

LINGÜÍSTICA COMPUTACIONAL: TEORIA E PRÁTICA

Marcelo Módolo*

OTHERO, Gabriel de Ávila & MENUZZI, Sérgio de Moura. *Linguística computacional: teoria e prática*. São Paulo, Parábola Editorial, 2005, 128 p. ISBN: 85-88456-39-X

A obra configura-se como uma introdução didática à linguística computacional, discutindo-se basicamente dois temas: i) “o que é linguística computacional e que tipos de trabalho ela tem desenvolvido” p. 12, nos capítulos 1 e 2, respectivamente *Tecnologia e linguagem; Trabalhos em linguística computacional* e ii) “como os conhecimentos da linguística podem colaborar no desenvolvimento de recursos computacionais” p. 12, nos capítulos 3, 4 e 5; por sua vez *Desenvolvendo um parser sintático I, Desenvolvendo um parser sintático II e Nossa gramática*. Há ainda três seções que finalizam esse livro: *Palavras finais, Referências bibliográficas* e um elucidativo *Glossário*.

No primeiro capítulo, *Tecnologia e linguagem*, relatam-se os avanços da informática, desde o início da década de oitenta, com a popularização dos primeiros computadores pessoais, até a recente criação de uma nova área de pesquisa, a *HCI* (*Human Computer Interaction* ou Interação Humano-Computador). O desenvolvimento de pesquisas em *HCI* é apontado como solução para tentar fazer com que máquinas sejam capazes de interagir com seres humanos e compreendê-los. Entretanto, para que esses estudos tenham êxito, a fim de criar máquinas “inteligentes”, ainda é necessário que se avance na análise das línguas naturais e na descrição formal dos sistemas linguísticos, bases indispensáveis para que se comece a falar de linguística computacional.

Tais pesquisas seriam, portanto, as molas propulsoras para o desenvolvimento da linguística computacional, que se divide em: a) linguística de *corpus* – trabalhos feitos a partir de *corpora* eletrônicos, isto é, grandes bancos de dados que contenham amostras de linguagem natural e b) processamento de linguagem natural (PLN) – estudo da linguagem voltado para a construção de *softwares* e sistemas computacionais específicos, como tradutores automáticos, *chatterbots*, *parsers*, reconhecedores automáticos de voz, geradores automáticos de resumos, etc.

No segundo capítulo, *Trabalhos em linguística computacional*, dão-se noções de linguística computacional e de programação. Primeiramente, em *Desenvolvimentos em linguística computacional*, historia-se o início da linguística computacional desde 1950 e o impulso que essa área ganhou nas décadas de 50-60 com os programas de tradução automática. Em seguida, destacam-se algumas aplicações que associam linguística à informática: os estudos de fonética e de fonologia empregados em aplicativos para reconhecimento de fala (como em aparelhos celulares, que fazem a discagem a partir do reconhecimento de voz do proprietário do aparelho), em programas na área da síntese de fala (como aqueles capazes de gerar fala, a partir de dados, caso do *Virtual Vision*) e em

* Pós-doutorando no Departamento de Linguística, Instituto de Estudos da Linguagem, Unicamp. Caixa Postal 6045, CEP 13084-971, Campinas, SP. E-mail do autor <modolo@usp.br>

outros muito mais sofisticados, que envolvem a interação entre ser humano e máquina por intermédio de diálogos orais em linguagem natural.

Além dos estudos de fonética e de fonologia, os estudos de sintaxe e de semântica são tidos como fundamentais para sistemas que envolvem a compreensão ou a geração automática de frases de uma língua, caso dos *chatterbots*, robôs (programas) que simulam a conversação humana. Como exemplo, detalha-se o funcionamento do primeiro *chatterbot* desenvolvido no mundo, o ELIZA, criado pelo pesquisador Joseph Weizenbaum, no MIT ([Massachusetts Institute of Technology](http://www.mit.edu)), em 1966. Além dos *chatterbots*, os tradutores eletrônicos são aplicativos que igualmente precisam de conhecimentos sintáticos e semânticos.

No subcapítulo *Processamento sintático computacional* esclarece-se o que é um *parser* automático, ou seja, um programa capaz de classificar morfossintaticamente as palavras e expressões de sentenças em uma dada língua e, principalmente, de atribuir às sentenças a sua estrutura de constituintes, baseando-se em um modelo formal de gramática. Tal ferramenta pode ser desenvolvida utilizando a linguagem *Prolog* (abreviação de *Programming in Logic*), uma linguagem de programação de natureza declarativa, baseada na lógica, tema do terceiro subcapítulo. Assim, no terceiro subcapítulo, *O Prolog*, esclarece-se que “para o *Prolog*, é importante que o programador dê ao programa dados para combinações — representados por proposições em um formato semelhante ao da lógica de primeira ordem —, e o *Prolog* resolverá, então, problemas pertinentes a esses dados, tentando demonstrar logicamente proposições derivadas a partir deles” p. 42.

No terceiro capítulo, *Desenvolvendo um parser sintático I: noções teóricas importantes*, examinam-se noções de teoria lingüística e teoria computacional, os primeiros passos para se desenvolver um *parser* sintático. Em *Teoria lingüística*, há noções de gramática formal baseadas em CHOMSKY (1956 e 1965). De posse dessas noções, é possível elaborar uma gramática robusta ou um conjunto de regras que — juntamente com o léxico — permitirá formar todas as sentenças gramaticais de uma dada língua. São examinadas, ainda nesse subcapítulo, noções de uma gramática sintagmática e teoria dos constituintes.

Em seguida, em *Teoria computacional*, elabora-se uma primeira descrição de como funciona a programação em *Prolog*, trabalhando com exemplos de frases em português e mostrando sua construção nessa linguagem. Juntam-se a esse subcapítulo esclarecimentos sobre *DCGs* (*Definite Clause Grammars*), presentes em quase todas as versões de *Prolog* distribuídas, “um formalismo para a representação de gramáticas livres de contexto” p. 76, isto é, a “*grosso modo*, *DCG* pode ser entendida como um recurso que torna mais fácil ao lingüista a implementação de regras sintagmáticas em *Prolog*” p. 77.

No quarto capítulo, *Desenvolvendo um parser sintático II: “colocando a mão na massa”*, aprofundam-se as considerações feitas no capítulo anterior por meio de dois subitens: *Sobre as regras gramaticais* e *Sobre o léxico*. Em *Sobre as regras gramaticais*, começa-se o desenvolvimento da gramática em *Prolog*, implementando regras sintagmáticas relativas à descrição de sentenças simples do português. Por sua vez, em *Sobre o léxico*, reforça-se a idéia de que o componente lexical é o fator primeiro e básico para que o *parser* possa fazer o reconhecimento das palavras que ele tem que etiquetar, ou seja, primeiramente é preciso “ensinar” a máquina, apresentar-lhe o léxico para que ela possa realizar as operações de etiquetagem.

No quinto e último capítulo, *Nossa gramática*, mostra-se de maneira prática como desenvolver um aplicativo computacional voltado para o tratamento da linguagem

natural. Com esse objetivo, são aprofundados e aplicados, com maiores detalhes, conhecimentos sobre *Prolog*.

O livro é recheado com sugestões de leitura e de visitas a *homepages*, as quais complementam ou exemplificam as discussões propostas em cada capítulo. O trabalho é uma boa oportunidade para professores conhecerem esse novo e promissor ramo da lingüística, que, respeitando-se as proporções, pode ser aplicado ao ensino fundamental e médio; haja vista empreendimentos já bem-sucedidos nesse setor como os de Eckhard Bick¹ na Dinamarca. Além da leitura da obra, também é recomendável conferir as orientações para montar seu próprio *parser*, exercício válido e igualmente sugerido pelos autores para bem compreender a teoria exposta.

¹ VISL (*Visual Interactive Syntax Learning*) <http://visl.sdu.dk/>