

AS ESTRUTURAS LÉXICO-CONCEITUAIS E A ELABORAÇÃO DE BASES LEXICAIS PARA TRADUÇÃO AUTOMÁTICA: O CASO DO VERBO APUNHALAR

Bento Carlos Dias-da-Silva¹, Mirna Fernanda de Oliveira²

¹Faculdade de Ciências e Letras – Universidade Estadual Paulista (UNESP)
Rodovia Araraquara-Jaú, km 01 – Araraquara – SP - Brasil

²Programa de Pós-Graduação em Lingüística e Língua Portuguesa – Faculdade de Ciências e Letras – Universidade Estadual Paulista - UNESP
Rodovia Araraquara-Jaú, km 01 – Araraquara – SP - Brasil
bento@fclar.unesp.br, mirna.oliveira@bol.com.br

***Abstract.** This paper sketches a computational linguistic representation for both the syntax and semantics of Portuguese through the analysis of the Portuguese verb “apunhalar” (English “to stab”). Lexical-Conceptual Structure Theory is used as the basis for the development of the interlingua that underpins the machine translation of Portuguese-English sentence pairs.*

***Resumo.** Este trabalho ilustra, com verbos "apunhalar" e "to stab", uma proposta de representação lingüístico-computacional da sintaxe e semântica de verbos do português, fundamentada na teoria das Estruturas Léxico-Conceituais, base de representação da interlíngua empregada no processo de tradução automática de pares de frases português-inglês.*

1. Introdução

O objetivo deste trabalho é compreender como se estabelece a relação entre a Estrutura de Argumentos e a Estrutura Léxico-Conceitual (ELC), proposta por Jackendoff (1990), com vistas ao planejamento de uma base lexical de verbos para um sistema de tradução automática. Neste trabalho, com base em Dorr (1992b), discute-se a ampliação dessa ELC, que passa a ser identificada com uma interlíngua, que exclui propriedades sintáticas específicas de uma língua em particular para incluir propriedades semântico-lexicais universais.

2. A Estrutura Léxico-Conceitual

Para Jackendoff, a ELC constitui um subconjunto da estrutura conceitual, e é concebida como uma linguagem mental. Esse modelo inclui tipos semânticos gerais como EVENT e STATE que são especializados em primitivos semânticos, ou funções semânticas. O tipo semântico EVENT, por exemplo, especializa-se em termos das funções GO, STAY, BE, GO-EXTENT e ORIENT, isto é, funções que representam a estruturação semântica de cada componente da frase. Assim, no exemplo (a), a ELC da frase "A bola rolou em direção a Beth" é dada por:

- (a) A bola rolou em direção a Beth.
 [Event GO ([Thing BALL],
 [Path TOWARD ([Position AT ([Thing BALL], [Thing BETH])]])])]

Nessa representação, a função semântica GO, especialização do tipo EVENT, associa o argumento 1 [Thing BALL] ao argumento complexo 2 [Path TOWARD ([Position AT ([Thing BALL], [Thing BETH])])], com as restrições léxico-semânticas e de ocorrência específicas a cada tipo.

Com o recurso da indexação, essa representação permite estabelecer a correspondência entre a ELC e a estrutura sintática da frase. Em (b) e (c), essa correspondência é ilustrada com os índices i e j, conforme sugere Jackendoff (1990, p.45):

- (b) Estrutura Sintática
 [S [SN João]i [SV entrou [PP em [SN o quarto]]j]].
- (c) Estrutura Conceitual
 [Event GO ([Thing JOHN]i, [Path TO ([Place IN ([Thing ROOM]j)])])]

A frase corresponde ao evento completo na estrutura conceitual. O verbo corresponde à função GO (Thing []i, [Path TO ([Place IN ([Thing]j)])]). O sujeito da frase corresponde ao primeiro argumento de GO e o locativo, ao seu segundo argumento, um argumento complexo. Esse argumento é complexo porque a função monoargumental TO possui um argumento que também é uma função monoargumental. Trata-se, portanto, de uma função de função. Tem-se: uma entidade do tipo Thing é argumento da função IN, que por sua vez é argumento da função TO.

3. Estrutura de argumentos, ambiente sintático e estrutura conceitual de *apunhalar*

Dada a relevância da explicitação dos argumentos do verbo para a ELC, recorre-se à Teoria do Léxico Gerativo (TLG) de Pustejovsky (1995) para nortear a especificação da **estrutura de argumentos** do verbo *apunhalar*, sem perder de vista a ELC:

Evento 1 = "processo de apunhalar"

Evento 2 = "o estado resultante de estar ferido"

Com a restrição: Evento1 deve preceder Evento 2.

Assim, a estrutura de argumentos de *apunhalar* é dada por:

apunhalar (Arg.1 – humano, Arg.2 – humano, Arg.S – objeto físico).

Quanto aos **ambientes sintáticos**, o levantamento da estrutura de argumentos do verbo *apunhalar* (cf. Oliveira, 2002) permite observar que tanto para o português como para o inglês, o tipo semântico dos argumentos é o mesmo para o sentido denotativo. A Tabela 1 mostra o resultado da análise dos tipos de ocorrências encontrados no corpus utilizado¹ e o padrão sintático do verbo.

¹ Ocorrências retiradas do *corpus* do NILC – Núcleo Interinstitucional de Linguística Computacional (<http://acdc.linguatca.pt/acesso/>).

Tabela 1: Tipos de ocorrências e o padrão sintático do verbo *apunhalar*.

Padrão Sintático	Ocorrências
V Apunhal- _____ <SNj>	“O escritor egípcio Naguib Mahfouz, 82, foi <u>apunhalado</u> em frente a sua casa no Cairo”
	“A polícia israelense disse ter prendido três palestinos que queriam <u>apunhalar um soldado</u> .”
	“Mahfuz, que está com 83 anos de idade, estava perto de sua casa, na capital, Cairo, quando <u>foi apunhalado</u> supostamente por um integralista islâmico.”

Em todos os casos, o evento é o mesmo: “alguém sofreu um ferimento causado por um instrumento pontiagudo à semelhança de um punhal”. O que muda é a entidade que é vista como agente/paciente e os locativos.

Quanto à **estrutura conceitual**, a ELC do verbo *to stab*, “apunhalar” em inglês, conforme propõe Dorr (1992b, p.132), é dada por:

to stab

[Event CAUSE (Thing * W), [Event GO Poss] ([Thing KNIFE_WOUND], [Path TOWARDS A Poss ([Position IN Poss ([Thing KNIFE_WOUND], [Thing * Z])])]), [WITH Instr * (Event *HEAD*), [Thing SHARP_OBJECT U]])]

O primeiro nível de descrição especifica o significado independentemente da expressão lingüística. Em termos nocionais, essa representação pode ser assim descrita: “trata-se de um evento em que uma entidade W (agente) faz com que Z (paciente) sofra uma punhalada Y por meio de um instrumento pontiagudo U”. O segundo nível de descrição na entrada da ELC é dado pela notação (*), leia-se "notação estrela", que é usada para especificar a realização sintática dos argumentos conceituais.

4. Conclusão

O levantamento dos argumentos do verbo *apunhalar* apresentado em 2, o levantamento dos padrões sintáticos e do esboço da estrutura conceitual da entrada do verbo *apunhalar*, permitem a representação individual das entradas em uma única representação que dê conta de frases do tipo abaixo:

- ◆ João apunhalou Maria. (tipo “W apunhal- Z”)

Desta forma, cada entrada do léxico teria a seguinte caracterização:

João = [Thing JOHN]
N

Maria = [Thing MARY]
N

apunhal-

V

 <SNj>
[Event CAUSE (Thing * W), [Event GO Poss] ([Thing KNIFE_WOUND], [Path TOWARDS A Poss ([Position IN Poss ([Thing KNIFE_WOUND], [Thing * Z]j))])], [WITH Instr * (Event *HEAD*), [Thing SHARP_OBJECT U]j)]]

em

P

 <SNj>
[Path TOWARDS A Poss ([Position IN Poss ([Thing Y], [Thing * Z]j))]]

dar

V

 <SNj> <PPk>
[Event CAUSE(Thing * W), [Event GO Poss] ([Thing Y j],[Path TOWARDS A Poss ([Position IN Poss ([Thing Y j], [Thing * Z]k))])]]

punhalada = [Thing KNIFE-WOUND]

N

Estas são amostras de representações estendidas de entradas de um léxico capaz de gerar pares de sentenças inglês/português do tipo “W apunhal- Z”; observe-se que as variáveis são atualizadas pelos valores dos argumentos.

Observa-se que a realização sintática da estrutura conceitual é a mesma para o português e para o inglês. Por meio dessa representação, é possível gerar tanto “*Paulo apunhalou a mulher*” como “*Paulo apunhalou a mulher com uma faca*”. (Cf. DORR, 1992a, p.19), o que atesta a operacionalidade da teoria das estruturas léxico-conceituais na construção de bases lexicais voltadas para tradução automática via interlíngua.

REFERÊNCIAS BIBLIOGRÁFICAS

DORR, B. J. The use of lexical semantics in interlingual machine translation. *Journal of Machine Translation*, 7:3, pp. 135-192, 1992a.

_____. *Machine Translation: a view from the lexicon*. Cambridge /Mass.: The MIT Press, 1992b.

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

OLIVEIRA, M. F. *A teoria do léxico gerativo e evidência de corpus: explorações*. Florianópolis: UFSC, 2002. Mimeo. (Monografia).

PUSTEJOVSKY, J. *The Generative Lexicon*. Cambridge (Mass.)/London: The MIT Press, 1995.