

# 中国的语言与民族<sup>1</sup>

王士元<sup>2</sup>

## 1.

中国地形的主要特征，是六千万年前两大板块撞击的结果。印度次大陆撞击亚洲大陆，形成了高耸的喜马拉雅山及与之相连的高原，并造成中国这种不断向东倾斜直到太平洋的阶梯状地形。正是这种阶梯状地形，使中国主要河流均向东流淌，即诗人们自古就咏唱的“东流水”<sup>3</sup>。

来自印度洋的暖湿气流受到这些高山的阻挡，使得中国西部的大部分地区地势高且干燥。在近一万年间，也就是农业出现以来，大部分人类活动都集中在中国东部地区，那里有充足的降雨及可用于农业耕作的河谷。中国的可耕地一直很稀少。最近一报导中的估算是9%，这个数字与传统农民歌谣“三山六水一分田”极为接近<sup>4</sup>。

西部-西北部的高山沙漠，东部-东南部的海洋，使中国在很大程度上处于隔离和自治的状态。在用马来作交通工具之前，这些自然障碍极大地限制了本地区与早期其它文明地区的交往。

目前，就近代人类的起源(近十万或二十万年前出现的人类)，在研究人类进化的学者中存在着较大的争论<sup>5</sup>。占有优势的观点，特别是遗传学家们所维护的观点，认为人类都是由非洲迁移过来的，这些来自非洲的移民之所以能够成功地征服世界，是因为他们拥有一种特殊的优势。

---

<sup>1</sup> 本文的初稿发表于台湾国立政治大学举办的第五届“中国语言及语言学国际研讨会”第一次会议上，英文原文刊于台湾中央研究院语言学研究所筹备处出版之《中国境内语言暨语言学第五辑：语言中的互动》(1999)。会议时得益于朋友们的讨论，译成中文稿时又增添了些新的资料，并先后有刘娟、汪锋及蔡雅菁的协助，在此一并致谢。

<sup>2</sup> 香港中文大学，台湾中央研究院，京都国际高等研究所。此项工作由香港研资局及台湾中研院赞助。

<sup>3</sup> 事实上，长江是一个特例。长江上游与澜沧江、怒江上游平行，如果不是受高地的阻碍，如在云南石鼓一地，它就会向南流入东南亚。在那儿，长江掉头向北，形成一个巨大的U字型，继而流向东方。长江蜿蜒向东海流去，一路上纳百川，添湖泊，成为亚洲第一大河流，滋润着辽阔的中国大地。一位作者最近对此作了生动的描述：“如果没有转弯，如果水流一直向前，那么它就会无情地流出中国...如果没有石鼓地区石灰石山群的阻挡，就不会有今天这条令人难以想象的河流，浩浩荡荡地穿过中国的心脏。如果没有这样一股汹涌澎湃的气势，中国是很难想象的。”Winchester (1996:3)。

<sup>4</sup> 见 Prosterman et al. (1996:90)。中国人口与耕地的比例仅为每人0.11公顷，而美国的人均耕地比例几乎是中国的七倍。

<sup>5</sup> 争论围绕着智人类起源的问题。关于人类血统中早期的种类，即原始直立人，起源于非洲的观点，是无可争议的，他们早在一百万年前就移居他地。见 Wu and Poirier (1995)。

许多持这种观点的人都坚持认为，这种特殊优势就是语言，这并不意味着被非洲征服之前，人类没有言语交际的初级形式，因为即使是猴子，也有成套的呼叫方式。此处是指与我们今天所使用的相近的语言系统，特别是指音段音系学所支持的系统<sup>6</sup>。这种观点暗示单一起源的构想，即发明了语言的非洲民族能够取代世界上所有其它民族<sup>7</sup>。正像遗传学家用粒线体 DNA 及 Y 染色体做研究一样，最终我们应该能够将所有语言追溯到单一的非洲源头。

如果这个构想正确的话，那么中国语言可能在非洲附近地区有远亲。的确，近期一些俄罗斯语言学家提出的德内-高加索语(Dene-Caucasian)的假设，也是以此为特征的，这是完全独立于遗传学家所做的另一项研究。随后我们将看到，他们声称汉藏语系不仅与高加索及里海沿岸的语言有关，甚至与西欧的巴斯克语(Basque)也有联系。

另一方面，反对“起源于非洲”的观点，则观察到地区连续性的大量事实，如中国科学院的吴新智，曾报导在中国发现的骨化石，表现出人体结构上发展的连续性，可从当前人类一直追溯到五十万年前的北京人。如果中国古代居民确实被来自非洲的移民完全消灭了，那么这种连续便是不可能的。

在我看来，全世界都被来自非洲的移民所完全取代，是一种太简单的想法，比较可能的是，这些移民与当地人的基因混合的比例，随着地区的不同而变化。吴新智等人提出的证据似乎说明，现代中国人在某些人体构造上的特征，可以追溯到相当久以前。

## 2.

虽然在围绕现代人类起源的问题上存有争议，但化石证据足以清楚地说明，很早以前在中国确实存在连续的灵长目动物的活动；见 Wu 和 Poirier (1995:4)。云南的元谋（58 号遗址）和北京周口店（62 号遗址）的著名人类化石，属于一种叫原始直立人的较早人种。他们可能没有音段式的语言。另一方面，在陕西大荔（8 号遗址）附近发现了距今约 200,000 年前的保存良好的头骨，该头骨也表现出很多现代人的特征，张光直(1986:53)曾做过这样描述：

“距今 200,000 年的铀系列，说明大荔头盖骨的年代紧接周口店之后，并为研究该地区人类从直立人向智人转变，提供了一个很方便的时间上的开端。”

一系列不同地区的考古发现，为研究古代中国各民族提供了线索。最早的陶器似乎发现于广西的甑皮岩中，约在距今 7000 到 10,000 年前。陶器的发现表明，人类的生活方式已脱离流浪或游牧而开始定居。水稻种植的最早记录出现在长江中游地区，大约在同一时间里。

---

<sup>6</sup> 详细讨论见 Wang 1998a。

<sup>7</sup> 关于不同观点的讨论，从严格的概率上支持多源说，参见 Freedman and Wang 1996。

张光直(1986:235)发表的一系列地图, 适切地捕捉了上述发现所提供的人类历时演变的全貌。早期农业时期的古代中国, 只有孤立、地域性的文化发展, 从东北角一直延伸到台湾, 遍布国土的东半部。到了距今 7000 年前, 较为成功的地域发展足以使这些文化扩展, 以致彼此相连。重大历史事件大约发生在 6000 年前。我们现在还不清楚, 这些事件的导火线是成功的农业生产所带来的人口扩张, 还是类似洪水的自然灾害。总之, 6000 年前是一个大规模融合的时期, 文化开始在相互间发挥强有力的影响, 以致张光直将这一融合的结果描述为“中国的开始”(initial China)。

这样一个如此大规模的文化工具及风俗习惯的扩散, 只能靠小规模、即由迁徙的人流一波接一波地传递而进行。的确, 这就是我所猜想的发生在中国 6000 年前的扩散, 它不仅仅是文化工具的扩散。此外, 还有一些建立在大量基本词汇分析(Wang 1998b) 基础上的初步证据, 也说明在这一时期, 即 6000 年前, 庞大的汉藏语系开始分化, 由此产生我们今天所拥有的几百种东亚及东南亚语言。荷兰语言学家 George van Driem, 近年来研究了许多喜马拉雅山区里的民族语言, 提供了不少宝贵的第一手资料(2001)。语言在地理上的分界, 往往跟政治分界是不一致的, 这些新材料让我们对藏缅语与汉语的关系, 有更深一层的认识。孙宏开主编的《中国新发现语言研究丛书》, 从 1997 年开始, 陆陆续续出版了约 30 册。这些专家著作也都是非常珍贵的材料。这套丛书的详细书目, 可以在网上浏览: <http://www.ethnic-languages.org.cn/sunhk2.htm>。

语言跟语言接触, 久而久之一定会产生混合的现象, 时间越久, 则混合越深入、越普遍。换句话说, 一个语言系统内的种种成分, 有两类不同的传递来源。有些成分是纵向传递, 由系统内部传下或发展出来; 另外一些成分来自横向传递, 是从接触的语言传来。我们得区分这两种来源, 才能正确地了解一个语言的结构, 并将它科学地归类。在这方面, 近年来, 国内有些杰出的研究, 包括陈保亚的基本词汇分层次, 汪锋的借词分层次, 意西微萨·阿错的混合语言研究。美国芝加哥大学的 Salikoko Mufwene 对语言接触的看法, 他所谓的语言演化的生态(2001), 跟国内这方面的研究基本上非常相似。

尽管考古学家不能把我们带到比商朝更久远的年代<sup>8</sup>, 历史学家也不能证实比公元前 841 年更古老的时期, 但是古代文献中有很多神话传说, 可以为我们提供有价值的线索。一方面对神话的诠释必须十分慎重, 另一方面, 一些神话传说确实表明, 中国民族的多样性可以追溯到其形成时期。有一个著名的传说, 讲的是炎帝求助于黄帝, 然后共同打败了由蚩尤率领的九黎部落。据推测, 这场战争发生在现在河北涿县附近, 即当时黄帝的统治区域, 那是中原的中心地带, 即汉文化的发源地。

---

<sup>8</sup> 虽然对夏朝的考古遗址至今仍没有统一的说法, 但一些观点倾向于认定二里头就是其代表之一。

之所以习惯把汉民族称作炎黄子孙，是因为据古书《国语》记载，炎黄本是兄弟<sup>9</sup>。九黎中的“九”字，是最大的个位数，但不能按照字面意思去理解，它仅仅是指有许多部落卷入了这场战争。另一方面，“黎”这个表示民族的术语，一直沿用至今。如今，主要居住在海南的黎族，用一个清声母的边音称呼自己[hlai]。然而却无从确定它们是否是蚩尤九黎的后裔。

另一些类似的证据也说明了大量的民族多样性，并进一步阐述了两个事实：即中国一直存在着多种民族，且中国境内的语言和民族，有着极为复杂的互相融合的根源。据清朝初期顾祖禹(1631-1692)的论著估计，在夏朝初期有 10,000 个国家，商朝初期约有 3,000 个国家，周朝初期有 1,800 个国家<sup>10</sup>，当顾祖禹用“国”这个词的时候，他并不是指当前世界上这些政治职能完备的国家，很可能他所说的国，多半只是一些规模、凝聚程度及语言亲密关系各异的部落。

且不追究顾祖禹文中数字的精确性，但有一点是清楚的，即中国的遗产充满多样性，且今天的汉语，涵盖了多种曾在这块土地上使用过、竞争过的语言。此外，通过几千年的密切交往，在汉民族与其相邻民族之间，也一定广泛地存在着语言的互动。由于与来自语言内部的演变有所接触，亦即由于有横向传播而非纵向传播，要想理出这些影响的来龙去脉是很困难的。

这困难一部份是因为历史文献中，有大量未予定义的种族名称，一部分则来自几千年间各民族的无数次迁移。一个民族的名称，往往不同于这个民族所说语言的名称。比如，巴玛(Bama)的瑶族所说的语言实际上是苗语，而海南的苗族却说瑶语<sup>11</sup>。民族名称也常常随行政区域的不同而变化。如云南的哈尼族与东南亚的阿卡(Akha)族是同一民族。然而，虽然有诸多的困难，这一方面的调查，仍对我们了解中国各语言的出现及其演化提供线索。

除了语言分析、文献研究、考古发掘、基因分析，还有一条重要的途径，可以让我们了解过去，那就是亲自下乡做田野调查。1920 年代，中研院成立后，在西部与西南部的研究上，做了些开端的工作，由李方桂、芮逸夫、凌纯声等学者做先锋。之后，早期有罗常培的小书(1950)，指出了方向。此后几十年，才有周振鹤与游汝杰的研究(1986)、邓晓华的导论(1993)、潘悟云主编的文集(2002)、及中研院的李壬癸对南岛语里某些词汇与文化关系的分析。近年来中研院的王明珂又作了不少深入的田野调查工作，特别是在四川、云南一带(1997, 2003, 2006)。从他的讨论里，我们可以了解一些民间传说的学术价值，并且深刻地感受到民族性的复杂，及辨别民族的困难有多高。

---

<sup>9</sup> 关于这些神话传说及其与历史语言学的延伸论述，详见徐(1991:34)，该文以俞(1980)的调查为基础。

<sup>10</sup> Chang (1983:27)曾引用顾祖禹文中的数据。

<sup>11</sup> 详细讨论见毛宗武等(1982)。一些瑶族地区人说拉珈语(Lajia)，属侗台语族，而另一些瑶族区已用汉语方言取代了自己的语言。

认真用这样多学科视野来研究人类语言及族群的，最早应算是史丹福大学的遗传学家 L. L. Cavalli-Sforza。他的科普小书(2001)已有中译本(2003)，更可以方便地供我们参考。

### 3.

当早期新石器文化在中国大部分地区出现时，各民族开始征战于黄河流域，最终统一并统治这片土地。在 1935 年出版的重要文献《夷夏东西说》中，傅斯年明确指出，在公元前几千年中，中国大部分社会、政治发展，都集中于黄河流域的一条东西轴线的两侧，当时战争主要是在东部的夷和商、与西部的夏和周之间展开。因此有一个向东融合的构想。直到汉朝长江流域才迅速地发展起来。中国历史开始在更大的地域内，沿南北轴线展开，由此开始向南融合。

黄河流域的民族成功的主要原因，是他们发明了文字。文字使一个社会在与毗邻社会的竞争中，取得了决定性的优势。的确，汉语中“文化”和“文明”两词，都是由汉字“文”构成的，“文”意为“文字”。大量的早期文字样本在 3400 年前就已存在，致使一些学者推测，文字的发明一定相当久远<sup>12</sup>。从这些文字样本的词汇和语法来看，很明显这些商朝和周朝民族的语言是汉语的直接祖先。

直到刘邦在公元 202 年建国，并以“汉”字命名国家，指一个民族群体的“汉”字才获得其显著性。在此之前，曾使用“华”和“夏”。在现代汉语中，我们可以找到诸如“汉子、男子汉、好汉、老汉”这样的词语，其中“汉”只是一个普通名词，意思是“男人”。有人怀疑这是否就是该字的初始意义，后来又被其它民族延伸用来指这个民族群体。在早期社会里，用一个民族的语言中，表示“人们”或“人”的词来指称一个民族的这种现象，并非罕见<sup>13</sup>。

不管“汉”这个名称如何演变，与中国第一个重大朝代同名的民族，早在该朝代建立以前，就已与其它民族相互影响了。传统上是以前中原为基准，根据这些民族所在的基本方位，来进行民族群体的命名。东边的夷也许说南岛语和南亚语(Austro-Asiatic)；西方的戎则可能是说藏缅语；北方居住着强悍的游牧民族—狄，汉朝势力减弱时，他们就会闯入，反之他们则隐入沙漠和山区之中。他们操阿尔泰语系的几种语言，这个语系包括一端的土耳其语<sup>14</sup>，和另一端的日语，像一条带子横贯中亚。在南方，“蛮”这个名称覆盖了种类最多的民族群体，他们的语言现在归入卡岱(Kadai)、苗瑶<sup>15</sup>及其它语族。当然，用古代名称识别现代民族，并非只是臆测，而

---

<sup>12</sup> 尽管这种猜测有一定的道理，但现在还没有发现什么毫无争议的早期样本。若要称得上是真正的文字，这些符号必须与口语密切相关，以便能够驾驭口语生动的表达力。

<sup>13</sup> 毛宗武等(1982:9)提出的一些例子包括：“Lakkia”一词指侗水族的一种语言，字面意思是“人山”。瑶族人自称“Mian”，意为“人”，根据不同的方言，可写作“勉、曼、门、敏”。不过我们应记得，“汉水”得名早于汉朝几个世纪。

<sup>14</sup> 郑张(1990)怀疑“狄”与“Turk”有关，因为“狄”已经用[r]和末音[g]构拟出来。

<sup>15</sup> 据说“蛮”可能与瑶族人的自称有关，许多瑶族人自称[mien]，常写作“勉”。

是根据目前人口分布进行外推。随着我们这一领域内知识的加深，这些名称也不需予以修正。

除了上一段落中提到的四个集合名称外，还有一个具历史渊源的名称 Yue，写作“越”或“粤”。很明显地，这一名称在商朝末期就已被用来指中国西北地区的民族<sup>16</sup>。从战国末期一直到基督教元年的几百年间，“越”开始用来指中国南方大范围内的民族，特别是“百越”这个名称。同样，“百”在这里不能作字面理解，只是指多种多样，从长江一直到越南北部，这些民族一直不同程度地与黄河流域的文明融合着。

公元开始的前几百年间，特别是汉朝瓦解以后，大规模的南迁人群，无疑是加强长江流域建设的主要力量。两方面的因素共同作用引发了此次迁徙，一是自然灾害，主要指黄河流域的洪水，一是持续的战乱及其破坏力。然而，对少数民族的另一称呼，“五胡乱华”的“胡”开始出现。这五个少数民族是：匈奴——与汉朝长期因争夺水草而发生战争的阿尔泰民族；鲜卑——辽、金及后来的满族的前身；戎；羯及羌。在这五个名称中，只有“羌”一直沿用至今，指中国西部一些藏缅语系的少数民族。

随着中原各民族的南迁，他们不断与当地居民接触，由此产生了深远的影响。这就是现代中国各语言形成的基本样貌。“汉化”这个词常用来指这种影响，意思是只有汉文化影响着当地民族。然而，这种影响并不总是单方面的。

#### 4.

中国的迁徙持续且不均衡。不均衡是指每一次迁徙人数都不相同。一次迁徙高潮发生在约公元 1100 年北宋向南宋过渡时期，那时，都城从开封迁到了杭州。由于这一变迁，在中国历史上华南的人口密度第一次超过了北方。迁徙的持续性，表现为总有人因洪水、干旱等自然灾害，或由于传染病、人口过密、战争等人为原因，而不断的迁移。尤其在中国历史每一次朝代更替的末期，这种迁徙十分普遍。中国的语言就像一条花毯，由无数次互相交错的迁徙编织而成，每次迁徙都增添一种颜色，留下一部历史。

今天语言各层面细致深入的分析，可将碎片恢复成画。中研院的何大安(1996)对四川达县湘语聚居区言语的分析，就是一个很好的例子，他的分析数据首先出现在崔荣昌(1989)的报道中。何首先观察了中古汉语中带有三声，即去声的语素，以三种不同的调型出现。如以下数据所示：

	A[55]	B[21]	C[24]
IIIu	59	3	22
IIIv	22	47	25
Total	81	50	47

<sup>16</sup> 关于这一称呼的历史研究，详见 Meacham 1996。

调 A 为[55]的高平调, B 为低降调, C 为升调。IIIu 指中古汉语中清声母, 等于传统分类中的阴去。IIIv 指中古汉语中浊声母, 与传统分法中的阳去一致, 虽然声母的清浊把中古汉语的声调分为两支, 是在各方言中可观察到的一个普遍现象, 但是有三种反映形式却很罕见。由于这种倾斜分布, 情况变得更有意思了, IIIu 在 A 中产生 59 种反映形式, 但在 B 中只有三种, 事实上其分布存在落差。

何仔细分辨各种书面语与口语层中一种叫做文白异读的现象, 最后得出这些声调发展的综合解释, 证明在口语层, 第三调根据声母的清浊分裂, IIIu 变为 A, IIIv 变为 B。另一方面, 在书面语层, 所有第三调语素在达县方言中都只有调 A, 因为书面语很可能来自北方, 在北方话中, IIIu 与 IIIv 已经合并了。

表中第一排 B 的落差, 与中古汉语官话第一调的落差相似, 这里的落差是因为在中古汉语 Iv 调中, 也就是阳平调中没有不送气的声母。因此官话中有送气声母如 páng 旁、táng 唐、以及 káng 扛, 但没有不送气的声母如 \*báng、\*dáng、\*gáng。分布中的落差对于解读历史来说是很好的线索。

以语料为基础, 何将这一语言追溯到元朝, 认为说湘语的居民当时住在湖北, 受到了附近赣方言的影响。在现代崇阳话中仍能找到赣方言的痕迹。元朝末年, 湖北居民群体移居到了湖南, 并一直居住到明朝末年。就是在这一时期, 湘语言环境使口语与书面语阶层开始产生差异, 最后, 在明朝末年, 该居民群体迁到目前的四川地区现在的住址, 其语言环境是西南官话。

何的研究是中国境内各语言及方言形成过程的一个很好的缩影。它证明了这样一个构想: 不同的语言群体常处于频繁迁移中, 并不断地与其它语言群体接触, 结果形成了今天的中国语言大拼图<sup>17</sup>。

## 5. 18

考古学研究化石和古人类的简单工具, 遗传学研究人体, 而对语言学来说, 其研究材料是古文献和现代语言。如前所述, 现存的汉民族古籍至少可以追溯到商朝的甲骨文。学者们普遍感到, 随着更多材料的发现, 时间可能推向更久远。例如, 几年前在山东发现的刻有文字的陶碎片, 据说约有 4000 年的历史。

另一方面, 古代中国其它语言的早期文献却极为少见, 而且由于当时重汉语音韵, 其它语言的文献也都用汉字转写, 致使对该语言的分析变得更复杂。然而尽管如此, 近年来在这一领域里仍取得了一定的进展。

对汉朝《白狼歌》的研究就是其中一例, 此文分别由 South Coblin、马学良、戴庆厦及郑张尚芳独立地进行分析, 这些学者得出了这样的结论: 文中所记录的语言

<sup>17</sup> 另外一个关于方言互动及方法讨论, 见王和连(1993), 其中涉及闽方言和北方官话。

<sup>18</sup> 关于本文第五部分的参考文献见王(1998b)。

属于藏缅语族，绝大部分与缅甸及彝支语很接近。作为藏缅语言最早的研究材料，《白狼歌》具有很高的价值。藏语七世纪出现文字，而缅语直到十二世纪才出现文字。

郑张尚芳发现《白狼歌》的汉字中，有 115 个跟缅语有关，而仅有约 40 个汉字与藏语有联繫。根据这一发现，他得出结论：藏缅语族的这两个分支从汉代起就已相距甚远。他进一步推测道：汉藏语系应该至少向前推 2000 年，这就把汉藏语系的日期推到了距今 4000 年前。

研究古代少数民族语言取得较大进展的另一个例子，是韦庆稳和郑张尚芳对《越人歌》的分析。该汉代文献被认为是用一种与壮语有亲属关系的语言写成的。另一研究领域涉及分散于各种古籍中的孤立词项。比如，李敬忠曾分析汉代《方言词典》中的许多词项，并提出这些方言词汇与壮语及中国其它南方语言之间的亲属关系。

同样，唐朝樊绰编辑的《蛮书》第八章中，据说有几十个词项与云南的白族语言一致。我们同样可以把对古代文献中跨语言纪录的研究，不论是专有名称还是整个词组，纳入这一类型。虽然有关发现并不总是系统的，且数量不多，但他们仍然可以和考古学家铲下掘出的文物媲美。这些珍品大部分尚未开发<sup>19</sup>。如果进行适当的分析，它们将逐步揭示中国古代语言的原貌。

上文中提到的研究，主要是关于历史上汉语与其它语言群体之间的互动，还有许多讨论现代语言中跨语言痕迹的作品。关于这方面有一篇罗常培早期撰写的论文，其中他描述了普通话来自通古斯语(Tungusic)，即满语的一些词。鉴于满族近期对中国长时期的统治，这并不奇怪。此外，还有大量关于南方方言与其邻近的少数民族语言互动的研究，特别值得一提的是在 1976 年，有两篇研究这一领域的重要文章问世：罗杰瑞和梅祖麟的<古代中国南部南亚语系><sup>20</sup>，及余霭芹的<中国南方语的底层：傣的连接>。余重点叙述了闽南各方言与少数民族语言之间几点重要的共同之处，并以这段有价值的观察作结：

“闽语与粤语都保留了许多古汉语的特征，并且它们也与周围的壮语、傣语等有着许多共同的特点，这些共同之处使它们有别于北方话。在继续对南方各语言的研究时，我们觉得必须考虑层次区别以及地区语言学。”

的确，不论是北方的阿尔泰语，还是南方的侗台语以及其它语言，汉语方言的研究不能与周围语言的发展相分离。

## 6.

---

<sup>19</sup> 李方桂和 S.W. Coblin 的工作(1987)在该领域中有示范性。

<sup>20</sup> 关于罗杰瑞和梅祖麟的不同观点，见 Zhang (1998)。



我本想在中国早期文献中找一些关于中国语言起源的讨论，却是徒劳无获<sup>21</sup>。但是在欧洲却有许多关于汉语起源的假设。有一些很荒谬，如 John Webb 在 1669 年出版的一篇有影响的论文<大清帝国的语言是原始语可能性的历史分析>，就是早期的一个例子，用这篇论文，他想说汉语是伊甸园中所说的语言。

十九世纪，伦敦大学学院的印度支那语文学教授及皇家语文学会主席 Lacouperie 曾很自信地说：

“中国语言的源头，来自西亚说乌拉尔-阿尔泰语系巴克语(Ugro-Altaic Bak)的移民...而这些移民的语言又融合了巴比伦语...这一综述现在看来是毋庸置疑的，因为已有许多有利的证据。”

对当时如此有名望的人来说，这些话的确是非常草率的。即使现在，关于中国起源的证据仍然很少，并且很有争议。

现在，基本上普遍认为汉语方言在谱系上与藏缅语系有关，并且构成了汉藏语系的核心。中研院的龚煌城对汉藏语系的韵母构拟(1995:41-92)，就是很好的证据。根据词汇比较，我猜想这些语系的分裂大约在 6000 年前<sup>22</sup>，然而，除了谱系关系，在其它问题上几乎没有共同意见。首先，关于侗台语和苗瑶语就有长达几十年的争论：李方桂把它们包括在汉藏语系中，而来自相反的意见，如 Benedict (1942, 1975)，则把它们归入其它语系中。潘悟云的华澳语系(1995:113-144)范围最大，如果这个假说可以成立，下一步迫切的工作是要把这个大语系的内部分类(subgrouping)仔细地证实出来。

蒲立本(Pulleyblank)讨论了汉语与印欧语系的一些关联，并提出了 24 组同词源词(cognate)，以供考虑<sup>23</sup>。其中举了一个“目”的例子(1995:175)，该词在广东话中读 [muk]，保留 k 这个尾音，声母 m 在这被看作前缀，他把词干与拉丁语中表示眼睛的词“oculus”的 oc 相比较。另一个例子 (1995:179) 是“狗” gǒu 的书面语“犬” quǎn；口语词汇“狗”有着不同的来源。很凑巧，广东话中的“犬” [hyn] 与英语中的“hound”（猎狗）有很强的相似性，因为这两种语言都把原先的 k 音变成 h 音。印欧语的 k 保留在英语的“canine”中，该词来自拉丁语，汉语的 k 保留在厦门话 k'ian 中。

狗从狼演化而来，早在一万多年前，就成了人类最早驯养的工作伙伴。近年来遗传学家证明，狗最早起源于东亚，见 Savolainen et al (2002)。有趣的是，2002 年 11 月 2 日的 Science 上，一条普通的中国狗居然成了封面人物。邓晓华已经开始从历史语言学的角度，探索与狗有关的一些词汇变化，以配合遗传学家的研究。

---

<sup>21</sup> 有关 Webb、Lacouperie 及其它学者观点的参考书目，见王(1995:iii)。

<sup>22</sup> 王(1998:530)。

<sup>23</sup> 蒲立本在 International Review of Chinese Linguistics 的第一期中进一步扩展了他的假说，文章后还有六篇评论。

做第二种假说的是法国语言学家 Sagart (1995:195-223), 他提出汉语与南岛语之间的谱系关系<sup>24</sup>。由于南岛语中的语素为典型的双音节, 而汉藏语素则是单音节, 因此 Sagart 必须做出这样的假设: 在原始语中存在许多丢失一个音节的现象。Sagart 提出另一个很有趣的假设, 是把原始语中的一个辅音与古汉语中一个构拟的介音相联系。这两种假设都受到了挑战。李壬癸特别提出六个很有说服力的批评, 以反对这一假说, 包括他所观察到的“这些同词源词没有严格的语义对等性”和“几乎没有一个同源词属于基础词汇”<sup>25</sup>。

然而, 为了捍卫他的例证, Sagart 提醒大家注意在远古时代, 海洋的侵袭造成中国海岸线的变化。比如说, 4300 年前, 北京处于渤海海域, 山东西部的大部分地区则是沼泽<sup>26</sup>。和后来一千年中的南岛民族一样, 这样的环境可能就是之后打算航行世界的远祖部落的家园。

Sagart 同时还提出了大汶口文化(山东, 距今 6000 年)与南岛习俗之间的一些有趣的联系。如: 小米为谷物之首, 及青春期拔除上犬齿的习俗<sup>27</sup>。Sagart 指出, 正如 Pietrausewsky 报导的那样, 安阳出土的商朝头盖骨, 与台湾现在的南岛民族之间有相似的地方。记得前面提到的傅斯年的论文中, 也有把商朝各族与夷族放在一起的叙述, 这暗示夷语与南岛语之间也许有亲属关系。虽然每一个例子中的证据都是间接的, 但它们竟然与语言学的假说相一致, 的确相当有趣。

Sergei Starostin 所讨论的第三种假说(1995:225-251), 是三种中最大胆的, 因为他把横跨三大洲的语言联系在一起, 德内-高加索(Dene-Caucasian)语据称包括西班牙北部的巴斯克语, 印度北部的 Burushaski 语, 北西伯利亚的 Ket 语, 高加索山区的一些语言, 北美洲的那德内(Na-Dene)诸语言以及汉藏语。事实上, 像萨丕尔(Edward Sapir)这么一位权威, 也观察到了后两个语群间的亲属关系。然而, 像 Starostin 把六个语群都归为一类的做法, 却是个未有人尝试过的挑战。如果真的存在这种衍生出分布距离如此遥远的语群的原始语, 它一定有 10,000 年或更久远的历史。

要评估这一假说是否合理, 需要核查 Starostin 比较过的所有构拟出的形式, 除了评估跨语系比较的可接受性, 也得评估特定语系中每一次构拟的内在合理性。由于我们所研究的这段时间相当悠久, 这些词在形式和意义上都积累了许多变化, 这就使对同源词的考证更加困难。然而, Starostin 最近比较了跨七个语系近 35 个基本词汇的构拟形式(1995:249)。我把他的原始数据改成了百分比, 如下表所示:

	OC	PTB	PNC	PY	PIE	PAN
古汉语(OC)	100	<b>74</b>	<b>43</b>	<b>34</b>	23	14

<sup>24</sup> 该假说的详细阐述见 Sagart (1993:1-62)。

<sup>25</sup> 李壬癸的批评见王(1995:94), 在同一本专刊中, Robert Blust 写了进一步的评述(283-298), 但支持南岛语构词学的语料则见 Stanley Starosta (393-404)。

<sup>26</sup> 见 Chang (1986)图 35、36 的阐述。

<sup>27</sup> 拔牙的习俗据史书中记载可溯及元朝, 在中国西南部也可见到。

原始藏缅语(PTB)	100	<b>51</b>	<b>40</b>	14	11
原始北高加索语(PNC)		100	<b>57</b>	17	11
原始叶尼塞语(PY)			100	11	11
原始印欧语(PIE)				100	14
原始南岛语(PAN)					100

我们可以从表中的第一排看到：最近的亲属关系是古汉语和原始藏缅语，有74%的同源词；其次在这四个德内-高加索语组中，其同源词百分比（上表中用斜体黑字表示）高于原始印欧语或原始南岛语中的百分比。

对于这几种汉语祖先的假说，李葆嘉曾做过综合性的讨论(2002)。

最近提出的另一组远距离语群是欧亚语群，格林伯格(Joseph Greenberg)在一本由斯坦福大学出版社出版的书中就此讨论过(Greenberg 2000)。这组远距离语群包括印欧语、阿尔泰语以及其它几个语系，和德内-高加索语组一样，这些大语群最终的有效性还需要长期详细的研究。显然，这是具有挑战性的。由于这两类语群覆盖了欧亚大陆的一些共同区域，若是探索在一万年前，甚至在农业出现之前，这两个古代社会语群如何互动，的确是很有意思的。

附带说明一下，上文中我所提到的三位有创新的学者都出自欧洲。我们期待着不久的将来，中国学者也能加入到寻找自己语言祖系的研究中来。事实上，由于构拟祖系的资源正在迅速消失，因此越来越迫切需要这一领域的研究。

今天，在近150种中国语言中，至少有几十种处于濒危状态，这些语言只有老年人还在讲，而且他们没有传给年轻的一代。对于语言学家来说，不论所说人数的多少及所处地域的大小，每种语言都具有同等的重要性。常常是那些比较不富裕、难以接触的语言，才更显其独特，且更添其作为语言知识的珍贵。随着这些濒危语言的灭绝，我们不知道哪些有趣的类型特征将随之消失，也不知道哪些有关过去的线索会随之消亡。为我们的后代把这些语言记录下来，对于保留我们的语言遗产来说是很关键的一步，同时，除了有意思的实地经验外，记录这些语言也会提供新的语料。

在濒危语言的研究上，中国云南具有独特优越的地位。位于喜马拉雅山山脚下的云南地理位置特殊，诸多发源于世界屋脊的河川，都是先穿过云南再往下游流去。河川在云南北部白雪皑皑的群山间切割而下，流经云南南部的热带丛林，为当地人民创造出赖以迁徙、运货的古道，多种语言及文化也以这些河川为传播渠道，传至华南和东南亚。包括藏缅、南亚、南岛与苗瑶等语系，都是经由这些河川两岸向外辐射传播，而广布世界各地。

这其中不少语言受到云南崎岖地形的隔绝保护，数千年来未受外力严重影响，因此保留了古形式，有助于我们了解古代语言，例如藏缅语系某些音的辅音群，就

有助于构拟 3000 多年前的汉语发音。许多今日汉语口语里已较少使用的字，在云南原住民的语言中依然可见，如代表“木柴”的“薪”，及代表“筷子”的“箸”。

这些较“未受污染”的语言，也提供了另一种世界观。熙熙攘攘的现代生活，让人类日益与自然世界疏离，赫然发现有人把“蝴蝶”称为“飞花”，“水”称为“梦游者”，更让人有耳目一新的惊喜。这是我在云南跟木霁弘、陈保亚几位老师一起探讨民族语言时，记下来的两个例子。有人说，学会一种语言，犹如习得另一种灵魂，因此我们更会记得，关照世界的方式其实有很多种，我们应该不忘随时睁开双眼，敞开心胸。

## 7.

十九世纪语言研究的中心在欧洲大陆。拉斯克(Rask)和格瑞姆(Grimm)的早期构拟，舒雷切(Schleicher)和舒梅特(Schmidt)的历史理论，以及语音变化规律的新语法假设，都为印欧语言的研究奠定了坚实的基础。二十世纪语言学的中心在美国。源于欧洲的观点，由保斯(Franz Boas)和雅可布逊(Roman Jakobson)等学者传到新大陆，导致当地语言研究的发展，如格林伯格(Joseph Greenberg)的类型学与普遍法则的综合，以及各种形式语法的探讨。

二十世纪初，中国语言学家在世界舞台上展露头角时，只有像马氏文通这样层次不高的作品。这是一部把印欧中心模式强加于中国语言的劣质语法。但愿今后的研究能避免这种全盘接受国外形式的做法<sup>28</sup>。

二十世纪中叶，赵元任、李方桂、罗常培等学者，把西方的观点与大量中国本土的学术思想相融合。他们从这些新方法论中吸收有用的东西，但并不放弃继承传统语文学中有价值的部分。

随着几代中国语言学家的不懈努力，我们的研究已经进入世界前列。学术的导向、目标的确定、独立方法论的创造<sup>29</sup>，都在朝着越来越健康的方向迈进，语言学研究也是如此。然而，他人的地图总会限制我们的视野，我们便注定要被他人领着路走。二十世纪已结束，二十一世纪的中国语言学更见延伸、充实。十九世纪的语言研究中心在欧洲，二十世纪在美国，若说二十一世纪在中国，但愿不是一种奢望<sup>30</sup>。

---

<sup>28</sup> 如果中国语言学家能够讲述 Panini 和 Grimm，却全然不知扬雄、陈第的贡献，那是很讽刺的。

<sup>29</sup> 与物理学不同，关于人类行为的学科，都相当受到认知和其赖以发展的社会环境影响。例如，在语言学讨论中，常有人对汉字的语素音节系统的本质存有偏见，并认为西洋字母本身存在优越性。另一个例子来自语言学权威梅耶(A. Meillet)的观察，他认为由于东方语言缺乏曲折变化形态，故其历史构拟受到限制(1925:26)。很明显地，一个均衡的人类语言性质理论，绝对不能基于这种以欧洲语言为中心的认识。

<sup>30</sup> “中国的语言与民族”这个课题，无论是在时间或空间上，都是门博大精深的学问，不是任何一个人在现有的知识背景下能交代清楚的。况且这个范畴里的很多问题，我们到现在才开始着手研究。在这篇宏观的短文中，我只能蜻蜓点水般略提一下，勾勒出一张初步的草图，文中难免有不少疏漏及错误，希望专家朋友们多多指正，让我们得以尽快充实这方面的知识。

## References

- Benedict, P. 1942. Thai, Kadai and Indonesian: A new alignment in Southeast Asia. *American Anthropologist* 44:576-601.
- 1975. *Austro-Thai language and culture with a glossary of roots*. New Haven: Human Relations Area Files Press.
- Cavalli-Sforza, Luigi Luca. 2001. *Genes, Peoples and Languages*. Penguin Books. 中译本 2003. 追踪亚当夏娃。吴一丰、郑谷苑、杨晓珮译。台北：远流出版社。
- Chang, K. C. (张光直) 1983. *Art, myth, and ritual: The path to political authority in ancient China*. Cambridge: Harvard University Press.
- 1986. *The archeology of ancient China*. 4<sup>th</sup> edition. New Haven: Yale University Press.
- Chen, B. Y. (陈保亚). 语言接触与语言联盟：汉越（侗台）语源关系的解释。语文出版社。
- Coblin, W. S. 1979. A new study of the Pai-lang songs. *Tsing Hua Journal of Chinese Studies* 72:135-78.
- Cui, R. C. (崔荣昌). 1989. Report on the Changsha speech of Daxian, Sichuan. *Fangyan* 1:20-29. (In Chinese)
- Deng, X. H. (邓晓华). 1993. *人类文化语言学*。厦门大学出版社。
- van Driem, George. 2001. *Language of the Himalayas*. 2 vols. Brill.
- Edmondson, J. A. and D. B. Solnit, eds. 1988. *Comparative Kadai: Linguistic studies beyond Tai*. Dallas: Summer Institute of Linguistics.
- Freedman, D. A. and W. S-Y. Wang. 1996. Language polygenesis: A probabilistic model. *Anthropological Science* 104.2:131-138.
- Fu, S. N. (傅斯年)1935<夷夏东西说>Reprinted in 傅斯年全集，傅孟真先生遗著编辑委员会编，1980年版，第三册，pp.86-157。台北：联经出版公司。
- Gong, H. C. (龚煌城). The system of finals in Sino-Tibetan. In Wang ed. 41-91.
- Greenberg, Joseph H. 2000, 2002. *Indo-European and its closest relatives: The Eurasiatic language family*. 2 vols. Stanford, Calif.: Stanford University Press.
- Ho, Dah-an. 1996. Stages and strata in dialectal history. *New horizons in Chinese linguistics*, ed. by J. C-T-Huang and A. Y-H. Li, pp.215-234. Dordrecht: Kluwer.
- Huang, J. C-T. and A. Y-H. Li, eds. 1996. *New horizons in Chinese Linguistics*. Dordrecht: Kluwer.
- Li, B. J. (李葆嘉). 2005. <从同源性到亲缘度>。汉语的祖先 1-102。北京：中华书局。Pp.681. [Based on W.S-Y.Wang. ed. 1995. *Ancestry of the Chinese Language*. *Journal of Chinese Linguistics* Monograph 8.]
- Li, J. Z. (李敬忠). 1994. *语言演变论*。广州：广州出版社。
- Li, F. K. and S. W. Coblin. 1987. A study of the old Tibetan inscriptions. Special publication no.91. Taipei: Institute of History and Philology, Academia Sinica.
- Lin, L. L. (林伦伦). 1990. Some Tai related words in Min dialect of Chinese (Han) language in Guangdong Province. *Minzu Yuwen* 3:78-79. (In Chinese)
- Liu, W. 1997. Dental continuity of humans in China from the Pleistocene. *Proc. 30<sup>th</sup> Intern. Geol. Congress*.
- Luo, C. P. (罗常培). 1950. *语言与文化*。北京大学出版社。（1989年语文出版社再版）。

- Manomaivibool, Prapin. 1976. Thai and Chinese - are they genetically related? *Computational Analyses of Asian and African Languages* 6:11-32.
- Mao, Z. W. et al. (毛宗武、蒙朝吉、郑宗泽). 1982. A sketch of the languages of the Yaozu. Beijing: Minzu Press. (In Chinese)
- Meacham, W. 1996. Defining the hundred Yue. *Indo-Pacific Prehistory Association Bulletin* 15:93-100.
- Meillet, A. 1925. *La méthode comparative en linguistique historique*. Paris.
- Minett, J.W. and W.S-Y. Wang eds. 2005. *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*. Hong Kong: City University of Hong Kong Press. Pp.xii+538.
- Mufwene, Salikoko S. 2001. *The Ecology of Language Evolution*. Cambridge University Press.
- Norman, J. and T. L. Mei. 1976. The Austroasiatics in ancient South China: Some lexical evidence. *Monumenta Serica* 32:274-301.
- Pan, W. Y. (潘悟云). 对华澳语系假说的若干支持材料。In Wang ed. 113-144.  
—ed. 2002. *东方语言与文化*。中方出版中心。
- Pietrusewsky, M. 1995. Taiwan aboriginals, Asians and Pacific Islanders: A multivariate investigation of skulls. *Austronesian studies relating to Taiwan*, ed. by Paul J-K. Li et al., pp.295-351. Taipei: Institute of History and Philology, Academia Sinica.
- Prosterman, R., T. Hanstad and P. Li. 1996. Can China feed itself? *Scientific American* 275.5:90-96.
- Pulleyblank, E. G. (蒲立本). 1995. The historical and prehistorical relationships of Chinese. In Wang ed., pp.145-197.  
—1996. Early contacts between Indo-Europeans and Chinese. *International Review of Chinese Linguistics* 1:1-24.
- Sagart, L. (1993). Chinese and Austronesian: evidence for a genetic relationship. *Journal of Chinese Linguistics* 21.1.1-62.  
—1995. Some remarks on the ancestry of Chinese. In Wang ed., 1995, pp.195-223.
- Savolainen, P. et al. 2002. Genetic evidence for an East Asian origin of dogs. *Science* 298.1610-3.
- Starostin, S. 1995. Old Chinese vocabulary: A historical perspective. In Wang ed., 1995, pp.225-251.
- Sun, H. K. (孙宏开)主编。1997-present. *中国新发现语言研究丛书*。共 30 册。
- Wang, Feng. 2006. *Comparison of Languages in Contact: The Distillation Method and the Case of Bai*. Academia Sinica, Institute of Linguistics. Pp. 227.
- Wang, M. K. (王明珂). 1997. *华夏边缘*。台湾：允晨丛刊 66。  
—2003. *羌在汉藏之间：一个华夏边缘的历史人类学研究*。台北：联经。  
—2006. *英雄祖先与弟兄民族*。台湾：允晨丛刊 110。
- Wang, W. S-Y. (王士元). 1987. Representing language relationships. *Biological metaphor and cladistic classification*, ed. by H. Hoenigswald and L. Wiener, pp.225-251. Philadelphia: University of Pennsylvania Press. Reprinted in Wang, 1991a.  
—ed. 1991a. *Explorations in language*. Taipei: Pyramid Press.  
—1991b. Languages and dialects of China. *Journal of Chinese Linguistics Monograph No.3*. Berkeley: University of California.  
—1994. Glottochronology, lexicostatistics, and other numerical methods. *Encyclopedia*

- of language and linguistics, pp.1445-1450. Edinburgh: Pergamon Press.
- ed. 1995. The ancestry of the Chinese language. *Journal of Chinese Linguistics* Monograph No.8. Berkeley: University of California.
- 1996. Linguistic diversity and language relationships. *New horizons in Chinese linguistics*, ed. by J. C-T. Huang and A. Y-H. Li, pp.235-268. Dordrecht: Kluwer.
- 1997. Languages or dialects? *The CUHK Journal of Humanities* 1:54-62. Hong-Kong: Chinese University of Hong Kong.
- 1998a. Language and the evolution of modern humans. *The origins and past of modern humans - toward a reconciliation*, ed. by K. Omoto and P. V. Tobias, pp.247-263. Singapore: World Scientific.
- 1998b. Three windows on the past. *The Bronze Age and early Iron Age peoples of eastern central Asia*, ed. by V. Mair, pp.508-534. Philadelphia: University of Pennsylvania Museum.
- 2006. 语言演化的探索。门内日与月：郑锦泉先生七秩寿庆论文集。语言暨语言学专刊外编之七。9-32. 钟荣富、刘显亲、胥嘉陵、何大安编辑。Taipei: Institute of Linguistics, Academia Sinica
- and Chinfa Lien. 1993. Bi-directional diffusion in sound change. *Historical linguistics: Problems and perspectives*, ed. by C. Jones, pp.345-400. Essex: Longman.
- 2002. 语言是云南的文化宝藏. 台湾: 科学人 8.58-59. [Linguistic treasures of Yunnan.]
- and J.W. Minett. 2005a. The invasion of language: emergence, change, and death. *Trends in Ecology and Evolution* 20.5.263-9.
- and J. W. Minett. 2005b. Vertical and horizontal transmission in language evolution. *Transactions of the Philological Society* 103.2.121-46.
- Wei, Q. W. (韦庆稳). 1981. An analysis of the relation between Yuerenge and the Zhuang language. *Minzu yuwen lunji*, pp.23-46. Beijing: Zhongguo Shehui Kexue Press. (In Chinese)
- Winchester, Simon. 1996. *The river at the centre of the world*. London: Penguin Books.
- Wu, X. Z. and F. E. Poirier. 1995. *Human evolution in China*. Oxford: Oxford University Press.
- Xu, T. Q. (徐通锵). 1991. *Historical linguistics*. Beijing: Commercial Press. (In Chinese)
- Ye-shes-vod-gsal A-tshogs (意西微萨·阿错). 2003. 藏、汉语言在“倒话”中的混合及语言深度接触研究。南开大学博士论文。
- Yu, M. (俞敏). 1980. An examination of the origins of the Han and the Zang peoples. *Journal of Beijing Normal University* 1980.1:45-53. (In Chinese)
- Yue-Hashimoto. A. O. K. (余霭芹). 1976. Substratum in southern Chinese - the Tai connection. *Computational Analyses of Asian and African Languages* 6:1-10.
- Zhang, H. M. 1998. Chinese etyma for "river". *Journal of Chinese Linguistics* 26:1-47.
- Zhengzhang, S. F. (郑张尚芳). 1990. Some Kam-Tai words in place names of ancient Wu and Yue states. *Minzu Yuwen* 6:16-18. (in Chinese)
- 1991. Decipherment of Yuerenge. *CLAO (Cahier de Linguistique d'Asie Orientale)* 20:159-168.
- Zhou, Z. H. and R. J. You (周振鹤、游汝杰). 1986. *方言与中国文化*. 上海人民出版社。

