

台湾南岛语的发生学模型构建及数理分类*

范志泉

[摘要] 我国台湾地区南岛语的内部分类一直是学术界具有争议性的研究课题。本文从跨学科协作视角，采用语言学词源统计分析法，结合生物学种系发生学网络模型生成理论和方法，重新对台湾南岛语的内部分类进行研究。新生成的发生学网络图显示，邹语群的内部关系主要通过横向的接触交流体现。传统的“排湾语群”或许可再分为三个分支，其中布农语为独立的一支；另外，由于历史上长期接触与融合，布农语与邵语更倾向于聚类在一起。系属地位有争议的邵语、赛夏语和巴则海语与泰雅语群聚类在一起。

[关键词] 台湾南岛语 发生学模型 数理分类

我国的南岛语主要分布在台湾岛、兰屿岛以及海南岛。

南岛语（Austronesian），旧称“马来—波利尼西亚语（Malayo-Polynesian）”。目前，学术界一般以“Austronesian”一词作为整个南岛语的统称，而“Malayo-Polynesian”一词则用来泛指中国台湾地区以外的所有南岛语。从地理分布区域来看，南岛语主要分布在亚洲的中国台湾、中南半岛和马来群岛以及大洋洲的美拉尼西亚（Melanesia）、密克罗尼西亚（Micronesia）、波利尼西亚（Polynesia）三大群岛等地区。分布在我国台湾地区的南岛语，亦称为“台湾南岛语（族）”；又因为位于南岛语空间分布的最北端，有时也称为“北部或北区南岛语（族）”。需要注意的是，所谓“台湾南岛语（族）”只是一个地理学的含义大于语言学的、笼统的说法，因为到目前为止找不出一种语言学的特征是所有台湾南岛语共有的，尤其是语言的创新特征（何大安、杨秀芳 2000:23-24）。为行文统一，以下将我国台湾地区的南岛语诸语统称为“台湾南岛语”，包括主要部分在台湾本岛上的泰雅语、赛德克语、赛夏语、邹语、卡那卡那富语、沙阿鲁阿语（也称“拉阿鲁哇语”）、布农语（也称“布嫩语”）、鲁凯语、排湾语、阿美语（也称“阿眉斯语”）、卑南语、噶玛兰语、巴则海语（也称“巴宰语”）、邵语以及兰屿岛的雅美语（也称“达悟语”）。此外，还包括凯达格兰语^①、道卡斯语、巴布拉语、巴布扎语、洪雅语、西拉雅语、猴猴语等已消亡的语言；其中，道卡斯语、巴布拉语、巴布扎语、洪雅语等又可统称为“西部平埔诸语”（李壬癸 2004a:1-48）。

台湾南岛语由于保存了较多早期南岛语的特征，其语言现象在整个南岛语系语言研究中具有非常重要的地位（李壬癸 2010:1）。有学者主张原始台湾南岛语在谱系树上基本和原始南岛语相等，所有台湾地区以外的南岛语在整个南岛语谱系上，都属于较低的层次（Starosta

* 本文系国家社科基金青年项目“台湾南岛语的语音对应规律数据库建设与研究（22CYY047）”的阶段性成果。审稿专家提出具体修改意见。谨致谢忱。

^① 凯达格兰语包括雷朗、南崁、埔顶、巴赛、哆罗美远、里脑等土语（李壬癸 2004b:859）。

1995); 也有学者认为台湾南岛语的确存在一些“非台湾”特征 (Starosta 1995; 何大安 1998; Li 2004)。然而, 有关台湾南岛语的内部分类研究却一直是学术界颇具争议的问题。

一 关于台湾南岛语内部分类的问题

(一) 诸家的分类

20世纪30年代, 浅井惠伦曾尝试将台湾南岛语分为5个分支, 即: ①北部语群, 包括泰雅语、赛德克语和赛夏语; ②布农语; ③邹语—排湾语群, 包括邹语、卡那卡那富语、沙阿鲁阿语、排湾语、卑南语、鲁凯语; ④阿美语; ⑤雅美语。其中, 布农语、阿美语及其方言各自为独立语群, 而雅美语属于菲律宾巴丹语群 (参见潘英 1998:83-85)。

1960年代初, Dyen (戴恩·伊塞多 1984) 将台湾南岛语分为3个分支, 即: ①泰雅语团, 包括泰雅语、赛德克语; ②邹语团, 包括邹语、卡那卡那富语、沙阿鲁阿语; ③第三语团^①, 包括布农语、鲁凯语、排湾语、卑南语、阿美语、巴则海语、噶玛兰语、邵语、费佛朗语。陈康 (1992) 和孙宏开等 (2007) 的分类基本直接沿用 Dyen 的三分法。前者收录了13种台湾南岛语, 后者新增另外2种台湾南岛语, 即巴则海语和噶玛兰语, 并将新增的巴则海语归入泰雅语群, 噶玛兰语归入排湾语群。

Ferrell (1969) 与 Dyen 的分类基本相似。不同的是, Ferrell 将 Dyen 的第三语团称为排湾语团, 并进一步下分为两支, 即: ①排湾第一分支 (Paiwanic I), 包括鲁凯语、巴则海语、赛夏语、雷朗语、巴布扎语、道卡斯语、巴布拉语、洪雅语、邵语、排湾语和卑南语; ②排湾第二分支 (Paiwanic II), 包括布农语、西拉雅语、阿美语、噶玛兰语和雅美语。

Blust (1977) 曾将台湾南岛语分成3个分支, 即泰雅语群、邹语群、排湾语群, 而台湾地区以外的南岛语均归至“马来—波利尼西亚语群”之下。后 Blust (1999) 又重新分类, 将其分为9个分支, 即泰雅语群、西北语群、西部平埔诸语、邹语群、鲁凯语、排湾语、卑南语、布农语、东台湾语群, 它们与马来—波利尼西亚语群共同组成南岛语系。其中, 鲁凯语、排湾语、卑南语、布农语及其方言各自为独立语群。

何大安 (1998) 将台湾南岛语分为“台湾类型”与“非台湾类型”两大类。

土田滋 (1998) 将台湾南岛语分为西北语群和南台湾语群两支。

李壬癸 (2006) 承继土田滋、Blust 等人的研究, 认为台湾南岛语主要有4个分支, 即: ①北部台湾语群, 包括泰雅语支和大西北语支, 大西北语支又有赛夏语、巴则海语、邵语和西部平埔诸语; ②中部台湾语群, 包括邹语、卡那卡那富语、沙阿鲁阿语; ③南部台湾语群, 包括布农语、鲁凯语、排湾语、卑南语; ④东部台湾语群, 包括阿美语、西拉雅语、噶玛兰语、巴赛语等。

Sagart (2004, 2014) 提出一个关于早期南岛语演化的谱系树模型, 该模型将南岛语一分为四, 即雷朗、赛夏、巴则海和 Pituish 语群^②。另外, 马来—波利尼西亚语与壮侗语 (Tai-Kadai) 则被归入一个新的“南部南岛语 (Southern Austronesian)”分支中, 该分支属于 Puluqish 的一部分。

^① 后改采用 Ferrell (1969) 的术语“排湾语团”。

^② 文中 Pituish 是由原始南岛语 (PAN, Proto-Austronesian) *pitū “7” 派生而来, 指蕴含数字“7”的一组语群。同理, PAN *puluq “10” > Puluqish; PAN *enem “6” > Enemish; PAN *lima “5” > Limaish。

Ross (2009) 认为迄今为止重建的原始南岛语实际上应位于南岛语谱系树中较低层级的一个位置，反映它的语言属于一个不包括卑南语、邹语和鲁凯语的分群，可称为“核心南岛语（Nuclear Austronesian）”，即“核心南岛语假设”。在该假设下，他将原始南岛语一分为四：卑南语、邹语以及鲁凯语各为南岛语的一个分支，而其他所有的南岛语则属于另一个分支。

吴安其（2009）认为台湾南岛语可分为2个分支，即：①泰雅—赛夏语群，包括泰雅语、赛德克语和赛夏语；②邹—卑南语群，包括邹—鲁凯语支和卑南—布农语支，前者又包括邹语、卡那卡那富语、沙阿鲁阿语和鲁凯语，后者又包括卑南语、邵语、阿美语、布农语、排湾语和巴则海语。它们与马来—他加禄语群、美拉—密克罗尼西亚语群一起构成南岛语系。

（二）诸家的主要分歧

具体来说，前人及时贤对台湾南岛语内部分类的主要分歧在于，台湾南岛语在整个南岛语中的位阶以及台湾南岛语下位分支数量的差异。关于台湾南岛语的位阶问题，主要涉及到台湾南岛语究竟整体是南岛语的一个分支，还是南岛语下的几个主要分支？李壬癸（1997）对此有较为详细的探讨，可参看。至于台湾南岛语下位分支数量差异的问题，除位于兰屿岛上的雅美语争议不大之外，台湾南岛语下位分支数量的差异实际具体表现在：

第一，关于邹语群的聚类问题。一直以来大多数学者普遍接受邹语、卡那卡那富语和沙阿鲁阿语为一个语群的分类，如李壬癸（Li 1972）和 Sagart (2004)。Ferrell (1969:70) 注意到卡那卡那富语和沙阿鲁阿语似乎并没有邹语那样的复杂结构，而且沙阿鲁阿语与邻近的西拉雅语和鲁凯语的相似性很大。土田滋（Tsuchida 1976:223）认为李壬癸（Li 1972）提出的邹语群共享的 *j 的脱落或消失现象，在卡那卡那富语和沙阿鲁阿语中，仅 *j 与 *i 相邻时才存在。Ross (2012) 认为尽管包括邹语、卡那卡那富语、沙阿鲁阿语在内的“邹语群假设”为许多人所接受，但并无证据支持此假设。相反，邹语群所反映的是邹语与卡那卡那富语、沙阿鲁阿语由于长期的接触交流，导致它们具有相似的语言特征。

第二，对排湾语群认定的差异。其中也会涉及有关邵语、赛夏语及巴则海语的系属地位问题。如前所述，三分法始于 Dyen 和 Ferrell，此后的分类研究大体围绕其展开。土田滋（1998）将排湾语群中的西部平埔诸语、赛夏语、巴则海语与泰雅语群合并，其余的与邹语群合并。而吴安其（2009）只将赛夏语与泰雅语群合并，其余的与邹语群合并。何大安（1998）把 Ferrell (1969) 的排湾语群第一分支诸语分立为一支，把排湾语群第二分支诸语与泰雅语群、邹语群合并为一支。Blust (1999) 将排湾语群一分为七，西部平埔诸语与邵语为一支，赛夏语与巴则海语为一支，但不与泰雅语群合并；阿美语、西拉雅语、噶玛兰语、凯达格兰语合并为一支；卑南语、鲁凯语、排湾语、布农语等单一语言为四个独立分支。李壬癸（2006）将排湾语群一分为三，西部平埔诸语、邵语、赛夏语、巴则海语等与泰雅语群合并为一支，与土田滋（1998）的不同点在于邵语的系属地位；剩余的排湾语群语言再分为两支。

（三）分类结果产生分歧的原因与思考

总的来说，前人及时贤在探讨台湾南岛语的内部分类研究方面取得了诸多成果，然而囿于不同的分类理论、方法以及材料，其分类研究依旧存在诸多分歧。

首先，在分类标准上，Ferrell (1969)、何大安 (1998)、Blust (1999)、李壬癸 (2006)、吴安其 (2009) 等采用历史语言学的标准。传统的语言分类主要是利用辅音、元音等的共享创新特征来划分的，是一种定性的分类方法。然而由于语音特征的选取完全依赖于学者个人的主观经验和学术积累，即使针对同一研究对象，不同学者选取不同的语音特征，其分类结

果往往也有所差异。除语音特征外, Starosta (1995) 和 Ross (2009, 2012) 均基于形态特征, 而 Sagart (2004) 则是基于词汇特征。

其次, 除定性分类研究外, 也有学者采用词汇统计的定量方法。如 Dyen (戴恩·伊塞多 1984) 所作的同源词百分比统计, 结果认为, 泰雅语群诸语之间以及邹语群诸语之间较高的同源词比例显示两个语群内部诸语言的关系是比较密切的, 所以各自组成了语群。但是排湾语群诸语之间的同源词比例却参差不齐, 那么是否还可归入一个语群? 何大安 (1999) 认为排湾语群不是一个发生学上同源的分支, 这个语群的内容需要重新界定。另外, 用作比较的核心词目数量越大, 就越难排除语言之间相互借用的成分。根据陈保亚 (1996) 划分词阶的方法, 斯瓦迪士的 200 核心词可一分为二: 前 100 词称为高阶词, 余下的 100 词为低阶词, 高阶词比低阶词更稳定, 更难以借用。Wang & Wang (2004) 也得出类似的结论。

最后, 自德国语言学家施莱赫尔首次运用谱系树形图描述印欧语系语言的谱系分类起, 谱系树模型便成为历史语言学的经典理论之一。谱系树形图简单明了, 不但能够清晰地表现整个语言纵向分化传递的历史与层次, 还能直观地显示出语言之间的亲缘关系程度。然而谱系树形图总归只是一种理想化的理论模型, 因为它容易忽略语言之间除了纵向分化外的横向接触与交流历史 (徐通锵 1991)。语言的横向传递接触势必会有更多的瓜葛牵连, 不可能完全整齐干净地表现为不同的树枝 (徐丹 2018)。

近年来, 生物学新提出了一种能够同时描述生物间的纵向分化和横向接触的种系发生学网络模型。目前, 国内外有不少语言学者已经尝试将这一模型用于印欧语 (Bryant et al. 2005)、大洋洲波利尼西亚语 (Bryant 2006)、汉语方言 (Hamed & Wang 2006)、藏缅语 (邓晓华等 2015)、壮侗语台语支 (张梦翰、韦远诚 2019) 等的研究中。可见, 这种发生学网络模型的特性同样也能适用于语言间的演化研究。

因此, 本文以台湾南岛语为研究对象, 利用生物学新近的种系发生学网络模型生成理论与构建方法, 生成台湾南岛语的发生学网络图, 以期较为系统地阐释其内部关系。并在此基础上, 通过网络结构的 Delta 值检测语言间的横向接触信息。在网络结构图中, Delta 值是用来衡量每一个语言点与其他语言点的交流程度 (Holland et al. 2002)。Delta 值的范围在 0 到 1 之间, 若数值是 0 则说明该网络结构为树形结构, 数值越接近于 1 则说明该网络结构无法回归到树形结构, 同时也能说明该网络结构中的横向交流事件多 (Gray et al. 2010)。

二 台湾南岛语发生学模型的构建

(一) 材料与方法

本文采用李壬癸、土田滋 (2001)、李壬癸 (2004b) 和何德华、董玛女 (2006) 收录的 15 种台湾南岛语为研究对象, 即泰雅语、赛德克语、邹语、卡那卡那富语、沙阿鲁阿语、鲁凯语、布农语、排湾语、卑南语、邵语、赛夏语、巴则海语、噶玛兰语、阿美语、雅美语。以其斯瓦迪士 100 核心词表为基础材料, 通过优选同源词, 编制同源词表, 计算每对语言的同源词百分比, 生成 15 种语言的同源词百分比相似矩阵。语言间的距离矩阵通过负对数计算获得, 进而将相似矩阵转换为距离矩阵 (见表 1), 作为发生学网络模型生成的基本数据材料 (具体步骤可参看邓晓华、王士元 2009)。

表1 15种台湾南岛语两两间的距离矩阵

语言	泰雅	赛德	邹	卡语	沙语	鲁凯	布农	排湾	卑南	邵	赛夏	巴则海	噶玛兰	阿美	雅美
泰雅	0.00	0.38	0.85	0.92	0.80	0.82	0.82	0.77	0.85	0.80	0.68	0.64	0.80	0.77	0.89
赛德	0.38	0.00	0.82	0.89	0.82	0.96	0.89	0.80	0.85	0.82	0.77	0.72	0.85	0.82	0.85
邹	0.85	0.82	0.00	0.43	0.38	0.70	0.62	0.68	0.66	0.59	0.62	0.62	0.70	0.55	0.70
卡语	0.92	0.89	0.43	0.00	0.31	0.70	0.66	0.66	0.70	0.57	0.70	0.70	0.72	0.60	0.66
沙语	0.80	0.82	0.38	0.31	0.00	0.64	0.59	0.64	0.62	0.68	0.64	0.68	0.57	0.57	0.62
鲁凯	0.82	0.96	0.70	0.70	0.64	0.00	0.60	0.46	0.55	0.66	0.70	0.74	0.59	0.57	0.74
布农	0.82	0.89	0.62	0.66	0.59	0.60	0.00	0.52	0.54	0.52	0.60	0.62	0.62	0.51	0.68
排湾	0.77	0.80	0.68	0.66	0.64	0.46	0.52	0.00	0.46	0.55	0.54	0.55	0.55	0.48	0.48
卑南	0.85	0.85	0.66	0.70	0.62	0.55	0.54	0.46	0.00	0.57	0.60	0.59	0.62	0.51	0.68
邵	0.80	0.82	0.59	0.57	0.68	0.66	0.52	0.55	0.57	0.00	0.51	0.47	0.57	0.52	0.66
赛夏	0.68	0.77	0.62	0.70	0.64	0.70	0.60	0.54	0.60	0.51	0.00	0.48	0.57	0.54	0.70
巴则海	0.64	0.72	0.62	0.70	0.68	0.74	0.62	0.55	0.59	0.47	0.48	0.00	0.64	0.62	0.70
噶玛兰	0.80	0.85	0.70	0.72	0.57	0.59	0.62	0.55	0.62	0.57	0.57	0.64	0.00	0.51	0.70
阿美	0.77	0.82	0.55	0.60	0.57	0.57	0.51	0.48	0.51	0.52	0.54	0.62	0.51	0.00	0.72
雅美	0.89	0.85	0.70	0.66	0.62	0.74	0.68	0.48	0.68	0.66	0.70	0.70	0.72	0.00	

(二) 台湾南岛语发生学模型的构建

基于以上材料与方法,本文采用SplitsTree 4软件包(Huson & Bryant 2006)中的邻接网络法(Neighbor-Net)(Bryant & Moulton 2004)生成15种台湾南岛语的发生学网络图(图1)。为了更好地描述台湾南岛语与其他南岛语的关系,本文在画图时加入马来—波利尼西亚语群的印尼语(陈枫、黄凤 1963)。

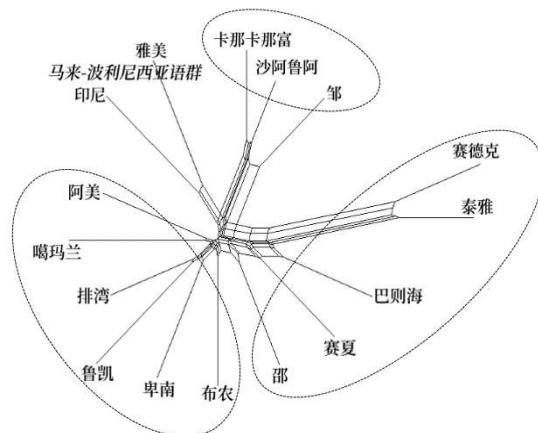


图1 15种台湾南岛语与印尼语的发生学网络图

另外，本文根据 Delta 值计算公式，计算出网络图中 15 种台湾南岛语的 Delta 值。根据前人的研究，印欧语（Bryant et al. 2005）、藏缅语（邓晓华等 2015）、壮侗语台语支（张梦翰、韦远诚 2019）以及大洋洲波利尼西亚语（Bryant 2006）的 Delta 值分别为 0.22、0.3078、0.25、0.41。图 1 的 Delta 值为 0.394，高于印欧语的 Delta 值，但又接近于大洋洲波利尼西亚语的数值。

为更好地考察台湾南岛语各区域 Delta 值的分布差异，本文还根据地理分布区域合并语言点，希望能够了解不同区域的语言接触交流程度。见表 2。

表 2 15 种台湾南岛语的 Delta 值

区域	分类点	Delta 值	区域	分类点	Delta 值
北部	泰雅	0.315	中部	邹	0.355
	赛德克	0.360		卡那卡那富	0.327
	赛夏	0.387		沙阿鲁阿	0.377
	巴则海	0.392		邵	0.446
南部	鲁凯	0.417	东部	阿美	0.449
	卑南	0.406		噶玛兰	0.438
	排湾	0.404	兰屿岛	雅美	0.431
	布农	0.411			

从表 2 数据可以看出，在台湾南岛语中，南部和东部的 Delta 值 (=0.421) 最高，去除东部后的 Delta 值 (=0.409) 有所下降；中部的 Delta 值 (=0.376) 次高，去除邵语后的 Delta 值 (=0.353) 略有下降；北部的 Delta 值 (=0.363) 最低。总的来说，Delta 值由南向北逐步递减。而单独各个语言点的 Delta 值来看，阿美 (=0.449)、邵 (=0.446) 等语言在整个发生学网络图中的 Delta 值较高。

三 台湾南岛语人群的遗传结构

语言与基因演化的相关性是语言学和分子人类学的研究领域之一，语言的谱系分类和族群的遗传谱系大致是平行演化关系（卡瓦利-斯福扎 2003）。分子人类学是一门利用基因组分析人群之间差异来研究人类演化历史的学科，其研究的材料便是由细胞核中的染色体和细胞质中的线粒体 DNA 组成的人类基因组（李辉、金力 2015）。根据遗传规则的不同，人类基因组可以分为常染色体 DNA、线粒体 DNA (Mitochondrion) 和 Y 染色体 DNA (Y-chromosome)。其中，常染色体 DNA 遵循父母双系遗传，一份来自父亲，一份来自母亲，在传代过程中会发生重组，且易受混血融合的影响；而线粒体 DNA (mt-DNA) 和 Y 染色体 DNA (Y-DNA) 在细胞内以单拷贝存在，属于单系遗传标记，且不会受到混血的影响，可以构成相对单纯的谱系树形图。随着不同人群 DNA 研究的不断深入开展，学者们发现线粒体 DNA 在东亚和东南亚不同语系人群之间的频率分布存在一定的差别，尤其在南北人群之间的差异较为明显，不过在同一语系语言的人群内部，学者们并没有发现其有非常明显的聚类现象（文波 2003；

Li et al. 2007), 也就是说母系线粒体 DNA 单倍群的分布缺乏语系特异性, 几乎没有特定语系独有的线粒体 DNA 单倍群 (Yao et al. 2002)。

与其他类型的遗传标记相比, 父系 Y 染色体 DNA 非重组区 (NRY) 可以有效地排除交换重组的干扰, 而突变率低的单核苷酸多态性标记 (SNP) 更能稳定地遗传, 并可以忠实地记录人群的迁移和演化历史, 是目前公认的研究早期人类起源、迁徙和演化的理想工具 (柯越海等 2001)。目前学者们完成了对东亚及东南亚地区人群 Y 染色体谱系的大量研究工作。随着有关语言与基因相关性研究的逐步推进, 人们发现, 比起母系线粒体 DNA, 该地区人群的父系 Y 染色体 DNA 不仅与其语言有着较强的相关性 (文波 2003), 而且还能反映与支持语言学者所提出的语言亲属分类关系 (王传超 2015)。

Trejaut et al. (2014) 的 Y 染色体研究显示, 单倍群 O1a-M119 是台湾南岛语的主要父系类型。Y 染色体单倍群由 O1a1-P203 和 O1a2-M110 两个分支组成。单倍群 O1a1-P203 在北部的泰雅、赛夏以及中部的邵语、邹语人群中广泛分布, 约为 88%-90%; 中频分布于南部的阿美、鲁凯、排湾、卑南以及台湾本岛外的雅美等人群中, 约有 50% 左右; 在布农人群中没有分布。单倍群 O2a1-M95 及其下游单倍群 O2a1a-M88 在台湾南岛语中, 只分布在雅美 (=13.3%) 和布农 (=37.5%) 等少数人群中。单倍群 O3-M122 的下游分支除 O3a2c-P164 在台湾东海岸人群 (阿美=35.9%, 卑南=13%, 排湾=4%) 中有分布外, 其他分支单倍群未见或少量零星存在 (<5%)。此外, 单倍群 O1a2-M110 是布农人群的主要父系类型, 约占 61%, 但在阿美、鲁凯、排湾、卑南等人群中约占 18%-28%, 在泰雅、赛夏、邵、邹等人群中低于 6%, 雅美人群甚至未见该单倍群分布。这一点在利用高清晰度 Y-SNP 基因型数据绘制群体遗传距离的多维标度图 (图 2) 中尤为明显, 显示了布农语人群在台湾南岛语人群中的独特性。与 Y 染色体 DNA 单倍群相比, 台湾南岛语人群的线粒体 DNA 单倍群在分布上呈现出明显的不均衡性 (Trejaut et al. 2005)。

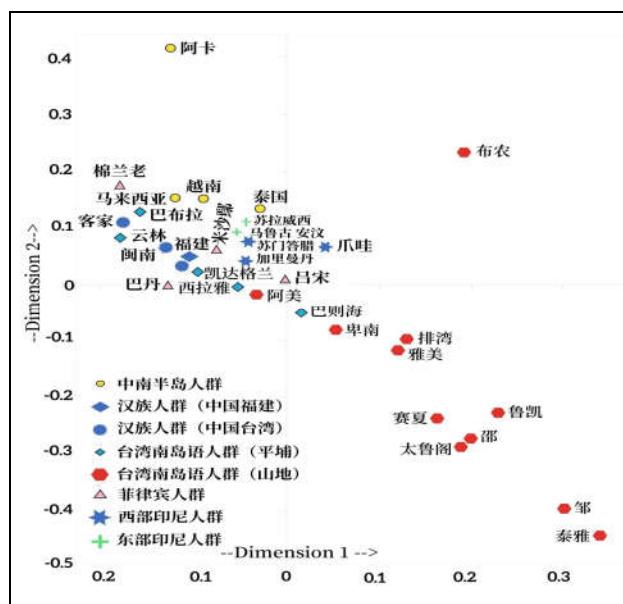


图 2 基于 Y-SNP 数据构建的群体遗传距离多维标度图 (改自 Trejaut et al. 2014)

四 结论与思考

我国台湾南岛语由于保存了较多早期南岛语的特征，其语言现象在整个南岛语系语言研究中具有非常重要的地位。然而有关台湾南岛语的内部分类研究一直以来在学术界都具有较大分歧。厘清台湾南岛语的内部分类，有利于客观认识其在整个南岛语谱系中的地位问题。为此，本文采用词源统计分析法，结合生物学的种系发生学网络模型生成理论和方法，从跨学科协作视角对台湾南岛语的内部分类重新进行了研究。本文构建的 15 种台湾南岛语发生学网络图显示，除位于兰屿岛的雅美语与印尼语聚类外，台湾南岛语大体可以呈现出三个不同的聚类，即：

- ①泰雅语、赛德克语、赛夏语、巴则海语、邵语；
- ②邹语、卡那卡那富语、沙阿鲁阿语；
- ③布农语、卑南语、鲁凯语、排湾语、阿美语、噶玛兰语。

总的来说，本文所显示出的三个聚类与 Dyen (戴恩·伊塞多 1984)、Ferrell (1969) 的分类结果基本一致。这不仅能够验证 Dyen 和 Ferrell 定性分类的科学性与合理性，而且还能反映出基于词源统计的定量分析在语言分类研究中的适用性和可操作性。相对于树形结构图而言，网状结构模型能够同时描述分化和接触两种演化模型，可为阐释有争议的语言系属问题提供一定的研究视角和理论依据。这在只能描述语言纵向传递式亲缘关系的传统树形图中是无法很好体现的，其争论的实质归根到底就是纵向分化和横向接触的断定问题。

从本文的研究还可以看到，台湾南岛语发生学网络图所呈现出的网络结构内部实际上存在着许多一定范围的盒状结构。盒状结构越大，代表语言特征蕴含的冲突信号越大，语言之间的横向接触交流就越多。如 Dyen 和 Ferrell 的邹语群，邹语与沙阿鲁阿语、卡那卡那富语虽聚类在一起，但本文的盒状结构显示它们也有一定横向接触，可证实 Ferrell (1969)、Ross (2012) 等人的观点。根据本文的研究，传统的“排湾语群”或许可以再下分三个分支，其中布农语可为独立的一支。布农语人群的遗传学证据显示确实如此。另外，一直以来，邵语、赛夏语、巴则海语等语言的系属问题也是学界争议的焦点。本文显示邵语、赛夏语和巴则海语等倾向于与泰雅语群聚类在一起，因此将其纳入李壬癸 (2006) 的“北部台湾语群”基本问题不大。此外，在图 1 中，邵语倾向于与布农语聚类在一起，究其原因是邵语不仅从布农语那里借用了自身所没有的音位，而且还吸收了大量布农语词汇，其中有些甚至已深入到基本核心词中 (Li 2013)。

本文所建构的台湾南岛语发生学网络图，其网状结构的 Delta 值由南向北递减，显示台湾南部的南岛语人群彼此之间有较多的横向交流历史。Delta 值为 0.394，高于印欧语的 Delta 值，但又接近于大洋洲波利尼西亚语的数值，是否也意味着台湾南岛语内部的发生学关系也可以用网络模型来描述。台湾南岛语的内部分类以及邵语、赛夏语等有争议语言的系属问题是南岛语学界讨论的焦点，其主要分歧在于如何描述纵向亲缘关系和横向接触关系。相对于传统的树形结构，网状结构能获取更多的语言群体的演化历史及关系信息。因此，生物学的演化网络生成理论与方法为语言演化提供了一个新的研究视角，当然还需要更多的实证研究，以验证该方法是否具有普适性。

参考文献

- [1] 陈保亚. 1996. 《论语言接触与语言联盟——汉越(侗台)语源关系的解释》, 北京: 语文出版社.
- [2] 陈 枫、黄 凤. 1963. 《印度尼西亚语汉语词典》, 北京: 商务印书馆.
- [3] 陈 康. 1992. 《台湾高山族语言》, 北京: 中央民族学院出版社.
- [4] 戴恩·伊塞多. 1984. 《台湾高山族语言》, 陈康译, 载中国社会科学院民族研究所语言室编《民族语文研究情报资料集(第3集)》(内部) 第77-106页.
- [5] 邓晓华、王士元. 2009. 《中国的语言及方言的分类》, 北京: 中华书局.
- [6] 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.
- [7] 何大安. 1998. 《台湾南岛语的语言关系》, 《汉学研究》第2期.
- [8] 何大安. 1999. 《论原始南岛语同源词》, 载石锋、潘悟云编《中国语言学的新拓展——庆祝王士元教授六十五岁华诞》第75-84页, 香港: 香港城市大学出版社.
- [9] 何大安、杨秀芳. 2000. 《丛书导论: 南岛语与台湾南岛语》, 载何大安主编《台湾南岛语言》丛书, 台北: 远流出版事业股份有限公司.
- [10] 何德华、董玛女. 2006. 《达悟语: 语料、参考语法及词汇》, 台北: 中研院语言学研究所.
- [11] 卡瓦利-斯福扎. 2003. 《追踪亚当夏娃——从演化历史看基因、民族和语言的关系》, 吴一丰、郑谷苑、杨晓佩译, 台北: 远流出版事业股份有限公司.
- [12] 柯越海、宿 兵、李宏宇等. 2001. 《Y染色体遗传学证据支持现代中国人起源于非洲》, 《科学通报》第5期.
- [13] 李 辉、金 力. 2015. 《Y染色体与东亚族群演化》, 上海: 上海科学技术出版社.
- [14] 李壬癸. 1997. 《台湾南岛民族的族群与迁徙》, 台北: 常民文化事业股份有限公司.
- [15] 李壬癸. 2004a. 《台湾南岛语言》, 载李壬癸著《台湾南岛语言论文选集(第一册)》第1-48页, 台北: 中研院语言学研究所.
- [16] 李壬癸. 2004b. 《台湾南岛语言论文选集(第二册)》, 台北: 中研院语言学研究所.
- [17] 李壬癸. 2006. 《从本土到国际: 台湾南岛语言研究的契机》, 载刘翠溶主编《四分溪论学集: 庆祝李远哲先生七十寿辰》第727-748页, 台北: 允晨文化实业股份有限公司.
- [18] 李壬癸. 2010. 《珍惜台湾南岛语言》, 台北: 前卫出版社.
- [19] 李壬癸、土田滋. 2001. 《巴宰语词典》, 台北: 中研院语言学研究所筹备处.
- [20] 潘 英. 1998. 《台湾原住民族的历史源流》, 台北: 台原出版社.
- [21] 孙宏开、胡增益、黄 行主编. 2007. 《中国的语言》, 北京: 商务印书馆.
- [22] 王传超. 2015. 《东亚人群的遗传历史》, 复旦大学博士学位论文.
- [23] 文 波. 2003. 《Y染色体、mtDNA多态性与东亚人群的遗传结构》, 复旦大学博士学位论文.
- [24] 吴安其. 2009. 《南岛语分类研究》, 北京: 商务印书馆.
- [25] 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.
- [26] 徐通锵. 1991. 《历史语言学》, 北京: 商务印书馆.
- [27] 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.
- [28] 土田滋. 1998. 「言語研究への誘い」. 日本順益台湾原住民研究会編『台湾原住民研究への招待』第137-139ページ. 東京: 風響社.
- [29] Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report.

- University of Hawaii Working Papers in Linguistics*, 9(2): 1-15.
- [30] Blust, Robert. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected Papers From the Eighth International Conference on Austronesian Linguistics*, pp. 31-94. Taipei: Institute of Linguistics, Academia Sinica.
- [31] Bryant, David. 2006. Radiation and network breaking in Polynesian linguistics. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages*, pp. 111-118. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.
- [32] Bryant, David & Vincent Moulton. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2): 255-265.
- [33] Bryant, David, Flavia Filimon & Russell D. Gray. 2005. Untangling our past: languages, trees, splits and networks. In Ruth Mace, Clare J. Holden & Stephen Shennan (eds.), *The Evolution of Cultural Diversity: Phylogenetic Approaches*, pp. 67-84. London: University College London Press.
- [34] Ferrell, Raleigh. 1969. *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. Taipei: Institute of Ethnology, Academia Sinica.
- [35] Gray, Russell D., David Bryant & Simon 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.
- [36] Hamed, Ben & Feng Wang. 2006. Stuck in the forest: trees, networks and Chinese dialects. *Diachronica*, 23(1): 29-60.
- [37] Holland, Barbara R., Katharina T. Huber, Andreas Dress, et al. 2002. δ Plots: A tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution*, 19(12): 2051-2059.
- [38] Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2): 254-267.
- [39] Li, Hui, Xiaoyun Cai, Elizabeth R. Winograd-Cort, et al. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *American Journal of Physical Anthropology*, 134(4): 481-488.
- [40] Li, Paul Jen-kuei. 1972. On comparative Tsou. *Bulletin of the Institute of History and Philology*, 44(2): 311-338.
- [41] Li, Paul Jen-kuei. 2004. Formosan vs. non-Formosan features in some Austronesian languages in Taiwan. *Selected Papers on Formosan Languages* (volume 2), pp. 953-976. Taipei: Institute of Linguistics, Academia Sinica.
- [42] Li, Paul Jen-kuei. 2013. Thao loans from Bunun. *Bulletin of Chinese Linguistics*, 7(2): 225-241.
- [43] Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. In Alexander K. Adelaar & Andrew Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, pp. 295-326. Canberra: Pacific Linguistics.
- [44] Ross, Malcolm. 2012. In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics*, 13(6): 1253-1330.
- [45] Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics*, 43(2): 411-444.
- [46] Sagart, Laurent. 2014. In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic.

- Language and Linguistics*, 15(6): 859-882.
- [47] Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.), *Austronesian Studies Relating to Taiwan*, pp. 683-726. Taipei: Institute of History and Philology, Academia Sinica.
- [48] Trejaut, Jean A., Toomas Kivisild, Jun Hun Loo, et al. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *The Public Library of Science Biology*, 3(10): 1838-1838.
- [49] Trejaut, Jean A., Estella S. Poloni, Ju-Chen Yen, et al. 2014. Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia. *BioMed Central Genetics*, 15(77): 1-23.
- [50] Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology*. Tokyo: Study of Languages & Cultures of Asia & Africa, Monograph Series No.5.
- [51] Wang, Feng & William S-Y. Wang. 2004. Basic words and language evolution. *Language and Linguistics*, 5(3): 643-662.
- [52] Yao, Yong-Gang, Long Nie, Henry Harpending, et al. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *American Journal of Physical Anthropology*, 118(1): 63-76.

The Construction of Phylogenetic Model and Mathematical Classification of the Austronesian Languages in Taiwan of China

FAN Zhiqian

[Abstract] The internal classification of the Austronesian languages in Taiwan, China has always been a controversial research topic. In this paper, from the perspective of interdisciplinary collaboration, this research topic is revisited by using both the linguistic analysis of etymology statistics and the theory and method of phylogenetic network model generation in biology. The newly generated phylogenetic network diagram shows that the internal relationship of the Tsouic group is primarily manifested as horizontal contact and transmission. The traditional “Paiwanic group” may be further divided into three branches, among which the Bunun language is an independent branch; in addition, Bunun and Thao tend to cluster together because of their longstanding contact and fusion in history. The languages of Thao, Saisiyat, and Pazih, whose genetic statuses are disputed, cluster together with Atayalic group.

[Keywords] Austronesian languages in Taiwan phylogenetic model mathematical classification

(通信地址: 350007 福州 福建师范大学闽台区域研究中心
350118 福州 福建工程学院南岛语族研究院)

【本文责编 吴雅萍】