

汉语词的

社会语言学研究

■ 王立著

HANYU CI DE SHEHUI YUYANXUE YANJIU

 商务印书馆

图书在版编目(CIP)数据

汉语词的社会语言学研究 / 王立著. —北京:商务印书馆, 2003
ISBN 7-100-03847-2

I. 汉... II. 王... III. 汉语—词语—社会语言学—研究
IV. H13

中国版本图书馆 CIP 数据核字(2003)第 057612 号

所有权利保留。

未经许可,不得以任何方式使用。

本书获江汉大学科研经费资助

HÀNYŪ CÍ DE SHÈHUI YŪYÁN XŪÉ YÁNJIŪ

汉语词的社会语言学研究

王立 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

北 京 冠 中 印 刷 厂 印 刷

ISBN 7-100-03847-2/H·972

2003年8月第1版

开本 850×1168 1/32

2003年8月北京第1次印刷

印张 8 7/8 插页 4

定价: 16.00 元

序

什么是汉语的“词”，长期以来是一个众说纷纭、莫衷一是的老大难问题。过去中国人研究写文章，只说“字”，所谓“积字成句，积句成文”，就是说，一个字一个字连接起来就成了一句句子，一句句子一句句子连接起来就成了一篇文章。即使是口头说话，也说“你一个字儿一个字儿慢慢地说”，“字”在这里就等于音节。在说汉语的人的语感中，“字”是最容易感知的“音义结合的最小的单位”，因此也是天然的语言单位，以汉语为母语的人很容易懂什么是“字”，不管是书面的独立的书写单位，还是口头的相当于音节的语音单位。应该说，“词”的概念是从西方语法引进的，西方语言的“词”由于有形态变化，并且形态变化一般都在“词”的尾巴上，既然每一个词都有一条晃动的小尾巴，那就很容易辨认，所以从小就会说有形态变化的西方语言的人很容易懂什么是“词”，就像我们很容易懂什么是“字”一样，对他们来说“词”也是最容易分辨的天然的语言单位。马建忠写《马氏文通》的时候就只说“字”和“字类”，不说“词”和“词类”。马建忠研究的是文言语法，而文言文中一个“字”绝大多数等于一个“词”，所以他这么说，也还勉强行得通；可是一旦要研究现代汉语，由于现代汉语多数的“词”是双音节的，单音节的“字”在多数情况下只是“词”的一部分，或者说是“词素”或“语

2 汉语词的社会语言学研究

素”，而不等于“词”，因此“字”和“词”必须有所区分。这样，语法学家就引进了西方语法理论中的“词”的概念。但是因为“词”在汉语中没有明显的形式标志，不是人人都很容易分辨的天然单位，即使是语法学家也不那么容易分辨清楚，而且分清楚了也说不清楚。

最早引进“词”的概念和术语的是章士钊的《中等国文典》（1907），第一次区分了“字”和“词”。他说：“句，集字而成者也。如《孟子》云：‘齐宣王见孟子于雪宫’，共九字为一句。分视之则为字，合观之则为句，此字与句之区别也。右所引句，共九字也，而自文法上视之，则‘孟子’‘齐宣王’‘雪宫’，皆名词；‘见’，动词；‘于’，前置介词。名词三，动词一，前置介词一，共五词也。是一字可为一词（如‘见’字为动词，‘于’字为前置介词之类），而一词不必为一字（如‘齐宣王’三字，‘孟子’两字，始为一名词之类），泛论之则为字，而以文法规定之则为词，此字与词之区别也。”（1907，1页）今天看来，从一开始，章士钊的“字”和“词”就区分得不太清楚，传统的“字”的概念在他身上的影响实在是太深了，所以他一开始就要说“句，集字而成者也”。黎锦熙的《新著国语文法》是一部系统全面的现代汉语语法的奠基性著作，明确地区分了4级语法单位：字、词、短语、句子。他给“词”下的定义是：“语词简称词，就是言语中间一个一个观念的表示，观念一名意象，英文 Idea。在此处是用它的广义：一切外界的感觉、反映的知觉，想像乃至概念等，凡是有认识作用而来的，都可以叫做观念。用声音或文字来代表这些单体的整个的意象，都叫做词。”（1924，15页）他还区分了字和词：“有时一个字就是一个词，如‘人’、‘马’、‘红’、‘来’等。有时要两个字以上组合起来才成功一个词，如‘鸚鵡’、‘老头子’、‘便宜’、‘吩咐’等。文法中组织句子，分别词类，是把词作单位；不问它是一个字

或是几个字,只要是表示一个观念的,就叫做词。”(1924,15页)黎锦熙说得具体一些,但是他给词下的定义是从意义着眼的,缺乏可操作性,因为一涉及意义,就免不了见仁见智,公说公有理,婆说婆有理,是非难分了。后来赵元任想从语音角度来给词下定义,吕叔湘想从语法角度来给词下定义,因此就有了所谓“语音词”、“词汇词”、“语法词”以及“形式词”、“理论词”等等的说法,这就是“众说纷纭”,实际上是承认实在解决不了什么是“词”的问题,只得限定在某个范围以内来说“词”的问题了。这以后计算机专家为了制定计算机用的词表,不得已又发明了“切分单位”、“信息词”、“工程词”等等术语,也同样说明了没法解决汉语中“词”的确切定义,只得在名词术语上变花样。1988年国家教育委员会和国家语言文字工作委员会联合公布了《汉语拼音正词法基本规则》,似乎什么是“词”的问题终于解决了。但是实际情况并非如此。正词法基本规则尽管规定了分词连写的规则,但是并没有得到社会的公认,更没有起到规范汉语拼音实践中分词连写的作用,也没有起到解决汉语中什么是“词”的问题。这并不令人感到意外,因为世界上所有分词连写的拼音文字,都是先有约定俗成的分词连写的文字,后有正词法的规范,古今中外还没有先由专家制定了正词法规则,然后老百姓遵照专家的规定去分词连写的先例。那么现有各种拼音文字分词连写的历史是怎样的呢?是谁给这些语言的“词”下的定义的呢?实际上使用拼音文字的语言,一开始,没有人给大家规定必须分词连写,也没有人关心什么是“词”的问题,谁觉得怎么写合适就怎么写,但是使用同一种语言的人什么地方分写,什么地方连写的想法和写法都差不多,经过一段时间的实践,也就约定俗成了,而由于使用同一语言的人的语感基本一致,所以分写的单位也

就基本一致,语言学家制定的所谓规范也只能是追认广大群众的实践而已,因为任何违反群众语感的规范都是行不通的。

汉语长期使用汉字,没有使用拼音文字,所以什么是“词”的问题就不容易解决了。但是这并不等于汉语中没有相当于英语的 word,或法语的 mot,德语的 Wort 那样的语言单位,语法学家分析语法从来没有由于对“词”的认识不同而引发争论,几乎所有的争论都是由对各个“词”所担任的句子成分功能的认识不同而引发的。这样看来语法学家对什么是“词”实际上存在一个共同的认识,尽管在理论上可以争论不休。例如在“台上坐着主席团”这句话中,哪是主语、哪是宾语可以争论不休,可是从来没有人说“台”是主语,“上坐”是谓语核心,“着主”是间接宾语,“席团”是直接宾语这一类胡话,而总是把“台上”、“坐着”、“主席团”认定是句法分析的最小单位,最多也就是进一步把“台上”分析为名词“台”加方位词或方位助词“上”,把“坐着”进一步分析为动词“坐”加动态助词“着”而已。由此可见,汉语中不是根本没有“词”,只是汉语中的“词”跟西方语言中的 word 等等不完全相同而已,而且也证明不同的语法学家对“词”的语感也是基本一致的;另外,连什么是“词”都不知道,可是词类问题的讨论却始终很热烈,各种词典照编不误,并且越编越多,真是咄咄怪事!这种似乎不可思议的怪现象只有一种解释,那就是语法学家也好,一般人也好,对相当于“词”一级的语言单位都有一种基本一致的语感,仅仅是说不清楚而已,而这正是语感的特点,知道怎么使用和分辨,但这只是一种感性认识,说不清楚,因为还没有上升到理性认识的阶段。至于这样一种语言单位用一个什么样的名称那倒并不重要。大概正由于这样的现实,20世纪50年代就有人认为可以从语感着手来解决什么是

“词”的问题。1956年6月赵恩柱在《中国语文》上发表《从语感就能判定词儿》的文章,指出:“就现代汉语的词儿看,使用现代汉语的人的看法大体一致,这意味着他们的感觉基本上是一致的。当然,我们也可以用语法形式来判定词儿,但这不是根本办法,而且,能不能有某种形式,还要决定于我们的语感。”

我在80年代初学习了社会语言学的调查统计分析方法之后萌生一种想法:能不能用社会语言学的调查统计分析方法在广大群众中进行调查,看看使用现代汉语的广大群众是怎么看待“词”这一级语言单位的。但是,我始终没有这样去做,因为既没有足够的勇气和精力,也缺乏必要的经费。因此当90年代末我看到武汉江汉大学王立同志在《语言文字应用》等刊物上发表的一系列有关汉语“词”的公众语感的社会语言学调查文章时就非常高兴,从心底里赞同和支持。后来我把她的调查统计分析成果告诉了陈松岑教授,她知道以后立即在北大进行了类似的调查,而且更为惊人的是得出了完全相同的结论,尽管陈松岑的调查对象包括了系统地受过语法教育的英文系和中文系的学生。

现在王立同志准备把她这几年来关于汉语“词”的公众语感的社会语言学调查统计分析成果著书出版,我当然举双手赞成,并很乐意为这样一本有特殊价值的著作作序。

王立同志是学社会学出身的,从事社会语言学的调查研究工作比我这种野狐禅式的人更有条件,做得也更细致科学。因为我国绝大多数从事语言研究的人都是纯文科出身的,对社会学和社会语言学的调查统计方法很不了解,因此当70年代社会语言学传入我国以后,从事宏观研究的人不少,而从事社会调查的微观研究工作的人却很少很少。王立同志的这部著作不仅公布了汉语社团

公众关于“词”的语感的调查数据,同时也是不了解社会语言学调查统计分析方法的读者了解和学习社会语言学调查和统计分析的具体方法的很好的教材,因为里边有大量的语言材料,而且作者详细地交代了她是怎样一步一步运用什么理论和方法来设计调查问卷的,是怎样收集和统计资料的,又是运用什么方法来核查调查数据的可靠性的,比起泛泛而论的介绍社会语言学方法的书来说,初学者更容易理解和掌握。这也是我支持她出版这部著作的一个重要原因。当然,社会语言学的调查数据还是没法跟自然科学的实验数据相比的,但是作为一种重要的参考数据还是有足够分量的,完全否定是缺乏科学根据的。过去有个别人完全否认社会调查数据的可信性,那是由于不了解社会调查的科学性和具体操作过程,这部著作的出版也许也有助于消除他们的疑虑。

胡 明 扬

2002年9月18日于北京

目 录

绪 论	1
第 1 章 汉语“词”的认识历程	11
1.1 汉语词观念的萌发(19 世纪末~20 世纪 20 年代)	12
1.1.1 马建忠:汉语文法的基本单位——“字”	12
1.1.2 章士钊:汉语文法的基本单位——“词”	14
1.1.3 刘 复:字是意义的最后的独立的单位	15
1.1.4 陈承泽:“词”乃“虚字”也	16
1.1.5 金兆梓:文法上的一个字并不就是一个个的方块字	18
1.1.6 词是表示观念、文法上有独立作用的一种符号	19
1.1.7 黎锦熙的文法单位体系——字、词、短语、句子	19
1.1.8 小 结	21
1.2 汉语词观念的形成(20 世纪 30~40 年代)	22
1.2.1 何 容:语言的成分是有表意作用的最小的声音单位	22
1.2.2 吕叔湘:复合的词只是最小的表现单位	24
1.2.3 王 力:“意义”包括“理解成分”和“语法成分”	25
1.2.4 高名凯:应当跑到文字后面去看词的构成	27
1.2.5 小 结	28
1.3 汉语词观念的确立(20 世纪 50~60 年代)	29
1.3.1 “字”和“词”是两个不同性质的单位	30

2 汉语词的社会语言学研究

1.3.2 应当建立词的观念	31
1.3.3 确认词的标准的讨论	35
1.3.4 区分词和词素的讨论	42
1.3.5 区分词和词组的讨论	45
1.3.6 小 结	49
1.4 汉语词观念的传播(20 世纪 70~80 年代)	51
1.4.1 界定词的标准逐渐聚焦形式	52
1.4.2 采用韵律成分做词的形式标记	55
1.4.3 现代汉语词的定量研究	56
1.4.4 古汉语复音词的专书研究	57
1.4.5 词的应用研究	63
1.4.6 社会语言学方法的初步运用	64
1.4.7 小 结	65
1.5 汉语词观念的深化(20 世纪 90 年代~21 世纪初)	66
1.5.1 字本位理论——汉语基本结构单位的再认识	66
1.5.2 用“音节基准”处理“两难”现象——双音定词基准	69
1.5.3 韵律对构词的限制	71
1.5.4 词的双音化的快速发展——内在规律和外部推力的相互 作用	74
1.5.5 中文信息处理中的分词研究	76
1.5.6 双字词在心理词典中的储存模式	79
1.5.7 小 结	81
第 2 章 从语感的角度认识汉语的“词”	87
2.1 百年研究的启迪	87
2.2 语感及语感测量	90

2.2.1 语感及其分类	90
2.2.2 语感测量	92
2.3 语感测量的工作程序	94
2.3.1 确定选题和建立假设	94
2.3.2 确立概念及测量方法	95
2.3.3 设计问卷及抽样方案	99
2.3.4 资料收集与数据整理	103
2.3.5 统计分析与结果推论	109
 第3章 汉语社团对于“词”的感知	 114
3.1 语感测量的先导研究	115
3.2 由正词法拼写规则引发的公众词感测量	117
3.2.1 引发语感测量研究的动因	117
3.2.2 研究方案的设计与实施	119
3.2.3 资料的整理与分析	125
3.2.4 首次词感测量的实际意义	129
3.3 两个独立样本的公众词感倾向测量	131
3.3.1 研究方案的设计	131
3.3.2 研究方案的实施	136
3.3.3 资料的整理与描述	139
3.3.4 两个独立样本测量结果的相关分析	143
 第4章 汉语社团的“字”“词”认知差异	 149
4.1 研究方案的设计与实施	150
4.1.1 建立假设	150
4.1.2 设计问卷	151

4 汉语词的社会语言学研究

4.1.3 收集资料	155
4.2 统计分析	156
4.2.1 两组调查对象指认字数结果的统计分析	156
4.2.2 两组调查对象切出词数结果的统计分析	160
4.2.3 同一调查对象分字与切词结果的统计分析	165
4.3 结果推论	168
4.3.1 字是文字单位	169
4.3.2 词是语言单位	171

第5章 汉语构词范式的确立

5.1 “双字格”成为汉语构词范式是汉语社团的共识	175
5.1.1 汉语社团对于“双字格”构词模式的普遍认同	175
5.1.2 汉语社团对于“双字格”构词模式的成功应用	179
5.2 “双字格”成为汉语构词范式是历史选择的结果	182
5.2.1 以联绵词为范例的组合式“双字格”的形成	182
5.2.2 简缩式“双字格”对汉语构词范式的选择	188
5.2.3 汉语构词法发展的必然结果	190
5.3 “双字格”成为汉语构词范式是其内在品质的实现	191
5.3.1 韵律要素与意义要素的完美结合	191
5.3.2 自身成词和参与构词的双重功能	193

第6章 三种结构词感倾向的讨论

6.1 汉语动趋结构词感倾向的讨论	198
6.1.1 学界对于V+D结构及D的认识	199
6.1.2 公众关于V+D结构成词与否的认知倾向	201
6.1.3 三音节V+D结构准词化倾向讨论	205

6.1.4 构建 V+D 结构词模的策略	211
6.2 汉语名方结构词感倾向的讨论	212
6.2.1 学界对于方位词(Localizer)的认识	212
6.2.2 公众关于 N+L(Noun+Localizer)结构的认知倾向	216
6.2.3 方位词身份的确认	220
6.3 汉语数量结构词感倾向的讨论	222
6.3.1 学界关于数词、量词、数量词的看法	222
6.3.2 公众关于数量结构的认知倾向	224
6.3.3 语言研究的分析单位与自然语言的使用单位	226
第 7 章 五次语感测量结果的分析	231
7.1 第 5 次语感测量结果的相关分析	232
7.1.1 第 5 次语感测量的设计与实施	232
7.1.2 地域因素与切词结果的相关分析	234
7.1.3 专业因素与切词结果的相关分析	238
7.2 五次语感测量结果的综合讨论	242
7.2.1 从双音结构词化倾向看“双字格”汉语构词范式	242
7.2.2 从三音结构成词走向看“双字格”汉语构词范式	249
第 8 章 汉语的词与中国社会语言学研究	260
后 记	265

绪 论

汉语的“词”及其研究是一个非常现实的课题。我国传统的“小学”以“字”为单位,从文字、音韵、训诂三方面对一个个形音义结合的具体单位进行了细致深入的研究,但却未能从整体上对汉语的基本结构单位及其存在形式作出理论上的概括(为叙述方便,我们暂且服从语言学界通行的说法,对汉语的这一基本单位以“词”相称)。词的问题早在汉语语言学草创之初即已引发众多语言学者的关注和讨论,历经百年,有关“词”的研究视角逐步拓展,但到底什么是汉语的词,至今仍是见仁见智,由此可见汉语词研究之艰辛。然而,词的问题又时时困扰我们,汉语教学以及自然语言处理中常常遇到需要对具体的字串(特别是双音节字串)作出词与非词的判断,而现有的语言理论常常使我们得到一些似是而非或者相互矛盾的结果,同时,学习者的汉语拼音分词连写结果所呈现的一致性倾向与正词法的某些规则又有较大出入,这便引起了我们对“词”的兴趣与关注,于是就有了我们用社会语言学的方法所进行的关于汉语“词”的研究,有了呈献给读者的这本小书。

各种语言都有其基本的结构单位,普通语言学将语言单位由小到大分为4级:语素(morpheme)、词(word)、短语(phrase)和句子(sentence)。每种语言都有其自身的特点,一种语言的基本单位的具体形式是怎样的,需视那种语言的特点和使用情况而定。汉

2 汉语词的社会语言学研究

语与西方语言有一定的差异,加上汉字记录汉语的独特方式,也就决定了汉语初始单位的研究很容易选择以“字”为单位,这也许更符合古代汉语的实际。但是随着系统的汉语语言学理论研究的深入以及应用语言学的需要,汉语的初始结构单位到底是什么便成为汉语研究的一个重要内容和亟待解决的实际问题。

印欧系语言的基本结构单位比较容易把握,比如英语中一个 ones 的 word,语音上有重音,词长不确定,形态上有变化,书写有间隔。汉语则没这么简单,语音上每个音节都是独立的,每个音节的书写符号也是独立的,而句子的书写则是连成一线没有间隔,这些都给确认及定义汉语的基本单位带来很大的麻烦。自马建忠开始的系统的汉语语法研究,借鉴建立在印欧语基础之上的句子结构分析方法,始触及汉语句法的基本结构单位。马氏所指的句法分析的基本单位显然不是一个音节或一个汉字,但他仍承袭了传统语文研究的术语,将这一结构单位冠之以“字”。可以说,自马建忠的汉语语法研究起,我们就对汉语基本结构单位有了语言学意义上的认识,并在语言系统内的各个不同层面上对这一单位进行分析。但一百多年以来,始终没有得出一个为语言学界公认的解释。

尽管从理论上确认“词”这一单位不容易,但不论是理论研究还是实际应用,汉语的基本单位都是个不可回避的问题,因为汉语的使用中客观地存在着这样一个单位,汉语的研究中需要这样一个基本的分析单位。因此,必须设法认识汉语的“词”,这一任务历史地落在了当代学人的肩上。

汉语的词应当是能够认识的。今天我们认识汉语的词,首先需要 对汉语词研究的理念和方法有一个清晰的了解,即一个世纪

以来,在认识词的漫长过程中,几代学者所使用的确认词和界定词的方法有哪些,其理论依据是什么,给予当代学人的启示是什么,我们应当如何审视这个问题。

百余年来,我们对汉语词的认识大体上采取的是中西结合的研究方法,即在我国传统的汉语研究基础上,借鉴西方现代语言学的理论和方法,从语音、词汇、句法、语义等方面对汉语的词进行描写和分析,力图从理论上对汉语的词作出解释。这些研究都是在语言系统内部不同层面上进行的,其结果往往是根据某一语言要素的特征对词加以定义,于是有了“语法词”、“词汇词”、“语音词”、“韵律词”等等。而随着汉语信息处理等语言应用研究的发展,又出现了“分词单位”、“拼写单位”、“信息处理词”、“工程词”等关于词的指称。

汉语语言学界也从语言的外部系统着手开展了一些与词相关的研究,如文化语言学,研究语言与文化的关系,在一些具体的言语单位与外界事物的对应或某些具体的语言单位产生的条件等方面进行了探讨,不过仍然没能回答什么是汉语的词。言语工程学界则不计较汉语学界关于词的争议,对于自然语言中存在的这个基本单位,语言学界认为它是“词”也好,是“语”也罢,只要便于工程上处理就把它作为一个单位。但这也只能是权宜之计,毕竟言语工程所面对的是自然语言,只有弄清了人脑使用自然语言的机制后,才有可能模仿人脑的机制进行自然语言的处理,所以从根本上说,言语工程需要语言学界对汉语的基本单位进行确认和作出解释。

既然理论上的争议无法确认汉语的词及其存在形式,我们就有必要调整一下观察的视角,走出固有的研究模式,从语言的本质

4 汉语词的社会语言学研究

属性出发,寻求一条有效的研究途径。对词的理解是一个复杂的社会心理过程,把寻求汉语基本结构单位的研究放到汉语使用的大背景中去进行,看看存在于汉语社团成员头脑中的基本语言单位是什么,也许能够达到认识词的目的。但这种研究困难很大。语言是物质的,也是观念的,进行这类研究需要有一套严格的科学的研究方法,需要做大量的实际工作。

通过对汉语使用情况的观察,我们感觉到,众多学者苦苦求索的那个基本的语言单位——词——其实早就存在于公众的语言直觉之中。当然,如果直接去询问汉语使用者“词是什么”,回答也许是茫然的。然而当你拿着一个现成的句子去问他们“这个句子有多少词”,他们中的绝大多数人会毫不迟疑地将一个一个存在于他们观念中的词切分开来。从人们对这一个个具体的词的迅速反映中,我们似乎进入了那个存在于人们大脑中的潜在词库。这个大脑词库中不仅贮存有许多已登录的词,而且还可以随时生成许多未登录的词。这就意味着,人类的大脑词库中积淀了丰富的语言知识,其实是一个语言单位与语言规则库。然而人们对自己已经掌握的语言知识并不是都有明确的理性认识的,但是凭着语言直觉,他们可以随时自如地组织调用这些语言单位和语言规则,在言语实践中下意识地使用它们。

这就启迪我们,虽然从理论上解释什么是词以及词是如何生成的有一定困难,但词是怎样的单位却不难从具体的已然话语中识别出来。从人们对已然话语的词切分结果中,我们可以观测到汉语社团成员头脑中词的基本存在形态及其生成模式。假如我们的调查能够成功地进行,也就有力地证实了汉语的词是客观存在的,因为只有客观存在的东西才能在人们的观念中反映出来。这