

追寻汉语的祖先

分子人类学关于东亚人群DNA的研究，给了语言学家们莫大的启示，这也许为破解汉语起源之谜提供更多的线索。

■本刊记者 孙展

图/华盖

汉语从哪里来?一直是个谜团。

17世纪,一名叫作韦伯的学者写道,“汉语是伊甸园中所讲的最初语言。”而另一位想象力丰富的学者干脆说,汉语是“被称为中国的那个伟大国家的某位圣贤一下子发明出来的,其目的是为了居住在周围的不同民族的人群之间能够用这种语言进行口头交际。”

实际上,在中国人内部,这样富有想象力的说法也同样代代相传。直到最近100多年,由于现代语言学的兴起,汉语才被用来和其他语言类比分析,但有关汉语起源的种种假设却总是扑朔迷离。而近年来,分子人类学者们有关东亚人群迁徙的分析,似乎正在接近这个谜底的答案。

语言的足迹

1600年,一艘荷兰商船途经马达加斯加前往印尼。在经过了数月的航行后,他们到达了目的地,令船员们感到意外的是,他们发现,虽然相聚遥远,但当地的马来语与他们途经的马达加斯加的土著语有着惊人的相似之处。

难道在大航海时代开始之前,这些分布在各个海岛上的人群就早已横渡大洋,穿梭在广袤的海域中了吗?难道这些相隔万里的海岛居民竟然是比自己更加能干的航海者?水手们百思不得其解。

水手们并不知道,自己无意中的这个发现,竟然触碰到了语言的核心秘密。400年来,一代一代的语言学家追随着这个秘密,开始了伟大的发现历程。

1786年,英国的琼斯爵士通过对比研究关注到:意大利语、凯尔特语、日耳曼语以及波罗的语、斯拉夫语,与远在亚洲的印度雅利安语有某种程度上的相似之处,会不会这些语言有同一个源头?他提出了自己大胆的假设——这些广泛分布在欧洲和印度的语言有一个共同的祖先,这个假说被称为“印欧语系假说”。

琼斯的假设,让语言学家们开始尝试着对比、分析这些有着共同关系的语言之间的亲疏远近。1861年,德国语言学者施莱歇尔在达尔文的进化论的启



发下,推导出语言也有着类似生物进化的过程,并给印欧语系画出了一棵“谱系树”,在这棵语言的谱系树上,原始的“印欧语”处于最底端,是“树干”,这个树干分出两个“树枝”,一枝是“斯拉夫-日耳曼语支”,另一枝是“雅利安-希腊-意大利-凯尔特语

支”,这两个“树枝”再分别分出众多“小树杈”,如日耳曼语、立陶宛语、斯拉夫语、凯尔特语、意大利语、希腊语、伊朗语等等语言。

尽管施莱歇尔推导的这棵“印欧语系大树”并不完全精确,但他的方法立刻成为语言学家们观察这个纷繁复杂世界的利器。他们反过来去研究印欧语系之外的语言。

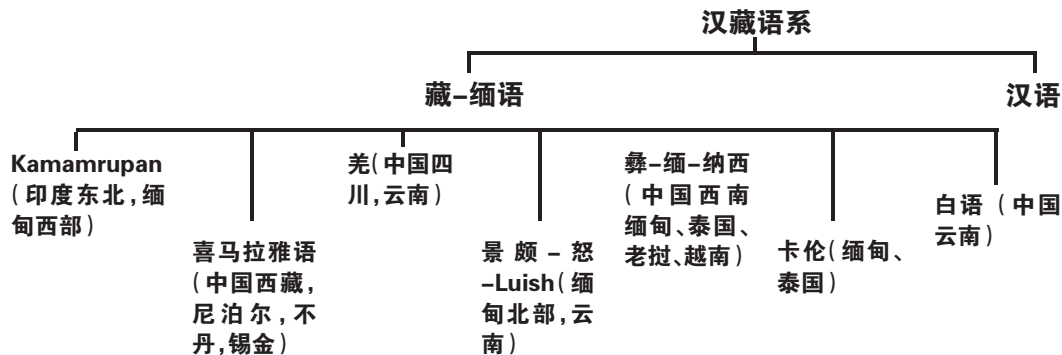
图/CFP

1600年那些荷兰水手们的疑惑,也得到了部分的解答。在北到台湾岛,南到新西兰,西到马达加斯加,东达复活节岛,横跨印度洋和太平洋的广袤海域上,的确存在一种共同渊源的海岛语系,尽管这种覆盖2亿人口,有着959种语言的语系内部形态千差万别,但它们却有着明确的同源关系,学者们将之称为“南岛语系”。在欧洲人进行海外殖民扩张之前,“南岛语”是世界上分布最广的语系。

那么古老而神秘的汉语是否也遵循这样的规律——同样是另外一种更加古老的语言分化而来,还是像学者们最初想象的那样,是由中国的某位圣贤独立创制而来?

从19世纪开始,语言学





家们逐步发现,这个地球上人口最多的国度里面的语言也能像其它语系一样,描绘出一棵“谱系树”。在这棵谱系树上,汉语、藏语、缅语等有着亲缘关系,语言学者们提出了“汉藏语系”的假说。

东亚同源?

如果汉语真的和藏缅语有同源关系,那么它们来自何方,又是如何分化而成的呢?

1950年代,美国语言学者斯瓦迪士从统计学的角度分析不同语言,得出了约200个核心词,几乎每种语言都包括这些核心词汇,它们具有超强的稳定性。根据对比分析不同语言之间核心词的同源关系,就能大致判断出语言之间的亲疏远近,同时,这些核心词汇的变化具有一定的速率,而分析两种语言中核心词汇的不同比率,则可以推算出它们之间分化的大致时间。

在对比汉藏之间的同源词时,学者们发现,汉语和藏缅语之间有共同的数词一至十,但“千”和“万”却各并不相同,而在汉语中,“千”和“万”在殷商时期就已经出现,那么汉语与藏缅语的分化必定早于殷商时期。语言学家们大致推断,这个时间距今约4000-5000年左右。

但在这种粗略的勾勒中,依然有很多矛盾难解之处。东亚地区语言的复杂性,远远超过了印欧语系。尽管经过长达一个多世纪的研究,语言学家们大致能够接受“汉藏语系”这种假说,但其间的分歧依然很大。

中国的语言学家李方桂早在1934年就提出,汉语不但与藏缅语有亲缘关系,同时

和苗瑶语、侗台语也有亲缘关系,而美国学者本尼迪克特则认为,苗瑶语、侗台语与汉语关系不大,反而与日语、南岛语有着亲缘关系。

颠覆性的观点来自1990年。当年,第23届国际汉藏语会议召开。就在会议正有条不紊地进行时,法国东亚语言学研究中心的学者沙加尔提交了一篇名为《汉语南岛语同一起源论》的文章。这个文章刚刚公布,就立即引发轩然大波。对于普通公众来说,这份标记有各种语音符号及绕口的学术名词的论文不亚于一篇“天书”,但在座的学者们却明白,这份论文的观点一旦成立,将是对传统的“汉藏语系假说”的最大挑战。

沙加尔认为,汉语与南岛语之间存在着不少的同源词,而这些同源词有着规则的对应关系。换句话说,汉语与广布在太平洋和印度洋海岛上的南岛语系,也就是400年前荷兰水手发现的马达加斯加和印尼人所讲的语言都来自于一种更加古老的“母语”。

沙加尔的观点极富挑战性——如果按照传统的理解,各个语系独立起源,互不隶属,那么该如何解释汉语和南岛语之间这种千丝万缕的联系?

沙加尔的观点得到了中国学者邢公畹的支持。他认为存在一种“汉藏-南岛”的同源关系。中国学者潘悟云也注意到,不仅仅是汉语存在这种联系,在东亚这片大陆上,侗台语就像是各种语系的交集,它既与汉藏语系有关,又与南岛语系、南亚语系有密切关系。沙加尔的汉语南岛语同源的说法,让潘悟云产生了一个更大胆的联想——这些语言会不会都拥有一个共同的起源?

稻米的密码

在如此广大的语言区域内,有一个词语引起了语言学者们的关注,那就是“稻米”。在东亚人群



中国语系的分布

▼图/CFP



中，稻米与人们的饮食和生产息息相关，占据着不可或缺的地位。

语言学家们发现，稻米在南岛语系中有着相同的来源，这个范围覆盖了马来-印尼语、西太平洋上帛琉群岛帛琉语、台湾的赛德语、卑南语、鲁凯语、越南的雷德语，菲律宾马诺波语，台湾巴则海语，道卡斯语，阿美语。而它们又与藏语的稻米有着明显的同源关系。另外，上古汉语的“稻”，则与侗台语和南亚语同源。

语言学家郑张尚芳也发现，在汉语里，“粝”本意为“粗米”，其同音词“痲”有“恶疮、麻风”的意思，而藏语中相对应的词，同样有两个含义，一个是“米”、一个是“疮、疖”，同样，印尼语中“米”与“麻风”也几乎同音。这种大跨度地区里连续出现的相似性难道只是一种偶然吗？

但是，如果单单凭借语言学本身来破解这些谜团，似乎是不可能的事情。语言总是在不断变动中，根据测算，每1000年左右，就能独立分化出来一种语言。尽管那些语言的核心词汇具有很强的稳定性，但超出2000年，甚至更加久远的时间，要想判断它们之间的亲缘关系，就非常困难了。

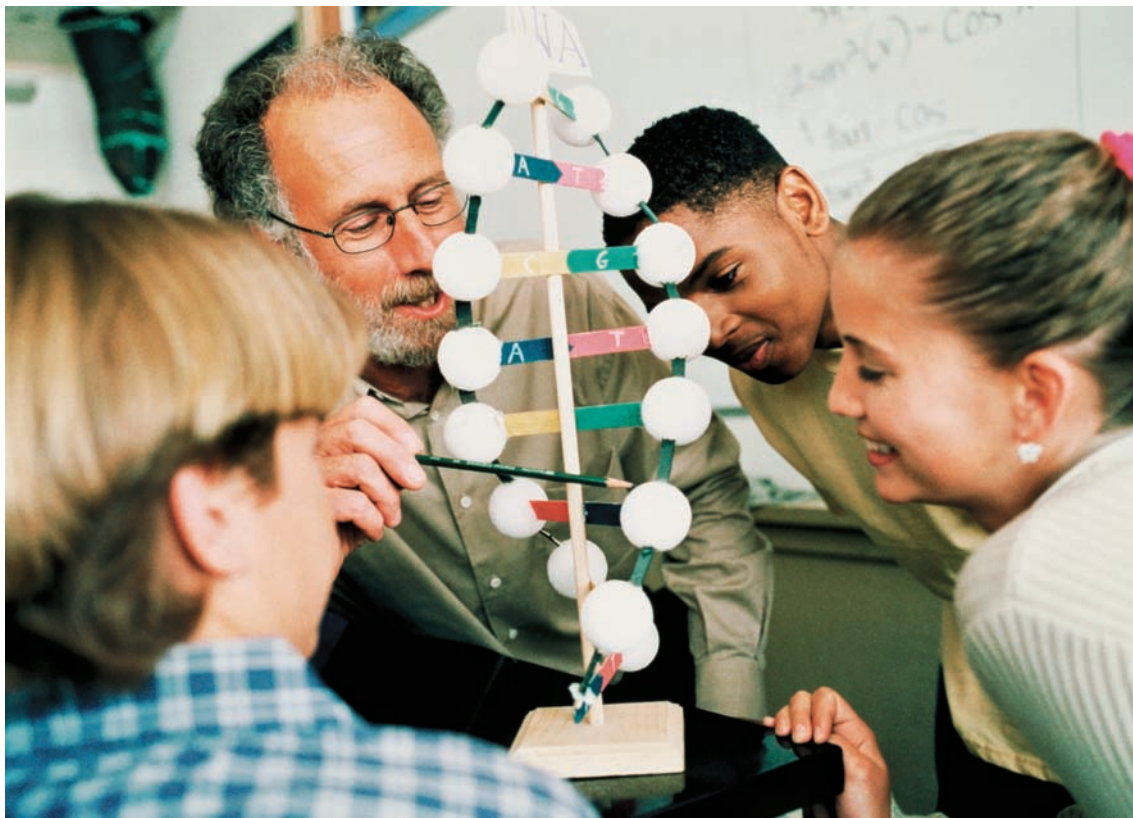
这也就是为什么语言学家总在争吵之中的原

因——那些在一名语言学家眼中，看上去具有密切关联的词汇，在另一名语言学家眼中，可能会是毫无意义。

这时候，其他学科的介入就变得不可避免。考古学的研究发现，在这些拥有稻米同源词汇的语言区域，存在着悠久的稻米种植历史。在浙江河姆渡遗址，曾经出土了距今约7000年的稻米遗存，长江中下游也因此被认为是稻米的起源地之一。而在其他区域同样也发现了稻米的遗存，在中国台湾和云南分别是距今5000年和4000年；泰国距今4000年，菲律宾距今3700年。考古学家们推测，会不会存在两条稻米传播路线，一条从长江中下游——台湾——菲律宾，一条是长江中下游——云南——缅甸、泰国？如果是这样，这又和语言学家推断的南岛和南亚语系的发展路线有许多重合之处。

语言学家解释说，之所以会出现稻米的传播与语言的传播有相似之处，是因为农业社会比采猎社会更能有效而稳定地增加人口。而随着人口的增加，就不可避免地带来种群的分化和迁徙，这也就是语言传播的动力之一。

除此之外，早在1959年，哈佛大学考古学家张光直就曾根据出土的早期考古资料论证，黄河流域前石



▲ 老师还在给不同肤色的学生讲解DNA双螺旋结构。图/华盖

器文化当中有一个尚未显出差别的“汉藏南岛综合体”。他认为从这个群体从新石器时代起,开始向中国南部移民的活动,结果分化出“汉藏群”和“南岛群”。难道这依然是个巧合?

来自基因的答案

越来越多的发现似乎在暗示着,在远古时期,东亚语言的确是从一种最古老的“母语”分化发展而来。但是由此产生的疑问却更多,如果存在这样一种假说的话,那么,这个“母语”的起源地到底在哪里?内部这种千差万别的语言又是从什么时候开始分化?它们之间的亲疏远亲如何?这似乎又完全超出了语言学目前所能解释的能力。

同时,语言学家内部反对这种假说的声音也从来没有停止过,每当支持东亚同源的证据多一分,似乎就会出现更多的相反意见。

1999年,一个偶然的机,潘悟云看到了一篇文章。这是一个生物学家介绍分子人类学家们关于东亚人群起源的最新观点。在这篇文章中,遗传学者用

DNA证实,东亚人群是距今4万年左右由非洲迁徙而来,然后从南向北迁徙分化,最后形成东亚大地丰富的民族。而此前,学界的普遍看法是,东亚现代人是本土直立人、智人连续进化而来。

潘悟云马上联想到,这与语言学家争论的东亚语言的问题息息相关。“同源”、“自南向北”这两个来自DNA的结论,居然和之前语言学界的那个东亚语言同源的假设一致。

早在1859年,达尔文就曾在《物种起源》中写到,“如果我们拥有一份完备的人类谱系图表,那么人种的形成序列将为全世界使用的各种不同的语言提供最好的分类。”而这句话,现在听上去,就好像是在为分子人类学家们今天所作的工作做注解。

随后不久,一篇从分子人类学角度更加详细的论述东亚人群起源和迁徙的论文出现了。在这篇文章中,学者们认为,东亚族群的祖先最初达到了东亚南部,约在在约20000到40000年前开始分化,一支自越南进入中国,并沿海岸线向北,成为百越和东夷的先祖,其中一个亚群在北部湾一带就转而南下,最终成



后人所绘巴比伦的通天塔。根据《圣经》记载，是挪亚的后人建造的，目的是通过它直上云霄，到达天堂。据说上帝耶和华看见人类居然拥有如此的智慧，不禁感到担忧。于是，他开始让人类使用不同的语言，令他们无法更好地沟通。因此，人类建造通天塔和理想城市的计划受阻，人类也因为语言不通而彼此分隔开来。资料图片

为南岛语族的先祖。

而另一支则向自云南穿过四川一直到达黄河中上游盆地，成为汉藏羌的祖先。其中一个亚群在约8000—6000年左右向东向南扩展，最终成为华夏族群的先祖。而另外的一些亚群则调头向西南发展，最终成为藏缅语族的祖先。

尽管这个来自基因的迁徙路线也仅仅是一个草图，但那些在语言学界长期悬而未决的难题似乎都能得到合理解释了。

例如，为什么差异巨大的东亚各个语系之间，却存在千丝万缕的联系？按照基因图谱来解释，那的确是因为他们拥有共同的祖先，但由于早在数万年前就开始分化，所以才会呈现出迥然相异的面貌。

而曾经长期存在归属争论的苗瑶语和侗台语看上去也各自有了结论——苗瑶语族群是约1.5万年前从北上过程中的汉藏羌那一支分化而来，最终达到了中国中部。而侗台语则是从海岸线北上那一支中分化而来，自然会与南岛语系有更紧密的关系。

那么汉语呢？那支从古汉藏羌分化而出，沿黄河

流域而下的亚群所讲的语言就是上古的汉语吗？显然没那么简单！还记得有一支沿南中国沿海一直上行达到山东的族群吗，他们成为东夷部落的先祖，而东夷部落的一部分沿黄河流域上行。最终两个族群在中原地区遭遇，在这其间，甚至还不可避免地遇到了从南方北上而来的部分苗瑶族群，以及从北方南下而来的部分阿尔泰语系的族群，他们征战并相互融合，无论是族群还是语言。

语言学者李葆嘉是这样描述这个复杂的过程的，“原始的华夏汉语就是一种伴随着多种文化撞击整合为强大中原文明的历史互动过程中，所产生的多元性、层叠性混合语言。”

而这就是为什么语言学家们在汉语中，既能与藏缅，又能与苗瑶、侗台、甚至与南岛、阿尔泰都找到缘关系的原因。

这听上去似乎清晰了不少。但对于那些仍然还未能解答的谜团来说，显然还有更漫长的路要走。 