

面向信息处理的

现代汉语

“名+名”逻辑语义研究

宋春阳 著

MIANXIANG XINXI CHULI DE
XIANDAI HANYU
MING JIA MING LUOJI YUYI YANJIU

学林出版社

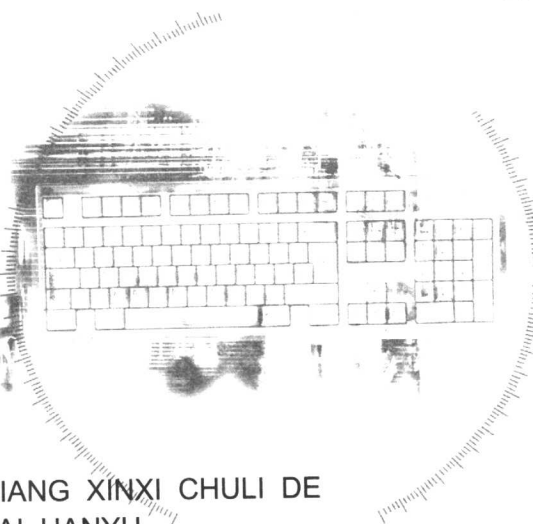
上海交通大学国际教育学院科研基金资助

面向信息处理的

现代汉语

“名+名”逻辑语义研究

宋春阳 著



MIANXIANG XINXI CHULI DE
XIANDAI HANYU
MING JIA MING LUOJI YUYI YANJIU

学林出版社

图书在版编目(CIP)数据

面向信息处理的现代汉语“名+名”逻辑语义研究 /
宋春阳著. —上海: 学林出版社, 2005. 5

ISBN 7 - 80668 - 907 - 9

I. 面... II. 宋... III. 汉语—名词—语义学—研究—现代 IV. H146.2

中国版本图书馆 CIP 数据核字(2005)第 024535 号

面向信息处理的现代汉语“名+名”逻辑语义研究



作 者—— 宋春阳

特约编辑—— 刘 娴

责任编辑—— 李晓梅

封面设计—— 周剑峰

出 版—— 上海世纪出版集团

学林出版社(上海钦州南路 81 号 3 楼)

电话: 64515005 传真: 64515005

发 行—— 新华书店上海发行所

学林图书发行部(钦州南路 81 号 1 楼)

电话: 64515012 传真: 64844088

照 排—— 南京展望文化发展有限公司

刷—— 上海港东印刷厂

开 本—— 850×1168 1/32

印 张—— 7.75

字 数—— 17 万

版 次—— 2005 年 5 月第 1 版

2005 年 5 月第 1 次印刷

印 数—— 4 000 册

书 号—— ISBN 7 - 80668 - 907 - 9/H · 38

定 价—— 16.00 元

序 言

本书的作者在攻读博士学位期间参与了多项国家、部委的有关中文信息处理如语料库切分标注、内涵逻辑语义词典等研究项目,在计算语言学的语义分析方面有切身的体会和经验积累,并始终坚持这个看来是十分艰难的但又富有很高学术价值和实用意义的研究方向。

自然语言处理当今尚处于低谷,尽管各种统计方法具有很广泛的适用性,但用到语言处理上依然不尽如人意。瓶颈不在于算法之类的关键技术上,而在于表达人类认知和思维的语言表达式与当代计算技术中间的衔接环节上。语言表达式这样的非结构化数据如何转换为计算机可以计算的结构化数据。五十年来主导语言学研究的结构主义所提供的结构分布理论如词类及其分布、替换方法都是为人这样的读者提供理据的,因为语言本身是人类使用的,所以这样的解释和理据并不会遭遇根本的质疑或困惑。近十年来,关于语义、功能、情景等方面各种理论的产生和活跃,使得语言研究大大地推进了一步。尤其是认知理论和模型理论的研究,试图在人类认知思维到语言表达式的形式化之间,寻找一种沟通,寻找合理的理据来解释,人是如何用语言来表达思维的,语言表达了哪些概念及概念之间的关系。比起单纯的词类分析和短语分析来说明白得多,深刻得多。

中文信息处理当前已从词处理逐步走向句处理,而且开发

了许多实用的应用系统,但是从概念分析的角度看,着重在于如何从小的结构颗粒所表达的概念意义组合成更大的合成结构所表达的概念意义,这种组合分析是关键之关键。

本书作者把研究重点放在名词及其所表达概念特征对应关系上,不仅是单个名词,更重要的是复合名词的概念的理据解释上。将结构形式与概念关系两者对应起来(数学上称为同态对应,最理想的情况是同构对应(一一对应关系))。作者是语言学出身,并不排斥语言学教学和研究中的优良传统和有益知识,但超脱了传统结构主义及格语法等众多熟知的现有理论,表达了一种新的内涵逻辑概念语义思想,用概念特征来解释纷纭复杂的汉语语义,尽量不是求繁而是求简,用化繁为简的思路来思考问题,表达了一种全新的研究汉语名词概念语义组合规律的思路和方法,给人一种清新的感觉。研究成果是汉语内涵逻辑模型研究的一个组成部分,已直接用于中文信息处理的多个应用项目中。作者的思想方法不仅充分表达了汉语在语义上概念直接组合的特点,在理论方法上也是可与国外的同行相媲美的。

宋春阳是我兼职于上海师范大学的博士生,语言学出身,从传统语言学立场视角转向计算语言学尤其是逻辑模型理论,从形象思维转向抽象思维,经历了一个痛苦艰辛的磨难过程,他能接受这样的挑战,而且用成果及著作来证明实实在在的业绩,而不是炒作,是值得祝贺的。现将这本书推荐给读者,让广大读者也有所知晓,引起兴趣和思考,是一件很高兴的事。

是为序。

陆汝占

2005年3月28日

前言

本书是面向中文信息处理的现代汉语名词研究,具体解决策略是以内涵逻辑理论为指导,吸取训诂学精华,通过提取词的抽象类义来把握名词的语义结构,以达到使汉语语义分析和解释形式化的目的。本书具体分析了有代表性的“关”、“木”、“刀”、“车”、“手”、“书”等单音节词的语义结构,初步建立了含处所义、材料义、部分义、人造物义以及符号义等词的语义结构模式,并以此来分别解释说明这些名词跟其他名词的语义组合结构、句法搭配以及所受语义限制。通过内涵特征来解释纷纭复杂的汉语语义,这是一种全新的研究汉语名词语义组合规律的思路和方法,可为中文信息处理提供新的理论支撑,解决单纯统计方法不能解决的问题。

第一章论述研究背景、现状和存在的问题以及选题的目的和解决策略。汉语属语义型语言,没有形式标记,因而需要语义知识的建设。当前中文信息处理单纯使用统计方法存在不足,基于语义规则的方法具有可观的实用价值。汉语语法学界和计算机界在名词的性质、定义和分类及功能上存在一些未解决的问题,有关“名+名”的语义研究也亟待加强。本文选题的目的是通过名词语义结构以及结构各成员间的搭配关系分析,建立“名+名”的语义组合模式,为机器切分、语义理解提供可形式化的规则。这可解决同形词的语义识别、多项 NP 语义结构分析、

未登录词的语义理解等等。分析了以往优秀的语义学研究成果,特别是吸取了义素分析法和训诂学理论的合理内核,创造性地提出了内涵特征和叠置原理、类义抽象解决汉语语义问题的策略。

第二章主要是建立内涵逻辑理论框架并用于词语语义解释和理解。文章首先定义了词、词义、概念、内涵、外延及相互关系。运用集合理论解释了概念内涵之间、外延之间的各自的关系。认为名词是指称实体的词语项,实体包括物理实体和抽象实体两类。名词具有内涵性质即语义特征,提出了“定义特征”和“情景区分特征”两个概念,并分析了名词定义特征的构成要素,介绍了“定义特征”的确定方法。通过实例论证了概念的内涵特征组合规律,解释了特征的层级性和方向性对句法组合的限制,以及两个名词的内涵外延组合方式。分析了概念内涵义和外延义与词类的对应问题,认为内涵特征算子是名词取内涵义的标记。最后讨论了具体语境中的有指与无指问题,指出理解语义离不开情景和上下文。

第三章主要论述抽象类义理论。文章首先探讨了汉语的基本单位、词的概念,由此引出要依靠叠置原理来作为词表收词的标准。进而提出运用抽象类义理论应对汉语语言单位界限不清、大量的多义词以及兼类词的语言事实。论述了抽象类义的原则、方法以及与内涵特征的关系。最后探讨了由语素义求取组合义的争论症结,认为处理内涵特征构词要区分频繁结构和非频繁结构,注意吸收认知语义学的研究成果。

第四章是本文的重点,通过分析实例建立汉语名词语义结构的模型。文章首先确定了选取例词的范围、依据。基本分析方法如下:首先通过收集相关辞典对例词解释,指出词典收词和释义存在的问题并对辞典中词的义项进行归并,然后提取抽象

类义、确定定义特征及分析构词的内涵特征,并排出例词的语义特征表,最后验证语义特征组合规律。以此来分别解释与名词有关的句法搭配以及所受语义限制。文章具体分析了有代表性的“关”、“木”、“刀”、“车”、“手”、“书”等单音节词的语义结构。

第五章是运用内涵逻辑理论对名词谓语句分析,论证名词谓语句的性质、分类,具体分析了名词概念间的关系对句法位置的限制,并加以形式化的规律总结。

第六章是总结。总结本文的研究特点、理论意义、提出后续研究方向和拟要解决的问题。

Preface

Based on the theory of Intensional Logic Semantics, absorbing the delicacy of traditional Chinese lexicology, this paper focus on the semantic analysis of nouns of modern Chinese for Chinese Information processing, attempting to study the semantic structure of nouns by extracting its abstract synonymity, in order to realize the formal expression of Chinese semantic analysis and explanation. After analyzing the semantic structure of the representative monosyllable nouns “guan” (barrier) “mu”(wood) “dao”(knife) “che”(vehicle) “shou”(hand) “shu”(book), we build the initial semantic structure model of nouns which contain the meaning of Location, material, part, artificial articles etc. Based on it, we can throw lights on the semantic structure, on the syntactic combination, on the semantic restriction of the lexical items respectively when the nouns collocate with other nouns. This is a new approach and method on the study of semantic combination rules of Chinese nouns which attempting to explain complicated Chinese meaning by intentional features. It also can provide new theory support to the Chinese Information processing and solve the problem which Statistical method fails to do. There are six

chapters in this paper.

Chapter one gives the background, the situation, the problems, the target and our strategy. Chinese belongs to the language of semantic type without formal mark, so it needs the construction of semantic informations. There's shortcoming in the statistical method for Chinese Information processing but there's great advantage in the method based on semantic rules. There are something unsolved on nouns' property, definition, classification and function in the Chinese grammarian circle and computerman circle, so it's necessary to enhance the semantic study on "noun+noun". This paper's target is build the semantic combination model of "noun+noun" by analyzing the semantic structure and collocation among the members in the structure of nouns, in order to provide the formal rules for machine syncopation and semantic understanding. It's also a new method for semantic differentiation of homomorphic words, semantic structure analysis of NP phrase, semantic understanding of newcome words. After analyzing the excellent study achievement of preceding semantics and absorbing the delicacy of plereme analysis method and traditional Chinese lexicology, we inventively set our strategy on intentional features and congruence principle, abstract synonymity which can solve semantic problems of Chinese.

Chapter two aims to build the framework with intentional logic theory for semantic explanation and understanding of words. As a first step, this paper gives the definition on words, meaning, concept, intention and denotation and ex-

plain the relation between the intentions and between the denotations by Set theory. We consider nouns as lexical items referring entities which include physical entities and abstract entities, so nouns own its intentional property, namely semantic feature. After founding the two concepts: “definitional feature” and “situational differentiating feature”, we analyze the composing elements of definitional feature of nouns and introduce the method for judging definitional feature. We illustrate the combining rules of intentions of concepts, giving explanation for the restrictions on syntactic collocation due to hierarchy and orientation of features, for the combining rules of intention and denotation between two nouns. After analyzing how intention and denotation of concepts correspond to word class, we argue that intentional feature operator is the mark for judging whether a noun adopts intentional meaning.

Chapter three discusses the theory of abstract synonymity. On defining the basic unit of Chinese and concept of word, we attempt to set congruence principle as our criterion for deciding whether a glossary includes a word on problem. We show the exact ways to deal with the problem of boundary of Chinese language units, the problem of polysemous words, the problem of the words with the multiple parts of speech. We also explain the principle and methods of the theory of abstract synonymity and its relation to intentional features. As for the arguments on whether we should get integral meaning from morphemes of a word, we argue we should differentiate high-frequency structure and low-frequency structure when we deal

with word-building by intentional features, paying more attention to absorbing the achievement of cognitive semantics.

Chapter four is the most important part of this paper. It is a sample of building the semantic structure model of Chinese nouns by analyzing examples. It also introduces the scope and basis of the selected words. Our analyzing methods are summarized as follows; at the first step, after collecting the explanations for example words in relative dictionary, we find out the problem in the explanation and inclusion of words of dictionaries. At the second step, we extract abstract synonymity of words and confirm definition features and analyze the intentional features for composing a word in order to give the details of the semantic features of example words. At last, we test and verify the combining rule of semantic features and apply it to the explanation for syntactic collocation and its restriction from meaning which is relative to nouns. As a demonstration, we analyze the semantic structures of the representative monosyllable nouns “guan” “mu” “dao” “che” “shou” “shu”.

Chapter five analyzes the sentence with nominal predicate by the theory of Intentional Logic. It gives the definition of the property and classification of the sentences with nominal predicate, analyzing the relations between noun concepts and its restriction on syntactic position. At last, it gives formal expression of rules for conclusion.

Chapter six is the conclusion of this paper. It concludes the study characteristics, the significance of the theory, the work to be done and the problem to be solved in the future.

目 录

序言	1
前言	1
第一章 绪论——现状与回顾	1
1.1 中文信息处理遇到的困难——问题与瓶颈	1
1.1.1 自然语言处理的过程	1
1.1.2 中文信息处理的困难	2
1.1.3 当前中文信息处理的流派和策略	3
1.1.4 信息处理工作原理及单纯使用统计方法的 不足	4
1.1.5 基于内涵语义规则信息处理的实用目标	6
1.1.6 基于内涵逻辑模型论的语义分析的总体 设计思想	8
1.2 面向信息处理的名词及“名+名”研究现状	11
1.2.1 汉语语法学界的研究	11
1.2.2 计算机界对名词的研究	12
1.2.3 名词定义、性质、分类研究存在的问题	14
1.2.4 “名+名”的研究亟待加强	18
1.3 选题的目的和理论依据	22

1.3.1	研究目的	22
1.3.2	依据的主要语义理论	26
1.3.3	语义理论与“名+名”研究	31
1.4	研究原则、方法和技术路线	34
1.4.1	原则	34
1.4.2	方法和技术路线——基于内涵逻辑的句法 分析	37
1.4.3	研究范围和步骤	41
第二章 内涵逻辑理论——用于词语语义解释和理解		43
2.0	引言	43
2.0.1	本课题的创新性	43
2.0.2	内涵逻辑解决语义理解的策略	44
2.1	名词内涵——语义特征	45
2.1.1	关于词语项、词义、概念、内涵、外延的 定义及关系	45
2.1.2	概念外延指称属性的形式化描写	48
2.1.3	概念之间的关系	50
2.2	内涵逻辑理论的运用	53
2.2.1	名词的定义	53
2.2.2	名词的内涵性质	54
2.2.3	名词与可能世界的联系——外延指称	55
2.2.4	专有名词内部的语义构成	55
2.3	内涵特征结构	56
2.3.1	关于语义特征的概念	56
2.3.2	“定义特征”和“情景区分特征”	57
2.3.3	“特征”概念提出的目的	58

2.3.4	概念的内涵特征——事物属性结构	59
2.3.5	名词定义特征的构成要素	60
2.3.6	“定义特征”的确定方法	61
2.3.7	特征层级性和方向性	65
2.3.8	概念的内涵特征组合规律	67
2.4	内涵外延对组合义及句法结构的影响	73
2.4.1	位置对概念间关系的影响	73
2.4.2	两个名词的内涵外延组合方式	74
2.4.3	关于概念内涵义和外延义与词类的 对应问题	76
2.4.4	内涵义与外延义的判定	78
2.4.5	名词取内涵义的表现形式——内涵特征 算子	79
2.4.6	外延义的提取算子	82
2.5	定指与不定指	83
2.5.1	定指与不定指定义	83
2.6	要在语境中理解词义	84
2.6.1	静态的词典义	84
2.6.2	动态语境义	85
第三章	类义抽象理论	87
3.1	现代汉语词与短语的讨论	88
3.1.1	汉语基本语法、语义单位	88
3.1.2	词的概念及定义词的标准	89
3.1.3	基于信息理解的标准	93
3.1.4	词的组成单位	95
3.1.5	词类的划分	96

3.2	叠置原理	99
3.2.1	什么是叠置原理	99
3.2.2	组合义的求解困难	100
3.2.3	叠置原理的操作策略和方法	102
3.3	抽象类义	102
3.3.0	抽象的理论依据	102
3.3.1	类义抽象的定义	103
3.3.2	类义抽象的目的	104
3.3.3	类义抽象的原则及方法	108
3.4	由语素义求取组合义的理论探讨	111
3.4.1	合成词产生的原因	111
3.4.2	传统构词理论的不足	112
3.4.3	关于依据语素理解词义的讨论	117
3.4.4	向心与离心	118
3.4.5	我们处理构词的方法	120
3.5	处理内涵特征构词要注意的几个问题	122
3.5.1	要区分频繁结构和非频繁结构	122
3.5.2	注意吸收认知语义学的研究成果	123
3.5.3	致力于构词的逻辑模型建立	127
第四章	基于实例的名+名语义结构研究	128
4.0	选词的范围、依据和对象	128
4.0.1	选词的范围、原则	128
4.0.2	抽象的依据辞书及语料	129
4.0.3	考察对象	129
4.1	“关”的类义抽象	130
4.1.1	对相关辞典解释进行收集	130

4.1.2	对辞典的义项的归并·····	131
4.1.3	抽象义提取,操作命令形式化·····	132
4.1.4	同一组概念(集合)的抽象·····	134
4.1.5	构词·····	135
4.1.6	构句·····	135
4.1.7	语义理解·····	136
4.2	“木”的内涵特征与构词·····	137
4.2.1	词典中的解释·····	137
4.2.2	词典收词和释义存在的问题·····	138
4.2.3	“木”的构词分析·····	139
4.2.4	“木”的内涵特征构成与复合词语的语义 计算·····	140
4.2.5	验证·····	143
4.3	“刀”的类义抽象·····	144
4.3.1	词典中的对“刀”解释·····	144
4.3.2	一个事件的构成要素·····	144
4.3.3	“刀”的定义特征及抽象·····	144
4.3.4	“刀”内涵特征构词分析·····	145
4.3.5	验证——如何求 $a + \text{刀}$ 或 $\text{刀} + b$ 的组合义 ·····	148
4.4	“车”的类义抽象·····	151
4.4.1	辞典对“车”的解释·····	151
4.4.2	“车”的类义抽象·····	153
4.4.3	“车”的内涵特征构词分析·····	154
4.4.4	与“车1”搭配的动词·····	156
4.4.5	与“车1”、“车2”相关的名词·····	156
4.4.6	验证——与“车”有关的概念组合的语义	