

[文章编号] 1671-881X(2010)04-0471-06

三字词的词法模式研究

曾立英

[摘要] 对《现代汉语语法信息词典》数据库的统计发现,三字词的构词具有一些比较固定的词法模式,该模式中那些能产性强的语素在构词时形成了一种“类义功能”,在语法结构上也可以单独切分,称之为“部件词”。“部件词”包括26个接头词,185个接尾词,考察部件词构词时的词性、义项、词频等,有利于进一步分析部件词意义的聚类。

[关键词] 三字词;部件词;词法模式;类义功能

[中图分类号] H109.4 [文献标识码] A

一、三字词构词的词法模式

我们对《现代汉语语法信息词典》数据库^①中8万多个词语进行了详细的考察,并按照“词语正序”和“词语逆序”两种方式进行统计,发现在以“词语逆序”的方式构成的词库中,词语的后字构词有一些很有规律的现象。如有一些语素经常出现在词尾,形成了一些比较凝固的词法模式,如“XY人、XY法、XY热、XY期、XY费、XY派、XY卡、XY迷”等结构。我们还将这些语素在构成三字词时的情况与构成二字词时的情况进行了对比考察,发现三字词的构词呈现出一种明显的规则性。

分别考察“X法”和“XY法”在《现代汉语语法信息词典》数据库中的构词情况,发现“法”构成二字词时,词性多样,遍布于名词、动词、形容词、区别词、副词及简称等,如“写法、执法、合法、非法、依法、高法”等;而“XY法”构成的三字词总共94个,其中属于名词词性的词就有92个,如“出版法、律师法、障眼法”等,“法”位于词尾构成的三字词词性趋同。另外,“期、费、派、卡、热”等语素位于词尾构成三字词时也出现和“法”类似的情况。由此,我们可以认识到三字词构词时的一些规律,主要表现在有些能产性强的语素构成三字词时词性趋同,词的结构以定中式居多,词的意义可以从字面意义推测等。

其中尤为值得注意的是,这些能产性强的语素在构成三字词时形成了一种“类义功能”。“热”在构成三字词时,在形式上形成了“XY热”的词法模式,在意义上也有一种聚合功能,如“股票热、回归热、电脑热”等,这些是比较新的词,是按照某种造词模式创造出来的,将来如果再出现一个新兴的行业,还可以另造新词。正是有了这种比较凝固的词法模式,新词才会源源不断地创造出来。

何谓词法模式?董秀芳认为词法模式具有以下特征:(1)其中一个成分具有固定性,另一个成分具有语法类别和语义类别的确定性。(2)构成成分之间的语义关系固定。(3)整体的意义基本可以预测^②(第101页)。

我们认为词法模式的形成应有两方面的基础,在形式上有一些固定成分,在意义上,有这个模式所凝聚的整体意义。因此,我们把汉语中具有一定规则性的产生词的格式称为词法模式,如“XY期”、“XY法”、“XY热”等三字词的格式,这些格式中有一个成分固定,词义比较透明,结构关系为定中式。

作者简介:曾立英,中央民族大学国际教育学院副教授,北京100081。

基金项目:教育部人文社会科学研究基金青年项目(07JJC740016);国家社会科学基金项目(09BYY032);国家自然科学基金项目

例如“XY 期”表示“某一个具体的时段”，以这个模式构成的三字词比较多，如“采收期、经销期、低潮期、巅峰期、合同期、繁荣期”等，调查 1998 年和 2000 年的《人民日报》的语料，发现“XY 期”模式构成的三字词共 312 个，说明这种词法模式的能产性还是很强的，有研究的价值。

三字词的词法模式具体有哪些，它的构成有哪些规则？笔者从对《现代汉语语法信息词典》数据库的逐条分析出发，认为三字词的词法模式主要可以从构成成分和语义类别两个方面来探讨。

二、词法模式中的“部件词”

三字词的词法模式中有一些成分比较固定的词，如“XY 法”、“XY 期”、“XY 热”中的“法、期、热”等，这一类词形式上比较固定且能产性强，我们把这些构词成分称之为“部件词”。部件词的抽取有助于三字词词法结构和语义结构的识别，因而对部件词的研究实际上也蕴含了对词法模式的研究。

部件词在构成三字词时相当于一个“造词模子”，可以生成很多的词语。董秀芳认为：“汉语词库中不仅存储一些高频词，而且同时存储一些高频语素，因为这些语素是构词要件，是词法规则所作用的对象。”^[1]（第 42 页）因此，我们有必要把这些高频语素和高频词都挖掘出来，作为造词的重要“部件”，分析词的构成。我们认为，词法研究的目标不仅仅是语言中现实的词的集合，也有语言中可能的词，即研究合法的词的生成规则。

如现代汉语中以“人”为词尾构成的词，能产性比较强，在《现代汉语语法信息词典》词库中以“人”为词尾构成的三字词有 161 次，其中有 155 次是“2+1”式的“XY 人”结构，例如“湖北人、证婚人、纳税人、年轻人、庄稼人”等，都是表示某种类别的人，基本上是定中结构，其中“湖北人、美国人”等是由“地名+人”的词法模式所构成的，表示“出生在某地或居住在某地的人”，“出生”或“居住”的含义就是词法模式所赋予的。该模式作用力最强，可以说具有一定的平行周遍性^[2]（第 355-365 页），由此，可以生成和理解“湖南人、德国人”等很多词。

我们注意到“XY 人”模式中，“人”是位于词尾的，那么“人”位于词首或词中时，有没有“XY 人”的规则？我们调查《现代汉语语法信息词典》词库，发现“人 XY”和“X 人 X”两种结构共有 96 次，如“人才库、人工湖、大人物、人贩子、二人转、人行路”等，其中“人”的意义比较分散，有指“人才”的，有指“人工”的，有指“人物”的等等，语义上不能形成一种类聚，所以不宜将“X 人 X”和“人 XY”看作词法模式。

像“XY 人”之类的词法模式还有很多，如“XY 队、XY 店、XY 山、XY 鱼、XY 油”等等，如“探险队、足球队、医疗队”等，“XY 队”也能形成表示“某种群体”这样一种词法模式义^②。调查了《现代汉语语法信息词典》词库，发现“店”位于词尾的三字词共 36 个，如“食品店、服装店、夫妻店、代销店”等；“店”位于词首的三字词只有 1 个，即“店小二”；“店”位于词中的三字词没有。从词频上考虑，“店”位于词尾时，因构词能力强，可形成一种“XY 店”的词法模式，表示“什么类型的店”，是定中结构。

类似于“XY 店”的词法模式，“XY 山、XY 鱼、XY 油”的构词能力也较强，在《现代汉语语法信息词典》中它们的词频分别为 37、40、52 个，它们的结构关系也基本上是定中式^③，中心语为“山、鱼、油”，中心语决定了“五台山、武昌鱼、食用油”这些三字词的整个组合的“类”范畴，是名词性的，表示某种类型的山、鱼、油等。

我们把“XY 人、XY 山、XY 鱼、XY 油”中的“人、山、鱼、油”等成词语素也列为部件词，一是重视这些语素在词中的位置，它们在词尾时才能凝固成词法模式，具有一定程度的定位性；二是强调这些语素的构词能力强；三是为了说明这些语素也能像“词缀”、“类词缀”一样，形成某些词法模式；第四是为了说明这些语素在整个组合中具有一种“类义功能”，因此我们把具备这些条件的语素称之为“部件词”。

三、基于《现代汉语语法信息词典》的部件词提取

三字词中究竟有哪些词可以称为部件词，我们在第 2 节中已考虑到首先基于词频选取部件词，认为构词词频高，是成为部件词的必要条件。我们首先基于《现代汉语语法信息词典》8 万多词的词库，进行

了三字词的提取,共有13 778个三字词,组成“三字词词库”。根据部件词在三字词的定位情况,我们分别提取“接尾词”和“接头词”。

(一)三字词的“接尾词”

我们对前面所介绍的“三字词词库”进行了逐一分析之后,对这13 778个三字词的位于词尾的词进行了再次抽取,共有1 539个字,把这1 539个单字在数据库中按照构词频率排列,建立三字词的“接尾词词库”。

这1 539个字按构词频率由高到低排列的前50位分别是:子、性、机、器、学、人、员、化、会、品、者、儿、率(1)、法、费、家、病、部、表、剂、线、权、车、力、量、站、队、式、室、花、片、物、纸、生、业、体、石、头、词、面、期、书、场、油、图、素、炎、点、虫、院,这前50个字的构词频率都在53次以上,其中前21个字的构词频率都在85次以上。

我们还对位于词尾的1 539个字的构词频率由低到高进行了排列,这排名后50位的单字分别是:妻、浅、签、千、传、汽、喘、棋、脐、喷、凄、抢、谱、浦、泼、坪、频、篇、屁、春、皖、戚、卿、染、犬、醛、丑、趣、渠、駿、屈、出、欠、囚、川、沁、勤、储、怯、俏、瞧、黜、蜈、跑、除、曼、朦、氓、闷、搭,这50个字的构词次数都是1,甚至于到527个字的构词频率都是1。这些字构词能力极弱,就不能作为部件词了。

成为三字词的部件词不仅要求构词词频比较高,还要求在语法结构的切分上能独立构词。如“物理学”和“伪科学”中都有“学”,但是“物理学”中的“学”指的是“指某种学科”义,可以构成“货币学、档案学”等词,而“伪科学”中的“学”首先和“科”合成“科学”,然后和“伪”这个语素构成三字词。从切分上说,如果对“伪科学”采取直接成分分析法的话,是“伪/科学”,而不是“伪科/学”,“学”并不是这个三字词的“直接成分”,所以我们把“物理学”中的“学”当作是三字词的部件词,而不把“伪科学”中的“学”作为该三字词的部件词。

因此,确定三字词的部件词,必须分析词的结构。若有一个三字词ABC,只有当C是被独立切分的时候,C才可能是部件词。如对ABC采取直接成分分析法,ABC被切分为“AB/C”时,C才可能是ABC的部件词,如“视力/表、体检/表、除臭/剂、清洁/剂”中的“表、剂”是部件词。若ABC被切分为“A/BC”时,C就不能被看作是三字词的部件词,如“女/主人、光/化学、伪/科学、总/动员、肝/硬化、黑/社会、副/产品、电/功率、零/利率”等中的“人、学、员、化、会、品、率”等就不是以上所举的三字词的部件词。

我们对按照词频排列的三字词“接尾词词库”,采取切分的分析法,提炼出部件词,得到按词频排列的部件词表。根据这个部件词表,去处理《人民日报》1998年和2000年已标注的语料库,补充少许有典型意义的部件词,进一步筛选部件词。最后筛选了185个接尾词,如下所示:

班、办、病、部、菜、层、厂、场、池、处、船、村、带、单、党、岛、地、点、店、调、度、队、额、儿、法、犯、方、房、费、风、服、感、工、股、关、观、官、馆、管、鬼、国、户、花、化、会、货、机、级、集、籍、计、记、剂、夹、家、价、间、件、江、奖、匠、街、节、界、金、局、剧、军、卡、科、课、口、库、款、狂、栏、佬、类、力、连、量、林、令、楼、录、路、率(1)、论、们、迷、面1、面2^④、能、年、牌、派、盘、旁、棚、片、票、品、瓶、期、气、器、枪、桥、球、区、曲、圈、权、券、群、热、人、日、赛、色、山、商、社、生、省、师、石、史、市、式、室、手、书、术、树、数、税、所、台、坛、体、厅、头、图、团、网、委、舞、物、席、系、险、县、线、腺、箱、鞋、心、星、形、型、性、学、炎、药、业、仪、于、员、园、院、月、展、战、站、者、证、症、值、制、状、子、族、组、罪

调查这185个接尾词,发现这些词大部分是常用词,如“班、病、菜、厂、船、村、党”等等,属于认知语言学中的基本层次(basic-level)范畴,所以以这些语素构成的词很多。

另外需注意的是,部件词有单音节的,也有双音节的。如“X巴巴、X油油、X灿灿、X乎乎、X光光、X溜溜”等叠音的后缀,还有“分子、主义”等双音节的部件词。

(二)三字词的“接头词”

为了调查三字词中的“接头词”的情况,我们按上一小节所操作“接尾词词库”的操作方法同样抽取

了《现代汉语语法信息词典》数据库的三字词中的“接头词”，按照词首的构词频率由高到低建立“三字词的接头词词库”。

我们在接头词词库中，提取按构词频率由高到低排列的 50 个语素，它们分别是：大、电、小、老、水、打、中、地、白、保、发、工、黑、高、一、生、自、三、外、不、红、分、黄、天、无、公、防、金、出、人、马、文、单、太、热、主、花、交、核、多、总、冷、通、海、火、上、开、反、有、二。这些字的构词频率都在 35 次以上。其中有些语素，如“大、小、老、水、打、地、白、一、二、黑、三、红、黄、天、人、热、花、冷、海、火”等是在 Swadesh (1952) 的 207 个基本词汇表中体现的，Swadesh 的基本词汇表是基本词汇库的原型，这个词表以概念上的必然性为其定义基准^[3]（第 61-72 页）。我们上面按词频排列的高频语素，如“电、核”等语素，Swadesh 的基本词汇表没有也是正常的，因为社会在进步和发展，语言中的词汇也在发展。

位于词首的部件词的确认，不仅要看词频，还要看部件词构成的三字词的结构，对结构的分析主要采取切分的方法。如“电”位于词首的频率虽然很高，但是构成的三字词 ABC 的结构有很多是 AB/C 式的，如“电视机”的切分是“电视/机”，说明“XY 机”是该词的词法模式，而“电 XY”不是“电视机”的词法模式，这是从直接成分分析法的角度来考虑的。确定一个语素 A 是不是三字词 ABC 的部件词，可以从语素的组合层次来看，A 应该被独立切分，形成 A/BC 式的格局。

考察了上面所提到的“三字词中位于词首的构词频率最高的前 10 个语素”，发现“水、中、地、白、保”等不是构成三字词的部件词，而是构成二字词的部件词，如“水烟斗、水利局”的切分分别是“水烟/斗、水利/局”，“中心词、中介人”的切分分别是“中心/词、中介/人”，“地形图、地区性”的切分分别是“地形/图、地区/性”，“白斑病、白日梦”的切分分别是“白斑/病、白日/梦”，“保温层、保健员”的切分分别是“保温/层、保健/员”，也就是说，“水、中、地、白、保”在词中没有被独立切分，不是部件词。而“大、小、老、打”等可以说是构成三字词的部件词，如“大/丈夫、小/花脸、老/病号、打/下手^⑤”等，因为这些三字词的切分格局是 A/BC 式，A 被独立切分，且 A 的构词频率高，所以 A 可以被提炼成为部件词。

我们对按照词频排列的三字词“接头词词库”，采取统计分析和切分分析相结合的方法，提炼出部件词，并在语料库中进行了检验，最后筛选了 26 个接头词：

副、打、第、多、泛、类、超、无、非、代、半、老、不、总、反、软、可、零、相、伪、第、单、负、大、老、小

这些接头词构成的三字词的词性有很多是区别词(b)和名词(n)，如“副厅级、超薄型、半自动、多层次”等是区别词，“代厂长、总书记、零距离”等是名词。有一些接头词本身是动词，如“超、反、无”等，在构词时由于使用频度高，渐渐向类词缀转化。另外，有一些接头词也可以说是和英语的前缀相对应的，如“超、反、不”等，下面以“超”为例来说明中文构词和英文的对译。

汉语中的“超”可对应于英语的 hyper-, ultra-, over-等词缀，比如“hypertonion(超概念)，hyper text(超文本)，hyperacoustic(超声波)，supergiant(超巨型的)，supernatural(超自然的)，superior(超一流)，ultraportable(超便携式)，ultrasonic(超音速的)，ultrafast(超速的)，ultracytochemistry(超显微生物学)，overweigh(超重)，over age(超龄的)”等，这些现象说明从外语翻译的外来词也对形成部件词产生了一定影响。语言在发展，汉语新词语的构词过程正兴起“词缀化”倾向，而这些词缀又不是真正的词缀，先纳入部件词的考察范围。

四、单音节部件词构词的类型功能

单音节部件词，包括接头词和接尾词，具有“类义功能(function of class meaning)”。有一些学者谈到过“类化作用”，如朱亚军^[3]（第 25 页）、王洪君、富丽^[4]（第 6 页），认为“类化作用”指的是词缀、类词缀和“助字”有决定整个组合的语法功能的类范畴。我们这里所说的单音节部件词的“类义功能”指的是，部件词在某一个义项或几个义项上构词能力强，构成的词形成一组表示某一种“类别”的意义相关的“词群”，在语法结构上也有决定整个组合的语法功能的类范畴作用。

有的语素有可能有几个义项，究竟是在哪一个或哪一些义项上这个语素能成为部件词，这就要考虑

语素的义项问题。比如“法”,《现代汉语词典》解释有7个义项,我们认为在2个义项上“法”可以形成词法模式,这2个义项我们分别归纳为“在某一个领域具体的法规”和“方法、方式”,形成的三字词组合分别有“广告法、票据法”和“结合法、嫁接法”等,语法结构都为定中式。

有的同音语素不只一个义项,义项之间有一定的关联,可以设同形1、同形2等。如以“族”为语素构成的能产性比较强的词群有两组,一组是如“布依族、俄罗斯族”之类的词,还有一组是如“追星族、工薪族”之类的词群,第一组中的“族”是“种族;民族”义,第二组中的“族”是指“具有某种共性的一大类事物或人”的意义。

我们把第3节中所介绍的26个“接头词”、185个“接尾词”整理成了“部件词知识库”,主要列举出部件词的读音、义项、词性、构词频率、构成的三字词、三字词的切分和语法结构,并参照刘源等的《信息处理用现代汉语分词规范及自动分词方法》^[5](第25-26页)以及台湾中研院^⑥的“接头词”和“接尾词”情况于库中,便于对比说明。

我们构建的“部件词”知识库,列出了每一个部件词的词性,构成的词是什么词性。单音节部件词在切分时,词性绝大部分标注为名词(n)、名语素(N_g)和词缀(k),三字词的语法结构以定中式和附加式为主,如“安家费、办公费、艺术性、可用性”等,部件词的词性在整个结构中起着主导作用,比如“安家费”中的“费”是名词,决定了整个组合也是名词性的;“艺术性”的“性”是词缀,决定了整个组合是附加式的。

语素或词的读音对能否成为部件词还是有很大作用的。比如“地”构成的三字词在整个5万多条的三字词库^⑦中共有258条,形成“XY地”结构,如:

A组:举办地、发案地、主办地、承包地、油菜地、生产地、成长地、碱场地、流出地……

B组:长长地、痴痴地、呆呆地、深深地、闷闷地、闪闪地、生生地、稳稳地……

这些三字词中,A组“举办地、发案地”中的“地”和B组“长长地、痴痴地”中的“地”读音和词性不一样。A组的“地”是名词,读“dì”;B组的“地”是助词,读轻声。像A组“举办地”中的词法模式是XY地(dì),表示“地方(dì fang)(1)”^⑧义,有265个;像B组“长长地”这种“XX地”的词法模式,第1个语素和第2个语素重叠,构成状态词,共20个。

我们在部件词知识库中非常重视部件词的意义及构成词的意义,发现部件词构成词的意义具有透明性、规则性,因而词典不会全部收录这种构词方式构成的词。由于部件词目前正处于能产活跃期,其新生类推潜能是无限的,如“显示器、充电器、山寨版、股票热、二手车”等,因此任何一部词典,即使是大型词典也不可能穷尽地收录部件词构成的所有词,而且不少部件词构成的词词频很低,如“减震器、测程仪、引座员、蛲虫病”等,它们是根据需要而临时创造的。三字词中的“部件词构词”具有一定程度的规则性和周遍性,是汉语新词得以出现的主要框架。

通过我们对部件词意义的归纳,发现部件词的意义大致可以归纳为下面15类:

1. 表示某种类型的人,这一类部件词比较多,如“犯、鬼、佬、家、户、狂、迷、群、商、生、手、星、员、者、族、佬、人、分子”等,这些部件词总是描写社会生活中的某一类人,表示人的类别。

2. 表示“物件、物体”的意义,如“件、品、体、物”都有表示“物品、事物”义,例如“件”的含义是“指总体中可以分开一一计算的事物”,如“焊接件、加工件”等;“品”的含义是“物品”,词频很高,有107次,如“收藏品、滋补品、洗涤品”等;“体”的意义是“物体”,如“棱柱体、立方体、结晶体”等;“物”的意义是“东西;事物”,如“图腾物、吉祥物”等。

3. 表示某一类具体的事物,如“机、菜、厂、场、单、店、额、房、费、国、花、会、货、籍”等,可以构成“锁边机、金针菜、合资厂、停车场、承包单、音像店、赔偿额、限价房、保安费、签字国、迎春花、招待会、原装货、法国籍”等。

4. 表示某个地方,如“点、处、地”等,可以构成“创新点、背阴处、举办地”等。

5. 表示事物的性质,有“化、度、气”等,如“普遍化、舒适度、书香气”等。

6. 表示观点、态度,有“论、观、感”等,可以构成“概率论、工作观、安定感”等。

7. 表示“部位或方面”,有“方、面”等,如“破产方、黑暗面”等。
8. 表示“方法”,有“术、法”等,如“移植术、结合法”等。
9. 表示某个组织,如“班、团、军、组、党、队、股”等等。
10. 表示某个机构,如“办、部、处、局、科、连、所、组”等。
11. 表示时间,如“年、月、日、周”等。
12. 表示处所,如“带、地、点、关、馆、库”等。
13. 表示计量、数量相关的意义,如“户、集、量”等,如“达标户、摄影集、生长量”等。
14. 表示某个特定领域,如医药领域的“病、炎、症、药”等,可构成“肠胃病、牙周炎、不孕症、特效药”等。
15. 表示地名,如“楼、路、省、市、县、村、岛”等。

本文基于《现代汉语语法信息词典》中对部件词的提炼,发现部件词构词具有“类义功能”,这说明在三字词的内部,实际上蕴含了语义驱动的原理,语义驱动着三字词的构词呈现出自己的规律,形成一系列的词法模式。所以我们很有必要了解部件词构词的模式,把我们所总结的“部件词”知识库中的规则应用于现代汉语文本的分词标注中,在对语料测试的基础上改进规则,处理文本中含有“部件词”的词语的外边界以及相关的歧义切分问题,将能加强对未登录词的词义及语法属性的判断,提高分词能力。

注 释:

- ① 参照俞士汶等著的《现代汉语语法信息词典》的电子版数据库,5万多词语于1995年底通过电子工业部的技术鉴定,2001年词语扩充到7.3万,2006年又扩充到80685个。
- ② 当“队”位于三字词词中时,共5例,如“总队长、卫队长、支队长”等,其中的“队”的意义在整个三字词的结构中,并不凸显,凸显的是表示“官职”义的“长”。
- ③ 需要注意的是,“XY山、XY鱼”结构并不都是定中式,也有例外,如“坐江山、炒鱿鱼”等,但这些动宾结构的词例毕竟是少数。“XY油”也有“绿油油、乌油油”等6例三字词,这正反映了3.1中所说的叠音后缀“油油”。
- ④ “面”分为“面1”和“面2”,是把义项“面条”等粮食作物和“部位或方面”分为两个词条来解释的,《现代汉语词典》也是分为两个条目的。
- ⑤ “打下手、打牙祭、打折扣、打主意”等三字词的结构都是述宾式,可以构成“打XY”的词法模式,这一类结构在《现代汉语语法信息词典》数据库中有45个。
- ⑥ 参照台湾“中央”研究院资讯科学研究所中文词知识库小组的《“搜”文解字——中文词界研究与资讯用分词标准》,1996年,未刊。
- ⑦ 这5万多条三字词词库,是在《现代汉语语法信息词典》13778个三字词的基础上扩展而成的,扩展《人民日报》1998年的经切分标注后的语料库中的三字词约38455个,共形成52210个三字词词库。
- ⑧ “地方(d fang)(1)”指《现代汉语词典》(第5版)所列词义的第(1)个义项,即“某一区域;空间的一部分”义。

[参 考 文 献]

- [1] 董秀芳:《汉语的词库与词法》,北京:北京大学出版社2004年版。
- [2] 陈保亚:《20世纪中国语言学方法论》,济南:山东教育出版社1999年版。
- [3] 朱亚军:《现代汉语词缀的性质及其分类研究》,载《汉语学习》2001年第2期。
- [4] 王洪君、富丽:《试论现代汉语的类词缀》,载《语言科学》2005年第5期。
- [5] 刘源、谭强、沈旭坤:《信息处理用现代汉语分词规范及自动分词方法》,北京:清华大学出版社1994年版。
- [6] 曾立英:《三字词中的类词缀》,载《语言文字应用》2008年第2期。

(责任编辑 何坤翁)