

面向对外汉语教学的汉语新词语信息库的建设*

亢世勇

鲁东大学文学院 / 山东省语言资源开发与应用重点实验室, 山东烟台 264205

kangsy64@163.com

摘要: 本文说明了汉语新词语词典在海外的需求, 介绍了作者目前拥有的新词语的语言资源, 提出面向对外汉语教学的汉语新词语信息库建设的理念, 即收词要全面系统、词语分级、描写信息要全面、便于检索使用, 最后重点介绍新词语的词性信息、例句信息、关联语汇信息以及词源信息。

关键词: 对外汉语教学 汉语新词语 信息库

The Construction of Chinese New Words Corpus Based on the Teaching Chinese to Foreigners

Kang Shiyong

College of Arts in Ludong University, Key Laboratory of Languages' Development and Application in Shandong Province

Yantai China 264025

Abstract: Introducing the need of Dictionary of Chinese New Words and the resources of this we have, the thoughts of creating this kind of language corpus should be considered on totally collecting the words, words' grading, and totally describing the information to convenient searching. Word class on the new words, example sentences, connecting words and the origin of words have been focused on studying.

Key words: teaching Chinese to the foreigners, Chinese new wrds, corpus

一、汉语新词语词典在海外的需求

随着中国社会的发展, 汉语新词语不断产生, 引起了人们的广泛关注。我们从 2001~2004 年主持并完成了国家哲学社会科学规划项目“《现代汉语新词语信息(电子)词典》的开发与应用”, 建立了《现代汉语新词语信息电子词典》及语料库等新词语资源, 编纂出版了《新词语大词典》等系列词典, 并进行理论研究, 由中国社会科学出版社出版《现代汉语新词语计量研究与应用》。我们了解到, 国外对汉语新词语词典需求比国内更迫切。原因是外国人学习汉语, 主要学习汉语的基本词汇; 大量海外华人由于久居海外, 他们掌握的只是汉语的基本词汇, 而他们现在接触到的都是活生生的现代汉语, 包括电视、报纸、网络等。这些媒体中的现代汉语实际上包括了很多新词语, 这些新词语给长期脱离汉语环

* 本文得到教育部人文社科项目(11YJA740040)的支持。

境的人造成一些阅读、理解上的困难，他们很需要有一本比较全面的汉语新词语词典来帮助查检、学习。为了服务海外读者，我们在原有词典的基础上，编纂出版系列汉外新词语学习词典，目前《汉英新词语学习词典》已经定稿，将由外语教学与研究出版社出版，《汉日新词语学习词典》中文稿已经完成，正由日本立命馆大学翻译为日文，将在日本出版。我们希望，还能出版韩文、德文、法文等版本。在总结这些辞典编纂经验的基础上，我们拟将综合国家哲学社会科学规划项目所取得的研究成果，开发面向对外汉语教学的汉语新词语信息库。

二、新词语目前的资源

在国家哲学社会科学规划项目的支持下，目前我们开发建设了以下资源。

1. 现代汉语新词语信息电子词典

目前电子词典已经收录了 1978 年以来近 4 万条新词语，按照语言信息处理的需要，依照语法功能进行了粗分类，分出名词、动词、形容词、区别词、副词等类，采用数据库文件格式，设立总库 1 个，名词分库、动词分库、形容词分库、区别词分库、副词分库各 1 个，设立了拼音、义项、词性以及语法特征等属性信息 200 多个，对每个新词语的语法属性进行了详细的描述。总库和每个分库之间有三个相同的属性字段，可以将这些库连接在一起进行全部信息的检索。

2. 新词语语料库

我们建立了新词语动态的、超大规模的语料库，具体涉及的语料有：《人民日报》1947 年以来所有的语料，人民网里人民日报报系 12 种报纸的近年语料，人民网网站群的 9 个网站近年的语料，《读者文摘》（后更名为《读者》）创刊以来的语料，《南方周末》创刊以来的语料，《光明日报》《中国教育报》《南方日报》网上的语料。这些语料从历时的角度看跨度达 50 年之久，从共时的角度看各个方面的语料都有，是一个系统的动态的语料库。为搜集整理新词语，调查新词语的流通度、稳定性、新颖性以及新词语的意义、用法提供了有力的保证。

3. 由上海辞书出版社出版了《新词语大词典》《学生新词语词典》《最新流行语小词典》《实用网络用语手册》

《新词语大词典》2003 年出版，按照全面性原则、规范性与描写性相结合的原则、必要性原则、普遍性原则、稳定性原则、音节原则从《现代汉语新词语信息电子词典》中遴选了 1978 年到 2002 年间出现 15000 多个词语，每个词条包括注音、词性、释义、例句。这些例句都是从语料库中选出的，是真实文本中的句子。释义除了说明词语的词汇意义外，还尽量进行词源的说明。即将出版的《新世纪新词语大词典》收录了 2000 年以来的 8000 多个新词语。今年我们拟将原来的《新词语大词典》和《新世纪新词语大词典》合并，出版《新词语大词典（1978—2013）》，收词 20000 多个。

《最新流行语小词典》2002 年出版，收录流通度较高又比较新的流行语 1100 多个，比如“CEO、CFO、CGO、CIO、CKO、COO、CPO、CTO、DIY、酷、酷毙、酷机、酷语、茶吧、迪吧、泡吧、书吧、陶吧、网吧、氧吧、话吧”等等。每个词条包括词目、注音、词性、释义、例句和出处等。每个词语（多义词的为每个义项）后列两个例句，一个是最早的，一个是靠近 2002 年的，目的在于说明该词语流行的时间范围。比如：洋教头

yàngjiàotou<名>对在中国执教的外国教练的戏称。例如：①明明是“教练”，偏要写成“教头”，外籍教练则是“洋教头”。（《人民日报》1995年8月4日）②您是否认为随着洋教头的到来，中国教练员也将逐步向智力型转变？（《人民日报》2000年2月2日）。

《实用网络用语手册》2003年出版，收录网络用语约3200条，其中汉字词语约1600条、字母词语约800条、数字词语约400条、符号约400条。这些网络用语从构成来看，主要包括谐音替代型、词汇缩略型、数字会意型、混合型、旧词赋新义、叠音赘语、新词新语、象形表意符号等等。

《学生新词语词典》2006年出版，在《新词语大词典》的基础上，删去了80年末以前产生的新词语，补充了2003年到2005年产生的新词语，共计10000多条。一般没有例句，只有个别不好理解的词条附上了例句。此后隔年修订一次，直至现在。

《现代汉语新词语词典》2007年出版，在《学生新词语词典》的基础上增加当年新出现的词语，隔年修订，直至现在。

《汉英新词语学习词典》收录1990年到以来的产生的、稳定的新词语10000多个立条，加上词汇关联共计收录新词语15000多个。除了注音、词性、释义、例句外，本词典还增加词目英译、关联语汇、知识窗等内容。

《汉日新词语大词典》完全按照描写与规范相结合的原则，收录20世纪80年代以来出现的新词语35000多条。这些词条除了一般的语文词语之外，还包括了一些术语、品牌等等。每个词条包括注音、词性、释义。这些内容都将在中文稿的基础上翻译为日文。

三、新词语信息库建设的理念

1. 收词要全面系统

以往编写出版的新词语词典由于受到各种原则的限制，尤其是规范原则的限制，很多新词语都因不被承认而不能进入词典。我们认为新词语词典除了具有规范的作用外，其实有更重要的作用，就是记录词汇的发展，它应该是词汇发展的“晴雨表”。只有新词语词典全面系统地收录了新词语，才能显示现代汉语词汇的发展，才能显示社会的发展。如果新词语词典不能够全面系统地记录新词语，那么这些新词语将永远沉没于语言实际使用当中，不被发现，就不能完整系统地显示词汇的发展变化，几十年之后，这些词语及其所代表的新事物、新观念就将无从查考。另外，新词语信息库的容量不受限制，可以不断扩充。基于以上考虑，我们拟采用描写性与规范性相结合的原则，全面系统地收录我们可能收集到的全部新词语。

2. 词语分级

李建国先生（1997）指出：新词语研究具有重要意义，“促使我们重新认识它，并适时地将研究成果吸收到辞书中，使之成为文化积累的一部分，以垂示将来”。“就新词新语的成长而言，至少分‘少儿期’‘青年期’和‘成年期’。少儿期的新词新语独立性弱，使用范围小，是具有语料性质的词语，能最迅速地反映社会政治、经济、文化各方面的变化，社会学、文化学的研究价值大。青年期的新词新语独立性强，使用范围大，最活跃，最敏感，很快进入共同语，文化学、语言学的实用价值大。成年期的新词新语是全民使用的基本词语，使用范围最广，构词能力最强，语言学的价值最大。”可以看出，新词语成长的不同时期，决定了新词语规范（或被广泛认可）的程度，有些新词语可能在“少儿期”就夭

折了,即所谓的昙花一现,还有部分所谓的“生造词”,即便是这些“词语”它毕竟是出现过。但我们也应该看到,在新词语这个大家族当中,这些成员毕竟不是在一个层面上,是有区别的,所以我们将利用语料库进行统计分析,将新词语分为三级。那些普遍认为规范的词语的,即我们收入《新词语大词典》《学生新词语词典》《汉英新词语学习词典》里面的词语可以定为一级词语。将那些正在“波动”中词语定为二级词语,比如2007年发布的2006年新词语“八荣八耻、白奴、白托、草根网民、车奴、倒扁、独二代、断背、法商、海缆断网、海绵路”等等;还有类似“酷装、酷言、酷臂、酷腿、酷手、酷脚、酷胸、酷臀、酷表达、酷家庭、酷保健、酷演讲、酷打扮、酷毙、酷机、酷语、酷头”等词语,以及一些比较被认可的网络词语,比如“QQ、E-book、DIY、美眉、聊天室、造砖、灌水、版主、恐龙、菜鸟、烘培机、爱虫”等,还有一些品牌名称等。将那些认为是不规范的词语定为三级词语,包括:不符合现代汉语规范化要求、表义含混或无任何“创新”的,如“白闯(白天闯入住宅行窃)、高买(在市场商店假装购物寻机盗窃商品)、搅理(胡搅蛮缠、强辞夺理)、敬戴(尊敬爱戴)”;没有必要的,如“老日(日元)、散装(很露的衣服)、大林(亲爱的,英语‘DARLING’的音译)、估揣(揣测、猜测)、鼓倡(宣传提倡、鼓吹)”;缺少普遍性的,如“吃豆腐、捣浆糊、同志(同性恋者)、波音的、苦瓜(因学习不好经常挨批的学生)、吃野食、大把进分(赚很多钱)、白菜(长得漂亮的女孩子)、蛋白质”;不稳定的,如“爱窝、爱资病、安死(安乐死)、摆揆丝、的高(舞蹈‘DISCO’音译词)、冬瓜(喻身体肥大)、抖骚、哈姆雷特(太高深)、换频道、寄托(美国GRE和TOFEL的合称)、拉蜜(勾引女人)、孔雀(自作多情的人)、抛海(挥霍)、看电影(洗澡)”等;还有网络上出现的“瘟都死(WINDOWS)、瘟酒吧(WIN98)、斑竹(版主)、老虫(点击器、‘MOUSE’的意译词)、Btw、JJ、GG、MM、7456、886、菌男(俊男)、霉女(美女)”等,以及不少的表情符号等。词语分级便于学习者区分新词语的规范性,学习掌握规范的词语;同时,也满足对不规范的词语的查检的需要。

3. 描写信息要全面

面向对外汉语教学的汉语新词语信息库主要目的是为对外汉语教学的服务的,这样就要尽量给学习者提供新词语系统完备的信息。因此,将继承《现代汉语新词语信息电子词典》的全部信息,此外,还要增加外文翻译、例句、关联语汇、词源信息等,这样新词语信息就相对完整,学习者能够比较全面地了解、掌握新词语。

4. 便于检索使用

新词语信息库收词量大、内容丰富,因此如何检索利用就成了一个很重要的问题,如果没有一个便捷、快速的检索工具,给使用带来一定的困难,那就大大降低了它的实用性。因此,我们拟采用数据的文件格式存储信息,可以进行有效地检索、分类、统计等。

四、新词语信息库的主要信息

新词语信息包括词语形、音、义以及语法属性等信息,本文主要介绍词性信息、例句信息、关联语汇信息和词源信息,其他信息参见文献《现代汉语新词语计量研究与应用》(中国社会科学出版社,2007)

1. 词性信息

《现代汉语词典》直到第五版才标注了词性,这是一个很大的进步。《现代汉语词典》

迟迟没有标注词性是有客观原因的。现代汉语词类的多功能性决定了为现代汉语常用词汇标注词性的复杂性。但新词的语法功能相对比较单一，标注词性的复杂度比较低一些。标注词性既可以帮助说明词义，又可以在一定程度上说明词的用法，所以，我们以目前中文信息处理和现代汉语教学中流行的语法体系的词类理论为指导，以词的“优势语法功能”为标准进行分类，以词语在大规模语料库中所表现出来的语法特征为根据具体归类，每一个词、多义词的每一个义项都标注了词性。

根据目前流行的中文信息处理和现代汉语教学语法体系，我们将词类分为名词、时间词、方位词、处所词、动词、形容词、区别词、状态词、副词、代词、数词、量词、介词、连词、助词、语气词、叹词、拟声词。而事实上，我们收录的词语中主要是名词、动词、形容词三类，区别词、状态词、副词都只有几十个到几百个，叹词两个“哇噻、yeah（吔）”，语气词1个“（吔）”。在这些词类中，区别词、状态词、副词的功能比较单一。但是名词、动词、形容词三类是多功能的，确定起来比较麻烦。好在新词的语法功能还比较单一，按照以上所说的“优势语法功能”可以比较准确地给每一个词标上词性。比如“另类”，我们采集到多数例子是出现在定语的位置上，由此看，“另类”应该定为区别词；但我们同时也注意到有“打入另类”“视为另类”这样的说法，而且意思也有细微差别，于是，我们把“另类”分为两个义项，前者为区别词，后者为名词。明显是短语的，不标词性。

2. 例句信息

“王了一先生几十年前曾说过，一本词典没有好例句就形同一具骷髅。例句作用之大，就在于它可以进一步具体地帮助读者理解词义，掌握词的用法。只有这样，才能给读者提供真正掌握这个词的条件，才能圆满地发挥词典的作用。”（李行健，2003）例句是词语存在的环境，环境有三个层面，即语法、语义、语用，需要在在这三个层面都具有典型性。即根据例句可以推出词义，根据例句可以看出词语的典型用法，根据例句可以看出词语使用的一般环境。只有满足这三个条件才算是一个典型的例句。（苏宝荣，2006）根据新词语的实际情况，我们确定如下选例原则：

（1）语义准确原则

所选例句能够准确地体现该词语的意义。因为词语进入句子之后，往往会因为上下文而发生意义的细微变化，所以在众多语料中要尽可能选择最能准确体现词语意义的例句。

（2）用法典型、全面原则

所选例句能够典型地体现词语的词类特征。汉语词类是多功能的，一个词类可以分布在多种句法位置，这就要求我们挑选例句时，首先要注意突出该词语的优势语法功能，同时也能兼及其他用法。具体做法是必须有一个例句能够体现其优势语法功能，另一个则选择体现其他用法的。比如前文提到“另类”，所采集的例句中，该词多出现在定语位置上，如果我们只选择出现在定语位置上的例句，那就突出的是区别词的特征而不是名词，所以，我们至少要选择一个该词出现在主宾语位置上的例句，这样该词语的用法（语法特征）就有了一个全面而突出的显示，便于学习者既全面而又有重点地了解掌握该词语的语法功能。

（3）时间原则

李行健（2003）指出：“新词语本来就是一种过渡性的词语，汉语中的‘葡萄、金刚、塔、站、革命、健康、逻辑、沙发’等都经过新词语阶段，现在不会有人认为它们是新词语了。也就是说，新词语进入汉语词汇（具有全民词语资格，方言性的、临时性的等等都

不能算)后,经过一段时间,能留下来并进入基本词语,它们也就不再是新词语了。是不是这样,诸如此类的问题都值得我们很好地研究。编写新词词典时,可否标明该词典收入的是什么年代到什么年代产生的新词语?这样也许更便于编者编纂和读者研究使用。”我们认为李先生的这个意见非常正确,每个新词语我们都特别注意考察了其使用时间范围,并把这个考察结果通过例句的选择表现出来。适当选取不同时期的例句,目的在于说明该词语使用的时间范围。从中也可以看到一个词语的兴衰。

(4)一词多例原则

郑锦全先生曾提出词汇学习的“一词多例”原则。为学习者提供多种多样的典型例句,学习者可以根据总结词语的意义及用法,可以补充词典的有关信息。按照这一思想,我们从语料库中抽取能够显示新词语各种属性信息的例句,为学习者提供一个恰当、丰富的语言环境,让学习者能够有效地掌握新词语。

3. 关联语汇信息

“类推”和“仿造”是新词语造词普遍使用的方法,随之出现大量的词群。设立“关联语汇”将这些同一词群中的词语串联起来,扩大了词典容量,具有连带的提示作用,方便读者使用。关联语汇当中收录的词语主要是1978年以来出现的新词语。关联语汇的处理办法如下:

(1)词群中共同构词语素单独出条的,在该条下把由该共同语素构成的词语全部列出。例如“吧”单独出条,则将由“吧”构成的两个词群的词语全部列全,如:冰吧—布吧—彩吧—餐吧—茶吧—迪吧—房吧—股吧—果吧—逛吧—话吧—街吧—酒吧—路吧—泡吧—琴吧—球吧—书吧—水吧—陶吧—玩吧—网吧—氧吧—浴吧—乐吧—纸吧—玻璃吧—彩妆吧—咖啡吧—木刻吧—玩具吧—文化吧—氧气吧—吧凳—吧弟—吧店—吧街—吧姐—吧客—吧娘—吧女—吧台—吧兄—吧椅—吧蝇—吧兄吧弟。

(2)包含在某个词群的词语中列出该词群的全部词语,比如在“美背”中列出“美鼻—美臀—美齿—美唇—美发—美甲—美肩—美睫—美颈—美眉—美脐—美容—美乳—美手—美体—美腿—美臀—美胸—美颜—美腰—美足”;在“空嫂”中列出“巴嫂—芭嫂—报嫂—菜嫂—的嫂—地嫂—房嫂—纺嫂—呼嫂—护嫂—警嫂—军嫂—商嫂—水嫂—铁嫂—物嫂—险嫂—月嫂—保安嫂—接送嫂—宴会嫂”等等。

(3)有些词语可能两个语素都有关联语汇,就将两个语素的关联语汇都列出。比如“款爷”中的两个语素都有关联语汇,这样就把两个语素的关联语汇都列出来,有:款哥—款姐—款娘—款女—款主—款族—班爷—板儿爷—板爷—包爷—导爷—倒儿爷—倒爷—的爷—卡爷—侃爷—揽爷—票爷,这样同一个词语将出现在两个关联语汇当中,可以扩大的词典的信息量,读者可以从两个关联当中解读该词语,理解得更加全面。

4. 词源信息

正如巢峰先生(2003)所言:“在治学之中,常常因一词之源,费尽周折。《辞源》要求‘结合书证,重在溯源’;《汉语大词典》要求‘古今兼收,源流并重’。可见语文辞书溯源的重要。”“词源就是知识,而辞书就是知识的结晶。重视辞书的知识性,尽可能交代新词语的出处,是编出出类拔萃的新词语辞典的重要条件。”“由于新词语产生的年代,一般就是编纂者生活的年代。因此,新词语溯源远易于古词语溯源。”“新词语一般都为编纂者所熟悉或容易找到根据。这项工作如果现在不做,若干年以后再做就难上加难。”同时,词

源信息对于学习者了解新词语产生的文化背景,掌握新词语的用法和意义以及通过新词语了解中国社会文化的发展都有积极意义。词源信息主要包括以下方面:

(1) 词语的理据,包括构词方式、语素义与词义的关系、语素意义的变化等。比如:

爱巢:“巢”本指“鸟窝”,很小但温暖。将相爱的或新婚的人的家比喻为“巢”,取其“小而温馨”之意。

吧蝇:苍蝇喜欢追逐腥臭,用来比喻沉溺于酒吧等各种娱乐场所的人,含贬义。

白魔(毒品):毒品就像魔鬼一样吞噬着人们的健康与生命,而且多数毒品都是白色的,故称,表现出人们对毒品的憎恨与恐惧。

大腕:“腕”就是“手腕”是人身体的一部分,用来代替人,以部分代整体,采用借代手法造词。

(2) 语源知识,包括外来词、方言词的来源介绍等。比如:

卡丁车:英文 carting 的音译,在英语中即小型赛车的意思。起源于美国,原本是农夫将农机设备改装而成的,用于娱乐。在英文原版里是小型单座汽车,后来指一项运动。

蔻:系日语转写形式“cute(可爱)”的音译,取“豆蔻年华”的“蔻”,起到了音译兼意译的作用,音义一致。

酷:系英文“cool(冷)”的音译。20世纪60年代开始成为美国青少年的街头流行语,初期指一种冷峻的、反主流的行为或态度,后来泛指时髦的、与众不同的。

炒鱿鱼:鱿鱼片儿炒熟后会卷成卷儿,样子像铺盖卷儿,所以“炒鱿鱼”喻指让员工卷铺盖卷儿走人,也就是解雇、开除。员工自动辞职,是炒老板的鱿鱼。

(3) 旧义和新义之间的联系。比如:

把脉:本义为切脉,是中医术语。这里的“脉”是指事物发展的趋势,新增比喻意义,取其调查、分析并对事物发展的趋势作出判断之意。

冲浪:原是一种水上运动项目,指运动员脚踏特制的冲浪板随海浪快速滑行,并保持身体的平衡。借指在网上浏览信息,一是喻指网络提供的海量信息,二是取冲浪的自由、爽心、随意等义。

充电:原为专业术语,指给电器补充电能,比喻产生新义,指通过学习增加能力,取增强能量义。

(4) 词语本身的相关文化象征含义。比如:

蓝道(出国留学之路):因为出国留学多半是飘洋过海的,即过去我们所说的“留洋”,取其有特征的颜色“蓝”而造出的词语。顺口溜:“从政之道红彤彤,从商之道金灿灿,蓝色之道明亮亮,升学之道黑沉沉”。

黑道(考学位、做学问):因为学士帽、硕士帽、博士帽都是黑色的,所以把考学位、做学问称为黑道。

红道(从政当官之道):由于我国政治往往与红色有关,所以将从政当官叫“红道”。

黄道(从商做生意之道):黄,指黄金,借指钱,从商做生意就是为了赚钱,故称。

补血(补充新物资或新成员):将新物资、新成员比喻为新鲜血液,预示着有好的发展,含褒义。

(5) 词语的专业知识背景介绍。比如:

考勤机:考勤机内置时钟,在线时与服务的时钟同步,显示刷卡人的姓名、留言、刷

卡时间、签到/签退状态等。有感应考勤机、IC卡考勤机、ID卡考勤机、网络考勤机等种类。

参考文献

- [1]巢峰. 新词语大词典·序一. 新词语大词典[M]. 上海: 上海辞书出版社, 2003: 1—6.
- [2]亢世勇. 现代汉语新词语计量研究与应用[M]. 北京: 中国社会科学出版社, 2007.
- [3]亢世勇、刘海润. 新词语大词典[Z]. 上海: 上海辞书出版社, 2003.
- [4]李建国. 新词新语研究与词典编纂[J]. 辞书研究, 1996(3).
- [5]李行健. 新词语大词典·序二. 新词语大词典[Z]. 上海: 上海辞书出版社, 2003: 7—13.
- [6]苏宝荣. 语境的多层面研究[J]. 中国语言学报. 2006(总12): 271—279.
- [7]周洪波. 新华新词语词典[Z]. 北京: 商务印书馆, 2003.