



司显柱 著

现代汉语转类词研究

——语料库视角

现代汉语转类词研究

——语料库视角

司显柱 著

復旦大學出版社

目 录

第一章 绪论	1
1.1 转类概念内涵: 活用与兼类	1
1.2 转类概念外延: 名词、动词、形容词	2
1.3 汉语转类词研究: 成绩与问题	3
1.4 批评与展望	5
1.5 研究目标与方法	5
1.6 研究意义	7
本章参考文献	8
第二章 关于《现代汉语词典》和“北京大学现代汉语语料库”	10
2.1 关于《现代汉语词典》	10
2.2 关于“北京大学现代汉语语料库”	11
第三章 现代汉语词类划分标准和汉语词语句法槽	13
3.1 现代汉语词类标准和转类词语判断标准	13
3.2 现代汉语词语句法槽	15
3.3 如何凭借句法槽类型, 快速查询转类案例	16
本章参考文献	18
第四章 现代汉语转类词语库建构	20
4.1 确立检索项	20
4.2 语料检索	21
4.3 检索语料解释	25
4.4 语料甄别	26

第五章 现代汉语名词转类动词研究	29
5.1 汉语名词转类动词研究综述	29
5.2 现代汉语名词、动词句法槽	31
5.3 现代汉语名转动词库建构	41
5.4 汉语名词转类动词描写: 分布与频次	66
5.5 概念合成理论与名转动词	74
小结	76
本章参考文献	77
第六章 现代汉语动词转类名词研究	79
6.1 汉语动词转类名词研究综述	79
6.2 现代汉语动转名词库建构	80
6.3 汉语动词转类名词描写: 分布与频次	205
6.4 语言类型学和系统功能语言学视角的动转 名词现象探索	211
小结	220
本章参考文献	221
第七章 现代汉语名词转类形容词研究	223
7.1 汉语名词转类形容词研究综述	223
7.2 现代汉语形容词语法特征和句法槽	227
7.3 现代汉语名转形词库建构	232
7.4 汉语名词转类形容词描写: 分布与频次	245
7.5 语法隐喻与名转形容词	249
小结	252
本章参考文献	253
第八章 现代汉语形容词转类名词研究	254
8.1 基于北大语料库的现代汉语形转名词库建构	254
8.2 汉语形容词转类名词描写: 分布与频次	398
8.3 认知语法和形转名词	406
小结	409
本章参考文献	410

第九章 现代汉语动词转类形容词研究	411
9.1 现代汉语动词转类形容词句法槽	411
9.2 现代汉语动词转类词库建构	413
9.3 汉语动词转类形容词描写: 分布与频次	448
9.4 语法隐喻和动词转类形容词	454
小结	455
本章参考文献	456
第十章 现代汉语形容词转类动词研究	457
10.1 现代汉语形容词转类词库建构	458
10.2 汉语形容词转类动词描写: 分布与频次	473
10.3 前景化和形容词转类动词	476
10.4 前景化下的形容词活用作动词在不同体裁 分布差异原因初探	479
10.5 前景化语言形转动词的翻译	482
小结	484
本章参考文献	484
后记	486

第一章 绪 论

语言不是处于一个一成不变的封闭系统中,而是处于一个持续变化的动态连续系统中。无论英语还是汉语,一个词并非总是囿于固有的词类,而是会发生一定的变化。转类词是指不经任何形式的变化,从一种词类转用为另一种词类的词。这是语言系统中一种语言偏离、变异的过程。转类现象也成为国内外的语言学学者研究的热门课题,并从语义、语用、修辞、文体学、功能语言学、认知语言学等多个角度对转类词进行研究。就现代汉语而言,词的功能及用法是相对固定的,但由于功能的游移,一个词会在词形不变的情况下直接被转用作另一词类,即转类,由于这一特征,又称为“零形派生”、“功能转换”。

1.1 转类概念内涵:活用与兼类

这里需要指出的是,根据谭景春对于词类活用与兼类甄别的论述,词语转类一般经历了“偶用→常用→固定”之发展路径,从历时的角度看是一连续统,处于这一连续统两端的典型情况是“典型活用”(词类正处于转变之中)和“兼类”(词类转变的完成,标志为已经收录于词典)。“我们把词类转变看作是一个过程,包括兼类和活用,兼类是词类转变的完成,活用是词类正处在转变之中,或仅仅是临时发生了转变。典型的兼类和活用虽然有本质的区别,但从活用到兼类是逐渐过渡的,在过渡中二者的界限不是很分明。活用的时间久了,使用的频率高了,就会被看作是兼类。兼类和活用都是词类转变,只是处于不同的阶段。”(谭景春,1998:368)

本书所讨论的转类词,内涵是指前述“词类正处于转变之中”的“偶用、常用”情形。如“八卦”一词,在《现代汉语词典》(2005版)标注的是名词,意思是“我国古代的一套有象征意义的符号。用‘—’

代表阳,用‘—’代表阴,用三个这样的符号,组成八种形式,叫做‘八卦’。”但是,在下面这个句子中,“八卦”是很是非、很桃色、很鬼怪的意思。“凡从事八卦算命的都会惹上是非、犯上桃花,激怒神鬼的!特别那种以大师自居号称能为人改命改运,消灾解难的,都很八卦!无非是骗财骗色骗名骗利,却装成一副善人的面孔,很八卦!”这里黑体“八卦”明显转类用作了形容词。再如“他很鲁迅,真是运气了你!幸福着你的幸福!”分别是名词活用为形容词,名词活用为动词和形容词临时转用为动词的情形。也就是胡裕树(1981:332)所说的“甲类词在特定的条件下,为了修辞上的需要,偶尔用作乙类词,这是活用”。

自然,对于已经完成词性转变的“兼类”,并不在我们的研究范围之内。如“编辑”、“翻译”既做名词,又做动词:他是个编辑/翻译//他编辑/翻译了一本书。

1.2 转类概念外延:名词、动词、形容词

以上,从词语转类连续统角度,通过对“活用”与“兼类”的区隔,我们对转类词语的研究限定在“活用”的范畴,排除“兼类”情形。但鉴于以下等因素,我们还得从概念外延的角度,对限定为“活用”情形的词语转类予以进一步限定,即我们只将现代汉语中的名词、动词和形容词之间时有发生但尚未被辞书认定和标注的转类词语确定为研究的范围。

之所以如此,这主要是因为,一方面,词语转类多发生在实词范畴之间,并主要集中在名词、动词、形容词之间(陈望道,1979:190—192;郭锡良,1983:125—126);另一方面,名词、动词和形容词是构成实词的主体。据尹庸《汉语词类的定量研究》(1986),在40 000多个现代汉语通用词中,三类实词占92%。据北京语言学院语言教学研究所《现代汉语频率词典》(1996)的统计,在使用频率最高的100、300与2 000词级中,三类实词所占的比重分别为38%、59%、78%(转引自张勇,2009:9)。

根据词类连续统的观点,词类转用的语言基础是词类的功能游移,且这种游移具有双向性特点,即“体词的谓词化方向”和“谓词的

体词化方向”。因而,在我们设定的“名词—动词—形容词”这一连续统中,词类功能游移既存在“名词→动词→形容词”的右向性运动,也存在“名词←动词←形容词”的左向性运动。这样,三类实词间的双向运动共可析出六组转类形态:名词→动词、名词→形容词,动词→名词、动词→形容词,形容词→名词、形容词→动词。

1.3 汉语转类词研究:成绩与问题

作为一种较为常见的语法和修辞现象,对转类词的观察、探索由来已久,概要如下:

自马建中首开先河,在《马氏文通》中谈及词的随文异用现象,并将之称为“字类假借”以来,已有 80 多年的历史。虽然汉语界对该类现象的探讨更多地针对古汉语,但对现代汉语的研究近年也日渐繁荣,据不完全统计,研究文献数有百计。经初步梳理,研究主要在以下方面展开:(1) 转类现象的定义与称谓。在这方面,早期的文献一般将之称为“词类活用”,如陈承泽在《国文法草创》(1982: 18)中指出“活用”是与“本用”相对而言的,从基本的语法性质出发的用法叫“本用”,不同于基本性质的用法称“活用”。吕叔湘在《中国文法要略》(1982: 24)里指出:一个词只应有一种用法(本用),若有其他用法则为“活用”。黎锦熙在《新著国文语法》(1951: 7)里谈到“活用”问题时认为“活用的时候,词类虽改,形体依旧。”王力在《汉语史稿》(1980: 373)中将词语转类解释为“词在句中的临时职务”或“修辞上的一种手段”。胡裕树(1981: 332)认为“甲类词在特定的条件下,为了修辞上的需要,偶尔用作乙类词,这是活用”。(2) 转类的范围。虽然绝大多数词类相互之间都可以转换,但从量或者频率的角度,主要集中在名词、动词、形容词之间。郭锡良等(1983: 125—126)认为“词类活用主要是有关名词、动词、形容词的问题……三类词的基本功能是比较固定的,但是某些词可以按照一定的语言习惯灵活运用,在句中临时改变它的语法功能,充当别的词类。”程望道(1979: 190—192)也说,在汉语里,词性之间的相互转换“多不过是名词和动词的转类”。不过,郑丽(同上: 5)认为“现代汉语中名词活用为形容词在活用现象中最普遍,占据最大比例”。(3) 关于转类词的定性与

价值问题。对于汉语中转类词的定性,存在两种截然不同的观点。《汉语语法修辞词典》(1988: 570) 将转类的定义为: 临时转用, 并不改变原来的词性。王力(1984: 21) 也持相同的看法“我们虽然承认汉语的词类有兼类现象, 但是我们不承认有转类现象, 承认有兼类现象, 是承认某些具体的词具有两个词类的语法特点; 不承认有转类现象, 是不承认‘词无定类’的说法。”许余龙(1997: 130) 也持类似的见解, 认为汉语是一种孤立型语言, 词的语法功能并不依赖于词尾的变化形式, 因而汉语中的许多词是兼类的, 谈不上转化不转化。另外一种以汪榕培等(1998: 55) 为代表, 他认为转化法是英语形成新词的重要方法之一, 各种词类经常相互转化, 一个词可用作几种词类, 并认为汉语情况相似。邢福义持类似的见解, 只不过他称之为“词性裂变”(1997, 转引自周领顺, 2003: 15)。如前所述, 由于汉语的“兼类”有两种不同的内涵: 一种是“转类”, 一种是原本就具有两种以上恒定的词性, 我们认为, 根据词类划分的“以义分类”, “以位归类”原则与标准, 汉语中属于“转类”内涵的那类“兼类”词就是“转类”。(4) 关于“转类”, 或者说, “活用”与“兼类”的甄别。谭景春(1998: 368—377) 指出“语言的演变是无休止的, 无时无刻不在发生着变化。我们把词类转变看作是一个过程, 包括兼类和活用, 兼类是词类转变的完成, 活用是词类正处在转变之中, 或仅仅是临时发生了转变。典型的兼类或活用虽然有本质的区别, 但从活用到兼类是逐渐过渡中, 在过渡中两者的界线不是很分明。”郑丽(同上: 7—8) 则根据罗施的“原型理论”, 从范畴成员之间的典型与非典型论述出发, 化解了汉语界对“活用”与“兼类”的争论, 对谭景春的观点作了背书。(5) 研究角度。从开展研究必须依托的学科或理论来看, 汉语界对词语转类或者说活用的研究主要从语法学和修辞学的角度展开, 不过近年也不乏从认知语法的新视角对汉语词语转类进行探讨, 如王冬梅(2001: 35) 指出“动词转指名词本质上是一种概念转喻”, 并根据转喻模型对动词名词化中的“转指”现象作了较为全面的归纳和解释。总体上, 前辈时贤对转类问题进行了长期、广泛的研究, 但研究主要集中在上述有限的领域, 鲜见系统的研究成果, 这从中国语法学界至今还没有一部关于转类问题的研究专著就可见一斑。

1.4 批评与展望

由此可见,词语转类研究取得了不菲的成果,但是有待解决的问题仍然不少,主要表现为:一、从现有文献来看,学者们虽然已从语法、语义、语用、功能、认知、修辞等多个角度对词类转用现象进行了观察、描写和解释,但都存在这么一个问题,即语料稀少而陈旧。文论中词类转用的语料多为学者们相互间的援引,“拿着小本钱做大买卖”,“引证的事例左右离不了大路边儿上的那些个”(吕叔湘,2002:354)。缺乏对此类语言现象的全面观察、系统描写和充分阐述,研究东鳞西爪,成果不成体系。可是,语言研究必须做到观察、描写和解释的三个充分,其中,观察充分又是描写、解释充分的前提与基础,所以,如果没有对现代汉语中正处于转类过程中的词语转类案例的充分性观察,就不大可能对其作出充分性描写与解释。其二,从方法论审视,缺乏文本类型学视角,导致了对该类词语在不同文本类型里分布与出现频率方面描写的缺失;其三,主要局限于名词转类动词的研究,对其他词类之间的转类着力不够,如动→名(向成东,1994;王冬梅,2001)、名→形(谭景春,1998)、形→动、副→动(崔应贤,2002)等,虽有所涉及,但浅尝辄止,缺乏纵深。

1.5 研究目标与方法

鉴于转类词研究的前述缺陷,为克服不足,时下和将来一段时间的研究,我们认为必须在研究内容上向前推进,在研究方法上进行创新。就前者来说,必须做到以下几点:(1) 将以往主要局限于名转动词、名转形容词的有限研究,扩展到汉语里名词、动词、形容词等主要词类之间相互转换的系统、全面描写;(2) 将以往主要局限于孤立、静止地在词与句子层面对转类词语在类型、结构、语义上的描写,延伸到从语篇、语境角度对转类词语使用的情景—功能语用分析;(3) 从文本类型的角度,描写转类词语的分布规律,并援用功能语言学、文体学等理论探索其差异背后之根源;(4) 综合演绎相关理论(如认知语言学、语用学),阐述转类词语生成、识读的语言、心理、认知理据与成因;(5) 加大对转类词语的应用研究,如对此类词语的英、汉翻译处

理和英、汉词典编撰中的词性标注问题等。

要突破的难点在于:(1)转类词语的搜集与统计问题。当下,在汉语界,尚没有建立起现代汉语转类词语库,因此对于汉语中到底有多少转类词,具体是哪些,缺乏搜集与统计,对于其分布、频率等方面的研究更是空白。开展此项工作的难度在于:一方面,当代汉语语料浩如烟海,另一方面目前尚未见能有效识别和标注此类词语的检索工具,因此,如何建立起一个能较全面地反映所涉语言的转类词语库自是有待克服的困难。(2)转类词状态判断与发展趋势预测问题。(3)对汉语转类词语的前述广度和纵深研究涉及了众多学科和理论,对这些理论的把握以及如何将之与本课题研究问题嫁接也是富有挑战性的难题。

就方法论而论,我们认为由于词语转类涉及语言学研究的诸多方面,要对其作出较为全面的描写与阐释自然不能囿于某个单一角度,局限于某一理论,而必须根据其所涉及的不同侧面,运用相关理论为指导开展研究。

当务之急是要以现代大型、权威性汉语词典和文本类型学指导下选取的代表性汉语文本为语料,以语料库语言学为指导搜集转类词案例,统计其频率,同时借助于已有研究文献中的转类词案例,完成现代汉语转类词库建设;在此基础上,就词语转类研究的前述内容,根据不同研究侧面,引入相关学科理论为指导,对之作出全面的描写与阐释。在研究方法上,要微观与宏观结合:既在微观层面上对其诸如结构、语义、类型等方面描写与归纳,又从文本类型、使用情景等高度阐述其功能,描写其分布与频率;定性 with 定量搭配:对转类词状态之判断、描写、预测,是在发展中,还是已经完成,既要以量的统计为基础,又要根据其在不同阶段的表现特征进行定性的判断;描写与解释并存:不仅要当下汉词语转类的状态、特征、趋势等进行系统、全面的描述,还要从“所以然”的角度对之作出阐述;分析与综合同用:分析主要表现于对转类词语的各个侧面研究,而综合运用则是对其作整体描写和得出结论。

基于此,本研究将以现代汉语权威词典《现代汉语词典》(商务印书馆2005年第5版)为蓝本,把在词典里完全没有转类标注(即只

有一种词性标注)的上述三类词语挑出,列为检索项,然后将之输入到北京大学汉语语言学研究中心的现代汉语语料库(网络版)进行检索,从包含检索项的语料里,依据现代汉语词类标准及其句法槽,辨识和分拣出转类词语,建立起现代汉语主要词语转类词库。在此基础上,统计和描写其在不同体裁里的分布与频率,最后,引入相关学科理论,对之作出阐释。

所以,不难看出,本研究的关键是基于前述《现代汉语词典》、汉语语料库和汉语词类标准及其句法槽,完成上述转类词语库的研制。

那么到底如何基于《现代汉语词典》、汉语语料库和汉语词类标准及其句法槽构建出上述词语库呢?对此,本书第三、第四和第五章将予以阐述。

1.6 研究意义

对前述汉语转类词语及其词库的研制涉及了多个学科和理论,既涉及词汇学、语法学、语义学、语用学、修辞学、文体学、功能语言学,还涉及认知语言学、语言类型学、词典学等诸多学科知识,因而转类词可谓是“一沙一世界,一花一天国”之小中蕴大的语言现象。也正因如此,中外语言学界对转类现象的研究持续不断,且随着各种语言学派的兴起繁荣,学者们借鉴各种理论对其进行的描写、分析和阐释都达到了前所未有的广度和深度。然而,诚如周领顺、李速立所言,与国外语言界对转类词的研究相比,国内无论是外语界还是汉语界(尤其是汉语界)都无法与之比肩(周领顺、李速立,2006: 19—20)。也正如司显柱所说,英汉转类词研究虽取得了不菲的成果,但仍在观察、描写、解释充分性,方法论以及词语转类范围等诸多方面存在缺陷。对此,他对转类词研究提出了六点展望,并认为“当务之急是要以现代大型、权威性英、汉语词典和文本类型学指导下选取的代表性英、汉文本为语料,以语料库语言学为指导搜集转类词案例,统计其频率,同时借助于已有研究文献中的转类词案例,完成现代英、汉语转类词库建设”(司显柱,2009: 71—73)。

事实上,根据系统功能语言学关于语言是特殊的符号系统的阐述,语言的功能在于产生意义,而意义是由语言形式体现的,离不开

使用的语境(Eggins, 1994: 53; Thompson, 1996/2000: 10)。这些观点能为我们对语言使用中词类的转类提供理论基础,有助于从语言的形式到功能,从词汇—语法、语义到语境等不同层次实现对这一重要的语言现象的描写和解释。同时,也将加深我们对汉语整体的认识与把握。换言之,对词语转类的研究,并不仅仅是将其作为语言中的一种构词方式或修辞手段的微观研究,而且还是一种由此及彼、由小见大对语言全貌进行探索的路径。

本章参考文献

- Eggins, S. *An Introduction to Systemic Functional Linguistics*. London & New York: Continuum, 1994.
- Thompson, G. *Introducing Functional Grammar*. London: Arnold, 1996.
- 陈承泽. 国文法草创. 北京: 商务印书馆, 1982.
- 程望道. 修辞学发凡. 上海: 上海教育出版社, 1979.
- 崔应贤. 简论鲁迅作品中“转类”辞格运用. 毕节师范高等专科学校学报, 2002(4).
- 郭锡良等. 古代汉语讲授纲要(上册). 北京: 中央广播电视大学出版社, 1983.
- 胡裕树. 现代汉语. 上海: 上海教育出版社, 1982.
- 黄伯荣、廖序东主编. 现代汉语(下册). 北京: 高等教育出版社, 1997.
- 黎锦熙. 新著国文语法. 北京: 商务印书馆, 1951. 吕叔湘. 汉语语法分析问题. 北京: 商务印书馆, 1979.
- 吕叔湘. 中国文法要略. 北京: 商务印书馆, 1982.
- 司显柱. 现代英汉语转类词研究: 回顾与展望. 外语学刊, 2009(1).
- 谭景春. 名形词类转变的语义基础及相关问题. 中国语文, 1998(12).
- 汪榕培、卢晓娟. 英语词汇学教程. 上海: 上海外语教育出版社, 1998.
- 王冬梅. 现代汉语动名互转的认知研究. 中国社会科学院研究生院博士学位论文, 2001.
- 王力. 词类. 上海: 上海教育出版社, 1984.
- 邢福义. 词类辨难(修订版). 北京: 商务印书馆, 2003.
- 许余龙. 对比语言学概论. 上海: 上海外语教育出版社, 1997.

张涤华、胡裕树、张斌、林祥楣主编. 汉语语法修辞词典. 合肥: 安徽教育出版社, 1988.

张勇. 现代汉语名、形、动词类活用情况考察. 首都师范大学硕士学位论文, 2009.

郑丽. 词类活用的多视角研究. 福建师范大学硕士学位论文, 2003.

周领顺. 名转动词变异的本质与度. 外语与外语教学, 2003(4).

周领顺、李速立. 我国的英汉转类词研究. 外语教学, 2006(4).

朱德熙. 语法讲义. 北京: 商务印书馆, 1982.

第二章 关于《现代汉语词典》和“北京大学现代汉语语料库”

由于本研究是建立在《现代汉语词典》(商务印书馆,2005 版)和北京大学汉语语言学研究センター研制的现代汉语语料库(网络版)基础之上完成的,所以有必要对该词典和该语料库作简要介绍。

2.1 关于《现代汉语词典》

《现代汉语词典》是由国务院下达编写指示,由国家级学术机构中国社会科学院语言研究所编写的以推广普通话、促进现代汉语规范化为宗旨的工具书,是中国第一部规范型现代汉语词典。它的两任主编吕叔湘和丁声树,皆为享誉中外的语言学家,审订人员均为国内一流的专家和学者。北京大学汉语语言学研究センター主任陆俭明教授肯定《现代汉语词典》“是读者最可放心使用的一部工具书”,在现代汉语语文类词典中有着崇高的地位。

作为一部权威的词典,《现代汉语词典》从编纂到修订都不是轻而易举的,更不是一朝一夕就能完成的。它 1956 年由国家立项,1958 年 6 月正式开编,1960 年印出“试印本”征求意见,1965 年印出“试用本”送审稿,1973 年内部发行,1978 年第 1 版,1983 年第 2 版,1996 年第 3 版,2002 年第 4 版,2005 年第 5 版。50 年来,《现代汉语词典》根据语言发展变化和国家颁布的语文规范不断精雕细琢,与时俱进,经过几代学人呕心沥血,刻苦钻研,在词典学和辞书工具书界获得了许多美誉:国家图书奖、国家辞书奖、中国社会科学院优秀科研成果奖、吴玉章人文社会科学奖,是语文词典的“经典之作”、“里程碑之作”,现代汉语语文工具书的“母本”、“国家的名片”等。

正是《现代汉语词典》的权威性,人们对《现代汉语词典》的研究,自它问世之日就开始了。自 20 世纪 50 年代至 1993 年共发表研

究论文 160 余篇,自 1993 年以来的研究论文逾 200 篇。在汉语词汇研究中,《现代汉语词典》已经成为一种有很高“公信力”的资源。在中文信息处理界,对《现代汉语词典》的利用已经成为一种普遍的现象,它已经被编排成不同格式的电子版,成为一种不可或缺的汉语词汇资源。在语文教学界,在新闻出版界,在人们的日常运用中,《现代汉语词典》的判断已经成为人们取舍的依据标准。

《现代汉语词典》(第 5 版)的修订从 1999 年开始,历时 6 年,是历次修订中修订幅度最大的一次。第 5 版共收字、词、词组、成语和其他熟语 65 000 余条,基本上反映了当前现代汉语词汇的面貌,能够满足读者查考的需要。这次修订重点体现在四个方面:

(1) 全面标注词类。新版《现代汉语词典》吸收了近年来汉语语法研究的新成果,第一次全面地对词类进行了标注,在区分词和非词的基础上,把词汇分成了名词、动词等 12 大类,给今后汉语研究和教学工作带来很大便利。

(2) 增收新词。在 1994 年修订本的基础上共增收新词语 6 000 余条,范围涉及语文、政治、社会生活、科学技术等诸方面。增收新词坚持既慎重又积极的原则,把已经在社会上广泛使用的或有影响力、有生命力的新词语及时收进词典。

(3) 删减旧词。共删减旧词约 2 000 余条。原则是删去那些过于古的、旧的、偏的和“方”的词语。包括纯文言词、使用地区狭窄的方言词、过时的音译词、反映过时的事物、现在已经不再使用的词等几个方面。

(4) 修改释义和例句。对释义、例句逐条作审查,对一些条目作必要的修改。主要包括三种情况:词义发生变化,词义所反映的客观事物发生了变化,原来的释义不够准确或完善。

鉴于《现代汉语词典》(第 5 版)的规范性、权威性及其词类标注的自身特点,故本研究以该词典为研究蓝本。

2.2 关于“北京大学现代汉语语料库”

北京大学现代汉语语料库是教育部“教育振兴专项资助项目”:现代汉语句子结构规则研究及其辅助软件环境的开发,是主要由北

京大学开发建立的大型现代汉语语料库。截止到 2009 年 7 月 20 日,该语料库共收录了 728 909 261 字节的语料,其中口语语料 259 506 字节,书面语料 728 649 755 字节。

北京大学现代汉语语料库以汉字为基本单位,未对中文文本进行分词处理,可以通过特殊符号与基本项的组合、过滤项的使用,实现简单检索、复杂检索以及特定标点符号的检索;以语言文字的学术研究、语文教育和语言文字的社会应用为主要服务对象;作为大型语料库,共收录了 728 909 261 字节的语料,且语料覆盖面广、抽样合理、分布均匀、比例适当;在自动加工的基础上进行了大量的人工加工,这一人机结合策略的运用保证了语料库系统建设的准确率;立足国内外的长远需要,选材有足够的时间跨度,时效性较强(目前更新至 2009 年 7 月),能够比较科学地反映现代汉语全貌。