valenciais projetam *estrutura de argumentos*. Diz-se que os argumentos projetados ou selecionados pelos adjetivos estabelecem com eles determinadas relações semânticas, as quais são descritas por meio de *papéis temáticos*. Além disso, o adjetivo pode projetar *restrições selecionais* aos seus argumentos.

Identificadas as propriedades dos adjetivos valenciais nos diferentes níveis de análise lingüística, prosseguiu-se com a projeção das mesmas no modelo de representação da informação lexical proposto na Seção 2.4. Esse modelo ou esquema de representação da informação lexical, elaborado com base no modelo cognitivo de processamento da linguagem difundido por Bock (1982), Levelt (1992; 1993); Bierwisch e Schreuder (1992) e Lowie (1998), estabelece que uma **unidade lexical** seja representada em um léxico lingüístico-computacional em função de quatro componentes: *forma gráfica, traços gramaticais, estrutura de argumentos* e *forma semântica*.

De acordo com as observações feitas na Seção 2.4.1, os três primeiros componentes sofreram algumas alterações e a *forma semântica*, inclusive, foi excluída da representação dos adjetivos valenciais. Dessa forma, para os fins deste trabalho, foram considerados apenas os componentes *forma gráfica, traços gramaticais* e *estrutura de argumentos*, sendo que: o componente FG especifica a forma gráfica (ou ortográfica) do item; o TG especifica suas propriedades sintáticas; o EA especifica o número e as funções semânticas dos argumentos exigidos por ele. O componente EA, aliás, é visto como uma estrutura de interface entre a semântica e a sintaxe.

Capítulo 4 Da representação lingüístico-computacional dos adjetivos valenciais

4.1. Dos formalismos

De acordo com a metodologia proposta por Dias-da-Silva (1996; 1998), as informações lingüísticas identificadas no *domínio lingüístico* são projetadas, no domínio representacional, em modelos formais de representação.

O equacionamento do *domínio representacional*, situado entre o levantamento das informações lingüísticas e a implementação do sistema de PLN, é estrategicamente positivo por duas razões (DIAS-DA-SILVA, 1996): (i) as representações formais contêm todas as informações necessárias para a construção do sistema de PLN; (ii) tais representações, por não estarem diretamente vinculadas a qualquer linguagem de programação, garantem transportabilidade das informações a diferentes linguagens de programação e algoritmos.

Vale ressaltar que a representação das informações lexicais precisa ser necessariamente explícita, consistente e, principalmente, não-ambígua, para que possa ser transformada em programas computacionais. Para isso, nada mais natural que se recorra a uma série de formalismos, dentre os quais, citam-se: **matriz de atributo-valor, lógica de predicados e *frame*** (ROSNER, JOHNSON, 1992; HANDKE, 1995; DIAS-DA-SILVA, 1996). Tais formalismos, entendidos aqui como sistemas abstratos, têm sido amplamente empregados para representar informações léxico-gramaticais. A seguir, descreve-se com mais detalhes cada um desses formalismos. O objetivo final da descrição apresentada nesta seção é propor a representação lingüístico-computacional das propriedades dos adjetivos valenciais.

4.1.1. Da *estrutura de traços* ou *matriz de atributo-valor*

Grosso modo, as estruturas de traços consistem na "quebra" de um dado lingüístico em partes menores, os chamados *atributos* ou *traços*. A cada traço ou atributo, associam-se *valores*. Por exemplo: assume-se que a *concordância* (CONC) *singular masculino* (*F sg_masc*), apresenta dois atributos: NUM(ero) e GEN(ero), sendo que os respectivos valores desses atributos são: *sg* (singular) e *masc* (masculino) (SHIEBER, 1986). Por meio de uma **matriz de atributo-valor** (MAV) (do inglês, "*attribute-value matrix*"), forma em que comumente são apresentadas as estruturas de traços (KOENIG, 1999), representa-se a estrutura de traços denominada *F sg_masc* (37).

(37) *F sg_masc* = $\left[\begin{array}{l}\text{CONC} \quad \left[\begin{array}{l}\text{NUM } \langle sg \rangle \\ \text{GEN } \langle masc \rangle\end{array}\right]\end{array}\right]$

Para fins de representação, ressalta-se que as estruturas de traços podem ser *simples* ou *complexas*. Uma estrutura de traços é simples quando composta por apenas um valor terminal, p.ex.: em (37), os traços NUM e GEN têm um único valor terminal: *sg* e *masc*, respectivamente. Uma estrutura é complexa quando contêm outras estruturas de traços. Um exemplo paradigmático de estrutura de traços complexa é a descrita em (37), pois a estrutura identificada como *F sg_masc* contém outras estruturas, no caso, CONC, NUM e GEN.

Uma das principais características das *estruturas de traços* é poder combinar informações; a operação ou mecanismo responsável por essa combinação ou união é a **Unificação** (SHIEBER, 1986).

Esse mecanismo é responsável por "unificar" estrutura de traços. Em outras palavras, entende-se que, dadas duas estruturas de traços ou matrizes de atributo-valor MAV1 e MAV2,

escolhe-se, por exemplo, MAV1 e, para cada atributo A de MAV1, procura-se uma instância de A em MAV2. Se A não ocorrer em MAV2, acrescenta-se o atributo A (de MAV1) e seu respectivo valor V em MAV2. Caso contrário, se A ocorrer em MAV2, e se o valor de V em MAV2 for V', então a unificação de V com V' passará a ser o novo valor de A em MAV2. Por exemplo: dadas as estruturas [NUM <*sg*>] e [PES <*3*>], a unificação dessas duas estruturas produz como resultado a estrutura em (38):

(38) $\begin{bmatrix} \text{NUM} <sg> \\ \text{PES} <3> \end{bmatrix}$

O mesmo já não acontece com as estruturas [NUM <*sg*>] e [NUM <*pl*>], pois os valores *sg* e *pl* do atributo NUM são distintos. Em outras palavras, a operação de unificação não se realiza quando a um mesmo atributo, especificado nas estruturas de traços a serem unificadas, são associados valores distintos como *sg* e *pl*.

A Figura 6 ilustra, de modo simplificado, as *estruturas de traços* representadas em uma *matriz de atributo-valor*. Mais especificamente, estão representadas as informações léxico-gramaticais de *comprou* em uma frase como João comprou um carro, segundo os princípios da HPSG (POLLARD, SAG, 1994).

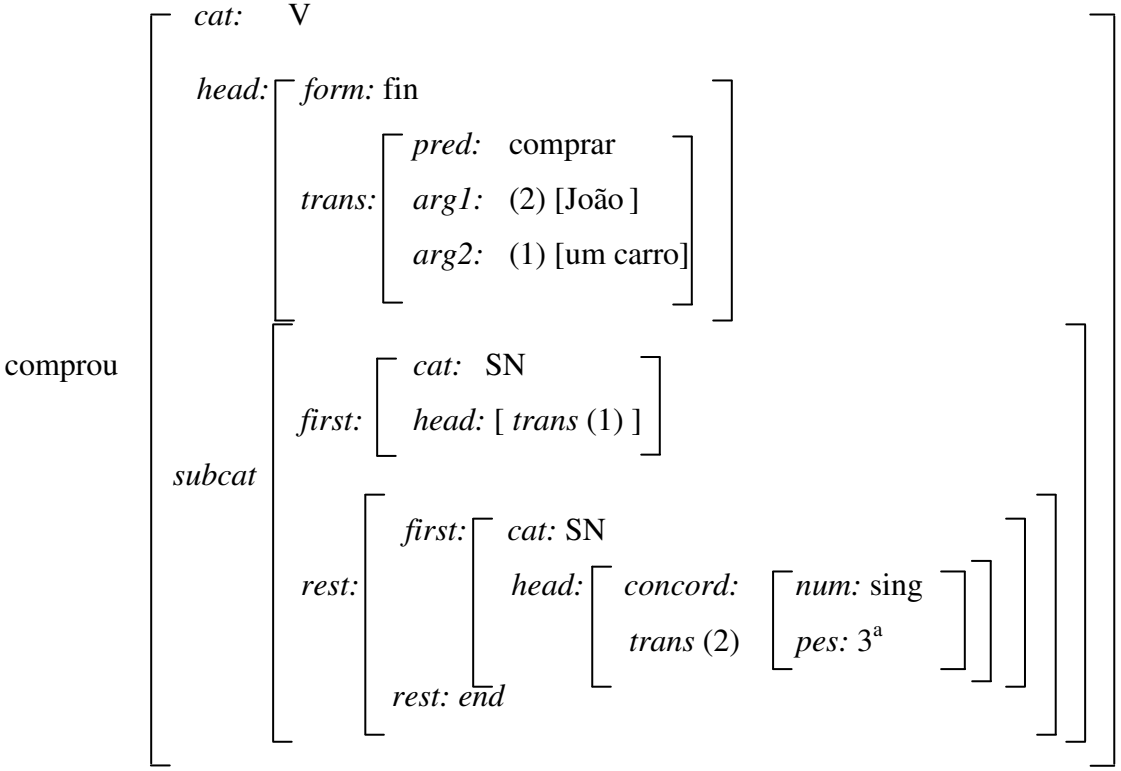


Figura 6: Representação em notação MAV, descrita com base na HPSG.

Com base na Figura 6, pode-se dizer que o verbo *comprou* está descrito por meio de uma estrutura de traços complexa, formada pelas estruturas **cat**, **head** e **subcat**. O valor da estrutura (ou traço) *cat* especifica a classe gramatical do item, no caso, V(erbo). O traço *head*, que especifica os aspectos morfossintáticos e a codificação da forma lógica do item, também é uma estrutura complexa, formada pelos traços **form** e **trans**. O valor do traço *form* indica que o item está na forma flexionada. A estrutura **trans**, por sua vez, codifica a forma lógica do item, ou seja, (i) o tipo do predicado e (ii) o número e tipo dos argumentos. A estrutura de traços **subcat** especifica o esquema de subcategorização item. Tal estrutura é composta por outras duas estruturas: **first** e **rest**, sendo que este último pode ser recursivamente especificado pelos traços *first* e *rest* até que se esgotem os argumentos do item. O primeiro traço *first* (de fora para dentro) especifica o argumento que ocorre na posição pós-verbal (no caso, um carro). A estrutura de traços *rest*, que especifica a argumento que ocorre em posição pré-verbal, no caso João, é formada pelos traços *first* e *rest*. Esse segundo traço *first*, em

específico, é formado pelos traços ***cat*** e ***head***. O valor do traço *cat* especifica a categoria sintagmática do argumento pré-verbal (no caso, SN = João). Já o traço *head* especifica a concordância verbal entre o argumento pré-verbal e o verbo.

Os valores dos traços *num* e *pes* do *arg1* (João) são herdados da entrada lexical do item João e, devem ser, respectivamente, *singular* e *3ª.* O segundo traço *rest* tem um valor atômico **end,** que indica o término dos argumentos.

4.1.2. Da lógica de predicados

Na Gramática Funcional (DIK, 1997), todo predicador está associado, no léxico, a um esquema denominado *estrutura de predicados* (cf. 2.4.1.3). A representação formal dessa *estrutura* baseia-se na *lógica de predicado* (SIEWIERSKA, 1991).

Basicamente, a **lógica de predicado** é um formalismo que permite construir a representação de mundos em termos de *objetos* e *propriedades* dos objetos. A fórmula básica de representação da lógica de predicado é *F(x)*, em que *F* representa uma propriedade qualquer (ou seja, uma variável) e *x* representa um objeto qualquer (isto é, uma variável). Comumente, na representação da lógica de predicados, o símbolo em "maiúsculo" (F) representa o *predicado* e o símbolo em "minúsculo e itálico" representa o *argumento*. De acordo com o número de argumentos, o predicado pode ser (ALLWOOD, et al., 1977):

(i) $F(x_1)$ = predicado de um lugar;

(ii) $F(x_1, x_2)$ = predicado de dois lugares;

(iii) $F(x_1, x_2, x_3)$ = predicado de três lugares;

(iv) $F(x_1, x_2, x_3, x_4)$ = predicado de quatro lugares;

(v) $F(x_1, x_2 \ldots x_n)$ = predicado de n-lugares.

A seguir, ilustra-se, com base em Dik (1997), a sentença *João comprou um carro* por meio da *lógica de predicados*:

(39) **comprou**$_{v}$ (x_1: **<***anim***>** (x_1))$_{Ag}$ (x_2)$_{Ob}$)

Do ponto de vista do formalismo empregado, ou seja, da lógica de predicados, diz-se que o verbo *comprar*, em sentenças como *João comprou um carro*, é um *predicado de dois lugares*, apresentando, conseqüentemente, dois argumentos, os quais são representados pelas variáveis $\mathbf{x}_1$ e $\mathbf{x}_2$. Observa-se que, no exemplo em (39), o argumento representado por $\mathbf{x}_1$ possui uma *restrição*, que está representada pelos símbolos "< >". No caso, o argumento $\mathbf{x}_1$ precisa ser do tipo +animado.

4.1.3. Dos *frames*

Originalmente proposta por Minsky (*apud* HANDKE, 1995), como uma base para a representação visual, a *teoria de frames* visa ao equacionamento da compreensão de diálogos das línguas naturais e outras atividades cognitivas igualmente complexas. A idéia básica de um *frame* é a integração de conceitos novos a conceitos adquiridos por experiência prévia. (HANDKE, 1995). Em termos formais, um *frame* é simplesmente uma estrutura formada por *atributos, valores* e *restrições* sobre os elementos que são os valores dos atributos.

A seguir, em (40), apresenta-se a entrada lexical do adjetivo ruim representada por um *frame,* tomando-se por base a proposta de Nirenburg *et al.* (1992).

```
(40)  (vermelho
          (vemelho-adj1
              (CAT adj)
              (ESTR-SINT
                  (1 ((raiz $var1)
                      (cat n)
                      (mods ((raiz $var0)))))
                  (2 ((raiz $var0)
                      (cat adj)
                      (suj ((raiz $var1)
                          (cat n))))))
              (ESTR-SEM
                  (PROJEÇÃO LEXICAL
                      ((1 2) (atributo-cor
                          (domínio (valor ^$var1)
                              (objeto-físico))
                          (escopo (valor vermelho))))))))
```

A entrada lexical em (40), pode ser descrita basicamente da seguinte maneira: o item lexical vermelho, que possui apenas um sentido (vemelho-adj1), tem como *atributos*: CAT(egoria), ESTR(utura) SINT(ática) e ESTR(utura) SEM(ântica). O atributo CAT tem como valor "adj". O atributo ESTR-SINT, em que estão representadas as propriedades sintáticas do item, tem dois valores: o primeiro especifica a construção N-Adj, ou seja, especifica a função de adjunto adnominal que o adjetivo vermelho pode desempenhar, e o segundo valor especifica a construção N-cópula-Adj, ou seja, especifica a função predicativa do adjetivo em questão. O atributo ESTR-SEM tem como valor outro atributo, PROJEÇÃO LEXICAL, que especifica a semântica lexical do item. No caso de vermelho, o atributo PROJEÇÃO LEXICAL tem como valores: o tipo de atributo ("cor"); o domínio ao qual ele é aplicável ou se associa ("objeto-físico") e o escopo do sentido expresso pelo adjetivo ("vermelho")[40].

[40] O *escopo* especifica o "valor" expresso pelo adjetivo dentro de uma escala. P.ex.: grande expressa um valor do atributo *graduável* TAMANHO; o escopo do adjetivo grande é de >0.75 na escala do atributo TAMANHO (NIRENBURG, 1992).

4.2. Do formalismo para a representação dos adjetivos valenciais

Dos formalismos brevemente descritos na seção anterior, ou seja, *estrutura de traços* ou *matriz atributo-valor, lógica de predicados, grafo* e *frames*, optou-se por empregar, na representação lingüístico-computacional dos adjetivos valenciais, a *estrutura de traços* ou *matriz atributo-valor*. Essa escolha pautou-se basicamente nos seguintes fatores:

(a) a utilização dessas estruturas em diversos modelos gramaticais para a representação das informações no léxico, tais como LFG (BRESNAN, 1982), GPSG (GAZDAR, 1985) e HPSG (POLLARD, SAG, 1987, 1994);

(b) a expressividade de tais estruturas ao descrever formalmente as informações lingüísticas, uma vez que essas informações podem ser descritas como uma *estrutura de dados*, no sentido computacional do termo (SHIEBER, 1986);

(c) a facilidade de se compreender e recuperar os relacionamentos estabelecidos entre, por exemplo, os itens lexicais e o conjunto de informações referente a cada item (KOENIG, 1999);

(d) o emprego da operação *unificação* na representação das estruturas de traços; mecanismo que torna mais econômica a tarefa de descrição dos dados lingüísticos (SHIEBER, 1986).

4.3. Da proposta de representação lingüístico-computacional

Partindo-se, por um lado, dos componentes previstos no esquema de representação da informação lexical proposto no Capítulo 2 e das propriedades dos adjetivos valenciais investigadas, e, por outro lado, da representação por meio de *matrizes de atributo-valor*, elaborou-se a representação lingüístico-computacional dos adjetivos valenciais, ou melhor, propôs-se um modelo básico de entrada lexical, o qual está ilustrado na Figura 7.

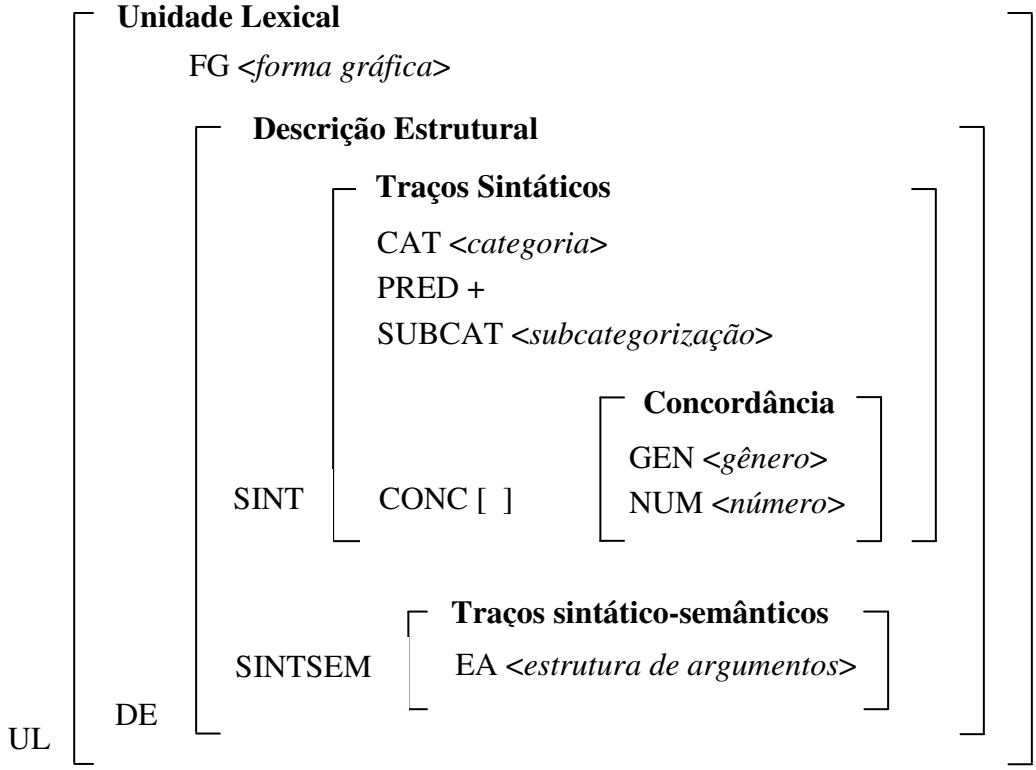


Figura 7: Modelo de entrada lexical para os adjetivos valenciais

Embora tenha sido feita uma descrição, ainda que breve, da representação por meio de *matrizes de atributo-valor,* julga-se conveniente fazer uma breve apresentação do sistema de notação empregado no modelo ilustrado na Figura 7.

Nota-se, primeiramente, que os *traços* ou *atributos* (ou *estrutura de traços*) são representados com letras maiúsculas (UL, DE, FG, SINT, CAT, PRED, SUBCAT, CONC, GEN, NUM, SINTSEM, EA) e na base inferior esquerda do colchete e, em segundo lugar, nota-se que os *valores* dos traços são representados com letras minúsculas e entre os símbolos "< >". A seguir, descreve-se mais detalhadamente cada informação contida nesse modelo, fazendo-se um paralelo com os componentes previstos pelo esquema de representação lexical proposto no Capítulo 2.

4.3.1. Das *estruturas de traços* que compõem o modelo de entrada lexical

A entrada lexical, ou seja, a *unidade de acesso* juntamente com a *especificação lexical* (cf. 2.2.1.2), foi denominada UL ("unidade lexical"). Uma UL, segundo o modelo proposto, é uma estrutura de traços complexa, formada pelas estruturas FG ("forma gráfica") e DE ("descrição estrutural"). A seguir, explicita-se o tipo de informação léxico-gramatical que cada uma dessas estruturas abrange.

4.3.1.1. Da estrutura de traço simples FG

O traço FG corresponde ao componente FG previsto pelo esquema de representação da informação lexical. De acordo com as restrições apresentadas na seção 2.4.1.1, o componente FG especifica apenas a *forma gráfica* (ou ortográfica) do item lexical. Não há, portanto, no modelo proposto, qualquer tipo de tratamento que pressuponha a representação fonológica das palavras. Além disso, observa-se que os adjetivos apresentam variações em sua forma gráfica, resultantes das flexões de *gênero* e *número* (p.ex.: alto, altos, alta, altas). Tais formas são consideradas, neste trabalho, realizações discursivas de um mesmo item lexical (p.ex.: ALTO) (cf. 2.2.1.2 (1)). Cabe ao traço FG, então, especificar a forma gráfica do item lexical (representado pela forma canônica, p.ex. <*alto*>). Dessa forma, o componente FG permite, além da identificação do item lexical, o acesso à informação lexical contida na entrada em questão.

4.3.1.2. Da estrutura de traços complexa DE

A estrutura de traços DE é formada por outras duas estruturas de traços complexas, a estrutura SINT ("sintática") e a estrutura SINTSEM ("sintático-semântica").

(1) Da estrutura SINT

A estrutura SINT abrange as informações previstas no componente TG do esquema de representação lexical (cf. 2.4.1.2). Mais especificamente, a estrutura de traços SINT é formada por outros quatro traços (ou estruturas): CAT, PRED, SUBCAT e CONC.

O traço CAT(egoria) especifica a *categoria gramatical primária* do item, ou seja, o valor do traço CAT é representado nesse modelo por meio da etiqueta *adjetivo*.

O traço PRED, que, na verdade, é um *valor* do traço ou estrutura de traços SINT, foi inserido, assim como sugerem Arnold e Sadler (1992), para indicar que o adjetivo em questão está em posição ou função predicativa e não modificadora, sendo que o valor do traço PRED é indicado por "+".

O traço SUBCAT especifica os argumentos sintaticamente realizados, ou melhor, especifica o *esquema* ou *quadro de subcategorização* dos adjetivos. Cabe ressaltar mais uma vez que o argumento em função sintática de sujeito é considerado argumento "subcategorizado".

O traço CONC, por sua vez, abrange as *categorias gramaticais secundárias* características do item lexical. As categorias secundárias de *gênero* (GEN) e de *número* (NUM) são atribuídas aos adjetivos por meio de regras de concordância, ou seja, o *gênero* e o *número* atribuídos aos adjetivos são, na verdade, do substantivo com o qual está ligado. Nesse traço, então, representa-se a concordância de número e de gênero entre o adjetivo e o substantivo; concordância essa responsável por atribuir tais categorias aos adjetivos. Ressalta-se que os valores dos traços NUM e GEN não herdados do substantivo pelo adjetivo; essa herança está representada no modelo de entrada por um índice numérico, que indica com qual dos argumentos subcategorizados o adjetivo está concordando.

(2) Da estrutura SINTSEM

A estrutura de traços SINTSEM, por sua vez, abrange as informações previstas pelo componente EA (cf. 2.1.4.3). A denominação SINTSEM é dada a essa estrutura por se entender que a *estrutura de argumentos* ou *valência* é uma estrutura de contato entre a sintaxe e a semântica. Cabe salientar que a estrutura de argumentos dos adjetivos está representada por meio de *papéis temáticos* e *restrições selecionais*. Além disso, faz-se uso de *índices* para "ligar" ou "associar" um papel temático a um determinado argumento subcategorizado. A exemplificam-se representações de alguns adjetivos valenciais segundo o modelo de entrada lexical proposto nesta seção.

4.4. Da aplicação do modelo: alguns exemplos de entradas lexicais

Para a montagem de uma entrada lexical, é preciso especificar as propriedades léxico-gramaticais do item em questão. Por exemplo, ao analisar o adjetivo valencial *pálido* em sentenças retiradas dos córpus de referência, verificou-se que ele pode subcategorizar dois argumentos, um sob a forma de SN e o outro sob a forma de SPrep [de SN]. Em (41), seguem alguns exemplos.

(41) a. Ramiro era gordo, *pálido*, bigodinho preto. (Q)

b. Mulder estava *pálido* de medo. (JO-FSP)

O adjetivo *pálido*, em (41a), tem sentido de "amarelado" e, em (41b), de "sem cor, branco". De acordo com as premissas deste trabalho (cf. 2.3.1.2), considera-se que *pálido*1 (41a) e *pálido*2 (41b) são itens lexicais distintos, os quais serão representados separadamente no léxico. A seguir, são ilustradas as entradas lexicais para *pálido*1 (Fig. 8) e *pálido*2 (10).

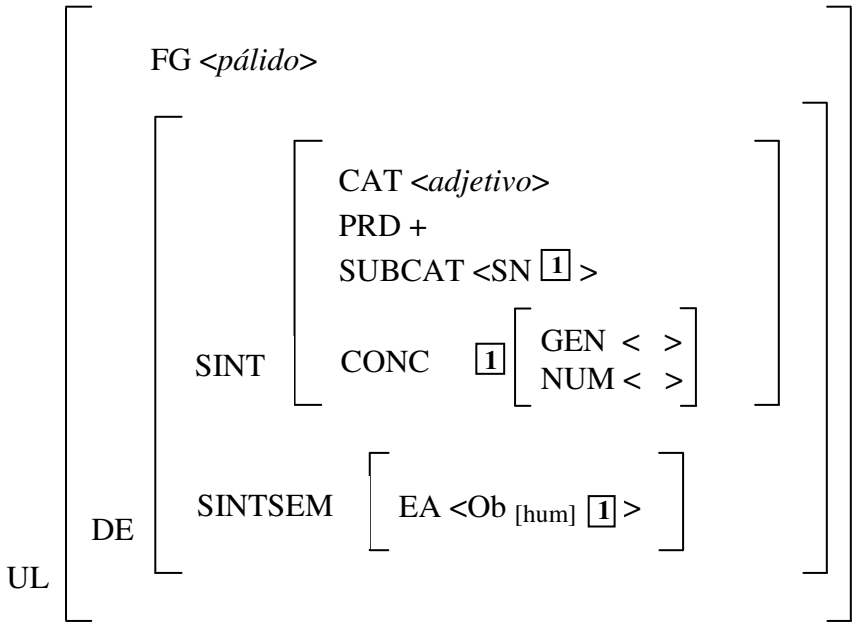


Figura 8: Proposta de representação do adjetivo *pálido*1

Interpretando a MAV da Figura 8, de fora para dentro, vê-se que a UL do item lexical *pálido*1 é representada por meio de duas estruturas principais: FG e DE. O traço FG especifica a forma gráfica do item lexical, no caso, *<pálido>.*

Quanto à estrutura de traços SINT, observa-se que nela estão especificadas: a categoria gramatical (traço CAT), a função (tipo ou posição) sintática (traço PRD), o esquema de subcategorização (traço SUBCAT) e a concordância entre adjetivo e substantivo (traço CONC).

O valor do traço CAT, como mencionado, é representado por meio da etiqueta <*adjetivo*>. No traço SUBCAT, especifica-se que *pálido* (no sentido de "amarelo") subcategoriza apenas um argumento na forma de SN, p.ex.: (28a) "[...] o seqüestrado estava *pálido* quando chegou na Polícia [...]"; (41a) Ramiro era gordo, *pálido*, bigodinho preto. Ao SN subcategorizado, associa-se o **índice [1]**, que é reescrito nos traços CONC e EA.

Em CONC, esse índice indica o argumento com o qual o adjetivo concorda em *gênero* e *número*. Ou melhor, indica o argumento do qual o adjetivo "herdará" os valores dos traços GEN e NUM, pois, no "léxico ou base lexical", os atributos GEN e NUM da entrada dos adjetivos são traços "à espera" de valores.

Para melhor entender o mecanismo de herança, faz-se necessário esclarecer a noção de *matching*. Informalmente, isso quer dizer que o sistema de PLN, ao acessar a "base lexical" ou "léxico" para analisar se uma seqüência é gramatical, verifica se as exigências especificadas na entrada do adjetivo são satisfeitas pelos constituintes com os quais o adjetivo ocorre na seqüência sob análise. Nesse momento, o adjetivo herda os valores dos traços GEN e NUM especificados na entrada do substantivo que é núcleo do SN subcategorizado pelo adjetivo. Por exemplo, na sentença "[...] o seqüestrado estava *pálido* quando chegou na Polícia [...]", o adjetivo herda os valores dos traços GEN e NUM da entrada de seqüestrado. Dessa forma, os traços GEN e NUM, que compõem a estrutura de traços CONC, herdam os valores *masc* e *sg*.

Tomando-se a sentença referida como exemplo, ilustra-se, na Figura 9, os valores dos traços GEN e NUM do núcleo do SN subcategorizado pelo adjetivo *pálido*1.

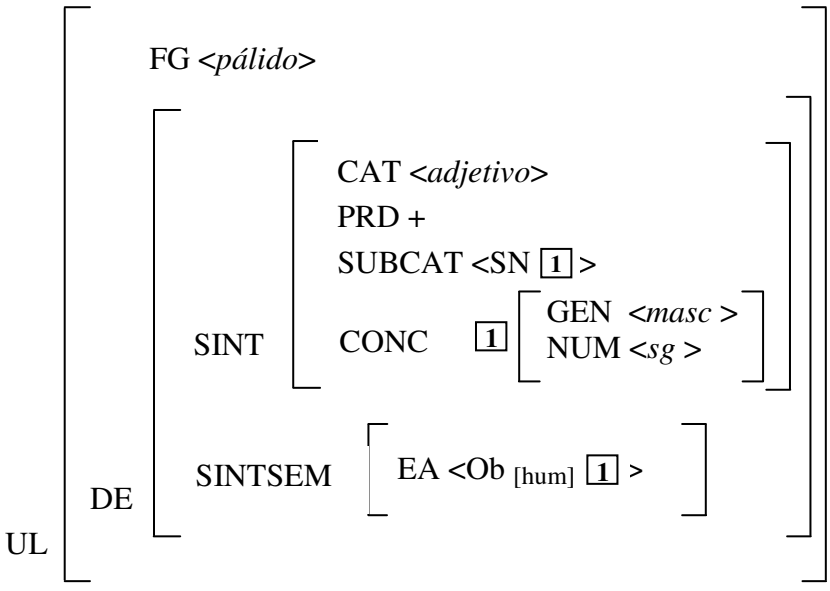


Figura 9: Exemplificação do traço CONC

A estrutura SINTSEM, como mencionado, é formada pelo traço EA, que especifica a *estrutura de argumento* ou *valência* do adjetivo. No caso de *pálido*1 (sentido de "amarelo"), a EA especifica que o único argumento projetado por esse adjetivo estabelece com ele uma relação que pode ser traduzida pelo papel temático Ob(jetivo). Além disso, esse argumento necessita ser do tipo +humano. No caso da sentença "[...] o seqüestrado estava *pálido* quando chegou na Polícia [...]", a restrição semântica é satisfeita, pois seqüestrado é do tipo +humano.

Dessa forma, observa-se que o índice [1], associado ao argumento no traço SUBCAT e presente no traço EA, funciona como "ligação" entre esses dois traços. Mais especificamente, o índice liga o argumento subcategorizado ao seu respectivo papel temático.

A seguir, na Figura 10, ilustra-se a entrada lexical para *pálido*2.

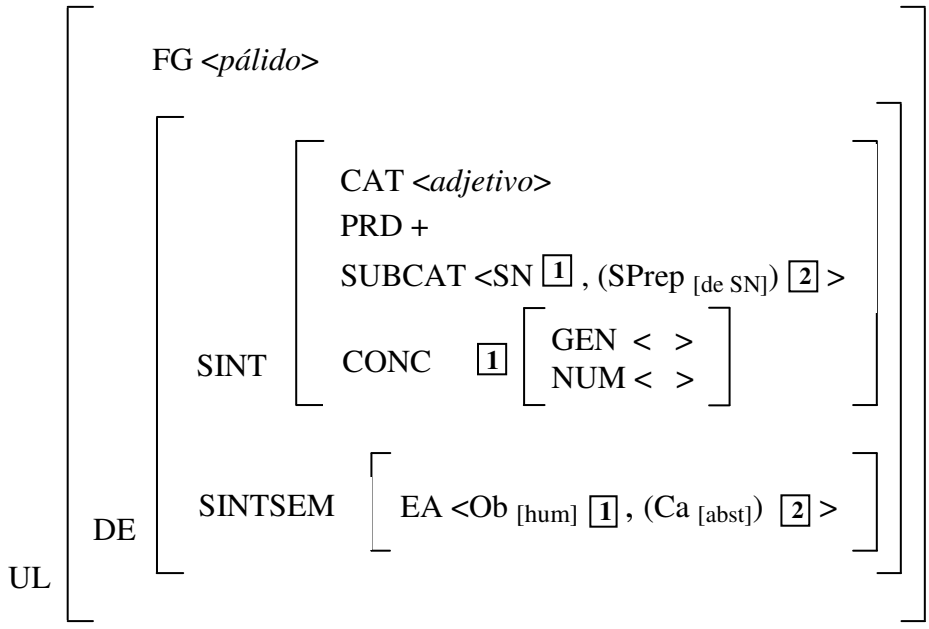


Figura 10: Representação do adjetivo *pálido*2

Diferentemente da entrada de *pálido*1, o traço SUBCAT da Figura 10 prevê que *pálido*2 (sentido de "sem cor, branco) pode subcategorizar um complemento facultativo na forma de SPrep [de SN]. A informação de que determinado argumento subcategorizado é

facultativo ou opcional está indicada pelos parênteses "( )". No caso de *pálido*2, tal argumento está associado ao papel temático Ca(usativo) e necesita ser do tipo +abstrato, p.ex.: (41b) Mulder estava *pálido* de medo. Assim, vê-se que *pálido*1 e *pálido*2 diferem quanto aos valores dos traços SUBCAT e EA.

A seguir, ilustra-se, na Figura 11, a entrada do adjetivo *descendente*.

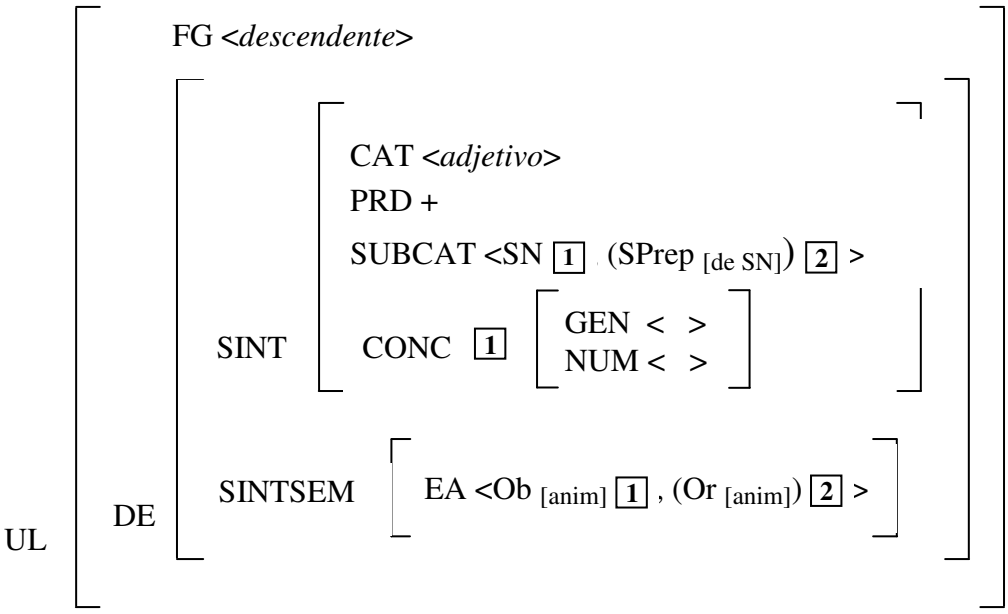


Figura 11: Representação da entrada do adjetivo *descendente*

Cabe salientar que o adjetivo *descendente* pode subcategorizar dois argumentos, um SN e um SPrep [de SN]. A ordem em que os argumentos são representados no traço SUBCAT é a ordem linear em que tais argumentos normalmente aparem nas sentenças, no caso, SN + adj + Sprep, p.ex.: (28b) "Os Edomitas eram *descendentes* de Esaú". Além disso, o argumento subtegorizado sob a forma de SN é obrigatório enquanto que o argumento que se realiza sob a forma SPrep [de SN] é facultativo.

Já a entrada do adjetivo *doador* ("aquele que doa") prevê três argumentos, um obrigatório e dois opcionais (Figura 12). Em outras palavras, *doador* subcategoriza obrigatoriamente um SN e facultativamente dois SPreps. O primeiro SPrep é introduzido pela preposição *de* e o outro pode ser introduzido pela preposição *a* ou *para*.

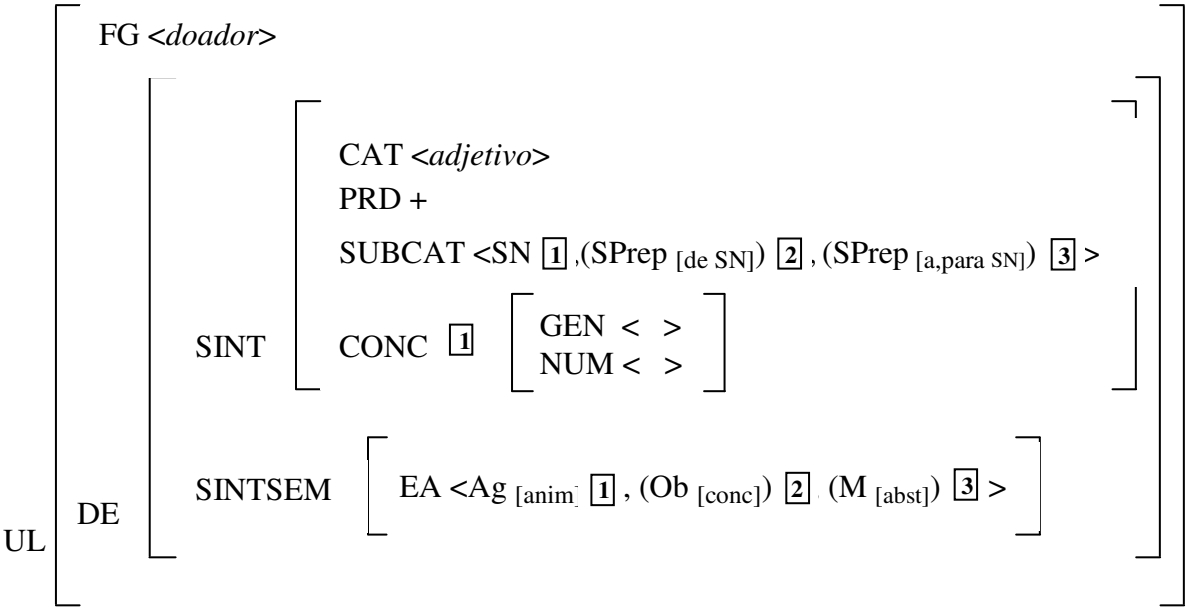


Figura 12: Representação da entrada lexical do adjetivo *doador*

A estrutura de traços SINTSEM, por sua vez, é responsável por explicitar o número, o tipo semântico e a função semântica dos argumentos projetados pelo item lexical. Como mencionado, o adjetivo *doador*, em sentenças como a referida, pode projetar três argumentos, os quais estão representados semanticamente da seguinte maneira na estrutura de traços SINTSEM: o argumento marcado com o índice [1] no traço SUBCAT é do tipo <animado> e exerce a função semântica de Ag(ente); o argumento marcado com o índice [2] é do tipo +concreto e exerce a função semântica de Ob(jetivo); por fim, o argumento marcado com o índice [3] é do tipo +abstrato e recebe o papel temático M(eta).

Cabe ainda ao traço SUBCAT especificar as diferentes categorias em que um argumento subcategorizado pode se realizar. Por exemplo, o adjetivo *temeroso* pode subcategorizar dois argumentos, um deles realiza-se sob a forma de SN e o outro, sob a forma de SPrep [com/de/por SN] ou SO (sintagma oracional) [de/em O]. Alguns exemplos estão em (41).

(41) a. Os comerciantes [...], *temerosos* com a queda [SPrep] do poder aquisitivo [...] (FI)

b. Os anunciantes estavam *temerosos* em aplicar [SPrep] suas verbas. (JO-FSP)

c. A cadela estava *temerosa* de ser ferida [SO] (JO-FSP)

Na Figura 13, ilustra-se a entrada lexical do adjetivo *temeroso*.

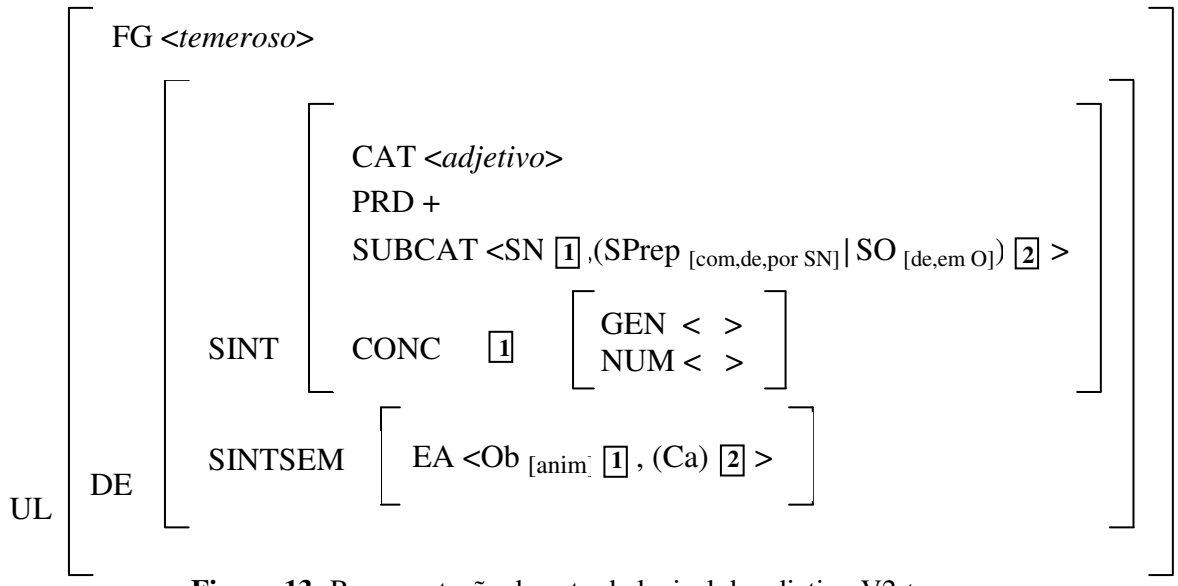


Figura 13: Representação da entrada lexical do adjetivo V2 *temeroso*

Para relacionar todas as categorias sintagmáticas nas quais o argumento A2 projetado pelo adjetivo em questão pudesse se realizar, foi inserido no modelo um tipo de notação, marcada pelo uso da barra vertical "|". Grosso modo, essa notação pode ser entendida como o *ou lógico*, que indica alternância de valores. No modelo de representação proposto, essa marcação assume a seguinte interpretação: o argumento A2 é tal que tanto poderá se realizar na forma de SPrep [com/de/por SN] ou SO [de/em O], sendo que ambas as formas encontram-se em disjunção exclusiva: uma delas pode ocorrer (SPrep ou SO), mas não ambas.

Como mencionado, todo argumento subcategorizado é associado a um índice numérico. Além de indicar o argumento do qual o adjetivo "herda" os valores dos traços GEN e NUM, esse recurso de indexação permite, como já se disse, "ligar" as estruturas SUBCAT e EA.

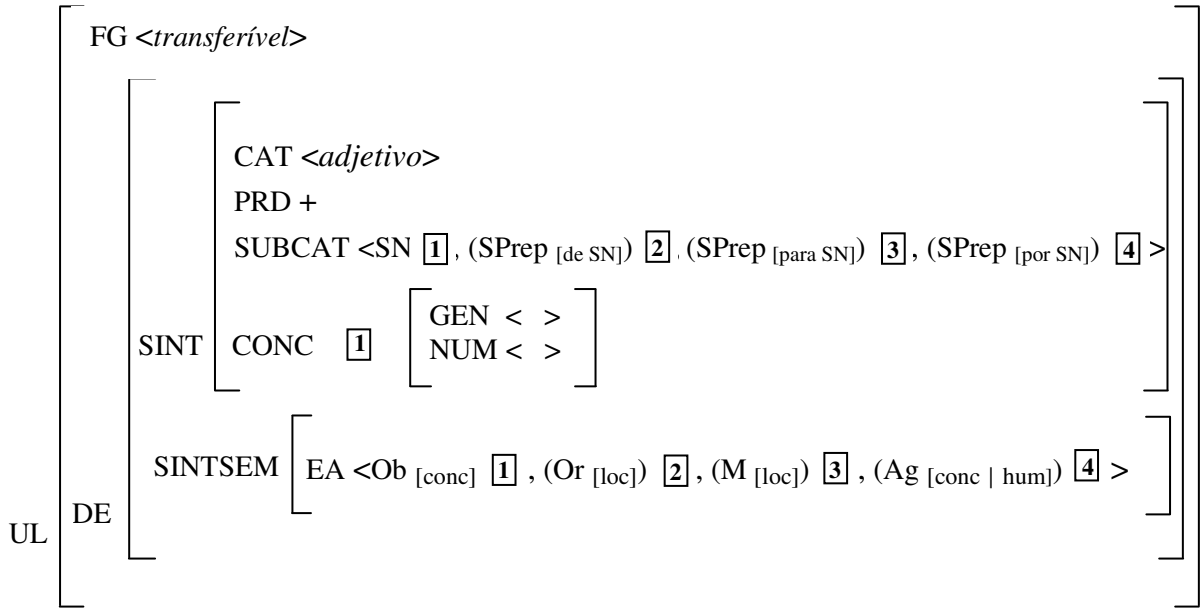


Figura 14: Representação do adjetivo V4 *transferível*

O valor do traço SUBCAT especifica que *transferível* pode apresentar quatro argumentos subcategorizados, um SN obrigatório e três SPreps facultativos (*de* + SN; *para* + SN; *por* + SN). Quanto à estrutura de traços SINTSEM, observa-se que o traço EA especifica que o argumento [1] é do tipo +concreto e recebe o papel temático Ob(jetivo); o argumento [2] é do tipo +locação e recebe o papel temático Or(igem); o argumento [3] também é do tipo +locação mas recebe o papel temático M(eta); e, por fim, o argumento [4] pode ser do tipo +concreto ou +humano e estabelece uma relação agentiva com o adjetivo *transferível*, recebendo, portanto, o papel temático Ag(ente).

No modelo proposto, ressalta-se que é possível representar, em um léxico lingüístico-computacional, as seguintes informações sobre os adjetivos valenciais:

- *realização gráfica;*

- *categoria gramatical;*

- *tipo sintático;*

- *esquema de subcategorização;*

- *valência* (estrutura de argumentos)*;*

- papéis temáticos;

- restrições selecionais;

- categorias de número e gênero.

Por fim, salienta-se também que o modelo de representação proposto na Figura 7 caracteriza-se por agrupar informações de natureza morfossintática e semântica. Representadas formalmente, tais informações podem ser utilizadas na compilação de léxicos (monolíngües) lingüístico-computacionais.

Capítulo 5 Da proposta de edição de informações na base da rede Wordnet.Br

Neste capítulo, apresenta-se a proposta de inserção de informações sintático-semânticas sobre os adjetivos valenciais na base da rede Wordnet.Br. Antes, porém, faz-se uma breve apresentação dessa base, destacando suas principais características, potencialidades e estrutura.

5.1. Das características da base da Wordnet.Br

A construção de *wordnets* tornou-se lugar comum no âmbito do Processamento Automático das Línguas Naturais. Vários países já possuem cada um a sua (p.ex.: Estados Unidos, Holanda, Itália, Espanha, Inglaterra, entre outros), integrando um projeto maior chamado **EuroWordNet** (VOSSEN, 1998). Esse projeto congrega bases lexicais das línguas citadas e de outras línguas com vistas a criar uma grande base plurilíngüe que sirva de apoio para a construção de ferramentas lingüístico-computacionais para as várias línguas envolvidas. No projeto EuroWordNet, todas as redes *wordnets* estão indexadas, conjunto a conjunto, com a rede WordNet, desenvolvida na Universidade de Princeton, para o inglês americano; essa, inclusive, é a "mãe" de todas as redes *wordnets*. A indexação das diferentes redes wordnets se faz por meio do **Índice-Inter-Lingüe** (ILI) (do inglês, *Inter-Lingual-Index*).

A rede WordNet é, na verdade, uma base relacional, em que unidades lexicais do inglês, pertencentes às categorias dos substantivos, verbos, adjetivos e advérbios, estão organizadas em termos de conjuntos de sinônimos (isto é, os *synsets*) que expressam conceitos lexicalizados. Tais conjuntos relacionam-se entre si em função das cinco relações de

sentido: antonímia, hiponímia, meronímia, acarretamento e causa (VOSSEN, 1998). Além disso, essa rede registra informações periféricas, associadas a cada sentido armazenado. São elas: frases-exemplo e glosas (isto é, definições informais).

A base da rede Wordnet.Br pode ser definida como uma base relacional de unidades lexicais do português (palavras e expressões) organizadas à semelhança de um *thesaurus*, dado que a estruturação das quatro categorias lexicais incluídas na base (substantivos; verbos; adjetivos; advérbios) lembram um *thesaurus* na medida em que foram organizadas em termos de conjuntos de sinônimos que expressam conceitos lexicalizados (DIAS-DA-SILVA; OLIVEIRA; MORAES, 2003). Na fase atual de desenvolvimento, estão sendo feitas a análise de consistência semântica dos synsets e a coleta e seleção de frases-exemplo[41].

Para os pesquisadores do PLN, essa base possibilita, por exemplo, a geração de parcelas de léxicos especiais, imprescindíveis para o desenvolvimento de diversos sistemas de PLN, tais como: sistemas de tradução automática, de sumarização automática, entre outros (SAINT-DIZIER, VIEGAS 1995; DIAS-DA-SILVA, 1998; PALMER 2001). Ao usuário da língua portuguesa, por sua vez, essa base, acoplada a ferramentas computacionais de auxílio à escrita, poderá oferecer a opção de escolha *on line* de palavras sinônimas e antônimas que, por motivos de estilo, precisão, adequação comunicativa, correção ou aprendizagem, o usuário queira substituir (ILARI, GERALDI, 1985).

5.2. Da proposta de inserção de novos dados

Atualmente, a base da rede Wordnet.Br armazena informações sobre as *relações* que se estabelecem entre os itens lexicais da língua. Retomando o que foi dito a respeito da estrutura do "léxico mental", vê-se que as informações contidas nessa base são, basicamente, referentes à *macroestrutura* do léxico, mais especificamente, às relações léxico-semânticas de

[41] CNPq 09/2001- Processo Nº 552057/01-0.

sinonímia e antonímia. Com a proposta de inserção de novos dados (no caso, relativos aos adjetivos valenciais), objetiva-se armazenar informações sobre a *microestrutura* do léxico, ou seja, informações sobre os itens lexicais da língua (e não sobre as relações que se estabelecem entre eles).

Como decorrência da investigação feita a respeito dos adjetivos do português, propõe-se uma extensão das informações lexicais para essa classe. Nesse sentido, além de abrigar as "relações" (léxico-semânticas) que se instauram entre os adjetivos, propõe-se que a estrutura da base da rede Wordnet.Br também possa abrigar informações sobre o *lema* de cada adjetivo, ou seja, sobre as propriedades sintático-semânticas responsáveis pela *microestrutura* do léxico (BIERWISCH, SCHREUDER, 1992; HANDKE, 1995). Nas seções subseqüentes, descreve-se a estrutura atual dessa base e esquematiza-se a extensão da mesma.

5.3. Da estrutura atual da base da Wordnet.Br

Na base da rede Wordnet.Br, a relação de sinonímia é representada formalmente pela relação lógica de pertença (x é sinônimo de y $\leftrightarrow$ x $\wedge$ y $\in$ A, em que A é um *synset*). A antonímia, por sua vez, é representada por uma relação entre conjuntos (x é antônimo de y $\leftrightarrow$ x $\in$ A e y $\in$ B, A e B são *synsets* e A e B estão relacionados pela relação de antonímia) (DIAS-DA-SILVA; OLIVEIRA; MORAES, 2002). O Esquema 1 ilustra a estrutura dessa base:

(1) Entrada n (categoria sintática *x*)
- Sentido n1
  - Synset
  - Synset de antônimos
- Sentido n.m

Nesse esquema, "n" é o número de identificação da entrada, "x" é uma variável que representa uma das quatro categorias gramaticais (substantivo, verbo, adjetivo ou advérbio), "n.1...n.m" são os números que identificam cada sentido da entrada n.

A partir da descrição sobre os adjetivos valenciais do português feita neste trabalho, propõe-se a ampliação da estrutura original para que as seguintes informações possam ser inseridas: (i) a tipologia quantitativa, (ii) o esquema de subcategorização e a (ii) estrutura de argumentos ou valência semântica.

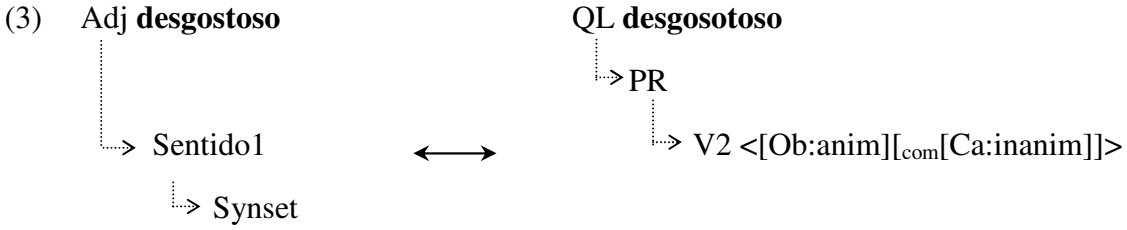
5.4. Da proposta de extensão da informação-base da rede Wordnet.Br

Paralelamente ao Esquema 1, propõe-se que a base da rede Wordnet.Br, além de conter os conjuntos de sinônimos e antônimos, deverá conter também informações sobre as propriedades sintáticas e semânticas dos adjetivos em *função predicativa*. O Esquema 2 exemplifica essa extensão.

(2) Unidade de Entrada n (Tipo semântico QL)
- Função sintático-semântica PR
  - Valência n.1 *<papéis temáticos dos As + restrições de seleção>*
  - Valência n.4

Nesse esquema, "n" é o número de identificação da unidade de entrada; "QL" indica que o adjetivo é do tipo "qualificador"; "PR" indica que a função sintático-semântica é a de "predicador"; "n.1...n.4" indica qual o subtipo lógico-semântico, (ou seja, V1, V2, V3 e V4), ao qual é associada a *valência semântica* (papéis temáticos + restrições selecionais) do adjetivo propriamente dito.

Salienta-se que indexações apropriadas poderão permitir, quando pertinente, o relacionamento entre as entradas estruturadas em termos das relações léxico-semânticas (sinônima e antonímia) e as entradas estruturadas em termos das informações sintático-semânticas aqui propostas. No caso específico do Esquema 2, indexações poderão permitir o relacionamento entre *valência* e *sentido*, este indiretamente armazenado na base da rede Wordnet.Br nos conjuntos de sinônimos. O Esquema 3 ilustra essa possível indexação.



No exemplo em (3), observa-se que a valência semântica (ou estrutura de argumentos) descrita entre os símbolos "< >" está associada ao Sentido1 do adjetivo *desgostoso*. Ressalta-se que para a especificação dessa valência, tomar-se-á, como base, a frase-exemplo. O príncipe Charles está desgostoso e escandalizado com as acusações contra sua mulher, a princesa Diana. A partir dessa frase-exemplo, identificam-se as relações semânticas que o PR estabelece com os argumentos e as restrições que ele impõem sobre os mesmos. Nessa frase, o adjetivo PR estabelece uma relação de "causa" com o argumento que se realiza na forma do SPrep com as acusações e, por isso, esse A recebe o papel temático Ca(usativo); já o A que se realiza na forma do SN sujeito o príncipe Charles, recebe papel temático Ob(jetivo). Dessas observações, elaborar-se-á uma estrutura do tipo <[Ob:anim][$_{com}$[Ca:inanim]]>.

Uma vez especificada a estrutura de argumentos que "traduz" o(s) sentido(s) de um adjetivo QL, poder-se-á generalizá-la para os demais membros do *synset*. Assim, os adjetivos {anojado; desagradado; descontente; desgostoso; dissaborido; dissaboroso;

malcontente; penalizado; triste} poderão herdar a valência especificada no exemplo em (3).

5.5. Da proposta de extensão da informação periférica da rede Wordnet.Br

Se, por um lado, as valências lógico-semântica e semântica poderão ser indexadas à estrutura original da base da Wordnet.Br (Esquema 1), por outro lado, a valência sintática, ou, como já se disse, *esquema de subcategorização*, poderá ser associada a uma das informações classificadas aqui como periféricas, no caso, a frase-exemplo. Diz-se periférica pelo fato de que essas frases, assim como as *glosas*, não constituem os *synsets*. Merece destaque, entretanto, o fato de a frase-exemplo fornecer o importante contexto de uso do adjetivo. Assim, na extensão que aqui se propõe, o *esquema de subcategorização*, ao estar associado à frase-exemplo, constitui o que poderia ser considerado o "comentário sintático" desta. Em (3), por exemplo, o esquema de subcategorização <[SN] [SPrep (com SN)]]> seria associado à frase-exemplo O príncipe Charles está desgostoso e escandalizado com as acusações contra sua mulher, a princesa Diana. Para a especificação do *esquema de subcategorização*, poder-se-á utilizar o conjunto de *padrões de subcategorização* descritos no Quadro 5.

Com isso, ressalta-se que, além de abrigar relações léxico-semânticas, a base da rede Wordnet.Br também poderá abrigar informações sobre o lema de cada adjetivo, ou seja, sobre as propriedades sintático-semânticas responsáveis pela microestrutura do léxico. Conseqüentemente, essa base poderá fornecer listas, para a compilação de léxicos monolingües, que contemple a valência e o esquema de subcategorização dos adjetivos. Informações essas pertinentes para léxicos que poderão integrar os mais variados tipos de sistemas de PLN.

Capítulo 6 Das considerações finais

Tendo como motivação principal a construção de léxicos lingüisticamente motivados para aplicação de PLN, apresentou-se, neste trabalho, uma proposta de representação lingüístico-computacional para os adjetivos valenciais do português do Brasil.

Para tanto, foram investigadas as principais propriedades desses adjetivos. Sendo valencial, uma das principais propriedades é a projeção de uma *estrutura de argumentos* ou *valência*. Essa estrutura tem sido amplamente empregada para representar parte das informações léxico-semânticas. Dessa forma, vê-se que a proposta de representação apresentada neste trabalho não envolve apenas propriedades morfossintáticas, como comumente acontece, mas também parte das propriedades semânticas sobre os adjetivos valenciais.

A inserção de informação semântica em léxicos lingüístico-computacional, aliás, tem sido um dos principais objetivos a ser alcançado na construção ou desenvolvimento de léxicos para fins do processamento automático do português.

Tal desenvolvimento, em específico, requer que as informações léxico-gramaticais sejam representadas de modo explícito, consistente e não-ambíguo, pois só assim elas podem ser transformadas em programas computacionais. Para os adjetivos valenciais, propôs-se um modelo de entrada lexical baseado no construto formal das *estruturas de traços* ou *matrizes de atributo-valor*, que garante a "transportabilidade" das informações representadas a diferentes linguagens de programação e algoritmos.

Do ponto de vista prático, apresentou-se, neste trabalho, uma proposta de inserção de informações sintático-semânticas sobre os adjetivos na base da rede Wordnet.Br. Sugeriu-se que essa base, além de armazenar informações sobre as relações léxico-semânticas típicas dessa subclasse de adjetivos, armazenasse também informações sobre a sua valência e

esquema de subcategorização. Essa proposta tem como objetivo propor um meio de refinamento dos dados armazenados nessa importante base de dados lexicais.

Dessa forma, acredita-se que este trabalho pode ser uma contribuição tanto para as pesquisas realizadas no âmbito do PLN quanto para aquelas desenvolvidas no domínio da teoria lingüística.

Referências bibliográficas

ALLEN, J. F. **Natural language understanding.** Menlo Park: Benjamin Cummings, 1987.

ALLWOOD, J.; ANDERSSON, L-G.; DAHL, O. **Logic in Linguistics.** Cambridge: Cambridge University Press, 1977.

AMAREL. S. Problem solving. In: Shapiro, E. (Ed.). **Encyclopedia of artificial intelligence.** New York: Wiley, 1990.

AMARO, R. F. **Posição do adjectivo e ligação seletiva: especificações para a computação do significado.** Lisboa, 2002, 159p. Dissertação (Mestrado em Lingüística). Faculdade de Letras da Universidade de Lisboa, Lisboa.

ANDERSON, S. **A-morphous Morphology**. Cambridge: Cambridge University Press, 1992.

ARNOLD, D. J. e SADLER, L. Noun-modifying adjectives in HPSG. **Technical report, Department of language and linguistics, University of Essex**, 1992.

ATKINS, S., et al. From resources to applications. Designing the multilingual ISLE lexical entry. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION - LREC, 3, 2002, Las Palmas, **Proceedings**… Las Palmas: University of Las Palmas, p. 687-683, 2002.

BAKHTIN, M. **Marxismo e filosofia da linguagem**. 3. ed. São Paulo: Hucitec, 1986.

BASÍLIO, M. **Teoria lexical**. São Paulo: Ática, 1987.

____. Aspectos da Representação Lexical de Formas Derivadas. **Palavra**, v.2, n.1, Rio de Janeiro: PUC-Rio, 1994.

____. Introdução: Questões clássicas e recentes da delimitação das unidades lexicais. In: Basílio, M. (ed.) **Palavra.** Departamento de Letras da PUC-Rio, pp. 9-18, 1999.

BASÍLIO, M., GAMARSKI, L. Adjetivos denominais no português falado. In: Castilho, A. T. de (Org.). **Gramática do Português Falado** – v. IV. 2ª ed. Campinas: UNICAMP, p. 629-650, 2002.

BECHARA, E. **Moderna Gramática Portuguesa**. São Paulo: Companhia Editora Nacional, 34a. ed., 1992.

____. **Moderna gramática da Língua Portuguesa.** Rio de Janeito: Editora Lucerna, 2001.

BHAT, D. N. S. **The adjectival category: criteria for differentiation and identification**. Studies in Language Companion Series (SLCS). Amsterdã/ Filadélfia: J. Benjamins, 1994.

BIDERMAN, M. T. C. A estrutural mental do léxico. In: **Estudos de Filologia e Lingüística -** Homenagem a Isaac Nicolau Salum. São Paulo: Editora da USP; T. A. Queiroz, p. 131-45, 1981.

____. Conceito Lingüístico de Palavra. In: Basílio, M. (ed.) **Palavra.** Departamento de Letras da PUC-Rio, pp. 81-97, 1999.

____. **Teoria lingüística: teoria lexical e lingüística computacional**. 2ª ed. São Paulo: Martins Fontes, 2001.

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

BLOOMFIELD, L. **Language**. London: Allen & Unwin, 1933.

BOCK, J. K. Towards a Cognitive Psychology of Syntax. **Psychological Review**, 89, p. 1-47, 1982.

BOGURAEV, B., BRISCOE, T. (Eds.). **Computational Lexicography for Natural Language Processing**, London: Longman, 1989.

BOLINGER, D. Adjectives in English: atributtion e predications. **Lingua***,* 18, p.1-34, 1967.

BORBA, F. S. **Introdução aos estudos lingüísticos**. 11ª ed. Campinas, São Paulo: Pontes, 1991.

____. **Uma gramática de valências para o português**. São Paulo: Editora Ática, 1996.

____. **Dicionário de usos do português do Brasil.** São Paulo: Editora da UNESP, 2002.

BORGES NETO, J. **Adjetivos: predicados extensionais e predicados intensionais**. Campinas: Editora UNICAMP, 1991.

BOSQUE, I.; PICALLO, C. Postnominal adjective in Spanish DPs. **Journal of Linguistics**, n.32, p.349-385, 1996.

BOUILLON, P. The adjective "vieux": The point of view of Generative Lexicon. In: Viegas, E. (ed.). **Breadth and depth of semantic lexicons**. Kluwer: Dordrecht, 148–166, 1999.

BRESNAN, J. An approach to Universal Grammar and the mental representation of language. **Cognition**, 10, p. 39-52, 1981.

____ (ed.). **The mental representation of grammatical relations**. Cambridge, Mass: The MIT Press, 1982.

____. **Lexical-functional syntax**. Stanford: University of Stanford, 2000.

BRISCOE, T. Lexical issues in natural language processing. In: Klein, E.; Veltman, F. (Eds.). **Natural language and speech**. Springer-Verlag, p.39-68, 1991.

BUSSE, W.; VILELA, M. **Gramática de valências**. Coimbra, Almedina, 1986.

BUTTERWORTH, B. (Ed.) **Language production volume 2: development, writing and other language processes**. London: Academic Press, 1983.

CALZOLARI, C., ZAMPOLLI, A., LENCI, A. Towards a standard for a multilingual lexical entry: the AGLES/ISLE initiative. **Computational Linguistics and Intelligent Text Processing: Third International Conference - CICLing**. Springer-Verlag Heidelberg, p. 264-279, 2002.

CARLSON, G. **Reference to kinds in English.** Tese de Doutorado, University of Massachusetts: Amhrest, 1977.

CASTELEIRO, J. M. **A sintaxe transformacional do adjetivo**. Lisboa: Instituto Nacional de Investigação Científica, 1981.

CASTILHO, A. T. Análise multissistêmica das preposições do eixo transversal no português brasileiro: espaço /anterior/~/posterior/. 2003. In: Jânia Ramos e Mônica Alckmim (Orgs.). Para a história do português brasileiro, Vol. IV. No prelo.

CHIERCHIA, G. **Semântica**. Tradução de Luiz Arthur Pagani, Ligia Negri e Rodolfo Ilari. Campinas, SP: Editora da UNICAMP; Londrina, Pr: EDUEL, 2003. Tradução de Semântica.

CHIERCHIA, G; MC-CONNELL-GINET, S. **Meaning and Grammar**. Cambridge: MIT Press, 1990.

____. Remarks on nominalizations. In: Jacobs, R. A.; Rosenbaum, P.S. (Eds). **Readings in English Transformational Grammar**. Ginn and Company. 1970.

____. **Lectures on government and binding**. Dordrecht: Foris, 1981.

CHOMSKY, N; LASNIK, H. Filters and control. **Linguistics Inquiry,** 8:3, p. 425-504, 1977.

COATES, J. Denominal adjectives: a study in syntactic relationships between modifier and head. **Lingua**, n.27, p.160-169, 1971.

COLE, R. A. (Ed.). **Survey of the State of the Art in Human Language Technology**. Cambridge: Cambridge University Press, 1995.

COOPER, R. et al. (Eds.). Using the framework. **The FRACAS Consortium**. Deliverable D16 – Final Draft, 1996.

CRUSE, D. A. **Lexical semantics**. Cambridge: Cambridge University Press, 1986.

CUNHA, C., CINTRA, L. **Nova gramática do português contemporâneo**. 2ª ed. Rio de Janeiro: Nova Fronteira, 1985.

DAVIS, A. R. **Linking by types in the hierarchical lexicon**. Stanford: CSLI Publications, 2001. Disponível em <http://www-linguistics.stanford.edu/~tdavis/thesis-ps.html>. Acesso em 1 agosto de 2002.

DEMONTE, V. Semántica composicional y gramatica: los adjetivos en la interficie léxico-sintaxis. **Revista Española de Lingüística**, 29, v.2, p. 283-316, 1999a.

____. El adjectivo: classes y usos. La posición del adjetivo en el sintagma nominal. In: Bosque, I., Demonte, V. (Orgs.). **Gramática Descriptiva de la Lengua Española.** Madrid: Real Academia Española - Espasa, 1999b.

DEESE, J. **The structure of associations in language and thought.** Baltimore: Johns Hopkins Press, 1964.

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.

____. Bridging the gap between linguistic theory and natural language processing. In: **Proceedings of the 16th international congress of linguistics**. Oxford: Elsevier Sciences, n. 16, p. 1-10, 1998.

DIAS-DA-SILVA, B. C.; OLIVEIRA, M. F. Inclusão de informação pragmático-discursiva na base lexical de um thesaurus eletrônico. **Estudos Lingüísticos**, 31, 2002. ISSN 1413 0939. 1 CD-ROM.

DIAS-DA-SILVA, B. C.; OLIVEIRA, M. F.; MORAES, H. R. Groundwork for the development of the Brazilian Portuguese Wordnet. **Advances in natural language processing**. Berlin: Springer-Verlag, p.189-196, 2002.

____. Reusability of dictionaries in the compilation of NLP lexicons. In: **Computational processing of the Portuguese language.** Berlim: Springer-Verlag, 2003.

DIK, S. C. **The theory of functional grammar**. Berlin, New York: Mouton de Gruyter, 1997.

DORR, B. J. **Machine translation: a view from the lexicon.** Massachussetts: Massachussetts Institute of Technology, 1993.

DUBOIS, J. et al. **Dicionário de Lingüística**. São Paulo, Cultrix, 1973.

FAULSTICH, E. L. de J. Descrição lexicográfica e dados de argumentatividade – leitura de um verbete: o adjetivo delicioso. **Estudos lingüísticos**. n. 19, p. 157-163, 1990.

FELLBAUM, C. (Ed.). **Wordnet: an electronic lexical database**. Cambridge: The MIT Press, 1999.

FERREIRA, A. B. H. **Dicionário Aurélio eletrônico Século XXI**. Versão 3.0, São Paulo: Lexikon Informática, 1999. 1 CD-ROM.

FILLMORE, C. J. The case for case. In: Bach, E., Harms, R. T. (Eds.). **Universals in linguistic theory**. Holt, Rinehart and Winston, Inc., p.1-88, 1968.

FRAZIER, L. Grammar and Language Processing. In: Newmeyer, F. (Ed.). **Linguistics: the Cambridge survey II: linguistic theory: extensions and implication**. Cambridge: CUP, p.15-34, 1989.

GAZDAR, G. et al. **Generalized phrase structure grammar**. Cambridge, Mass.: Harvard University Press, 1985.

GRIMSHAW, J. **Argument structure.** Cambridge: The MIT Press, 1992.

GROSS, D.; MILLER, K. J. Adjectives in WordNet. **International Journal of Lexicography**, 3 (4), 1990, p. 265 - 277. Disponível em <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>. Acesso em 20 fev. 2001.

GROSS, D.; FISCHER, U.; MILLER, A. The organization of the adjectival meaning. **Journal of Memory and Language**, 28, p. 92-106, 1989.

GRUBER, J. **Studies in lexical relations**. Tese de Doutorado, Cambridge, Mass., 1965.

HACKING, I. Aristotelian categories and cognitive domains. **Synthese**, 126, p.473-515, 2001.

HANDKE, J. **The structure of the Lexicon: human versus machine**. Berlin: Mouton de Gruyter, 1995.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa (versão 1.0).** Rio de Janeiro: Editora Objetiva, 2001. CD-ROM.

HUTCHINS, W. J.; SOMERS, H. L. **An introduction to machine translation**. London: Academic Press, 1997.

ILARI, R.; GERALDI, J. W. **Semântica**. São Paulo: Editora Ática, 1985.

JACKENDOFF, R. Morphological and semantics regularities in the lexicon. **Language**, Baltimore: v. 51, 3, 1975.

____. **Semantic structures**. Cambridge: Mass.: The MIT Press, 1991.

____. **The architecture of the language faculty**. Cambridge: Mass.: The MIT Press, 1997.

JUSTESON, J. KATZ, S. **Co-occurrences of antonymous adjectives and their contexts**. Yorktown Heights, NY: IBM Research Report, 1989.

KAPLAN, R. e BRESNAN, J. Lexical-functional grammar: a formal system for grammatical representation. In: Bresnan, J.(Ed.). **The mental representation of grammatical relations**. Cambridge, MA: MIT Press, 1982.

KOENIG, J-P. **Lexical relations.** Stanford, CA: 1999.

KRATZER, A. An investigation into the lumps of thought. **Linguistics and Philosophy**, 12, p. 607-653, 1989.

LAKOFF, G. **Irregularity in syntax**. Holt: Rinehart and Winston, Inc., 1970.

LANGACKER, R. **Fundamentals of Linguistic Analysis**. New York: Harcort, Brace, Jovanovich, 1972.

LEVELT, W .J. M. Accessing words in speech production: stages, processes and representations. **Cognition**, 42, p.1-22, 1992.

____. **Speaking: to intention to articulation**. Cambridge, Mass.: The MIT Press, 1993.

LEVI, J. **The syntax and semantics of complex nominals**. New York: Academic Press, 1978.

LEVIN, B., PINKER, S. Introduction to special issue of *Cognition* on lexical and conceptual semantics. **Cognition**, 41, p.1-7, 1991.

LOBATO, L. **Adjetivos: tipologia e interpretação semântica**. Boletim da ABRALIN, n.14, p.9-25, São Paulo: ABRALIN, 1993.

LOWIE, W. **The acquisition of interlanguage morphology: a study into the role of morphology in the L2 learner's mental lexicon**. Groningen: University Library Groningen, 1998.

LYONS, J. **Semantics**. Cambridge: Cambridge University Press, vol. 2, 1977.

____. **Introdução à lingüística teórica**. Tradução de Rosa Virgínia Mattos e Silva e Hélio Pimentel; revisão e supervisão de Isaac Nicolau Salum – São Paulo: Ed. Nacional, Editora da Universidade de São Paulo, 1979.

____. **Language and linguistics**. An introduction. Cambridge: CUP, 1981.

MALING, J. Transitive adjectives: a case of categorial reanalysis. In: Henry, F., Richards, B. (Eds.). **Linguistic Categories: Auxiliaries and Related Puzzles**. Dordrecht: D. Reidel Publishing Company, vol. 1, pp. 253-89, 1983.

MANI, I. **Automatic Summarization**. Amsterdam: John Benjamins Publishing Co., 2000.

MATTOSO CÂMARA JR., J. **Estrutura da língua portuguesa**. Petrópolis: Vozes, 1970.

MARANTZ, A. On the nature of gramatical relations. **Linguistic Inquiry**. Monography 10. Cambridge: The MIT Press, 1984.

MEL'CUK, I. **Dependency Syntax: theory and practice.** The SUNY Press, Albany, N.Y, 428p. 1988.

MENUZZI, S. de M. **Sobre a modificação adjetival do português.** Araraquara, 1992. 200p. Dissertação (Mestrado). Universidade de Campinas, Campinas.

MILLER, G. A., FELLBAUM, C. Semantic networks of English. **Cognition***,* v. 41, n. 1-2, p. 197-229, 1991.

MILLER, G. A. et al. Introduction to WordNet: an on-line lexical database. **Journal of Lexicography**, 3(4), p. 235-244, 1990.

____. **Five Papers on WordNet**. 1993. Disponível em <http://www.cogsci.princeton.edu/~wn>. Acesso em: 30 mar. 2001.

MIRA MATEUS, M. H. et al. **Gramática da Língua Portuguesa**. 4ª ed., Lisboa: Caminho, 1994.

MULLER, C. **Essai de statistique lexicale.** L'IC, Klincksieck, 1964.

MURPHY, M. L. Knowledge of words versus knowledge about words: the conceptual basis of lexical relations. In: Peeters, B. (Org.) **The lexicon-encyclopedia interface**. New York/Amsterdam: Elsevier, p.317-349, 2000.

NEVES, M. H. M. **A vertente grega da gramática tradicional**. São Paulo: HUCITEC, 1987.

____. **A gramática funcional**. São Paulo: Martins Fontes, 1997.

____. **Gramática de usos do português**. São Paulo: Editora UNESP, 2000.

NEWMEYER, F. J. **Language Form and Language Function**. MIT Press, 1998.

NIQUE, C. **Initiation méthodique à la grammaire générative.** : Paris: Armand Colin, 1974.

NIRENBURG, S. **Machine translation.** San Mateo, Morgan Kaufmann, 1992.

PALMER, F. R. **Grammatical roles and relations**. Cambridge: Cambridge University Press, 1994.

PALMER, M Miltilingual resources – Chapter 1. In: Hovy, E., et al. (Eds.). **Linguistica Computazionale***,* v.14-15, 2001.

PEREIRA, E. C. **Gramática expositiva**. 8ª ed. São Paulo, 1918.

PERINI, M. A. **Gramática descritiva do português.** São Paulo: Editora Ática, 1998.

PESETSKY, D., TORREGO, E. Tense, case, and the nature of syntactic categories. 2002. In: Gueròn, J; Lacarme, J. (Eds.). **The Syntax of Time**. MIT Press. No prelo.

PINHEIRO, G. M., ALUÍSIO, S. M. Corpus Nilc: descrição e análise crítica com vistas ao projeto Lacio-Web. **Relat. Téc. ICMC/USP**, NILC RT-69, 2003.

POLLARD, C., SAG, I. A. **Information-based Syntax and Semantics**. Volume I: Fundamentals. Stanford: CSLI Publications, 1987.

____. **Head-driven phrase structure grammar**. Chicago: University of Chicago Press, 1994.

POSTAL, P. M. Anaphoric Islands. In: Binnick, R. I, et al. (Eds). **Papers from the Fifth Regional Meeting Chicago Linguistic Society.** Chicago: University of Chicago, p. 209-239, 1969.

PUSTEJOVSKY, J. **The generative lexicon**. 2ª ed. Cambridge: Mass.: The MIT Press, 1996.

QUIRK, R. et al. A **Comprehensive Grammar of the English Language.** London: Longman, 1991.

RAPOSO, E.P. **Teoria da gramática: a faculdade da linguagem**. Lisboa: Caminho, 1992.

RASKIN, V., NIRENBURG, S. Lexical Semantics of Adjectives: a microtheory of adjectival meaning. **Technical Reporty Computing Research Laboratory (CRL)/New Mexico State University** (NMSU), MCCS-288, 1995.

ROCHA, L. H. P. da. **A sintaxe do adjetivo verbal**. Araraquara, 1994. 154p. Dissertação (Mestrado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara.

ROCHA LIMA, C. H. **Gramática Normativa da Língua Portuguesa**. Rio de Janeiro: José Olympio Editora, 31a ed., 1972.

ROSNER, M, JOHNSON, R (Eds.). **Computational linguistics and formal semantics.** Cambridge: CUP, 1992.

SAG, I. A. English relative clause constructions. **Journal of Linguistics**, v.33, p.431-484, 1997.

SAG, I., WASOW, T. **Syntactic theory: a formal introduction**. Stanford: CSLI Publication, 1999.

SAINT-DIZIER, P., VIEGAS, E. **Computational lexical semantics**. Cambridge: Cambridge University Press, 1995.

SANFILIPPO, A. Lexicons for constraint-based grammars. In: Cole, R. A (Ed). **Survey of the state of the art in human language technology**. Oregon: Graduate Institute, p. 118-121, 1995.

SAPIR, E. **Language**. An introduction to the study of speech. New York: Harcourt, Brace & World, Inc., 1949.1949.

SELLS, P. **Lectures on contemporary syntactic theories**. Stanford: CSLI Publications, 1985.

SHIEBER, S. M. **An introduction to unification-based approaches to grammar**. CSLI Lecture Notes 4. Center for the Study of Language and Information. Stanford: CA, 1986.

SIEWIERSKA, A. **Functional grammar**. London-New York: Routledge, 1991.

SPENCER, A. On the redundancy of lexical categories. Paper presented at **Lexical Functional Grammar'97**, San Diego, June 1997.

STOWEL, T. **Origins of phrase structure.** Tese de Doutorado, MIT Press, 1981.

STUBBS, M. **Words and phrases: corpus studies of lexical semantics.** Great Britain: Blackwell Publishers, 2001.

TEYSSIER, J. Notes on the syntax of the adjective in modern English. **Lingua**, n. 20, p. 225-249, 1968.

VIEGAS, E., RASKIN, V. Computational semantic lexicon acquisition: methodology and guidelines. **Relat. Téc. CRL/NMSU**, MCCS-315, 1998.

VILELA, M. **Ensino da língua portuguesa: léxico, dicionário, gramática**. Coimbra: Almedina, 1995.

VOSSEN, P. et al **The EuroWorNet base concepts and top ontology.** Deliverable...1998. Disponível em: <http://www.illc.uva.nl/EuroWordNet/docs.html>. Acesso em: 17 fev. 2003.

WASOW, T. Postscript. In: Sells, P. **Lectures on contemporary syntactic theories**. Chicago: The University of Chicago Press, 1985.

WEISZFLOG, W. (Ed.). **Michaelis português: moderno dicionário da língua portuguesa.** Versão 1.0. São Paulo: DTS Software Brasil, 1998. 1 CD-ROM.

WILLIAMS, E. Argument structure and morphology. **Linguistic Review**, 1, p.81-114, 1981.

WILKS, Y. D. et al. Machine tractable dictionary as tools and resources for natural language processing. In: **Proceedings of Colling´88**, p. 750-55, 1988.

Anexo 1 Textos consultados: siglas e obras

Córpus do NILC	
Abreviaturas	**Texto correspondente**
LI-LT JO-FSP JO-JB	Livro – Literatura Texto jornalístico – Folha de São Paulo Texto jornalístico – Jornal do Brasil
Córpus do LL	
Abreviatura	**Texto correspondente**
A	*Ângela ou as areias do mundo*. FARIA, O. Rio de Janeiro: José Olímpio.
AV	*A viúva branca*. LEITE, A.
AVL	*A velhinha de Taubaté*. VERÍSSIMO, J.F. Porto Alegre: L&PM, 1983.
B	*A borboleta amarela*. BRAGA, R. Rio de Janeiro: Record, 1955.
BP	*Brasileiro perplexo*. QUEIROZ, R. Rio de Janeiro: Ed. do Autor, 1963.
CAR-O	*Discursos da Academia*
CG	*Contos gauchescos*. NETO, S.L. 5ª ed. São Paulo: Globo, 1957.
FI	*Ficção e ideologia*. CUNHA, F.W. Rio de Janeiro: Pongetti,1972.
FIA	*Fisiologia animal comparada*. PINSETTA, S.E. São Paulo: Anglo, 1985. (Livro-texto, 43.)
FSP	*Folha de São Paulo*.
GAN	*Grupos animais*: *embriologia dos cordados*. PINSETA, D.E. São Paulo: Anglo, 1985. (Livro-texto, 39.)
N	*Noite*. VERÍSSIMO, E. Porto Alegre: Globo, 1957.
NP	*Noções práticas de estatística.* NUNES, M. R. Rio de Janeiro: 1971.
Q	*Quarup*. CALLADO, A. 2.ed. São Paulo, Círculo do Livro, 1974.
VEJ	*Revista Veja*.