

语言、考古和遗传学多学科视角下的 印欧语系起源与扩散

杨文姣 孙瑾 郭健新 王传超

摘要 印欧语的起源与扩散研究从首次提出至今已走过了近百年的历程，当前学术界主要集中于库尔干假说和安纳托利亚假说，对于印欧语的起源及扩散路径，国内外研究丰富且说法庞杂，目前尚未达成共识。本文从考古学、语言学和遗传学多角度出发，梳理了有关印欧语系起源和扩散问题的主要假说，评述和对比了各学科学者的研究成果及其局限性。综合多学科证据，我们推断印欧语起源于高加索山南部地区到安纳托利亚一带，经由东欧早期的农业人群传入东欧大草原，尔后经由以颜那亚文化为代表的青铜时代早中期的东欧草原游牧人群传播至欧洲、中国西北（吐火罗语）和南亚。

关键词 印欧语系 考古学 语言学 遗传学 人群迁徙

作者杨文姣，厦门大学社会与人类学院人类学研究所博士生宁夏大学外国语学院讲师（福建厦门 361005）；孙瑾，厦门大学社会与人类学院博士，兴义民族师范学院讲师（贵州兴义 562400）；郭健新，厦门大学社会与人类学院博士后（福建厦门 361005）；王传超，厦门大学社会与人类学院人类学研究所教授（福建厦门 361005）。

中图分类号 H7

文献标识码 A

文章编号 0439-8041(2021)05-0175-11

一、引言

印欧语系（Indo-European languages）是世界上最大的语系，主要分布在欧亚大陆西部和南部，涵盖了欧洲的大部分、印度次大陆北半部以及伊朗高原的语言，共包括 445^① 种语言和方言，其中较为常用的英语、法语、葡萄牙语和西班牙语，在殖民主义扩张中逐步发展壮大，现已遍及世界各大洲。世界人口的 46%（32 亿）^② 以印欧语系为第一语言，是所有语系中使用人数最多的。印欧语系又被分为若干语族，包括日耳曼语族（Germanic）、意大利语族（Italic）、凯尔特语族（Celtic）、波罗的-斯拉夫语族（Balto-Slavic）、印度-伊朗语族（Indo-Iranian）、希腊语族（Greek）、阿尔巴尼亚语族（Albanian）、亚美尼亚语族（Armenian）、安纳托利亚语族（已消亡）、吐火罗语族（已消亡）等。

印欧语起源是一个古老而又历久弥新的经典话题，对其研究首先来自语言学的猜想与实践，逐步转向多学科的实证研究。荷兰学者马库斯·冯·鲍霍恩（Marcus Zuerius van Boxhorn）首先注意到欧洲各种语言、梵语和波斯语之间的相似之处。一个多世纪后，威廉·琼斯（William Jones）爵士在印度学习梵文

① 数据来源：Ethnologue, <https://www.ethnologue.com/subgroups/indo-european>.

② 数据来源：Ethnologue list of language families 22th. Ethnologue.com. 25 May 2019. https://www.webcitation.org/61AIADZgn?url=http://www.ethnologue.com/ethno_docs/distribution.asp?by=family.

后,发现梵语、希腊语、拉丁语之间存在相似对应关系,并于 1786 年孟加拉亚洲协会的年度演讲中发表他的结论,即梵文与希腊文、拉丁文来自同一源头。德国语言学家、梵语专家弗朗兹·葆朴(Franz Bopp)对此进行了系统论证,并于 1816 年出版《论梵语动词变位系统—与希腊语、拉丁语、波斯语和日耳曼语相比较》。琼斯的重要发现开启了历史比较语言学的研究,语言学家开始尝试重建原始印欧语并确认各语言之间的亲属关系,在对欧洲语言和古代的梵语、拉丁语、希腊语等进行大量深入比较后,不仅发现了许多同源词(cognate words),语法构词和语音上也存在相似之处,故此推断这些相关语言属于同一个完整的语系,即印欧语系。

对印欧语起源的研究最早来自语言学的探索,旨在为印欧语系确定系属地位。随着研究不断深入,印欧语起源的谜团受到了神话、语言学、人类学、遗传学和考古学等领域的关注,也成为各个学科较有争议的一个重大而复杂的课题。对印欧语起源地的推测众说纷纭,有“波罗的海—黑海起源说”“安纳托利亚起源说”“中欧—巴尔干起源说”“东欧大草原起源说”^①。此外还出现了“印度起源说”^②“旧石器连续范式(The Paleolithic Continuity Paradigm)”^③等。目前学术界主要集中于库尔干假说(东欧大草原起源说)和安纳托利亚假说(农业起源说),对于印欧语的起源及扩散路径,国内外研究丰富且说法庞杂,目前并未达成共识。为此,本文将对关于印欧语起源与扩散的两种主要假说从语言学、考古学、遗传学角度进行梳理,试图总结目前最新的跨学科协同研究如何为印欧语源流解读提供新参考。

二、印欧语系起源和扩散假说

长期以来,印欧语的起源和扩散问题一直是多学科争论的焦点。考古学家、语言学家、遗传学家从物质文化考古材料、历史语言学的语系重构、古 DNA 全基因组数据分析切入,对印欧语的起源地和迁徙路线进行了深入阐释。根据学界对印欧语起源地的猜想和假设,目前以库尔干假说和安纳托利亚假说为最具争议性的热点。

(一) 库尔干(Kurgan)^④假说

1. 考古学

库尔干假说是最为广泛接受并长期占统治地位的关于印欧语起源的模型。该假说认为印欧人起源于“坟冢文化”,即东欧大草原上的颜那亚文化(Yamnaya culture)。19 世纪至 20 世纪,考古学受到了语言学对印欧语系和各语支亲属关系如火如荼研究的影响,也开始探究印欧语和印欧人的起源问题,主要采用纯考古学的方法。考古学的研究是基于文化传播理论而展开,该理论认为某种考古文化可能与某个特定人群的活动相对应,而该文化发展与传播则与此人群的迁徙活动密切相关;如果能够找到并确定属于印欧人的最早的考古文化及其发展轨迹,那么其起源和迁徙问题便迎刃而解。^⑤

英国考古学家戈登·柴尔德(Gordon Childe)^⑥1926 年提出印欧语系起源于乌克兰黑海以北、俄罗斯西部的东欧大草原地带,他用印欧语中的动植物名称来描绘印欧人的居住环境,又以“马匹”“车轮”等词确定其所处年代。柴尔德的观点在立陶宛裔考古学家玛丽亚·金布塔斯(Marija Gimbutas)的著作《史前的东欧》(*The Prehistory of Eastern Europe*)^⑦中得到进一步引申和论证,并于 1963 年首次提出“库尔干假说”,根据考古墓穴中发现的冶金术、青铜兵器、战士和马匹遗骸,推断早期操印欧语的游牧

① Mallory, J. P. & Adams, D. Q., *Encyclopedia of Indo-European Culture*, London: Fitzroy Dearborn Publishers, 1997, pp. 297-298.

② Elst, K., *Update on the Aryan Invasion Debate*, New Delhi: Aditya Prakashan, 1999.

③ Alinei Mario, "Interdisciplinary and Linguistic Evidence for Palaeolithic Continuity of European, Uralic and Altaic Populations in Eurasia," *Quaderni di Semantica*, 2003(24).

④ 库尔干(Kurgans): 东欧大草原人建造的土坟墓。

⑤ 王欣:《印欧人的起源与吐火罗人的迁徙:学术史的回顾与方法论的思考》,《暨南史学》2013 年。

⑥ V. Gordon Childe, *The Aryans: A Study of Indo-European Origins*. London: K. Paul, Trench, Trubner & Co.; New York, A.A. Knopf, Inc. 1926.

⑦ Gimbutas, Marija, "Part I: Mesolithic, Neolithic and Copper Age Cultures in Russia and the Baltic Area," *The Prehistory of Eastern Europe*, American School of Prehistoric Research, Harvard University Bulletin, No. 20, Cambridge, MA: Peabody Museum, 1956.

民族在公元前 4500 年到公元前 2500 年之间从东欧大草原迁徙而来，他们向西扩张进入东欧、中欧和北欧地带，向南到达了北高加索、小亚细亚一带，向东抵达乌拉尔山东部。

考古学家大卫·安东尼（David Anthony）结合考古学和语言学材料，试图寻找欧亚大草原早期马车和马匹驯化的证据，以期探索印欧人的起源与迁徙。《马匹、车轮与语言——青铜时代来自欧亚草原的骑士如何构建现代世界》（*The Horse, the Wheel and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* 2007）被认为是对库尔干假说的进一步修订。所有重大的扩张都是由于某种技术上的创新所造成的。^①因此，考古学家以马拉战车作为参照标准来研究操印欧语民族的物质文化传播。安东尼对近东地区的马匹驯养过程、骑马的历史、制作马车工具等历史的研究进一步论证了印欧语言的起源地在欧亚草原，操印欧语的游牧民族是马拉战车的创造者和使用者，有了交通工具的助力，操印欧语的人经过长途跋涉，实现了印欧语横扫欧亚大陆的扩散。其中，最早的安纳托利亚分支由草原人群穿过巴尔干半岛带到了小亚细亚。此外，草原社会政治经济的一个特点是牛、马或羊的祭祀被阐释为盛宴，可能存在于葬礼或其他宗教仪式中，这些盛宴可能是古印欧语发展的一个重要载体，它是考古学和语言学证据之间的联系。更具体来说，大草原的动物祭祀和社会等级与古印欧语有关，如古印欧语中含有表达“给与、赠予、分配、奖赏、合法接受”等完整循环的词语，且这些词在安纳托利亚语、托查利亚语、吠陀语和亚美尼亚语中能找到同源词。两位学者（Anthony & Ringe）将古印欧语中表达政治经济的语言学证据和草原考古相结合，发现二者相当吻合。对于印度-伊朗语为何系属印欧语大家族，还要看印度-雅利安民族的主要成分是否来自游牧民族。印度史诗中也记载了那些骑着高头大马的民族就是草原上的游牧人群。此外，一些学者（Anthony, Mallory, Kuzmina）认为印度-伊朗语也起源于黑海以北的大草原，向东迁徙至里海，在公元前 2000 年左右向伊朗和印度迁徙。要理清印欧语的扩散路线，还需要回答吐火罗语和吐火罗人的来源问题。研究印欧语起源的学者对吐火罗语的资料都很重视（Henning, Gamkrelidze, Ivanov）。^②考古学家认为，吐火罗人起源的关键是阿尔泰山与天山之间的克尔木齐文化，该文化源于东欧大草原的颜那亚文化，其中的一支向东迁徙至阿尔泰山南麓并带来了小麦和吐火罗语。^③与颜那亚文化一脉相承的阿凡纳谢沃文化（Afnasievo culture）、中亚青铜时代的安德罗诺沃文化（Andronovo culture）对印欧语扩散至伊朗高原和印度北部、新疆塔里木盆地也产生了重要影响。语言不仅能通过人群的流动和迁徙而传播，还能随着动物的驯化和交通工具的发明而加速传播，在欧亚大草原的考古墓穴中发现的马匹、马车等重要遗骸，为我们提供了那个时代的物质文化细节，包括生活习惯、墓葬方式、图腾崇拜等，这些都是操印欧语的草原人群所共有的、明显的深层次文化特征，正是这些考古文化展现了不同地区共享一种文化的连续性，将操印欧语的游牧人群联系在一起，为我们展示出印欧语扩散的路线。语言区域、考古文化、人群之间的关系是相互对应又互相影响的。

2. 语言学

受到达尔文进化论的启发，施莱歇尔（Schleicher）将生物学的方法应用到印欧语的谱系研究中，为语言学引入了谱系树结构。他的学生施密特（Schmidt）认为谱系树图只能表现语言的纵向发展，肯定存在语言间互相接触而带来的横向扩展，所以，施密特提出了波浪理论，这些都是早期对印欧语系探索的语言学研究成果。此外，施莱格尔（Friedrich Schlegel）、格林（Jacob Grimm）、德尔布吕克（Delbrick）、惠特尼（Whitney）等欧美学者也都致力于印欧语研究，寻找印欧语的共同祖先。语言学家通过建立以发音和句法变化为基础的模式，来确定不同支系语言之间的关系，以此重构原始印欧语系。单纯对语言进行重构来理解语言扩散过程是片面的，因此，语言学家和考古学家重点结合能够反映各个族群物质文化生活的词汇，一个有力旁证就是将马匹、马拉战车与原始印欧人联系在一起。早期印欧语系民族尽管相距遥远，但

① 卡瓦利·斯福尔扎（Cavalli-Sforza L. L.）：《追踪亚当与夏娃——从演化历史看基因、民族和语言的关系》，吴一丰、郑谷苑、杨晓珮等译，台北：远流出版事业股份有限公司，2003年，第138页。

② 徐文堪：《吐火罗人起源研究》，北京：昆仑出版社，2005年。

③ 林梅村：《吐火罗人的起源与迁徙》，《西域研究》2003年第3期。

是他们使用的与马匹、马具、马拉战车有关的词汇基本相同，如与“轴（axis）”同源的词也出现在日耳曼语族、意大利语族、凯尔特语族、波罗的-斯拉夫语族、印度-伊朗语族、希腊语族的部分语言中，且“轴”不能离开轮式车辆而单独存在^①，因此有力地证明了马车在早期印欧民族中广泛存在。库尔干假说的支持者采用“语言古生物学”的研究方法，通过分析总结各分支语族中关于气候、动物和植物分布的共同词根来推断印欧语系的起源地，如印欧语中都有“雪”和“冬”的同源词，由此推断那里的气候比较寒冷。“狼、熊、鲑鱼”等属于北温带动物的词汇也广泛存在于印欧语系各族的语言中。此外，古印欧语中有柳树（Willow）和橡树（Oak）等词，但没有橄榄树（Olive）和棕榈树（Palm），因此可以界定，印欧人的生活范围大致是在欧洲大陆到俄国南部地区^②。如果是草原游牧民族发明了马具、缰绳、车轮和车厢，说明马拉战车极有可能产生于草原与农业社会接壤的地区，这个地区应该出产树木和青铜^③，这与考古学家安东尼的设想一致。印欧语起源于东欧大草原的另一重要证据是在重建印欧语时发现了许多与游牧民族有关的词汇，如羊毛、牲畜、马、乳制品等。马洛里（Mallory）也曾是东欧大草原假说的坚定支持者之一，他将词汇地理分析法、词汇考古法应用在印欧语地理位置的研究上，通过重建的家畜词汇“绵羊”“山羊”“牛”“猪”“狗”和“谷物”以及“镰刀”“磨石”“陶器”等农业工具词汇，来证明在公元前 7000 年前，农业或新石器时代的经济不可能存在。^④换言之，通过对词汇的构拟和重建，语言学的例证更倾向于将印欧语系的起源地定为畜牧业发达的东欧大草原。

柴尔德在语言学成果的基础上，建立了一个“核心”词汇库，并推测这种核心语言是从本土使用的原始印欧语中保留下来的，核心词汇中有动植物的词描绘当时的生活环境，铁和青铜没有核心词，但有马和车轮的词。因此，印欧人最初的散布似乎在青铜时代开始之前就有了地位。随后，柴尔德把这些语言学观点与考古学证据联系起来，他认为绳纹陶器等文物是游牧民族的遗物，这些装备齐全的游牧民族驾着马车在青铜时代初期从他们的草原家园迁移到黑海北部。^⑤也就是说，他们是东欧大草原上的印欧人。受进化论影响的还有语言学家（Chang, Cathcart, Hall & Garrett），他们借用生物学的方法来构建进化树，分析了所有语言后发现印欧语的起源可以追溯到大约公元前 4000 年前^⑥，与东欧大草原假说中考古学根据墓穴挖掘器物所推断的时间一致。

3. 遗传学

21 世纪以来，随着遗传学的发展和古基因组学新领域的崛起，古人类和动植物的 DNA 分析成为目前研究人类史前历史的变革性技术，古基因组学通过对古人类遗骸进行 DNA 分析，整合古今人群遗传学数据，为探索人口迁徙、人群间相互关系、人类起源和演化历史等开辟了新思路。来自古 DNA 的信息与考古学和语言学证据一样举足轻重，尤其在解决一些像语言起源、人类迁徙等难以衡量的问题上能够提供更加科学的证据。生物学家和遗传学家可以利用线粒体 DNA 和 Y 染色体追寻母系或父系任何一方的祖先，族群的迁徙势必会带动语言的流动，这必定为寻找印欧语起源地打开了另一扇大门。

越来越多的遗传学家、分子人类学家开始关注语言和考古，随着古基因分析技术的突飞猛进，古 DNA 样本数据的增加，古基因组学也取得了一批突破性的成果，为印欧语源流、南岛语族起源、东亚人群迁徙等科学问题提出新证据。哈佛医学院的进化和人口遗传学家大卫·赖克（David Reich）团队发布了生活在 8000 至 3000 年前欧洲的 94 个古人的 DNA 数据，证实了中东农民在 8000 至 7000 年前来到欧洲。早期安纳托利亚假说中支持的新石器时代早期农业人群在欧洲立足，数量如此之大以至于后面几波迁徙无法

① Mallory, J. P., & Adams, D. Q. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*, Oxford University Press, 2006.

② 转引自刘先宽：《印欧语系的起源及特点》，《锦州医科大学学报（社会科学版）》2008 年第 6 期。

③ 刘欣如：《从雅利安人到欧亚游牧民族：探索印欧语系的起源》，《历史研究》2011 年第 6 期。

④ Mallory, J. P., & Adams, D. Q. *Encyclopedia of Indo-European Culture*, London: Fitzroy Dearborn, 1997, p. 295.

⑤ 转引自 Renfrew, Colin. “The Origins of Indo-European Languages,” *Scientific American*, 1989(261), pp. 106-115.

⑥ Chang, W., Cathcart, C., Hall, D., & Garrett, A. “Ancestry-constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis,” *Language*, 2015 (1).

替代。但是哈克 (Haak) 等学者^①通过对颜那亚绳纹器文化有关的草原游牧人群古基因组进行分析,判断这些人曾居住在约 5000 年前的俄罗斯和乌克兰地带,他们在遗传学上与德国 4500 年前(属绳纹器文化)的人群极为相似,由此推断大量移民从欧洲东部进入到欧洲中心地带,替换掉 75% 的中欧祖先,同时可能携带了一种早期印欧语言,这一研究结果为草原假说及印欧语系扩散到欧洲提供了强有力的证据。艾伦托夫特 (Allentoft) 等学者^②也发布了来自欧亚大陆的 101 个古人的低覆盖率基因组,数据分析表明,在青铜时代早期,颜那亚人从东欧大草原迁徙到北欧和中亚,其基因证据与假设的印欧语扩散路线一致。颜那亚人早期分开后,一支向东迁徙至阿尔泰山,一支向西迁徙至欧洲西部,旁证了新疆塔里木盆地发现的吐火罗语为何与原始印欧语的西部分支,如日耳曼语等具有较多相似性,而与东部语支几乎没有共同点。青铜时代是一个高度动态的时期,大规模的人口在这阶段迁徙和更替,形成了今天欧洲和亚洲人口结构的主要部分。

古 DNA 证据考释了欧洲经历了两次人群大融合,第一次是 9000 年前安纳托利亚地区的农民随着农业扩张进行的大规模移民,第一次移民并没有伴随大规模的语言扩散。第二次是 5000 年前来自欧洲大草原的游牧民族大迁徙,将语言传播至欧洲大陆以及南亚。在此推断上,赖克团队又提出一个较为折中的假设,即印欧语系可能起源于高加索地区以南,即今天的伊朗或亚美尼亚,那里是颜那亚人和古安纳托利亚人的共同源头,其中一支迁去了欧亚大草原,在那里以 1:1 的比例与采集狩猎者融合,另一支去了安纳托利亚^③。此外,王传超等学者^④发布了北高加索地区自 6500 年前到 3200 年前跨越 3000 多年的 45 个古代遗骸的高质量全基因组水平的数据,发现与该地区 6300 年前铜石并用时代的古人群相比较,颜那亚文化人群和所有后来的草原游牧人群都带有先前未被发现的与周边的东欧农业人群相关的祖源成分,该农业人群的遗传成分最可能是欧洲早期的农业人群经由西侧路线带入到欧亚草原上的,同时也可能把原始的印欧语带到了草原上,这一研究为高加索山南部是印欧语的起源地提供了可能性,并为印欧语系的早期分支是安纳托利亚语作出了解释。对于印欧语东北方向扩散路径的研究,宁超、王传超等学者^⑤发布了新疆东北部距今约 2200 年的天山北麓石人子沟遗址 10 个古人个体的全基因组数据,发现该遗址的古人是典型的混合人群,其一部分祖源成分主要来自东北亚地区,而另一部分主要是来自阿凡纳谢沃、颜那亚等青铜时代早中期欧亚草原游牧人群,表明至少一部分铁器时代的新疆人群来自欧亚大陆西部的草原^⑥。研究语言的起源和演化历史离不开对人群及其迁徙路线的探讨,新疆石人子沟遗址古 DNA 的发现为印欧语系的起源和扩散至新疆的路径提供了关键线索。古 DNA 的研究为印欧语在欧洲的源流探索提供了更加科学的证据,而宁超和王传超等中国学者则为印欧语向欧洲东部至亚洲的扩散打开了新的通道,来自古人类基因的信息与考古学文物、语言学材料相契合,属于印欧语的吐火罗语被发现于新疆塔里木盆地也直接证明了印欧语系人群向东迁徙的路径。2021 年王传超团队又在《自然》(Nature) 上发文继续证实了在 5000 年前青铜时代早中期颜那亚草原游牧人群从欧洲向东扩张进入蒙古,揭示了颜那亚人群向东扩散的范围非常广,直达蒙古高原,这一发现为颜那亚人来过阿尔泰山,并向中国新疆传播印欧语(吐火罗语)提供了强有力的支持。^⑦另外,纳拉辛汉 (Narasimhan)^⑧等学者对来自中亚和南亚最北部的 523 个 8000 年前的古代 DNA

① Haak, W., Lazaridis, I., Patterson, N. et al. "Massive Migration from the Steppe was a Source for Indo-European Languages in Europe," *Nature*, 2015, 522(7555), pp. 207-211.

② Allentoft, M., Sikora, M., Sjögren, K.G. et al. "Population genomics of Bronze Age Eurasia," *Nature*, 2015 (522), pp. 167-172.

③ 大卫·赖克 (David Reich):《人类起源的故事》,叶凯雄、胡正飞译,杭州:浙江人民出版社,2019年,第144页。

④ Wang, C.C., Reinhold, S., Kalmykov, A. et al. "Ancient Human Genome-wide Data from a 3000-year Interval in the Caucasus Corresponds with Ecogeographic Regions," *Nat Communication*, 2019 (10).

⑤ Wang, C.C., Reinhold, S., Kalmykov, A. et al. "Ancient Human Genome-wide Data from a 3000-year Interval in the Caucasus Corresponds with Ecogeographic Regions," *Nat Communication*, 2019 (10).

⑥ Ning, C., & Wang, C.C. "Ancient Genomes Reveal Yamnaya-related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan," *Current Biology*, 2019 (29).

⑦ Wang, C.C., Yeh, H.Y., Popov, A.N. et al. "Genomic Insights into the Formation of Human Populations in East Asia," *Nature*, 2021, pp. 413-419.

⑧ Narasimhan, V. M., Patterson, N. J., et al., "The Formation of Human Populations in South and Central Asia," *Science*, 2019 (365).

全基因组水平的数据进行分析后,推断出公元前 3300 年开始横跨欧亚大草原的颜那亚人分两路迁徙,一路向西达到欧洲西部和南部,另一路向东直抵阿尔泰山,其后裔在一千年后开始出现在大草原以南的中亚地区,随后在公元前 2000 年前半期又向南亚传播。东西方学者从古基因的角度全方位演示了颜那亚草原人群涌入欧洲、阿尔泰山和南亚的路线(如图 1),同时也为我们勾画了印欧语扩散的图景。



图 1 颜那亚草原游牧人群的迁徙路线(改自: Narasimhan et al 2019)

在高加索山脉、新疆、蒙古国发现的古 DNA 样本中含有颜那亚人祖源成分,这从基因的角度反映了颜那亚草原游牧人群在欧亚大陆的连续性,操印欧语的人群出现在这些地区,带来了吐火罗语,创造产生了物质文化,与考古学所描述的颜那亚文化和阿凡纳谢沃文化的连续性相吻合。在北高加索,遗传矩阵与语言矩阵完全相关,甚至群体亲缘关系的遗传树也几乎与他们的语言树相同。^① 语言学和考古学推论在顺序和年代方面的吻合度非常高,语言与基因的演变是类似的,从遗传学角度测算的印欧语人群扩散时间与语言学和考古学时间基本匹配。在最初探索印欧语起源的进程中,缺少文献材料是语言学家面临的一大挑战。语言学材料的缺乏,必然需要借用考古学资料来旁证,古 DNA 的研究也提供了更为科学的依据。同时,基因数据的分析也离不开语言学和考古学的参照。在多学科群策群力的贡献下,印欧语的起源似乎有了更充实的论证。

(二) 安纳托利亚假说

1. 考古学

安纳托利亚假说(农业扩散假说)是由英国考古学家科林·伦福儒(Collin Renfrew)在其著作《考古和语言》(*Archaeology and Language*)中首次提出,他认为印欧语的家乡在新石器时代的安纳托利亚,并且通过和平而不是马拉战车征服的方式扩散至欧洲^②。伦福儒从考古学角度出发,推断印欧语系的扩散与公元前 6500 年开始的农业经济扩张有关,随着语言传播者的人口增长和经济发展而散布出去,替换了沿途的狩猎采集者语言。这一想法从理论人类学的角度来看非常勇敢。伦福儒将印欧语的扩散具化为两个阶段:1) 公元前 6500 年左右,位于安纳托利亚的前原始印欧语分裂为“安纳托利亚语”和“古原始印欧语(Archaic Proto-Indo-European)”两个分支。说前原始印欧语的农耕者携其农业技术进入欧洲;说古原始印欧语的居民出现在巴尔干、多瑙河流域。2) 公元前 5000 年左右,古原始印欧语又分裂为“西北印欧

^① Oleg B, Khadizhat D, Anna D, et al. "Parallel Evolution of Genes and Languages in the Caucasus Region," *Molecular Biology & Evolution*, 2011(10): 2905-2920.

^② Renfrew, Colin., "The Origins of Indo-European Languages," *Scientific American*, 1989.

语”“巴尔干原始印欧语”和“早期草原原始印欧语”（吐火罗语的祖先）^①。农业起源说的成熟离不开澳大利亚考古学家彼得·贝尔伍德（Peter Bellwood）的研究，他结合考古学和语言学材料来考察新石器时代农业人群的扩散，认为语系的起源地应该在地理上与农业起源地重合或交叉，确信早期农业人群确实传播了其语系的原始语言。^② 8500年前印欧语从安纳托利亚来到欧洲，支持这一说法最好的证据认为，主体语言的替换条件需要大规模的人口迁徙，在新石器时代早期农民定居欧洲后，人口基数可能非常大，以至于后来的迁徙不会产生太大的影响^③，但此说法受到支持草原假说的学者的挑战。此外，巴尔干半岛新石器时代早期文化与安纳托利亚文化有着十分密切的联系，新石器时期人群开始沿着地中海海岸线向西迁徙，早期安纳托利亚的农业发展和人口增长推动了农民向欧洲大陆的迁徙（如图2）。一旦农业人群向外迁徙，必然会带动语言的传播，印欧语伴随着农民的迁徙向外扩散。人口的大量增长将农民和他们的语言推向了更广阔的领域。



图2 前安纳托利亚早期农民进入欧洲的路线（改编自 Wolfgang Haak, et al. 2015）

2. 语言学

根据历史语言学家对原始印欧语构拟的谱系树观察，安纳托利亚语族最早与原始印欧语分化，为操安纳托利亚语的人群首先迁徙提供了直接证据。安纳托利亚语族作为原始印欧语的一个单独分支，与日耳曼语族等同列第二层级。因此，贝尔伍德推断安纳托利亚语必定与印欧语系的起源地有关。此外，印欧语系的安纳托利亚语族中最古老的赫梯语（已消亡）保留着原始印欧语的喉音、颞音、（准）施动格等非常古老的特点，辅证了安纳托利亚是古印欧语的起源地这一假说。^④

与传统的历史比较语言学对原始语系的构拟不同，语言学家和生物学家（Gray, Atkinson, Gray, Drummond & Greenhill, Kitchen, Ehret, Assefa & Mulligan）借用生物学谱系发生学原理与方法，使用计算机算法程序，结合同源词和词法距离矩阵等对印欧语的谱系树构建及其分化进行研究。

俄罗斯语言学家（Gamkrelidze & Ivanov; Gamkrelidze; Dolgopolsky; Klimov）根据原始印欧语系、原

① 转引自徐文堪：《欧亚大陆语言及其研究说略》，兰州：兰州大学出版社，2013年，第114页。

② 彼得·贝尔伍德（Peter Bellwood）：《最早的农人：农业社会的起源》，陈洪波等译，上海：上海古籍出版社，2020年，第268页。

③ Renfrew, Colin. *Archaeology and Language: The Puzzle of Indo-European Origins*, New York: Cambridge University Press, 1987. Bellwood, Peter. *First Farmers: The Origins of Agricultural Societies*, Malden, MA: Blackwell Pub, 2005.

④ 周流溪：《印欧语言起源新说》，《外语教学与研究》1991年第2期。

始闪米特语（可能位于黎凡特北部）和原始卡特为利语（高加索地区三个语族之一）间的借词所发生的地理位置推测安纳托利亚的东部为原始印欧语系的起源地^①，其分支语言在先向东后向西的迁徙过程中分化。甘克里雷泽和伊万诺夫（Gamkrelidze & Ivanov）从语音、语法构词、语义等层面对原始印欧语进行全面重构，将印欧语的传播路线归纳为从安纳托利亚向西进入希腊和巴尔干半岛，向东经高加索再向西进入黑海以北的欧洲地区。^②

新西兰学者格雷（R. D. Gray）和阿特金森（Q. Atkinson）^③用贝叶斯分析法对 87 种语言中的 2449 个词汇的矩阵进行分析，估算出最初印欧语系分化的时间大约在 7800 年到 9800 年之间，与安纳托利亚假说所表述的印欧语系在 8000 年至 9500 年前左右随着农业的发展从安纳托利亚开始扩散的说法相吻合，此外，根据所构建的进化树和分化时间来看（参见图 3），赫梯语（安纳托利亚语最古老的分支）早在 8700 年前左右与原始印欧语最先脱离，其次是吐火罗语，因分化时间太早，所以吐火罗语和其他印欧语东支几乎没有同源词。格雷等学者将类似遗传学家计算 DNA 序列的方法应用到语言计算中，他们对印欧语系的研究是继对南岛语族起源与扩散研究^④后的又一次新探索，进一步验证了遗传学和生物学方法对测算语言分化年代的适用性。

布兰查德（Blanchard）基于斯瓦迪士在 20 世纪 50 年代提出的同源核心词，分析了印欧语中的词法距离矩阵，测出原始印欧语的最初分化时间可以追溯到公元前 7400 年^⑤。雷姆科·布卡特（Remco Bouckaert）团队使用两种来自进化生物学随机模型的定量系统地理推断工具，结合 103 种古代和现代印欧语言的基本词汇数据来确定印欧语系的扩散模型，所得印欧语言树的分化时间和分化位置都与 8000 年至 9500 年前安纳托利亚的农业扩张相吻合^⑥。

在支持草原假说的语言学家看来，语言学构拟和考古学资料似乎并不支持安纳托利亚是印欧语的起源地，因为根据对原始印欧语系的构拟和重建，发现原始印欧人已经懂得驯养马匹和制造马车。但伦福儒 2003 年提出，安纳托利亚分支先从原始印欧语共同体中分离出去，所以并不必须具备“轮”“车”这一类词汇。农业起源假说与草原假说并非完全对立与矛盾，也存在一定的交叉与重叠。安纳托利亚假说的优势在于把印欧语系的传播与考古学上重要事件即农业的传播联系起来。在研究方法上，支持农业扩散假说的学者在语言学研究方面采用了更科学、更大数据化的方法，如利用贝叶斯系统发生学方法分析出更精确的证据^⑦，当传统语言学的研究方法不足以提供充分信息来直接证明语言演化与扩散，来自生物学和遗传学的方法则创造了新的可能性。

3. 遗传学

谈到用基因解密语言起源，遗传学家卡瓦利·斯福尔扎（Cavalli-Sforza）是最早利用遗传学来追溯人类进化和迁徙模式的学者之一，他也赞同印欧语的传播与农业扩张息息相关，认为来自近东的农耕者与沿途的采集狩猎者之间发生了接触与融合，换言之，操原始印欧语的中东农耕者在新石器时代，把印欧语的前身带进了欧洲，包括库尔干地区，这种语言随后才演变为原始印欧语。但斯福尔扎也肯定了库尔干假说和安纳托利亚假说并不必然互相冲突，甚至可以互相增援^⑧。古 DNA 研究显示大多数现代欧洲人的祖先

① 彼得·贝尔伍德（Peter Bellwood）：《最早的农人：农业社会的起源》，陈洪波等译，上海：上海古籍出版社，2020 年，第 274 页。

② Gamkrelidze, Thomas V., and Vjaceslav V. Ivanov. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and Proto-Culture*. Boston: De Gruyter Mouton, 1995.

③ Gray, R. D., & Atkinson, Q. D. "Language-tree Divergence Times Support the Anatolian Theory of Indo-European Origin," *Nature*, 2003(426), pp. 435-439.

④ Gray Russell D. et al, "Language Trees Support the Express-train Sequence of Austronesian Expansion," *Nature*, 2000(405), pp. 1052-1055.

⑤ Blanchard, P., Petroni, F., Serva, M. et al. "Networking Phylogeny for Indo-European and Austronesian Languages," *Nat Prec.* 2009.

⑥ Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J. et al. "Mapping the Origins and Expansion of the Indo-European Language Family," *Science*, 2012(337), P957-960.

⑦ Ryder, Robin J., and Geoff K. Nicholls, "Missing Data in a Stochastic Dollo Model for Binary Trait Data, and Its Application to the Dating of Proto-Indo-European," *Journal of the Royal Statistical Society, Series C (Applied Statistics)?*, 2011(1), pp. 71-92.

⑧ 卡瓦利·斯福尔扎（Cavalli-Sforza L. L.）：《追踪亚当与夏娃—从演化历史看基因、民族和语言的关系》，2003 年，第 220 页。

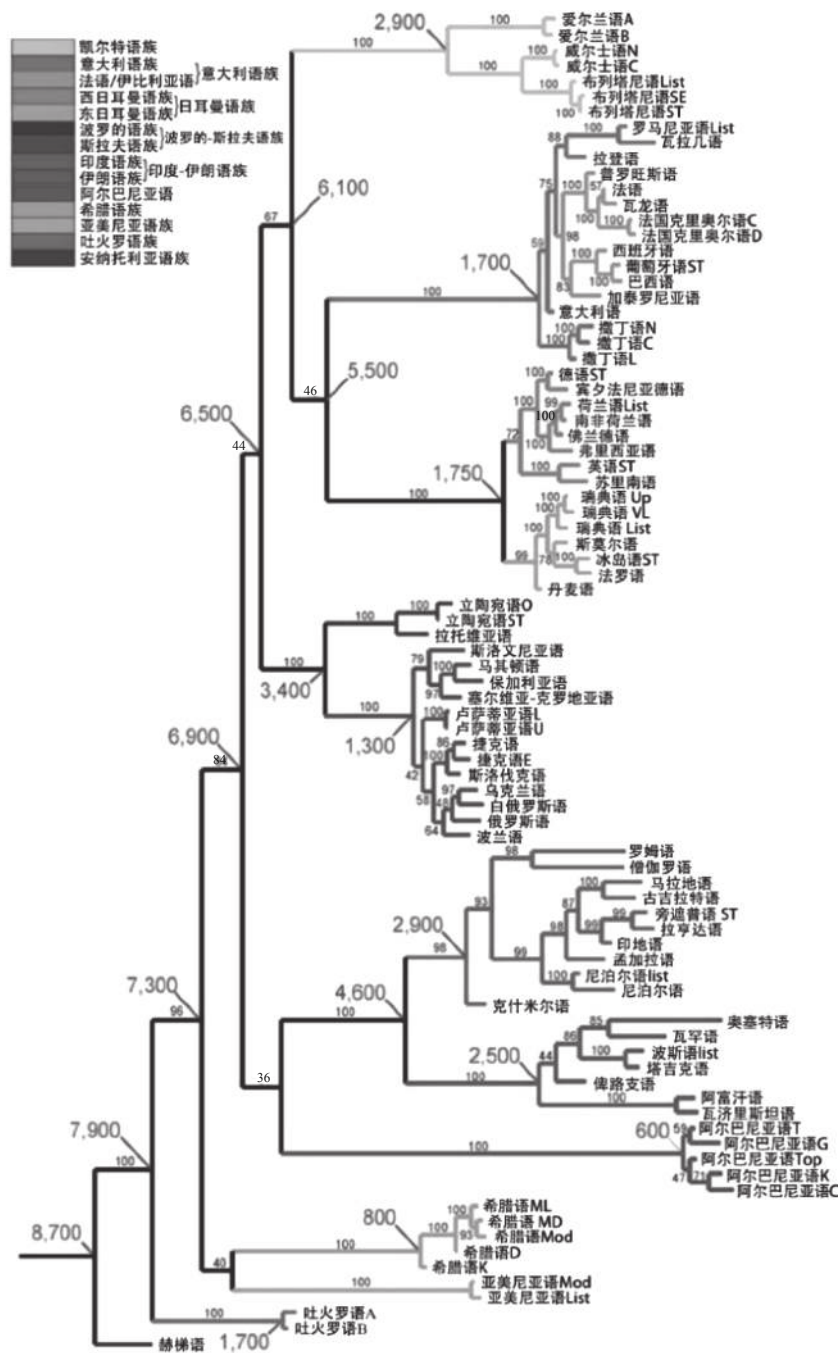


图3 进化树和分化时间预测 (引自 Gray & Q. Atkinson 2003)

来自三个高度分化的群体，即中石器时代的欧洲采集狩猎者、新石器时代安纳托利亚西北部的农民、新石器时代晚期到青铜时代过渡时期的草原游牧人群。^①有学者进一步从古 DNA 角度对安纳托利亚起源及其农业扩散进行了论证，证实了与新石器时代安纳托利亚农业人群相关的人口迁徙推动了农业在欧洲的扩张。其中，马西森 (Mathieson)^②等利用迄今为止最大的古 DNA 数据库，对 230 个生活在公元前 6500 年至公元前 300 年欧亚大陆西部人群的样本 (样本包括新石器时代的安纳托利亚农民的首个全基因组

① Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., & Krause, J. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-day Europeans," *Nature*, 2014.
 ② Mathieson, I., Lazaridis, I., Rohland, N. et al. "Genome-wide Patterns of Selection in 230 Ancient Eurasians," *Nature*, 2015(528), pp. 499-503.

DNA 样本) 进行全基因组分析, 证明了他们是欧洲最早的农民来源。马西森 (Mathieson)^① 随后又提出, 要验证草原假说支持的印欧语传播路径, 首先要关注巴尔干半岛与草原祖先相关的基因, 其次是安纳托利亚。因此, 该课题组对 225 个来自巴尔干半岛、喀尔巴阡盆地、北庞蒂克草原和邻近地区的古人进行古基因组测序分析, 结果并没有发现青铜时代晚期的这些样本里存在大量与草原人群有关的血统, 目前关于青铜时代安纳托利亚人的基因数据非常少, 并且没有显示与操印欧语的人群有关, 间接否定了印欧语起源于东欧大草原并经由巴尔干半岛扩散至安纳托利亚的说法。达姆加德 (Damgaard)^② 等认为青铜时代早期颜那亚文化的游牧人群对安纳托利亚, 甚至中亚和南亚的遗传影响很有限, 安纳托利亚样本中缺乏草原血统表明, 印欧语系最早分支在该地区的传播与来自草原的主要人群迁徙并没有关系。语言随着人群的迁徙和移动而传播, 通过古基因可以检测遗传漂移、人群迁徙以及演化历史, 因此对数千年前的古人类 DNA 进行分析, 不仅能发现其源流及扩散历史, 还能窥见这群人所讲语言的情况, 虽然埋在土下骨骼不能讲话, 但其反映的基因信息却能与语言的谱系、考古文化相吻合。

虽然早期的语言学 and 考古学资料似乎使得库尔干假说更受欢迎, 但是支持安纳托利亚假说的学者另辟蹊径, 从农业扩散入手, 因为农业人口的迁徙, 必定会带动语言的流动。伦福儒提出的安纳托利亚农业扩散假说中所推测的起源时间要比库尔干假说早得多, 虽论证过程严谨, 但在考古和基因方面稍显薄弱, 比如根据伦福儒的农业扩散假说所推算的时间, 同源词所指的羊毛、轮车、马、轴等印欧文化特征在公元前 6000 年之前尚未出现。对于计算语言学家来说, 测定原始语言的年代仍存在很大难度和很多疑点, 比如缺乏更为详实的语言史前信息, 语言变化速度、词汇替换率等。语言学家基于现代语言重建原始语系, 必定需要考古学和遗传学提供的证据来验证每个时期操该语言的人群情况来支持和参考。借助古 DNA 的信息, 考古学也迎来了第二次科学革命, 为人类探索史前文明带来了新方法。因此, 整合多学科相关研究成果, 才能勾画出一个更准确、更科学的印欧语起源与扩散概貌, 最终实现语言学、考古学和遗传学的多学科联手, 共同攻关历史难题。

三、结语

对语言演化的探索在一定程度上也是追寻人类演化的问题。语言的起源地和语族间的亲缘关系不仅能从考古学和语言学材料中寻迹, 还可以从古 DNA 分析中得到线索。语言是文化的载体, 考古文物可以为语言的源流与扩散提供可靠的史料, 以反映每个历史时期的物质文化。虽然出土骨骼和器物并不能说话, 但通过对其重建和分析, 能够重现历史上操该语言人群的真实生活面貌, 捕捉到人群之间的交际方式及交际范围^③。而遗传学则提供了一种强大的新工具, 用古基因中的遗传特征来追溯人类迁徙。古 DNA 数据的优势在于打破时间和空间的限制, 能够更加科学地分析当时的人群之间的关系, 从而为人类及其语言的起源和扩散提供强有力的证据。

综合语言、考古和遗传学多学科的证据, 我们推断印欧语起源于高加索山南部地区到安纳托利亚一带, 因为安纳托利亚语是印欧语系的最早期分支; 尔后, 随着农业的扩张, 安纳托利亚农业人群开始向西扩散进入欧洲东部, 形成欧洲早期的农业人群; 欧洲东部早期的农业人群带着原始印欧语随即向东进入东欧大草原, 与那里的早期游牧人群融合而形成了青铜时代早中期以颜那亚文化为代表的草原游牧人群; 以颜那亚文化为代表的草原游牧人群向西涌入欧洲, 向东直达阿尔泰山区, 并深入塔里木盆地遗留下了吐火罗语, 向东南进入南亚, 不断地传播印欧语。

对印欧语系源流与扩散这一热点话题的研究具有重要的学术意义和现实意义。将遗传学、分子人类学的新进展和新方法应用在印欧语系的语言年代测算和谱系分类上, 对我国的汉藏语系、阿尔泰语系等语言

① Mathieson, I., Alpaslan-Roodenberg, S., Posth, C. et al. "The Genomic History of Southeastern Europe," *Nature*, 2018(555), pp. 197-203.

② Peter de Barros Damgaard et al., "The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia," *Science*, 2018(360).

③ 孙娜, 王传超:《从三维视域看语言的扩散》,《中国社会科学报》2020 年 12 月 15 日第 3 版。

的研究有重要参考价值。近年来古基因组学在印欧、东亚人群源流、人群形成历史、迁徙路线等科学问题上取得的进展也对吐火罗语及其人群起源和迁徙的研究具有启示意义，目前哪个新疆的考古文化与其相对应还缺乏充足证据，并且吐火罗人的古 DNA 研究还只是起步阶段，且研究以颜那亚文化、阿凡纳谢沃文化、西伯利亚等人群基因为主，阿凡纳谢沃人群与颜那亚人群确实有基因联系，但要判断阿凡纳谢沃人群是否迁徙到塔里木盆地，还需要进一步挖掘更多的基因证据来论证，也必定需要相关领域的成果给以借鉴。在语言起源的问题上，单一学科的贡献显然势单力薄，虽然各个学科在研究理论、方法、材料上存在差异，但所得出的研究结果却存在交叉和吻合之处，也论证了跨学科合作的必然性和有效性。我们根据对前人时贤研究的梳理，发现目前对印欧语起源的研究是考古学文化、语言学理论及材料和遗传学方法与证据的成功有机结合，进一步印证了各学科的假设，推进了对人类史前史的认识。

（本文为厦门大学南强青年拔尖人才支持计划（X2123302）；中央高校基本科研业务费（ZK1144）；国家自然科学基金青年项目“中国东南各族群的遗传混合”（31801040）；福建省社科基金一般项目“台湾南岛语族的起源与形成”（FJ2020B131）。厦门大学人类学研究所王传超教授为本文通讯作者）

（责任编辑：周奇）

The Origin and Dispersal of Indo-European Languages from the Perspectives of Linguistics, Archeology and Genetics

YANG Wenjiao, SUN Jin, GUO Jianxin, WANG Chuanchao

Abstract: The research on the origin and dispersal of Indo-European languages has gone through nearly a hundred years since it was first proposed. At present, the academic circles mainly focus on the Kurgan hypothesis and Anatolian hypothesis. The research on this topic is abundant yet complex at home and abroad. There is no consensus on its controversial origin and diffusion path. This paper reviews the main hypotheses on the origin and diffusion of Indo-European languages, and compares the research achievements and limitations of scholars from the perspectives of archaeology, linguistics and genetics. We infer the Indo-European languages originated in a region from the south of the Caucasus Mountains to Anatolia, then spread to the Eastern Steppe by early European farmers, and the nomads associated with Yamnaya culture in the Eastern Steppe in early and middle Bronze Age spread the languages to Europe, Northwest China (Tocharian) and South Asia.

Key words: Indo-European languages, archeology, linguistics, genetics, population migration