

## CO-INDEXAÇÃO SEMÂNTICA ENTRE VERBOS DO PORTUGUÊS E DO INGLÊS. Aline Camila Lenharo, Bento Carlos Dias-da-Silva. – Lingüística – Letras – Departamento de Letras Modernas – Faculdade de Ciências e Letras – Campus Araraquara.

Este trabalho discute a co-indexação entre parte dos synsets, “conjunto de sinônimos” (do inglês *synonym sets*) de verbos do português e parte dos synsets do inglês, por meio do alinhamento dos synsets semanticamente equivalentes das duas línguas e da especificação de glosas, isto é, definições que explicitam o conceito lexicalizado e compartilhado pelos verbos que formam cada synset da base da rede WordNet.Br (WN.Br). Por exemplo, a glosa do português “cobrir com telhas” define o conceito lexicalizado e compartilhado pelos verbos do synset {atellar, telhar} e a glosa correspondente do inglês “*cover with tiles*”, o conceito lexicalizado pelo verbo {tile} (DIAS-DA-SILVA; LENHARO, 2006; LENHARO, 2006).

A partir da proposição dessas definições, do exame de definições lexicográficas e do estudo da metodologia de construção de wordnets, em especial, da WordNet de Princeton (WN), analisam-se os tipos de co-indexação que se estabelecem entre os synsets de verbos das duas línguas (MILLER; FELLBAUM, 1991; VOSSSEN, 1998; MARRAFA, 2001; DIAS-DA-SILVA et al, 2002; DIAS-DA-SILVA, 2006, 2004).

A co-indexação mais direta (a EQ\_SYNONYM) verifica-se entre synsets semanticamente equivalentes nas duas línguas, como o exemplo mencionado, no qual, o índice 01299130 do ILI –isto é, o Índice Interlíngüe (*Inter-Lingual-Index*), composto pela coleção de synsets da WN– relaciona as wordnets do português e do inglês. Esta é a relação semanticamente mais direta e, por isso, não recebe nenhuma especificação especial na base da WN.Br. Porém, devido às reconhecidas diferenças entre as línguas, oriundas de diferentes padrões de lexicalização e da existência de lacunas lexicais (CRUSE, 1986; VOSSSEN, 1998), relações complexas de equivalência podem ocorrer.

Tais relações complexas são comparáveis aos diferentes tipos de correspondência entre os significados das palavras. Elas são especificadas e sinalizadas no desenvolvimento da base da WN.Br quando há lacunas lexicais no português ou no inglês, ou quando a lexicalização dos conceitos nas duas línguas não se correspondem. As relações complexas discriminadas por Peters et al (1998) são: a de semelhança semântica (EQ\_NEAR\_SYNONYM) e a de inclusão semântica (EQ\_HAS\_HYPERONYM e EQ\_HAS\_HYPONYM).

Verifica-se a relação EQ\_NEAR\_SYNONYM quando um synset da WN.Br corresponde a diferentes ILIs ou quando mais de um synset da WN.Br correspondem a um único ILI como, por exemplo, {arriscar, aventurar}, glosado por “correr risco diante de resultado favorável”, relaciona-se por EQ\_NEAR\_SYNONYM aos synsets do inglês {risk, put on the line, lay on the line} e {gamble, chance, risk1, hazard, take chances, adventure1, run a risk, take a chance}, glosados, respectivamente, por “*expose to a chance of loss or damage*” e “*take a risk in the hope of a favorable outcome*”, através dos ILIs 02470374 e 02469144.

A relação EQ\_HAS\_HYPERONYM é especificada quando um synset da WN.Br lexicaliza um conceito mais específico que qualquer conceito lexicalizado em um ILI. Por esse motivo, o synset da WN.Br é ligado a um ILI hiperônimo. Por exemplo, o synset {dicionarizar, vocabularizar}, “incluir ou registrar em dicionário”, é co-indexado, através do ILI 00930267, ao synset {lexicalize, lexicalise}, glosado por “*make or coin into a word or accept a new word into the lexicon of a language*”, “fazer ou inventar uma palavra, ou aceitar uma nova palavra no léxico de uma língua”.

Este tipo de relação é normalmente utilizado quando há lacunas lexicais (*gaps*) no inglês ou quando há incompletude na WN. Essas lacunas lexicais podem ser “genuínas”, isto é, decorrentes da não existência de um determinado conceito em uma determinada cultura. Por exemplo, o conceito CAIPIRINHA (o conceito de uma bebida típica brasileira que consiste na mistura de limão, açúcar, pinga e gelo) tipicamente brasileiro lexicaliza-se no português na forma **caipirinha**. Ou elas podem ser “não-genuínas”, decorrentes das diferenças de padrões de lexicalização que se verificam entre línguas diferentes. É o que ocorre, por exemplo, com a lexicalização do conceito SAUDADE: dizem-se, no português, “Eu tenho saudades de você” e, no inglês, “I miss you”, o que demonstra que esse conceito, no português, é lexicalizado pelo substantivo **saudade**, ao passo que, no inglês, o mesmo conceito é lexicalizado pelo verbo **miss**.

A relação EQ\_HAS\_HYPONYM é especificada quando o synset da WN.Br lexicaliza um conceito mais geral que qualquer conceito lexicalizado em um ILI. Por esse motivo, o synset da WN.Br é ligado a um ILI hipônimo. Até o momento, não foram encontrados exemplos para ilustrar esta situação.

Conclui-se que a sistematização dos diferentes tipos de co-indexação possibilitará a descrição e o estabelecimento de generalizações dos diferentes modos de expressão a que as duas línguas recorrem para representar os conceitos.

### **Referências Bibliográficas**

- CRUSE, D.A. **Lexical semantics**. Cambridge, Mass: Cambridge University Press, 1986.
- DIAS-DA-SILVA, B.C. O estudo lingüístico-computacional da linguagem. **Letras de hoje**. Curso de pós-graduação em Letras, PUCRS, Porto Alegre, 2006, vol. 41. p. 103-138. ISSN 0101-3335
- \_\_\_\_\_. Montagem da Base da Wordnet para o Português do Brasil. **Relatório Técnico da Chamada CNPq 09/2001 – Conteúdos Digitais/Edital SocInfo/ProTeM 01/2001, 01/03/2002 a 31/05/2004**. Araraquara: CELiC, FCL, UNESP, 2004, p.50.
- DIAS-DA-SILVA, B.C.; OLIVEIRA, M.F., MORAES, H.R. Groundwork for the development of the Brazilian Portuguese Wordnet. In: RANCHHOD, E. M.; MAMEDE, N. J. (eds.) **Advances in natural language processing**. Berlin: Springer-Verlag, 2002, p.189-196. ISBN 3-540-43829-7
- FELLBAUM, C. (ed.) **WordNet**: an electronic lexical database. Cambridge (Mass.)/London: The MIT Press, 1998.
- LENHARO, A.C. Estudo do mapeamento conceitual entre wordnets de línguas distintas: estudo exploratório da co-indexação de synsets de verbos do português e do inglês. In: **Caderno de Resumos do 54º Seminário do Gel**. Araraquara: Ferrari Editora e Artes Gráficas, 2006. p.395 - 396
- LENHARO, A.C.; DIAS-DA-SILVA, B.C. **Relatório anual de pesquisa**: Face Lingüística da Rede Wordnet para o Português do Brasil, 2006.
- MARRAFA, P. **WordNet do Português**: uma base de dados de conhecimento lingüístico. Lisboa: Instituto Camões, 2001.
- MILLER, G.A., FELLBAUM, C. Semantic networks of English. **Cognition**, 1991, vol. 41, p.197-229.
- PETERS, W.; VOSSEN, P.; DÍEZ-ORZAS, P.; ADRIAENS, G. Cross-linguistic Alignment of Wordnets with an Inter-Lingual-Index. In: **Computers and the Humanities**. Dordrecht/Boston/London: 1998, vol. 32, p.221–251.
- VOSSEN, P. Special issue on EuroWordNet. In: **Computers and the Humanities**. Dordrecht/Boston/London: 1998, vol. 32.

**Bolsa:** CNPq/PIBIC.