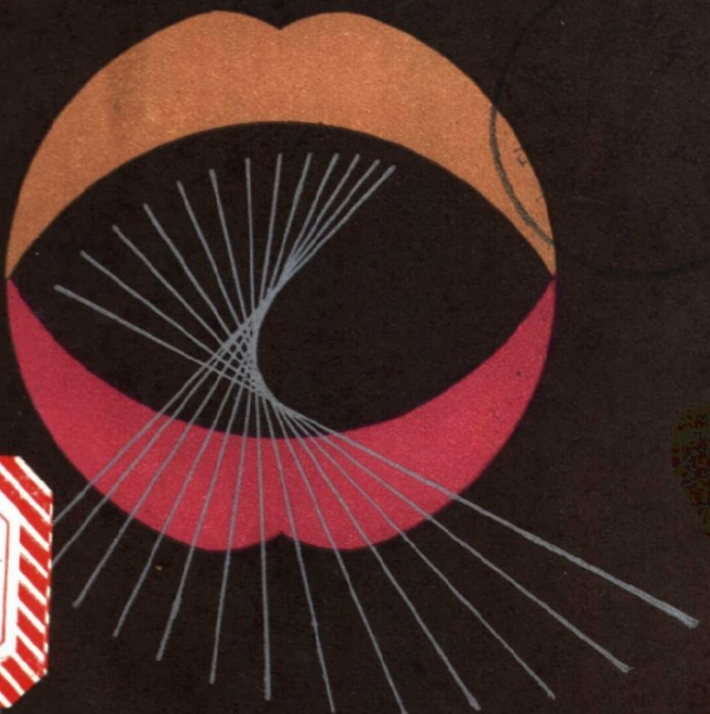
钱锋著

计算语言学引论

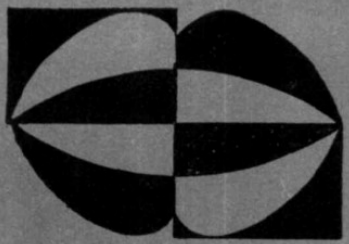
Jisuan Yuyanxue Yinlun



学林出版社

ISBN 7-80510-550-2/T·6

定价：（平）7.85 元



● 钱 锋 著

计算语言学引论

学林出版社

责任编辑：余　石
封面设计：甘晓培

计算语言学引论
——计算机理解自然语言的基本方法　　钱锋

学林出版社出版　　上海文庙路120号
新华书店上海发行所发行　　江苏丹阳新华印刷厂印刷
开本 850×1156　1/32　印张12.25　插页 4　字数 287,000
1990 年 6 月第 1 版　1990 年 6 月第 1 次印刷　印数(精)150册 (平)850册

ISBN 7-80510-550-2/T.6　　定价(精) 9.85 元
(平) 7.85 元

Qian Feng

Introduction to Computational Linguistics—A Multilingual Approach to Some Fundamental Theories and Basic Methods for Computer Understanding of Natural Languages

Xielin Publishing Company, Shanghai, China, 1989

The book is dedicated, in respect and admiration, to the following people:

Prof. Lü Shu-xiang
of the Institute of Linguistics, Chinese Academy of Social Sciences, China who, as a pioneer and a veteran, always encourges me to poke into new things; Prof. Yaohan Chu of Department of Computer Science, University of Maryland College Park

Campus, USA who has introduced me into the world of Computer Science, first through his books, later, his colorful lectures; Prof. I. S. Bátori of Department of Computational Linguistics, EWH Koblenz, Federal Republic of Germany who has provided me with a productive and friendly environment of teaching and research; Prof. U. Müller of Deparment of German Studies, University of Salzburg, Austria who has done all miracle to introduce me to the two German-speaking countries; Prof. P. Zinterhof of Department of Computer Science and System Analysis, University of Salzburg, Austria, without him I could not have had so bright a group of young people sharing and enhancing my knowledge and Dr. D. Walker of Bell Communications Research, NJ. USA who has secured an access for me to the ever challenging and ever growing communities of ACL

序

目前，中国的中文信息处理、机器翻译和自然语言理解等研究，正在八方协力之中迅猛发展。一本试图给这些学科提供基本理论及实用方法的书，得以应时出版，是否可说是躬逢胜伐？

本书受胎于在美、德游学之时。1987 年母校复旦中文系徐志民教授约我去讲“计算语言学”，讲稿始得初具雏型。不久，旋应奥地利萨尔茨堡大学德文系主任 Ulrich Müller 教授，计算机科学及系统分析系主任 Peter Zinterhof 教授以及联邦德国柯布伦茨大学计算语言学系 István Bátori 教授之约，先后担任客座，讲“计算语言学”、“自然语言理解”及“人工智能与专家系统”。仍用此稿，遂边教边改，而得以成书。

计算语言学日新月异，其生命力(Libido)之强劲，文、理、工交叉之特质(Idiosyncrasy)，向其他领域之渗透(Permeability)和其与它们交融的亲和力(Affinity)及其发展的高速度，在当代前沿科学(Science frontiers)之中，恐怕罕有其匹。然而，其有关论文，则常分布于名称、性质、旨趣各不相干的种种出版物，学者搜求，颇为不易。这不像数、理、化，有一大批“经典”教材，用以传授莘莘学子。联邦德国以 Bátori 为首三教授，发

一狠心，于1989年编辑出版了洋洋一巨著，集全世界数十位专家之力，对计算语言学各个方面进行了介绍。但这是一本综览。[1]据我所知，目前"引论"或教科书，只得五本，[2]而其书名、内容、陈述风格则各有千秋。尤令人憾者，是它们全用英文写，也只论英语。其实，汉语，德语或日语等，对计算语言学可能更富"挑战性"。有感于此，本书立意面向汉语(Chinese language oriented)，而兼及其他语言，如英、德、法、日、俄等。从语系看，一为汉藏，一为印欧，一为暂无系属(日语)；从类型看，也有分析，屈折及粘着，皆鼎足而三。这是本书略可一道的特点之一。

第二点是关于算法或处理方案的陈述。计算语言学家在作程序设计时，已广泛采用计算机框图(Flow-chart)，但写作时仍偏爱文字陈述(如[2]所引各书)。本书则在可能之处，都用框图详示之；一便理解，二便学者据此编制程序实现(Implementation)。这一课程，我是与 Lisp 或 Prolog 课程同时教的，学生得以进行程序实习，颇有成效。

书中尽量采用图示，以唤起学者将形象(Picture;Image)与逻辑(Logic)相配合而思维，这是第三点。认知科学(Cognitive Sciences)已经验明，即最长于逻辑思维的科学家，亦常使用恰当的图形以辅助其思考。申言之，善于利用形象也往往是一些自然科学家独出奇思、作出创造的契机(其中最出名的例子，是德国化学家 V. Kekulé 梦一蛇自啮其尾而发现苯环)。特别须指明的是，这一点对文、理、工皆然。在中国或外国，听本课者有来自中文(汉学)、现代语言(英、法、德……)、语言学(Linguistics)、数学、计算机、心理乃至神学(Theology)各系学生，而有的文科学生比理、工科学生学得好，因而获得博士(Ph.D)学位者。所以，本书的"读者群"可以范围很广。形象地说，它可以作为文科(中文，现代语言，文学……)与理、工科(计算机科学，信息科学，数学……)学生"相互访

问"的一座桥梁。

第四点是例句之选用。汉语语法研究向有一优良传统，即例句尽量从作家之语言典范的作品中选取。计算语言学则反之，例句向来都是研究者自己应景所制。本书中凡是较重要的例句，都采自各自语言的名家，如汉语之取钱锺书，德语之取 S.Zweig，法语之取 M.Proust，日语之取川端康成，等等。这样，也容易引起兴味，不知读者以为然否？

第五点，译名之立。计算语言学常变常新，新名词层出不穷，而皆来自英语。其他语言之翻译，常捉襟见肘。本书之译名，多数出自笔者，其中有的未见经传，有的与坊间所译不尽相同，请举数例。ATN 译为"增广迁跃网络"，不译"扩充转移网络"，因 Augment 本意为"增强"，"拓广"，而不仅为单方向的"扩充"；Transition 意为从一状态跃迁至另一状态，而并不是地点之"转移"。再如，Conceptual Dependency 之译为"概念依存"，而不译为"从属"；Grammar 之译为"语法"，而不译"文法"；"Word Processing"之译为"文书处理"，而不译为"字处理"；乃至"Handle"之译为"语柄"，等等。皆出此机杼。再者，本书中凡人名一律不译，已有中文名者（如"司马贺"——Herbert Simon，韩礼德——M.A.K.Halliday），亦不例外。笔者认为，如只记住"尚克"、"盖兹达"或"马尔库斯"等，而欲查原文来读，或在国际场合与人交流，岂非风马牛不相及？且不说还有 Montague，既译"孟德鸠"（与"孟德斯鸠"瓜葛？），复译"蒙太格"（受"蒙太奇"之赐？）等等例子，叫人难识庐山真面目，译尤不如不译。然因学力、时间有限，严复的"一名之立，旬月踌躇"，则非所能及。

本书仅是一次尝试。在编写时，自然参考了多种时贤之作，以取他山之石。目前全世界有一百多所大学皆开有计算语言学或有关课程，其中大部分都无自编讲义，而有相当多大学采用[2]所

引美国斯丹福大学 T.Winograd 的书。本书第一至第四章，许多材料也取自该书。第五章参考了 M.Marcus 的主要论文，写第六章时则阅读了 J.Bresnan 的有关著作。

最后，讲究的教材书，很多是讲授多年，披览数载而成。从这点说，本书是急就之章，错漏之处，恐未必少，望海内外大家、本书读者教正。本书赋形，全赖徐志民、陈光磊兄作曹；出版，全仗雷群明兄鼎力；我的助手王洁硕士在我教这门课时，无论中、外，一直是我的助教，大部分 Lisp 程序出自她手；萨大计算机科学与系统分析系 H.Clausen 教授，德文系 A.Weiss 博士在教学安排、图书供应及计算机设备等方面提供了各种方便，在此一并向他们致谢。

钱 锋

1990.1.

萨尔茨堡-柯布伦茨途中

---

[1] Batori, I. S., W. Lenders, W. Putschke 1989 Computerlinguistik, Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen. Walter de Gruyter

[2] Allen, J. 1987 Natural Language Understanding, Benjamin/Cummings

Grishman, R. 1986 Computational Linguistics, Cambridge Univ. Press

Harris, M.D. 1985 Introduction to Natural Language Processing, Reston Publishing Company

Tennant, H. 1981 Natural Language Processing, Petrocelli Books

Winograd, T. 1983 Language as a Cognitive Process, Vol. 1:Syntax, Addison-Wesley

CONTENTS

目　录