

本书获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 “名垣名”逻辑语义研究

宋春阳摇著

学林出版社

目 录

序言	员
前言	员
第一章 绪论——现状与回顾	员
中文信息处理遇到的困难——问题与瓶颈	员
自然语言处理的过程	员
中文信息处理的困难	圆
当前中文信息处理的流派和策略	猿
信息处理工作原理及单纯使用统计方法的 摇不足	源
基于内涵语义规则信息处理的实用目标 摇	远
基于内涵逻辑模型论的语义分析的总体 设计思想	愿
面向信息处理的名词及“名 垣名”研究现状	员
汉语语法学界的研究	员
计算机界对名词的研究	圆
名词定义、性质、分类研究存在的问题	猿
“名 垣名”的研究亟待加强	愿
选题的目的和理论依据	圆

摇一摇研究目的	圆
摇二摇依据的主要语义理论	圆
摇三摇语义理论与“名垣名”研究	猿
摇一摇研究原则、方法和技术路线	猿
摇一摇原则	猿
摇二摇方法和技术路线——基于内涵逻辑的句法 分析	猿
摇三摇研究范围和步骤	源
第二章摇内涵逻辑理论——用于词语语义解释和理解	源
摇一摇引言	源
摇一摇本课题的创新性	源
摇二摇内涵逻辑解决语义理解的策略	源
摇二摇名词内涵——语义特征	源
摇一摇关于词语项、词义、概念、内涵、外延的 定义及关系	源
摇二摇概念外延指称属性的形式化描写	源
摇三摇概念之间的关系	缘
摇三摇内涵逻辑理论的运用	缘
摇一摇名词的定义	缘
摇二摇名词的内涵性质	缘
摇三摇名词与可能世界的联系——外延指称	缘
摇四摇专有名词内部的语义构成	缘
摇四摇内涵特征结构	缘
摇一摇关于语义特征的概念	缘
摇二摇“定义特征”和“情景区分特征”	缘
摇三摇“特征”概念提出的目的	缘

(一)摇概念的内涵特征——事物属性结构	缘
(二)摇名词定义特征的构成要素	远
(三)摇“定义特征”的确定方法	远
(四)摇特征层级性和方向性	缘
(五)摇概念的内涵特征组合规律	远
摇(一)摇内涵外延对组合义及句法结构的影响	苑
(一)摇位置对概念间关系的影响	苑
(二)摇两个名词的内涵外延组合方式	苑
(三)摇关于概念内涵义和外延义与词类的 对应问题	苑
(四)摇内涵义与外延义的判定	苑
(五)摇名词取内涵义的表现形式——内涵特征 算子	苑
(六)摇外延义的提取算子	愿
摇(二)摇定指与不定指	愿
(一)摇定指与不定指定义	愿
摇(三)摇要在语境中理解词义	愿
(一)摇静态的词典义	愿
(二)摇动态语境义	愿
第三章摇类义抽象理论	愿
摇(一)摇现代汉语词与短语的讨论	愿
(一)摇汉语基本语法、语义单位	愿
(二)摇词的概念及定义词的标准	愿
(三)摇基于信息理解的标准	愿
(四)摇词的组成单位	愿
(五)摇词类的划分	愿

摇猿圆叠置原理	怨
猿圆什么是叠置原理	怨
猿圆组合义的求解困难	员
猿圆叠置原理的操作策略和方法	员
摇猿猿抽象类义	员
猿猿抽象的理论依据	员
猿猿类义抽象的定义	员
猿猿类义抽象的目的	员
猿猿类义抽象的原则及方法	员
摇猿源由语义义求取组合义的理论探讨	员
猿源合成词产生的原因	员
猿源传统构词理论的不足	员
猿源关于依据语素理解词义的讨论	员
猿源向心与离心	员
猿源我们处理构词的方法	员
摇猿缘处理内涵特征构词要注意的几个问题	员
猿缘要区分频繁结构和非频繁结构	员
猿缘注意吸收认知语义学的研究成果	员
猿缘致力于构词的逻辑模型建立	员
 第四章摇基于实例的名 垣名语义结构研究	 员
摇源圆选词的范围、依据和对象	员
源圆选词的范围、原则	员
源圆抽象的依据辞书及语料	员
源圆考察对象	员
摇源猿“关”的类义抽象	员
源猿对相关辞典解释进行收集	员

灑灑圓摇对辞典的义项的归并	员员
灑灑陸摇抽象义提取 操作命令形式化	员圆
灑灑源摇同一组概念(集合)的抽象	员源
灑灑缘摇构词	员缘
灑灑近摇构句	员缘
灑灑近摇语义理解	员远
摇灑灑圆摇“木”的内涵特征与构词	员苑
灑灑员摇词典中的解释	员苑
灑灑圆摇词典收词和释义存在的问题	员愿
灑灑陸摇“木”的构词分析	员愿
灑灑源摇“木”的内涵特征构成与复合词语的语义 计算	员园
灑灑缘摇验证	员猿
摇灑灑陸摇“刀”的类义抽象	员源
灑灑员摇词典中的对“刀”解释	员源
灑灑圆摇一个事件的构成要素	员源
灑灑陸摇“刀”的定义特征及抽象	员源
灑灑源摇“刀”内涵特征构词分析	员缘
灑灑缘摇验证——如何求 葬刀或刀 垣豐的组合义	员愿
摇灑灑源摇“车”的类义抽象	员貌
灑灑员摇辞典对“车”的解释	员貌
灑灑圆摇“车”的类义抽象	员貌
灑灑陸摇“车”的内涵特征构词分析	员貌
灑灑源摇与“车员”搭配的动词	员远
灑灑缘摇与“车员”、“车圆”相关的名词	员远
灑灑近摇验证——与“车”有关的概念组合的语义	

计算	员愿
摇灞缘瑶“书”的类义抽象	员愿
灞缘员瑶“书”的词典义项	员愿
灞缘圆瑶义项归并及类义抽象	员愿
灞缘陵瑶“书”的语义求解	员园
灞缘源瑶语境中语义排歧	员源
摇灞陵瑶“手”的类义抽象及构词	员缘
灞缘员瑶词典对“手”解释	员缘
灞缘圆瑶“手”的类义抽象及定义特征	员缘
灞缘陵瑶验证——“手”内涵特征构词分析	员苑
第五章摇基于逻辑语义特征的名词谓语句分析	员猿
摇缘员瑶名词谓语句的分类	员猿
缘员员瑶名词谓语的句法成分的类型	员猿
缘员圆瑶谓语句名词的语义类型	员缘
摇缘圆瑶名词谓语句的性质	员缘
摇缘陵瑶名词谓语句中名名语义关系对其句法功能的影响	员愿
缘陵员瑶粤 \in 月	员愿
缘陵圆瑶粤 $<$ 月	员员
缘陵陵瑶粤越月	员园
缘陵源瑶粤 \neq 月	员猿
缘陵缘瑶结论	员源
第六章摇余论	员缘
摇缘员瑶语言的二义性	员缘
摇缘圆瑶语义分析的策略和目标	员愿

摇源原摇语义分析要简化,可排歧,可做形式化操作	员愿
摇源原摇关于语义工程的一点说明	员愿
参考文献	员员
摇附录一摇贺阳与《语法信息词典》形名兼类比照表	圆苑
摇附录二摇前 员愿个高频单音节名词排序表	圆愿
摇附录三摇名词概念分类表	圆员
后摇记	圆愿

序 言

摇摇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 research 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

圆年猿月圆日

前 言

摇摇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“名垣名”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“名垣名”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 羸孕语义结构分

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。

