

语言是一个复杂适应系统

王士元

(香港中文大学)

我有幸受贵校邀请,与大家一起参加清华大学国学研究院80周年纪念活动,感到非常高兴。大家知道,清华大学在1925—1929年间,一度是国学圣地。王国维、陈寅恪、梁启超、赵元任四位国学大师渊博的中外历史文化知识及中西相结合的治学之风,依然影响着一代代清华学子。在这样的育人环境和人文精神熏陶下,清华大学培育出一批德智兼备的人才,他们的学术成就奠定了中国许多领域里的重要思想基础,相信今后清华大学也定将人才辈出。

在这四位著名导师中,赵元任先生是所认识并受到直接指导的老师,我有幸跟他在美国柏克莱同事几十年,赵先生在东亚系,我在语言学系。大家都知道赵先生对中国语言学的研究具有不可磨灭的贡献,把他称为“现代汉语研究之父”是完全恰当的。^①因为他早期受过自然科学的训练,所以研究语言的时候就带进了一些科学的学术态度与方法,并且一直是从跨学科的宏观视野去思考问题,这是非常值得我们仿效的。赵先生很早就注意到汉语里的声调,以及一些有趣的语言歧义现象,在这两方面都奠定了很好的基础。今天我会在这个基础上从另一个角度开始进行讨论。

一、复杂适应系统——CAS

近年来有许多学科都在研究一种结构,叫做 Complex Adaptive System,也许可以翻译为复杂适应系统,简称为 CAS。^②讨论 CAS 本来是一些物理学家的兴趣,一部分是来自对气候变化的研究。比如说,煮一锅水的时候,炉上的火均匀地给水加热,这样锅里水的温度便一度一度地线性上升。当温度达到一百度的时候,锅里的水会突然非线性地从液体转变为气体。这种从量变到质变的非线性过程叫做相变(phase transition),也可以说是一种涌现现象(emergence),也就是一种 CAS。

研究 CAS 的一个基本观点,就是有些系统能从一个极为简单的初始条件开始,逐步地变化,在某些环

① “现代非汉语研究之父”大家公认是李方桂先生。

② John Holland 是一位研究 CAS 的权威,他 1998 年的书已有中译版,参见约翰·霍兰:《涌现:从混沌到有序》,陈禹等译,上海:上海科学技术出版社,2001年。郝柏林的文集对 CAS 也有很多非常有启发性的介绍与讨论,参见郝柏林:《混沌与分形:郝柏林科普文集》,上海:上海科学技术出版社,2004年。最近 CAS 的观点也用于讨论语言习得,参见 John Holland, *Language Acquisition as A Complex Adaptive System*. in Minett and Wang eds., 2005, pp. 411-436.

收稿日期:2006-05-10

作者简介:王士元,香港中文大学教授,台湾中研院院士。

境里,会产生预料不到的非常奥妙、非常复杂的结果。CAS、相变、涌现、稳定(stability)、平衡(equilibrium),这些概念已经传入很多学科,包括生命科学及社会科学。自然环境里的生物竞争是一个CAS,市场的经济贸易也是一个CAS,目前有很多人用这些概念来分析、建模、预测很多不同的现象。

每个人从很小就学会了语言,因此我们就会把语言视为理所当然,跟呼吸与走路一样,像孔子所说“如入芝兰之室,久而不闻其香”。其实我们的语言的确是一个极为奥妙、极为复杂的系统,而且是一个时时刻刻适应周围环境的系统,所以语言是一个典型的CAS。目前我们无法知道语言是什么时候、怎么样涌现的,但是可以推想到它是从手势、身体的动作、脸上的表情、嘴里发出的声音逐步变化出来的。根据远古的化石分析观察,现代人的身体,包括大脑的形状、体积以及基因组的内容,大约是在15—20万年前形成的,而人类的原始文化,包括艺术、音乐、抽象思维,据考古学的考证,大概是在三四万年前相当突然地发展起来的。很可能在这十几万年的过程中,发生了不同的相变,一套很简单的沟通手势及声音,转变成人类所独有的语言,进而导致文化的发展。自从人类有了原始的语言,由于环境变得越来越复杂,语言也同步变得越来越复杂。在这个复杂化的过程里,语言不断地自我调整 and 适应。这几个不同的相变,究竟是哪些过程,是什么时候发生的,这些都是语言学家研究的范畴,也许图1可以提供一点初步的假设。

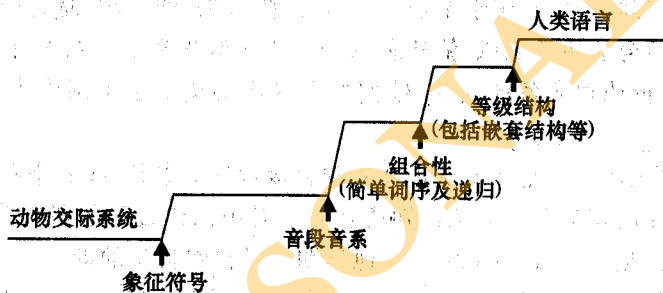


图1 语言形成中的相变(一个初步的假设)

人类目前有五六千种语言,现在我举的例子主要来自汉语和英语。我们的目标是要了解一般语言的复杂系统,了解它在使用过程中怎样不断变化和调整。英语的历史比较短,也比较明确。从上古英语算起只有一千多年。它最初是由北欧的几个日耳曼语混合而成的,而上古英语的时间相当于唐朝末叶,是最早有文字记载的英语,在构词和句法上很像现代的德语。比方说,上古英语名词的构词,区别性(gender)、数(number),也区别于格(case)。可是因为英国长期被别的民族统治,先被北欧的一些航海民族,尤其是丹麦人,后被法国统治,语言在这一千多年的混合里发生了很大的变化。现代英语跟其他的日耳曼语言很不一样,名词基本没有性和格的标记了。尤其是词汇,英语里有大量来自北欧和法语的词。可是有些不规则的特征,例如,英语形容词 good; better 相对于德语的 gut; besser, 英语动词 think; thought 相对于德语的 denken; dachte, 都是有利的证据,说明英德是同源的。

二、音变与歧义

英语的音系在一千多年里也发生了很多变化,而文字往往会保留很多年以前的发音,这一点可以从英文的文字上直接观察到。比方说,上古英语里有很多词头的辅音串,如 gn 及 kn 这两个辅音串,要是前头有个词缀, g 和 k 在现代英语还是发音的,如 a-gnostic, ac-knowledge。但要是这两个舌根塞音前头没有词缀的保护,它们就消失了,动词“know”于是跟否定词“no”就变为同音。像这样舌根塞音的消失,在上古汉语到中古汉语的演化中也类似地发生过,不过因为时代早得多,条件也比较复杂,所以相对来讲,我们对这个音变过程的认识并不像英语那么清楚。

音变当然不只限于辅音,英语历史里,有更大量的元音演变。其中之一是把写为“ea”的元音并入写为

“ee”的元音，因此 meat: meet, heal: heel, steal: steel 都变为同音。总而言之，由于这种音变，产生了一批同音词，这样的例子在每个语言里都不胜枚举。音变就是同音词最主要的来源。接着，同音词也就成为语言里的歧义现象最主要的来源。

汉语的历史比英语悠久得多。最早的记载是三四千年前的甲骨文和金文，也就是上古汉语，诗经的押韵，以及汉字里的声符，都能让我们了解到一些商周朝代的古音。拟构上古音韵这一门学问，可以从明朝末期的陈第(1541—1617)算起，到了清朝，一些学者像顾炎武、钱大昕、段玉裁已经做得很好，很有科学价值。这是一门了不起的学问，很值得我们中国语言学界骄傲。^①

这几千年里，汉语当然也有不少的变化。现今的语言学家公认汉语是跟藏缅语同源的。可是这几百种藏缅语绝大多数都是 SOV 的语序，就是说在一般的陈述句里，动词在宾语之后。可是汉语的语序是 SVO。这个语序应该是汉语跟别的藏缅语言分开以后才产生的。至于这个重大变化的原因是什么，过程是怎样，发生在什么时间，都是一些非常有趣的问题，希望尽早有人把它研究清楚。

我们相信上古汉语的声母是有很多辅音串的。可是在现代汉语的方言里，声母一个辅音串都没有。拟构辅音串一个有力的证据就是汉字里的声旁。比方说，“各”是“洛”的声旁，可是“各”的声母是 /k/，而“洛”的声母是 /l/。又如“京”的声母是 /tɕ/，而“凉”的声母是 /l/，如果进一步追究，就会发现“京”字的 /tɕ/ 在几百年前本来也是个 /k/。几百年中，北方方言发生了一个大音变，叫做颞化，即把 /k/ 变成了 /tɕ/。可是在南方方言里，/k/ 还是保留着。根据同样道理，“监”在南方方言里还是 /k/，而在北方话里已经颞化为 /tɕ/。如果把“监”字上面加个草字头变成“蓝”，或者加竹字头成为“篮”，它也就变成 /l/ 的声母。这种颞化音变在语言中非常普遍。相信每个语言的历史中都至少有过一次这样的音变。不管我们说的是什么语言，大家都是用同样的器官来发音、来听音的。而音变就是藉由运用这种发音器官所产生的变化而形成的。

像刚才举的几个例子，声符本身的声母是 /k/，而加了形符后的字声母却是 /l/，类似的例子也不难找。怎样解释这些字呢？一般语言学家的看法是，这些字在上古汉语里的声母是个辅音串 /kl/，后来其中的 /k/ 在某些语音环境里消失了，只剩下来 /l/，就像刚才谈到的上古英语 /kn/ 一样，现在只剩下 /n/。可是在一些别的字里，消失的却是 /l/，而剩下的是 /k/，在北方话里有时候会变成 /tɕ/。也许用下列公式来表达这两个类似的音变，可以更容易看明白，因为人类的发音器官都大同小异，类似的音变就会经常在不相干的语言里出现：

$$\text{英 } k \rightarrow \emptyset / _n \quad \text{汉 } k \rightarrow \emptyset / _l$$

有趣的是这些辅音串消失的时间，声母简化的时候，也是汉语里声调开始起作用的时候。这两个变化看来是有关系的，因为声母简化的结果是音节种类的减少。语言是一个 CAS，也就是一个会自我调整的系统。利用声调来区别词以弥补消失的辅音串，让我们能够清楚地了解，这个上古音韵系统调整的过程，可能就是这种 CAS 调整的表现。这个问题很值得进一步作更系统的研究。

接着再来看从隋唐时代的中古汉语到现代方言的转变。那个时代的韵书，以及一些后期的韵图，把汉字归纳为四个声调类，即平、上、去、入。宋朝人制作的韵图，又把声母分为清、浊两大类。平声调配上清声母叫做阴平调，配上浊声母叫做阳平调。这样，四个调类乘上两种声母，就有八个调形。语音学告诉我们，声调的主要声学成分是声带抖动的基频，而这个基频是很受声母的清浊影响的。

1973 年我发表了一张关于普通话里四个声调的图片。^②“骂”字是去声，它的声调，一般都说是高降

① 国外语言学界因为没有很多外语资料可以参考，所以对中国的语言学史不够了解。1989 年适逢纪念李方桂先生的会议，我借机在这方面稍微做了介绍。近年来国内出版了不少这方面的书，例如何九盈：《中国古代语言学史》，广州：广东教育出版社，1995 年；何九盈：《中国现代语言学史》，广州：广东教育出版社，1995 年。

② Wang, W. S-Y., The Chinese language. *Scientific American*, February Issue, 1973.

调 51。可是从那个图可以很明显地看到,“骂”的基频是 451,是从 4 升上去的,这就是/m/这个浊声母造成的影响。要是换个清声母,例如,“怕”、“踏”、“萨”,那就没有前头上升的那段,而是一个真正的 51 了。这当然不只是汉语如此,每个语言里浊辅音的音节,都会比它相对的清辅音音节要低。例如英语里的 bin 这个音节,肯定要比 pin 低几个赫兹。这是能根据发这些音时的生理动作来解释的。

51 跟 451 这样细微的音质上的不同,一般人是不会注意到的,甚至连受过语音训练的人也不容易听清楚。可是阴阳乘以平上去入的这八个调形,在音质上肯定是不一样的。隋唐时代的人跟现代人发音的器官不会差得太远,所以我想这些音质上的差异,中古汉语里也一定存在。

对于这个问题我在 1986 年提到过一点间接的证据,主要是在杜甫《丽人行》里,这首诗可以从语义上分成三段,我把里头每一句的韵脚按照这三段分别列出来,如表 1 所示。“u”(unvoiced)代表清声母,也就是阴调;“v”(voiced)代表浊声母,也就是阳调。第一段和第三段句尾押韵的字都是非常整齐地阴平阳平相对,没有一个例外。这不可能是偶然的巧合。我们可以算出这个现象的概率,绝对不是偶然的,说明杜甫这位伟大的诗人,远在等韵学发达以前,已经对语音有很敏锐的感觉。^①说明他不但注意到了当时语音里阴平和阳平的差别,而且还很有效地把这个语音系统里的成分,用在他的诗里。

表 1 杜甫《丽人行》的押韵字及其声调

I			II		III		
1	u	新	7	有	17	v	尘
2	v	人	8	唇	18	u	珍
3	u	真	9	见	19	v	神
4	v	匀	10	身	20	u	津
5	u	春	11	亲	21	v	巡
6	v	麟	12	秦	22	u	茵
			13	釜	23	v	苹
			14	锦	24	u	巾
			15	下	25	v	伦
			16	纶	26	u	曛

资料来源:[6]。

现在由于电子科技的发达,语音上的一些细微区别,可以很快地大量显示以及分析出来。^②最近,我们就把北京话里的四个声调,跟香港话里的六个舒声调,利用一个很大的语音数据库画出来。图 2 里用的 X 轴是基频的斜率,平调就是零, Y 轴是从基频的频率所算出的五度。^③一点一点的声调被椭圆包围起来,好像一个个声调气球。

从这几个声调气球可以看到几件有趣的事情。北京话的四个声调分配相当均匀,在这两维的声调空间里,每个声调都有自己的地方,并且每个气球的中心跟其他的中心距离都差不多。

可是香港话的气球,除了 T1(阴平调),别的五个声调几乎都挤在右下角。研究香港话的人都知道,香港话里没有一个地地道道的降调,这是比较特殊的。因为在中国的方言里,降调比升调更常见。^④如果把

① 当然我们也可以说杜甫所注意的不是调形的区别,而是清浊的关系。

② 参见王士元、彭刚:《语言、语音与技术》,上海:上海教育出版社,2006 年。书中介绍了一些这方面的发展。

③ 详细的说明参见: Peng, Gang and Wang, W. S-Y., Tone Recognition of Continuous Cantonese Speech Based on Support Vector Machines. *Speech Communication*, 2005, (45), pp. 49-62; Peng, Gang, Temporal and Tonal Aspects of Chinese Syllables: A Corpus-based Comparative Study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, 2006, (34)1; pp. 134-154。图 2 出自 Peng, Gang 一文。

④ Cheng, C. C., A Quantitative Study of Chinese Tones. *Journal of Chinese Linguistics*, 1973, 1, pp. 93-110.

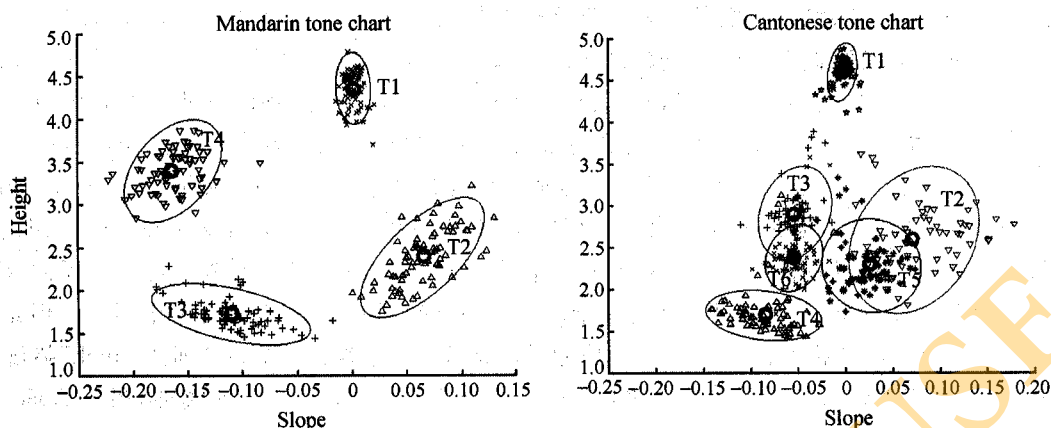


图2 普通话及香港话的声调气球

大批的数据量化成这些气球，就能把事情看得更清楚、更生动。在现在的香港话里，区别还是很清楚。但是 T2 跟 T5 的距离，以及 T3 跟 T6 的距离已经很小。我们做过一些初步的听觉实验，在有些词里面，这两对声调已经根本不分而变成同音了。

刚才谈到上古汉语到中古汉语的声母辅音串的消失，以及声调的兴起以弥补这个消失，是一种 CAS 适应的现象。现在来谈一下韵尾。在香港话里，一套 -p、-t、-k 的塞音韵尾保留得相当完整。像“十”字的 -p，“八”字的 -t，还有“六”字的 -k。可是在北京话里，这些人声字都丢了韵尾，转移到别的声调里去了。“十”字本来是阴入调，现在在北京话里是阳平调，“六”字本来也是阴入调，现在在北京话里是去声调，而“八”字本来是阴入调，现在在北京话里是阴平调。这些就是传统音韵学里所谓“入派三声”的例子。而每当一个人声字失去它的韵尾而并入另一个声调，就会增加一批同音字。

大家也许看过赵元任先生一个有名的小故事：施氏食狮史。^① 这个故事是只用“shi”这个音节配上不同的声调硬造出来的，整个故事生动地说明了北方话有很多同音词。可要把这个小故事里的人声字都找出来，并且把它们按照韵尾 -p、-t、-k 来分类，那么在词汇上的同音歧义就会减少很多。

语言是一个沟通工具，因此一定要适应沟通上的需要。若要弥补这些韵尾的消失，以减少同音词的混淆，有什么办法呢？一个办法就是给单音节的词加上另一个音节。两个双音节的词完全同音的机率当然会低得多。把北京话和香港话对比一下，就能看到这种适应的趋势。很多语义在普通话里要用双音节表示，而香港话只要一个音节就够了。

三、语法与歧义

词的同音跟句子的同音都是歧义现象，而句子的歧义往往是由于词的同音而产生的。早在 1948 年，赵元任先生就提出了“鸡不吃了”这样的例子。这个句子的歧义，来自几个不同的因素。一个因素是主语可以省略，另一个因素是宾语可以移到前面。两个因素加在一起，就可以从“张三不吃鸡了”省略到“不吃鸡了”，再移位成“鸡不吃了”。这样就跟用“鸡”作主语的“鸡不吃了”变成完全同音了。^②

从词汇的角度看来，还有一个因素，就是“鸡”这个词的语义延伸。“鸡”既能用来代表一种家禽，也能代表家禽的肉。同样的，也可以说“鱼不吃了”有歧义，因为“鱼”可以指一种动物，也可以指这种动物的肉。

① 赵元任：《语言问题》，北京：商务印书馆，1968 年，1980 年再版，第 149 页。

② 有些语言的主语不能轻易省略，或者宾语前移的条件要比汉语苛刻得多。像在英语的“The chicken is not eating”这样的句子里，“chicken”只能做主语，因此句子就没有歧义了。

但是,如果说“牛不吃了”或者“马不吃了”,就没有这样的歧义了,因为“牛”和“马”只能指动物,而不能指这些动物的肉。例如,能说“我们今天吃鸡”或者“我们今天吃鱼”,但是不说“我们今天吃牛”而要说“我们今天吃牛肉”。有些语言就是用不同的词来表达动物与肉。英语在这儿也特别有意思,指动物的词“cow”是古英语传下来的,而指肉的“beef”却是来自法语的借词。英语里很多跟饮食有关的词来自法语。

现在再谈一个同音词的例子。在讨论英语的文献里,有一个常用的句子,就是“I saw her duck”。这个句子的一个意思是“我看见她的鸭子”,另一个意思是“我看见她躲开”。造成这个歧义的原因之一,当然是名词的“duck”(鸭子)跟动词的“duck”(躲开)是同音词。这一对词的同音当然不是语义延伸,而是偶然的。要是换一个名词或动词,歧义就不存在了。例如“I saw her elephant”就只有一个语义。但是,与此同时,“her”这个词也要有两个搭配的功能。一个功能是属格代词,这样就可以跟名词的“duck”搭配;另一个功能是受格代词,可以跟动词“saw”搭配。这样“her”就可以一方面作“saw”的宾语,另一方面也作动词“duck”的主语。换一个代词,也就没有歧义了。比方说,“I saw his duck”,“duck”只能是名词,而“I saw him duck”,“duck”就只能是动词。所以也可以说这两个不同格的“her”是同音词。

与此相对的例子在汉语里也能找得到。英语中“his”跟“her”的区别,是古英语性和格的区别在现代英语构词中的一点残留,汉语里当然没有。不过汉语里经常要用“的”来表示属格。像“他的鸭子”,说“他鸭子”就有点奇怪。只有使用频率很高的词组,像“他爸”、“他姐”、“他家”等,才可以把“的”省略掉。这样有时候就会造成一些同音歧义。比方说“我看见他 jia⁵⁵了”,在这里“jia⁵⁵”可以是名词“家”——我看见他家了;但我们也可以理解为动词“加”——我看见他加了。造成这个歧义的因素跟英语相同,一个是因为“家”跟“加”是同音词(jia⁵⁵),另外也因为“他”可以是属格代词,也可以是主格及受格代词。同音词与语法歧义是每个语言里都有的普遍的现象。几十年前,美国政府大幅投资,希望能让计算机来做自动翻译,把一种语言译成另一种语言。那时候我在柏克莱也主持了两个项目,分别把俄语和汉语译为英语,往往一个较长的汉语句子能译成好多英语句子,但其中只有一句是正确的。可是一直到现在,还是没有人能够突破歧义这一瓶颈。这也充分说明了语言是一个非常奥妙的 CAS。

四、语言起源与建模研究

我们当然不可能把时间倒转,回到几万年前,来直接观察语言究竟是怎样涌现、怎样演变,有哪些相变,这些相变是多久以前发生的。光阴是唤不回的,但我们可以用仿真的方法,在计算机上探索一些相关的有趣问题。^①

这几年来我们做了一些初步的工作,用程序先设定一群个体(agent),每一个个体有两个方阵,一个是发音表达的方阵,一个是听音取意的方阵。这两个方阵内部都包含语音和意义两种元素,若干的音义之间形成各种可能的配对,每种配对都具有一定的概率。在实验开始时,这些概率都是随机提供的,没有任何系统。这就代表原始人群里,每个个体都有自己独特的音和义的配合,及独特的语言,问题是怎样才能让这一群个体透过互相沟通,产生出一套共同的语言。战国时代荀子说过,“名无固宜”,但是可以“约定俗成”,实验可以说是替荀子的这句话做了个小验证。用 CAS 的术语来说,怎样从混沌变为有序。

沟通开始时,随机抽取两个个体,比方说:个体张三和个体李四。张三在自己发音表达的方阵中抽取某个语音 U_1 来表达某个意义 M_1 ,李四听到 U_1 ,就会在自己听音取意的方阵中抽取对应的意义来理解这个语音。如果李四恰好在听音取意的方阵中选择了 M_1 ,便说明两个个体沟通成功, $M_1 - U_1$ 在两个方阵

^① Gong, T., J. W. Minett, W. S.-Y. Wang, J. Ke and J. H. Holland, Co-evolution of Lexicon and Syntax from a Simulation Perspective, *Complexity*, 2005, (10)6, pp. 50 - 62; 王士元:《演化语言学中的计算机建模》,《北京大学学报》2006年第2期,第17—22页。

啊本来配对的概率都将上升。如果李四在听音取意的方阵中选择了其他的意义例如 M_2 ，则说明两个个体沟通不成功， $M_1 - U_1$ 和 $U_1 - M_2$ 在各自的方阵中本来配对的概率都将下降。通过几千几万次的沟通后，慢慢地这两个个体的两个方阵之元素和概率逐渐相同，就慢慢形成一个共同的音和义联系的词汇群，这也就达到了荀子所说的“约定俗成”。图 3 就是由这样的计算机仿真所画出来的。^① X 轴是沟通的次数，Y 轴是这个群体之中的沟通度。我们可以看到在这个实验中，开始沟通后有很长一阵子沟通度一直都偏低，可是在 3 000 次左右，沟通度突然上升，这就是相变，仿佛水煮开了，这个原始人群也开始建立他们的语言了。

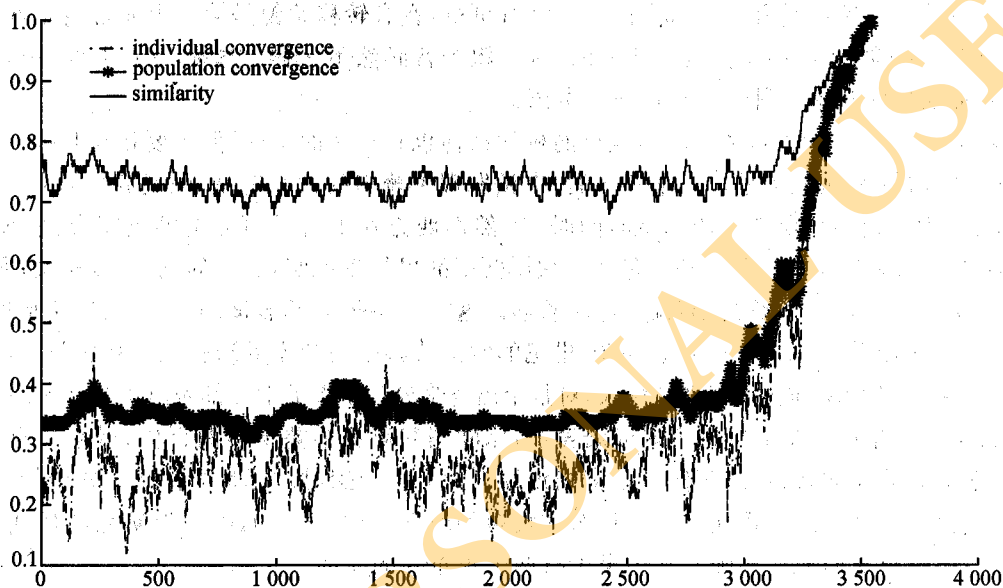


图 3 词汇形成中的相变^[1]

语言是能迅速沟通的工具，几十秒的时间我们就可以说出几十个词。但仅仅依靠词汇来表达是不够的，而时间是一维的，只能依靠词汇出现的前后次序来表达更多的意义，这就是语法。现在我们也开始这方面的研究：如何让共同的词汇，通过相互的影响，达成用相同的次序来表达同一个意思。比方说用“狗”、“咬”、“人”这三个共同的词汇，来表达被咬的是人这个意思。刚开始不同的个体用不同的次序来表达这个意思，三个词出现的次序可以是“狗咬人”、“人咬狗”、“咬人狗”、“咬狗人”、“人狗咬”、“狗人咬”六个次序，不同的个体刚刚开始沟通时并不知道谁被咬，后来慢慢地通过多次沟通尝试，用刚才多个个体沟通的方法达成统一的认知，形成统一的 SVO 格式“狗咬人”，这就形成了简单的语法。随着沟通内容的增加，语言系统越来越复杂，简单的语法也不够用了，慢慢地又产生了更加复杂的语法。共同词汇的形成是一个相变，语法的产生又是一个相变。

在语言学里，用建模仿真来研究语言的起源及习得，是一条比较新的途径。西方学者在这个领域起步比较早，我们则是在 2001 年发表了第一篇文章，介绍这个途径。^② 现在已有几篇博士论文完成，如香港城

① Ke, J., J. W. Minett, C. F. Au and W. S.-Y. Wang, Self-organization and Selection in the Emergence of Vocabulary. *Complexity*, 2002, (7)4, pp. 1-14.

② 王士元、柯津云，《语言的起源及建模仿真初探》，《中国语文》2001 年，第 282、195—200 页，收入《王士元语言学论文集》，北京：商务印书馆，2002 年，第 280—298 页。

市大学的柯津云、区靖邦和南开大学的陈瑞雪。^① 目前香港中文大学的龚涛及黄俊杰正在做这方面的博士研究。刚才谈到的语法次序涌现,正是龚涛的论文题目。

五、语言学的视野

用科学的精神来研究语言,中国人算是比较早的。明朝末年的时候就有陈第。之前的人们谈论到语言的时候,总认为古人的东西才是好的,而非议当时新出现的语言变化,说这个不对,那个错了。但陈第在他的《毛诗古音考》中说道:“时有古今,地有南北,字有更革,音有转移。”他是第一个非常清楚地说明,因为有时时间和空间上的跨度,所以语言总是变化着的,古音跟今音必然不一样。这是一个非常了不起的觉悟。所以中国人的确很早就开始用科学的精神来思考语言了。

正当陈第写作他的《毛诗古音考》之际,欧洲的科学精神也有很大的发展,但主要是在其他一些领域:那时伽利略正在发明望远镜观察天象。几十年之后,牛顿发现了著名的力学定理,顾炎武(1613—1682)的《音学五书》就差不多跟牛顿的《定理》(Principia)同时。^② 然而西方在语言学研究上的真正突破,要算英国人 William Jones 1786 年在印度所作的讲演。他第一次用概率的思想来证明,亚洲的梵文、波斯文,跟欧洲的希腊文、拉丁文都是来自于同一个源头,那就是古印欧语。这个发现非常了不起,很多人认为这是现代语言学的开端。从 Jones 开始所建立的比较方法,在 19 世纪的时候又跟生物学家的演化论互相影响、互相丰富,使得西方的学者对语言的探索和认识越来越深入,对整个语言学界乃至人类学界都做出了非常大的贡献。

中国人很早就开始用科学的精神来研究语言了,比欧洲起码要早差不多两百年,后来又有像顾炎武、钱大昕、段玉裁那样的大学者,对语言做了很多细致深入的分析和探讨。可是为什么现在中国的语言学研究,跟西方比起来有那么大的差距呢?又为什么在短短的时间之内,西方的语言学研究能如此迅速发展,不但迎头赶上,还超越我们那么多呢?其中至少有两个很重要的原因。

第一,中国传统的学者在语言的研究上只重视我们自己的汉语,而西方的学者非常注重研究各种不同的语言。对语言学来说,一个语言重要与否有它内在的标准,跟它的人口多少、科技经济的强弱或一些其他的临时文明状况,都没有很大的关系。^③ 陈第、顾炎武、钱大昕、段玉裁这些非常聪明的明清学者,他们研究的都是汉语。然而世界上有六千多种语言,如果我们只了解其中一种,那是远远不够的。语言的研究最重要的是一种一般规律性的研究,就是说,我们要通过研究很多不同的 langue,来获得对 language 更深入的认识。^④ 只有观察的语言多了,我们才能站在更高的位置、具有更广的视野,对整个语言学体系,或者对其中的某一种具体的语言,有更详尽的观察。西方在这方面是做得很好的,在 20 世纪六七十年代,仅柏克莱一所学校就开设了 70 多种语言的课程。当时从赵元任先生那儿获悉此事,我是既惊讶又敬佩,这也是吸引我从俄亥俄迁到柏克莱的一个原因。他们不但要了解自己境内的语言,还派人到世界各地去研究各种特别的语言。所以我们一定要开拓视野,这是非常重要的。

① Ke, Jinyun, *Self Organization and Language Evolution: System, Population and Individual*. City University of Hong Kong. PhD dissertation. 2004. Au, Chingpong, *Acquisition and Evolution of Phonological Systems*. City University of Hong Kong. PHD dissertation. 2005. 陈瑞雪:《语言系统演化与语言习得建模》,南开大学博士论文,2005 年。

② Hu, Shih., *The Chinese Renaissance*. Chicago: University of Chicago Press, 1934.

③ 例如 Sapir Edward 早就很生动地说过:“When it comes to linguistic form, Plato walks with the Macedonian shepherd, Confucius with the heading-hunting savage of Assam.”也就是说:论及语言形式,柏拉图与马其顿的牧羊人、孔子与阿萨姆猎人头的野人,都可以平起平坐。参见 Sapir Edward, *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World, 1921.

④ 法语的 langue 是指个别的语言,而 language 却是人类语言的总称,这个区别是很重要的。语言学的目标当然是要了解世界上所有 langue 之中的异同,而基于这种了解来建立对 language 的理论。

第二，西方的语言学虽然起步比较晚，但非常注重跟其他学科的沟通与合作，采用跨学科的研究方法，所以它的研究领域得以不断拓展，对于语言的了解和认识也更加深入，获得很多新的发现。西方人在做研究的时候，并不把学科跟学科之间的界线划得那么死。比方在19世纪，达尔文的演化论问世的时候，有位非常出色的语言学家 Schleicher，经常跟达尔文通信。Schleicher 看了 *Origin of Species* 之后非常兴奋，很快就写了一本小书，利用达尔文的演化论来分析语言。可见在西方，生物界、语言学界还有其他的学科，很早就有非常密切的互动。尤其是近几十年来，特别在遗传学、神经学、心理学以及人工智能几方面，有许多空前的大突破，他们得到的一些新知识，对语言学都有很大的启发和助益。

因此，一方面研究语言的视野要扩大，要注意种种不同的语言。汉语和英语当然是很重要的语言，可是其他语言里的现象也不能忽视，尤其是我们自己国内的很多少数民族语言。另一方面，我们一定要有跨学科的研究态度。我经常鼓励朋友们要 work locally, think globally, 也就是大处着眼小处着手，这样才能获得较好的学术成果。

19世纪欧洲在语言学的研究上成果斐然，20世纪美国又占尽优势，而在这一二十年内，欧洲的语言学又可能急起直追，居于领先的地位。虽然目前西方语言学的研究比我们广、比我们深，但悠久的历史给我们留下了宝贵的研究素材。^① 我们现在应该利用这些宝贵的历史资料及现存的语言与方言，结合现代科学和先进的研究方法，使语言学研究迎头赶上，并且要比西方做得更好。我们要好好思考当今的语言研究到底该往哪个方向走，才最有希望能突破和创新。相信21世纪的语言学研究，中国有能力做得更好。

（该文为作者应黄国营教授之邀于2005年9月在清华大学做的报告，由陈卉、蔡雅菁整理。本文的思想是在香港中文大学的语言工程实验室(LEL)与同事一起研究的，包括 James Minett、李行德、帅兰、彭刚、郑洪英和龚涛，尚春峰、汪锋曾对内容提出修改意见。LEL的资助主要来自香港地区政府研究基金会(RGC)及台湾中研院，一并致谢）

参考文献：

- [1] 黄国营. 现代汉语的歧义短语 [J]. 语言研究, 1985, 1.
- [2] 刘丹青. 语言共性与歧义结构 [A]. 中国语言学的新拓展 [M]. 香港: 香港城市大学出版社, 1999.
- [3] Chao, Y. R. *Ambiguity in Chinese* [M]. Reprinted in *Aspects of Chinese Sociolinguistics*. Stanford University Press. 1976.
- [4] Gell-Mann, Murrey. *Language and Complexity* [M]. Minett and Wang eds. 2005.
- [5] Minett, J. W. and Wang, W. S-Y. eds. *Language Acquisition, Change and Emergence; Essays in Evolutionary Linguistics* [M]. Hong Kong: City University Press. 2005.
- [6] Wang, W. S-Y. *A Note on Tone Development* [M]. *Papers in Honor of Wang Li on His Eightieth Birthday*. Hong Kong: Chinese Language Society. 1986.
- [7] Wang, W. S-Y. *Language in China; A Chapter in the History of Linguistics* [J]. *Journal of Chinese Linguistics*, 1989.
- [8] Wang, W. S.-Y., J. Ke and J. W. Minett. *Computational Studies of Language Evolution* [M]. Huang C. R. and W. Lenders eds. *Computational Linguistics and Beyond*. Academia Sinica; Institute of Linguistics. 2004. 俞士汶, 黄居仁主编. 计算语言学前瞻 [M]. 北京: 商务印书馆, 2005.

(责任编辑: 匡云)

① 我们当然不要把“中国语言学”这门学科看得太窄，语言是人类的特征，因此是一门纯粹国际性的学科。我们的目的是要利用我们手头上的材料，经过充分的分析与思考，让我们更进一步地了解人类本身。