

Edição de informações sintático-semânticas dos adjetivos na base da rede Wordnet.Br

Ariani Di Felippo, Bento Carlos Dias-da-Silva¹

¹Centro de Estudos Lingüísticos e Computacionais da Linguagem (CELiC)

FCL – UNESP – Caixa Postal 174 – Araraquara – SP – Brazil

Núcleo Interinstitucional de Lingüística Computacional (NILC)

ICMC – USP – Caixa Postal 668 – São Carlos – SP – Brazil

{arianidf@uol.com.br,bento@fclar.unesp.br}

Abstract. *This paper proposes the formal inclusion of the “argument structure” and the “subcategorization frame” of adjectives in the Wordnet.Br lexical database. The conclusion outlines such an extension.*

Resumo. *Neste trabalho, propõe-se a inclusão da “estrutura de argumentos” e do “esquema de subcategorização” dos adjetivos na base da rede Wordnet.Br. A conclusão esquematiza essa proposta de extensão.*

1. Introdução

A partir do desenvolvimento da rede WordNet¹ da Universidade de Princeton [Fellbaum 1999], EUA, vários países construíram ou estão construindo suas próprias *wordnets*, dada a importância desse tipo de base lexical na compilação de parcelas de léxicos para o desenvolvimento de diversos sistemas de processamento automático de línguas naturais (PLN). 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, a rede WordNet registra informações periféricas, associadas a cada sentido armazenado. São elas: frases-exemplo e glosas (isto é, definições informais).

A base da rede *wordnet* brasileira (doravante, Wordnet.Br), em desenvolvimento a partir do aplicativo *Thesaurus Eletrônico para o Português do Brasil – TeP* [Dias-da-Silva, 2003], apresenta um total de 17.416 substantivos, 11.078 verbos, 15.073 adjetivos e 1.139 advérbios, estruturados em função das relações de *sinonímia* e *antonímia* [Dias-da-Silva et al. 2002; Dias-da-Silva 2003; Dias-da-Silva e Moraes 2003]. Na fase atual de desenvolvimento da base lexical da Wordnet.Br, estão sendo feitas a análise de

¹ O nome da rede americana é grafado com “N” maiúsculo para diferenciá-la das demais, caracterizando-a, como diz Fellbaum, como “a mãe de todas as Wordnets”, construída para essa variante do inglês. Atualmente, várias comunidades de PLN já possuem seus aplicativos no formato *wordnet*. Dentre eles, citam-se as redes originalmente propostas para integrar o núcleo da EuroWordNet: as redes para o inglês britânico, holandês, espanhol, italiano, alemão, sueco, francês, tcheco e estoniano. Recentemente, também em fase de construção, cita-se a rede Wordnet.Pr, para o português europeu [Marrafa 2001].

consistência semântica dos synsets e a coleta e seleção das frases-exemplo². Para os pesquisadores do PLN, a base da Wordnet.Br possibilita, por exemplo, a geração de parcelas de léxicos especiais, munidos de conhecimento léxico-semântico, imprescindíveis para o desenvolvimento de diversos sistemas de PLN, tais como: sistemas de tradução automática, de sumarização automática, entre outros [Briscoe e Boguraev 1989], [Saint-Dizier e Viegas 1995], [Dias-da-Silva 1998] e [Palmer 2001]. Ao usuário da língua portuguesa, por sua vez, a base da Wordnet.Br, acoplada a ferramentas computacionais de auxílio à escrita, deverá oferecer a opção de seleção *on line* de palavras sinônimas e antônimas que, por motivos de estilo, precisão, adequação comunicativa, correção ou aprendizagem, o usuário queira substituir [Ilari e Geraldi 1985].

Neste trabalho, em particular, propõe-se a inserção, nessa base, de informações sobre a função de *predicador* do adjetivo do português brasileiro. Nas seções subseqüentes, delineiam-se (i) a motivação (psico)lingüística para a inclusão desse tipo de informação na base da rede Wordnet.Br, (ii) a estrutura atual da base, (iii) o tratamento dado aos adjetivos no projeto Wordnet.Br e (iv) as informações relativas à função de *predicador* que poderão ser especificadas na base. Na conclusão, esquematisa-se a extensão resultante do acréscimo dessas informações.

2. Da motivação (psico)lingüística

Os estudos realizados no domínio da Psicolingüística têm contribuído consideravelmente para a construção de léxicos lingüístico-computacionais [Handke 1995]. A seguir, são feitas considerações a respeito da estrutura global e interna do *léxico mental*³ com o objetivo de delimitar os subsídios (psico)lingüísticos para a proposta de extensão da base da Wordnet.Br.

2.1. Da Macroestrutura do “léxico mental”

O *léxico mental* (LM) 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 [Levelt 1993] e as associações estabelecidas entre os itens lexicais distribuem-se em associações (i) paradigmáticas e (ii) sintagmáticas [Biderman 1981].

(1) *Das relações paradigmáticas*: diz-se que as unidades lexicais pertencem ao mesmo paradigma quando uma puder ser substituída pela outra em um mesmo ponto da cadeia sintagmática; tais relações podem ser: a) morfossemânticas: relações entre os itens que apresentam a mesma raiz (p.ex.: embalar, embalado, embalador); b) léxico-conceituais: relações que se estabelecem entre conceitos lexicalizados: a *hiponímia* (p.ex.: laranja é hipônimo de árvore) e a *hiperonímia* (p.ex.: árvore é hiperônimo de laranja); c) léxico-semânticas: relações que se estabelecem entre unidades (= formas) lexicais e não entre conceitos; são elas: as relações de *sinonímia* (p.ex.: asfixiado, sufocado) e as de *antonímia* (p.ex.: grande/pequeno).

² CNPq 09/2001- Processo N° 552057/01-0.

³ Entende-se por *léxico mental* a parte do conhecimento lexical do indivíduo delimitada pela sua língua [Bierwisch e Schreuder 1992], [Levelt 1993].

(2) *Das relações sintagmáticas*: são relações resultantes da combinatória freqüente entre itens lexicais; a principal relação sintagmática da macroestrutura do léxico é a *colocação* (do Inglês: “*collocation*”). O termo *colocação* é aqui entendido como a relação que se verifica entre seqüências de unidades lexicais que co-ocorrem habitualmente como, por exemplo, *feliz aniversário*, *dias de sol*.

De acordo com o modelo de processamento cognitivo da linguagem de Levelt (1992, 1993), tais relações podem ser consideradas *diretas*, e, conseqüentemente, representadas no interior das entradas do LM. Uma relação léxico-semântica “direta”, por exemplo, é aquela em que os sinônimos de um item *x* são listados na entrada lexical de *x*. Essa abordagem, inclusive, é a adotada no modelo *wordnet*.

2.2. Da Microestrutura do “léxico mental”

Além das informações referentes à *macroestrutura* do léxico, as quais podem (ou não) compor as entradas do LM, ressalta-se que a entrada lexical (E) de um item *x* armazena os *lemas* e os *lexemas*, que estão interligados por um ponteiro lexical, isto é, cada *lema* “aponta” para seu *lexema* correspondente (Levelt 1993).

Os **lemas** são representações das propriedades semânticas e sintáticas dos itens lexicais. Observe-se que não se trata, aqui, do sentido em que esse termo é empregado no âmbito da Lexicografia, isto é, a forma canônica de uma unidade lexical. A forma *entrar*, em *João entrou na casa*, pertence à classe V(erbo) e projeta uma estrutura semântica de dois argumentos: [(Agente <*anim*>) (Meta <*loc*>)]; esses argumentos se realizam sintaticamente como um SN (*João*) e um SPrep (*na casa*). A construção de um SV (p.ex.: *entrar na casa*) ou de uma sentença (*O gato entrou na casa*) depende da informação sintática contida no *lema*.

Já os **lexemas** são representações das estruturas morfológica e fonológica das unidades lexicais. Por exemplo, o *lexema* de *entrar* especifica que esse item é formado pelos seguintes segmentos morfológicos: o radical /entr-/, a vogal temática /-a/ e a flexão /-r/; e pelos cinco segmentos fonológicos: /eN/, /t/, /r/, /a/ e /R/. Dessa forma, pode-se dizer que, do ponto de vista (psico)lingüístico, as informações semânticas armazenadas na base da rede Wordnet.Br são referentes às relações que se estabelecem entre os itens do LM. Essa rede equaciona, do ponto de vista computacional, parte das relações semânticas responsáveis pela estrutura global ou *macroestrutura* do LM.

A extensão dessa base aqui proposta consiste na inserção da informação sintático-semântica de *predicador* que parte dos adjetivos da língua desempenha. Nesse sentido, além de abrigar as relações léxico-semânticas que se instauram entre os adjetivos, ela deverá também abrigar as informações sobre o *lema* de cada adjetivo predicador, ou seja, informações responsáveis pela *microestrutura* do LM [Bierwisch e Schreuder 1992], [Handke 1995].

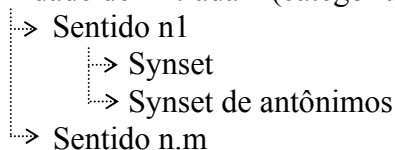
3. Da estrutura atual da base da Wordnet.Br

Como mencionado, a elaboração da base da Wordnet.Br teve como ponto de partida o TeP, que foi elaborado segundo os princípios da rede americana. Da metodologia proposta por Miller e Fellbaum (1991) para a construção da rede WordNet, foram utilizadas três noções básicas no desenvolvimento do Tep [Dias-da-Silva et al 2002]: (i) o *método diferencial*, que pressupõe o princípio de ativação de conceitos por meio de

um conjunto de formas lexicais relacionadas pela sinonímia, o que elimina a necessidade de especificação do valor semântico para o sentido de uma entrada lexical; (ii) a noção constitutiva de *synset*, isto é, o conjunto de sinônimos; (iii) a noção de *matriz lexical*, que postula uma correspondência biunívoca entre sentido e *synset*.

Com essa metodologia, a relação de sinonímia passa a ser 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). O Esquema 1 ilustra a estrutura da base do TeP e, conseqüentemente, da Wordnet.Br:

(1) Unidade de Entrada n (categoria sintática x)



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

4. Do tratamento dado aos adjetivos na Wordnet.Br

Para a compilação dos conjuntos de sinônimos e antônimos de adjetivos do TeP, partiu-se do princípio de que os adjetivos do português, assim como os do inglês e do espanhol, podem ser divididos em duas classes: os *qualificadores* (QLs) e os *classificadores* (CLs) [Quirk et al 1991], [Borba 1996], [Demonte 1999] e [Neves 2000].

Os **qualificadores** indicam o valor de uma propriedade ou atributo do substantivo com o qual se liga. Dessa forma, dizer “ X QL” ou “ X é QL” pressupõe um atributo A , tal que $A(x) = QL$. Por exemplo, dizer “torre *alta*” ou “a torre é *alta*”, pressupõe um atributo ALTURA, tal que $ALTURA(torre) = alta$. Além disso, pressupõe-se que: (i) os atributos são *bipolares*, isto é, os adjetivos *alto* e *baixo* são antônimos e expressam os valores dos pólos do atributo ALTURA; (ii) atributos podem ser *graduáveis* (contínuos) ou *não-graduáveis* (dicotômicos), por exemplo: o atributo graduável ALTURA varia em um contínuo de “alturas” entre os valores polares “alto” e “baixo”, isto é, os valores dos extremos do atributo ALTURA; já o atributo SEXO, não-graduável, apresenta apenas dois valores: *macho* e *fêmea*; e, por isso, são denominados *dicotômicos*. Dentre as principais características dos QLs estão [Casteleiro 1981], [Gross, Fischer e Miller 1995]: (i) a nominalização da propriedade que expressam (p.ex.: muro *alto* $\rightarrow$ a *altura* do muro); (ii) a gradação (p.ex.: o muro é muito *alto*); (iii) a comparação (p.ex.: a torre é mais *alta* do que a pirâmide).

Já os **classificadores** colocam a denotação do substantivo com o qual ocorrem numa subclasse, nomeando-a, p.ex.: o adjetivo classificador *cambial* no SN *a reforma cambial*. Observe-se que a paráfrase *do câmbio* sinaliza que *cambial* liga a entidade “reforma” a outra, exterior a ela: o “câmbio” [Borba 1996], [Neves 2000], [Basílio e

Gamarski 2002]. Enquanto os QLS possuem as propriedades descritas em (i), (ii) e (iii), o mesmo não ocorre com os CLs (p.ex.: *a cambiabilidade da reforma/ *a reforma é muito cambial/ *a reforma é mais cambial do que a crise.)⁴. 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 e Fellbaum 1991].

5. Da sistematização das “novas” informações sobre os adjetivos

As duas funções sintático-semânticas básicas dos adjetivos relacionam-se com sua posição: **posição adnominal** (Padn) e **posição predicativa** (Ppred). Tanto os QLS quanto os CLs podem ocorrer em Padn; a Ppred, no entanto, não é exclusiva, mas sim característica da subclasse dos QLS. Isso porque há certos CLs que admitem a função predicativa quando em condições contextuais específicas [Casteleiro 1981]: (i) com 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*). Os CLs, que comumente ocorrem apenas em Padn, são, na verdade, complementos dos substantivos (p.ex.: câmara municipal > câmara do município) com os quais ocorrem ou têm valores adverbiais (p.ex.: matador profissional).

Focalizando os QLS, ressalta-se que esses, quando em Padn, expressam o valor de atributo *preexistente* ao julgamento do falante (p.ex.: rapaz *pobre*). 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. Os QLS, quando em Ppred, instauram verdadeiro processo de *predicação* (p.ex.: Aquele rapaz é *pobre*).

5.1. Os adjetivos *predicadores* ou *valenciais*

Sendo, portanto, **predicador** (PR), o QL designa um “estado-de-coisas”, isto é, algo que pode ocorrer em algum mundo (real ou mental) (Dik, 1997). No interior da predicação, estabelecem-se as propriedades ou relações especificadas pelo PR. A *valência*, então, pode ser entendida como a relação abstrata do PR com os argumentos (As) que dele dependem [Neves 1996].

O termo *valência* pode ser usado em três níveis: *valência lógico-semântica*, *valência sintática* e *valência semântica*.

a) Quanto à *valência lógico-semântica* dos adjetivos

Sendo o nível mais abstrato, diz respeito ao número de As que um PR pode ter. Há duas interpretações possíveis para a valência adjetival; na primeira, consideram-se apenas os constituintes diretamente dependentes dos adjetivos; na segunda, considera-se, como argumento adjetival, o constituinte em função de sujeito [Busse e Vilela 1986]. Neste trabalho, optou-se pela segunda interpretação. Dessa forma, considera-se que os adjetivos PRs ou valenciais do português podem ser de quatro tipos [Borba 1996], como descrito na Tabela 1.

⁴ O símbolo “*” indica agramaticalidade.

Tabela 1. Valência lógico-semântica: tipologia

Tipologia	Descrição e exemplificação
Valência 1 (V₁)	Projeta um argumento lógico-semântico. P.ex.: <u>Meu pai</u> (A1) era <i>alto</i> , loiro e de olhos azuis.
Valência 2 (V₂)	Projeta dois argumentos lógico-semânticos. P.ex.: Renunciou à convicção porque <u>ela</u> (A1) não era <i>útil</i> <u>a seus propósitos</u> (A2).
Valência 3 (V₃)	Projeta três argumentos lógico-semânticos. P.ex.: <u>O réu</u> (A2) era <i>condenável</i> <u>à morte</u> (A3) pelo juiz. (A1).
Valência 4 (V₄)	Projeta quatro argumentos lógico-semânticos. P.ex.: <u>A carga</u> (A2) era <i>transportável</i> do <u>estaleiro</u> (A3) <u>para o navio</u> (A4) <u>por guindastes</u> (A1).

b) Quanto à valência sintática dos adjetivos

Esse nível trata da função sintática e/ou da categoria sintagmática (e/ou morfossintática) dos As realizados na sintaxe, p.ex.: em *Aquele rapaz é pobre*, o adjetivo *pobre* requer (ou projeta) um argumento lógico-semântico (A1), que, sintaticamente, realiza-se sob a forma do sintagma nominal sujeito (“aquele rapaz”). A valência sintática também pode ser entendida como **esquema de subcategorização** [Raposo 1992]. Salienta-se que nem todos os As lógico-semânticos são realizados na sintaxe. No caso dos adjetivos valenciais (ou seja, aqueles em Ppred), observa-se que eles partem sempre de um índice *I*, o sujeito. Os demais argumentos são opcionais.

c) Quanto à valência semântica dos adjetivos

Esse nível trata das relações semânticas que se estabelecem entre o PR e os As. A valência semântica é também designada **estrutura de argumentos** [Grimshaw 1992]. Mais especificamente, são observadas, nesse nível, as funções temáticas (= papéis) dos As e as restrições seletivas que o PR impõe sobre seus argumentos. Por exemplo: em *Paulo era descendente de italianos*, o adjetivo *descendente* projeta dois argumentos lógico-semânticos, que se realizam sob a forma de SN (“Paulo”) e de SPrep (“de italianos”) e que, do ponto de vista semântico, recebem papel temático *Objetivo* (“Paulo”) e *Origem* (“de italianos”), sendo que ambos precisam ser do tipo semântico <humano> [Borba 1996].

6. Conclusão

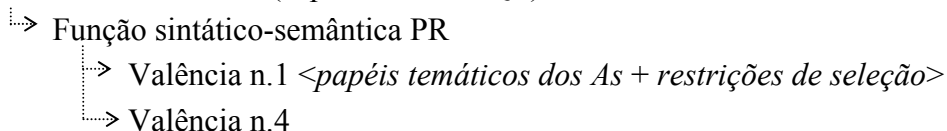
A partir da breve exposição sobre os adjetivos valenciais do português, propõe-se a ampliação da estrutura original da base da Wordnet.Br para que as seguintes informações possam ser inseridas: (i) a valência lógico-semântica; (ii) a valência sintática ou esquema de subcategorização; (iii) a estrutura de argumentos ou valência semântica.

6.1 Da extensão da informação-base da Wordnet.Br

Em suma, a classificação esboçada na seção anterior permite estender a informação original (Esquema 1) relativa aos QLS da Wordnet.Br com os seguintes tipos: QL = {PR} = {Adj_V1} + {Adj_V2} + {Adj_V3} + {Adj_V4}. Assim, paralelamente ao

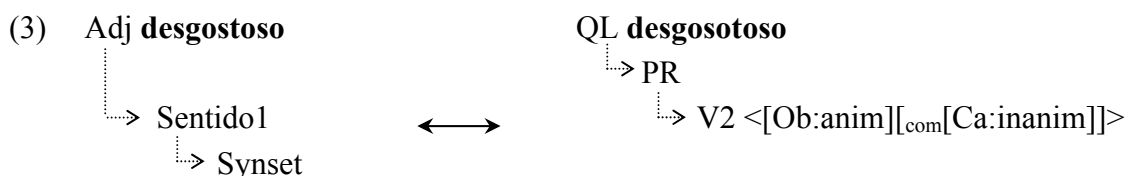
Esquema 1, propõe-se que aos QL da base da Wordnet.Br estejam associadas informações sobre a *valência lógico-semântica* e sobre a *estrutura de argumentos* ou *valência semântica*. O Esquema 2 exemplifica essa extensão.

(2) Unidade de Entrada n (Tipo semântico QL)



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” indicam qual o subtipo lógico-semântico, (ou seja, V1, V2, V3 e V4), ao qual é associada a *valência semântica* (papéis temáticos + restrições seletivas) do adjetivo propriamente dito.

Salienta-se que indexações apropriadas deverão 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 deverão permitir o relacionamento entre *valência* e *sentido*, este armazenado na base da Wordnet.Br em função dos 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. Ao adjetivo *desgostoso*, no Sentido1, por exemplo, está associada a frase “O príncipe Charles está desgostoso e escandalizado com as acusações contra sua mulher, a princesa Diana”. A partir da 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).

6.2 Da extensão da informação periférica da 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 da 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”.

Assim, com esse acréscimo, a manipulação da base da rede Wordnet.Br poderá gerar listas de formas, para a compilação de léxicos monolíngües, que, além de fornecerem as relações (léxico-semânticas) que se instauram entre os adjetivos, fornecerão também sua valência e seu esquema de subcategorização, isto é, informações sobre o *lema* do adjetivo, ou seja, sobre propriedades sintático-semânticas responsáveis pela *microestrutura* do léxico da língua [Bierwisch e Schreuder 1992], [Handke 1995].

Referências

- Basílio, M. e 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.
- Bierwisch, M. e Schreuder, R. (1992) “From concepts to lexical items”. *Cognition*, 42, p.23-60.
- Biderman, M. T. C. “A estrutura 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.
- Borba, F. S. “Uma gramática de valências para o português”. São Paulo, Editora Ática, 1996.
- Busse, W. e Vilela, M. “Gramática de valências”. Coimbra, Almedina, 1986.
- Briscoe, E. J. e Boguraev, B. (eds) “Computational lexicography for natural language processing”. London/New York, Longman, 1989.
- Casteleiro, J. M. “A sintaxe transformacional do adjetivo”. Lisboa, Instituto Nacional de Investigação Científica, 1981.
- Demonte, V. (1999) “Semántica composicional y gramática: los adjetivos en la interfície léxico-sintaxis”. *Revista Española de Lingüística*, 29, v.2, p. 283-316.
- Dias-da-Silva, B. C. (1998) “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.
- Dias-da-Silva, B. C. (2003) “O TeP: construção de um thesaurus eletrônico para o português do Brasil”. *Boletim da Associação Brasileira de Lingüística (ABRALIN)*.

- Fortaleza: Imprensa Universitária, v.26, número especial, p.86 - 89.
- Dias-da-Silva, B. C. e Moraes, H. R. (2003) "A construção de thesaurus eletrônico para o português do Brasil". *Alfa*, v.47, n.2, p.101 - 115.
- Dias-da-Silva, B. C., Oliveira, M. F. e Moraes, H. R. (2002) "Groundwork for the development of the Brazilian Portuguese Wordnet". *Advances in natural language processing*. Berlin, Springer-Verlag, p.189-196.
- 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.
- Grimshaw, J. "Argument structure". Cambridge, The MIT Press, 1992.
- Gross, D., Fischer, U. e Miller, A. (1989) "The organization of the adjectival meaning". *Journal of Memory and Language*, 28, p. 92-106.
- Handke, J. "The structure of the Lexicon: human versus machine". Berlin, Mouton de Gruyter, 1995.
- Ilari, R. e Geraldi, J. W. "Semântica". São Paulo, Editora Ática, 1985.
- Levelt, W. J. M. (1992) "Accessing words in speech production: stages, processes and representations". *Cognition*, 42, p.1-22.
- _____. "Speaking: to intention to articulation". Cambridge, The MIT Press, 1993.
- Lyons, J. "Language and linguistics. An introduction". Cambridge, CUP, 1981.
- Marrafa, P. (2001) "WordNet do Português – Uma base de dados de conhecimento lingüístico". Lisboa: Instituto Camões, 2001.
- Mel'čuk, I. "Dependency Syntax: theory and practice". The SUNY Press, Albany, N.Y, 428p. 1988.
- Miller, G. A. e Fellbaum, C. (1991) "Semantic networks of English". *Cognition*, v.41, n.1-3, p.197-229.
- Neves, M. H. M. "Gramática de usos do português". São Paulo, Editora UNESP, 2000.
- _____. "Estudo da estrutura argumental dos nomes". In: Kato, M. A. (org.) *Gramática do Português Falado V: Convergências*. Campinas: Ed. Unicamp/FAPESP, p. 119-154, 1996.
- Palmer, M. (2001) "Multilingual resources – Chapter 1". In: Hovy, E., et al. (eds.). *Linguistica Computazionale*, v.14-15.
- 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.
- Saint-Dizier, P. e Viega, E. "Computational lexical semantics". Cambridge, Cambridge University Press, 1995.
- Vossen, P. et al. (1998). "The EuroWorNet base concepts and top ontology". <<http://www.illc.uva.nl/EuroWordNet/docs.html>>. Fev. 2003.