

Ho-11961 ① 钱 锋 著

计算语言学引论

Jisuan
Yuyanxue
Yinlun



学林出版社

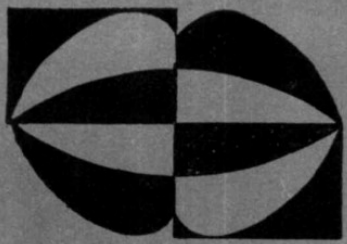
ISBN 7-80510-550-2/T·0

定价：(平) 7.85 元

41

H10-119C1

200741



● 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 Fundamen-
tal 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. Lǚ Shu-xiang
of the Institute of
Linguistics, Chinese Academy of Social Sciences, China who,
as a pioneer and a veteran, always encourages 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,

Germany who has

productive and

teaching and

Müller of

Federal Republic of

provided me with a

friendly environment of

research; Prof. U.

Department 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，是否可说是躬逢胜饯？

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合力(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、旨趣各不相同的种种出版物，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，旬月踟躇”，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

Prolog: Computational Linguistics within the Realm of Contemporary Sciences 1

0.1 A trend in the contemporary sciences and technologies 2

0.2 Natural language as an object of scientific investigation 5

0.3 Linguistics as a cornerstone for Computer Science11

0.4 Linguistics in Artificial Intelligence14

0.5 Linguistics in Communication19

0.6 Linguistics is a must for computing in the Humanities21

0.7 Linguistics plays a role in Office Automation23

0.8 Competence of Computational linguists on the job market24

0.9 Computational Linguistics vs Mathematical Linguistics25

0.10 Placing Computational Linguistics in the realm of contemporary science frontiers28

CHAPTER 1 A Reflexion of the Development of Modern Linguistics29

1.1 A macro outlook of the evolution of science29

每

1.2	Some critic stages in modern linguistics	31
1.2.1	Traditional grammar	31
1.2.2	Comparative Linguistics	32
1.2.3	Structural linguistics.....	32
1.2.4	Transformational generative linguistics	34
1.3	Basic Models for Computational Linguistics	36
1.3.1	A communication model	37
1.3.2	A processing model	39

CHAPTER 2 Linguistic Pattern Matching and Formal Grammar 42

2.1	Linguistic Pattern and their matching	43
2.1.1	Introducing linguistic pattern	43
2.1.2	A formal definition of linguistic pattern.....	45
2.1.3	Variable pattern	49
2.1.4	Application of variable pattern	53
2.2	Categories and dictionary in Computational Linguistics	54
2.2.1	Lexical pattern and lexical categories	54
2.2.2	A discussion of categorization in various languages	55
2.2.3	On dictionary in Computational Linguistics	59
2.3	Formal grammar — a brief introduction	59
2.3.1	Introducing formal grammar	59
2.3.2	Formal grammar, formal languages, programming language and natural language	61
2.4	Transition Network	63
2.4.1	From formal grammar to transition network	63
2.4.2	Generating sentences using transition network.....	68
2.4.3	Recognizing sentences using transition network	70
2.4.4	Parsing English sentence using transition network	72

2.4.5	Transition network with parallel processing	74
2.4.6	Transition network with backtracking mechanism	81
CHAPTER 3	Context-Free Grammar(CFG) and	
	Parsing	89
3.1	Two approaches to structural description	89
3.1.1	Tree structure.....	90
3.1.2	Functional description	92
3.1.3	Tesnière's dependency analysis	96
3.2	Rules and Derivations	98
3.2.1	A formal description of CFG	98
3.2.2	The derivation of sentence	99
3.2.3	Sentence derivation without tree structure	106
3.2.4	Parsing sentence with CFG.....	108
3.3	Sentence recognition with CFG	115
3.3.1	Top-down recognition.....	115
3.3.2	Bottom-up recognition	122
3.4	Serial processing by backtracking	128
3.4.1	A description of the algorithms	129
3.4.2	Case studies languages as in Chinese and German	130
3.5	Parsing with CFG-Chinese, German, French,	
	Russian and Japanese.....	141
3.6	A general parsing scheme-active charts	145
3.6.1	A concept of active charts.....	146
3.6.2	A active chart parser	148
3.6.3	Parsing sentence with active charts.....	152
CHAPTER 4	Augmented Transition Network	
	(ATN)	164
4.1	Recursive transition network(RTN).....	165
4.1.1	RTN and CFG.....	165

4.1.2	Formal description and matching of RTN.....	169
4.1.3	Parsing with RTN	174
4.2	Augmenting RTN	182
4.2.1	The deficiencies of RTN.....	182
4.2.2	Augmenting RTN—considerations from Chinese and German	186
4.3	The structure of ATN and its applications.....	194
4.3.1	The structure of ATN.....	194
4.3.2	Feature register—considerations from Chinese and German.....	196
4.3.3	Features and semantic analysis.....	204
4.3.4	ATN dictionary	207
4.3.5	Conditions and actions	210
4.3.6	Role register—considerations from Chinese and German ..	215
4.3.7	Rules and semantic analysis	222
4.4	Formal definition for ATN structures and processes	227
4.4.1	ATN and its arcs	228
4.4.2	Parsing with an ATN.....	234
4.5	Parsing English with ATN.....	238
CHAPTER 5 Deterministic Parsing and Wait-		
 And-See Parsing (WASP)		
 247		
5.1	WASP parsing fundamentals	252
5.1.1	The philosophy of deterministic parsing.....	252
5.1.2	The structure of the grammar interpreter.....	254
5.1.3	The structure and the process of the rules.....	264
5.2	A case study—parsing a declarative English sentence	273
5.2.1	Rule INITIAL in the packet NOWHERE	274
• 4 •		

5.2.2	Packet SS-START	276
5.2.3	Packet PARSE-SUBJ.....	278
5.2.4	Packet PARSE-AUX.....	279
5.2.5	Packet BUILD-AUX.....	281
5.2.6	Packet PARSE-AUX again	283
5.2.7	Packet PARSE-VP	285
5.2.8	Packet SUBJ-VERB	288
5.2.9	Packets INF-COMP and SS-VP	291
5.2.10	Packet SS-FINAL	292

CHAPTER 6 Functional Grammar and Lexical-

	Functional Grammar.....	294
6.1	Functional Grammar.....	298
6.1.1	Introducing functional grammar	298
6.1.2	Parsing with functional grammar.....	304
6.1.3	The dictionary in functional grammar.....	308
6.2	Lexical-Functional Grammar (LFG).....	309
6.2.1	The principles of LFG.....	309
6.2.2	The process of parsing with LFG	319
6.2.3	Long-distance dependency in LFG	324
6.3	Parsing Chinese language with LFG.....	335
	EPILOG: Computational Linguistics in the 80's	341
7.1	Improvements of the available disciplines	342
7.1.1	Refinements to WASP parsing.....	342
7.1.2	Other WASP researches and their results	350
7.1.3	Improvements of ATN	352
7.2	Generalized Phrase Structure Grammar (GPSG) ...	354
7.2.1	GPSG as compared with other phrase structure grammars.....	354
7.2.2	Syntactic features and grammatical rules	355
7.2.3	Metarules	358

7.2.4	Semantic interpretation	360
7.2.5	GPSG and its axiomatization method	360
7.3	Recent advances in Computational Semantics.....	361
7.3.1	Case grammar in the semantic system of Japanese Language	363
7.3.2	Montague grammar and semantic system in Machine Translation	366
7.3.3	Computational semantics and Mathematical Logic.....	369
7.4	Computational Linguistics and knowledge Engineering	370
7.4.1	Semantics and knowledge representation	370
7.4.2	Incorporation of knowledge engineering techniques in Computational Linguistics.....	371
7.4.3	Computational Linguistics and the Fifth Generation Computer projects	373

目 录

序

代绪论 计算语言学与当代科学技术	1
0.1 当代科学技术的一个特点	2
0.2 作为科学研究对象的语言	5
0.3 计算机科学中的语言学	11
0.4 人工智能中的语言学	14
0.5 通信技术中的语言学	19
0.6 人文科学应用计算机也需要语言学	21
0.7 现代化办公室与语言学	23
0.8 当代技术职业和语言学	24
0.9 计算语言学和数理语言学	25
0.10 计算语言学在当代前沿科学中的位置	28
第一章 现代语言学发展的反思	29
1.1 科学发展的宏观规律	29
1.2 现代语言学发展过程中的几次重心转移	31
1.2.1 传统语法——看作法律的语言学	31
1.2.2 历史—比较语言学——看作生物学的语言学	32