

A Cognitive Investigation and Fuzzy
Classification of Word-class in
Mandarin Chinese

中国当代语言学丛书

汉语词类的 认知研究和模糊划分

上海教育出版社

SHANGHAI
EDUCATIONAL
PUBLISHING
HOUSE

袁毓林 著

中国当代语言学丛书

汉语词类的 认知研究和模糊划分

袁毓林 著

图书在版编目(CIP)数据

汉语词类的认知研究和模糊划分 / 袁毓林著. —上海:上海教育出版社,2010.10
ISBN 978-7-5444-2841-5

I. ①汉... II. ①袁... III. ①汉语—词类—研究
IV. ①H146.2

中国版本图书馆CIP数据核字(2010)第189381号

中国当代语言学丛书

汉语词类的认知研究和模糊划分

袁毓林 著

上海世纪出版股份有限公司出版发行
上海教育出版社

易文网: www.ewen.cc

(上海永福路123号 邮政编码: 200031)

各地新华书店经销 上海中华印刷有限公司印刷

开本 960×640 1/16 印张 28.5 插页 2

2010年10月第1版 2010年10月第1次印刷

印数 1-2,000本

ISBN 978-7-5444-2841-5/H·0147 定价: 36.00元

(如发现质量问题,读者可向工厂调换)

出版者前言

《中国当代语言学》丛书是上海教育出版社的重点出版项目之一。本丛书于 1990 年春由游汝杰(复旦大学)、张洪明(美国威斯康辛大学麦迪逊校区)和唐发铨(本社)策划,并开始组稿和编辑工作。当初拟定的丛书编辑宗旨如下:

中国语言学在 20 世纪二三十年代开始摆脱传统小学的樊篱,进入现代语言学的新阶段。半个多世纪以来,中国语言学已经积累了可观的研究成果,特别是最近十多年来,许多领域在海内外又有了长足的发展。这套丛书希望总结中国当代语言学各个分支学科领域的研究成果,特别是反映最新的研究进展,以期收到承前启后、继往开来的效果,促进中国语言学的现代化。丛书作者则不限国别地域,不限门户学派,唯求高明独到,力争每一本书都能达到当代该学科的最高水平。

1992 年 6 月组稿者将丛书的编辑宗旨、计划和撰稿人名单告知当时在美国访问的朱德熙先生,请他为本丛书撰写总序。朱先生十分赞赏丛书的编辑宗旨,并且认为撰稿者也都是“一时之选”,欣然答允为序。孰料朱先生病情日益加剧,天不假年,未及提笔就不幸逝世。丛书的总序也因此暂付阙如。

从 2000 年开始,刘丹青(中国社会科学院语言研究所)、张荣(本社)也参加了丛书的编辑工作,编委会工作由游汝杰主持,编辑和出版的方针也有所调整。本丛书原拟五年内出齐,结果未能如愿,因为有的作者忙于其他工作,未能按计划完成书稿;有的作者虽然已经完成书稿,但是希望有时间反复修改,使之完善,而不想匆匆交稿。考虑到学术研究需要艰苦的劳动和大量的时间,限定出版时间,不利保证书稿质量。又考虑到学术研究的特点是学无止境、与时俱进、推陈出新,丛书的出版工作也应该是册数开放、不设时限、常出常新。基于上述认识,我们将不设本丛书终止出版的时限,即没有出完的一天。我们不想追

求轰动效应,只要优秀书稿不断出现,我们就不断出版。

本丛书将成为一套长期延续出版的丛书。希望本丛书的编辑和出版方针,能对学术著作的出版工作走上健康发展的道路有所贡献。

上海教育出版社

2003 年 10 月

Preface

The series *Contemporary Chinese Linguistics* is one of the important projects of Shanghai Educational Publishing House. The planning of the series and the soliciting of contributions began in the spring of 1990 with the joint efforts of Rujie You (Fudan University), Hongming Zhang (University of Wisconsin at Madison) and Fanao Tang (Shanghai Educational Publishing House), who were brought together to edit the series by such following common grounds.

Not until the twenties and thirties of 20th century could Chinese linguistics break down the barriers of the traditional Chinese philology and enter its modern stage. Since then, and especially in the last ten years, rapid progress has been made in various different fields of Chinese linguistics and considerable wealth of research achievements have been accumulated. The series tries to present these achievements so as to stimulate the further research.

In June 1992 the editorial committee apprised Prof. Dexi Zhu of the target and the policy of the series with a name list of contributors and invited him to write a preface for the series. Prof. Zhu appreciated the target of the series and the contributors, and promised to write a preface. But his cancer situation turned worse and worse day by day, and did not allow him to write it. So the preface remains unfinished, it is a great pity.

Prof. Liu Danqing of Sciences and Mr. Zhang Rong, the editor of the Shanghai Educational Publishing House, joined the editorial committee from the year of 2000, and the policy of editing and

publication has been adjusted since then. We planned to publish the serials within 5 years at the beginning, but the plan was not realized because some authors were too busy with some projects else, and did not finish writing according to the schedule, while others who had finished the manuscripts would like to revise them to perfect. Considering academic study needs hard work and a plenty of time, if we set deadline, the quality could not be guaranteed, and it is the feature of academic study that there is not limit to knowledge and the old should be weeded through while the new should be brought forth, we will not restrict the number of series volumes and their dates of publication. We would not like to pursuit sensational effort, and what we want to do is to publish qualified manuscripts whenever we have.

This series will be published successively in China. We hope our policy and publication would make contribution to the publication of academic works healthily in China.

Shanghai Educational Publishing House

Oct. 2003

目 录

第一章 汉语词类的认知研究和模糊划分	1
第一节 从原型观念到公理系统和隶属度量表	1
第二节 认识的起点：什么是词类？	2
第三节 工作的前提：汉语有没有词类？	5
第四节 汉语的现实：词类界限和词类归属的模糊性	5
第五节 寻找出路：澄清词类的范畴性质并建立分类程序	8
第六节 词类模糊划分和隶属度分析的意义	11
第二章 词类性质的证伪性测试和本体论检讨	14
第一节 词类性质的证伪性测试	15
2.1.1 “词类是语法功能类”的可证伪性	16
2.1.2 “词类是(语法)意义类”通不过证伪测试	19
第二节 能不能根据分布相似性来划分词类？	22
第三节 鸡和蛋的问题：词的意义和用法哪个更本质？	28
第四节 从分类目的看词类功能观的本体论预设	33
第五节 但书：调和词类划分中意义和功能的矛盾	42
第六节 并非悖论：词类是语法功能类	47
第三章 汉语词类划分的公理化方法	50
第一节 词类的范畴性质	50
3.1.1 作为个体的词和作为范畴的词类	50
3.1.2 词的用法、功能和分布	51
3.1.3 词类是一种原型范畴	51
第二节 词类的发现程序	53
3.2.1 词类的典型成员和非典型成员	53

3.2.2 “框架—槽”和替换操作	54
3.2.3 怎样设置分布框架的值	55
第三节 各别词类的分布框架及其形式表示	55
3.3.1 名词的分布框架和形式定义	55
3.3.2 处所词的分布框架和形式定义	57
3.3.3 时间词的分布框架和形式定义	57
3.3.4 方位词的分布框架和形式定义	58
3.3.5 数词的分布框架和形式定义	59
3.3.6 量词的分布框架和形式定义	61
3.3.7 区别词的分布框架和形式定义	62
3.3.8 动词的分布框架和形式定义	63
3.3.9 形容词的分布框架和形式定义	64
3.3.10 状态词的分布框架和形式定义	65
3.3.11 体词和谓词的分布框架和形式定义	65
3.3.12 数量词的分布框架和形式定义	67
3.3.13 代词的分布框架和形式定义	69
3.3.14 副词的分布框架和形式定义	72
3.3.15 介词的分布框架和形式定义	73
3.3.16 连词的分布框架和形式定义	74
3.3.17 助词的分布框架和形式定义	75
3.3.18 语气词的分布框架和形式定义	76
3.3.19 感叹词的分布框架和形式定义	77
3.3.20 拟声词的分布框架和形式定义	78
第四节 走向公理系统和公理化方法	79
3.4.1 公理系统和公理化方法	79
3.4.2 直观的公理系统和形式的公理系统	81
3.4.3 词类定义离公理系统有多远	83
第五节 汉语词类的准公理系统	85
3.5.1 定义词类的初始概念、公设、形式工具和先行理论	85
3.5.2 各别词类的公理化定义和形式表示	86
3.5.3 句法成分的先行性和独立性	92
第六节 结语：词类范畴的原型效应	95

第四章 汉语词类模糊划分的数学基础	97
第一节 词类划分工作呼唤数学工具	97
第二节 普通集合和模糊集合	98
第三节 子集合和模糊子集	100
第四节 系统聚类法和逐步聚类法	102
第五节 利用模糊关系的模糊聚类方法	105
第六节 硬划分(清晰分类)和软划分(模糊分类)	113
第七节 汉语词类划分的模糊分类策略	118
第五章 汉语词类模糊划分的隶属度量表	120
第一节 引言: 汉语词类模糊划分的基本步骤	120
第二节 权值设定和隶属度计算的约定及有关说明	121
第三节 各别词类的分布特征和权值设定	122
5.3.1 名词的分布特征和权值设定	123
5.3.2 时间词的分布特征和权值设定	126
5.3.3 方位词的分布特征和权值设定	129
5.3.3.1 单纯方位词的分布特征和权值设定	129
5.3.3.2 合成方位词的分布特征和权值设定	132
5.3.3.3 方位词的分布特征和权值设定	135
5.3.4 处所词的分布特征和权值设定	138
5.3.5 数词的分布特征和权值设定	141
5.3.5.1 系数词的分布特征和权值设定	141
5.3.5.2 位数词的分布特征和权值设定	143
5.3.5.3 由系数词和位数词构成的合成数词	145
5.3.5.4 合成数词的分布特征和权值设定	147
5.3.5.5 数词的分布特征和权值设定	149
5.3.6 量词的分布特征和权值设定	151
5.3.6.1 简单量词的分布特征和权值设定	151
5.3.6.2 复合量词的分布特征和权值设定	154
5.3.6.3 量词的分布特征和权值设定	156
5.3.7 准数量词的分布特征和权值设定	157
5.3.7.1 真数量词和准数量词	157

5.3.7.2 准数量词的分布特征和权值设定	158
5.3.8 动词的分布特征和权值设定	161
5.3.9 形容词的分布特征和权值设定	164
5.3.10 状态词的分布特征和权值设定	166
5.3.11 区别词的分布特征和权值设定	169
5.3.12 代词的分布特征和权值设定	171
5.3.12.1 体代词的分布特征和权值设定	171
5.3.12.2 谓代词的分布特征和权值设定	173
5.3.12.3 代词的分布特征和权值设定	174
5.3.13 副词的分布特征和权值设定	176
5.3.14 介词的分布特征和权值设定	178
5.3.15 连词的分布特征和权值设定	180
5.3.16 助词的分布特征和权值设定	182
5.3.17 语气词的分布特征和权值设定	184
5.3.18 感叹词的分布特征和权值设定	186
5.3.19 拟声词的分布特征和权值设定	187
第四节 结语:怎样运用这套量表来进行词类判定	189
第六章 量表中特征选择和权值设定的根据	202
第一节 对立和统计:特征选择和权值设定的原则	202
第二节 各别词类的特征选择和权值设定的根据	203
6.2.1 名词的特征选择和权值设定的说明	203
6.2.1.1 关于名词的分布特征的统计和分析	204
6.2.1.2 名词的特征选择和权值设定的说明	206
6.2.2 时间词的特征选择和权值设定的说明	208
6.2.3 方位词的特征选择和权值设定的说明	210
6.2.3.1 单纯方位词的特征选择和权值设定的说明	210
6.2.3.2 合成方位词的特征选择和权值设定的说明	211
6.2.3.3 方位词的特征选择和权值设定的说明	213
6.2.3.4 关于方位词的进一步说明	214
6.2.4 处所词的特征选择和权值设定的说明	215
6.2.5 数词的特征选择和权值设定的说明	217
6.2.5.1 系数词的特征选择和权值设定的说明	217

6.2.5.2	位数词的特征选择和权值设定的说明	218
6.2.5.3	合成数词的特征选择和权值设定的说明	219
6.2.5.4	数词的特征选择和权值设定的说明	220
6.2.6	量词的特征选择和权值设定的说明	221
6.2.6.1	简单量词的特征选择和权值设定的说明	221
6.2.6.2	复合量词的特征选择和权值设定的说明	222
6.2.6.3	量词的特征选择和权值设定的说明	223
6.2.7	准数量词的特征选择和权值设定的说明	224
6.2.8	动词的特征选择和权值设定的说明	225
6.2.8.1	关于动词的分布特征的统计和分析	226
6.2.8.2	动词的特征选择和权值设定的说明	227
6.2.9	形容词的特征选择和权值设定的说明	229
6.2.9.1	关于形容词的分布特征的统计和分析	229
6.2.9.2	形容词的特征选择和权值设定的说明	230
6.2.10	状态词的特征选择和权值设定的说明	232
6.2.10.1	关于状态词的分布特征的统计和分析	232
6.2.10.2	状态词的特征选择和权值设定的说明	233
6.2.11	区别词的特征选择和权值设定的说明	234
6.2.12	代词的特征选择和权值设定的说明	235
6.2.12.1	体代词的特征选择和权值设定的说明	235
6.2.12.2	谓代词的特征选择和权值设定的说明	237
6.2.12.3	代词的特征选择和权值设定的说明	238
6.2.13	副词的特征选择和权值设定的说明	239
6.2.14	介词的特征选择和权值设定的说明	240
6.2.15	连词的特征选择和权值设定的说明	242
6.2.16	助词的特征选择和权值设定的说明	243
6.2.17	语气词的特征选择和权值设定的说明	244
6.2.18	感叹词的特征选择和权值设定的说明	245
6.2.19	拟声词的特征选择和权值设定的说明	246
第七章 特殊小类的隶属度量表及其说明		248
第一节	引言：大类中分布特殊的小类	248
第二节	七种小类的分布特征和权值设定	249

7.2.1 表人名词的分布特征和权值设定	250
7.2.2 助动词的分布特征和权值设定	252
7.2.3 形式动词的分布特征和权值设定	255
7.2.4 名动词的分布特征和权值设定	258
7.2.5 名形容词的分布特征和权值设定	261
7.2.6 后补词的分布特征和权值设定	264
7.2.7 修饰词的分布特征和权值设定	265
第三节 特征选取和权值设定的根据	267
7.3.1 表人名词的特征选择和权值设定的说明	268
7.3.2 助动词的分布特征和权值设定的说明	269
7.3.3 形式动词的分布特征和权值设定的说明	271
7.3.4 名动词的分布特征和权值设定的说明	273
7.3.5 名形容词的分布特征和权值设定的说明	275
7.3.6 后补词的分布特征和权值设定的说明	277
7.3.7 修饰词的分布特征和权值设定的说明	278
第四节 实例测试和相关讨论	280
第八章 关于词性与表述功能的关系	293
第一节 引言	293
第二节 表述功能和语法位置对词语的选择限制	293
8.2.1 语法位置对词语进行选择限制的依据是什么?	294
8.2.2 什么是表述功能?	298
8.2.3 修饰是不是一种表述功能?	302
第三节 关于表述功能和词性的关系	308
8.3.1 关于表述功能的两个层面	308
8.3.2 关于“表述功能实际上就是词性”的论证方式	310
8.3.3 “句法层面的词性”到底是什么?	313
8.3.4 关于“这本书的出版”的句法分析	314
8.3.5 关于饰词化标记“的”和“地”	319
第九章 关于词类与表述功能类的关系	322
第一节 引言	322
第二节 关于词类的本质的问题	323

9.2.1 关于词类的本质的论证	323
9.2.2 关于表述功能的类别	325
9.2.3 关于词类的泛时、跨语言可比性	327
9.2.4 关于表述功能跟语义的关系	328
第三节 表述功能对于词类划分的不充分性	330
9.3.1 表述功能需要句法功能来确认	331
9.3.2 表述功能对于词类分辨的不充分性	332
9.3.3 表述功能跟句法功能的不完全对应性	335
9.3.4 表述功能的语法地位有限	337
第四节 结语: 语法功能压倒表述功能	340
第十章 关于等价功能和词类划分的标准	341
第一节 引言	341
第二节 分类的依据应该有可观察的表征	342
第三节 关于“功能相容性”和“等价功能”	344
第四节 关于“等价功能的传递性”	345
第五节 修饰语位置对于区分词类的作用	348
第六节 词的使用频率跟语法功能的关系	351
第七节 等价束的句法成分功能解释	353
第八节 唯谓性谓词的词类归属	355
第九节 结语: 语法功能压倒语法意义	356
附录一 现代汉语词类量表的效度调查	360
0. 说明	360
1. 词类模糊划分量表的效度调查表	361
2. 用以检验量表效度的疑难词表	364
3. 词类模糊划分量表效度调查的结果	367
4. 词类模糊划分实施中遇到的问题	372
附录二 《汉语的词类问题》札记	383
0. 说明	383
1. 贺重《词的分类有哪些不同?》札记	383
2. 王力《汉语的词类》札记	386

3. 黎锦熙《中国语法中的“词法”研讨》札记	389
4. 陆宗达《汉语的词的分类》札记	390
5. 周祖谟《划分词类的标准》札记	391
6. 高名凯《关于汉语的词类分别》札记	393
7. 曹伯韩《关于词的形态和词类的意见》札记	395
8. 文炼、胡附《谈词的分类》札记	397
9. Б. Г. 穆德洛夫《汉语是有词类分别的》札记	402
10. 高名凯《再论汉语的词类分别——答 Б. Г. 穆德洛夫同志》札记	402
11. 钟稔、赵淑华、金德厚、王还《汉语的词类问题》札记	403
12. 陈乃凡《汉语里没有词类分别吗?》札记	404
13. 俞敏《形态变化和语法环境》札记	405
14. 曹伯韩《汉语的词类分别问题》札记	406
15. 吕叔湘《关于汉语词类的一些原则性问题》札记	407
附录三 《日本文法事典》词类问题札记	421
0. 说明	421
1. 体词的定义和解说	421
2. 谓词的定义和解说	423
3. 副词的定义和解说	425
4. 虚词的定义和解说	427
5. 附属词的定义和解说	428
参考文献	430
后记	438

第一章 汉语词类的认知 研究和模糊划分

本章概略地介绍我们对汉语词类进行认知研究和模糊划分的操作程序:基于词类是原型范畴(因而属于同一词类的词有典型成员和非典型成员之别)这种认识,对于典型成员,采取公理化的方法,用具有区别性的分布特征下严格定义;对于非典型成员,采取模糊数学的办法,建立各别词类的隶属度量表,据此对有关的词进行隶属度分析,定量化地确定它们的词类归属,从而实现汉语词类的模糊划分。本章还介绍了这种研究的有关理论背景,包括对词类的性质的认识、对汉语有无词类的论证、汉语词类划分的现实困境、寻找走出困境的相关办法,以及这项工作的意义等内容。

第一节 从原型观念到公理 系统和隶属度量表

汉语词类一直是语法理论上的老大难问题。虽然大多数学者承认词类是语法功能类,是根据词的语法功能(分布)而分出来的类;但是,由于词的数量十分巨大,不同的词在分布上的关系十分复杂,突出地表现为人们在语感上认为是同一类的词并不一定有某项/组为其他词类所无的分布特征(即区别性特征),因而目前还没有一个为多数学者公认,并且符合汉语实际的词类划分标准和系统。这直接给汉语信息处理和对外汉语教学等应用领域带来了不便,因为这两种工作都需要一个可供参照的分类体系和一个精确的词类标注系统。

从实际操作上看,汉语词类划分是个相当棘手的问题,表现为:脱离了分布就没法分类,但是完全依靠分布就分不出类来。为了解决这

个问题,我们在1993年提出了汉语词类是一种原型范畴的理论(详见袁毓林(1995a))。但是,光有这样一个笼统和抽象的原型观念是远远不够的;必须有一套行之有效的操作程序,以保证我们可以有根有据地把词类的非典型成员归入合适的类中去。为了在操作上实现这一思想,我们在1999年通过建立公理系统来为各别词类的典型成员设立严格的区别性分布框架(详见袁毓林(2000a));在此基础上,我们进一步用模糊数学的办法来为各别词类的非典型成员进行隶属度分析,并参考心理学上的心理测量量表(scale),建立了词类模糊划分的隶属度量表,基本上发展出了一套形式化程度高,又富有弹性的、适合于汉语词类的分析体系和操作方法(详见袁毓林(2003;2005a、b))。

下面,我们分别说明进行这项研究的认识起点、工作前提、现实困境和突破困境的方法,特别是实际的操作程序,以及这项工作的意义。

第二节 认识的起点:什么是词类?

我们相信,词类是一种基于词的用法(usage)的分类。当然,“用法”是一种通俗的说法,用语言学专门术语来说就是词的语法功能(grammatical function)。但是,“语法功能”是一种比较抽象的说法,需要若干先行概念才能说清楚。比如,Bloomfield (1926) § VI: Construction, Categories, Parts of Speech 给出了这样的定义和假设:

Def. Such recurrent sameness of order are construction. (p. 28)

定义:这种反复出现的顺序上相同的东西就是结构。

Def. Each of the ordered units in a construction is a position. (p. 28)

定义:结构中每一个按顺序[排列]的单元就是位置。

Assumption. Each position in a construction can be filled only by certain forms. (p. 28)

假设:结构中每一个位置只能由某些形式来填充。

Def. The positions in which a form occurs are its function. (p. 29)