

面向中文信息处理的现代汉语动宾语义搭配研究

MIANXIANG ZHONGWEN XINXI CHULI DE XIANDAI HANYU
DONGBIN YUYI DAPEI YANJIU

周卫华 著

『求索』语言文学学术文库

CENTRAL CHINA NORMAL UNIVERSITY PRESS



华中师范大学出版社

面向中文信息处理的现代汉语动宾语义搭配研究

MIANXIANG ZHONGWEN XINXI CHULI DE XIANDAI HANYU
DONGBIN YUYI DAPEI YANJIU

周卫华 著



『求索』语言文学学术文库



华中师范大学出版社

新出图证(鄂)字 10 号

图书在版编目(CIP)数据

面向中文信息处理的现代汉语动宾语义搭配研究/周卫华著.
—武汉:华中师范大学出版社,2009.4
(“求索”语言文学学术文库)

ISBN 978-7-5622-3886-7

I. 面… II. 周… III. 汉语—动宾—研究—现代 IV. H146.3

中国版本图书馆 CIP 数据核字(2009)第 025472 号

面向中文信息处理的现代汉语动宾语义搭配研究

© 周卫华 著

责任编辑:肖鹏飞

责任校对:罗 艺

封面设计:罗明波

编辑室:第五编辑室

电话:027-67867364

出版发行:华中师范大学出版社

社址:湖北省武汉市珞喻路 152 号

邮编:430079

电话:027-67867076(发行部) 027-67861321(邮购)

传真:027-67863291

网址:<http://www.ccupress.com>

电子信箱:hscbs@public.wh.hb.cn

经销:新华书店湖北发行所

印刷:武汉理工大印刷厂

督印:章光琼

字数:150 千字

开本:880mm×1230mm 1/32

印张:6.5

版次:2009 年 4 月第 1 版

印次:2009 年 4 月第 1 次印刷

印数:1—1000

定价:20.00 元

欢迎上网查询、购书

敬告读者:欢迎举报盗版,请打举报电话 027-67861321

三峡大学文学院
“求索”语言文学学术文库
编委会

主任：刘德富

副主任：李建林

编委：（以姓氏笔画为序）

王作新	王前程	邓新华	田斌
刘月新	吴卫华	陈和春	李建国
张朔	范军	赵宏	胡绍华
段维	曾巍		

总 序

王先霭

三峡大学文学院组织出版“‘求索’语言文学学术文库”，整体地展示本院教师最新的重要学术成果，第一批书稿已经交付出版社，即将面世。这是他们学科建设的重大举措。学院领导向我介绍了他们的构想和当前的进展，嘱咐我就此写一点文字。我和三峡大学文学院及其前身宜昌师专中文系的多位同行交往几十年，对学院的发展轨迹有所了解，可以借此谈谈我对这个学术团队的观察和由此而生的感想。

从宜昌师专所在的北山坡，到今天三峡大学的葛洲坝与三峡大坝近边依山傍水的主校区，地理距离并不远，而几十年来几代人所经历的创业历程，却是并不平坦单纯，而有着十分丰富厚重的内涵。在学科建设中，传统是一种无形的资源，无形的力量，文科尤其如此。没有相当时间的积累，一个学术团队不可能于朝夕间突兀而起，更难有持续发展的后劲。二十世纪五六十年代到八十年代初的宜昌师专中文系的老师们，是一批非常敬业的园丁，倾注心力于教学，同时也在各自选定的领域学有专攻，功力扎实，学风朴实，撰写过不少很有见地的论著，带出了一批批后继学者。本丛书的作者们，出生于上世纪五十年代至七十年代，而今正是治学的黄金年龄。他们曾分别到国内多所大学攻读学位、进修和访学，既仰承本校先行者们的纯朴作风，又广泛吸纳校外、省外和国外多种学术营养，最近十年左右，陆续地开始崭露头角。他们大都有稳定的学术方向，从我十几年来的接触，从这些作者的著作目录，不难看出他们在求索之途跋涉的清晰脚印，而本文库里的专著则多是作者在稳定方向上的新成果。我比较赞赏稳步渐进的治学方式，我以为，

《荀子·劝学》说的“螭蛇无足而飞，鼫鼠五技而穷”，在今天依然值得我们深思。学无定向，常常导致学无定见。趋时跟风，可以煊赫于一时，最终则消歇于长久。治学首先是个体独立的精神劳作，求索者要准备着经受孤独。但这绝不是主张单兵独人各自为战，恰恰相反，在一个院系、一个专业里，大家都应有整体意识，有学术上的交流、交融和碰撞。然而，也只有各人有专攻、有特长才能有效地交流、交融和碰撞，交流、交融和碰撞必须建立在大家各自独立思考和张扬学术个性的基础之上。本文库的作者们，分属若干二级学科，就我的了解，其中，文艺学和中国现当代文学力量较强，而语言类和文学类彼此呼应，有竞争更有协作。我觉得，中国文学一级学科的建设，文学专业和语言专业的配合相当重要。在自然界，物种多样、结构多层的环境有利于动植物的生存。在一个院系，所属二级学科既要有重点，又要有相对的和动态的均衡发展，这才有利于各学科的健康发展，有助于各个学科和各个成员适应当代社会的急速变化而不断调整自身的研究方向和研究思路。三峡大学文学院的进步，与他们学术生态的合理是大有关系的，这套文库将有利于他们学术生态的进一步合理化。

现在看到的收入文库里的第一批著作，其学术含量彼此还是有些区别的。这也使我们可以期望，后续的作品给我们更多的惊喜。文库以“求索”为名，也表达了这样的意思：学术之路是无尽的，学者生命的价值在无休止的探求之中，最灿烂的明珠永远在我们的前面闪耀。

2007年3月26日

于武昌桂子山

前 言

从 20 世纪 80 年代开始,中文信息处理经历了一个从“字处理”到“词处理”,再到当前“句处理”的发展阶段。对自然语言进行处理,仅仅进行词法和句法分析,尚不能满足机器翻译、自然语言理解、情报自动检索等自然语言处理系统的要求,还要进行语义分析。在自然语言处理中,语义研究具有非常重要的意义。汉语中,词和词之间没有明确的边界,词或词组的类别与它们的句法功能没有一种明确的一一对应关系,主语、宾语等句法成分与它们在句子中担任的语义角色也没有明确的对应关系,这些因素导致了汉语的语义结构和句法结构、语义成分和句法成分之间的关系极为复杂。在现阶段,对汉语的句子做一个完全的语义分析还面临着很大的困难。

本书对动宾语义搭配展开研究,源于近年来自然语言处理领域出现的一个新的语言处理策略——浅层句法分析^①。浅层句法分析(shallow parsing)也叫部分句法分析(partial parsing)或语块分析(chunk parsing),它是跟完全句法分析相对的。浅层句法分析不要求得到完全的句法分析树,只要求识别句子中某些相对简单,但又在句子组成中意义重要的成分。

动宾搭配是一个句子的核心成分,处于优势地位,它实际上映射了整个句子的轮廓。这是因为动词本身就带有丰富的句法和语义信息,它不但决定了一个句子可能的句法结构和语义框架,也决定了它对那些与之共现的名词性成分有语义选择限制。从语义上来看,动词不但

^① 有关浅层句法分析的详细内容,可以参考:孙宏林、俞士汶:《浅层句法分析方法概述》,载《当代语言学》2000年第2期,第74~83页。

对宾语的语义角色有一定的规定性,还对承担语义角色的名词性成分在语义类别上有选择限制。比如,动词“吃”表示“把食物等放到嘴里经过咀嚼咽下去(包括吸、喝)”这一意义时,动词的意义规定其宾语的语义角色是受事,充当受事的词语在语义上必须具有[可食性]的语义特征。所以,本书从两个方面研究动词和宾语之间的语义搭配关系,一是动词和宾语之间的语义角色关系,二是动词对充当宾语的名词在语义类别上的选择限制关系。

本书以“小句中枢说”与“核心语法和边缘语法”作为理论背景。“小句中枢说”是邢福义先生(1995)提出的关于汉语语法学的新的表述理论。该理论的基本内容包括五个方面:其一,认定小句在汉语各类各级语法实体中占据中枢地位。其二,认定小句有“成活律”、“包容律”和“联结律”。了解小句三律,可以加深对汉语语法系统中小句的中枢地位的认识。其三,认定汉语语法重视“句管控”,即小句在中枢地位上对汉语语法规则的方方面面发挥其管束控制的作用。其四,认定汉语句法结构具有兼容性和趋简性。结构形式的趋简,导致结构语义的兼容;语义兼容的可能性,又提供结构趋简的可能性。其五,认定要在研究工作中贯彻和实践“小句中枢说”,主要的思路和方法是“两个三角”,即“表一里一值”小三角和“普一方一古”大三角^①。

“小句中枢说”以小句为基点研究汉语语法事实的语法学价值,不仅在于它是一套适合于汉语语法研究的语法描写理论,更重要的还在于它在语法实体及其相互关系的本质认识方面具有重要的指导意义。“小句中枢说”,特别是其“句管控”的思想,在中文信息处理中也具有其独到的优势。比如,温锁林(2004)将以小句为本位得出的规则用于汉语词性标注,在开放测试中达到了95%的准确率。同样,“句管控”思想在本书的研究中也具有很重要的指导意义,动宾搭配中动词词义的选择和动宾语义关系的确定,都要以小句为观察站。比如,动宾搭配

^① 邢福义:《研究观测点的一种选择——写在“小句中枢”问题讨论之前》,载《汉语学报》2004年第1期。

“挂了电话”中,“挂”的意义可以是“挂(3):把话筒放回电话机上使电路断开”,也可以是“挂(4):指打电话”,只有在这个搭配进入“他已经给小王挂了电话”和“他还没说完就挂了电话”这两个句子中,接受了“句管控”之后,才能确定动词的意义和动宾之间的语义关系。

“核心语法和边缘语法”理论是吴振国先生在《现代汉语的边缘语法及对外汉语教学问题》一文中提出的^①。他认为,一种语言系统的语法规则可以分为两部分:核心部分和边缘部分。吴振国先生把核心部分的语法规则称为核心语法,边缘部分的语法规则称为边缘语法。核心部分的语法规则具有普遍性、一般性和自由性;边缘部分的语法规则具有特殊性、熟语性和固定性。普遍性是从不同语言的异同来看,核心语法具有人类语言的普遍性,或者至少是大多数人类语言都具有的语法规则,属于核心语法的格式,翻译成别的语言时,一般可以直接对译;特殊性则是某种或某些较少的语言(或方言)所特有的语法规则,属于边缘语法的格式,一般不具备对译性。一般性是就一种语言系统内部来说,核心语法可以按照一般语法规则进行自由类推;熟语性是指边缘语法一般不能自由类推。自由性是指核心语法格式是自由的,可以自由扩展;固定性是指边缘语法一般不能自由扩展。

在语法研究中,常常有学者感叹:任何语法规则都有例外。在中文信息处理中也同样如此,我们常常对一些不规则的“例外”现象感到束手无策。“核心语法和边缘语法”理论为我们在中文信息处理中处理“规则中总有例外”的语言现象时,提供了一个新的思路。我们可以将自然语言中的核心语法和边缘语法区别对待,对二者采取不同的处理策略:为能够自由类推的核心语法现象建立规则,将不能自由类推的边缘语法现象直接放入词库。本书就是运用“核心语法和边缘语法”理论,区分动宾语义搭配中的核心语法现象和边缘语法现象,为属于核心的动宾搭配建立规则,将边缘的动宾搭配放入词库,以提高自然语言处

^① 吴振国:《现代汉语的边缘语法及对外汉语教学问题》,2002年12月在“对外汉语—商务汉语教学与研究专家讨论会”(上海财经大学)上宣读。

理的效率。

就研究范围而言,本书的研究并不是面向所有的动宾搭配。动词主要指单音节动词,宾语只指名词作宾语的情况。本书以单音节动词作为研究对象,主要是基于以下几个方面的因素:

一、单音节动词和双音节动词在不同文本中的使用频率存在很大的差异。一般来讲,单音节动词在文学作品中的使用频率要高于在其他文本中的使用频率。根据我们的不完全统计,单音节动词在文学作品中的使用频率大约是其他文体(新闻、科技、政论等)的3~5倍。

二、单音节动词中表示动作行为的动词较多,词的意义比较具体,充当其宾语的名词在意义上一般也比较具体,抽象名词较少。单音节动词和其名词宾语在词义上比较具体的特点,有利于归纳动宾语搭配规律,也有利于对充当宾语的名词进行语义分类。

三、单音节动词的义项较多,在中文信息处理中,对含有多个义项的单音节动词进行词义标注和排歧就显得尤为重要。在本书考察的500个单音节动词中,具有多个义项的词有348个。而且,使用频率越高,动词的义项往往越多。在本书的统计中,使用频率处在前100位的单音节动词,有86个动词具有多个义项,共有436个义项。

本书从一个大约一千万字的语料库中,提取了使用频率较高的500个单音节动词(见附录一),然后参考《现代汉语词典》^①的释义,并依据一定的原则,确定500个单音节动词的1337个义项作为本书的研究对象。本书所考察的2339个离合词是以语料库中提取的500个单音节动词为参照,从《现代汉语词典》中收集而来的。对离合词的研究,主要是考察其扩展形式,提出在中文信息处理中离合词的处理策略。

本书的研究主要采用了以下几种资源:

一、语料的运用。本书在考察单音节动词和双音节动词在不同文本中的使用频率时,采用了1998年1月的《人民日报》语料。该语料由北京大学计算语言研究所研制开发,经过了词语切分和词性标注。本

① 为《现代汉语词典》第5版,下文中如无说明则亦为第5版,不再注出。

书语料库中的文学作品和其他类型文本,全部由华中师范大学吴振国先生提供。本书采用中科院计算语言所张华平、刘群两位先生开发的FreeICTCLAS自动分词软件,对语料库进行了自动分词和词性标注的处理。该软件采用的词性标注符号说明,见附录二。

二、词典的运用。本书对动宾语义关系的研究,主要参考了孟琮等(1999)编写的《汉语动词用法词典》。本书500个单音节动词的1337个义项,是根据《现代汉语词典》的解释确定的;2339个离合词也是从《现代汉语词典》中选取的。本书还参考了林杏光等(1994)编写的《现代汉语动词大词典》。

目 录

前言	(1)
第一章 核心语法和边缘语法	(1)
第一节 理论的提出	(2)
第二节 区分核心语法和边缘语法的依据	(3)
一、从普遍语法和个别语法看核心语法和边缘语法的区分	(3)
二、从语言习得看核心语法和边缘语法的区分	(5)
第三节 核心语法和边缘语法的区分标准	(8)
一、核心语法的普遍性和边缘语法的特殊性	(9)
二、核心语法的一般性和边缘语法的熟语性	(11)
三、核心语法的自由性和边缘语法的固定性	(12)
四、核心语法的语义组合性和边缘语法的语义凝固性	(14)
五、标准的运用	(14)
第四节 核心语法和边缘语法的相对性	(15)
一、核心语法和边缘语法是一个连续统	(15)
二、核心语法和边缘语法是一种相对性的划分	(16)
第五节 核心语法和边缘语法与中文信息处理	(17)
一、在中文信息处理中区分核心语法和边缘语法的必要性	(17)
二、核心语法和边缘语法在中文信息处理中的应用	(19)
第二章 现代汉语动词与宾语的语义关系	(21)
第一节 现代汉语动宾语义关系研究概述	(21)
一、现代汉语动宾语义关系研究成果介绍	(21)

二、西方语法理论与现代汉语动宾语义关系研究·····	(28)
三、现代汉语动宾语义关系研究评介·····	(30)
第二节 现代汉语动宾语义关系的性质和分类原则·····	(34)
一、动宾语义关系的性质·····	(34)
二、划分动宾语义关系的原则·····	(39)
第三节 现代汉语动宾语义关系的分类及相对性·····	(42)
一、动宾语义关系的分类·····	(42)
二、动宾语义关系的相对性·····	(45)
第四节 从信息处理的角度看动宾语义关系的分类·····	(46)
一、动词的意义对宾语语义角色的规定性·····	(46)
二、自由类推宾语和不可自由类推宾语·····	(50)
第三章 单音节动词与各类宾语的搭配考察·····	(54)
第一节 语料的准备·····	(54)
一、单音节动词的选取与处理·····	(54)
二、动宾搭配的提取·····	(59)
三、动宾语义搭配框架的建立·····	(62)
第二节 十三类动宾语义搭配框架的考察·····	(63)
一、动词+[受事]·····	(63)
二、动词+[结果]·····	(74)
三、动词+[对象]·····	(82)
四、动词+[致使]·····	(89)
五、动词+[施事]·····	(94)
六、动词+[处所]·····	(102)
七、动词+[材料]·····	(110)
八、动词+[工具]·····	(116)
九、动词+[时间]·····	(120)
十、动词+[关系]·····	(123)
十一、动词+[目的]、动词+[原因]、动词+[方式]·····	(126)
第三节 小结·····	(126)

第四章 单音节动词对宾语的语义选择限制	(129)
第一节 动宾搭配与语义选择限制	(129)
一、选择限制	(129)
二、动词和宾语之间存在语义选择限制	(131)
三、在动宾搭配框架中考察动宾语义选择限制的必要性	(132)
第二节 现代汉语名词的语义分类	(136)
一、现有名词语义分类系统	(137)
二、本书对现代汉语名词的语义分类	(141)
第三节 动词对宾语的选择限制与中文信息处理	(146)
一、语言知识库的建立	(146)
二、动词对宾语的选择限制在中文信息处理中的应用	(147)
第五章 离合词的研究与处理策略	(150)
第一节 离合词的界定与结构类型	(150)
一、离合词的界定	(150)
二、离合词的结构类型	(155)
第二节 离合词的扩展形式及其特点	(158)
一、离合词的扩展形式	(158)
二、离合词扩展形式的特点	(163)
第三节 离合词的处理策略	(166)
一、建立离合词词库	(166)
二、离合词的切分与标注	(169)
第六章 实验与分析	(171)
第一节 动宾搭配的认可规则	(171)
一、语料的选取及预处理	(171)
二、动宾搭配认可规则的描述	(171)
第二节 实验与分析	(173)
一、实验一:简单句的动宾搭配认可	(173)
二、实验二:动词“打”构成的动宾搭配认可	(174)

三、实验结果分析	(175)
附录一:500 单音节动词词表	(179)
附录二:词性标注符号集	(181)
参考文献	(182)
后 记	(191)

第一章 核心语法和边缘语法

世界上任何一种自然语言,都包含着纷繁复杂的语言现象。我们以人类自然语言为对象进行语言研究的时候,也就很难建立一个没有“例外”现象的语言规则系统。语言研究的目的,不仅是为了揭示语言现象的本来面貌,更是为语言在各个领域的应用提供基础,比如对外语言教学和自然语言处理等。如果我们能够透过繁芜的语言现象,在一种语言系统的语法规则中,把位于核心位置的语言现象和处于边缘的语言现象区分开,为核心语言运算系统建立规则,把处于边缘的语言现象放入词库,就有利于简化语言系统中的语言规则,为各个应用领域的语言研究提供便捷。那么,在人类自然语言或某一种自然语言的规则系统中,是否存在着核心语法和边缘语法的区别?如果存在,应该用什么原则和标准加以区分?这是我们首先要回答的问题。

“核心语法和边缘语法”理论是本书研究动宾语义搭配的理论基础。本章主要包括五个部分:(一)理论的提出。本节对核心语法和边缘语法理论作了简要介绍;(二)区分核心语法和边缘语法的依据。本节从普遍语法和个别语法、语言习得两个方面论述了自然语言中存在核心语法现象和边缘语法现象;(三)核心语法和边缘语法的区分标准。本节讨论了区分核心语法和边缘语法的四条标准以及标准的运用;(四)核心语法和边缘语法的相对性。本节主要说明核心语法和边缘语法是一个连续统,二者的划分具有相对性;(五)核心语法和边缘语法与中文信息处理。本节讨论了在中文信息处理中区分核心语法和边缘语法的必要性,以及该理论在中文信息处理中的应用。

第一节 理论的提出

核心语法和边缘语法的观点是吴振国先生在《现代汉语的边缘语法及对外汉语教学问题》一文中提出的。他认为,一种语言系统的语法规则可以分为两部分:核心部分和边缘部分。吴振国先生把核心部分的语法规则称为核心语法,边缘部分的语法规则称为边缘语法。核心部分的语法规则具有普遍性、一般性和自由性;边缘部分的语法规则具有特殊性、熟语性和固定性。普遍性指的是,从不同语言的异同来看,核心语法具有人类语言的普遍性,或者至少是大多数人类语言都具有的语法规则,因此,属于核心语法的格式,翻译成别的语言时,一般可以直接对译;特殊性则是某种或某些较少的语言(或方言)所特有的语法规则,属于边缘语法的格式,一般不具备对译性。一般性是指在一种语言系统内部,核心语法可以按照一般语法规则进行自由类推;熟语性是指边缘语法一般不能自由类推。自由性是指核心语法格式是自由的,可以自由扩展;固定性是指边缘语法一般不能自由扩展。该文主要探讨在对外汉语教学中,应该区分核心语法和边缘语法,针对核心语法现象和边缘语法现象要分别采取不同的教学措施,以提高对外汉语教学的效率。

我们认为,把汉语中的语法规则分为核心语法和边缘语法两个部分,不仅适用于对外汉语教学,也适用于中文信息处理。计算机在处理自然语言时,和我们人类学习语言的情况是一致的。一个人在习得某种语言的时候,必须知道某种语言的各种语法现象的规则和条件。同样,“不论哪一种自然语言处理的语法,其最根本、最关键的问题,是要指出各种语言形式出现和变换的条件。只有指出了条件,计算机才有可能根据有关的条件,执行相应的动作,从而使整个系统成为一个可以动态地执行的过程”^①。计算机在执行有关的程序时,不但要给它指出条件,而且要明确是什么条件,它才能有效地执行相应的动作。但事实

① 冯志伟:《自然语言的计算机处理》,上海外语教育出版社,1996年,第10页。