

纵向分化与横向接触

——中国民族语言 592 个语档的系统发生网络分析*

王 帅 冉启斌

[提要] 本文以中国境内的 592 个民族语言语档为对象, 基于 ASJP 模式的语言距离计算, 结合系统发生网络理论与方法, 考察了中国民族语言的系统发生网状关系。结果显示, 民族语言内部横向接触程度和同质性要低于汉语方言。从语系层面来看, 南岛语系语言间的横向接触最多; 阿尔泰语系语言间的横向接触最少。阿尔泰语系内部的同质性最高, 差异最小; 汉藏语系内部的同质性最低, 差异最大。从语族的层面看, 印度尼西亚语族语言间的横向接触最多; 满一通古斯语族语言间的横向接触最少。突厥语族内部的同质性最高; 藏缅语族内部的同质性最低。基于距离计算的系统发生网络分析为中国民族语言的复杂关系提供了数据。通过可视化和量化的手段, 我们能够相对容易地从总体上了解语言的纵向分化和横向接触。

[关键词] 中国民族语言 系统发生网络 Delta 值 Q-residual 值

一 引 言

同生物物种一样, 语言也是演化的实体。许多生物学理论和方法已被广泛运用于语言研究。自 Schleicher (1853) 首次使用谱系树模型描述印欧语系诸语言的谱系分类以来, 谱系树模型已成为历史语言学经典理论之一。然而, 在语言演化的过程中, 除了树状关系外, 语言间的关系有时还会呈现出网状结构, 谱系树模型容易忽略语言的横向接触与交流历史。随着分子人类学的飞速发展, 生物学家开始关注如何建立一个既能描述物种之间纵向分化, 又能兼容横向接触的模型, 种系演化网络模型应运而生 (邓晓华等 2015)。

在语言学研究领域, 越来越多的学者开始采用系统发生网络理论与方法来研究语言的纵向分化与横向接触。目前, 生物学中的种系网络模型已经成功应用于印欧语 (Bryant et al. 2005)、藏缅语 (邓晓华等 2015)、汉语方言 (Szeto et al. 2018)、混合语 (徐丹 2018)、壮侗语 (张梦翰、韦远诚 2019)、南岛语 (范志泉 2023a, 2023b) 等研究中。研究表明, 对于呈现树状分化的语言群, 网状方法能够准确还原其树状特征; 而在更为复杂的语言关系中, 则能揭示更为复杂的网状结构。因此, 这类方法既可用于验证谱系树模型的可靠性, 还能识别语言间的横向接触等动态演化因素, 为深入理解语言演化机制提供新的视角和方法支持。

* 本文为国家社科基金重大项目“中国境内语言核心词汇声学数据库及计算研究 (19ZDA300)”的成果之一。论文曾在“中国语言学会第二十二届学术年会”(长春 2024.9.21-22) 上宣读, 感谢与会专家的指教。匿名审稿专家和编辑部提出了宝贵的修改意见, 特此致谢!

本文拟以较大规模的中国民族语言材料为研究对象，利用系统发生网络理论与方法，建立中国民族语言的系统发生网络。通过计算语言群的横向交流度量值，考察不同民族语言在演变过程中的横向接触信息，从更广泛的样本角度较为全面系统地展示民族语言内部的关系。

二 研究材料与研究方法

（一）研究材料

本研究的主要材料来源于“ASJP 模式的中国民族语语档数据库”（数据库的情况参见冉启斌、王帅 2023）。该数据库包含了中国境内的 592 个民族语语档，涉及 5 个语系 9 个语族，其中还有部分语档属于混合语。592 个语档的分类及数量如表 1 所示。

表 1 592 个中国境内民族语语档分类统计

语系	语族	语档数量	语系	语族	语档数量
汉藏语系	藏缅语族	173	阿尔泰语系	蒙古语族	28
	苗瑶语族	148		满—通古斯语族	16
	壮侗语族	169		突厥语族	12
南亚语系	孟—高棉语族	34	印欧语系	印度—伊朗语族	3
南岛语系	印度尼西亚语族	6	其他	混合语	3

从语档数量的分布来看，不同语系和语族之间存在较大的差异。汉藏语系语档的数量最多，占据了全部民族语语档数量的 82.77%。相比之下，南岛语系、印欧语系、混合语的语档数量较为有限，都不足 10 个。这种数量上的差异实际上反映了中国境内各语系和语族的语言多样性以及语言分布的不均衡性。在后续分析中，我们将主要聚焦于这 592 个民族语语档，以期全面考察和研究中国境内民族语言的关系。

（二）研究方法

目前，各类系统发生网络方法已广泛应用于揭示物种间复杂的网状进化关系以及种内个体和群体间的关联，如统计简约法（Templeton et al. 1992）、中间网法（Bandelt et al. 2000）、邻接网法（Bryant & Moulton 2004）等。其中，邻接网法在处理复杂性状间冲突时表现尤为突出，且性状越复杂，分析效果越显著。该方法基于距离数据进行计算，且已经被集成于 SplitsTree4 软件包中。本文采用邻接网法构建中国民族语言的系统发生网络。

从可视化角度来看，使用邻接网法构建的系统发生网络中会出现很多放射状枝条，每支枝条代表对应语言与其他语言之间的横向接触情况，语言对应的终端枝条长短能够反映该语言与网络中其他语言之间的距离关系远近。此外，网络图中还会出现盒状网格信号，它体现了当语言特征或性状不满足树形结构时所呈现出的冲突信息。盒状信号越大，表示组成盒状信号的各语言之间可能有更为广泛的语言接触历史；反之，较小的盒状信号则表明各语言之间的语言接触程度较浅。系统发生网络在展示语言群的网状或树状分化时直观明了，但其中也包含了难以直接观察的信息。因此，采用一些定量的方法对网状结构进行考察至关重要。实践证明，Delta 值（Holland et al. 2002）和 Q-residual 值（Gray et al. 2010）可用于衡量网络中的盒状信号，量化冲突信息。Delta 值主要体现群体之间的横向接触程度，其取值范围在 0

到 1 之间, 数值趋近 0 说明该网络结构为树形结构, 数值趋近 1 说明该网络结构中存在较多的横向交流事件。Q-residual 值主要反映群体的异质性的信息, 它会受到构成网状结构的末端分支长度的影响 (Gray et al. 2010)。一般而言, Q-residual 值越小、终端枝条越长, 表示相应语言群内部的同质性越低。我们采用 ASJP 模式的语言距离计算方法, 对 592 个民族语语档进行了两两之间的语言距离计算^①。基于所得距离矩阵, 我们利用 SplitsTree4 软件中的邻接网法绘制了系统发生网络, 同时在软件中一并完成了 Delta 值和 Q-residual 值的计算。

三 592 个民族语语档的系统发生网络与横向交流度量

语言通常根据其系属关系或地理分布划分为不同的语言群。以往的研究在运用系统发生学理论和方法探讨中国境内民族语言关系时, 多聚焦某一特定语言群, 缺乏整体性分析。这主要是因为研究者通常采用词汇年代学方法, 通过优选同源词建立词汇表, 计算同源词比例, 以此为基础构建演化网络的距离矩阵 (邓晓华等 2015)。然而, 不同语系因历史演变和词汇替换速率等差异, 常难以实现跨语系比较。相比之下, ASJP 模式的语言距离计算是通过“莱文斯坦编辑距离”量化语言的基本词汇差异, 不依赖同源词的识别, 适用于跨语系的比较。该方法已成功应用于世界语言系统分类 (Müller et al. 2010)、世界语言分化的起源地和扩散模式 (Wichmann et al. 2010) 等研究。在考察中国境内民族语言演化关系时, 我们首先从宏观角度出发, 构建了 592 个民族语语档的系统发生网络, 如图 1 所示。

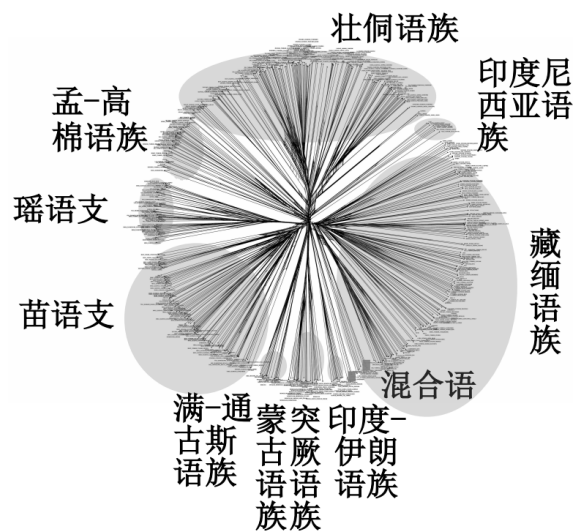


图 1 中国民族语言 592 个语档的系统发生网络

从图 1 来看, 592 个民族语语档的系统发生网络呈现由中心向四周辐射的结构, 同一语

^① ASJP 模式的距离计算提供了“归一化莱文斯坦距离”(Levenshtein Distance Normalized, LDN)和“归一化莱文斯坦距离商”(Levenshtein Distance Normalized Divided, LDND)等不同计算方法。LDN 算法可以消除因词长差异引起的偏差, 而 LDND 算法则可进一步消除不同语言或方言中因某些词的偶然相似带来的影响。本文采用的是 LDND 算法。

系、语族或语支的语档倾向于集聚成独立的放射簇。其中汉藏语系语言的语档分化为 4 个典型的放射簇，分别对应壮侗语族、藏缅语族、瑶语支和苗语支。这些放射簇在网络中被其他语言群的放射簇分隔开来。例如，壮侗语族语档放射簇和藏缅语族语档放射簇被印度尼西亚语族语档放射簇分隔开。阿尔泰语系语言的语档则分化为 3 个典型的放射簇，即满一通古斯语族、蒙古语族和突厥语族。此外，3 个混合语语档和藏缅语族语档集聚在同一个较大的放射簇中。系统发生网络中的每个枝条代表一个民族语语档，语档终端枝条的长短反映其与其他民族语语档之间的距离远近。相比其他语档，瑶语支语档的终端枝条较短，表明其与其他民族语言语档的距离较近，差异较小。阿尔泰语系和孟—高棉语族的语档终端枝条较长，表明它们与其他民族语言语档的距离较远，差异较大。此外，图 1 还显示，系统发生网络中存在明显的盒状信号，反映了这些民族语言之间存在广泛而复杂的接触现象。

从系统发生网络图上，我们只能大致看到民族语言群的聚合情况。为进一步了解中国境内民族语言的相互接触情况，我们计算了 592 个民族语语档的 Delta 值和 Q-residual 值，以考察其横向交流程度。同时，我们还计算了 300 个经过分类与地理平衡的汉语方言语档的 Delta 值和 Q-residual 值，在这里和民族语一并进行对比，如图 2 所示。

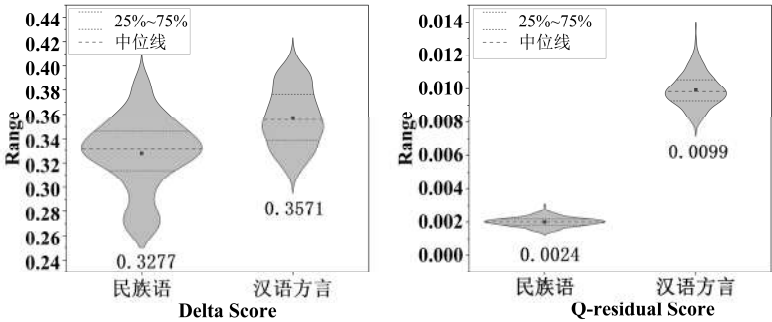


图 2 中国民族语言 592 个语档和汉语方言 300 个语档 Delta 值和 Q-residual 值小提琴图

图 2 显示了 592 个民族语语档和 300 个汉语方言语档 Delta 值和 Q-residual 值的分布范围。从数据浮动范围来看，民族语 Delta 值的浮动范围大于汉语方言，Q-residual 值的浮动范围小于汉语方言。从均值来看，民族语的 Delta 值（0.3277）要低于汉语方言（0.3571），表明汉语方言内部的横向交流和接触更为频繁；而民族语的 Q-residual 值（0.0024）低于汉语方言（0.0099），反映出民族语内部同质性较低，差异较大。使用曼—惠特尼（Mann-Whitney）U 检验对这两个指标进行分析，结果显示民族语与汉语方言在 Delta 值和 Q-residual 值方面存在显著差异，两项指标的显著性水平均为 $p=0.000$ 。

相较于汉语方言，中国境内的民族语言之间存在更大的差异，这一点不难理解。中国自古以来就是一个多民族、多语言的国家，各民族分布呈现出“大杂居，小聚居”的特点。不同民族的迁徙与政权更替，对语言的发展演变产生了深远的影响。无论是民族语言还是汉语方言，都经历了多层次、多类型的语言接触和深度交融。通过构建系统发生网络并结合 Delta 值和 Q-residual 值等量化指标，我们能够清晰地观察不同语言群体之间的演化关系。

四 不同语系的系统发生网络与横向交流度量

上文我们以 592 个民族语语档为研究对象，从整体视角考察了中国民族语言的系统发生网络与横向接触情况。下面我们从语系的角度出发，逐一考察不同语系内民族语言的系统发生网络与横向接触情况。由于 Delta 值和 Q-residual 值的计算至少需要任意 4 个分类点 (Taxon)，而印欧语系语档和混合语语档数均低于 4 个，因此在分语系和语族考察时，将不对印欧语系语档和混合语语档进行单独考察。图 3 展示了其他四个语系的系统发生网络。

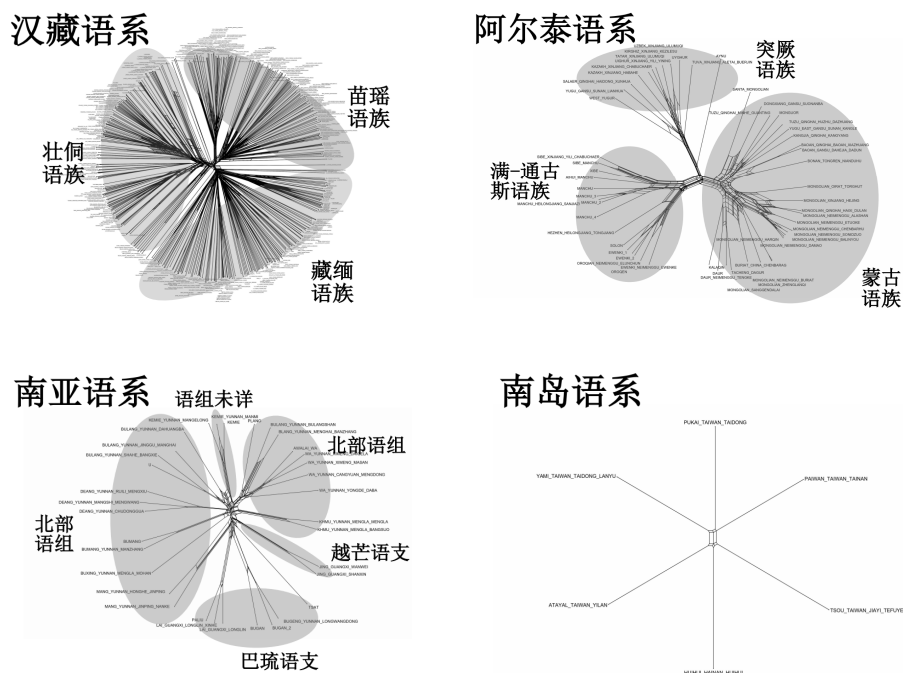


图 3 不同语系的系统发生网络

观察图 3，各语系内部普遍存在不同程度的语言放射簇和盒状信号。特别是汉藏语系、阿尔泰语系和南亚语系，由于语档数量较多且涉及不同层级的语言群，其放射群簇更为明显，分支清晰、层次分明。相比之下，南岛语系的语档较少，仅包含一个语族的语档，因此其放射簇不如其他语系明显。尽管南岛语系只包含 6 个语档，但各语档的终端枝条较长，且网络中盒状信号明显，表明该语系语言间可能存在较频繁的横向接触与交流。为了进一步分析不同语系内部的横向交流程度，我们计算了各语系的 Delta 值和 Q-residual 值，结果如表 2 所示。

表 2 不同语系内部横向交流度量值数据概要

语系	Delta 值	Q-residual 值	语系	Delta 值	Q-residual 值
汉藏语系	0.2980	0.0023	南亚语系	0.3130	0.0032
阿尔泰语系	0.1824	0.0040	南岛语系	0.5006	0.0031

据表 2, 四个语系中, 南岛语系的 Delta 值最大, 为 0.5006, 反映出其内部语言之间的横向接触和交流最为频繁; 阿尔泰语系的 Delta 值最小, 仅为 0.1824, 说明其内部语言间的横向接触和交流较少。Delta 值呈现出的排序为: 南岛语系 > 南亚语系 > 汉藏语系 > 阿尔泰语系。观察 Q-residual 值, 阿尔泰语系的 Q-residual 值最大, 为 0.0040, 表明其内部语言的同质性较高, 差异较小; 汉藏语系的 Q-residual 值最小, 为 0.0023, 显示其内部语言的同质性最低, 差异最大。Q-residual 值呈现出的排序为: 阿尔泰语系 > 南亚语系 > 南岛语系 > 汉藏语系。

先前研究表明, 汉藏语系的演化历史与印欧语系和南岛语系显著不同, 这与其独特的演化历程和地理环境密切相关。在东亚和东南亚地区, 各民族长期共存, 导致语言频繁接触, 语言分化与融合相互交织, 难以明确区分。此外, 在过去两千多年中, 该地区的各民族语言始终受到汉文化的影响, 频繁的语言接触极大程度改变了这一地区的语言关系面貌(邓晓华等 2015)。从 Delta 值和 Q-residual 值来看, 尽管汉藏语系的 Delta 值低于南岛语系和南亚语系, 但其 Q-residual 值在所有语系中最小, 反映了汉藏语系语言的同质性低, 内部差异大。通过计算 592 个民族语语档的两两相似度, 结果显示相似度最小的语言对出现在汉藏语系内部。不同语系内部的相似度计算也表明, 汉藏语系的内部差异最为显著(冉启斌、王帅 2023)。

阿尔泰语系语言的 Delta 值和 Q-residual 值显示, 其内部横向接触和交流程度较弱, 呈现出较高的同质性。通过对突厥语族、蒙古语族和满一通古斯语族的比较研究, 可以发现这些语言无论是在音系、词汇、形态还是句法等方面都具有显著的相似性(黄行 2018)。此外, 根据 Bryant et al. (2005) 的研究, 印欧语的 Delta 值为 0.22, 而我们计算得到的阿尔泰语系语言的 Delta 值为 0.1824, 略低于印欧语。印欧语通常又被视为具有同源关系语言的典范, 其分化模式呈树状结构。根据 Delta 值, 阿尔泰语系语言也趋向于类似的分化模式。

中国的南亚语主要分布在云南和广西两省区, 其中京语和侬语集中分布在广西境内, 其余南亚语则主要分布在云南。在考察的四个语系中, 通过对 Delta 值和 Q-residual 值的降序排列分析发现, 南亚语系在这两项指标上均位列第二, 这一量化结果表明南亚语系内部各语言之间具有较为密切的接触关系。进一步观察其系统发生网络可以发现, 网络中存在密集的盒状结构, 这一现象直观地反映了南亚语言之间存在频繁的横向接触与交流历史。

中国的南岛语系语言主要分布在台湾岛、兰屿岛和海南岛。李壬癸(2011:10)等学者指出, 台湾南岛语保留了最多的原始南岛语特征, 且语言间的分歧显著, 彼此接触和影响深刻。根据 Delta 值和 Q-residual 值的计算结果, 南岛语系语言展现出了最高的横向接触和交流程度, 且内部同质性低, 仅次于汉藏语系。范志泉(2023b)以斯瓦迪士 100 核心词表为基础材料, 通过优选同源词, 对 15 种台湾南岛语计算得到的 Delta 值为 0.394, 略低于本文结果, 这可能与样本的选取和数量有关。总体而言, 南岛语系较高的 Delta 值反映了该语言群语言间丰富的横向接触交流历史。

五 不同语族的系统发生网络与横向交流度量

在前一节中, 我们从语系层面考察了民族语言的纵向分化与横向接触机制, 本节将从语族层面探讨各语言群内部的演化关系。由于南亚语系和南岛语系的相关分析主要基于孟—高棉语族和印度尼西亚语族的语档材料, 这两大语族的分析内容已在前一节中有所涉及, 因此本节将重点考察汉藏语系和阿尔泰语系的语族。各语族的系统发生网络如图 4 所示。

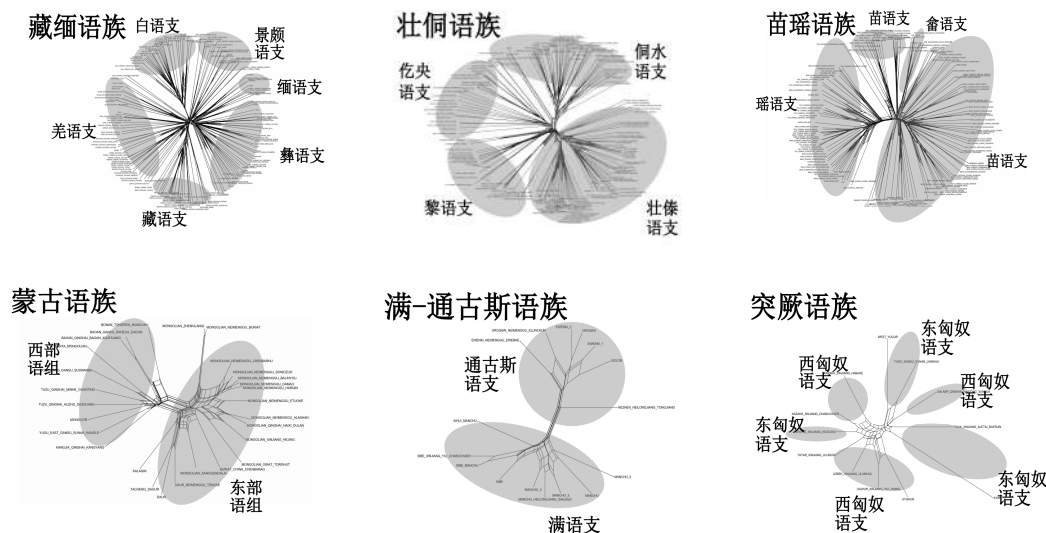


图 4 不同语族的系统发生网络

从图 4 可以看出，在各语族的系统发生网络中，同一语支或语组的语档倾向于集聚在一起，形成放射簇。在藏缅语族的演化网络图中，可以清晰观察到 6 个放射簇，分别是景颇语支、缅语支、彝语支、藏语支、羌语支，以及一个由白语支构成的放射簇。既往研究表明，在藏缅语族的次级分类中，缅语支与彝语支应视为两个独立的语支，与藏语支、景颇语支等并列。从系统发生网络的集聚模式看，缅语支语档和彝语支语档相互邻近，各自独立成簇，并与藏语支语档和景颇语支语档分离，这与传统分类研究结果一致。此外，网络图显示，羌语支语档集聚成一个独立的放射簇，与其他语支的语档分离，进一步支持了羌语支作为独立语支的分类地位。同时，白语支语档集聚成一个放射簇，与其他语支的语档分离，其单一性和独立性尤为突出。

关于壮侗语族，自 20 世纪以来，国内学界普遍采用壮傣、侗水、黎、仡央四大语支的分类体系。观察壮侗语族的系统发生网络，可以发现其中有四个典型的放射簇，分别对应侗水、壮傣、黎和仡央语支，结果与传统分类一致。

在苗瑶语族的系统发生网络中，瑶语支语档集聚成独立的放射簇，与其他语支的语档明显分离。尽管畲语支语档被苗语支语档包围，但仍集聚成一个相对独立的放射簇。研究表明，苗语和彝语是汉藏语系方言可懂度最低的语言（黄行 2018），这一点在图中得到了体现，苗语支语档内部形成多个不同的放射簇，反映了其复杂性。

蒙古语族的系统发生网络显示，蒙古语族的语档分为左右两个明显的放射簇，左侧为西部语组语言的语档，右侧为东部语组语言的语档。满-通古斯语族的系统发生网络显示，该语族的语档形成了上下两个放射簇，上面是通古斯语支语言的语档，下面是满语支语言的语档。相比之下，突厥语族的系统发生网络呈现出较为复杂的结构，其语言并未严格按照西匈奴语支和东匈奴语支形成两个泾渭分明的放射簇，而是呈现出一定程度的交织分布。

为了更全面地了解各语族内部的横向交流程度，我们分别计算了它们的 Delta 值和 Q-residual 值。不同语族内部横向交流度量值数据如表 3 所示。

表 3 不同语族内部横向交流度量值数据概要

语族	Delta 值	Q-residual 值	语族	Delta 值	Q-residual 值
藏缅语族	0.3484	0.0030	满一通古斯语族	0.2572	0.0119
壮侗语族	0.2937	0.0049	突厥语族	0.3842	0.0126
苗瑶语族	0.2916	0.0043	孟—高棉语族	0.3130	0.0033
蒙古语族	0.2869	0.0123	印度尼西亚语族	0.5006	0.0031

观察表 3，各语族中印度尼西亚语族的 Delta 值最大，为 0.5006；满一通古斯语族的 Delta 值最小，为 0.2572。汉藏语系中，藏缅语族的 Delta 值最大，其次是壮侗语族和苗瑶语族。阿尔泰语系中，突厥语族的 Delta 值最大，其次是蒙古语族和满一通古斯语族。观察各语族的 Q-residual 值，突厥语族的 Q-residual 值最大，为 0.0126；藏缅语族的 Q-residual 值最小，为 0.0030。这表明突厥语族语言的同质性最高，差异最小；而藏缅语族语言的同质性最低，差异最大。在汉藏语系中，壮侗语族语言的同质性最高，其次是苗瑶语族，藏缅语族最低。在阿尔泰语系中，突厥语族语言的同质性最高，其次是蒙古语族，最后是满一通古斯语族。

藏缅语族语言的同质性较低，这一现象可从地理分布、历史演化及内部语言差异等多方面解释。作为汉藏语系中语种最多、分布最广且内部差异最大的语族，藏缅语族的语言特征深受地理环境、人群迁徙和语言接触的影响（邓晓华等 2015），其内部各语支差异显著，甚至部分语言的方言之间也难以通话。戴庆厦等（1989）通过比较藏语支、缅语支和彝语支，发现缅语支与彝语支语言许多基本特征相似，但与藏语支差异较大。此外，尽管景颇语支与缅语支因长期接触而趋于一致，但从语言亲缘关系看，两者并不密切（杜冠明 2008）。

在所有语族中，突厥语族语言表现出较多的横向接触交流和同质性较强的特点。前人研究表明，突厥民族的复杂形成过程使其语言发展模式有别于那些相对稳定、拥有固定文化中心的语言。民族流动性促使突厥语族内部经历了持续的分化与融合（王远新 1994）。突厥语族的各语言中，不仅存在大量同源词汇，而且在语音系统中也体现出许多相似、相近的特点。此外，社会环境、人际交往和宗教影响，在语言中都会有所反映。在中国的 9 种突厥语族语言中，维吾尔语、哈萨克语、柯尔克孜语、乌兹别克语、塔塔尔语和撒拉语的共同点更多；西部裕固语和图瓦语、东北柯尔克孜语的共同点更多。就语音和语法特点而言，维吾尔语和乌兹别克语比较相近；哈萨克语和塔塔尔语、柯尔克孜语比较接近；原属乌古斯语的撒拉语，受维吾尔语和汉语的影响较大（陈宗振 2009）。图 4 中突厥语族系统发生网络进一步佐证了上述关系。如网络下端的西匈奴语支放射簇中，维吾尔语和乌兹别克语集聚在一起。网络左侧西匈奴语支的两个语档分别是哈萨克语哈巴河话和察布查尔话，左侧东匈奴语支的语档是柯尔克孜语，其下方的语档是塔塔尔语。这 4 个语档若按照 Simons & Fennig（2017）的分类，均属于突厥语族的西部语组。

六 结 语

本文基于系统发生网络方法，以较大规模的中国境内民族语言材料为研究对象，从多个层面对民族语言的纵向传递与横向接触进行了深入分析，并提出了一些初步见解。从语系层

面来看,南岛语系内部诸语言间表现出显著的横向接触交流特点;相比之下,阿尔泰语系语言的横向接触程度较浅,内部同质性高。从语族的层面来看,印度尼西亚语族语言之间的横向接触交流较多,而满一通古斯语族的语言横向接触交流相对较少。突厥语族内部语言的同质性最高,语言间差异小;而藏缅语族语言内部的同质性最低,语言间差异显著。

语言的演变,是内外因素相互作用的结果。语言作为人群的社会性产物,人群的接触势必会导致语言之间的接触。语言的纵向式演变与横向式演变不是截然对立的,在关注语言纵向分化的同源关系时,同时也应该考虑到语言间横向传递的接触关系。系统发生网络理论与方法在语言演化研究中的应用为我们提供了一种更全面、更综合的视角。通过考察语言演化网络,我们可以更深入地理解语言之间的相互关系,并揭示语言演化的复杂性和多样性。

参考文献

- 陈宗振. 2009. 《中国现代突厥语族语言研究概况》,《语言与翻译》第3期。
- 戴庆厦、刘菊黄、傅爱兰. 1989. 《关于我国藏缅语族系属分类问题》,《云南民族学院学报》第3期。
- 邓晓华、杨晓霞、高天俊. 2015. 《试论语言演化网络——以藏缅语为例》,《语言研究》第3期。
- 杜冠明. 2008. 《汉藏语言的谱系》,《民族语文》第2期。
- 范志泉. 2023a. 《台湾南岛语的定量分类及其横向交流探析》,《海峡人文学刊》第2期。
- 范志泉. 2023b. 《台湾南岛语的发生学模型构建及数理分类》,《民族语文》第3期。
- 黄行. 2018. 《中国民族语言识别: 分歧及成因》,《语言战略研究》第2期。
- 李壬癸. 2011. 《台湾南岛民族的族群与迁徙》,台湾本铺: 前卫出版社。
- 冉启斌、王帅. 2023. 《内部差异与外部关联——中国民族语言 592 个语档的相似度计算分析》,《民族语文》第3期。
- 王远新. 1994. 《突厥语的分类及历史分期问题》,《满语研究》第2期。
- 徐丹. 2018. 《中国境内的混合语及语言混合的机制》,《语言战略研究》第2期。
- 张梦翰、韦远诚. 2019. 《台语支系统发生模式及横向交流度量》,载徐丹、傅京起主编《语言接触与语言变异》第376-391页,北京: 商务印书馆。
- Bandelt, H. J., V. Macaulay & M. Richards. 2000. Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution* 16(1): 8-28.
- Bryant, D. & V. Moulton. 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2): 255-265.
- Bryant, D., F. Filimon & R. D. Gray. 2005. Untangling our past: Languages, trees, splits and networks. In R. Mace et al. (eds.), *The Evolution of Cultural Diversity: Phylogenetic Approaches*, pp. 67-84. London: University College London Press.
- Gray, R. D., D. Bryant & S. J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 365(1559): 3923-3933.
- Holland, B. R., K. T. Huber, A. Dress, et al. 2002. δ plots: A tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution* 19(12): 2051-2059.
- Müller, A., S. Wichmann, V. Velupillai, et al. 2010. ASJP World Language Tree of Lexical Similarity: Version 3. <https://asjp.clld.org/static/WorldLanguageTree-003.pdf>.
- Schleicher, A. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für*

Wissenschaft und Literatur 3: 786-787.

Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the World* (20th edition). Dallas: SIL International. <http://www.ethnologue.com> [2023-12-25].

Szeto, P. Y., U. Ansaldi & S. Matthews. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22(2): 233-275.

Templeton, A. R., K. A. Crandall & C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132(2): 619-633.

Wichmann, S., A. Müller & V. Velupillai. 2010. Homelands of the world's language families: A quantitative approach. *Diachronica* 27(2): 247-276.

Longitudinal Divergence and Horizontal Contact: A Phylogenetic Network Analysis of 592 Doculects in China's Ethnic Minority Languages

WANG Shuai and RAN Qibin

[Abstract] This study examines the phylogenetic network relationships of 592 doculects of ethnic minority language in China, based on the distance calculation method of ASJP (Automated Similarity Judgment Program) and applying phylogenetic network theory and methods. The results show that the degree of horizontal contact and homogeneity within ethnic minority languages is lower than that within Chinese dialects. From the perspective of language families, the Austronesian languages exhibit the highest level of horizontal contact, while the Altaic languages show the least. The Altaic languages demonstrate the highest homogeneity and the smallest differences, whereas the Sino-Tibetan languages display the lowest homogeneity and the greatest differences. From the perspective of language branches, the Indonesian languages exhibit the most horizontal contact, while the Manchu-Tungusic languages show the least. The Turkic languages exhibit the highest homogeneity, whereas the Tibeto-Burman languages exhibit the lowest. Based on distance calculation, phylogenetic network provides a series of data for China's ethnic minority languages. Through visualization and quantification, it enables a relatively straightforward and comprehensive understanding of longitudinal divergence and horizontal contact of these languages.

[Keywords] ethnic minority languages in China phylogenetic network delta value Q-residual value

(通信地址: 300071 天津 南开大学文学院)

【本文责编 吴雅萍】