



中国书籍文库
China Books Library

汇集优秀原创学术论著
推动科研成果转化交流

现代汉语指人名词研究

XianDai HanYu ZhiRenMingCi YanJiu

韩 蕾 著



中国书籍出版社
China Book Press

013035304

 中国书集
China Book

H146.2
22

现代汉语指人名词研究

XianDai HanYu ZhiRenMingCi YanJiu

韩 蕾 著



H146.2
22



北航

C1642516



中国书籍出版社
China Book Press

图书在版编目(CIP)数据

现代汉语指人名词研究/韩蕾著. —北京:中国书籍出版社,

2012. 9

ISBN 978 - 7 - 5068 - 3077 - 5

I. ①现… II. ①韩… III. ①现代汉语—名词—研究

IV. ①H146. 2

中国版本图书馆 CIP 数据核字(2012)第 203070 号

责任编辑/ 牧 人

责任印制/ 孙马飞 张智勇

封面设计/ 中联学林

出版发行/ 中国书籍出版社

地 址: 北京市丰台区三路居路 97 号(邮编:100073)

电 话: (010)52257143(总编室) (010)52257153(发行部)

电子邮箱: chinabp@ vip. sina. com

经 销/ 全国新华书店

印 刷/ 三河市华东印刷有限公司

开 本/ 710 毫米 × 1000 毫米 1/16

印 张/ 16. 5

字 数/ 297 千字

版 次/ 2013 年 1 月第 1 版 2013 年 1 月第 1 次印刷

书 号/ ISBN 978 - 7 - 5068 - 3077 - 5

定 价/ 48. 00 元

版权所有 翻印必究

序 言

(一)

应该公正地说，是计算机科学家的探索，开辟出了自然语言信息处理这一具有重大突破意义的领域，开创了计算语言学这一重要的分支学科。也正是因为计算语言学的探索进入了一个关键性的阶段，急切地寻求语言学家的协助和合作；这一合乎科学发展的动向得到了语言学界的迅速反响，计算语言学成了计算机科学与语言学研究相结合的一门交叉科学，成了应用语言学研究中的一门显学。

从上个世纪八九十年代起，中国国家语委就开始了汉语语料库的建设，开始了一系列计算语言学的理论建设和人才培养工作。本书的作者就是在这样的大背景下成长起来的一位应用语言学的青年才俊。她参加了国家语委、语言文字应用研究所许嘉璐教授、傅永和教授领导的国家社科九五重大项目《信息处理用现代汉语词汇研究》的研究工作，参加了北京大学计算语言学研究俞士汶教授领导的《人民日报标注语料库》的研制工作，参加了上海师范大学《当代汉语语料库（一期）》的研制工作，完成了她的博士论文《信息处理人名词研究》。

正式工作后，她又在博士论文的基础上，继续深入研究，对有关现代汉语指人名词的方方面面的属性描写和分析规则进行了一系列的研究。本书就是这一系列研究工作的集大成性质的成果。

(二)

语言信息处理的词切分、词性标注、短语结构标注和句子分析都会遇到语言单位的分类和句法语义属性的描写的问题，需要探讨语言单位的分类机制和原则，语言结构的组配规律和应用限制问题，这些都涉及句法、语义、语用三个平面的属性描写与规则探讨。目前的研究主要是在句法语义属性的描写上，语用层面的研究比较少见。

汉语语言学在句法层面上的研究受到了汉语信息处理需求的严重挑战。传统的词分类研究的分析深度和句法语义属性描写的细度远远达不到汉语信息处理的精度要求。这就要求开展面向信息处理的理论、方法、机制、规则的研究，开展细致的句法语义语用属性分析描写的研究。本书就是这类面向汉语信息处理的现代汉语基础研究，旨在对现代汉语指人名词的再分类规则作深度探讨，对现代汉语指人名词的句法语义属性作细致的描写。更具有创新意义的是在进行指人名词的分类属性描写时，纳入了“指称性”的标准，探讨了句法组构中的指称性的原理和识别方法，深入到了句法——语用属性描写的层面。

(三)

跟传统的汉语本体研究相比，面向信息处理的汉语基础研究具有以下几方面的特征：

可验证性是面向信息处理的汉语基础研究的第一个特点。现代语言学认为操本族语者的语言直觉（Intuition）是判断一种语言形式是否合乎语法，是否可以接受的根本标准。但是这种语言直觉是整个言语社团的社会语言心理，它是一种客观存在，但无法验证，所谓“只可意会，

不可言传”。传统的语言研究实际上是依赖于语言学家的直觉 (Linguistic Intuition)，即“内省法”。内省法提高了语言分析的精确性和精致性，但仍没有解决可验证性这个根本问题，因此对某一个语言形式的合法性、可接受性，不同的语言学家往往有不同的判断和解释，这是一种常见的现象。面向信息处理的汉语基础研究，强调从实际语料出发，运用归纳法抽绎出规则规律来，这就有了词切分、词性标注（辨歧）、短语结构（结构层次、结构功能、结构关系）分析，等等一系列的工作程序，运用的是“概率统计法”。概率统计法大大加强了可验证性。但是，即使是大规模的语料统计，也免不了遇到言语中的种种变异：临时性的出错、言语应用领域（语域）差异造成的统计偏差等等。从理论上讲，全域性的或者超大规模的语料统计应该可以消解这些偏差，但实际上无法做到真正的“全域”和“超大规模”。这就涉及可操作性了。

可操作性是面向信息处理的汉语基础研究的第二个重要特点。本书运用概率统计法和内省法相结合研究方法，用各自的研究的结果互校，作对比分析，这是一种切实可行且有实效的好方法。实际上最强调“从言语实际出发”的语言学家在语言研究中也不排斥“内省式”语言学分析的“干预”。话语语言学家是最强调第一手口语语料的真实性和完整性的了，通过隐蔽性的现场录音、电话窃听、非暗示性的调查对话等等方法，收集到大量语料，然后进行分析。然而在统计分析以前，还必须对录音记录稿进行整理，把“嗯、啊、这个这个”之类的无意义的“停顿性垫话”、反复啰嗦的赘余成分，说了半截子话再重新起头说产生的“非结构性成分”等等都删除。这种“整理”，实际上就是语言学家的“内省式”干预。此外，本书中采用的“多层次分类标注法”、“分（小）类的属性描写法”等等，也都是加强面向信息处理的汉语基础研究的操作性的方法。

严密的系统性是面向信息处理的汉语基础研究的第三个重要特点。人们在听说读写的言语实践中得到自身的语言知识库的全力支持，得到在长年的语言交际中积累起来的言语交际知识和语境知识的大力支持，

这种知识库的建设和应用是随机的、开放式的。而计算机的运行则是一个相对封闭的系统，面向信息处理的汉语基础研究必须更强调严密的系统性。本书运用组合分析的原理进行分类的聚合分析，提出分类的句法语义属性描写的“框架测试法”，对关系类指人名词分析时运用了名词配价分析的原理；在对指人名词同位结构进行“内部限制”和“外部限制”的分析时，注意到了句法语义属性的结合处理，注意到了上下文因素、说话语境因素，这些都体现了汉语信息处理工程的系统性，是一种很有启发的探索。

(四)

诸多方家都说面向信息处理的汉语基础研究“很重要”，但是也“很难”。

既然“很重要”，就要有人来做；因为“很难”，更需要不畏难的志士仁人来做。

现在这方面的探索已经不再是“筚路蓝缕”的阶段了，“深入”、“细化”的难度更大。

本书作者的辛勤劳动，只是面向信息处理的汉语基础研究的一个小课题，只是汉语信息处理工程的一块砖石。但是，宏伟大厦就是在这样的一块块砖石的基础上建筑起来的。

成果有价值，人才正年轻，前途灿烂辉煌。

范开泰

前 言

本书运用结构主义语言学、语料库语言学、认知语言学、功能语言学等多种理论,对现代汉语指人名词进行研究。

第一章介绍中文信息处理的基本背景,阐明现代汉语语料库建设与汉语语法研究相互依存的关系。

第二章指出面向中文信息处理的“名·名”词语串研究还比较薄弱,本研究的主旨是帮助计算机正确完成跟指人名词构成的双名词语串有关的句处理任务。此外,还阐明了研究的选题依据、性质定位和基本思路。

第三章强调语料库加工过程中应对名词进行多层级分类标注,并对提前分类标注的可行性、必要性等问题进行论证。

第四章在以往研究的基础上,阐明指人名词的判定标准及分类的必要性,提出框架测试法,论证该方法的理据并进行分类实践,归纳出现代汉语指人名词分类系统。

第五章探讨双项指人名词组合的内部制约。先是确立同位短语独立的语法地位,接着描写指人同位组构模式、模式间的连续统及组构规律,然后探讨同位短语的认知语义基础及作为独立类的理据。

第六章从大规模语料中提取对短语定界有用的外部规则,主要考察

词语串邻近词语的属性、一定的句式对双名组构的外部限制。并论证边界确认规则与否认规则、外部规则与内部规则之间的关系。

第七章总结本书研究特点及其理论价值和实用价值，并探讨了由计算机处理汉语引发的语言工程的系统性、两可现象与语法规则、语法/非语法形式与静态/动态等问题。

作者

《中国书籍文库》部分书目

一、政治与哲学

1	马克思主义大众化——基于国际金融危机视野下的研究	2	马克思主义哲学前沿理论研究
3	社会关系与和谐社会——马克思社会关系视域中的“和谐社会”解读	4	孙中山民生社会主义思想研究
5	保守主义：一种审慎的政治哲学	6	复杂性科学研究
7	九鬼周造的哲学——漂泊之魂	8	论黑格尔哲学

二、历史与文化

9	《老子》与现代人生	10	二十五史梦文化解读
11	桂海越裔文化钩沉	12	历史文化村镇景观保护与开发利用
13	民俗信仰与双向认知	14	透视大众文化
15	文化观与翻译观——鲁迅、林语堂文化翻译对比研究	16	行政法视野下非物质文化遗产保护研究
17	中国动漫文化：本体与心理论	18	中国货币文化简史

19	《四书》微揽	20	边疆民族史探究
21	东亚坐标中的跨国人物研究	22	东亚坐标中的遣隋唐使研究
23	东亚坐标中的书籍之路研究	24	两汉之际社会与文学
25	训诂通论与实践	26	雅典海上帝国研究

三、文学与艺术

27	《尤利西斯》的小说艺术	28	历代中国画技法之美
29	美术教育质的研究案例	30	美术考古文存
31	商周青铜器与青铜器雕塑艺术	32	舞蹈创作思维
33	歌谣的多学科研究	34	美学理论视野中的文学翻译研究
35	《红楼梦》研究新论	36	谁为情种——《红楼梦》精神生态论

四、法律与社会

37	城乡一体化之现代农业形态	38	当代都市报研究
39	当代中国科技进步与低碳社会构建	40	地方治理创新视角下的地方政府债务危机防范研究
41	电子政府与服务型政府	42	犯罪空间分析与治安系统优化
43	服务行政与服务型政府	44	公共安全管理研究
45	公共选择理论探索	46	农村劳动者素质与现代化

47	生态价值取向研究	48	现代性批判的技术与方法
49	中国和平发展战略实施的国际环境	50	公益诉讼——基于经济法视野下的研究
51	经济法基础理论与实务问题研究	52	侵权责任法案解
53	物流法律制度研究		

五、经济与管理

54	信息资源获取与应用	55	资本论的方法研究
56	数据挖掘模式下的审计风险预警系统研究	57	公平与效率不可兼得吗——美国、瑞典模式的比较与借鉴
58	网络信息资源理论与实践研究	59	比较：制度经济和产权理论
60	资源型城市产业兴衰与转化之规律	61	现代人力资源开发与 E 时代
62	市场经济与区域发展	63	经济全球化与社会主义经济体制
64	北京发展连锁经营理论及对策研究	65	中国教育经济与管理研究
66	教学档案的管理与信息化建设	67	人力资源：高校无形资产管理危机的核心要素研究
68	绩效导向型公共预算管理研究	69	精细化管理
70	银行风险管理研究——以民营商业为例	71	政府海洋产业管理研究

六、教育与语言

72	超文本写作论	73	对应阅读心理的表达意识
74	教师校本培训项目制	75	教师在校本教研中成长
76	高等学校管理新视野——基于师资队伍建设与教学质量研究	77	教育伦理探微
78	培养学生创新精神和实践能力的支持系统研究	79	多维大学校园文化研究
80	高校数字图书馆建设评估研究	81	图书馆核心价值及其实现策略
82	图书馆科学发展的理念与实践	83	现代图书馆及数字资源利用
84	汉语方言地理学——入门与实践	85	汉语交际中的得体性
86	现代汉语指人名词研究	87	言语交际新思维
88	语言问题八讲	89	语言与逻辑
90	语用学研究与运用		

七、其他

91	技术认识范畴研究	92	数字信息检索与创新
93	运动性心理疲劳研究	94	现代旅游业应用型人才培养研究
95	创新整合论——科技创新与文化创新的整合机制	96	竞技体育与科技前沿
97	钱谦益年谱		



北航

C1642516

13.1-15-15

目 录

CONTENTS

第一章 引言:语料库与语言学研究	1
1. 自然语言处理与语料库	/ 1
2. 语料库发展与语言学理论	/ 3
3. 汉语语料库与中文信息处理	/ 4
4. 现代汉语语料库与语法研究	/ 11
第二章 研究背景	23
1. 选题的目的	/ 23
2. 选题的依据	/ 25
3. 面向信息处理的“名·名”研究现状简介	/ 29
4. 选题的研究定位	/ 30
5. 选题的研究思路	/ 34

第三章 关于名词多层级分类标注的构想	38
1. 理论架构	/ 38
2. 名词多层级加工实践	/ 43
3. 对 MCT 法的理论反思	/ 46
4. 小 结	/ 50
第四章 指人名词的分类研究	51
1. 已有的分类研究	/ 51
2. 指人名词的确定	/ 54
3. 指人名词分类的必要性及方法	/ 55
4. 框架测试法的理据及运用	/ 57
5. 指人名词分类结果	/ 62
6. 余 论	/ 77
第五章 指人名词同位组构的内部限制	80
1. “同位”概念的历史	/ 81
2. 同位短语的地位	/ 83
3. 现存的问题	/ 86
4. 指人名词同位组构模式	/ 87
5. 余 论	/ 112
第六章 指人名词同位组构的外部限制	116
1. 外部研究要解决的问题	/ 116
2. 外部定界规则的类型、表述形式和特点	/ 119
3. 确认规则	/ 123
4. 否认规则	/ 137
5. 有待进一步研究的若干问题	/ 147
6. 余论	/ 154

第七章 小结	156
1. 本书研究的总结	/ 156
2. 与本书研究相关的若干问题	/ 158
参考文献	166
附录 1:指人名词表	180
附录 2:名词研究札记二则	187
附录 3:现代汉语否定肯定对用格式研究	214
后 记	244

第一章

引言：语料库与语言学研究

1. 自然语言处理与语料库

自然语言处理（Natural Language Processing，简称 NLP）就是以电子计算机为工具对自然语言信息进行各种类型处理和加工的技术。

1946 年第一台电子计算机诞生后不久，人们就想用计算机来研究和处理自然语言。从上个世纪 50 年代初期到 60 年代中期，机器翻译一直是研究的中心。当时采用的主要是“词对词”的翻译方式，译文效果很差。机器翻译的困境使人们意识到：要让计算机真正具备类似于人那样的处理自然语言的能力，就必须对语言自身的规律进行深入的挖掘。因此，自 60 年代中期以后，人们便开始重视研究自然语言的语法、语义和语用等基本问题，并尝试实现计算机的自然语言理解，即人机对话，也就是人用自然语言向计算机提出问题，相应的，计算机也能够理解并用自然语言做出回答。当前，除了机器翻译和自然语言理解之外，自然语言处理的内容还涉及情报自动检索、语音自动识别与合成、文字自动识别、词典自动编纂、自动文摘、计算机辅助教学等众多领域。（冯志伟 1996）