

王士元

语言学论文集

◎ 王士元 著



商務印書館

王士元

语言学论文集

（增订本）



商务印书馆

王士元语言学论文集

王 士 元 著

商 务 印 书 馆

2002 年·北京

图书在版编目(CIP)数据

王士元语言学论文集/王士元著. —北京: 商务印书馆, 2002

ISBN 7-100-03220-2

I. 王… II. 王… III. 语言学-文集 IV. H0-53

中国版本图书馆 CIP 数据核字(2000)第 54594 号

所有权利保留。

未经许可, 不得以任何方式使用。

WÁNGSHÌYUÁN YǔYÁN XUÉ LÙNWÉNJÍ

王士元语言学论文集

王士元 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

中 国 科 学 院 印 刷 厂 印 刷

ISBN 7-100-03220-2 / H·816

2002 年 3 月第 1 版

开本 850 × 1168 1/32

2002 年 3 月北京第 1 次印刷

印张 10 1/8

定价: 16.00 元

自序

收入这本书里的多数文章原先是用英文在美国发表的,中国的读者也许不大容易直接看到。当石锋教授建议把文章的译文和中文的文章收集起来出版一本文集的时候,我就欣然同意了。李行德教授最近在《中国语言学报》里(JCL 28, 116—162, 2000)分析了中国和西方的语言学,他说这两个传统的距离,还是太远。我很赞同他的看法。如果这本文集能够为沟通中国和西方的语言学起一点促进作用,我也就心满意足了。

这十几篇文章的写作时间跨越了三十多年,其中有的是源自我早期在语音学方面的工作。有人曾把语音学称为语言学“不可缺少的基础”,这是颇有道理的。书中也有些文章是有关文字的。任何一个人都会对我们独特的文字系统产生浓厚的兴趣,其余的文章主要是讨论语言演变的本质的,特别是从方法论的观点对这个问题进行的探讨。

这些论文涉及好几个学科领域,这正好体现了我的这样一个信念:如果我们的目的是要真正理解语言是什么,那么我们就不能把自己的视野限制得太狭窄,就不能因已有的学科界限而墨守成规。不同学科之间的边界犹如画在沙滩上的线条,随着每一次先进知识的波涛到来,这边界就会发生变化,甚至完全消失。人类的知识,特别是研究语言的知识,应该是彼此相连的,并且最终是相互贯通的。

尤其是研究语言变化,我们不得不借鉴生物学里发展出来的演化论。语言变化跟生物变化有很多相同的地方,这一点达尔文早就讨论过。他在 1859 年出版的书的第 14 章,提出过这样的假设:世界上人群的综合谱系也就是世界上语言的谱系。大家知道一群人在一起的时候,他们的基因是经常交换而混合成一体的,同时,这群人的语言也是经常相互影响而大致相同的,可是要是有一部分人离开主群很久,他们的基因跟他们的语言在新的环境里就会改变,时间越久,差异就越大:语言变得快,几百年后他们的话就可能跟主群的话互不通音了,基因变得比较慢,但是久而久之世界上所有的人群中的差异就逐渐积累出来了。

因此研究基因时所发明的一些方法及模型,例如计算距离以及画树形图,都可以用来研究语言。基因跟语言的共同之处是 Cavalli Sforza 教授几十年来继承达尔文的假设所探讨的焦点,他最近出版的《基因、人群与语言》已有中译本,是很值得参考的。

可是,这两种变化也有一些很基本的差异,也是我们不得不注意的。首先,“适者生存”这个演化论里选择的概念所指的对象是生物竞争,而语言只是寄生在人群里的一种文化系统。一个语言的利或弊当然可能间接地影响一群人的命运,可是人群的繁荣或灭亡肯定有很多更直接的因素跟语言并不发生关系,因此演化论的有些思想不能完全照搬地用在语言上。

另一点基本的差异是传达的方式。一个婴儿的基因是完全从父母得来的——父亲母亲各供给一半染色体,精子和卵子结合的那一刹那就全部规定,不可能再有别的来源,这跟语言的传达很不一样。婴儿的话是由他主动地观察他周围的语言环境,一步一步地归纳出来的。语言环境里的材料,很可能有父母的话,但是来源

繁多,并且是经常变的。家里听到的话,街上的话,老师及同学的话,都会影响他的语言生长。一个人的语言主要部分是在青春期待型的,这十几年的语言环境里的语料,或多或少地都是他语言的来源。

总而言之,小孩学的语言,跟他接触的环境里的语言的差异,一代一代地积累起来,就是历史语言学所研究的语言变迁。要是我们追本穷源,再扩大我们的视野,一直问到数万年前最早的语言,也不过就是一些基本词汇,而目前世界上的如此复杂的语言,还就是经过数千代的小孩学语积累而逐步形成的,所以必须把这几样东西的关系搞清楚,就是小孩学语,语言接触以及语言演化,才能了解到语言的真面目。要达到这个目标,我们有很多很有意义的研究要作。我非常希望中国和西方的语言学家能够齐心协力地在新世纪里求出一些道理来。

这二三十年来由于有大量的计算机运用在建模仿真,我们对演化论的了解又迈进了一大步。霍兰(John H. Holland)1995年的书现在有中译本,由上海科技教育出版社发行,书名里的三个词——隐秩序、适应、复杂——都是目前很多学科正在研究的课题。这些课题再加上混沌论里的一些基本概念,很可能是21世纪的一门重要的新科学。

“适者生存”这个想法在20世纪的演化论里扮了极重要的角色,作了很多贡献。可是单靠它来解释所有的演化似乎是不够的,尤其是一些非线形的复杂现象,包括屡次像冰化成水那样的相位改变。语言究竟是怎样在短短的十万年左右的时间内产生的:从无变到有,从个别的词变到无限多的句子,产生出这么多多种奇妙的语法,这些语言学里最基本的问题希望也可以借助这门新科学来

研究出些道理来。这也是语言学的一个新方向。工作量当然是很大,更希望国内的同行们也会对这些问题有兴趣,来参与这项工作。

书中有的文章是合作的成果,我在这里要感谢北京大学林焘教授,麻省大学沈钟伟教授,阳明大学曾志朗教授,香港城市大学柯津云女士,慷慨诚挚地与我分享他们的劳动和创见。编成这样一本书是一个艰巨的任务,石锋教授花费了大量时间和精力做完了这件事。校对的时候,又由汪锋先生及林千哲先生帮助,商务印书馆的金欣欣先生关照。无论这本书对于语言研究最终会有什么用处,我们都不会忘记这几位朋友曾经付出的努力。

原本,我来香港是为了能够看到她回归祖国的怀抱,身临其境地体验中国人的欣喜与骄傲,打算 1997 年之后就直接返回加州。但是我留下来了,在香港一年又一年,因为我实在不想离开这些志同道合的学生和同事,我们在一起分享着那么多求学的欢愉。所以,最后,我要感谢香港城市大学以及这里的一些新朋友。

王 士 元

2001 年 6 月

香港,马鞍山

目 录

自序	(1)
语言的变异及语言的关系	刘娟 石锋 译(1)
观察历史的三个窗口	焦立为 向倩 译(41)
语言的出现与传播	梁洁 译(75)
竞争性演变是残留的原因	石锋 廖荣蓉 译(88)
词汇扩散的动态描写	与沈钟伟 合作(116)
声调的音系特征	石锋 译(147)
声调感知问题	与林焘 合作(170)
北京话的第三调	与李功普合作 焦立为 译(185)
吴语浊塞音的研究	与沈钟伟 合作(196)
人脑对书面文字的处理	与曾志朗 合作 沈钟伟 译(225)
方言关系的计量表述	与沈钟伟 合作(240)
词源统计分析法、词汇统计学和其他 数理统计方法	陈保亚 周政后 译(262)
语言的起源及建模仿真初探	与柯津云 合作(280)
研究之乐,乐在其中	卢少光 译(299)

语言的变异及语言的关系

一、引言

中国语言学的起源可以追溯到很久以前。众所周知,直到现代,中国传统学问的精髓还是“小学”,而对语言方面的研究正是总称为“小学”的各门学科的主要内容。当柏拉图在争论语言究竟是一种社会规约还是一种神的赐予的时候,荀子也在思考同样的问题,并且有类似的分析。^①

荀子用了三条四言短语来概括他对词的本质的认识,即“名无固宜,名无固实,名有固善。”前两个短语是说“词没有其固有的适宜的声音形式,也没有其固有的意义内容”,这是对语言符号任意性的最早论述之一。而第三个短语则是说明词可以有其固有的内在价值。在我们所发现的每一种语言中,各种声音符号体系都是如此。事实上,广告代理商们对第三点一定是烂熟于胸的。因为他们总是挖空心思去为每一种新产品寻求一个最为有利可图的名称。

在中国传统语言学诸多来自经验的成就当中,我们可以提及扬雄对语言地理学的研究和许慎在公元 121 年完成的字典。我们对 2000 多年前中国语言情形的了解大多是从前者得来的。后面我们会更详细地说到扬雄。在许慎的著作中我们看到第一个明确的分析汉字字形的系统,其中的规则一直延用至今。

在这些早期的开创之后,随之而来的是一个接一个的世纪里面源源不断纷至沓来的语言学成就。这些成就包括沈约对词的声律的分析、宋代在发音分析的基础上建立的韵图、陈第对汉语音韵学的开拓性研究等,在此提及的只是几个主要的成果。纵览一个又一个世纪以来积淀的学术成就,无论以什么样的标准来衡量都堪称一流。所以当我们从一本旨在描写“语言学通史”(Itkonen, 1991:2)^②的出版物中读到下面这样的话时不能不感到诧异:

“从严格科学的观点分析,中国的语言学传统没有多少价值……汉语里相当多的有价值的东西也都借鉴于普遍的(或全球)的观点,这正说明一种发达的文化并不必然造就一流的语言学传统,二者之间没有内在的必然性。”

当然,一个言行毫无教养的人说出如此肆无忌惮的话语是武断的和不能原谅的。而从另一方面来说,我们期望这样的论调会给那些致力于中国语言学研究的学者以反面的激励,使他们在向西方介绍他们的学术领域方面做出更多的努力。

现在,中国语言学已开始全面地与国际学术界接轨。汉语本身具有独特而悠久的传统,又是类型多样的和有史可查的历时演变现象的丰富宝库。有鉴于此,我们完全有理由相信,在不久的将来,中国语言学将不只是吸收普通语言学的成果,还要为普通语言学的发展作出重要的贡献,丰富普通语言学的理论和方法,并将有助于确定普通语言学的发展方向。

我的论题分为两部分。第一部分是讨论从历史时代早期以来中国的语言变异。在这一部分中我将举出一些例证以说明这种变异不仅是语言本身的变异,而且还有使用这些语言的民族的变异。第二部分我将说明通过考察历史关系来分析这种变异的一些新方

法。在此,我将鼓起勇气进入一个尚未展开的领域,我们可以把这个领域称为“语言学体系”。^③

二、汉族和其他民族

我们先看看,关于现在一共有多少种汉语方言的问题,吕叔湘是怎样讲的:(1980:85 页)

“大家都知道汉语的方言很多,可究竟有多少呢?……如果只要口音有些不同,就算两种方言,那就多得数不清……要是一定要语音系统有出入……才算不同的方言,大概会有好几百,或者一二千。要是只抓住几个重要特点的异同,不管其他差别,那就可能只有十种八种。”

上述引文中“特点”这个术语,我在下面对特征状态分析的讨论中,把它译成“特征”。“特征”和“特征状态”在生物学体系中用得很广^④。例如,在汉语语言学中起过重要作用的一个特征就是中古汉语全浊音的演变。这个特征表现在现代方言中有几种不同的状态。例如,中古汉语的 *b 在苏州话里仍保留为 b;在别的地方还有其他状态,包括送气的和不送气的 p。一共有多少种方言,以及都有哪些方言,这要取决于我们所选取的特征。要选出那些最能识别历史亲属关系的特征实在不是一件容易的事情。

我们知道,吕叔湘所讨论的方言是汉语的变体^⑤。这种语言和讲这种语言的民族有着跟历史上的汉王朝相同的名称。这是在华夏辽阔疆域上建立起来的最早的一个长期存在的朝代,并且使得跟我们有密切联系的文化传统结为一体。一千多年来,“汉”的观念在中国人的心目中根深蒂固,比如对于“男人”这个概念就有

“男子汉”、“好汉”和“汉子”等词语。人类学家告诉我们,一个民族用他们称呼男人的词来作为自己民族的名称是很常见的。从所有名字叫做“汉”的民族、朝代,以及河流和城镇来研究它们在历史上的确切联系是饶有兴味的。^⑥

根据最近的人口普查结果,现在中国人口的 90% 是汉族。然而,要理解中国的语言变异,关键是要知道这种汉族占绝大多数的情况还只是出现得相当晚的现象。如果我们按现在的边界线对中国做一个粗略的描述,这块土地上生活着众多的民族,每个民族都有不同的文化和语言传统^⑦。目前,在政府行政区划中正式承认的有 55 个少数民族。实际尚未详细列出的比这个数字还要多。如果考虑到那些居住在汉族地区,千百年来已在某一个方面失去了本来特色的群体,那么这个数字要成倍增加。简而言之,我们今天的各种方言,最初都是以从中原地区^⑧扩散开来的汉语为表层语言,各少数民族的语言为底层语言,这样相互作用、相互影响而形成的。

这些少数民族的力量过去曾经比今天强大得多。回溯过去一千六百多年的历史,大约有一半的时间里,北部中国是在阿尔泰系的统治者建立的王朝的控制之下。元朝和清朝是两个主要的例子,在此之前还有很多其他的朝代。如果翻阅一下谭其骧和他的同事最近在上海编纂的那种历史地图集,我们可以看到有的少数民族曾经有着强大的势力并控制了大片的疆土。《苏武》、《王昭君》、《杨家将》、《武家坡》、《岳飞》,这些历代流传的故事讲的都是汉族对于少数民族处于防御地位的历史事实。

汉族以外的民族所用的名称很混乱^⑨。早期的名称主要是指居住在以汉族为中心的四周的民族。“中国”这一名词早在《诗经》

时代就出现了^⑩。据《礼记》所载,这是指被东夷、西戎、南蛮、北狄包围的一个国家。尽管有证据表明,周代的王室是戎族,楚王讲的是他所出身的蛮夷语,但这些到底是什么民族,我们却不清楚。还有许多笼统的名称,比如“五胡乱华”中的“五胡”,尽管那个时代交战的民族实际上不止五个。对于西北的民族常常是用“番”来指称,如“番邦”^⑪。对于南方和东南的民族,喜欢用“越”来指称,如“百越”。同样的,这个数字“百”也是一个概数,指很多民族。

有些在早期文献上发现的名称已经存在了数百年,如彝和羌。羌字的上半部分是个羊字头。这常常被认为和周朝姓姜的贵族有联系。但是在这数百年中发生过很多征伐和迁徙,我们实在无法认定今天的彝族和羌族确实就是那些在古代具有相同名称的民族的后裔。很多群体在现代已经没有踪迹可寻。他们可能已经离开了这个疆域,也可能已经被征服者消灭或同化了。同时,也有很多现代的群体,我们完全不知道它们历史上和地理上的来源。

三、中国民族的遗传学背景

我们可以肯定的只有一件事,这就是历史上不仅少数民族,也包括汉族,一定曾经发生过大规模的遗传学的交互混合。这在日本的斋藤(Saitou)和他的同事们及上海的赵桐茂和他的同事们最近对中国人口的生物学研究中可以非常清楚地看到。

仔细观察一下上海的研究对于语言的历史方面所具有的意义是很有启发的。这项研究是以在 74 个地点,从约 10000 个人中所获取的免疫学数据为基础进行的。每个点都保证超过 100 个人的有效数据。其中有 44 个点代表汉族人,其余 30 个点少数民族,

包括蒙古族、藏族、朝鲜族等。这些人口的生物学联系可在树形图中表现出来。(见图1)

树形图上最突出的特征就是,跟一个汉族群体有着最密切关系的往往并不是汉人,而是某个其他民族的群体。于是,在树形图中,东北长春的汉族人(15)跟朝鲜族(16)生物学关系最亲近。在青海,西宁的汉族人(23)和蒙古族人(22)关系最密切。在湖北,黄石的汉族人(50)有大理的白族人(51)作为最近的亲族。在湖南,邵阳的汉族人(62)和四川西昌的彝族人(63)关系最密切。很显然,他们在地理位置上的接近起了重要作用。因为一个群体更可能和邻近的群体在语言上和生物学上相互影响。事实上,产生具有相同基因这种更为亲近的关系,常常是以共用一种语言作为最重要的前提。

对于以地理位置作为条件,有一些例外情况令人惊奇。例如,跟江苏南通(33)的汉族人关系最密切的是甘肃的保安人(32),而不是同在江苏的徐州汉族人(31)。同样地,斋藤等人独立进行的研究也表明了徐州汉族人跟藏族人、蒙古族人和朝鲜族人的关系比跟北京汉族人的关系更接近。我们可望在考虑到全部有关的人口学因素之后,会对所有这些生物学关系有更深入的认识。

然而,从最近的全部生物学研究中得出的一个结论是不容置疑的:“汉族”这个概念并不是以任何内部遗传学上的结合为基础的、种族的概念。从我们刚才所谈到的所有这些事实中自然可以产生出这句俗语:“远亲不如近邻”。更确切地说,“汉族”是一个以文化为基础的民族的概念。并且在这种共同的文化里面首先是共同的语言,特别是汉族人所创制发展的汉字。很多世纪以来,尽管不同的群体在口语中有着巨大的差异,可是汉字作为非常重要的

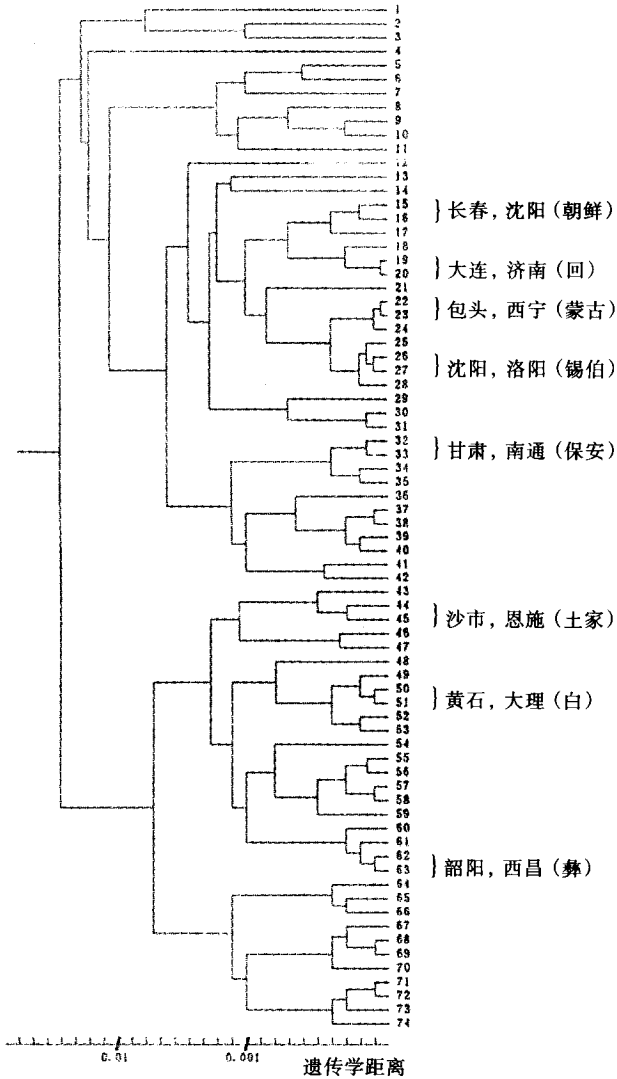


图1 遗传学距离的树形图, 引自 Zhao et al(1991)。在图中加注的每一对都是: 跟一个汉族社群在生物学上相邻最近的不是另一个汉族社群, 而是某个附近的少数民族社群。

纽带把他们联结在一起。同时,汉民族从中原的一个小部落兴起,到今天不仅在中国,而且在全世界占有显著地位,这个纽带必定是起到了极其重要的作用。^⑩

四、来自《方言》的教益

正是这种口语形式的差异促成了中国最早的一种语言学研究,这刚好是在纪元之前^⑪。《方言》这本著作,一般认为是扬雄所写^⑫。它以民众的实际经验为依据的方向是很了不起的。传统的中国学问往往只限于研究贵族的语言和活动,然而,扬雄则记录了来往于西汉都城长安的人们所讲的语言,不管他们是学者、士兵还是普通的老百姓。之前秦朝的统一对于中国的语言也产生了巨大的混同的影响。因此到了西汉时代,人们彼此就更容易听懂了。根据周祖谟的说法,《方言》中的大多数词条都来自先前两个王朝的中心地区:秦语分布在汉朝的都城;晋语在山西地区,反映出更早的周代地域。

然而,我们绝不可以从方言这个词的现代意义来对待《方言》这本书^⑬。在扬雄的时代,当然不可能有语言和方言的区分、语言间的亲属关系这些观念。他所记录的内容一定来自汉语的口语以及阿尔泰语、藏缅语、壮侗语,还有许多其他非汉语和非汉藏语的语言系族中的各种不同形式。

这部著作现在见到的版本约有 12000 字,分为 13 卷。李敬忠(1987)最近为了找出其中非汉语的词,把《方言》的词条逐一做了考察。他发现,全书 13 卷几乎每一卷都有这种词。他是用现在的壮、侗、水、布依、毛南以及其他的南方少数民族语言中的词作比较