

基于单字的现代汉语词法研究

王洪君 著

商務印書館

2011年·北京

图书在版编目(CIP)数据

基于单字的现代汉语词法研究/王洪君著.—北京:
商务印书馆,2011

ISBN 978-7-100-07526-8

I. ①基… II. ①王… III. ①汉语—词法—文集
IV. ①H14-53

中国版本图书馆 CIP 数据核字(2010)第 236141 号

所有权利保留。

未经许可,不得以任何方式使用。

基于单字的现代汉语词法研究

王洪君 著

商 务 印 书 馆 出 版

(北京王府井大街36号 邮政编码100710)

商 务 印 书 馆 发 行

印 刷 厂 印 刷

ISBN 978-7-100-07526-8

2011年4月第1版

开本 850×1168 1/32

2011年 月北京第 次印刷

印张

定价: 元

序 言

不知不觉地,发表的词法论文有十好几篇了。它们散落在多个杂志上,不易收集。不少同仁、朋友、学生向我建议,出个文集吧。

我也有这个念想。词法研究是我自己看得很重、用功也甚勤的一个方面。我的词法研究,特点是“基于单字”。它的核心思路是:跨层面的单位关联点是研究单层面的结构基点。比如跨语音、语法两层面的结构关联点(像汉语的字和英语的词)是研究一个语言音系或词法的基点。这一思路的形成,有徐通锵师字本位理论的影响,也源于自己对汉语语音词法接面规则的研究,还得益于 Halliday 对于层级单位划分的理论论述和朱德熙等先生关于形式和意义必定互为表里的理念。宋柔先生关于汉语两字组的研究和基于大规模语料库开发的两字组中单字组配的各种统计材料,更开拓了我对汉语词法的认识和研究思路。把多年来的词法论文放到一起,基于字的词法研究思路和以上理论背景可以看得更加清楚。

文集中论文的编排可以有两种不同的方式。一种是按研究内容与学科分支的关系分门别类排列,另一种是以论文发表的先后为序。

2 基于单字的现代汉语词法研究

第一种方式有利于显示内容上的组织性,文集最终选择了这种方式。也即,先语法构词,次语音构词,再综观汉语词法以至整个汉语文系统的类型特点。语法构词中,先复合词、再派生词、再凝固度和语义特征等问题。语音构词中,先论基于字的分音和合音,次论汉语联绵语音词的构造方式,再论汉语一般韵律词和韵律类词的特点和单双音节在汉语词法句法中的重要作用,最后讨论与韵律词/类词/短语有关的节奏问题。在类型特点的讨论中,先从语音和词法两层面的接面论音系的基本单位,再从语音、文字和语汇-语法层面的关系论汉语文大系统的基本单位,最后是对语法基本单位的不成熟讨论。最后的讨论已经溢出了词法、句法,进入了汉语篇法的基本单位问题。

第二种编排方式其实也有它的优点——有利于展示个人的研究历程:观察词法现象的视角、思路和研究资料是如何随着研究的进展而逐步扩展的。

无奈,一本文集不可能同时采取两种编排方式,只能放弃后一种选择。我想了个补救的方法:在不少论文的后面追加了补记。补记一方面对文章存在的问题或发表后有所改动的观念做出说明,另一方面介绍文章写作的背景,特别是文章的观察视角、思路如何形成和研究资料如何取得。希望这些补记能够使以前发表的文章都“跟”上我现在的思路,同时又能更好地揭示我在语音构词方面的研究如何深刻地影响了我的语法构词研究和对汉语文大系统的理解,也能够更好地显示参与汉语信息处理的部分工作给了我多么大的收益。

我喜欢亲自触摸大量的真实语料,从各种不同的角度进行

统计和排比。像文集中《汉语的类词缀》一文,为了说明类词缀的最大的搭配对象与词缀、与虚字有什么不同,我从8亿字符语料库中穷尽检索到303104个含“式”的句子,经人工鉴别和去重后实得“-式”缀的语段3466例。由于其他工作繁多,仅这一项统计就花去我整整6个月的时间。但在材料的徜徉中,从隐隐地体验到其中的规律,到通过某一角度的统计而抓住其中的规律,这一过程令人着迷。通过分析这些语料得出了词缀的最大搭配对象为双音词,类词缀的最大搭配对象为类词(4音节以上的黏合字组和少数凝固词组),虚词的最大搭配对象为自由短语的结论。其他篇语法构词的论文,我也都利用了对上亿字符语料库的检索。我始终认为,统计与规则不是对立的,语言中并没有物理学那样的规则,有的只是倾向。统计规律中蕴藏着语言学的道理,语言学的规律。

本项研究曾先后得到教育部人文社科重点研究基地北京大学汉语语言学研究中心自立项目《现代汉语两字组结构研究》、国家社科一般项目《现代汉语韵律与语法语用的关联》(01BYY022)的资助,谨致谢忱。

文集的出版使得我多年来基于字的词法研究成果和我心目中的字本位得以完整地展示。感谢商务印书馆,助我实现多年的念想。感谢文集的责任编辑、我大学的同窗密友刘一玲,她细心、专业的审读,更正了我书稿中不少错误,也使颇难安排的表格能够尽可能美观地呈现在读者面前。

科学探讨无止境,语言学的探讨无止境。随着计算机和互联网技术的飞速发展,语言学可资利用的语料越来越丰富,语言

4 基于单字的现代汉语词法研究

本身也眼见着出现了许多新的文体、新的造词手段。文集中所收录的研究有它的局限,但我相信,基于字本位的汉语词法研究在新的时代一定会有新的发展,新的用武之地。

王洪君

2010年2月28日

于北大承泽园

目 录

序言·····	1
从字和字组看词和短语	
——也谈汉语中词的划分标准·····	1
从与自由短语的类比看“打拳”、“养伤”的内部结构·····	30
“逆序定中”辨析·····	54
动物、身体两义场单字组构两字的结构模式·····	62
动物、身体两义场单字及两字组转义模式比较·····	92
试论现代汉语的类词缀·····	108
从两个同级义场代表单字搭配的异同看语义特征和语义层级	
——以“锅”和“碗”为例·····	142
二字短语凝固度分级初探·····	163
《信息处理用现代汉语分词词表》的内部构造和汉语的 结构特点·····	183
汉语常用的两种语音构词法	
——从平定儿化和太原嵌1词谈起·····	204

2 基于单字的现代汉语词法研究

汉语联绵语音词的韵律类型

- 兼评石毓智的《论汉语的大音节结构》…………… 235
- 汉语的韵律词和韵律短语…………… 249
- 音节单双、音域展敛(重音)与语法结构类型和成分次序 …… 277
- 普通话中节律边界与节律模式、语法、语用的关联…………… 303

汉语的特点与语言的普遍性

- 从语言研究的立足点看中西音系理论的发展…………… 331
- “字本位”与汉语二语教学…………… 346
- 语言的层面与“字本位”的不同层面…………… 367
- 汉语语法的基本单位与研究策略…………… 393

从字和字组看词和短语^{*}

——也谈汉语中词的划分标准

提要 一、“字”的名称比“语素”更能确切地反映该级语法单位在汉语系统中的特殊地位——语音单位和语法单位的交汇点。二、语言中各级单位的切分应该以语音、语法单位的交汇点为基点。即，汉语应该从字组中字与字的结合关系出发，先找出字与字自由组合的短语规则，再以排除法确定词。三、具体给出与成词多有交涉的两字主谓、述宾、述补、联合、偏正式短语及其成分字序列在功能和语法—语义两方面的对应规则及这些短语的 a、b、c 三级鉴定式。四、以这些规则和鉴定式为标准判定了近三百个两字组成词与否及成词性、类词性的高低。

一 为“字”正名

中国语言学界通常用“语素”和“词”对译英语的 morpheme

^{*} 原文发表于《中国语文》1994年第2期102—111页，本次发表有个别改动。原文受韵律音系学(prosodic phonology)、赵元任(1975)及徐通锵“字本位”说的启发而作。写作过程中不少问题曾向学友郭锐请教。初稿曾蒙徐通锵先生通阅，二稿曾在第四届全国现代语言学研讨会及中国语言学年会的分组会上宣读，得到不

2 基于单字的现代汉语词法研究

少宝贵意见。

和 word,汉语语法分析中设立语素和词这两级单位似也已成惯例。然而正如赵元任(1975)所指出的:“研究现代语言学的学者都同意,对于所研究的对象语,不应该刻意地去寻找那些在我们从前刚好会说的那种语言中十分熟悉的东西,而应该确定我们实际上碰到了什么,并给它们以合适的名称。”那么,让我们先看看汉语中被称为“语素”的成分与英语的 morpheme 有什么相同之处,又有什么不同的特点,然后再考虑它的名称是否合适。

两者的共同点在于都是最小的音义结合体。两者的不同点则至少有如下几点:1)汉语的最小音义结合体大多是单音节的,有声调作为明显的标志。不合此例的基本上分属两种情况:a.不是语言中的基本形式(如双声、叠韵、重言、嵌1、特殊四字格、儿化韵、D变韵、Z变韵等,都是经“一生二”、“二合一”等叠音或合音构词法派生的次生形式)。b.外语音译词,因连音带义一起舶来,所以保留了外语中的音义关联。舶来时间长了的,往往要按汉语的音义关联改造(参看徐1993)。英语的 morpheme 与语音单位没有大致重合的关系。它可能是单音节,也可能是多音节,还可能跨着音节界线。可能不足一拍,可能多节拍,也可能跨节拍。它也没有重音(英语的重音到词这一级才标定)。2)汉语的最小音义结合体大多是自由的。而且不管是自由还是黏着,音形大多是稳定的。即以不变的形式加入以下所有层次的组合。英语的 morpheme 则大多是黏着的,语音形式是不确定的。要先通过构词或构形规则组合成有重音的词,形式才稳定下来,才能进入短语结构的构造。

以上两点区别不是无足轻重的,它说明了汉语中最小的音

义结合体与英语的 morpheme 在各自语法体系中地位不同。英语中(也许所有语言都同样)并非没有与语音单位大致重合的语法单位,只是它落在了 word 上:除个别虚词外,每个 word 都有一个、也只有一个主重音。由一个主重音联系在一起的若干节拍,形成语音上的一级单位,称作 phonological word(语音词)。英语的 word 是以不变之音形参加以下各层次组合的最小成分。

可以看出,汉语中所谓“语素”仅在“最小的音义结合体”这一意义上与英语的 morpheme 相当,而在是否是语音、语法单位的交汇,是否是以不变之音形参与更大结构的构造等特点上都不同于 morpheme,而与 word 相当。我们认为,其不同点是更重要的,它们是一个语法系统最基础的单位所应具有的特征。把英语的 morpheme 译为“语素”,把汉语中这种一音一义的小单位也称为“语素”,实在是太委屈后者了。

一个语言中语音、语法单位的交汇点,是该语法系统最基础的单位,也是母语者最容易意识到的单位。因而,也就往往成为文字的一级单位,形成语音、语法、文字三种单位在某一级上大致重合的事实。是语言的特点决定了文字,而不是相反。与存在语音、语法、文字三种单位交汇点这一事实相适应,自然语言中也就往往有专门指称该级交汇单位的名称。如英语中的这个 word,在语言学家专门定义之前就既指语法单位,也指语音单位、文字单位。英语语言学家并没有因为 word 能同时指三种单位而抛弃它,保留这一名称也更能反映出该级单位在语言、文字体系中的特殊地位。汉语中本来也有指称汉语中语音、语法、

文字三种单位大致重合之特殊单位的名称,这就是“字”这个字。在语言学家专门定义之前,汉语的“字”既可指文字单位(如《说文解字》),也可以指语音单位(如“吐字清楚”)或语法单位(“因字而成句”,徐通锵 1993)。为什么我们要放着能反映汉语系统中语法、语音单位在何处交汇这一重要事实的“字”这一名称不用,而要用与“字”并不相当的 morpheme 去类比而另起“语素”这一新的名称呢?应该为“字”恢复名誉了。

选用“字”这一术语,绝不意味着要混淆文字、语音、语法三种不同性质的单位。要为“字”正名,明确定义这三种单位也是必不可少的任务。我们认为,文字上的“字”定义为“占据一个方形空间的小单位”,语音上的“字”定义为“一个音节”,大约是没有异议的。关键在于语法的“字”如何定义。有两种可能的选择:1)定义为“最小的音义结合体”。这是对于所有语言都适用的定义,但未必适用于汉语。因为“辗转”和“转”要平列都算一个“字”,而实际上“辗转”是由“转”派生的非基本单位^①。2)定义为“与单个音节结合在一起的、意义上有同一性的语法单位”。并可再具体细分为:a.基础字:一音节一义的结合体。b.表音字:单音节时无义,需进入更大的语音组合才有意义。比如,“辗转”是在基础字的基础上根据重叠加变形的规则而派生的,表层由两个表音字组成,构成两字联绵词。c.合音字:一个音节两个

^① 原文是:“因为‘匍匐’和‘爬’要平列都算一个‘字’。”现改以“辗转”和“转”为例并加了一句话,是吸收了孙景涛博士论文(Sun Jingtao 1999 *Reduplication in Old Chinese* 古汉语重叠之研究, Ph.D. dissertation, Vancouver: University of British Columbia)的研究成果。

意义的结合体。这样定义的“语法字”与目前学界所说的“语素”所指不完全重合^①：语法字是一音节一义的最小单位，是不能由词法规则推导出来的基本单位，由重叠加变形派生来的联绵词不是单纯的基本单位而是派生性的，其特定的联绵词形框架也承担了拟声摹态、具体化等派生义。“语法字”与“语音字”也不重合，“语法字”要有意义，“语音字”不一定要有意义。定义上没有音节数目限制的“语素”，用于汉语分析已有几十年历史了，总给人削足适履之感。本文要试试以“一音节一义”为定义的“字”，看看它能不能为汉语分析带来一些方便。

本文讨论语法问题，下文中不另加说明的“字”均指语法单位，不指文字或语音单位。

二 汉语中的词及其切分策略

一个语言中各级语法单位的最佳区分策略，是先确定那些与语音单位大致重合，母语者最容易意识到的单位，然后再按一定的原则、程序确定其他各级单位。以英语为例。尽管美国结构主义者号称是按“音素→音位→语素→词→短语→句子”的程序操作的，但细观其工作过程就可以发现，它时时处处都离不开词这级基础。王理嘉(1988)已指出，英语的音位其实并非直接对比长段语流而发现的，而是对比单音节的词而发现的。语法

^① 后面的几句有改动，吸收了笔者(王洪君 1994《汉语常用的两种语音构词法》，1996《汉语语音词的韵律类型》，本书收录)、孙景涛(Sun 1999)的研究成果。

单位的发现也同样。按结构主义的发现程序,对比若干语段而直接得到的“以大致相同的意义重复出现的小片段”(霍凯特 1958),其实只有词(如[ˈneɪʃən]、[ˈnæʃənəl]、[ˈnæʃənæli])或语子(如[ˈneɪʃən]、[ˈnæʃən]、[ˈnæʃən]、[əl]、[ˈæɪ]、[iti])。接着,无论是把词切分为语素还是把语子归纳为语素,都需要以词作为比较的基点。需要对比足够数量的词,才能发现其中某些意义大致相同的小片段,尽管语音形式不同,但却彼此间有规则性的语音联系,而其他意义相同的小片段也有类似的规则性语音联系,从而再切出或归纳出语素。

汉语的情况有所不同。汉语的词、短语甚至可能连句子都没有稳定的语音标记。但是从分界明确的单位出发找出一套程序来确定分界不明确的单位,这一原则应该是有普遍意义的。汉语中词和短语的界线模糊,字和字组却是一目了然的。以字和字组为基点观察字组中成分字间的语法-语义关系,不难发现几种典型的情况:1)字组只在语音上可线性地分析为更小的成分及联系规则,语法-语义上不可再做线性分析。2)字组从语法-语义的角度可以分析为成分及其构造方式,但这些构造方式 a.不能无限地反复套用。b.不能类推到所有同类成分上,甚或只是个体特异的。c.只适用于较短的片段之间(单字或双字)。3)字组从语法-语义的角度可以分析为成分及其构造方式,这些构造方式 a.可以反复或彼此套用。b.可以类推在一类成员中使用。c.既适用于较短的成分之间,也适用于很长的成分之间。

如果从最小的结构单位到无限长的话语,用一套规则就可

以控制,那么就只需要设立一级语法单位,然后按照结合层次的不同,反复套用该套规则来组合该级语法单位就可以了。做不到这一点,就必须设立中间层次的单位,就必须设立不同层级的几套规则。上述 1)2)类字组的内部组合规则与 3)类字组的内部组合规则明显的属于两套,而 1)2)类字组又可以按 3)的规则参与更大的组合。这样,相对于 3)类规则而言,1)2)类字组是不可再分析的小单位(凝固字组),而 3)类字组是可分析的自由组合体(自由字组)。它们应是不同层级的单位。下面把 3)类字组称作“短语”(与通行的用法一致),把 1)2)类字组称作“词”。请注意我们的“词”与通行的用法不完全相同,它是指“有内部结构但结构方式不符合短语规则或句子规则的多音节的音义结合体”。也就是说,“词”是单音节的“字”与多音节的“短语”的中间层级:由词向下,可按一套规则分析到字,由词往上,可按另一套规则组合成短语。这样定义的“字”和“词”,分别等于赵元任(1975)所说的“音节词”和“结构词”。这样的单位区分似更适合汉语的特点。

综上所述,我们认为,汉语中词切分的最佳策略是以字和字组为基点,观察字组内各成分间是否有语法-语义关系。如果有,再判断此构造是属于词的结构方式呢,还是短语的结构方式。

这就又引出了新的问题,汉语中哪些是构词规则,哪些是短语规则?我们认为,构词方式数目众多,有的方式只是几个甚至一个字组特有的,很难从正面穷尽列举。而短语规则、短语结构类型是有限的,这就为用排除法确定词提供了可能。如果能够

穷尽地列举短语规则,那么,对于任何一个有意义的多音节小片段,其直接成分的组合不符合短语(或句法规则)的,就是词。也就是说,解决汉语中词的区分问题,要依赖于搞清楚短语的规则。

三 关于短语规则

短语规则应包括以下几方面内容:1)确定短语结构的类型及每种类型的各个组合位置。2)确定每个字的释义(把字再细分为义项)。3)确定每种短语类型整体的语法-语义类别以及在更大结构体中的分布功能、扩展转换功能。4)确定每个字各个义项的语法-语义类别以及在上级短语体中的分布功能。5)确定字类的组合序列与短语结构类型的对应关系。这种对应一般应是顺向唯一的。即可以允许 $a + b \rightarrow A$ 、 $a + c \rightarrow A$ 、 $d + e \rightarrow A$ 共存,但不允许 $a + b \rightarrow A$ 、 $a + b \rightarrow B$ 共存。除非 a、b 两小类所有成员的组合都能形成 A、B 两种结构。

在这几项内容中,目前只有 1)2)有基本可用的研究成果。各家语法对 1)的描述都相差不多。2)则有《现代汉语词典》可资利用。3)4)只有部分可利用的成果。对各类短语的典型扩展转换式已经研究得较充分,而对短语整体功能的研究还很不够。字类的划分也还没人系统地做过。但是北大新近推出的词类划分标准可以给字和短语的分类提供大部分标准。另有不少关于词的小类划分的成果可以利用。5)是最薄弱的环节。汉语的研究至今未对成分搭配与整体结构间的限制作出描述(如动、形在