

从历史演化看汉语的基本结构单位

汪 锋

(北京大学 中文系, 北京 100871)

摘 要: 先秦汉语的语素组合都具有递归性,语素即为基本结构单位;现代汉语中的语素组合既有递归性的,也有非递归性的,后一类组合中的语素不是基本结构单位,只是基本结构单位的组成成分。从历史演化的角度看,汉语的基本结构单位的构成层次发生了变化,由单层发展为双层,在这一变化中,“一个字·一个音节·一个概念”的强势关联制约着整个语法格局的形成,通过字化现象和用字策略表现出来。英语等其他多音节语言常用的“增音节策略”在汉语中不用。

关键词: 递归性;基本结构单位;用字策略;增音节策略;字化

中图分类号: H1-09 **文献标识码:** A **文章编号:** 1000-5919(2015)01-0136-07

语言符号包含形式和意义两个方面。索绪尔在《普通语言学教程》中指出:符号的形式和意义犹如一个硬币的两面,不可分割。常见的语言单位如语素、词、词组和句子是不同层级的符号。一般说法是,语素是最小的音义结合体;词是最小的能独立运用的音义结合体;词组是词和词的组合;句子是带有一定语调能独立表达意思的词或词组。在这四级最常用的单位中,最容易把握的是语素,通过最小同一性对比,提取语素并不困难。(陈保亚 1999)最难界定的是“词”,对于汉语研究来说,“词”是个舶来品,是要在汉语中找英语 word 的对应物。

理想的语言符号是一个形式与一个意义相配,如下所示:

形式 1	形式 2	形式 3	……	形式 n
意义 1	意义 2	意义 3	……	意义 n

然而,从语言符号的形式方面来看,存在着大小不同的语音单位,从小到大,依次有语音特征、音位、音节、多音节;从语言符号的意义方面来看,有语义特征、概念、概念组合。因此,形式和意义结合的时候就有很多种可能,比如,可以在单音节层面赋予意义,也可以在多音节层面赋予意义。前者的实例如汉语,绝大多数汉语的单音节都有意义;后者的实例如英语,很多单音节并不表达任何意思,只有组合成多音节片段后才有意义,比如:America 中 a、me、ri、ca 四个音节都没有独立的意思,即使可以单独表意(如:a 可以是不定冠词“一个”的意思),也与其整体要表达的意思“美洲”没有任何关联,不能通过同一性对比来提取“a”是一个语素(陈保亚,1999)。

由于意义不是物质形式,不能直接把握,因此,在语言符号的切分和认定方面,语音形式是最具操作性的出发点。语言符号最根本的性质就是约定俗成性,一个语言形成什么样的语音与意义的关联,有自

收稿日期:2014-09-20

作者简介:汪锋,男,湖北广水人,北京大学中文系副教授。

基金项目:国家社科基金重大项目“基于中国语言及方言的语言接触类型和演化建模研究(项目批准号:14ZBD102)”。国家社科基金重点项目“基于严格语音对应的汉语与民族语关系字专题研究”(项目批准号:13AZD051)。

己的历史渊源。语音与语义关联形成语言符号,而这些符号又是为语言的交际功能服务的,在这一点上,世界语言都是共同的,这又可以推知,尽管不同的语言形成语言符号的路径各有不同,但受制于交际功能,又必然有一些共同的特点。

这样,考察汉藏语言单位的性质就有两个密切相关的角度,一个角度是类型学,通过当今语言中语言单位构成的不同模式,可以更确切的把握汉藏语中各语言单位的性质;另一个是从历史演化的角度来看这些基本单位是如何形成的,期间,语音如何与语义关联起来,形成不同的语言符号,而这些语言符号如何在使用中又构成更大的单位,当然,也可能随着语言的发展,语音与语义关联的方式也发生转变。

二

很多学者都谈到,现代汉语是双音节词占优,古汉语是单音节词占优。但是,什么是词,却众说纷纭,那么所谓双音节词化的汉语历史演变究竟要说明什么问题就值得重新审视。

问题是由“词”这个术语引起的,那么,我们可以考虑先不用这个容易招致争议的概念,而从语素这个概念出发来观察古今汉语的历史演变,或许可以把问题看得更清楚。如果以先秦汉语为古汉语的代表,我们随机选取《论语词典》(杨伯峻)中被判定为“词”的双字组合,如下:

九夷 子欲居九夷
八佾 八佾舞于庭
下流 是以君子恶居下流
公室 禄之去公室
凶服 凶服者式之
天子 天子穆穆
木铎 天将以木铎

上述双音节“词”分析如下:

- 1.【九夷】何晏集解引马融曰:“东方之夷有九种。”《后汉书·东夷传》:“夷有九种。曰:畎夷、于夷、方夷、黄夷、白夷、赤夷、玄夷、风夷、阳夷。”可见,九夷就是九种夷人的意思。
- 2.【八佾】按周礼只有天子才能用。诸侯用六佾,大夫用四佾,士用二佾。八佾是八行共六十四人。“佾”就是一行八人组成的舞队,有几个这样的舞队,就是几佾。可见,八佾就是八个佾的意思。
- 3.【下流】在下之流,即为下处。
- 4.【公室】公共之室,即为朝廷。
- 5.【凶服】凶的服装,不吉利的服装,即为丧服。
- 6.【天子】上天之子,即为皇帝。
- 7.【木铎】木舌之铎,即为木铎。古代还有金铎。

从上面分析可以看出,这些“词”在先秦其实都是语素组合。如果知道其中每个语素的意义,其整体意义就可以推导出来,这样的“词”自然是不能算“词”了。

这样看来,在先秦文献中,语言符号的单位就是从语素到语素组合,而语素组合直接实现为句子,并没有“词”这个中间环节。也可以说,在这种情形下,语素与词是合一的。

还可以从另外一个角度来看,语言最核心的特性就是递归性,也就是洪堡特所谓“有限规则,无限运用”。指的是语言单位能自由运用,生成无限话语。因此,一个语言系统最基本的构造就是一种基本语言单位再加上有限的单位组合规则,而这些有限的组合规则可以自由运用。具体说,从基本单位 A,基本单位 B,一直到基本单位 N,同时有一套有限的基本组合规则 R1、R2 到 Rx。经过 R1(A,B) 这样的操作,就可以生成语言表达需要的话语。现在我们的问题是这些基本单位如何认定,是否可以考虑依据递归性来看?也即,这些递归性的规则作用于什么样的单位,这样的单位就是语言的基本单位。

这里有一个前提,就是先确定什么组合是递归性的?最基本的就是判断语素组合 AB 是否有递归性,陈保亚(1999)提出“平行周遍”对比来分析,如果一个语素组合 AB 能进行平行周遍对比,那么该组合就具有递归性,A 和 B 就都是该语言的基本单位,而 AB 是语言单位的组合,不是基本的语言单位。所谓平行周遍对比,指的是“1. 参加对比的言语片断在构造上是平行的,即它们的意义都是成分的意义加组合关系的意义;2. 对比的言语片断是周遍的,即所有同类的字都可以参加对比。”

以平行周遍性原则来分析上文中《论语词典》中的双字组合,可以提取具有生成性的模式,格式如下:

九夷→X + 夷 [X = 数词],义为多少数量的夷。“X = 数词”说明可以周遍的类,该类成员都可以出现在该组合中;“义为多少数量的夷”说明这些周遍替换共同的平行性意义。其他几组分析如下:

八佾→X + 佾 [X = 数词],义为多少数量的佾

下流→X + 流 [X = 垂直向方位词(上、中、下)],义为垂直向的某种方位

公室→公 + X [X = 任何物质名词],义为公家的物品

凶服→X + 服 [X = 表示命运性质的形容词,如:吉、凶],义为某种性质的服装

天子→X + 子[X = 有人性的名词],义为某某的儿子

木铎→X + 铎 [X = 质料],义为某某质料的铎

可见,这些组合在先秦时期都可进行平行周遍对比,是递归性结构,而不是基本语言单位^①。需要注意的是,一旦社会的语言环境逐渐变化,某些组合不再周遍,或者意义发生了转变,不再保持原初的平行意义,这些变化了的组合就转变成基本语言单位了。有意思的是,操同一语言的不同文化背景的人或许在判断平行周遍性上有不同看法。比如,“天子”在很多人心目中就等于“皇帝”,并不等于“天”与“子”组合起来的意思,也就成为一个基本语言单位了,而很多具备中国文言知识的人,了解其本初的意思就是“天”之“子”,则可认为该组合具有平行周遍性,是递归性结构。

三

值得注意的是,这些语素的形式标记是单音节,这与世界很多语言都不一样,在古代汉语中可以说没有多音节语素(数量不多的双声、叠韵其实是同一个音节的变形,相关研究参见孙景涛(2008),也就是说,这些双音节的变形是很受限制的,而且这个限制框架源自单音节格局。)这样通过单音节形式把一个语义概念包裹起来,构成最小的离散的语言单位,造成了古汉语的根本属性。徐通锵(1991)称之为“字”,用他的表述方式就是“一个字·一个音节·一个概念”。

根据陈保亚、汪锋(2006)的英汉 100 核心词/语素材料^②,可以发现无论先秦汉语还是英语,其中绝大部分都是单音节,如下:

英语	先秦汉语	北京话
all	皆	整个儿
ash	灰	灰
bark	樹皮	树皮
belly	腹	肚子
big	大	大
bird	鳥	鸟儿

英语	先秦汉语	北京话
bite	噬/齧	咬
black	黑	黑
blood	血	血
bone	骨	骨头
breasts	乳	嘤儿/嘤嘤儿/妈妈
burn	熬/燒	烧/着

① 《论语词典》中的语素组合是否全部都是递归性结构值得另文深入探讨,而且,如果不是递归性结构,其中的机制如何,也是十分重要的议题。

② 陈保亚、汪锋:《论核心语素表的确定——以上古汉语为例》,《语言学论丛》第 33 辑,2006 年。

英语	先秦汉语	北京话
claw	爪	爪子/爪儿
cloud	雲	云彩
cold	寒	冷
come	來	来
die	死	死/故去/走了
dog	犬/狗	狗
drink	飲	喝
dry	乾/燥	干
ear	耳	耳朵
earth	地	地
eat	食	吃
egg	卵	鸡子儿,鸡蛋
eye	目	眼睛
feather	羽	毛儿
fire	火	火
fish	魚	鱼
flesh	肉/肌	肉
fly	飛	飞
foot	足	脚/脚(把)丫儿
full	盈/滿	满
give	與/予	给
good	善/吉	好/强
grease	膏/脂	大油/猪油/荤油
green	綠	绿
hair	髮	头发
hand	手	手
head	首	脑袋(瓜儿/子)/脑壳
hear	聽	听
heart	心	心
horn	角	犄角
hot	熱	热
I	我/吾	我
kill	殺	杀
knee	膝	膊楞盖儿/胳膊*瓣儿/磕膝盖儿
know	知/曉	知道
leaf	葉	叶子/叶儿
lie	臥	躺
liver	肝	肝儿
long	長	长
louse	虱	虱子
man	夫	男的/爷们儿
many	多	多
moon	月	月亮

英语	先秦汉语	北京话
mountain	山	山
mouth	口	嘴
name	名	名字/名儿
neck	領/頸	脖子
new	新	新
night	夜	夜里(le)/黑下(xia)
nose	鼻	鼻子
not	不/勿/毋	不
one	一	一
path	道/路	道儿/路
person	人	人
rain	雨	雨
red	赤	红
root	本/根	根儿
round	圓	圆
sand	沙	沙/沙子
say	曰/謂	说
see	見	看/瞧/瞅/瞟
seed	種	种子
sit	坐	坐
skin	膚;皮	皮/皮儿
sleep	寐/寢	睡
small	小	小
smoke	煙	烟
stand	立	站/戳
star	星	星星
stone	石	石头
sun	日	太阳
swim	游	凫水儿,游泳
tail	尾	尾巴
that	彼	那个
this	是/此	这个
tongue	舌	舌头
tooth	牙/齒	牙
tree	木/樹	树
two	二	二
walk	行	走
water	水	水
we	我	我们
what	何	什么
white	白	白
who	誰	谁
woman	女	女的/娘们儿
yellow	黃	黄
You	爾	你/您(尊称)

简单统计一下,则有如下对比:

英语单音节数	先秦汉语单音节数	北京话单音节数
91	99	69

如果我们从核心词的角度来看语言的类型,可以看到先秦汉语确实是一个音节打包一个概念的。也就是,徐通锵先生所说的“一个字·一个音节·一个概念”。在 Swadesh 的 100 核心词的对应上,除了

“bark”一项,先秦汉语要用语素组合(树皮)来对应,其他都是单音节,而这个例外的一项,也不是反例,因为它不是多音节的单语素,而是两个单音节的语素组合,因此,可以说,先秦汉语的核心概念在语音和意义的关联上是百分之百的“字本位”。但是,值得注意的是,表面上看,到现代北京话时,符合字本位的似乎下降了30%,只占约70%。但不符合的主要是两类,一类是“云彩、月亮”(先秦是“云、月”);一类是“肚子、石头”(先秦汉语是“腹、石”)。细究起来,“云彩”“月亮”“肚子”“石头”都是“语素+语素”的组合,也就是,语素仍然是单音节的,并没有出现多音节的语素。也就是,从先秦汉语到现代北京话,字本位的强势关联没有发生变化。对比古今汉语,可以发现变化的是概念的认知模式发生了变化,同一个所指‘moon’,古代用一个语素“月”表达,而在北京话中则多了一个语素“亮”,可以说明所指的一个重要性质,因此在表达上更为细致,有更多的语义信息提示(相当于通常所说构词法中的“复合”);同一个所指‘stone’,除了继续保持“石”这一核心语素外,又增加了一个语素“头”,这样的语素更多的功能是用来说明所指所属的类别(相当于通常所说构词法中的“附加”)。这样的变化,可以看作语言表达内容日渐丰富的需要。而这样的变化更重要的是引起了语言单位层次构成的变化,先秦汉语的语言基本单位直接体现为单音节语素,语素组合按照生成性规则进行,产生无限话语,但在北京话中,语言基本单位的构成不仅体现为单音节语素,还包括双语素组合乃至多语素组合,这些语素组合构成的是基本语言单位,而不是生成性结构,这样就导致该语言系统中产生了两种性质的语素组合,一种语素组合构成基本语言单位,相当于之前的单音节语素的功能,一类语素组合构成生成性语言结构,相当于之前的语素组的功能。前者如:石头,木头;后者如石刀,木刀。鉴于在语言中的功能不同,这两类组合必须区分开来。前者可以叫作词,后者可以叫作词组。陈保亚(2014)明确论证了区分这两类单位的必要性,他称之为规则组合与不规则组合。反过来说,只有语言中存在这样两类不同的语素组合时,才有必要区分语素、词、词组这三级单位。至于这三级单位叫什么术语其实并不重要。比如,在北京话中,可以把这三级单位分别叫字、字结、字组(陈保亚,1999)。

四

从上文汉语和英语的 Swadesh 100 核心语素对照表中,可以看到英语的主要倾向也是单音节语素为主,占91%。如果假设这些核心词先在语言系统中产生的话,最初的语言符号编码倾向确实是一个音节打包一个概念的。但与汉语很不同的是,其中9%的例外并不是语素组合,而是多音节语素,也就是,直接增加形式上的成分来增加符号的多样性,以适应表达内容的增长。比如:belly, many, liver, mountain, person 等,都是双音节语素,但每一个单音节本身并不是语素。这与汉语在语言基本单位扩充上采取了完全不同的策略。或许这可以称之为“增音节策略”。相比较而言,汉语的基本单位扩充策略主要是利用已有的单音节语素,或许可以称之为“用字策略”。王洪君(2000)提出从备用单位和使用单位来考察现代汉语的基本语法单位,论证了从字出发更容易说清楚汉语备用单位的构成情况(字→词)。这是从共时角度得出的结论,而本文从历时角度的观察更说明了汉语如何从单层备用单位发展到双层备用单位的情况。

汉语中“用字策略”的一个最重要的证据就是“字化”现象。也就是,因为汉语中存在着“一个音节·一个概念”这样的形式-意义的强势关联,汉语中不符合这样要求的符号就会受到压力,逐渐变得符合这一要求。在汉语中,常见的“字化”有两种情况,即:1. 一个音节两个字融合为一个字,可称为融合字化;2. 无意义音节获得意义,可称为获义字化。(陈保亚,1999)

融合字化最常见的现象是儿化现象。最初,“儿”是表小、表亲,加在名词后面,如:狗儿、猫儿、花儿,等,但由于语音连读,“儿”字的读音弱化,变成一个卷舌动作,与前接音节融合,这样,如:“狗儿”/gou/ + /er/这样的两个音节两个语素,即,两个字,就变成了一个音节/gour/。一个音节里有两个语素,是汉语中形成的强势关联“一个音节·一个概念”不能容许的,人们在使用中也会感觉到这样的音节跟

其他常规音节不一样,这样就推动“儿化”在意义上发生变化,不再有表小、表亲的意义了,在普通话中开始向一个音位发展(陈保亚,1990)。^①在湖北应山话中,由于儿化的意义丧失,又增加一个“娃”与“儿化”一起构成一个新的小称词缀,“娃儿”/war/,也符合字化的要求(汪锋,2001)。

获义字化最常见的现象是一些汉语早期的连绵语中,其中一个获得代表整体的意义,或者多个音节的音译外来语中,其中一个音节获得了整体的意思,从而变成“字”。前者如:“蝶”之于“蝴蝶”,在现代汉语中,经过“彩蝶、粉蝶、蝶衣、蝶翼”的对比组合,可以提出“蝶”为单音节语素,也就是“字”;后者如“的”之于“的士(taxi)”,“巴”之于“巴士(bus)”。“的”的语素表现可以通过对比“打的、黑的、的哥、的姐”看出,“巴”的语素表现可以通过“中巴、小巴、大巴”提取出来。

有人把获义字化称为语素化,这只是对非语素变成语素这一表面现象做出了概括,并没有解释为什么在汉语中这些单音节非语素一定要变成语素。从世界语言来看,多音节语素的情况很普遍,为什么这些多音节语素“蝴蝶”“的士”“巴士”等要变化呢?但如果从汉语中单音节与语义的强势关联来看,这些问题就迎刃而解了。再者,语素化只能概括获义字化,对于融合字化却要另找解释。而“字化”可以统一解释汉语中普遍存在的这两类现象。

汉语中一直坚持“用字策略”还表现在给新事物或者新概念命名时,往往用古字^②。比如,表示“foot”意义的汉语,先秦是“足”,北京话是“脚”。football 对应的是“足球”,而不是“脚球”。如果同一个概念,古今用字不同,并不是说古字就被完全淘汰了,而是被用来构造基本语言单位,相当于构词语素,也就是被降级处理了。相较而言,古字不再具有生成能力,不能用作基本语言单位,与其他基本语言单位来构成递归性结构。在现代汉语的递归性结构中,可以有“脚疼、臭脚、手和脚、洗脚”等,不能换成其对应的古字形式,如“*足疼、*臭足、*手和足、*洗足”。从这个角度来看,汉语古今字的变化并不是完全替代,古字并不完全在语言使用中退场,而是成为“用字策略”的一个重要组成部分。如前文所说,由于语言表达内容的扩充,汉语又要坚持字本位的原则,一个途径是创造新字,也就是,用新的单音节形式来包装新出现的事物。但单音节本身的表达能力是有限的,假设单音节由声、韵、调三个部分组成,30 个声母、60 个韵母、4 个声调,能表达 7200 个事物。这对越来越丰富的物质文化生活来说是不够的。这就让汉语走上另一个途径,就是“双字化”(参见徐通锵,2001)。这样的话,一个事物可以用两个字来表达,那么,两个字表达的能力是多大呢?如上假设,双字化的表达能力是 $7200 * 7200 = 51840000$ 。这种几何级数的增长就彻底解决了“字”本位的困境。而且由于有古今字的区分,古字从不用来构造递归性结构,只用来构造基本语言单位,这就为区分两类语素组合或者说两类字组提供了重要的标示。陈保亚(2014)用来论证规则和不规则语素组的例子“视”和“看”正好也可以用说明古今字的这一重要区分功能,如下:

	左项语素的 分布性质	右项语素的 分布性质	语素组 的分布性质
远视	A	V'	N
近视	A	V'	N/A
斜视	A	V'	N/V
重视	A	V'	V
轻视	A	V'	V
弱视	A	V'	N

	左项语素的 分布性质	右项语素的 分布性质	语素组的 分布性质
远看	A	V	V
近看	A	V	V
细看	A	V	V
斜看	A	V	V

可见,古字“视”主要与其他语素一起来构成现代汉语的基本结构单位,而今字“看”则直接实现为基本构成单位。

① 陈保亚:《语言演变结构基础》,《缀玉集》第一集,北京:北京大学出版社 1990 年版。
② 这一重要现象是跟北大中文系邵永海老师讨论时,由他首先提出来的。

五

语言是一个复杂适应系统(王士元 2006)。^① 为了表达的需要,语言结构在历史演变中会不断调整,通过观察语言的变与不变,能更清楚的了解语言这个复杂系统是如何适应的。在汉语的研究中,一些基本概念常常引起很多争议,其中一个原因就是没有把汉语放在动态适应的维度上。在上文中,我们试图从先秦汉语到现代汉语的变化中,观察其中基本结构单位的性质,发现在先秦汉语的系统中,基本结构单位是单层的,语素即为基本结构单位,如果 Harris(1946)以先秦汉语为材料,或许可以实现其从语素直接到话语的语言生成机制。^② 在现代汉语中,基本结构单位变为双层,有许多语素不能直接实现为基本结构单位,需要同其他语素一起组合才能构成基本结构单位,这样就产生了双层单位:基本结构单位和基本结构单位的构成单位。这一双层单位的产生大大增加了概念表达的能力,满足了语言社会表达的需要。在这一产生过程中,单位的基础始终是单音节与语义结合的模式,也就是“字本位”强调的强势关联。这也决定了汉语发展过程中与其他类型语言不同的关键,字化现象和用字策略等与其他语言如英语中的增音节策略构成对比。汉语中的用字策略在适应过程中是如何深刻影响汉语的面貌值得进一步深入研究。

(附记:承胡素华教授相邀,本文部分内容在中央民族大学以《“词”的故事》为题报告过,得到许多有益建议;成稿后,陈保亚教授、潘秋平教授指正多处,谨此致谢。)

参 考 文 献:

- [1] 陈保亚,《20 世纪中国语言学方法论》,济南:山东教育出版社,1999。
- [2] 陈保亚,《语法描写的必要条件:双层语法单位》,《语言学论丛》第 49 辑,2014。
- [3] 孙景涛,《古汉语重叠构词法研究》,上海:上海教育出版社,2008。
- [4] 索绪尔,《普通语言学教程》,高名凯译,北京:商务印书馆,1980。
- [5] 汪锋,《应山话小称词缀演变规律初探》,《语言学论丛》第 24 辑,2001。
- [6] 王洪君,《汉语语法的基本单位与研究策略》,《语言教学与研究》第 2 期,2000。
- [7] 徐通锵,《语义句法刍议》,《语言教学与研究》第 3 期,1991。

What is the Basic Structural Unit of the Chinese Language from the Evolutionary Perspective?

Wang Feng

(Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: The combinations of morphemes in Pre-Qin Chinese were all recursive, which means that morpheme is the basic structural unit. However, the combinations of morphemes in modern Chinese may be either recursive or non-recursive. Those morphemes in the non-recursive combinations are not basic structural units, but they are components of basic structural units. From the evolutionary perspective, the component hierarchy of the basic structural units in the Chinese language has changed from single tier to double tier. In the course of this change, a strong connection of “one character with one syllable and one concept” has constrained the formation of Chinese grammar as a whole. This can be denoted by the phenomenon of characterization and the strategy of using characters to form compound words. The strategy of adding more syllables in such multi-syllable languages as English has been rarely used in the Chinese language.

Key words: recursiveness, basic structural unit, strategy of using characters, strategy of adding more syllables, characterization

(本栏责任编辑 管 琴)

① 王士元:《语言是一个复杂适应系统》,《清华大学学报》2006 年第 6 期。

② Harris, Z. S, 1946, From Morpheme to Utterance, *Language*. 22, pp. 166—183.