基于数据库的新造词语的构词法研究*

徐艳华　亢世勇

烟台师范学院汉语言文学院 （264025）

kangsy46@sohu.com

摘要：我们利用成熟的关系数据库，构造了新造词语构词法信息库，在此基础上进行分类归纳统计，总结了新造词语的构词规律，为中文信息处理未登录词的识别提供了一个基本依据。

关键词：新造词 ； 构词方式； 中文信息处理

新造词语是“汉语利用原有的构词材料，按照固有的构词方式构造而成的表示新事物、新概念的新的词语形式”。[①]我们按照①适用性原则②规范性原则③明确性原则④习惯性原则⑤稳定性原则收录新造词语3万多个，利用成熟的关系数据库描述了新造词语和其构词法信息库的二维关系，建立了新造词语构词法信息库，在此基础上对新造词语构词法进行了研究、总结，归纳了构词的规则，相信这样的成果，能够为计算机未登录词的识别提供一个基础依据。限于篇幅，本文只讨论二音节、三音节词语。

1．新造词语构词法信息库的研究与实现

1．1 新造词属性的确立

新造词构词法信息库旨在研究用何种构词材料以何种方式组合构成何种词，从而总结新造词语的构词规律，为了满足这一方面的需要，在确立属性时主要考虑了以下几个方面：

（1）构词部件。词的构造成分，有语素、词、短语三种成分。新词语，双音节词居多，三音节、四音节次之，四音节以上的较少。构词部件设立了三个，分别称为构件1、构件2、构件3。

（2）构词方式。主要有主谓、动宾、联合、偏正、补充、前接、后加等形式。

（3）词性。标注该词语的词性。

（4）音节。标注各词语的音节数，即双音节词标“2”，三音节的标“3”，四音节词标“4”。

（5）单纯词。标注新造词中的单纯词。

1．2 构词法信息库的构件类型、词类体系及相关标记

本库的词性标注采用的是信息处理用现代汉语词类体系及标记集的研究成果，其如下：名词 n，动词 v，形容词 a，时间词 t，处所词 s，方位词 f，区别词 b，副词 d，状态词 z，代词 r，数词 m，量词 q，叹词 e，拟声词 o，介词 p，连词，c 助词 u，语气词 y。以上是基本词类的标记，此外还有：词素 g，前接成分 h，后接成分 k，简称略语 j，人名 nr，地名 ns。有些构件是短语，按短语结构标注为：介词结构 pp，定中结构 dp，状中结构 zp，联合结构 lp，动宾结构 bp，补充结构 cp 主谓结构 wp，数量结构 mq。新造词的构词方式及标记为：联合式 L，动宾式 B，定中式 D，状中式 Z，主谓式 W，补充式 C，附加式（前缀 Q，后缀 H）。

2．双音节词的构词方式

* 本项研究得到国家社科规划项目（01CYY002）的支持

双音节词共15719个，占新造词的51.7%。其中：

2．1 状中式（Z）

状中式（Z）的有3378个，占双音节词语的21.5%，其中：Vg+Vg→V（共1838个，占63.2%）如：俯看、勾写、混居。Ag+Vg→V（共774个，占23.6%）如：广交、冷冻、恶斗。R+Vg→V（共473个，占14.6%）如：自慰、自负、本销。Ng+Vg→V（共229个，占6.8%）如：海聊、面授、火葬 。Dg+Vg→V（共118个，占3.5%）如：复建、互保、重塑。B+Vg→V（共24个，占0.7%）如：负增、单恋、反拍。F+Vg→V（共24个，占0.7%）如：内保、南下、中转。Tg+Vg→V（共13个，占0.4%）如：晨练、春训、冬钓。M+Vg→V（共7个，占0.2%）如：半退、二审、初评。Q+Vg→V（共7个，占0.2%）如：倍感、双控、条播。

双音节词语以状中式构成的新造词都是动词。在这10种结构模式中，Vg+Vg，Ag+Vg，R+Vg，Ng+Vg，Dg+Vg占优势，其他几类比例甚少。

2．2 定中式（D）

定中式的有5956个，占双音节词语的31.9%，其中：Ag+Ng（共1477个，占24.8%）其中：Ag+Ng→N （99.5%）如：钝木、高墙、大盘；Ag+Ng→A（0.4% ）如：准时、红火、长足；Ag+Ng→D（0.1%）如：全力。M+Ng→N（共59个，占1%） 如：两岸、零料、首例。Vg+Ng→N（共869个，占14.6%）如：堕力、飞弹、呼机。Ng+Ng→N（共3378个，占56.7%）如：饭局、草婚、光卡 。B+Ng→N（共83个，占1.4%） 如：公房、女桥、主课。F+Ng→N（共11个，占0.3%） 如：东风、内政、左祸。J+Ng→N（共6个，占0.1%）如：黔剧、京味儿、美商。Ng+Q→N（共3个，占0.06%）如：车次、车吨、贝堆。Tg+Ng→N（共8个，占0.13%）如：晨脉、秋葵、午市。X+Ng→N（共7个，占0.12%）如：鸳盟、檬果、芭星。R+Ng→N（共12个，占0.2%）如：本岗、另类、啥时。M+Q→N（共8个，占0.13%）如：两打、一方。Q+Q→Q（共1个，占0.02%）如：人次。Nr+Ng→N（共6个，占0.1%）如：莎剧、苗家、魏稻。Tg+Vg→N（共3个，占0.06%）如：晨泳、冬泳、晨祷。Q+Ng→N（共11个，占0.3%）如：套菜、条律、听啤。

定中结构的双音节词名词占绝对优势，Vg+Ng，Ng+Ng这两类结构模式是定中结构的主要模式。

2．3 动宾式（B）

动宾式的有3332个，占双音节词语的21.2%，其中：Vg+Ng（共2436个，占73.13%）→V（共2428个，占99.3%）如：堕胎、翻天、服气；→A（共18个，占0.7%）如：绝情、耐苦、走运。Vg+Vg→V（共667个，占20.04%）如：躲生、防爆、助学 。Vg+Ag→V （共213个，占6.4%）如：发沉、告急、淘劣。Vg+F→V（共6个，占0.2%）如：支内、仇外、跑外。Vg+Q→V（共3个，占0.1%）如：解捆、掉份、开打。Vg+B→V（共3个，占0.1%）如：盗公、转公。Vg+M→V（共1个，占0.03%）如：刮三。

动宾式结构构成的新造词，动词居多，以Vg+Ng，Vg+Vg，Vg+Ag三种结构模式为主。

2．4 联合式（L）

联合式的有1998个，占双音节词语的12.73%，其中：Vg+Vg→V（共1285个，占64.3%）如：罚扣、封堵、观眺。Ag+Ag→A（共531个，占26.6%）如：皓洁、恭虔、嘈吵。Ng+Ng→N（共171个，占8.6%）如：伴侣、棱角、始末。Q+Q→N（共10个，占0.38%）如：轮回、场次、档次。Fg+Fg→N（共1个，占0.05%）如：南北。Tg+Tg→N（共1个，占0.05%）如：初始。

以联合式构成的新造词，动词最多，占64.3%，形容词次之，占26.6%，名词只占9.1%。

2．4 主谓式（W）

主谓式的有211个，占双音节词的1.3%，其中：Ng+Vg（共134个，占63.7%）→V（共117个，占88%）如：耳读、机写、盘升；→N（共15个，占11.3%）如：果冻；→A（共2个，占0.7%）如：火爆。R+Vg（共44个，占21.1%）→V（共43个，占97.8%）如：自测、自保、他杀；→A（共1个，占2.2%）如：自由。J+Vg→V（共2个，占0.9%）如：美援。F+Ag→A（共2个，占0.9%）如：内涝。B+Vg→V（共2个，占0.9%）如：阳痿。Vg+Ag→A（共2个，占0.9%）如：喘促。R+Ag→A（共5个，占2.7%）如：自愧、自傲。Ng+Ag（共20个，占8.9%）→A（共15个，占73.7%）如：资深、丛密；→N（共5个，占26.3%）如：口红。

主谓式的新造词，动词占优势，以Ng+Vg，R+Vg，Ng+Ag三种结构模式为主。

2．5 补充式（C）

补充式的有341个，占双音节词的2.1%。其中：Vg+Ag→V（共120个，占35.2%）如：扶正、搞臭、走红。Vg+Vg→V（共209个，占61.3%）如：点穿、卡死、打翻。Vg+F→V（共1个，占0.3%）如：落后。Vg+U→V（共3个，占1%）如：爬着、掰了。Vg+P→V（共4个，占1.1%）如：朝向、位于。Ag+Vg→A（共4个，占1.1%）如：凉透、破露、雄绝。

补充式的新造词，动词占98.9%，形容词占1.1%，动词占绝对优势，以Vg+Vg，Vg+Ag两种结构模式为主。

2．6 加前缀（Q）

加前缀的共57个，占0.4%。其中构成名词的共48个，占84.2%，如：阿飞、老抠、老板；构成形容词的共6个，占10.5%，如：老辣、老旧；构成动词的共3个，占5.3%，如：从严、从宽。

2．7 加后缀（H）

加后缀的共453个，占2.9%。其中构成动词的共59个，占13%）如：钝化、黄化、火化；构成名词的共373个，占82.2%，如：对子、剑手、警坛；构成形容词的共18个，占4%，如：煌然、紧巴、悸然；构成区别词的共3个，占0.7%，如：恶性、硬性。

从以上描述可见：（1）新造词语构词方式，定中式最多，加前缀最少，依次是定中式、状中式、动宾式、联合式、加后缀、补充式、主谓式、加前缀。（2）状中式、动宾式、补充式、主谓式构成的主要是动词，定中式构成的主要是名词，联合式由其联合的构件类型而定。

3．三音节词的构词方式

三音节词语共6502个，占新造词语的21.4%。其中：

3．1 定中式（D）

定中式的有4206个，占三音节词语的64.7%。其中：Ng+V→N（共21个，占0.5%）如：性侵犯、鱼采购、核扩散。PP+Ng→N（共2个，占0.05%）如：对台戏、在室女。N+Ng→N（共1800个，占42.8%）如：风筝城、国情车、核垃圾。Ag+N→N（共753个，占17.9%）如：多层次、全方位、方便菜。DP+Ng→N（共105个，占2.5%）如：白褶裙、大篷车、白眼病。Bg+Ng→N（共72个，占1.7%）如：副班长、女强人、自动伞。V+Ng→N（共845个，占20.1%）如：发射场、分手饭、按摩袜。BP+Ng→N（共286个，占6.8%）如：翻身仗、防寒服、扶贫款。ZP+Ng→N（共101个，占2.4%）如：公休日、公用筷、优生法。LP+Ng→N（共29个，占0.7%）如：港台剧、推拉车、吃喝风。NS+Ng→N（共38个，占0.9%）如：中国风、欧洲军、印度风。WP+Ng→N（共29个，占0.7%）如：自留山、人行道、手抛

道。Ag+NR→ N（共2个，占0.05%） 如：活雷锋、活鲁班。F+Ng→ N（共6个，占0.15%）如：地下军、掌上机、炉前工。M+N→ N（共46个，占1.1%） 如：半成品、二老外、零缺陷。Ag+DP→ N（共13个，占0.3%） 如：红大院、长防林、新税制。Ng+DP→ N（共4个，占0.1%） 如：核废料、车月票、泵排量。NR+ Ng→ N（共16个，占0.4%） 如：琼瑶迷、雷锋卡、江青裙。MQ+ Ng→ N（共32个，占0.8%）如：二次文献、二号文件、五年计划。

定中式的三音节新造词全部是名词，在构词模式上，以Ng+Ng ，Vg+Ng ，Ag+Ng 三类为主。

3．2 状中式（D）

状中式的有190个，占三音节词的2.9%。其中：Ag+Vg → V（共72个，占38.4%）如：活读书、假出口、粗加工。Ag+Ag → A（共8个，占4.7%）如：穷大方、全自动、富小气。PP+Vg → V（共7个，占3.7%）如：朝钱看、往外挤、向右转。Fg+Vg → V（共3个，占1.6%）如：锅下愁、场外招、雪上飞。ZP+Vg → V（1个，占0.5%）如：过劳死。BP+Vg （共11个，占5.8%）→V（共10个，占99.9%）如：没戏唱、回头看；→N（共1个，占0.1%）如：随身听。Vg+BP → V（共2个，占1.1%）如：会来事、伙种地。Ag+BP→V（共5个，占2.8%）如：假夺权、大换肩。Mg+Vg → V（共6个，占3.3%）如：半跳槽、二进宫、半残废。Dg+BP → V（共8个，占4.2%）如：不起眼、不走样、不称霸。Dg+Vg → V（共9个，占4.7%）如：不介入、不起诉、再教育。Vg+Vg → V（共22个，占11.6%）如：倒接班、倒发奖、热身唱。Ng+Vg → V（共22个，占11.6%）如：电捕鱼、双肩挑、口头纠。O+Vg→ V（共2个，占1.1%）如：乓乓响、碰碰响。MQ+Vg→V（共3个，占1.6%）如：一刀砍、一刀切、一日游。Bg+Vg →V（共1个，占0.5%） 如：总动员。Ag+U+Vg→V（共2个，占1.1%）如：黑着干。Vg+U+Vg→V（共2个，占1.1%）如：走着瞧、对着干。Tg+Vg → V（共1个，占0.5%）如：生前葬。WP+Ag → A（共1个，占0.5%）如：自来红。CP+Vg→V（共1个，占0.5%）如：绑紧跳。D+Ag→ A（共1个，占0.5%） 如：共同美。BP+Ag →A(共3个，1.6%) 如：上场慌、上场昏、上场怯。

状中式的三音节词，以动词居多，形容词甚少，在结构模式上以Ag+Vg，Vg+Vg，Ng+Vg为主。

3．3 动宾式（B）

动宾式的有693个，占三音节词的10.7%。其中：Vg+DP → V（共68个，占9.8%） 如：翻老帐、赶热浪、走弯路。Vg+Ng →V（共575个，占83.1%）如：够意思、打电话、点码子。Vg+Ag → V（共4个，占0.6%）如：反腐败、够慈气、玩深沉。Vg+Vg → V（共12个，占1.8%）如：反分工、拉赞助、打埋伏。Vg+ZP → V（共4个，占0.6%）如：反冒进、放单飞、吃大富。Vg+MQ→V（共4个，占0.6%）如：顾一头、翻一番、送一程。Vg+U+Ng→V（共3个，占0.4%）如：挂了帅、乱了套、揭了壳。Vg+BP→V（共2个，占0.3%）如：反跳槽。Vg+J→V（共3个，占0.4%）如：除六害。CP+Ng→V（共9个，占1.3%）如：拖下水、拉下马、吃错药。Vg+Tg→V（共2个，占0.3%）如：赌明天、抢农时。Vg+Ag+U→V（共1个，占0.15%）如：说白了。ZP+Ng→V（共4个，占0.6%）如：不信邪、不懂电、不进鳞。Vg+Q+Ng→V（共1个，占0.15%）如：打把伞。

以动宾方式构成的三音节词全部是动词，在构词模式中，Vg+Ng 占绝对优势，Vg+DP 次之。

3．4 主谓式（W）

主谓式的有104个，占三音节词的1.6%。其中：Vg+Ag →A（共3个，占2.9%）如：抓药难、劳动美、开门红。Ng+Ag →A（共32个，占30.8%） 如：根子正、性开放、生活美。Ng+Vg+Ng→V（共31个，占29.8%）如：房改房、手拉手、利改税。Ng+ZP →A（共4个，占3.8%）如：核聚变、情未了、心太软。Ng+Vg （共24个，占23.1%）→V（共16个，占66.7%）如：肝儿颤；→N（共8个，占33.3%）如：同窗恋、肠梗阻。R+Vg → V（共1个，占1%）如：大家拿。Bg+Vg+Bg→V（共1个，占1.9%）如：单改双、专转本。Fg+Vg+Fg→V（共1个，占1.9%）如：内转外。Mg+Vg+Mg→V（共1个，占1.9%）如：二

过一。Mg+Ng+Vg→V（共1个，占1.9%）如：两手抓。Mg+Ng+Ag→A（共2个，占3.8%）如：一头热、一头沉。

该类型动词最多，形容词次之，名词最少，以Ng+Ag，Ng+Vg+Ng，Ng+Vg三种结构模式为主。

3．5 并列式（L）

并列式的有50个，占三音节词的0.8%。其中：Ag+Ag+Ag→A(共19个，占39.6%)如：高精尖、快准狠、脏乱差。Vg+Vg+Vg→V(共9个，占18.8%)如：产运销。R+R+R→R(共1个，占2.1%)如：你我他。Ng+Ng+Ng →N(共8个，占16.7%)如：山江湖、责权利、人财物。Ag+C+Ag→A(共5个，占8.3%)如：威而刚、少而精、大而全。Ng+Ng →N(共7个，占9.4%) 如：葱韭菜、工副业。Vg+Vg → V(共3个，占3.2%)如：离退休、玩儿闹。

以并列方式构成的新词中，形容词最多，占总数的47.9%，动词次之，占20.9%再者是名词，还有个别的代词。

3．6 补充式（C）

补充式的有47个，占三音节词的0.7%。其中：Vg+U+Ag（共11个，占23.8%）→V（共10个，占90%）如：过得硬、黑得好；→N（共1个，占10%）如：热得快。Vg+Dg+Ag→V(共3个，占7.1%)如：搞不通、摆不平、说不准。Vg+U+Vg→V(共14个，占31.0%)如：信得过、谈得拢、玩得转。Vg+LP →V(共2个，占2.4%)如：洗洁净。Vg+Vg →V(共10个，占19.1%) 如：热昏头、逗咳嗽、站起来。Vg+Ag→V(共3个，占7.1%) 如：说清楚、气不公。Vg+ZP →V(共4个，占9.5%)如：谈不拢、搞不通、吃不开。

补充式的三音节词动词占98.7%，名词只占1.3%；结构模式以Vg+U+Vg，Vg+U+Ag， Vg+Vg为主。

3．7 加前缀

加前缀的共29个，占0.4%；其中构成名词的26个，占89.7%，如：类继父、非集团、超高温；构成动词的共3个，占10.3%，如：负反馈、负建设、洋冒进；

3．8 加后缀

加后缀的共1183个，占18.2%。其中构成名词的共1087个，占91.9%，如：敦煌学、方法论、追星族；构成动词的共96个，占8.1%，如：正规化、国际化、轨道化。

由以上描述可见：（1）定中式最多，加前缀最少，依次是定中式、加后缀、动宾式、状中式、主谓式、联合式、补充式、加前缀。（2）定中式构成的全部是名词，状中、动宾、主谓、补充构成的全部是动词，联合式构成的形容词最多。

尽管新造词的形式更加多样化，但其构词方法与基本词汇一致，基本没有新的突破。其中偏正式是能产性最强的一类。不论是双音节词，还是三音节、四音节，偏正式所占比例最大，其次动宾式和联合式的能产性也较强。在三音节中，附加式中的前缀式所占比例大于动宾式和联合式，可见，前缀式在三音节词中占优势。

附注：

① 徐国庆《现代汉语词汇系统论》，北京大学出版社，1999年4月第50页

参考文献：

①亢世勇《汉语数据库建设及其应用》，作家出版社，2000年9月

②徐国庆《现代汉语词汇系统论》，北京大学出版社，1999年4月

③葛本仪《汉语词汇论》，山东大学出版社，1997年9月

④周荐《汉语词汇研究史纲》，语文出版社，1995年

⑤王德春《汉语新词语的社会文化背景》，世界汉语教学，1990 年第三期
⑥宋玉柱《现代汉语语法十讲》，南开大学出版社，1986 年 3 月
致谢：本文研究对象——新词语来自于国家社科规划项目“《现代汉语新词语信息电子词典》的开发与应用”所开发的电子词典当中，构词法信息库的标注，慕亚芹同学做了不少工作，谨表谢忱。
作者简介：徐艳华，（1976—），女，山东烟台人，硕士生，主要研究领域为汉语信息处理、计算语言学；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、计算语言学。

Researches on word-formation of New word Based on The Corpus

Xuyanhua kangshiyong

Yantai normal university shandong 204625 china

Email :kangsy46@sohu.com

Abstract: With the help of the perfect relation data base, we build the word-formation message data base of newly-created words, on which we have the induction and statistics on different aspects to master the word-formation rules of newly-created words in order to provide scientific basis for the recognization of words which are not logged with the Chinese processing.

Key word: newly-created word word-formation pattern word-formation feature