
基于语料库的现代汉语新词语动词语法特征的研究*

颜伟 亢世勇

烟台师范学院汉语言文学学院 山东 264025

Email:kangsy46@sohu.com

摘要: 本文以《现代汉语新词语信息电子词典》的已有成果为基础,以大规模真实语料的统计数据为依据,用统计数字来描述动词的各项语法属性,说明新词语动词的特点。

关键词: 新词语动词; 语法属性; 计量研究

0 引言

现代汉语新词语大量涌现,动词又是其中重要的一部分。对其进行整理研究,这无论是对现代汉语词汇的研究还是对中文信息处理都有积极的意义。

“现代汉语新词语信息电子词典动词库”是我们主持的国家社科规划项目“《现代汉语新词语信息电子词典》的开发与应用”的一部分。该库收录的主要是新词语动词,即主要是指具有新意义或新形式的动词。目前已收录动词 11413 个。本文选取了动词库中 3159 个词条建立了统计用库,根据这个统计用库来对新词语中动词语法信息进行初步的统计和分析,并与《现代汉语语法信息词典》的统计进行比较,说明新词语动词语法特征的特点。

1. 动词数据库的研究与实现

1.1 词库语法属性字段的设计和分析

《现代汉语新词语信息电子词典》是按照北京大学计算语言所的《现代汉语语法信息词典》的规格进行开发的,在新词语信息的描写上采用的是分类与属性描述相结合,以属性描述为主的方法。在粗分类的基础上进行语法属性的描述,属性描述尽可能详尽。动词库中共确定了 46 项属性,分别是同形、义项、粘着、系词、助动词、趋向动词、形式动词、自主动词、动态动词、准谓宾、有宾、前名、后名、介宾的后、外内、体谓准、双宾、兼语句、后动量词、后时量词、存现、动介、动结、动趋、不、没、很、着了过、在正在、重叠、AABB、V—V、V了V、V了一V、离合、VVO、复数主语、单作主语、单作谓语、单作宾语、单作状语、单作补语、情态词、谓词性主语、兼类、备注。

1.2、《现代汉语新词语信息电子词典动词子库》的结构和规模

“动词子库”采用成熟的关系数据库(在 Microsoft Access 环境下实现)文件格式详细描述词语及其语法属性的二维关系。设立属性字段 46 个,描写新词语动词 3159 个。

* 本研究得到国家社科规划项目(01CYY002)的支持

二、新词语动词语法信息的统计及与基本词动词的比较

2.1 新词语库中动词语法信息的统计与分析

2.1.1 动词的次范畴

动词统计用库中离合词有 219 个, 占 6.93%; 自主动词有 1343 个, 占 42.51%; 动态动词有 1717 个, 占 54.35%, 能带宾语的动词有 667 个, 占 21.11%;

2.1.2 句法功能

(1) 动词的组合特征

① 动词有无名词的特点

能作“有”的宾语的有 839 个, 占 26.56%; 能受名词直接修饰的有 234 个, 占 7.4%; 能直接修饰名词的有 749 个, 占 23.71%; 兼类词有 8 个, 占 0.25%;

② 动词同一些虚词的关系

能受“不”修饰的有 1318 个, 占 41.72%; 能受“没”修饰的有 2036 个, 占 64.45%; 能受“很”修饰的有 99 个, 占 3.13%; 能带“着、了、过”的有 2239 个, 占 70.88%, 其中能带“着”的或能带“了”的或能带“过”的分别有 426 个, 1272 个, 541 个, 分别占 13.49%, 40.27%, 17.13%, 只能带“着”或只能带“了”或只能带“过”分别有 118 个, 822 个, 98 个, 分别占 3.74%, 26.02%, 3.1%, 能同时带“着、了”或能同时带“了、过”或同时带“着、过”的分别有 170 个, 440 个, 156 个, 分别占 5.38%, 13.93%, 4.94%, 能同时带“着、了、过”的有 435 个, 占 13.77%; 能受“在、正在”修饰的有 1523 个, 占 48.21%; 能受“介宾的”修饰的有 467 个, 占 14.78%;

③ 动词与后继成分的关系

能后接结果补语的有 503 个, 占 15.92%; 能后带趋向补语的有 503 个, 占 15.92%; 能后接时量词的有 476 个, 占 15.07%; 能后接动量词有 737 个, 占 23.33%; 能带双宾语的有 28 个, 占 0.89%; 能组成兼语句的有 47 个, 占 1.49%; 后面能接介词的有 391 个, 占 12.38%, 其中能带“在”“往”“向”“于”“自”的分别有 170 个, 95 个, 56 个, 62 个, 8 个, 占 5.38%, 3.01%, 1.77%, 1.96%, 0.25%; 及物动词(外动词)所带的宾语当中体词性的有 660 个, 占 20.89%, 谓词性的有 22 个, 占 0.69%, 准宾语有 3 个, 占 0.09%, 能同时带体谓准有 9 个, 占 0.19%;

(2) 动词的造句功能

能单作主语的有 1042 个, 占 32.99%; 能单作谓语的有 734 个, 占 24%; 能单作宾语的有 987 个, 占 31.24%; 能单作状语的有 157 个, 占 4.97%; 能单作补语的有 4 个, 占 0.12%; 能作准谓宾的有 667 个, 占 21.11%; 能组成存现句的有 14 个, 占 0.44%; 要求复数主语的有 52 个, 占 1.65%;

2.1.3 形态特征

能重叠(“VV”或“ABAB”)的有 314 个, 占 9.94%, 其中能以“VV”形式重叠的有 7 个, 占 0.22%; 能以“ABAB”形式重叠的有 307 个, 占 9.72%; 能以“AABB”形式重叠的有 14 个, 占 0.44%; 能以“V—V”形式重叠的有 7 个, 占 0.22%; 能以“V了V”形式重叠的有 7 个, 占 0.22%; 能以“V了一V”形式重叠的有 7 个, 占 0.22%; 能以“VVO”形式重叠的有 7 个, 占 0.22%;

数据表明: 新词语动词的优势语法功能主要体现在动作性较强、能受“没”的修饰和能带动态助词“着、了、过”几个方面。当然其中任何一种都不能囊括所有的词。通过统计, 我们发现: 有 617 个动词动作性比较差、不能受“没”的修饰、不能带“着、了、过”而且大部分不能带“动时量”补语。这些动词中有 141 个通过接受“在、正在”的修饰保证了动词身份, 有 34 个动词通过进入祈使句保证了其动词身份, 有 56 个动词通过接受“不”的修饰保证了其动词身份, 有 31 个通过单独作谓语保证了其动词身份, 只有 4 个动词通过带宾语保证了其动词的身份, 其余的动词几乎不能重叠, 大部分是可以作主语或宾语的词。虽然库中有些动词基本不体现上述优势语法功能, 但它们有一个共同的特征即作谓语。

2.2 基本词语动词*与新词语动词语法特征的比较

2.2.1 动词次范畴信息的比较

基本词动词中离合词有 2055 个, 占 19.98%; 新词语中离合词的比例较基本词减少了近三分之二。新词语动词离合词数量大大减少了。

基本词中能带宾语的有 6148 个, 占了 59.78%。新词语中自足性动词(不带真宾语的动词)占据了主要地位, 而及物动词退居了“二线”, 能带真宾语由动词的优势语法功能退化为劣势语法功能, 比例下降了一半多。

2.2.2 句法信息的比较

句法信息是新词语和基本词语差异的主要方面, 体现出了新词语的发展和变化。

从组合特征上看: ①新词语动词的名词性大大加强, 主要表现在“有宾”动词大量涌现, 基本词动词中能作“有”的宾语的有 341 个, 占 3.32%, 而新词语中有宾动词的比例增长到近 27%, 变化十分显著; 基本词动词中能受名词直接修饰的有 632 个, 占 6.51%; 能直接修饰名词的有 3011 个, 占 29.28%, 而新词语动词的比例则有所上升; 同时兼类词的数量也增加了许多。②能受一些特征词的修饰仍是新词语动词的优势语法功能。基本词动词中能受“不”修饰的有 7585 个, 占 73.76%; 能受“没”修饰的有 9116 个, 占 88.65%; 能受“很”修饰的有 248 个, 占 2.41%; 能带“着、了、过”的有 8507 个, 占 82.73%; 能被“在、正在”修饰的有 4353 个, 占 42.33%。动词的这些优势语法功能在新词语动词中比例虽有所下降但仍是新词语动词最优势的语法功能。③新词语动词后接的成分呈现出蜕化趋势, 除双宾语的数量有所增加之外(基本词动词中能构成双宾句的有 161 个, 占 0.16%), 其他的后接成分的比例都有很大的下降。基本词动词中能构成兼语句的有 334 个, 占 3.25%; 能后带体谓准的动词有 6418 个, 占 59.78%; 后跟结果补语的有 2306 个, 占 22.43%; 后跟趋向补语的有 4402 个, 占 42.18%; 后接时量的有 4221 个, 占 41.05%; 后接动量的有 4336 个, 占 42.17%。新词语中能组成兼语句的动词的数量下降了一半多, 能带体谓准宾语的数量(即及物动词的数量)下降了近三分之二, 而能带结果补语和趋向补语下降率分别为 29.02%, 62.25%; 能带时量补语和动量补语的下降率分别为 63.28%, 44.68%。

从造句功能上看: ①新词语动词也有明显的变化。基本词动词中能单作主语的有 2009 个, 占 19.54%; 能单作谓语的有 5214 个, 占 50.71%; 能单作宾语的有 4118 个, 占 40.05%; 能作准谓宾的有 2119 个, 占 20.61%; 能单作状语的有 558 个, 占 5.52%; 能单作补语的有 133 个, 占 1.29%。新词语与基本词语动词在单作状语、宾语和作准谓宾方面没有明显的变化, 而在单作主语和单作谓语方面却发生了巨大的变化。“单作谓语”这一基本词动词的优势语法功能在新词语中变为弱势功能, 其比例由基本词动词的 50.71%降为新词语动词的 23.24%, 整整“损失”了一半多; 而与此同时, 作为劣势语法功能的“单作主语”却由基本词中的 19.54%上升到 32.99%, 增长率为 68.83%。此外, 单作补语的比例也有较大下降。②新词语动词的其他特征正逐渐减少, 语法属性较单一化。基本词动词中要求复数主语的有 340 个, 占 3.31%; 能构成存现句的有 221 个, 占 2.15%; 可以有谓词性主语的有 120 个, 占 1.17%。新词语动词前面必须出现复数主语的比例下降了 50.15%, 组成存现句的动词的比例下降了 79.53%, 能带谓词性主语的动词在新词语动词中中没有。

2.2.3 形态信息的比较

新词语动词中双音节动词的重叠形式基本没变或稍有所增加, 而单音节动词的重叠形式大大简化。基本词动词中能以“VV”、“ABAB”、“AABB”形式重叠的分别有 569 个, 746 个, 45 个, 占 5.53%、7.25%、0.44%; 能以“V—V”重叠的有 589 个, 占 5.73%; 能“V了V”的有 425 个, 占 4.13%; 能“V了一V”的有 399 个, 占 3.88%。而新词语动词库中双音节动词的重叠形式“ABAB”和“AABB”分别有 307 个, 14 个, 占 9.72%, 0.44%, 其中“ABAB”式增长率为 34.06%, “AABB”式与基本词持

*“基本词语动词”即指《现代汉语语法信息词典》收录的动词, “新词语动词”即指《现代汉语新词语信息电子词典》中收录的动词。

平：单音节动词的重叠形式“VV”“V—V”“V了V”“V了一V”都只有7个，占0.22%，这较基本词动词有了巨大的变化，其减少率为96.02%，96.16%，94.67%，94.33%。根据对《新词语信息电子词典》的统计（亢世勇，2001），整个新词语库中单音节动词只有176个，占0.44%，双音节动词有17251个，占44.62%。新词语中双音节词占了绝对优势，同时整个新词语的发展有多音节化趋势，所以，上述现象比较容易理解。

3. 现代汉语新词语动词语法特征的特点

3.1 新词语动词的优势语法功能

新词语动词的优势语法功能主要体现在动作性较强、能受“没”的修饰和能带动态助词“着、了、过”，这一点体现了新词语动词对基本词动词的承继。同时，这也是新词语动词词类归属判定的主要标准。

现在的词类划分工作基本上是以经典的范畴化理论为逻辑背景的。通常认为某一类词类的全体成员一定共有或没有某一分布特征，并且据此来给不同的词类下定义；往往希望一类词与另一类词有明确的界限，并且默认同一类词的地位相等，没有典型成员和非典型成员之别。这种追求简单、整齐的做法在实际的词类划分工作中往往行不通，因为词类不完全是特征范畴。我们同意词类是一种原型范畴，是人们根据词与词之间在语法性质上的种种相似性而概括出来的（袁毓林，1995）。原型论通过某一词类的典型成员来给整个词类下定义。这种方法既简便又直观，比较切近人们对于词类的语感。这种定义方式既反映了词类作为一种原型范畴的一些本质特点，又照顾到词类之间边界模糊的特点。按照原型论观点，有较强动作性，能受“没”的修饰，能带“着、了、过”的动词是动词中的典型成员。统计数字表明：动词中有2542个典型成员，占80.47%。其实，正是这80%左右的典型动词奠定了新词语动词词类的总体框架，也决定新词语动词词类的总体语法特征。

3.2 新词语动词的语法属性有单纯化倾向

基本词中动词的各种语法属性虽有优势和劣势之分，但其不平衡性的差异是有限的，而新词语中动词的语法属性却有畸轻和畸重的现象。新词语动词中动态动词和自主动词分别占54.35%，42.51%，而其他的如系词、助动词、趋向动词等在统计用库的3159个词条当中却没有收录。形态信息中“ABAB”式和“AABB”式重叠比较丰富而单音节动词的“VV”“V—V”“V了V”“V了一V”重叠却只占了极小的份额。

3.3 新词语动词的系列化倾向比较明显

汉字是音节文字，大多数情况下是一个字一个音节，一个语素，只要两个音节之间语义上可以搭配，事理上能讲得通，符合语言结构习惯，通过仿造就可以构成系列化的双音节新词。而这种搭配因构造上的“血缘”关系，其对某种语法属性往往具有相同的排斥性或吸引力，即它们要么都具有某种语法属性要么都不具有某种语法属性，具有很强的系列化倾向。新词语动词数据库中“以XVX”方式（如“以公养私”）造出73个动语，其中所有的都受“不”的修饰，可单作主语、谓语和宾语，93.15%的可受“在、正在”的修饰，而其它的语法属性则都不具有了。以“研X”方式（如“研审”）造出19条“近亲”动词，其中全部19条词语都是自主动词、动态动词，能作有宾，能重叠，能受“不、没、着、了、过、在、正在”修饰，有21.05%可作准谓宾，73.69%可放在“介宾的”后，43.37%可后加动量词，21.05%可后加时量词，57.89%可带结果宾语，78.95%可带趋向补语，但所有19个词条都不能组成兼语句、双宾句和存现句，不受“很”修饰，不带介词，不单作状语和补语。

3.4 新词语动词名词性增强动词性减弱

基本词中的名动词具有名词的功能，它们在动词中所占的比例也较大（约 39%）（俞士汶等，1998）。而新词语动词的一个较大的特点就是名词性加强动词性减弱。这表现在①“有宾”的比重显著上升，由原来的 3.32% 上升到 26.56%，比例是基本词动词的八倍多。同时能受名词修饰的动词和能修饰名词的动词也占了一定的比例。②能单作主语的动词的比例显著上升，由 19.54% 上升到 32.99%，上升率为 68.83%。同时单作谓语的动词的比例大大下降，由 50.71% 下降为 23.24%，下降率为 54.17%。③形态中单音节动词的重叠形式“VV”“V—V”“V了V”“V了一V”的比例全面下降。

4. 余论

面对现今各种新词语词典实际应用价值不高的现状，电子词典为我们对新词语进行定量的、穷尽性的全面的研究提供了可能。但如何把这种可能变成实际还需要我们积极地探索和思考。

数据库的统计显示：动词从发展的绝对速度来看是最快的。新词语的产生是和社会的发展变化密不可分的。这其中的原因究竟在哪里？这值得我们思考。

对于词汇，我们应该承认其是发展的而非静止的，对于词汇的发展，新词语的出现，我们应该加以引导和规范而非禁止。新词新语的整理，将是一个长期的过程。

参考文献

- 俞士汶、朱学锋等：《现代汉语语法信息词典详解》，清华大学出版社，1998。
亢世勇：《汉语数据库建设及其应用》，作家出版社，2000。
陈原：《新词语》，语文出版社，2000。
陈建民：《汉语新词语与社会生活》，语文出版社，2000。
于根元：《二十世纪的中国语言应用研究》，书海出版社，2000。
朱德熙：《语法讲义》，商务印书馆，2000。
亢世勇：基于语料库的现代汉语新词语研究，第 34 届汉藏语会议论文，2001。
俞士汶、段慧明、朱学锋：汉语词的概率语法属性描述，《应用语言学》，2001（11）。
王铁昆：新词语的判定标准与新词新语词典编纂的原则，《语言文字应用》1992（4）。
王铁昆：简评《1991 汉语新词语》，《语言文字应用》1992（4）。
袁毓林：词类范畴的家族相似性，《中国社会科学》1995（1）。

作者简介：颜伟（1979—），男，山东临沂人，硕士研究生，主要研究领域为汉语信息处理。亢世勇（1964—），男，陕西延安人，硕士，教授，硕士生导师，主要研究领域为汉语信息处理和计算语言学。

Researches on The New Verbs of Contemporary Chinese Based on The Corpus

Yan Wei Kang Shiyong

Yantai normal university shandong 204625 china

Email:kangsy46@sohu.com

Abstract: Based on the New Words And Phrases Dictionary,the grammatical characteristics of the new verbs will be described with statistical data acquired from the corpus . This article introduces the new project , some results and further trend .

Keywords:the new words and phrases verbs; grammatical attribute probable; grammatical attribute description