当代语言学理论丛书 Contemporary Linguistic Theory Series

计算语言学导论

Introduction to Computational Linguistics

翁富良 王野翊 著

Weng Fu-Liang Wang Ye-Yi

中国社会科学出版社

主编 黄正德 许国璋

Chief Editors James Huang De Bao Xu

当代语言学理论丛书
主编　黄正德　许德宝

计算语言学导论

Introduction to Computational Linguistics

翁富良　王野翊　著
Weng Fu-Liang Wang Ye-Yi

中国社会科学出版社

(京)新登字030号

图书在版编目(CIP)数据

计算语言学导论/翁富良,王野翊著. —北京:中国社会科学出版社,1998.9
(当代语言学理论丛书)
ISBN 7-5004-2080-3

Ⅰ.计… Ⅱ.①翁…②王… Ⅲ.统计语言学 Ⅳ.H087

中国版本图书馆CIP数据核字(98)第01118号

中国社会科学出版社出版发行
(北京鼓楼西大街甲158号)
北京奥隆印刷厂印刷 新华书店经销
1998年9月第1版 1998年9月第1次印刷
开本:850×1168毫米 1/32 印张:7 插页:2
字数:178千字 印数:1—3000册
定价:16.00元

序　言

语言学自乔姆斯基以来，对认知科学、心理学、医学、电子计算机以及人工智能等学科都产生了巨大的影响，成为人文科学的带头学科。只要在国外走一走，就会发现几乎所有的大学都设有语言学系或语言学专业。语言学理论不但对语言学系的学生至关重要，而且也是心理系、教育系、社会学系、认知学理论乃至计算机系的学生必修的基础理论课。乔姆斯基的语言学理论为什么对人文科学和社会科学的影响如此之大？他的什么变革使本来默默无闻的语言学（理论）一跃而成为认知科学、心理学、电子计算机以及人工智能等学科的奠基理论？这不是一句话能说清楚的。要回答这个问题，得从现代语言学的立足点说起，系统介绍现代语言学的基本理论和研究方法、研究对象、研究范围以及研究结果等。不说清楚这些问题，现代语言学在人文科学中的带头作用和对社会科学的巨大影响也就无法说清楚。有系统有深度地介绍现代语言学理论，这就是我们这套丛书的编写目的。

要系统介绍现代语言学，各种理论的来龙去脉都得交代清楚，某种理论的发生、发展、不同阶段以及各个流派之间的关系都要说清楚。不能只把一种理论搬来，不管它的过去和与其他理论的联系，那样会让人不知所云。在系统介绍的同时，也要把各种理论的最新研究成果写进去，并评价其优劣不同以及对现代语言学研究的贡献等，做到有深度、有系统，这是我们介绍的第一个原则。介绍的起点一般是以乔姆斯基与哈利的《英语语音系统》

(1968) 为始，介绍的终点就是今天，介绍时以八九十年代发展起来的语言学理论为主，所以这套书叫作《当代语言学理论丛书》。

要介绍现代语言学并不容易。台湾、新加坡、香港等地的学者有很好的经验。他们介绍的特点就是把现代语言学理论与汉语的研究结合起来。这样理解起来方便得多，效果也就比较好。单纯介绍，不谈在汉语中的应用，结果理论还是死的东西。我们这套丛书也本着这一点，在选材和编写上都强调在汉语中的应用，尽量用汉语说明。汉语与某种理论不相关的时候，才用其他语言中的例子。这是我们介绍的第二个原则。

我们的第三个原则是以介绍美国语言学理论为主。美国是现代语言学研究的中心，也是生成语言学的发源地。要介绍现代语言学就离不开这个发源地。所以从选材上来讲，我们以美国语言学系研究生（博士和硕士）的必修课为标准，包括语言学史、句法学、音系学、语义学、心理语言学、社会语言学、历史语言学、语言获得理论、计算机语言学与人工智能等。有些新兴学科和边缘学科就放在主要学科中介绍。比如神经语言学归入了心理语言学，音系与句法的交叉研究归入了音系学，语义和句法的交叉研究归入了语义学等。

应该指出，有些学者一直在致力于现代语言学的介绍工作，比如黑龙江大学、上海复旦大学、天津师范大学的学者等。我们希望这套丛书能与他们的研究结合起来，起到使国内外语言学研究接轨的作用。

《当代语言学理论丛书》的编写开始于1993 年，由著名句法学家黄正德教授全面负责，许德宝协助作主编工作。编委大都是在美国读语言学博士而且有教授语言学经验的学者，一般是在讲义的基础上增删整理成书。但即使是如此，也都得付出很多的劳动。我们也请了在美国教授多年的语言学家、汉学家和有在国内外介

绍现代语言学经验的学者作为顾问，帮助我们把这一套丛书出好。在此向他们谨致谢意。我们还得感谢中国社会科学出版社对这套丛书的大力支持，特别是责任编辑及其他有关同志的辛苦工作，不然这套丛书也不能和读者见面，在此也一并致以谢意。

《当代语言学理论丛书》编委会

1996年7月于纽约

当代语言学理论丛书
Contemporary Linguistic Theory Series

主　编
Chief Editors

编辑委员会
Editorial Board

吴建慧(伊利诺大学)
Mary Wu(University of Illinois,Champaign—Urbana)
谢天蔚(戴维斯加州大学)
Tianwei Xie(University of California,Davis)
徐大明(新加坡南洋理工大学)
Daming Xu(Nanyang Technological University,Singapore)
许德宝(纽约汉弥尔顿文理学院)
De Bao Xu(Hamilton College,New York)
张敏(新加坡国立大学)
Min Zhang(National University of Singapore)

顾问编辑委员会
Advisory Editorial Board

作者介绍

翁富良 斯坦福国际研究所语音技术和研究实验室研究工程师。1984年毕业于复旦大学计算机科学系。在1984—1989年间，师从吴立德教授进行模式识别和自然语言理解方面的研究。1989年，赴卡内基—梅隆大学机器翻译中心继续自然语言理解的研究。1993年，在新墨西哥州立大学获硕士学位。自1994年起，在斯坦福国际研究所的语音技术和研究实验室从事语言、语音模型研究。先后单独或与同事合作在一些专业杂志和会议上发表论文20余篇，曾获1986年国家教委科技进步一等奖，第三届中国国家自然科学四等奖。

王野翊 1985年于上海交通大学计算机科学与工程系获学士学位。后师从上海交通大学孙永强教授和中科院数学研究所陆汝钤研究员进行自然语言处理的研究，并于1988年获上海交通大学计算机科学与工程系硕士学位。1992年于美国卡内基—梅隆大学(Carnegie Mellon University)获计算语言学硕士学位，现为卡内基—梅隆大学计算机科学学院语言技术研究所(Language Technologies Institute, School of Computer Science)博士候选人。主要研究课题包括统计学机器翻译、语言模型、语言学习、神经网络。

前　言

语言是反映人的思维的最重要的一面镜子，又是人与人之间交流的最重要的媒介。对语言的研究，是一个经久不衰的古老课题。几千年来，中外学者从语言与思维、语言与现实的关系等不同角度，在词源、注释、分类、语法等各个方面开展了广泛的研究。近一两百年来，西方学者在逻辑学、数学和分析哲学方面的成果，大大推动了语言形式化的研究。而随着计算机科学的发展，建立语言的形式化计算模型成为语言学的重要课题。计算语言学作为以形式化的计算模型来分析、理解和处理语言的科学也就应运而生。而信息革命的展开，更使计算语言学的研究达到了一个空前的程度。

如果说纸的发明对人类文明的继承光大有着巨大作用的话，那么语音和语言技术的发展，将对人类各语种之间的交流，各文化体系间的促进与提高至关重要。语言和语音技术为语言文本和会话的检查、理解、合成、翻译、重组，提供了有效的自动化工具，使得靠人工进行的信息交流和信息处理能够逐步地为具有智能的语言技术所取代。在信息革命使世界日新月异的今天，计算语言学的发展成为我们是否能够跟上世界潮流的一个重要因素。这一点应该激起广大研究人员的高度紧迫感，同时也应该获得科研基金组织和工商界有识之士的重视，对计算语言学的研究给予长期的大力支持。

本书的作者希望此书的出版能够对国内计算语言学的发展起到一定的促进作用。由于篇幅和作者水平的限制，我们不可能面

面俱到地覆盖整个领域。我们在选材时一方面注重本领域的基础性的经典工作，希望读者阅读理解后能够举一反三，用于解决实际问题；另一方面我们侧重于介绍一些当前国际计算语言学界的研究重心，希望有关的研究人员能够站在该领域的前沿。

本书的对象是大学计算机专业、数理统计专业及语言学专业的高年级学生或研究生，与计算语言学有关的科研人员，以及其他有兴趣的读者。由于计算语言学的综合性特点，如有条件，作者建议组织多学科的兴趣小组，相互交流，共同提高。

本书第2章第1节、第3章、第4章、第5章、第6章、第7章由翁富良撰写，第1章、第2章第2、3节、第8章、第9章、第10章由王野翊撰写。

目　录