

# OS ADJETIVOS VALENCIAIS DO PORTUGUÊS E SUA REPRESENTAÇÃO LINGÜÍSTICO- COMPUTACIONAL

Ariani DI FELIPPO<sup>1</sup>

Bento Carlos DIAS-DA-SILVA<sup>2</sup>

■ **RESUMO:** Em função de aplicações reais como, por exemplo, os sistemas de tradução automática, para as quais sistemas de processamento automático de língua natural têm sido construídos, é essencial a compilação de léxicos (monolíngües e bilíngües) que sejam computacionalmente tratáveis e, sobretudo, lingüisticamente motivados. Essa tarefa, por sua vez, exige a descrição minuciosa das propriedades gramaticais e semânticas dos itens lexicais. Neste artigo, em particular, discutem-se as características léxico-gramaticais dos adjetivos valenciais do português do Brasil e os modelos relevantes para a representação formal específicas dessas características, com vistas à proposição da representação lingüístico-computacional para desse tipo de adjetivo em termos do formalismo denominado **estrutura de traços**.

■ **PALAVRAS-CHAVE:** Adjetivo. Valência. Semântica lexical. Léxico. Processamento Automático de Línguas Naturais.

---

<sup>1</sup> Doutoranda do Programa de Pós-Graduação em Lingüística e Língua Portuguesa – Faculdade de Ciências e Letras-UNESP – 14800-901 – Araraquara-SP, Brasil. E-mail: arianidf@uol.com.br.

<sup>2</sup> Departamento de Letras Modernas – Faculdade de Ciências e Letras-UNESP – 14800-901 – Araraquara-SP, Brasil. E-mail: bento@fclar.unesp.br.

## Introdução

Atualmente, após cinquenta anos de pesquisas acumuladas, o Processamento Automático das Línguas Naturais (PLN) é um domínio em plena ebulição, no qual sistemas computacionais são projetados, analisados, implementados e mesmo comercializados, haja vista o acelerado crescimento da indústria da informática. Dentre os vários tipos de sistemas de PLN (aplicações/produtos), estão (COLE, 1995): os sistemas de correção gramatical e ortográfica, de tradução, de sumarização de textos, de manipulação de unidades lexicais e de processamento de fala.

Todos esses sistemas de PLN, diferentes dos demais tipos de sistemas computacionais, manipulam o código lingüístico e, por isso, pressupõem um tipo de “arquivo” em que são armazenadas as unidades lexicais (palavras e expressões) que serão manipuladas pelo sistema durante os procedimentos de interpretação e/ou produção de língua natural. Esse “arquivo” é concebido como uma base de dados em que são especificadas, para cada unidade nela contida, informações de natureza lexical, morfológica, sintática, semântica e, até mesmo, pragmático-discursiva, dependendo das especificidades do sistema de PLN para o qual essa base de dados é desenvolvida (PALMER, 1999). Do ponto de vista do PLN, esse tipo de “mega arquivo” é definido como o “léxico” do sistema. .

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): (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) **lingüísticamente 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. Dessa forma, a construção de léxicos para fins do PLN requer, sobretudo, a investigação de questões como os diferentes formalismos de representação da informação lexical e as propriedades dos itens lexicais.

Neste trabalho, investiga-se uma pequena parte dessas questões, relacionada à representação lingüístico-computacional dos adjetivos. A razão para a escolha dessa classe de palavras pautou-se, basicamente, pelo

fato de que os adjetivos ocupam lugar de destaque na exteriorização da **visão de mundo** do falante, pois parece ser por meio dos adjetivos que se manifestam com bastante clareza as opiniões do locutor (relacionadas à crença, valores, afetividades e registro do que ocorre no mundo exterior) (BORBA, 1996). Naturalmente, qualquer palavra lexical se destina a esse papel, porém, o adjetivo ocupa lugar de destaque. Conseqüentemente, para melhor emular a análise e síntese dos enunciados lingüísticos, os sistemas de PLN não podem prescindir da manipulação das unidades que integram essa classe (RASKIN; NIRENBURG, 1995). Na verdade, propõe-se aqui a representação lingüístico-computacional dos adjetivos **valenciais**. Ressalta-se que a estrutura valencial tem sido amplamente empregada para representar parte das informações léxico-gramaticais das unidades do léxico (SAINT-DIZIER; VIEGAS, 1995; PUSTEJOVSKY, 1996).

Na busca de equacionamento metodológico para a proposta de representação lingüístico-computacional dos adjetivos valenciais, tomou-se por base Dias-da-Silva (1998), que fornece os subsídios essenciais para o desenvolvimento de projetos multidisciplinares na área do PLN. 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: **lingüístico**, **representacional** e **implementacional**. A proposta de representação aqui exposta, no entanto, pressupõe investigações apenas nos dois primeiros domínios.

As atividades do domínio lingüístico concentraram-se nas seguintes tarefas: (i) **sob o ponto de vista da Psicolingüística**: investigação sobre as questões da delimitação de **item lexical** e da estruturação interna das entradas no léxico; (ii) **sob o ponto de vista da Lingüística**: investigação teórica sobre a categoria adjetival, ou seja, identificação dos adjetivos valenciais do português e de suas principais propriedades.

Com base em pressupostos (psico)lingüísticos, foi possível (i) especificar os tipos de informação lexical lingüisticamente relevantes para o processamento automático das línguas naturais e (ii) delimitar, para os fins deste trabalho, o conceito de **item lexical**. A partir das considerações feitas a respeito de (i) e (ii), propôs-se a representação lingüístico-computacional dos adjetivos valenciais. Da investigação da classe dos adjetivos do português, foi possível identificar os adjetivos valenciais e, principalmente, as suas propriedades sintáticas e semânticas. Tal investigação fundamentou-se basicamente nas informações fornecidas pela literatura, sendo que, para a

exemplificação das propriedades desses adjetivos, foram utilizadas ocorrências retiradas dos *corpus* do Laboratório de Lexicografia (LL) da Faculdade de Ciências e Letras – UNESP/Araraquara e do Núcleo Interinstitucional de Lingüística Computacional (NILC) – USP/São Carlos<sup>3</sup>.

No domínio representacional, foram investigados os seguintes modelos de representação formal das informações léxico-gramaticais: **estrutura de traços** ou **matrizes de atributo-valor**, **lógica de predicados** e *frames*.

Em termos formais, este trabalho está assim organizado: primeiro, são feitas algumas considerações a respeito do conceito de **item lexical** e da estrutura interna das entradas no léxico; segundo, são descritas as principais propriedades dos adjetivos valenciais do português; terceiro, são descritos alguns dos principais modelos formais de representação das informações léxico-gramaticais; quarto, apresenta-se a proposta de representação lingüístico-computacional para os adjetivos valenciais, ou seja, um modelo de entrada lexical canônica que possa ser processado por sistemas de PLN; e, por fim, são feitas algumas considerações finais a respeito deste trabalho.

## **Considerações sobre a microestrutura do léxico e a delimitação de item lexical**

### **Da microestrutura do léxico ou estrutura interna das entradas lexicais**

Para propor a representação formal dos adjetivos valenciais para fins do PLN, era preciso delimitar os tipos de informação lexical a serem associados às entradas no léxico (no caso, monolíngüe) e as unidades a serem consideradas **entradas** lexicais.

A delimitação dos tipos de informação lexical a serem associados às entradas foi feita com base no modelo de processamento cognitivo de Levelt

---

<sup>3</sup> A todo uso extraído do *corpus* do LL, associa-se uma sigla que indica a referência do texto do qual o uso foi extraído, p.ex.: AF (ANGELO, I. *A festa*. 5.ed. São Paulo: Summus, 1978). E a todo uso extraído do *corpus* do NILC, associa-se um conjunto de informações que especifica o gênero e o tipo de texto, p.ex.: LI-LT = Livro – Literatura.

(1992, 1993). Nesse modelo, o acesso aos itens no léxico mental (LM)<sup>4</sup> é realizado em duas etapas: **seleção do lema** e **codificação da forma** da palavra (ou **lexema**) a ser verbalizada<sup>5</sup>. 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)<sup>6</sup>.

Tendo em vista que o acesso aos itens lexicais realiza-se nas etapas de seleção do lema e codificação da forma da palavra, conclui-se que a **especificação lexical**<sup>7</sup> (ou seja, o conjunto de informações associadas às entradas) está subdividida em duas dimensões: **lemática** e **lexemática**. Em outras palavras, diz-se que 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).

Partindo-se dessa concepção de acesso ao LM, concebe-se uma entrada lexical canônica (E) como uma estrutura de dados, no sentido computacional desse termo, contendo quatro **componentes** lingüísticos, os quais estão especificados no Quadro 1.

---

<sup>4</sup> É a parte do conhecimento lexical do indivíduo delimitada por sua língua (BIERWISCH; SCHREUDER, 1992; LOWIE, 1998).

<sup>5</sup> 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; especifica as condições conceituais que garantem o uso apropriado do item, indicando, entre outras coisas, sua classe gramatical e seus argumentos, e lexema é a representação das estruturas morfológica e fonológica de um item lexical.

<sup>6</sup> O Esq da estrutura silábica de gato, por exemplo, é preenchida pelos segmentos  $[/g/ /a/_{\mu}]_{\sigma}$  e  $[/t/ /u/_{\mu}]_{\sigma}$ . Os símbolos  $\sigma$  (sigma) e  $\mu$  (mi) são usados para representar, em fonologia métrica, a estrutura de uma palavra em termos dos elementos abstratos: sílaba ( $\sigma$ ) e mora ( $\mu$ ). 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).

<sup>7</sup> No modelo de processamento cognitivo de Levelt (1992, 1993), uma entrada lexical (representação de um item no LM) é composta pela unidade de acesso (unidade abstrata que pertence ao léxico) e pela especificação lexical (conjunto de informações) correspondente a essa unidade.

Quadro 1 – Esquema geral de representação das informações lexicais

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

**Do componente FG:** especificam-se a forma **gráfica** (ortográfica) e as estruturas morfológica e fonológica do item. Para a representação dos adjetivos valenciais, no entanto, não foram investigadas as informações referentes às estruturas morfológica e fonológica. Neste trabalho, considera-se que no componente FG representa-se apenas a forma **gráfica** dos itens. Essa forma **gráfica** nada mais é do que a seqüência de **grafemas** que constitui um item lexical. Os **grafemas** são, na verdade, os símbolos gráficos unos, constituídos por traços gráficos distintivos, que permitem ao falante entender visualmente os itens da língua, da mesma forma que os **fonemas** permitem ao falante entender 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 cavalgada).

**Do componente TG:** especificam-se as propriedades sintáticas desse item. Presume-se, aqui, que as propriedades sintáticas realizam-se por meio das **categorias gramaticais primárias** e das **categorias gramaticais secundárias**. Por categoria gramatical primária, entende-se a classe gramatical a que o item pertence, por exemplo, **substantivo**, **verbo** e **adjetivo**. A especificação da categoria primária do item engloba, além da classe, a descrição dos traços subcategoriais (isto é, o quadro de

subcategorização). No esquema de subcategorização, representa-se a forma sintagmática em que os argumentos lógico-semânticos podem ocorrer. Por categorias gramaticais secundárias, presumem-se as categorias de **gênero**, **número**, **modo**, **caso**, **aspecto**, **voz**, entre outras (LYONS, 1979).

**Do componente EA:** especificam-se o número e o tipo semântico dos argumentos projetados por um predicador (PR). Sendo predicador (PR), o item lexical tem a propriedade de poder ligar-se a um certo número de elementos exigidos pela sua semântica, **os argumentos** (As). Por exemplo:

(1) O rapaz era descendente de portugueses. (JO-FSP)<sup>8</sup>

Em (1), **descendente** estabelece relação com dois argumentos: **o rapaz** (A1) e **de portugueses** (A2). A essa propriedade, é dado o nome **valência** (BORBA, 1996; NEVES, 1997, 2000). Essa propriedade pode ser descrita em três níveis: lógico-semântico, sintático e semântico. O primeiro diz respeito ao número de As exigidos pela semântica do predicador. Por exemplo, **descendente** é um adjetivo de valência 2, pois pode projetar até dois argumentos. O segundo diz respeito ao esquema de subcategorização, ou melhor, à realização sintagmática dos argumentos lógico-semânticos. Por exemplo, o A1 de **descendente** realiza-se sintaticamente sob a forma de sintagma nominal (SN) (= **o rapaz**) e o A2, sob a forma de sintagma preposicional (SPrep) (= **de portugueses**). O terceiro nível, também denominado **estrutura de argumentos**, diz respeito às relações semânticas estabelecidas entre o predicador e seus argumentos; tais relações podem ser representadas por meio de **papéis temáticos**, p.ex.: em (1), o A1 tem a função semântica de Objetivo e o A2, de Origem<sup>9</sup>. Além dos **papéis temáticos**, também são empregadas as chamadas **restrições seletivas** na descrição da estrutura de argumentos. Tais restrições são traços de natureza semântica caracterizadores dos argumentos selecionados pelo predicador. O princípio básico do emprego dessas **restrições** é associar a cada

<sup>8</sup> JO-FSP = Jornal - Folha de São Paulo.

<sup>9</sup> Para efeitos práticos, a literatura tem demonstrado serem suficientes, para a descrição da estrutura de argumentos, 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) (BORBA, 1996). Dessa forma, são esses os papéis temáticos empregados na representação proposta neste trabalho.

argumento do predicador uma lista de **traços** ( $F_i$ ) que restringe o conteúdo semântico dos argumentos (SAINT-DIZIER; VIEGAS, 1995). O verbo **comer**, por exemplo, projeta dois argumentos, A1 e A2, cujos papéis temáticos são: Agente e Objetivo, respectivamente. Em outras palavras, os traços **animado** e **concreto** restringem o conteúdo semântico do A1 e A2 do verbo **comer**, respectivamente. O exemplo em (2), elaborado com base na Gramática Funcional (FG) (DIK, 1997), ilustra o uso de **papéis temáticos** e de **restrições seletivas** na representação da estrutura de argumentos:

(2) COMER [V] ( $x_1$ : <anim> ( $x_1$ ))<sub>Ag</sub> ( $x_2$ : <conc> ( $x_2$ ))<sub>Objetivo</sub>

Em (2), a **estrutura de predicado**<sup>10</sup> 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 (Ob). Os argumentos indicados pelas variáveis  $x_1$  e  $x_2$  apresentam as **restrições seletivas** <animado> e <concreto>, respectivamente.

**Do componente FS:** especifica-se a **forma semântica** (FS) do item lexical. Segundo Bierwisch e Schreuder (1992), a FS de um item representa a contribuição que esse item faz para o significado 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) **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;
- b) **estrutura qualia:** responsável pela especificação dos “modos de significação”, apresenta os atributos e valores de um objeto em função

---

<sup>10</sup> 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.



- dos **qualia**: FORMAL (de que  $x$  feito); CONSTITUTIVO (as partes de  $x$ ); TÉLICO (a função ou finalidade de  $x$ ) e AGENTIVO (como  $x$  origina-se);
- c) **estrutura de eventos**: responsável pela descrição dos eventos, estados e transições; fornece os elementos para a representação semântica dos predicados;
- d) **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; imprime 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. 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.

### Da concepção de item lexical ou unidade de acesso

De acordo com o modelo de acesso ao LM apresentado na seção anterior, uma entrada lexical é dividida em **unidade de acesso e especificação lexical**. A unidade de acesso, em especial, é a unidade abstrata que pertence ao léxico. Como bem salientam Langacker (1972) e Basílio (1999), a questão da delimitação das unidades abstratas que pertencem ao léxico é antiga e tem sido discutida sob diferentes perspectivas. Os especialistas (seja na Lingüística, na Psicolingüística ou no PLN) divergem principalmente quanto a se a representação lexical se faz por **palavras** ou por **morfemas** (radicais ou raízes).

De um modo geral, pode-se dizer que há duas concepções básicas e opostas de unidade ou item lexical. Na primeira, de cunho (psico)lingüístico, apenas as formas que servem de base para a formação de outras formas são consideradas pertencentes ao léxico, isto é, as raízes ou radicais. Dessa concepção, resulta um tipo de léxico denominado *root lexicon* ("léxico de raízes" ou "léxico de morfemas") (BUTTERWORTH, 1983). Do ponto de vista (psico)lingüístico, esse tipo de léxico é altamente econômico e eficiente, uma vez que a redundância de informações no léxico é evitada com a formação dos itens derivados por meio de processos

morfolexicais. Do ponto de vista do PLN, a implementação desse tipo de léxico, no entanto, implica a construção de um complexo módulo de processamento morfológico, composto por afixos e regras de formação de palavras, responsável pela geração dos itens derivados. Na segunda concepção, são listadas todas as formas possíveis da língua, inclusive as flexionadas; a esse tipo de léxico é dado o nome *full-form lexicon* ("léxico de formas plenas") (BUTTERWORTH, 1983). Os argumentos a favor desse tipo de léxico advêm da computação, uma vez que a grande extensão (tamanho) de um léxico de formas plenas pode ser tratável por meio das técnicas computacionais de compactação de arquivos.

Duas outras concepções, no entanto, podem ser identificadas, ambas de cunho lingüístico-computacional. Na primeira, listam-se a raiz e outras formas-base idiossincrásicas, isto é, que são empregadas em processos morfológicos não regulares. 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 õ (ajo, aja, ajas, ajamos, ajais, ajam). Pesquisas têm mostrado, no entanto, que essa concepção é viável para línguas que apresentam pouca variação ortográfica e morfológica (HANDKE, 1995). Na segunda, listam-se as chamadas formas **canônicas**. Mais especificamente, as formas pertencentes ao paradigma flexional são tidas como realizações discursivas (p.ex.: formas **embalar**, **embalou**, **embalando**) de um único item lexical canônico EMBALAR e as formas pertencentes ao paradigma derivacional (p.ex.: **embalar**, **embaladeira**, **embalado**) são tidas como itens lexicais distintos e, conseqüentemente, com entradas lexicais também distintas (LYONS, 1979).

Neste trabalho, optou-se pela segunda concepção de cunho lingüístico-computacional. Dessa forma, é possível estabelecer que **embalado**, **embalados**, **embalada** e **embaladas** têm como entrada lexical canônica<sup>11</sup> EMBALADO. A escolha por essa concepção pautou-se pelos seguintes fatores: (i) o número de entradas é reduzido, em relação a um "léxico de formas plenas", e (ii) a implementação não implica a construção de um módulo de processamento morfológico tão complexo quanto aquele que é necessário para um "léxico de raiz".

---

<sup>11</sup> Entende-se por "item lexical canônico" a forma lexical abstrata que pertence ao léxico e a partir da qual o módulo morfológico do sistema de PLN geraria as formas flexionadas.

Ainda sobre a questão da delimitação das unidades pertencentes ao léxico, ressalta-se que foram tomados como base os pressupostos do projeto WordNet (FELLBAUM, 1999). De acordo com esses pressupostos, é possível distinguir, com base nos exemplos em (3), três itens lexicais distintos, **embalado1**, **embalado2** e **embalado3**.

- (3) a. O bebê dormiu *embalado* por um barulhinho bom. (JO-FSP)  
b. Pacote *embalado* a vácuo. (JO-FSP)  
c. O São Paulo é o time mais *embalado* do campeonato. (JO-FSP)

O critério empregado para a identificação de itens distintos é o da substituição por sinônimos. De fato, **embalado1**, **embalado2** e **embalado3** podem ser substituídos, sem que haja alteração substancial do significado das sentenças em que ocorrem, por seus respectivos sinônimos: p.ex.: **ninado**, **embruhlado** e **animado**.

## Dos adjetivos valenciais do português

As duas funções sintáticas básicas desempenhadas pelos adjetivos relacionam-se com sua posição: posição adnominal (Padn) e posição predicativa (Ppred). Em Padn, os adjetivos desempenham a função denominada **adjunto adnominal** e, em Ppred, a de **predicativo**. Em função adnominal, como em (4), o adjetivo é periférico no sintagma nominal; diz-se que o adjetivo associa-se ao substantivo de forma direta.

- (4) a. O marido, homem *alto* e exato burocrata. (LI-LT)<sup>12</sup>  
b. [...] alguns deputados defendem texto da reforma *previdenciária*. (JO-FSP)

Em função predicativa, como em (5), o adjetivo é núcleo no sintagma verbal, portanto, núcleo do predicado. Mais especificamente, quando em Ppred, o adjetivo associa-se ao substantivo de forma indireta, ou seja, liga-

---

<sup>12</sup> LI-LT = Livro – Literatura.

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; BORBA, 1996; NEVES, 2000).

- (5) a. André é *alto* e sabe sair jogando com muita facilidade. (JO-FSP)  
b. O menino Luisinho que era *preto* [...]. (JO-FSP)

Do ponto de vista semântico, os adjetivos em posição adnominal ou expressam o que seria o **valor** de um atributo (do substantivo) preexistente ao julgamento do falante (FELLBAUM, 1999) ou relacionam propriedades denotativas de um substantivo com as propriedades denotativas de outro substantivo (BASÍLIO; GAMARSKI, 2002).

No primeiro caso, típico dos adjetivos **qualificadores**, postula-se a existência de um atributo A, tal que  $A(x) = \text{adj}$ , isto é, o adjetivo é o valor da função A(x). Por exemplo, dizer **homem alto** (4a) significa postular a existência prévia do atributo ALTURA, tal que  $\text{ALTURA}(\text{homem}) = \text{alto}$ . 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 qualificador em Padn ou (em função modificadora) é a de preencher o valor de um desses atributos. O termo "atributo" faz referência a um construto que pode ser igualmente denominado "qualidade", "traço" ou "dimensão". Tomando-se como ilustração o modelo de Léxico Gerativo (LG) de Pustejovsky (1996), pode-se dizer que **alto** preenche o valor do *qualia* CONSTITUTIVO do substantivo **homem**.

No segundo caso, típico dos adjetivos **classificadores**, a função semântica desempenhada é a de combinar ou relacionar propriedades denotativas de um substantivo com as propriedades denotativas de outro substantivo. Por exemplo: em **reforma previdenciária** > reforma da **previdência** (4b), observa-se que a paráfrase "da previdência" sinaliza que **previdenciária** liga a entidade **reforma** à entidade **previdência**, ou melhor, o adjetivo **previdenciária** relaciona propriedades denotativas do substantivo **reforma** e do substantivo **previdência**<sup>13</sup>.

<sup>13</sup> Alguns adjetivos classificadores, no entanto, funcionam como meros circunstanciais, p.ex.: Lílian disse que é ouvinte *regular* [= x ouve regularmente] de nossos programas [...] (JO-FSP).

Já em Pprep (5), entende-se que o adjetivo é, do ponto de vista semântico, verdadeiro predicador, expressando, assim, um **estado-de-coisas** (predicação). Sendo predicador, o adjetivo tem a propriedade de poder ligar-se a um certo número de elementos exigidos pela sua semântica, os **argumentos (As)**<sup>14</sup>. Por exemplo: em (1), **descendente** estabelece relação com dois argumentos: **o rapaz (A1)** e **de portugueses (A2)**; em (5), **alto e preto** estabelecem relações com apenas um elemento, o A1, que se realiza em “André” e “o menino Luisinho”. A essa propriedade, é dado o nome **valência** (BORBA, 1996). Dessa forma, a representação aqui proposta especifica as informações sintático-semânticas relativas à posição/função predicativa dos adjetivos.

## Da valência adjetival

A valência pode ser descrita em três níveis: **lógico-semântico, sintático e semântico**.

O primeiro diz respeito ao número de argumentos exigidos pela semântica do predicador. Os adjetivos podem ter quatro tipos de valência lógico-semântica (Quadro 2):

Quadro 2 – Tipologia da valência lógico-semântica dos adjetivos

Tipos	Exemplos
Valência 1	(6) <u>João</u> (A1) era <i>bonito</i> . (JO-FSP)
Valência 2	(7) <u>O homem</u> (A1) era <i>descendente de portugueses</i> (A2). (JO-FSP)
Valência 3	(8) <u>O rapaz</u> (A1) era <i>doador de órgão</i> (A2) <u>para transplantes</u> (A3). (JO-FSP)
Valência 4	(9) <u>A carga</u> (A2) era <i>transportável do navio</i> (A3) <u>para o cais</u> (A4) <u>pelos quindastes</u> (A1). (BORBA, 1996)

O nível da **valência sintática** (ou subcategorização) diz respeito à realização sintática dos argumentos lógico-semânticos. Em outras palavras,

<sup>14</sup> Quanto à função predicativa, especificamente, ressalta-se que esta não é exclusiva, mas sim característica dos adjetivos **qualificadores**. Os adjetivos **classificadores** comumente não aceitam a posição predicativa. No entanto, alguns deles admitem essa posição 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, p.54-55).

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, é dada a denominação **quadro de subcategorização** (RAPOSO, 1992) ou valência morfossintática (ou sintática) (BORBA, 1996). Por exemplo, em (1), o A1 de **descendente** realiza-se na forma de sintagma nominal (SN), “o rapaz”; o A2, na forma de sintagma preposicional (SPrep), “de portugueses”. Muitas vezes, os argumentos podem ocorrer sob diferentes formas sintagmáticas, p.ex.:

- (10) a. Os comerciantes maçons do Rio de Janeiro, *temerosos* com a queda do poder aquisitivo. (FI)<sup>15</sup>  
 b. Marialva partia, de repente, *temerosa* de Martim. (PN)<sup>16</sup>  
 c. [...] nós, brasileiros, conscientes e *temerosos* pela destruição da pátria naqueles idos perversos de 1963-64. (ME-O)<sup>17</sup>
- (11) a. O médico, *temeroso* de revelar o verdadeiro sentido da palavra, mente [...]. (LI-LT)  
 b. O terceiro, *temeroso* em perder o que lhe foi confiado, simplesmente enterrou o dinheiro. (JO-FSP)

Com base nos exemplos em (10) e (11), é possível observar que o A2 do adjetivo **temeroso** pode ocorrer sob as formas de SPrep e sintagma oracional (SO), sendo que o SPrep pode ser introduzido pelas preposições **com**, **de** e **por** e o SO pode ser introduzido pelas preposições **de** e **em**. Essas possibilidades devem constar do quadro de subcategorização do item predicador, no caso, do adjetivo **temeroso**.

Com relação à valência sintática, vale ressaltar que os adjetivos sempre terão um argumento subcategorizado **obrigatório**, o sujeito (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

<sup>15</sup> FI: CUNHA, F.W. *Ficção e ideologia*. Rio de Janeiro: Pongetti, 1972.

<sup>16</sup> PN: NUNES, M. R. *Noções práticas de estatística*. 2.ed. Rio de Janeiro: IBGE, 1971.

<sup>17</sup> ME-O: MÉDICI, E.A. *O jogo da verdade*. Brasília: Secretaria de Imprensa da Presidência da República, 1971.

os argumentos são omitidos, verifica-se que a significação intrínseca do adjetivo é expressa e os argumentos ausentes são interpretados de modo “subentendido” (PUSTEJOVSKY, 1996; NEWMAYER, 1998).

O nível da **valência semântica** diz respeito às relações semânticas estabelecidas entre o predicador e seus argumentos. Em outras palavras, pode-se dizer que o sentido ou valor semântico de um adjetivo valencial é determinado pela combinatória (semântica) estabelecida entre ele e seus argumentos (PUSTEJOVSKY, 1996).

No caso dos adjetivos monovalentes, o significado adjetival é determinado pela relação semântica estabelecida entre o adjetivo e seu único argumento (12).

(12) a. O garoto (A1) é *esperto*. (JO-FSP)

b. A água (A1) estava *esperta*. (JO-FSP)

Na sentença em (12a), por exemplo, em que **esperto** tem sentido “astuto”, pode-se associar o papel temático Objetivo (indica a entidade à qual se verifica uma situação) ao argumento realizado sintaticamente na forma 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, é uma **restrição seletional** imposta pelo adjetivo ao argumento. Na sentença em (12b), em que **esperto** tem sentido “quase quente”, associa-se o mesmo papel temático ao A1; 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 “astuto, finório, matreiro” (12a); quando em combinação com argumento de traço –**animado**, **esperto** tem o sentido “quase quente” (12b).

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.

(13) 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. (JO-FSP)

Em (13a), o argumento A1, realizado sintaticamente sob a forma do sintagma nominal “o som alto”, recebe o papel temático Causativo, pois a sentença relaciona-se a “o som alto causa desagrado a todos”. O sintagma preposicional “a todos”, por sua vez, recebe o papel temático Experienciador, que traduz uma experiência ou uma disposição mental. Em (13b), o A1, introduzido pela preposição **por**, recebe o papel temático Agente, pois a sentença em (13b) 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 papéis temáticos Origem (ponto de partida) e Meta (ponto de chegada).

## De alguns formalismos léxico-gramaticais

De acordo com a metodologia proposta por Dias-da-Silva (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, 1998): (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). 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 dessa seção é propor a representação lingüístico-computacional das propriedades dos adjetivos valenciais.

**Estrutura de traços ou matriz de atributo-valor:** consiste 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 singular masculino** (*F sg\_masc*), apresenta dois atributos: NÚM(ero) e GÊN(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, na Figura 1, a estrutura de traços denominada *F sg\_masc*.

$$\left[ \text{CONC} \left[ \begin{array}{l} \text{NUM} \langle \text{sg} \rangle \\ \text{GEN} \langle \text{masc} \rangle \end{array} \right] \right]$$

Figura 1 – Exemplo de representação em **matriz de atributo-valor**

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.: os traços NUM e GEN, que têm como valores terminais **sg** e **masc**, respectivamente. Uma estrutura é complexa quando contém outras estruturas de traços. Um exemplo paradigmático de estrutura de traços complexa é a descrita na Figura 1, pois a estrutura identificada como *F sg\_masc* contém outras estruturas, no caso, CONC, NUM e GEN. As principais características das **estruturas de traços** é poder combinar informações; a operação ou mecanismo responsável por essa combinação 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 descrita na Figura 2:

$$\left[ \begin{array}{l} \text{NUM} \langle \text{sg} \rangle \\ \text{PES} \langle 3 \rangle \end{array} \right]$$

Figura 2 – Exemplo de estrutura de traços resultante da operação de Unificação

**Lógica de predicados:** 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. 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** (ALLWOOD; ANDERSON; DAHL, 1977). A seguir, ilustra-se em (14) a estrutura de predicados do verbo **comprou** (DIK, 1997).

$$(14) \text{COMProu} \downarrow (x_1: \langle \text{anim} \rangle (x_1))_{Ag} (x_2)_{Ob}$$

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  $x_1$  e  $x_2$ . Observa-se que, no exemplo em (14), o argumento representado por  $x_1$  possui uma **restrição**, que está representada entre os símbolos “< >”. No caso, o argumento  $x_1$  precisa ser do tipo “animado”.

**Frames:** 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 *frames* 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**. Exemplifica-se uma estrutura de *frame* em (15) tomando por base a proposta de Nirenburg (1992):

(15) (vermelho

(vermelho-adj1

(CATEGORIA adj)

(ESTR-SINT

(1 ((raiz \$var1)

(cat n)

(mods ((raiz \$var0))))))

(2 ((raiz \$var0)

(cat adj)

(subj ((raiz \$var1)

(cat n))))))

Em (15), representa-se o adjetivo **vermelho**, que possui apenas um sentido (vermelho-adj1), e seus **atributos**: CAT(egoria), ESTR(utura)-SINT(ática). 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, a função de adjunto adnominal, e o segundo valor especifica a construção N-cópula-Adj, ou seja, a função predicativa do adjetivo em questão.

## Do formalismo para a representação dos adjetivos valenciais

Dos formalismos descritos na seção anterior, optou-se por empregar, na representação lingüístico-computacional dos adjetivos valenciais, a **estrutura de traços** ou **matriz atributo-valor**. Essa escolha pautou-se basicamente pelos seguintes fatores: (a) 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; (b) 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; (c) 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; KOENIG, 1999).

## Da representação lingüístico-computacional dos adjetivos valenciais

Partindo-se, por um lado, dos componentes previstos no esquema de representação da informação lexical e das propriedades dos adjetivos valenciais e, por outro lado, da representação por meio de **matrizes de atributo-valor**, propõe-se um modelo de entrada lexical canônica para os adjetivos valenciais. Para ilustrar esse modelo, apresenta-se, na Figura 3, a entrada lexical canônica de **temeroso**.

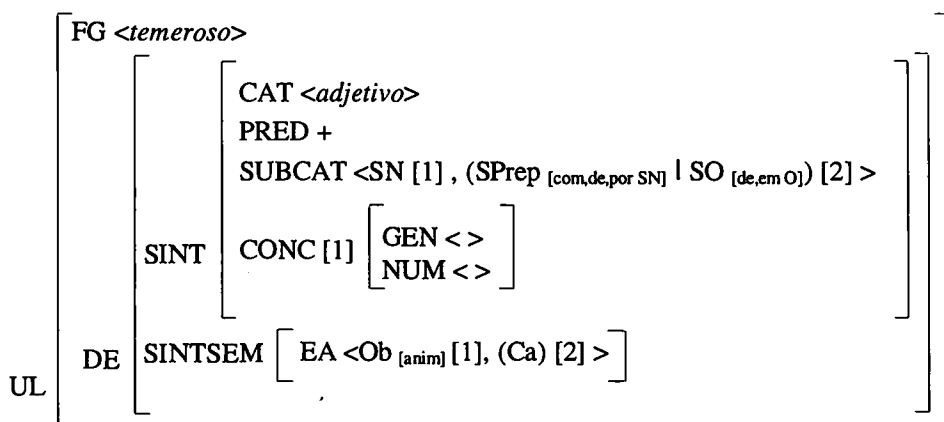


Figura 3 – Modelo de entrada lexical canônica para os adjetivos valenciais

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 1. Cabe ressaltar, especificamente, 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 os **valores** dos traços são representados com letras minúsculas e entre os símbolos "<>". A seguir, descreve-se detalhadamente e exemplifica-se, com base na Figura 3, cada tipo de informação lexical contida nesse modelo de entrada canônica.

## 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**, 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.

**A estrutura de traço simples FG:** abrange parte das informações previstas no componente FG do esquema de representação da informação lexical (cf. Quadro 1). Mais especificamente, 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.: temeroso, temerosos, temerosa, temerosas). Tais formas são consideradas, neste trabalho, realizações discursivas de um mesmo item lexical (p.ex.: TEMEROSO). Cabe ao traço FG, então, especificar a forma gráfica do **item lexical** (representado pela forma canônica, p.ex. <temeroso>).

**A estrutura de traços complexa DE:** é formada por outras duas estruturas de traços complexas: a estrutura SINT (“sintática”) e a estrutura SINTSEM (“sintático-semântica”). A estrutura SINT abrange as informações previstas no componente TG do esquema de representação lexical (Quadro 1). 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 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 pelo símbolo “+”. O traço SUBCAT especifica o **quadro de subcategorização** dos adjetivos, sendo que os argumentos são

representados segundo a ordem linear em que normalmente aparecem nas sentenças. O adjetivo **temeroso**, por exemplo, pode subcategorizar dois argumentos, A1 e A2; o A1 realiza-se sob a forma de SN e o A2, sob a forma de SPrep [com/de/por SN] ou SO (sintagma oracional) [de/em O]. Para especificar todas as categorias sintagmáticas nas quais o argumento A2 projetado pelo adjetivo em questão possa se realizar, foi inserido 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. Além disso, ressalta-se que o argumento que se realiza somente na forma de SN, o sujeito, é obrigatório, sendo que o argumento 2 é facultativo. A informação de que determinado argumento subcategorizado é facultativo ou opcional está indicada pelos parênteses "()". O traço CONC, por sua vez, abrange as **categorias gramaticais secundárias** de gênero (GEN) e de **número** (NUM), atribuídas aos adjetivos por meio de regras de concordância. Ressalta-se que os valores dos traços NUM e GEN são herdados do substantivo pelo adjetivo; essa herança acontece no momento em que as propriedades do adjetivo são **unificadas** às propriedades dos itens com os quais ocorre na sentença sob análise.

**A estrutura SINTSEM:** abrange as informações previstas pelo componente EA. A denominação SINTSEM foi dada como consequência de se considerar a estrutura de argumentos como 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 seletivas**. No caso de **temeroso**, a estrutura SINTSEM especifica que o argumento marcado com o índice [1] na estrutura SUBCAT recebe o papel temático Ob e necessita ser do tipo **+animado** e que o argumento marcado com o índice [2] recebe o papel temático Ca. Ressalta-se que os índices numéricos presentes das estruturas SUBCAT e SINTSEM são responsáveis por "ligar" ou "associar" um papel temático a um determinado argumento subcategorizado.

## Considerações finais

Apresentou-se, neste trabalho, uma proposta de representação lingüístico-computacional para os adjetivos valenciais do português do Brasil. Diz-se “lingüístico” porque tal representação baseou-se em princípios (psico)lingüísticos sobre o léxico mental. Diz-se “computacional” porque as propriedades desses adjetivos foram explicitamente especificadas por meio de um esquema de representação formal de modo que os sistemas de PLN possam processá-las. Salienta-se que a representação proposta caracteriza-se por agrupar informações de natureza morfossintática e semântica, tais como: **realização gráfica; categoria gramatical; tipo sintático; esquema de subcategorização; valência semântica** (estrutura de argumentos); **papéis temáticos; restrições seletivas; categorias de número e gênero**. Dessa forma, essa representação dá ao léxico um papel de centralidade nos sistemas de PLN. Tal centralidade, inclusive, segue a tendência dos modelos gramaticais **lexicalistas**, em que o léxico engloba uma carga considerável de informações a ponto de gerar a “sintaxe”. Por fim, ressalta-se que trabalhos como este contribuem para a construção de sistemas de PLN lingüisticamente motivados e capazes de simular aspectos da competência e do desempenho lingüísticos humanos e para a proposição de modelos lingüísticos mais completos, explícitos e, conseqüentemente, mais operacionais.

Esta pesquisa sobre a valência adjetival resultou, aliás, na proposta de inserção da mesma na base da Wordnet.Br (DI FELIPPO; DIAS-DA-SILVA, 2004). Nessa base, unidades lexicais do português (substantivos, verbos, adjetivos e advérbios) estão organizadas em termos de conjuntos de sinônimos que expressam conceitos lexicalizados. Atualmente, estão sendo inseridas informações periféricas que são associadas a cada conceito (sentido) armazenado, são elas: frases-exemplo e glosas (definições informais). A proposta de inserção formal da valência consiste basicamente: (i) na associação das valências lógico-semântica (quantitativa) e semântica aos conceitos ou sentidos expressos pelos adjetivos e (ii) na associação da valência sintática às frases-exemplo. Assim, com esse acréscimo, a base Wordnet.Br, que já é um importante recurso lexical para pesquisadores do PLN e lingüistas, poderá ser enriquecida.

DI FELIPPO, A.; DIAS-DA-SILVA, B.C. The valency frames of Brazilian Portuguese adjectives: a computational linguistic model. *Revista do GEL*, Araraquara, v.2, p.55-81, 2005.

■ **ABSTRACT:** *Considering urgent the demand for linguistically-motivated lexicons for Natural Language Processing systems, this paper proposes a computational linguistic model for the valency frames of Brazilian Portuguese adjectives. This representation model emerges from the analysis of (i) the mental lexicon microstructure, (ii) the adjective valency frames, and (iii) the lexical-grammatical formalisms relevant to represent such type of grammatical information. The paper concludes with the description of a specific feature structure-based template for adjective lexical entries.*

■ **KEYWORDS:** *Adjective. Valency. Lexical semantics. Lexicon. Natural language processing.*

## Referências

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

ANGELO, I. **A festa**. 5.ed. São Paulo: Summus, 1978.

BASÍLIO, M. Introdução: questões clássicas e recentes da delimitação das unidades lexicais. *Palavra*, Rio de Janeiro, n.5, p.9-18, 1999.

BASÍLIO, M.; GAMARSKI, L. Adjetivos denominais no português falado. In: CASTILHO, A.T. (Org.). **Gramática do português falado**. 2.ed. Campinas: Unicamp, 2002. v.4, p.629-650.

BIDERMAN, M.T.C. Conceito lingüístico de palavra. *Palavra*, Rio de Janeiro, n.5, p.81-97, 1999.

BIERWISCH, M.; SCHREUDER, R. From concepts to lexical items. **Cognition**, Amsterdam: Elsevier Science Publisher B.V., n.42, p.23-60, 1992.



BORBA, F.S. **Uma gramática de valências para o português**. São Paulo: Ed. Ática, 1996.

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

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

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

CUNHA, F.W. **Ficção e ideologia**. Rio de Janeiro: Pongetti, 1972.

DIAS-DA-SILVA, B.C. Bridging the gap between linguistic theory and natural language processing. In: INTERNATIONAL CONGRESS OF LINGUISTICS, 16<sup>th</sup>, 1998, Oxford. **Proceedings...** Oxford: Elsevier Sciences, 1998. p.1-10.

DI FELIPPO, A.; DIAS-DA-SILVA, B.C. Edição de informações sintático-semânticas dos adjetivos na base da rede Wordnet.Br. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO-SBC, 24., 2004, WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA-TIL, 2., 2004, Salvador. **Anais...** Salvador, 2004. v.1. 1 CD-ROM.

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

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

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

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

LANGACKER, R. **Fundamentals of linguistic analysis**. New York: Harcourt, Brace, Jovanovich, 1972.

LEVELT, W.J.M. Accessing words in speech production: stages, processes and representations. **Cognition**, Amsterdam: Elsevier Science Publisher B.V., n.42, p.1-22, 1992.

LEVELT, W.J.M. **Speaking: to intention to articulation.** Cambridge, Mass.: The MIT Press, 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. **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.

\_\_\_\_\_. **Semantics.** Cambridge: Cambridge University Press, 1977. v.2.

MÉDICI, E.A. **O jogo da verdade.** Brasília: Secretaria de Imprensa da Presidência da República, 1971.

NEVES, M.H.M. **Gramática de usos do português.** São Paulo: Editora UNESP, 2000.

\_\_\_\_\_. **A gramática funcional.** São Paulo: Martins Fontes, 1977.

NEWMAYER, F.J. **Language form and language function.** Cambridge, Mas.: The MIT Press, 1998.

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

NUNES, M.R. **Noções práticas de estatística.** 2.ed. Rio de Janeiro: IBGE, 1971.

PALMER, M. Multilingual resources – Chapter 1. In: HOVY, E. et al. (Eds.). **Multilingual information management: current levels and future abilities,** 1999. Disponível em: <<http://www.cs.cmu.edu/~ref/mlim/>>. Acesso em: 1 abril 2005.

POLLARD, C.; SAG, I.A. **Head-driven phrase structure grammar.** Chicago: University of Chicago Press, 1994.

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

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. In: **TECHNICAL Reporty Computing Research Laboratory**. New Mexico State: NMSU, 1995. MCCS-288.

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

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

SHIEBER, S.M. **An introduction to unification-based approaches to grammar**. Number 4 in CSLI Lecture Notes. Stanford, CA: CSLI, 1986.

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