

词汇语义量化研究

Исследования
семантической структуры слов

◆ 薛恩奎 / 著

黑龙江人民出版社

Исследования
семантической
структуры



词汇语义量化研究

Исследование
семантической структуры слов

李海明 著

商务印书馆出版

Исследование
семантической
структуры

教育部人文社会科学百所重点研究基地 学术丛书
黑龙江大学俄语语言文学研究中心

词汇语义量化研究

Исследования семантической структуры слов

薛恩奎 著

黑龙江人民出版社

图书在版编目(CIP)数据

词汇语义量化研究/薛恩奎著. —哈尔滨:黑龙江人民出版社, 2006. 11
ISBN 7-207-07223-6

I. 词... II. 薛... III. 语言学—研究 IV. H0

中国版本图书馆 CIP 数据核字(2006)第 148737 号

责任编辑: 张晔明
封面设计: 于克广

词汇语义量化研究

Cihui Yuyi Lianghua Yanjiu

薛恩奎 著

出版发行 黑龙江人民出版社
通讯地址 哈尔滨市南岗区宜庆小区 1 号楼 (150008)
网 址 www. longpress. com E-mail hljrmcbs@yeah. net
印 刷 黑龙江省地质测绘印制中心印刷厂
经 销 全国新华书店
开 本 880 × 1230 毫米 1/32
印 张 8. 5
字 数 200 000
版 次 2006 年 11 月第 1 版 2006 年 11 月第 1 次印刷
书 号 ISBN 7-207-07223-6/H · 251

定价: 28.00 元

(如发现本书有印制质量问题, 印刷厂负责调换)

目 录

绪论	(1)
第一章 语言中的时间系统	(6)
一、语言中的时间系统与时间定位	(8)
1.1 语言时间范畴	(8)
1.2 语言时间结构	(10)
1.3 时相、时序、时位	(14)
1.4 结束语	(23)
二、时点与时段的概念与功能辨析	(24)
2.1 引言	(24)
2.2 事件时间分布的词汇化现象	(25)
2.3 关于时间的两个基本概念	(28)
2.4 关于时间定位的相关参数	(34)
2.5 结束语	(38)
三、俄语动词时间范畴的语义描述	(39)
3.0 引言	(39)
3.1 动词语义结构呈现的时间性特征	(40)
3.2 动词形态变化体现的时间特征	(43)
3.3 事件时间与参照时间的相互关系	(44)
3.4 关于时间成分的分类及其语义功能	(47)
3.5 结束语	(50)
四、空间前置词语义情景结构	(51)
4.1 引言	(51)

4.2 语义关系的常量与变量	(51)
4.3 语义情景结构模式	(53)
4.4 标准理解模式	(59)
4.5 结束语	(63)
第二章 语义配价与句法题元	(64)
一、语义价与句法题元的对应关系	(66)
1.1 语言是形式与意义的对应关系	(66)
1.2 形式与意义的不对称现象	(68)
1.3 语义价是词义的基础	(71)
1.4 句法题元与语义价的不对称现象	(73)
1.5 词的配价与句的配价	(76)
1.6 结束语	(85)
二、配价、词义、句式	(87)
2.1 引言	(87)
2.2 配价性质和配价体系结构	(89)
2.3 语义价、句法价的相互关系	(95)
2.4 结束语	(97)
三、俄语关系动词的语义句法结构	(99)
3.1 引言	(99)
3.2 关系动词分类	(100)
3.3 关系动词的句法结构	(103)
3.4 结束语	(108)
四、俄语语序的核心结构效应	(110)
4.1 词序的语义基础	(110)
4.2 句子构成单位与语序变化的关系	(113)
4.3 核心结构对语序的基本限制	(117)
4.4 语句主要成分的分布	(120)

4.5 核心结构与实际切分	(123)
4.6 结束语	(125)
五、 H_0 的转折关系语义阐释	(126)
5.1 概述	(126)
5.2 客观事件的主观认识	(128)
5.3 $X, H_0 Y$ 转折句意思类型	(135)
第三章 词典释义与词义解析	(143)
一、电子化词典释义的语言学探索	(144)
1.0 概述	(144)
1.1 电子词典的类型	(145)
1.2 理论语言学发展对词典学的影响	(150)
1.3 词汇函数在词典学中的应用	(153)
1.4 结束语	(156)
二、“意思 $\Leftrightarrow$ 文本”语言学与词典学	(157)
2.1 “意思 $\Leftrightarrow$ 文本”语言学的背景	(157)
2.2 “意思 $\Leftrightarrow$ 文本”语言学的层次	(159)
2.3 语义元语言的结构	(162)
2.4 语义元语言的词典学价值	(168)
2.5 结束语	(176)
三、模式联想思维与词汇语义理据	(177)
3.1 理论概述	(177)
3.2 新词的语义理据	(180)
3.3 联想映射	(183)
3.4 结束语	(185)
四、动词义素聚合场与组合场	(187)
4.1 语义聚合场与组合场概述	(187)
4.2 语义聚合场类型	(189)

4.3 语义组合场模式	(192)
4.4 俄语动词概念意义类别及其形式组合模型	(194)
4.5 结束语	(197)
第四章 建立语言信息处理的描写方法	(198)
一、面向俄汉机器翻译的语义研究	(200)
1.1 关于计算语言学	(200)
1.2 词汇语义的量化分析	(203)
1.3 语言规则和关于语言规则的规则	(208)
1.4 结束语	(217)
二、基于“意思”的语义描写方法	(218)
2.1 方法选取以应用为先导	(218)
2.2 知识表示以应用操作为目的	(221)
2.3 语义结构是现实情景的映射	(228)
2.4 结束语	(231)
三、“ $M \Leftrightarrow T$ ”文法转换规则与机器翻译	(233)
3.1 “ $M \Leftrightarrow T$ ”工作原理	(233)
3.2 深层句法的构成	(237)
3.3 句法转换规则	(239)
3.4 面向俄汉机器翻译的汉语研究	(247)
3.5 结束语	(252)
主要参考书目	(253)

绪 论

自从有了人类,就有了语言。人类依靠语言进行思维,依靠语言认识世界,依靠语言表达情感,依靠交流思想,语言是人类的基本属性。世界上任何一个民族、任何一个部落无论是发达或落后都有自己的语言。但是,未必每一个民族的语言都有文字。从现代语言学的角度来看,语言是音、形、义的综合体。但是,早期的语言并没有文字系统。一个民族的语言文字系统是在长期的生产过程中逐渐建立起来的,并且与这个民族的科学技术水平密切相关。因此,应该从历时的观点和共时的观点两个方面看待语言和文字的关系。

1)语言的起源伴随着人类的起源,语言的进化与发展伴随着人类的进化与发展。正因为人类有了语言,借助语言这一工具表达思想、交流信息,人类社会才得以发展。物质是世界的前提,人类是语言的前提。因为有了物质,才有世界的存在;因为有了人类,才有语言的起源;因为有了能量,才有世界的变化;因为有了信息交流,世界才有意义。这是语言的功能,是语言对人类社会发展的作用。同样,人类社会的发展也促进了语言的发展,文字就是人类社会文化发展到一定阶段的产物。从语言的发展过程来看,可以分为有文字的语言和无文字的语言。无文字的语言是音、义结

合法,有文字的语言是音、形、义结合法。语音伴随着一定的意义才称其为语言,“形”是将包含一定意义的语音单位固定在一定介质上的符号,并通过一定的规则构成一定的系统才称其为语言。因此,语言和文字的关系是指有文字的语言和没有文字的语言,当今世界上还是有很多语言是没有文字的语言。从历时观点看,语言与文字有先后时间关系,文字是语言的形式结构系统。

2)任何语言都脱离不开声音,都必须具有语音系统,但不见得每一种语言都有文字。文字的起源要远晚于语音系统。在文字出现以前人类曾经历一个很长的摸索时期。一般认为,语言自起源至今大概经历三个发展阶段。第一个阶段是采用图形、符号表义时期,如我国古代有“上古结绳而治”之说。这种采用图画符号表义,可以说是利用实物帮助记忆,并不能表达抽象的概念,是语言发展的萌芽时期。这种简单的图形、符号还不能称其为语言;第二阶段是音、义语言时期。音、义语言能够表达各种复杂的概念,具备了语言的本质特征;第三阶段是音、形、义语言时期。语音发展到语音、语形,完善了语言的功能,克服了“音不能传于异地,留于异时”的弊端。文字是用以记录音、义语言的书面符号系统。文字在语言科学中的功能不仅克服了信息承传的时空限制,还表明了一个民族的科学技术水平和传播、继承一个民族的科学技术水平作用。

从本质上说,语言是传递、接收、存储信息的工具。早期的语言信息主要通过人类自身的器官实现传递、接收、存储的功能。这也是最早期的语言信息处理方式。文字的起源、生存和发展需要一定的物质基础。作为一种文字符号系统需要有一定的物质媒介作为存储的介质和书写的工具、材料。语言信息存储介质的出现

意味着语言信息的传播由人体器官向人体器官以外的物质媒体发展。我国早期存储文字的介质主要有龟甲、兽骨、金石、竹木简、布帛等,主要以“刀”为书写工具,以“刻”为书写方式。造纸术的发明标志着书写材料的一场革命,书写工具、书写方式也都发生了根本的变革。活字印刷术的发明标志着语言信息处理第一次革命,它为我国的文化、经济的发展开辟了广阔的道路,为推动世界文明的发展作出了重大贡献。自造纸术发明以后,语言信息的传播经历了人工手抄阶段、雕版印刷阶段和活字印刷阶段。计算机的诞生是20世纪科学技术的伟大成就之一。计算机激光照排技术是继活字印刷术之后,语言信息复制技术的又一次革命。人们充分利用电和光的作用,彻底改变了语言信息的存储介质、传播方式。什么“电子文件”、“电子图书馆”、“光盘版”、“电脑版”、“信息高速公路”等都是计算机技术在语言信息处理领域的具体运用。

自从计算机发明以后,人们对语言与文字的关系又有新的不同的认识。人们利用计算机技术彻底改变了语言信息的存储介质与传播方式。这种利用方式只是将语言的形式结构按原样复制、传播,并不是改变语言的“义”的处理方式。人们希望计算机能代替人类大脑,实现人类思维的电子化,试图利用计算机代替人来处理语言的“义”。因此,与计算机技术和语言学密切相关的交叉学科“计算语言学”诞生。计算语言学是一门研究如何运用计算机技术研究自然语言、处理自然语言的学问。从宏观角度看,利用计算机处理自然语言可分为两大领域:一是文字的输入、输出、复制、传输等,这一领域的研究技术已趋于成熟;二是词汇层次的定性、定量研究和句子层次的语法模型研究,以支持自然语言的计算机分

析与合成。这一领域的研究不仅涉及语言的“形”,还涉及语言的“音”和“义”。语言的形式结构和意义结构是计算语言学的两大前沿研究课题。

形式结构主要研究句子的结构模式、结构类型规则以及和意义结构之间的对应关系规则,属于语法学;意义结构主要研究词的概念分类,词义结构规则以及与形式结构之间的对应关系规则,属于语义学。词义描写是句法分析、合成,建立句子意义表达结构的基础。

本书是笔者近 10 年来在词汇语义方面的研究成果。通过这些年的学习和研究,对俄罗斯的词汇语义学、计算语言学的研究方法和所研究的内容有了一个大体了解,并结合国内在句法学、语义学领域的研究状况发表了一些文章。为保障内容的完整性,此次出版对有关内容作了适当的调整,对部分内容作了不同程度的修改和补充,全书分为四章。

第一章 语言中的时间系统

本章主要从自然时间、自我时间和事件时间三个层次论述语言中的各种时间范畴。

第二章 语义配价与句法题元

本章主要从词汇的语义结构和句子的句法结构两个方面论述语义配价与句法题元的对称与不对称关系,论述语义配价与句式的对应关系。

第三章 词典释义与词义解析

本章主要利用理论语言学的研究成果论述用于语言信息处理的电子词典释义方法和释义内容,讨论建构语言知识库的方法论体系。

第四章 建立语言信息处理的描写方法

本章主要从分析词的意义结构,句子的句法结构角度入手,讨论如何提取词的意义元素及其组合规则和句子结构元素及其组合规则的方法,以建立支持计算机运算的语义与句法对应规则体系和控制这些语言规则的运算规则体系。

第一章 语言中的时间系统

时间是一种无形的客观存在,时间是物质存在的基本形式。就自然时间而言,人类与自然界物质一样存在于这个时间之中。但是,人类是理性的生命体,他能够感知时间、认识时间、利用时间,把时间与自身的生存方式紧密地结合在一起。人类是通过语言文字系统来反映、表达时间观念的。所以,“时间”作为语言表述单位中的重要因素而成为语言学中的一个重要范畴和研究对象。但自然时间与语言中的“时”是两个不同的概念。自然时间是一维单向的运动流,无所谓“过去、现在、将来”的区分。但是,人类具有认识世界的能力,人类世界是一个超越自然世界的文化世界、精神世界。人类所生存的世界是一个自我的时空世界。人类以自我为中心,将自然时间轴切分为“过去、现在、将来”,从而构拟成人类自我的主观时间。主观时间通过语言系统中的形式单位表述,又构成了语言表述的“过去、现在、将来”的表述时间范畴。客观事件的发生、变化总是随自然时间而流逝,与自然时间不同的是,事件总是与人类相关的事件,事件的时间总是以人类自我为中心而与自然时间分离。自然时间、自我时间、事件时间三个层次的时间与“过去、现在、将来”构成九种时间范畴。这九种时间范畴形成人类通过语言表达自然、表述自我的语言时间系统。

“时间”在语言中通过一定的形式结构来表达,构成语言结构的“时”范畴,也称语法“时”范畴,“时间”是语法“时”范畴表达的内容。当描述一个事件的发生或事件过程的状况所涉及的时间性质是不同的。事件的发生、状态的变化与过程的持续涉及两个不同的时间概念。一种是与事件的发生、状态的变化在时间轴上的位置相关的时间,一般称“时间点”。“时间点”用来描述事件在某一时刻瞬间的状态,状态针对“时间点”而存在的。例如:速度、温度、价格的状况都是针对某一时间点而存在的。“时间点”描述事件在某一时刻瞬间状况的一种“存量”。另一种是与事件的过程持续相关的时间。一般称“时间段”。“时间段”描述事件从某一起点时刻到某一终点时刻积累效果的“流量”。例如,要描述一个国家的经济增长速度,描述一个家庭的用电量、用水量等,必须对应于一个具体的、有起、止时间点的“时间段”。一个有起点时间 t^1 和终点时间 t^2 构成的时间段 T 对应于无数个下位时间点和时间段:点的移动即构成线段。因此,时间点与时间段是两个相对的概念,但是两个不同的概念。它们之间是上下位关系,时间段以时间点为基础。从理论上讲,一个时间点可切分为无数个时间段。我们对时间的认识一是依据我们所描述的事件与时间的关系,即事件对时间要求的精度,二是依据人类对时间认识的技术水平,即科学技术对时间精度认识的能力。

一、语言中的时间系统与时间定位

1.1 语言时间范畴

时间是物质存在的基本形式。人类最初的时间观念来源于对客观世界物质显性的运动方式和人类自身的生理变化的感知,如日升日落、月圆月缺、寒暑变换、昼夜交替等等。这些自然现象的周期性运动形成了人们对时间量度的认识,根据不同自然现象变化的周期,确定出年、月、日、时、分、秒等时间度量单位。人类主体的存在形式是更为本质的生存时间,是一种本质上区别于自然时间的人类时间。人类以个体存在的形式形成自我中心说,“我、此时、此地”构成指示时间,把自我融入到自然时间之中,从而构拟了人类存在于之盍的主观时间系统,将一维的时间流区分为“过去、现在、将来”,构成三时主观时间系统。物质运动的持续性形成一定的“量”,在时间轴上占据一定的“时间段”,一般也称“时段”,时段具有对时间的计量功能。对时间的指示将时间区分为过去、现在、将来,确定了时间的先后顺序,从而形成“时间点”的概念,也称“时点”。时点具有时间定位功能。

从系统几何学角度看,“点”是一个初始的概念,“点”的延伸就形成一个线段。从A点延伸到B点形成线段C,C是A、B的上位量,A、B是C的下位量,“点”和线段是上下位“量”的关系。

时间点与时间段是两个相对的概念,“段”中有“点”,“点”中有

“段”,从“量”的角度看,1小时是由60个以“分”为单位的“点”构成,同时,1分钟又是以60个以“秒”为单位的点构成。“分”对“时”而言是“点”,对“秒”而言是“段”。从功能角度看,“时段”与“时点”是两个具有不同功能的单位。时点除了“量”的范畴,还具有“位”的范畴,具有定位的功能,这也反映了人类对时间的不同认知方式。正因为有了“点”的定位功能,才有可能在时间轴上划分出长短不等的“时间段”,才可能按主观方式确定事件在时间轴上的先后顺序。当“时间点”表示定位功能时,“时间量”的范畴意义已退居背景,进入前景的只是作为定位功能的“点”;过去、现在、将来;昨天、今天、明天。

客观世界的各种事物、现象都存在于一定的时间和空间之中,我们把客观事物的变化,各种现象的产生与消亡等统称为“事件”。任何事件都有产生、持续、结束的阶段性的,体现了时点与时段的范畴意义,任何事件都与其它事件相关,体现了事件之间的顺序关系。人类认识世界,同时也要通过语言形式表述所认识的世界。当人类使用语言表述事件的顺序、持续等时间范畴时,无意识地将自身“说话人或观察者”作为一个参数融入事件过程,使事件的一维时间流转化为三时的时间位。语言中的时间系统包含两个方面的内容,一方面是事件本身所具有的时间属性特征在语言中的反映,另一方面是表述事件时间特征的语言形式。事件的时间与语言形式表述的时间是两个不同的概念,属于两个不同的序列。事件的时间与客观的自然时间是一致的单向时间流,发生在“先”的永远在“先”,发生在“后”的永远在“后”,只表现为由“先”到“后”的顺序关系。语言形式表述事件的时间不仅涉及事件之间的先后顺序关系,还涉及事件