

ABORDAGENS
COMPUTACIONAIS
da teoria da
gramática

Leonel Figueiredo de Alencar
Gabriel de Ávila Othero
(organizadores)

ABORDAGENS
COMPUTACIONAIS
da teoria da
gramática

MERCADO[®]
LETRAS

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Abordagens computacionais da teoria da gramática / Leonel Figueiredo de Alencar, Gabriel de Ávila Othero, (organizadores). – Campinas, SP : Mercado de Letras, 2011.

Vários autores

ISBN 978-85-7591-204-1

1. Português – Gramática – Abordagens computacionais 2. Português – Gramática – Teoria I. Alencar, Leonel Figueiredo de. II. Othero, Gabriel de Ávila.

11-11254

CDD-469.5

Índices para catálogo sistemático:

1. Gramática : Teoria : Abordagens computacionais : Linguística 469.5

capa e gerência editorial: Vande Rotta Gomide

revisão: Editora Mercado de Letras

DIREITOS RESERVADOS PARA A LÍNGUA PORTUGUESA:

© MERCADO DE LETRAS® EDIÇÕES E LIVRARIA LTDA.

Rua Dr. João da Cruz e Souza, 53

Telefax: (19) 3241-7514

13070-116 - Campinas SP Brasil

www.mercado-de-letras.com.br

livros@mercado-de-letras.com.br

1ª EDIÇÃO

*texto adaptado para as novas
normas de ortografia da
língua portuguesa*

2 0 1 2

IMPRESSÃO DIGITAL

Esta obra está protegida pela Lei 9610/98.

É proibida sua reprodução parcial ou total
sem a autorização prévia do Editor. O infrator
estará sujeito às penalidades previstas na Lei.

Agradecimentos

Este livro só pôde ser publicado graças a esforços conjuntos de diversos colegas. Devemos agradecer primeiramente aos autores dos capítulos aqui publicados, por terem compreendido a proposta do livro e nos terem enviado seus textos dentro dos prazos programados. Além deles, destacamos o papel de Carmen Kelling, que integrou a equipe organizadora no início do projeto e contribuiu com valiosas sugestões para o resultado final do livro.

Por fim, gostaríamos de agradecer a todos os revisores dos capítulos, que leram cuidadosamente os textos aqui publicados e enviaram suas observações para os autores. Entre esses, destacamos Luiz Arthur Pagani, Emanuel de Souza Quadros e Adriana R. Quinelo.

SUMÁRIO

INTRODUÇÃO	9
<i>Leonel Figueiredo de Alencar e Gabriel de Ávila Othero</i>	
1. LÍNGUAS FORMAIS, GRAMÁTICAS E AUTÔMATOS NO PROCESSAMENTO AUTOMÁTICO DAS PALAVRAS	13
<i>Leonel Figueiredo de Alencar</i>	
2. AGRUPAMENTOS NOMINAIS NO PORTUGUÊS BRASILEIRO: UMA ANÁLISE FORMAL E SUA APLICAÇÃO COMPUTACIONAL	77
<i>Karine Alves David e Gabriel de Ávila Othero</i>	
3. PARSING DO PORTUGUÊS	99
<i>Ronaldo Teixeira Martins e Gabriel de Ávila Othero</i>	
4. O PROGRAMA MINIMALISTA EM SUA PRIMEIRA VERSÃO	127
<i>Marcello Modesto</i>	
5. QUESTÕES RECENTES DA TEORIA MINIMALISTA	153
<i>Christoph Gabriel</i>	

6. HPSG: ARQUITECTURA	219
<i>Antônio Branco e Francisco Costa</i>	
7. REPRESENTAÇÃO SINTÁCTICA EM HPSG.	237
<i>Antônio Branco e Francisco Costa</i>	
8. REPRESENTAÇÃO DO SIGNIFICADO EM HPSG.	269
<i>Antônio Branco e Francisco Costa</i>	
SOBRE OS AUTORES	303

INTRODUÇÃO

O objetivo central deste livro é preencher a lacuna existente de obras que explorem, de forma mais aprofundada, a aplicação computacional de pesquisas em sintaxe formal. Esta publicação busca oferecer aos leitores de língua portuguesa um material teoricamente bem fundamentado e, ao mesmo tempo, de leitura acessível na área de interface entre gramática e computação, sob a perspectiva de alguns dos principais modelos linguísticos dos últimos 25 anos caracterizadas por um *design* computacional. Diferentemente do que ocorre para o inglês, para o francês e para o alemão, inexistia em português um livro-texto sobre abordagens formais da gramática e sua interface com a Linguística Computacional que pudesse ser utilizado em cursos de Letras, Linguística, Ciências da Computação e Informática, em disciplinas como, por exemplo, Linguística, Sintaxe, Morfologia, Semântica, Linguística Computacional, Processamento da Linguagem Natural e Linguística Formal. Uma vez que alguns dos manuais introdutórios disponíveis, como Othero e Menuzzi (2005) e Othero (2006, 2009), se restringem à análise automática no nível sintático, lacunas análogas se constatavam na área tanto da morfologia formal e sua aplicação no processamento automático das palavras quanto da interface entre gramática e semântica.

Na Europa e nos Estados Unidos, a área da Linguística Computacional está em extrema expansão e goza de muita popularidade, tanto nos cursos

de Ciências da Computação quanto nos de Linguística. De fato, o estudo do processamento automático da linguagem natural é considerado nesses países como relevante não só para a indústria de *software*, mas também para a elaboração de teorias gramaticais formalmente mais consistentes e psicolinguisticamente mais realistas. Como em várias outras ciências, também na linguística teórica tem-se usado, cada vez mais, o computador para simular vários aspectos do objeto de estudo (no caso, a linguagem natural) e, assim, testar, com um grau de sofisticação que dificilmente poderia ser atingido por seres humanos, a adequação dos modelos postulados.

No Brasil, a área é relativamente nova, mas vem se desenvolvendo em diversas universidades do país. Mesmo assim, como se ressaltou acima, ainda há uma enorme carência de materiais bibliográficos facilmente acessíveis em língua portuguesa que possam servir de embasamento a esses estudos. Este livro tem a pretensão de preencher essa lacuna, contribuindo, desse modo, tanto para uma formação mais sólida em teoria gramatical nos cursos de Letras e Linguística, quanto para a construção de programas de processamento automático da linguagem natural mais eficientes e robustos, o que só é possível, como sugere Sérgio Menuzzi, no prefácio de Othero (2006), se os programadores levarem em conta os resultados das pesquisas desenvolvidas em teoria gramatical sob uma perspectiva computacional.

Este livro consiste de 8 capítulos que abordam aspectos teóricos e práticos da interface entre estudos gramaticais e computacionais, elaborados por especialistas com experiência comprovada em suas áreas e vinculados a conceituadas universidades brasileiras e europeias. A maioria dos autores atua tanto na pesquisa quanto no ensino em nível de graduação e pós-graduação.

O primeiro capítulo, de Leonel Figueiredo de Alencar, focaliza especialmente o processamento computacional no nível da palavra, sob a perspectiva do modelo da Morfologia de Estados Finitos. Essa abordagem constitui a tecnologia mais avançada atualmente para a construção de programas como analisadores morfológicos e corretores ortográficos. Esse capítulo expõe, de forma didática, as noções fundamentais da matemática e da computação necessárias para a implementação de programas de

processamento automático não só no nível da palavra, mas também da sintaxe, servindo, portanto, para embasar a leitura dos capítulos subsequentes.

O segundo capítulo, de Gabriel de Ávila Othero e Karine Alves David, descreve a estrutura dos agrupamentos nominais em português, mostrando uma implementação computacional em Prolog, sob a perspectiva teórica de uma gramática de estrutura sintagmática que segue alguns princípios da Teoria X-barra. Este capítulo traz os resultados combinados dos estudos sobre a estrutura dos agrupamentos nominais em português que serviram de foco para grande parte das investigações de Mestrado (no caso de David) e Doutorado (no caso de Othero) desses autores. Além de analisar formalmente a estrutura interna dos agrupamentos nominais, os autores trabalharam com a implementação computacional de suas regras de estrutura sintagmática.

O terceiro capítulo, escrito por Ronaldo Teixeira Martins e Gabriel de Ávila Othero, focaliza os estudos de análise sintática automática (*parsing*) em língua portuguesa, discute alguns conceitos-chave nessa área e apresenta brevemente dois analisadores (*parsers*) desenvolvidos para processar computacionalmente o português brasileiro, o Curupira e o Grammar Play. O capítulo serve como uma introdução ao processamento sintático do português, levando o leitor da área de Letras a pensar numa interface possível (entre muitas) entre investigação sintática e processamento computacional.

Os capítulos 4 e 5, respectivamente de Marcello Modesto e Christoph Gabriel, expõem diferentes aspectos da sintaxe no quadro do Programa Minimalista, a mais difundida concepção da linguagem natural, no âmbito da linguística teórica, como um sistema computacional autônomo (e geneticamente determinado) da mente dos falantes. O Programa Minimalista é a versão atual da gramática gerativa transformacional de Noam Chomsky, lançada nos anos 50 do século passado, cuja preocupação central tem sido descrever, numa linguagem próxima da matemática, as representações mentais e as operações cognitivas que as manipulam para produzir as sentenças das línguas naturais. Embora mais estudada por linguistas, essa abordagem também tem sido utilizada para construir aplicativos de processamento automático da linguagem natural, ao mesmo tempo em que tem servido como ponto de partida para a construção de modelos alternativos

que, dispensando as transformações sintáticas, alcançaram ampla difusão na Linguística Computacional, como a LFG e a HPSCG.

Finalmente, os capítulos 6, 7 e 8, de Antônio Branco e Francisco Costa, delineiam o modelo da HPSCG, o qual, ao lado da LFG, constitui a teoria gramatical computacionalmente implementada mais difundida atualmente no campo da engenharia da linguagem natural, tendo constituído a base, por exemplo, de programas de extração automática de informações a partir da Internet e do projeto Verbmobil, um programa de computador capaz de traduzir textos tanto escritos quanto falados em japonês, inglês e alemão. Enquanto o primeiro capítulo expõe os fundamentos teórico-epistemológicos do modelo, os dois restantes debruçam-se sobre o tratamento computacional, respectivamente, da sintaxe e da semântica.

Este livro destina-se, sobretudo, aos estudantes avançados de graduação em Letras, Linguística, Computação, Sistemas de Informação e Informática; estudantes de cursos de pós-graduação (*lato e stricto sensu*) nessas áreas; pesquisadores, autores de *software* e demais interessados nos fundamentos teóricos linguísticos do processamento automático da gramática das línguas naturais e nos mais recentes formalismos gramaticais e sistemas para o desenvolvimento de gramáticas computacionais. Os leitores poderão encontrar mais informações sobre o livro no blog *Abordagens computacionais da teoria da gramática*, em <http://teoria-da-gramatica.blogspot.com/>

Os organizadores

OTHERO, G. A. e MENUZZI, S. M. (2005). *Linguística Computacional: teoria & prática*. São Paulo: Parábola.

OTHERO, G. A. (2006). *Teoria X-barra: descrição do português e aplicação computacional*. São Paulo: Contexto.

OTHERO, G. A. (2009). *A gramática da frase em português: algumas reflexões para a formalização da estrutura frasal em português*. Porto Alegre: EDIPUCRS.