



## 试论核心词及其类型<sup>①</sup>

赵世举

**摘 要:** 所谓核心词,就是词汇系统的核心部分。但对其性质、范围、标准等根本问题的认识,分歧颇大。主要症结在于把核心词绝对化了,忽视了其相对性和多样性。其实,核心词有不同的层次和类型。从语言基础来看,有单语核心词、多语核心词和全语核心词;从存在状态看,有语言核心词和语用核心词;从时间维度看,有共时核心词和历时核心词;从语域看,有通语核心词和方言核心词;从词汇地位看,有轴心核心词和边缘核心词等。这些不同类型的核心词既相联系,又相区别。

**关键词:** 核心词;常用词;高频词;基本词汇

在任何语言的词汇系统中,词都不是等价的。不同的词往往在地位、功能、价值等方面具有差异性。有些词地位高、功能强、用途广、价值大,它们构成词汇系统的核心部分,也是最重要的部分。那么这些词该怎么称呼?依据什么标准来确认?可谓见仁见智。不同的人由于视角不同,或依据的标准不同,或采用的筛选方法不同,因而确认的范围及具体对象也就不同,所叫的名称也各异。已有的关于“基本词汇”、“常用词”、“高频词”、“核心词”的研究大体与此相关,但它们各有侧重,并不能简单地相提并论。从总体上看,对这一问题的认识在理论上还较为模糊,实践上也有困惑,颇有争议。其主要症结在于,对词汇核心部分的认识有些绝对化和片面性,忽视了其相对性和多样性,因而当不同学者立足点不同、判断标准不同时,自然就出现了不同看法,得出了不同结论,随之也就难免各执一端,是己而非他。

其实,上述相关研究并不相互矛盾,只是从不同角度、按照不同标准对不同性质的重要词汇所做的不同考察而已,各有其旨趣和价值。不过问题在于,有些研究毕竟只是某一视角下的局部性研究,或对重要词汇的某一方面的研究,并非对词汇核心部分的全面认识,因而不具备充分的解释力和全面的涵盖力。

我们认为,词汇的核心部分是相对的,并没有固定的范围和绝对的对象。在不同的条件下,词汇的核心部分是不同的,不能一以定之。我们把词汇系统的核心部分——常用度高、分布面广、稳定性强、流通域宽、组合力强的那些词叫做“核心词”。之所以叫核心词,意在表明它与其他词相比较处于核心地位,是词汇中最重要的部分;自然也表明它的作用、价值比其他词更大。还有一点需要说明的是,“核心词”这一概念有一个非常明确的预设:它是整个词汇系统或词汇系统中某个局部的组成部分,是相对于“非核心词”而言的。这也是我们讨论核心词的前提和基础。

<sup>①</sup> 本文的主要观点曾在“词汇学国际学术会议暨第九届全国汉语词汇学学术研讨会”(2012,山东大学)大会上报告过。

因此,要确认核心词,必须根据具体情况,综合不同的条件来筛选。不过一般来说,无论是哪类核心词,大体都有如下共性特征:

- (1)常用度高。即经常被不同的人所使用。
- (2)分布面广。可在不同语体、不同题材、不同文体中使用。
- (3)稳定性强。即在一定时期内使用稳定,而非昙花一现。
- (4)流通域宽。通行于同一语言集团的不同地域、不同人群之中。
- (5)组合力强。能较自由地与其他词语进行较广泛的组合。

以上几点既是核心词的基本特点,也是筛选核心词必须依据的因素和条件。据我们初步考察,核心词大致可以分为如下主要层次和类型:

## 一、从语言基础来看,有单语核心词、多语核心词和全语核心词

基于不同语言基础的核心词具有很大的差异性。有的是以一种语言的词汇为对象来考察其核心词,有的则基于多种语言,甚至跨语族跨语系来提取不同语言的共有核心词,还有的采取各种手段试图探寻人类所有语言的共有核心词。这就形成了不同类型的核心词。

### (一) 单语核心词

即某一种语言词汇系统的核心部分。比如汉语核心词。

单语核心词实际上是个非常宽泛的概念。它既可以指一种语言的跨时空核心词,也可以指一种语言某一类型的核心词。例如“汉语核心词”,作为跨时空概念,指汉语词汇系统中历代共有的、并且共同语和各方言兼具的核心词汇。对于这种严格意义上的汉语核心词,目前还未见系统的研究成果。汉语核心词更多的是用来指汉语某一类型的核心词,即从不同角度出发筛选出的核心词。例如《尔雅》的《释诂》《释言》篇中所收录的词语大多属于上古汉语核心词;当代一些学者所研究的汉语基本词汇中的重要词汇,也具有核心词的性质,类似的研究成果如郑林曦等所编《普通话三千常用词表》(1959年),何克抗等主编的《现代汉语三千常用词表》(1987年)等。另如英国语言学家奥格登20世纪30年代研制的《基础英语词汇表》、韦斯特1953年编制的《英语通用词表》,属于英语核心词的一类;1935年加拿大学者贝克编纂的《法语词汇手册》,1954年法国基础法语研究中心古根内姆等研制的《基础法语》词表;1964年,德国学者普菲费尔编制的《基础德语(口语)词表》等。

### (二) 多语(或跨语)核心词

即多种语言共有的核心词,是不同语言词汇的共核。这类核心词必须通过对不同语言进行比较研究才能确定。历史比较语言学界所构拟的亲属语言共有原始词汇大体属此类。例如俄罗斯学者多尔戈波尔斯基于1964年筛选出的北欧语言15个核心词;美国学者马提索夫1978年提出的东南亚语言(藏缅语)的200词;我国学者郑张尚芳1995年提出的《原始华澳语三百核心词(征求意见稿)》;黄布凡1997年提出的《藏缅语300核心词词表》;江荻2000年提出的《汉藏语言语义词表》(200词);美国学者伊通1940年编纂的《英语、法语、德语、西班牙语比较频率词表》;新西兰学者从大约998种太平洋诸语言中提取的210个最基本词等<sup>①</sup>。

### (三) 全语核心词

即人类所有语言共有的核心词。这是一个假设性存在,因为目前还没有,事实上也难于得到充分性验证,而且这种意义上的“词”本来就是一种宽泛的界定。但也不乏研究成果,例如美国莫里斯·斯瓦迪士在20世纪50年代初为推算语言的发展年代,运用统计学等方法,主要依据印欧语言所研制的《二百词表》(后来修订为《百词表》)大致属此类。它的选词原则是:“‘语素项目必须是世界共同的,非文化方面的,容易辨认的广阔概念,在多数语言中有一个单词可以对应的’,也就是选择一些与不同的生活环

<sup>①</sup>参见 <http://language.psy.auckland.ac.nz/austronesian/>。

境、不同的物质文化条件无关、不易受另一语言影响的基本词根语素。”<sup>①</sup>该词表经过了汉语和非洲科普特语等语言的初步检验,后来众多学者利用它来研究各种不同的语言或方言,其普适性不断得到验证,因而成为国际上被广泛认可和使用的准全语核心词表。此外,美国学者伯林和凯通过对近百种语言的比较研究,提取的 11 个共有颜色词<sup>②</sup>,也大体具有全语核心词的性质。

## 二、从存在状态看,有语言核心词和语用核心词

语言有两种存在状态,一种是静态符号系统,一种是动态运用。在这两种不同的状态下,核心词也是不同的。我们把静态语言系统的核心词叫语言核心词,把动态运用状态下的核心词叫语用核心词。

### (一) 语言核心词

即语言系统在非使用状态下的核心词。它是静态词汇系统最核心的部分,也是最基础的部分,它决定一种语言词汇系统的基本面貌。语言核心词具有抽象性和稳定性。

所谓抽象性,即语言核心词是对客观对象高度概括的抽象符号,而不与任何特定对象相联系。例如“人”,并不指任何特定的人,而是对人类共性特征的抽象,指“能制造工具并使用工具进行劳动的高等动物”<sup>③</sup>。

所谓稳定性,表现在两个层面:一是在共时层面,语言核心词不受语境影响,基本上是固定的,尽管不同的语用范畴有不同的语用核心词,但各类语用核心词都是以语言核心词为基础和主体的;从历时层面看,虽然词汇系统也会演化,但其演化是渐进的,在一定时期内,其核心词也总是相对稳定的,即使稍有变化,其主体部分还是稳定不变的。也正是由于语言核心词的稳定,才保证了语言的传承、发展和语言交际的顺利进行。

对词汇核心部分进行的泛时空研究涉及的大都是语言核心词。例如英国学者斯达伯司研究的核心词,大体上就是语言核心词。他提出了判断核心词的十二条标准<sup>④</sup>:

(1)在含义上是中性的。具有纯概念的、逻辑的、认识的和陈述性的意义,不具有观点、感情或判断上的含义。

(2)不带文化色彩,不受文化的约束。

(3)在使用上也是中性的,不暗示或隐含上下文的内容。

(4)在会话的场合方面也是中性的,它们不受是正式还是非正式场合的限制。

(5)在总结性的表达中人们往往优先选用核心词。

(6)一般更具类别性,它常常在一些构成上下位关系的词群中充当其中的上位词。

(7)相对于其他词语,核心词是更一般化的词,所以核心词常常可以用来替代非核心词,而反过来则不行。

(8)核心词可以用来定义非核心词,反之则不大可能。

(9)在语义类型或功能相同的一组词中,核心词一般有较强的组合能力,非核心词相对而言其搭配要受到更多语义和句法上的限制。

(10)由于核心词更具一般性,所以它们在扩展方面的特点必然是:生成新义的能力更强。

(11)核心词能构成更多的复合词。

(12)核心词并不是简单无序的堆积,而是一个有机联系的整体。例如,核心词通常有明确的反义词与其对应,非核心词则不太容易联想出与之对应的反义词。

以上这些大体上反映了语言核心词的一般特征。

### (二) 语用核心词

即语言在运用状态下(亦即言语中)的核心词,也可叫言语核心词。语用核心词是动态的,因使用领

①徐通锵:《历史语言学》,商务印书馆 1991 年,第 414 页。

②这 11 个颜色词及其层级是:黑、白;红;绿、黄;蓝;棕(褐);紫、粉红、橙、灰。

③《现代汉语词典》(第 6 版),商务印书馆 2012 年,第 1090 页。

④翟颖华:《核心词研究述评》,载《长江学术》2012 年第 3 期。

域、语体、运用目的等不同而有差异。比如,同一种语言的口语核心词和书面语核心词往往就有不同。有学者对英语的情况进行过较具体的研究,并在《朗曼当代英语词典》(第三版,1995)的相应词条中做了标示。

语用核心词又有不同的层次和类型,可再分为一般交际核心词和特定语用核心词。

### 1. 一般交际核心词

一般交际核心词,指人们日常基本交际使用最频繁、用途最广泛的重要词汇。这类核心词较为稳定,是词汇最基础的部分,大体上与语言核心词的主要部分相交集。例如前面提到的奥格登的《基础英语词汇表》,就被认为是“通向通用英语的基础”;1953年韦斯特的《英语通用词表》,也同样强调其“通用”性质,据称,掌握了这些词,就可以读懂80%的文献。一般频率词典、常用词词典、通用词词典所收录的主要词汇都是一般交际核心词。

### 2. 特定语用核心词

特定语用核心词,指基于不同领域、针对不同对象、服务不同目标所使用的核心词。这类核心词类型繁多,千差万别,这里略举数端。

#### (1) 语言教学核心词

即出于语言教学的需要筛选的核心词。我国古代精选汉字编纂的识字课本,如上古时期的《史籀篇》及后来的《三字经》、《千字文》等,就类似于这种性质。国外最初的核心词研究,也主要是出于教学的需要,如德国凯丁编纂的《德语频率词典》(1898年),美国桑代克编纂的《教师两万词词书》(1921年)等。

语言教学核心词,因服务于不同教学对象,有母语教学核心词和非母语教学核心词的区别。这两者有明显的不同,因为母语教学对象,都有一定的语言基础,教学大都在母语环境中进行,而且兼有全面的语文教育乃至人文教育的任务;而非母语教学的对象,一般缺乏语言基础,大多在非目标语环境中进行,一般也未必要求兼具全面的语文教育和人文教育的任务,而且通常要借助学习者的第一语言来辅助教学,难免会有第一语言的干扰和跨文化的障碍,因此,在核心词的筛选上往往就会有差异。

母语教学核心词:一般以学习者第一语言的语言核心词为基础进行筛选,它既是一种语言最重要、最基础的词汇,也是学习者应该掌握的最基本词汇,是进一步学习和运用的基础。例如郑林曦等所编《普通话三千常用词表》就是运用统计方法筛选出的普通话核心词汇,所选词语都是“日常生活和文化学习上最常用的,这三千多个词在一般读物中常用率约在70%以上……可以供学习普通话和学习文化以及编辑教材等方面参考”<sup>①</sup>,国家语委2008年发布的《语文新课标教材3000基本词语表》等。

非母语教学核心词:就是供第二语言/外语教学使用的最重要词汇。这类核心词的筛选除了考虑具体词汇在词汇系统中的核心地位之外,也要考虑跨文化交际因素,因而跟母语教学核心词往往有差异。不过,其主要部分是相同的。

国内外关于非母语教学核心词的研究较多,已有不少成果。我国的相关研究始于20世纪五六十年代,并且先后研制了不少供教学或水平检测使用的词表。尽管筛选标准不尽一致,做法不同,名称也各异,但多数词表大体上还是能够反映核心词的基本面貌的。例如1964年原北京语言学院的《外国学生用四千词表》;1988年中国对外汉语教学学会的《汉语水平等级标准和等级大纲·词汇等级大纲》;1992年国家汉办的《汉语水平词汇与汉字等级大纲》;1999年杨寄周主编的《对外汉语教学初级阶段教学大纲·词汇大纲》;2010年国家汉办又推出的《汉语国际教育用音节汉字词汇等级划分》;翟颖华2012年推出的《面向第二语言教学的现代汉语核心词表》<sup>②</sup>等。非母语教学核心词具体选取多少为宜?多数人主张在1000—5000词之间。英国劳特利奇出版社陆续推出的“供学习者使用的核心词”,为八种语言各选取了5000个左右的词。美国之音的“特别英语”学习节目确定的核心词为1500个。

语言教学核心词,因服务于不同的教学目标,又有基础教学核心词和主题教学核心词的不同。

①《普通话三千常用词表·内容说明》,文字改革出版社1959年。

②翟颖华:《面向第二语言教学的现代汉语核心词研究》,武汉大学博士学位论文2012年。

基础教学核心词:是为满足最基本教学目标和日常最基本运用而筛选的重要词汇。这些词汇其实就是一般交际核心词汇。例如1988年中国对外汉语教学学会组织研制的《汉语水平等级标准和等级大纲·词汇等级大纲》中的甲级词表,有学者对国内12套对外汉语教材用词情况进行抽取编制的《前1500条高频词语表》<sup>①</sup>,普菲费尔编制的《基础德语(口语)词表》。有学者提出的“最低词汇量”,更应属于基础教学核心词。例如北京语言大学编制的《外国留学生汉语教学大纲(长期进修)词汇表·初等阶段词汇(最常用)》为764个词,史有为等主张最低词汇量以500—600之间为宜<sup>②</sup>。国家汉办《汉语国际教育用音节汉字词汇等级划分》设定的普及化等级词汇中的“最低入门等级词汇”505个,这些都属于基础教学核心词,适用于语言的起步教学。

主题教学核心词:是针对特定学习目的或服务特定教学对象的需要而筛选的核心词。这类核心词主题集中,关联性强,明显体现出语言交际的实用性、学科专业的一致性和教学对象的针对性。如上海师大研制的《旅游汉语词汇大纲》,根据“旅游图式”选取汉语旅游核心词汇4674条;国家汉办和北京大学联合制定的《商务汉语考试大纲》中的《商务汉语常用词语表》,针对商务目的收词2457个。另如埃伊诺阿·拉腊尼亚加·多明盖斯等编《商务西班牙语词汇》,服务于特定教学对象的《剑桥少儿英语词汇表》等。

## (2) 辞书释义核心词

词汇是巨量的、开放的,但人们都希图运用有限的词语来解释无限的词汇,这就形成了释义核心词。也就是Dailey所说的“能定义词典中所有其他词汇的一套语义原语”<sup>③</sup>。释义核心词也是“元语言”研究的内容之一,苏新春(2003)称之为“释义元语言”。Wierzbicka论述说:1)任何语言的词典中都存在不可定义的词,它们的数量较少,自成系统,它们的作用是用来定义其它的词语。2)不可定义的词是可列举的,语言中的其它词可以用它们来定义。3)不可定义的词在不同的语言中虽然各有所不同,但却是相互对应的,在语义上是等价的。因此,不可定义的词在各种语言中可视为“普遍词汇”<sup>④</sup>。释义核心词最突出的特点是通用性、简明性和有限性。

对于辞书释义核心词,无论是理论研究还是实际应用都有不少成果。例如韦斯特等所编《英语教学词典》(第4版)选用了1490个词,用以解释24000个词项;《朗曼当代英语词典》选用了2000个词,用以解释56000个词项,连用例用词也在其中。我国在辞书释义核心词研究方面也取得了不少成果。1996年张津、黄昌宁筛选出“定义原语”3857条<sup>⑤</sup>;李葆嘉提出了“释义元语言系统”概念<sup>⑥</sup>;苏新春提取出释义元语言4000条<sup>⑦</sup>。安华林提取现代汉语词典释义元语言近3000个“词元”<sup>⑧</sup>。

## (3) 专科核心词

指一定专业领域词汇中的最重要词汇。它反映一定领域最基本的概念和相关的重要事物及现象,主要为一定的学科专业、行业领域服务。前面提到的《旅游英语词汇2000》、《剑桥商务英语核心词汇》等收录的词汇,也可看做专科核心词。另如原苏联学者别列斯涅夫等编纂的《德语畜牧学频率词典》(收词2050个),阿列克谢耶夫和图里金娜合编的《英俄频率词典—最低限量的新闻词汇》<sup>⑨</sup>等。

## (4) 专库核心词

为了某种特定需要而筛选的封闭性专用核心词。这类核心词针对性强、功能明确。手机、翻译器、学习机等电子终端产品的内置词(字)库即属此类。例如网易公司推出的有道手机词典内置的本地词库,提供了常用中英文词汇,可满足离线查询的基本需要。

① 国家语言资源监测与研究中心:《中国语言生活状况报告》(2006)下编,商务印书馆2007年,第205页。

② 史有为:《对外汉语教学最低量基础词汇试探》,载《语言教学与研究》2008年第1期。

③ 张津、黄昌宁:《从单语词典中获取定义原语的一种方法》,载《清华大学学报(自然科学版)》1997年第3期。

④ 张津、黄昌宁:《从单语词典中获取定义原语方法的研究及现代汉语定义原语的获取》(项目研究报告),1996年。

⑤ 苏新春:《元语言研究的三种理解及释义型元语言研究评述》,载《江西师范大学学报》(哲学社会科学版)2003年第6期。

⑥ 李葆嘉:《语义语法学理论和元语言系统研究》,载《深圳大学学报(人文社会科学版)》2003年第2期。

⑦ 苏新春:《汉语释义元语言研究》,上海教育出版社2005年。

⑧ 安华林:《现代汉语释义元语言研究》,南京师范大学博士学位论文2004年。

⑨ 冯志伟:《频率词典》,载中国百科网·百科词条·教育百科。

### 三、从时间维度看,有共时核心词和历时核心词

词汇犹如生物的细胞,总是处于不断的新陈代谢之中。因此,一种语言的词汇系统并非一成不变,在时间维度上总会有不同的面貌,表现出很强的时代性;不言而喻,核心词也会发生变化。大而言之,有共时核心词和历时核心词之分。

#### (一) 共时核心词

即一种语言在一定的历史阶段所具有的核心词。例如汉语核心词,简单来说,有古代汉语核心词和现代汉语核心词之分;稍稍细分一下,有上古汉语核心词、中古汉语核心词、近代汉语核心词和现代汉语核心词之分。还可以依据一定的条件做更细的历史阶段的划分。例如苏培成在研究汉语基本词汇时把现代汉语分为四个阶段:1919—1949;1950—1965;1966—1976;1977—现在<sup>①</sup>。各个历时阶段的核心词系统及其成员不尽相同。以具体的词为例,“王”、“曰”无疑是上古汉语核心词,但不是现代汉语核心词;“计算机”、“电”无疑是现代汉语核心词,但不是上古汉语核心词。

前面所提到的国内外关于各种核心词的研究成果大多属于共时核心词,主要是现当代语言核心词。汉语学界对于历史上的共时核心词的研究也有一些成果。例如徐朝华《上古汉语词汇史》对上古汉语核心词有较多较系统的研究,吴宝安《西汉核心词研究》提取西汉核心词 866 个。

#### (二) 历时核心词

即一种语言跨时代的核心词。这类核心词是词汇系统中最核心最稳定的部分,也是一般所谓基本词汇中最具生命力的重要部分,一般变化不大。也可以叫做根核心词。正是由于历时核心词的存在,保证了语言系统的稳定、传承和发展。

符淮青曾指出:“基本词汇是从古代到现代,在实际运用中必不可少的词,它们表达的是对于人们交际最不能缺少的概念。”他列举了基本词汇的八种类型<sup>②</sup>。其中使用频度高、分布面广、通行域宽、组合能力强、且古今通用的词当为历时核心词。此外,古今通用的重要虚词也是历时核心词的组成部分。

要确认一种语言的历时核心词,首先必须对该语言各个历史时期的共时核心词进行系统研究,然后抽取各个不同历史时期共时核心词的交集部分,即为该语言的历时核心词。比如要确认汉语历时核心词,就必须以先秦至当代各个历史时期的共时核心词为基础,从中抽取共有的部分。不过,历时核心词也有相对性。还是以汉语为例:既有从先秦至当代的全程绝对历时核心词,也有某一时段内的相对历时核心词,例如先秦至隋唐时期的历时核心词。一般而言,选取的时段越长,得到的历时核心词的数量就越少。

### 四、从语域看,有通语核心词和方言核心词

由于地域和社会阶层行业的差异,语言往往都会形成通语和方言的差异。我国先秦时期就有“雅言”和“齐语”、“楚语”等方俗之语的不同,汉代扬雄的《方言》就是上古汉语“通语”与“别国方言”的记录。可以断言,通语核心词和方言核心词也会有一定的差异性。

#### (一) 通语核心词

就是在操同一种语言的全体成员中通用的重要词汇。它是一种语言最重要的内核,为四方之民、各色人等所常用。先秦所谓“雅言”中的重要词汇、《尔雅》所收录的主要词汇大体都是上古汉语的通语核心词。扬雄《方言》中的训释词和无地域限定的“通语”、“凡语”大多属于那时的通语核心词。例如:

党、晓、哲,知也。楚谓之党,或曰晓,齐宋之间谓之哲。(《方言》卷一)

娥、嬖,好也。秦曰娥,宋魏之间谓之嬖,秦晋之间凡好而轻者谓之娥。自关而东河济之间谓之媼,或谓之姁。赵魏燕代之间谓之姝,或曰姝。自关而西秦晋之故都曰妍。好,其通语

① 苏培成:《关于基本词汇的一些思考》,载《词汇学新研究》,语文出版社 1995 年。

② 符淮青:《现代汉语词汇》(增订本),北京大学出版社 2004 年,第 157 页。

也。(《方言》卷一)

上例中训释词“知”、“好”都是通语核心词,其中“好”还被明确交代为“通语”。

现代汉语普通话核心词是现代汉语通语核心词,为操现代汉语的人所通用。郑林曦等所编《普通话三千常用词表》,翟颖华研制的《现代汉语通用核心词表》(收词 3000 个)大体属于现代汉语通语核心词。奥格登的 850 词应属于英语通语核心词。

通语核心词也是个相对概念,既有严格意义的通行于整个语言集团的全域通语核心词,也有一个语言集团内部只通行于若干方言区的区域通语核心词。这类核心词也古已有之,例如:

悼、怒、悴、愁,伤也。[伤]自关而东汝颖陈楚之间通语也。汝谓之愁,秦谓之悼,宋谓之悴,楚颖之间谓之愁。(《方言》卷一)

嫁、逝、徂、适,往也。自家而出谓之嫁,由女而出为嫁也。逝,秦晋语也。徂,齐语也。适,宋鲁语也。往,凡语也。(《方言》卷一)

其中的“伤”、“愁”、“逝”、“适”为区域通语核心词,“徂”为方言核心词,“往”为全域通语核心词。

## (二) 方言核心词

有两个不同层面的方言核心词:一是地域方言核心词,二是社会方言核心词。

地域方言核心词是指在同一个语言集团中通行于不同地域的核心词。例如汉语的北方方言核心词、吴方言核心词、粤方言核心词等。刘俐李等编著的《现代汉语方言核心词·特征词集》是目前见到的关于方言核心词研究的最系统的成果。

社会方言核心词是指在同一个语言集团中通行于不同社会群体的核心词。因职业、阶层、年龄、性别、宗教信仰、受教育程度等因素影响而形成的不同社会群体,都会形成自己的社会方言,因而在核心词的构成上都可能有差异。

因为方言是语言的变体,因此,无论是地域方言核心词,还是社会方言核心词,都由通语核心词和方言个性核心词组成。方言核心词是通语核心词的基础,通语核心词是方言核心词的内核。

## 五、从词汇地位看,有轴心核心词和边缘核心词

语言的核心词并不是一个简单的并列集合,而是一个有规律分布的连续统。1916 年法国速记专家艾思杜发现,在一定量的文本中,词出现的频次与频序之积为一个常数,可表示为  $n_r \cdot r = K$  (其中  $n$  表示频次,  $r$  表示频序,  $K$  为常数)。后来经过美国物理学家贡东的进一步研究(1928)和美国语言学家齐普夫的验证(1935),形成了著名的齐普夫定律,认定  $K$  (有学者记作  $C$ ) 是一个参数:  $0 < K < 0.1$ <sup>①</sup>。这说明核心词的地位并不是均等的,但又是有序分布的。

核心词在实际语料中的覆盖状况也表明它是一个连续统。《现代汉语频率词典》显示,175 个高频词可以覆盖语料的 49%,1230 个词可以覆盖 76%,3000 词可覆盖 86%,5000 个词可覆盖 91%,8000 个词可覆盖 95%<sup>②</sup>。据《中国语言生活状况报告(2008)》对 5.56 亿词次(纯词语,不含标点、符号、纯西文、纯阿拉伯数字、数字与西文混合式、网址)的语料统计显示,总共使用的词种达 2261272 个,各词种的覆盖率也明显不同,这里可把覆盖率在 80% 以下的情况列表如表 1<sup>③</sup>。

从上表可以看出,少量的核心词具有很高的覆盖率,而且也呈不均等分布。

根据核心词在其连续统中的地位,可分为轴心核心词和边缘核心词两类。所谓轴心核心词,即处于核心词连续统内中心部位的那些词,它们最具典型性,是词汇最重要的部分;所谓边缘核心词,是核心词连续统内渐离中心、逐步临近非核心词汇的部分。就目前的情况看,从词汇地位角度研究轴心核心词和边缘核心词的还极少。

① 冯志伟:《齐普夫定律的来龙去脉》,载《情报科学》1983 年第 2 期。

② 北京语言学院语言教学研究所:《现代汉语频率词典》,北京语言学院出版社 1986 年,第 27 页。

③ 国家语言资源监测与研究中心:《中国语言生活状况报告(2008)》,商务印书馆 2009 年,第 16 页。

表 1 词种覆盖率简表

| 词种数  | 占总词种比例 | 覆盖率 |
|------|--------|-----|
| 5    | 0.00   | 10  |
| 34   | 0.00   | 20  |
| 108  | 0.00   | 30  |
| 278  | 0.01   | 40  |
| 595  | 0.03   | 50  |
| 1159 | 0.05   | 60  |
| 2230 | 0.10   | 70  |
| 4640 | 0.21   | 80  |

综上所述,我们大体有如下主要看法:

- (1)核心词具有相对性和多样性。有不同性质和不同类型的核心词。
- (2)不同类型的核心词,其确认标准不同,范围、数量及具体成员也不尽相同。
- (3)核心词是连续统,其内部是有层级的。
- (4)不同类型的核心词,其功能是不同的。
- (5)不同类型的核心词既相联系,又相区别。

因此,我们应根据不同需要,确立相应的标准,采用不同的方法,来研究提取核心词。

On Core Words and Their Types

Zhao Shiju (Professor, Wuhan University)

**Abstract:**Briefly speaking,the so-called core words are the essential part of lexicon system. But the essential knowledge on their features,range,criterion,and so on,differs widely and thus causing ambiguity in theory and difficulty in operation. The major problem lies in the absolutization of core words,neglecting their relativity and diversity. As a matter of fact,there exit different layers and types in core words. According to our primary exploration,from the aspect of language basics,there are monolingual core words,multi-lingual core words and whole language core words; from the aspect of their existence,there are language core words and pragmatic core words; from the aspect of time dimension,there are synchronic core words and diachronic core words; from the aspect of register,there are universal core words and dialectal core words; from the aspect of lexicon status,there are axis core words and marginal core words,etc. These various core words are related to each other as well as different from each other.

**Key words:**core words; common words; high frequency words; basic vocabulary

● 作者简介:赵世举,武汉大学文学院教授,博士生导师;湖北 武汉 430072。Email:zsj@whu.edu.cn。

● 基金项目:国家社会科学基金一般项目(08BYY034)

● 责任编辑:何坤翁

◆