



语(巴哼方言、唔奈方言),优诺语,炯奈语,畚语,其中巴哼语和优诺语为一语丛,炯奈语和畚语为一语丛。瑶语支仅勉语(勉方言、金门方言、标敏方言、藻敏方言)。

1988年出版的《中国语言地图集》(以下简称“地图”)将苗瑶语族分为苗语支:苗语(黔东南方言、湘西方言、川黔滇方言),布努语(布瑙方言、巴哼方言、唔奈方言、炯奈方言、优诺方言5个方言),畚语;瑶语支:勉语(勉金方言、标交方言、藻敏方言)。本文具体分析苗瑶语族的12个语言(方言)。<sup>②</sup>

传统苗瑶语分类的标准,语支分类的依据主要是音韵和同源词。张琨认为,瑶语是指完全保留或大部保留 p、t、k、m、n、ŋ 这套辅音韵尾的;苗语没有这些辅音韵尾,或至多有 n 和 ŋ。(但他认为这种分类法一般同这两个汉语名称所指的民族学上的分类一致的观点却是错误的)此外,苗语支的语言声母比韵母多,苗、布努、畚语元音不分长短;而瑶语一般韵母较声母复杂,元音分长短。瑶语浊的塞音和塞擦音声母可以出现在各个声调。按照张琨的观点,苗语支内部的语言和方言之间的分类主要依据是:(1)带前置鼻音的塞音和塞擦音声母 mp、nt、ŋk、ŋq、nts、ntr、nt 出现的情况。(2)有没有复辅音声母。(3)小舌音 q 是否与舌根音 k 对立。(4)各种鼻音的对立和各种边音的对立。

### 1.2 亲疏关系的计量分析

有关苗瑶语的分类,除了传统的做法外,还有最近黄行做的苗瑶语方言亲疏关系的计量分析研究,黄行的研究出色,并取得成绩。黄先统计出苗瑶语方言声韵调平均的相关系数,后作方言亲疏关系的聚类统计。我们跟黄虽然都依据语言学的计量统计方法来作苗瑶语的分类,但不同之处在于:

1)使用的材料不同。黄依据王辅世、毛宗武《苗瑶语古音构拟》的800词作声韵调音类的相关统计;我们则主要依据 Swadesh 的100词表作同源词的数理分析。而 Swadesh 的100词至少是目前国际语言学界公认的作历史语言学比较的最佳优选词目,同时具有较强的可操作性。Swadesh 的100基本词是经过许多语言检验过的一组能反映语言纵向分化的词表,这组基本词不容易受到借用的影响,词义有较高的稳定性。所以做统计时可以保证因借用而造成统计的干扰很小。当我们扩大统计的范围,比如王、毛的800词,要剔除其中借词的难度就大多了,特别是来自语族外部的借用。我们发现苗瑶语中有许多汉语的借词,即使在100词里我们也发现了“沙”、“眼”等词来自古汉语。当统计的词汇表扩大到大量的文化词时,这些借用对统计的影响就会造成很大的影响。更重要的是我们这里进行的是亲缘关系的分析,需要考虑的是共享词汇的保留率,所以基本词就可以反映这个关系,我们在 Swadesh 的100词的基础上针对苗瑶语的情况做了尽可能小的修改,把统计的词汇表增加至110词,目的就是既保证使用一个普适的统计根据,又不失考虑苗瑶语的特殊情况。从词源统计法的角度看,选取的词目越多,则相互借用的可能概率就会越高。

2)得到树形图的方法的不同。黄用聚类分析中的平均连结的方法,我们用构造种系发生树形图的距离法中的 Fitch-Margoliash 和 Neighbor joining 法。黄行的工作得到的是语言之间亲疏关系的表示。而我们采用的是广泛应用于生物学上研究种系发生学关系的方法,目的是得到一个亲缘关系的表示。不仅得到语支的分化顺序,而且从树图上可以用树枝的长短表示语言之间的距离,可以看出语言变化的不同速度。而且我们考虑了汉语以一个外部语言作为基准来看苗瑶内部的分化。这是前面研究没有做的。<sup>③</sup>我们所使用的距离法的优点是,树上的所有语言都能相互比较,而不仅仅是一次只能考察一对语言。而且语言离根部路径的长度通

常是有差别的,表明它们彼此都有一个共同的起点,即树根。而较早的距离法得出的树的末端离根部是等距离的,如平均值联系和不加权每对集合的数学均值(UPGMA)。因为这些较早的方法假定变化率是恒定的,我们在此不考虑这些方法。

3) 研究的目的不同。黄着重于方言亲疏关系,即方言相似性比较,这是一种平面性的比较。我们则强调方言的历时的发生学比较,运用生物学谱系分类方法,作苗瑶语言亲缘关系分类。“亲疏关系分类是分析分类对象的相似程度的,任何一对方言间的关系都可以用一个相关系数来表示。”(王、沈 1992)。语言的亲疏关系不同于语言的亲缘关系,方言的亲疏关系和亲缘关系的差别从原则上讲是共时关系和历时关系的差别。方言间有差异,从共时角度来看,这些差异表现了方言间的相似程度。不过,方言间的相似程度有不少是接触造成的,而跟语言的发生学没有关系。

## 二 词源统计分析方法的基本原则

本文主要运用词源统计分析的理论和方法来作苗瑶语族语言的亲缘关系分类。词源统计分析法的基本观念是两种具有亲缘关系的语言分离的时间深度,可以通过它们继承的词共享程度来判断。词源统计分析法不同于词汇统计学,词源统计分析法强调的是历史方面,词汇统计学主要关心的是使用词汇中相似的数量表示描述的目的,如估计可相互理解的程度(王、沈 1992)。词源统计分析法最初由 Swadesh 提出,他受到化学上发明的碳-14 年代测定法的启示,认为,一定存在一组“基本的词”,来描述一定存在于各种语言中的普遍现象。这种基本词汇包括表示身体部分的词(头、手),表示自然物体的词(月亮、山脉),表示共同活动的词(来、睡觉),以及表示一些其他范畴的词。实际上,要确定这样一种基本词汇是相当困难的。Swadesh 先提出了 200 词,后又减少到 100 词。词源统计分析法自 50 年代初提出后,用来测定一对语言分离的年代,由于其本身固有的一些严重缺陷,使得该方法未得到广泛运用。但近年来,由于生物学种系发生树理论的不完善和计算机的巨大进步,词源统计分析法又开始受到历史语言学家的极大重视。对此,国内大多数语言学者并未引起应有的关注,而过分强调汉藏语言的特殊性,往往自立一套词表,忽略 Swadesh 100 词的国际性、可比性和计量原则。

### 2.1 优选核心同源词的原则和问题

如何优选核心同源词,即设计一种通用基本词汇表是词源统计分析的最重要步骤之一。这个问题一直存在较大争论,例如蒂特(Teeter 1963)就认为“根本就不存在理想的词汇表。”一种语言中不同的词汇范畴具有不同的变化程度,一种语言中具有许多这类范畴。词汇中这种不同的变化程度可通过词汇传播理论来预测。埃勒加德(Ellegard 1959)认为,更好的研究策略是根据经验计算不同的语言,以及不同的词范畴的不同的变化程度提出方法,而不是假定一种相同的值。

我们认为,虽然同是同源词,但不同的同源词,在统计中的地位是不同的,即它的“权重”不同。同源词同样有历史文化层次的差别,有的同源词较容易被借用,有的同源词则被借用的概率较低。我们在作苗瑶语族语言比较时,选择 800 个同源词来作比较,跟选择 100 同源词作比较的结果是不同的。所以,选择多少数目的同源词才较适合作亲缘关系的分类比较,这个问题很重要。同时,还要考虑计算统计上的方便。若所选词目太少,其误差必大,所以,我们不选择雅洪托夫的 35 词表,经过比较和分析,我们认为 Swadesh 的 100 词表可操作性强,最为合适;他的后 100 词,虽较容易被借用,但仍有比较价值,可作选词参考。

## 2.2 语言关系表述的计量原则

传统的语言分类主要是通过建立语言特征和规则,例如声韵调、词汇、音变类型等不同特点来区划的。语言的差异局限于语言现象的定性的描写,例如前面提到的张琨、王辅世等对苗瑶语族的分类标准。但“这种知识并不能提供一种测量方言间亲疏程度的准则。”(郑,1988)。

传统的语言分类虽然也有一些简单的统计,例如作同源词百分比的统计。但这些简单化的统计,充其量只能够了解每一对语言间的关系,而无法从整体上把握语言集团的内部的错综复杂的关系。而数量的研究则提供一种计算语言亲疏以及亲缘关系和语言分类的科学手段。特别是生物学家发明的一些研究生物种系发生分类的程序,对语言学家很有用。因为生物学的分类与语言学分类很类似,而科学研究的一个重要特点,就是可用公式来反复验证和测量研究对象。当然,语言学与生物学除了有相似的生物遗传基因系统外,语言学还要考虑文化基因即横向的文化传播问题。历史语言学家最困难的在于,如何把历时的纵向传承不同于当代的横向传播的特征分离出来。

## 三 词源统计分析方法在苗瑶语族语言的亲缘关系的比较研究中的应用

词源统计分析法的步骤是:(1)编制同源词统计表;(2)计算每对语言的同源百分比;(3)整理相似矩阵和距离矩阵;(4)画出与距离矩阵最匹配的树形图。词源统计分析法研究的最终目的是画出语言种系发生树形图,树形图包含两个重要的信息:(1)语言集团的呈阶级式的聚合分类;(2)树枝的长度可以反映语言从祖语分离的时间距离以及各语言间的亲缘程度。

### 3.1 相似矩阵(Similarity Matrix)

我们的具体做法是:首先优选出同源词,编制同源词表,<sup>④</sup>然后计算出每对语言的同源百分比。这部分工作量大,难度大。首先,如果我们不能很好地排除借词,则画出来的树图会与事实相去甚远;其次,词目与义项往往纠葛不清,同一个词目下,不同的语言会有不同义项对应的反映形式,如何正确处理,这取决于对古代音韵以及对该语言的熟知程度。我们采用的是分词目计算,即采用较严格的语义对应原则。这项工作的结果是编制出语言的相似矩阵,如表一。这是计量研究的基础,是决定各语言之间的关系的最关键一步。

表一

	湘西	黔东南	川黔滇	滇东北	布努	巴哼	优诺	炯奈	畲	勉瑶	标敏	藻敏
湘西		68	68	69	66	64	59	63	63	50	47	46
黔东南			73	73	75	65	55	66	66	54	50	52
川黔滇				84	77	65	54	64	63	56	52	53
滇东北					76	66	57	62	65	57	52	53
布努						65	59	64	63	56	51	51
巴哼							64	63	58	49	47	48
优诺								57	57	45	43	44
炯奈									60	53	48	50
畲										53	50	50
勉瑶											64	67
标敏												61
藻敏												

### 3.2 距离矩阵(Distance Matrix)

由于数理树形图是通过分枝的长度来反映语言间的距离的,所以,我们必须把上面的相似矩阵转换为距离矩阵(Distance Matrix)。这种转换可以通过下面公式完成:

$$d = -\log s$$

d代表距离,s代表相似数字;负对数值百分比越大,则距离越小;这表明时间距离越长,同样的词汇就越有机会发生变化。

通过转换,反映类似情况的相似数字可以转换成反映距离的不相似数字,计算出距离矩阵,见表二。为了视觉上的好看,我们将每个数目乘以100。

表二

	湘西	黔东	川黔滇	滇东北	布努	巴哼	优诺	炯奈	畲	勉瑶	标敏	藻敏
湘西		17	16	16	18	19	23	20	20	30	33	34
黔东			14	14	13	19	26	18	18	27	30	28
川黔滇				8	12	19	27	19	20	25	28	27
滇东北					12	18	25	21	19	25	28	27
布努						19	23	19	20	25	29	29
巴哼							19	20	24	31	33	32
优诺								25	25	35	36	36
炯奈									22	27	32	30
畲										27	30	30
勉瑶											19	18
标敏												21
藻敏												

### 3.3 从无根树到有根树

计算出距离矩阵后,就可以计算反映亲缘关系的树形图了。从距离矩阵转换成无根树有许多种方法,所有这些方法都需要大量的计算。有幸的是,我们可以借用生物学家为生物种系发生分类设计出很好的计算程序。其中最有影响的方法之一是1967年由Fitch和Margoliash发明的以及1987年由Saitou和Nei发明的Neighbor joining程序。<sup>⑤</sup>Qiao & Wang(1998)又在此基础上重新设计了一套可以穷尽地分析所有无根树的新方法。为了理解许多计算过程的一些观念,首先需要定义一些和树形图相关的基本术语。一棵树表示一种连结分枝、节点和末端的集合。末端是通过每条末端只连接一条单枝来区别的。末端是通过分析分类单位的特有标记,在我们的例子中,是通过分析语言来标记。每个节点恰好连接三条分枝。三个末端恰好有一棵无根的树,一棵无根的树只有一种表述三种语言的方式,在树的中部,只有唯一的节点N1,连接A、B、C语言。如果我们要在树上添加第4种语言D,D可加在A、B、C三个分枝的任一分支上,把原来的分枝一分为二,这就新创建一棵带一个新节点N2的无根树。3语言树有3条分枝,而4语言树则有5条分枝,三种表示方式。如果我们向该语言集团另添加一种语言E,可以把E添加到5条分枝的任一条上,形成一个新节点N3,会有7条分枝,可以有15种表示方法,因此,5种语言正好有15棵无根树。

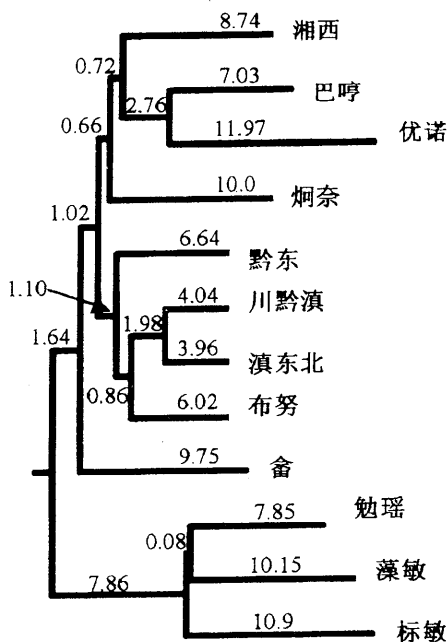
我们可以通过在任一条分枝上插入根,使其生根的方法,来形成一棵有根树。一般的过程是把根置于分离两个末端最远路径的中点。这样,用 $R(L) = B(L) * U(L)$ 的等式来计算从1开始的L-1奇整数的连乘值R(L)。例如,L=12,那么R(12)等于最前面11个整数的积,其值达1374931多万。这就是说这12种语言存在1374931多万种子群方式。因此,在历史语言

学研究中必须用计算机程序来处理这类数据。

任何一对语言之间的距离是分离它们的分枝距离的总和。我们的目的是挑选最合理的语言树形图,根据这棵树我们可以重建另一种距离矩阵,即外向矩阵,外向矩阵必须和原来的距离矩阵即内向矩阵最为相似,应当尽量接近它们的值。

下面即我们根据数理分析得出的苗瑶语言树形图。

图示一



图示说明:

树图上的数字表示距离的长短,树枝的距离只计算每一树枝的端口到根部的横向距离,以及各个树枝横向距离的相加。而不管纵向的关系;属于同簇内的各语言比簇外的各语言关系更密切。我们分析了8种、10种、12种苗瑶语语言数据,用 Qiao 和王士元的穷尽法画出8种语言的最佳的树,分别用 fitch 跟 neighbor joining 两种方法计算出10种和12种的最佳树图,这三种方法所画出树图的结果大同小异,内部结构关系大致相同。特别值得指出的是,不管使用哪一种方法,如果添加了古代汉语的数据,则畲归入瑶语大簇内,而其他结构关系不变,这其中的原因值得研究。可能反映畲语地位正好在苗、瑶语之间。

#### 四 苗瑶语族语言的亲缘关系的分析

##### 4.1 有关同源词数量统计的分析

语言的相似矩阵是建立在同源词计算的基础上的,所以,同源词的确认至关重要。这有两个工作必须做;1,排除借词;2,建立同源词语音对应规律的数据库。由于借词历史层次不同,往往也有较完整的语音对应系统,所以如何正确地区别开借词和同源词,这取决于我们的历史音韵学知识体系以及对整个东亚区域的相关民族的历史文化及其变迁的理解。首先,我们根据苗瑶语的特点,在 Swadesh 的 100 词的基础上,对选词作了修改调整,结果为 111 词。下面我们逐一作出说明:

词目第 12“沙”,苗瑶语反映为明显的汉语借词,替换为 Swadesh 的后 100 词的“盐”。第 26“毛”和第 27“发”,苗瑶语为同一个同源词。第 29“眼”苗瑶语反映形式为明显的古汉语借词,所以,我们用 Swadesh 后 100 词的“(大)腿”替换。第 41“腹”,苗瑶语中,“肚子”、“腹部”、“肠子”,词义交叉,苗语指称“肠子”跟瑶语指称“腹部”的词对应,而瑶语则另有指称“肠子”的词,苗语另有指称“腹部”的词,这样,我们就把“腹部”、“肠子 1”和“肠子 2”分词目单立;第 50 苗瑶语没有指称“女人”的词,我们用“姑娘”替换;第 53“听”,苗瑶语“听”跟“听见”是两个不同来源的词,所以分立;第 54“看”,苗瑶语“看”跟“看见”也是两个不同来源的词,所以分立。第 63“泗”,苗瑶语反映为明显的晚期汉语借词,替换为“潜水”。第 68“睡”和第 100“躺”,苗瑶语

为同一个同源词。第 80“冷”瑶语支分别有两个不同来源的形式,所以,分立为“冷 1”和“冷 2”。此外,我们增补雅洪托夫 35 词中有、而 Swadesh 的 100 词没有的“盐、风、年龄”,还有几个较重要的跟苗瑶语区域经济生态环境密切相关的基本词汇“猪、水獭、穿山甲、摘猪草”,这些都反映苗瑶语民族原始的采集生活和自然环境的特点。

其次,我们排列出用作比较的 111 词,并讨论它们在苗瑶语中的反映形式。经过分析,我们有如下几个发现:

1) 凡 111 词中的名词,即 Swadesh 100 词的前 52 词以及后增补的名词部分,在苗瑶各语言中,同源比例很高,可说是基本同源。而动词、形容词部分则反映较大差异,不同源比例高。

2) 瑶语支语言的汉语借词比例远远高于苗语支,例如勉语大坪江有 13 个借词,标敏有 14 个借词,藻敏大坪则有 18 个借词,相反,湘西、黔东南仅只有个别借词,而川黔滇、布努甚至没有发现借词。而苗语支族属于瑶族的,除了布努外,借词比例略高些。例如优诺有 8 个借词、炯奈有 6 个借词、畲语陈湖有 7 个借词、巴哼白文有 3 个借词。这些借词的音韵系统反映中古音以后的历史层次。这反映了苗瑶语族诸语言“汉化”的程度以及“汉化”速度的不平衡性,呈阶级性表现,瑶语支汉化最早、汉化程度最深,其次是优诺、畲语、炯奈和巴哼;而苗语三大方言和布努语则汉化最迟,比较而言,保留传统最多。

3) 这只是我们在 111 词中观察到的现象,但如果我们把比较的范围扩大到 Swadesh 的后 100 词或更多的词目,情况很可能发生变化,据我们的研究,畲语以及炯奈、巴哼等有相当数量的一批词汇跟客家话同源。另外,据王辅世、陈其光等人的研究,苗瑶语还有相当数量的一批词汇跟古汉语同源。这就论证了我们在前面提到的理论:不同语言的词汇系统可以分为不同的词汇范畴,不同的词汇范畴的词变化的速度和程度是不同的,所以,即使同样是同源词,也不能假定一种相同的变化的值。

#### 4.2 同传统分类的比较

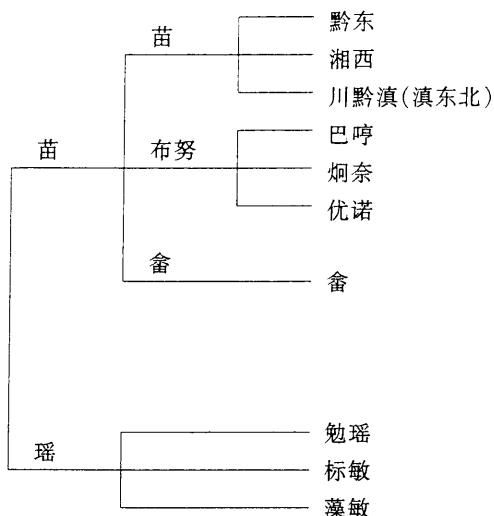
本文前面已列举迄今为止各种主要的苗瑶语分类,为了比较的方便,我们选择了两种最有代表性的苗瑶语分类,即 1988 年的《中国语言地图集》(下文简称《地图》)和 1995 年的王辅世、毛宗武的《苗瑶语古音构拟》(下文简称《构拟》)的分类来作比较(请见下页图示二和图示三)。我们将这两种分类转换为树形图,但这种树形图不同于我们的数理树形图,它的树枝没有任何意义,不表示距离远近,仅仅反映语言聚类而已。

我们的树形图显示,有两个大的簇类,即苗语支和瑶语支,畲语属于苗语支,但畲横跨于苗、瑶之间,与瑶关系距离较近。在苗的大簇内,包含 3 个簇类,即畲簇类,黔东南、川黔滇跟滇东北、布努簇类,湘西、巴哼、优诺、炯奈簇类。苗瑶两大簇分类跟《地图》相同,但跟《构拟》不同,《构拟》是苗、畲、瑶三语支并列,反映三个平行等距离的聚类。

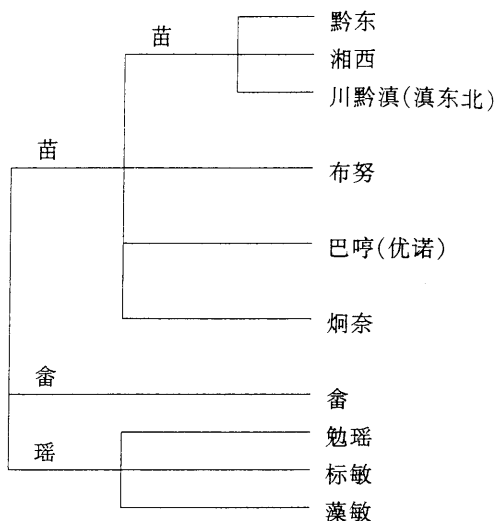
方言和语言的定位不同。《地图》和《构拟》都一致认为苗语包括湘西、黔东南、川黔滇三大方言,内部关系密切程度超过布努语,其界限分明,湘西、黔东南、川黔滇三大方言反映的是一种方言之间的关系,而苗语跟布努语反映的则是一种语言之间的关系。然而,我们的数理树形图表明:川黔滇与滇东北距离最近,关系最密切,可以理解为方言之间的关系。其次,布努语跟川黔滇与滇东北距离最近,相形之下,黔东南和湘西跟川黔滇与滇东北的距离都远于布努语跟川黔滇与滇东北距离,也就是说,布努语跟川黔滇与滇东北属于同一个聚类,而黔东南和湘西则属于另外一个更远一层的聚类。这是我们跟传统分类的一个较大的分歧。

下面具体比较我们跟传统分类的异同:

图示二 《中国语言地图集》分类



图示三 《苗瑶语古音构拟》分类



《地图》将巴哼、优诺、炯奈都归于布努语，当作布努语的方言，这显然是没有道理的。因为不论从族群认同，还是语言结构上看，都不应作此种分类。从族群认同看，他们分属于瑶族不同的支系，各有自己的自称；从语言数理分析看，巴哼、优诺、炯奈跟湘西近，甚至跟黔东的距离近于跟布努的距离。《构拟》将布努、巴哼、炯奈分立为苗语支下的不同语言，但却把优诺当作巴哼的方言看待。这种分类也是有问题的。从我们的树图看，巴哼、优诺、炯奈跟湘西的距离较近，而畲语则跟布努关系最近，畲语是相对独立的一支，但《构拟》却将畲语单立为一个独立的语支，与苗语支平行。这显然是有问题的。

《地图》和《构拟》都没有细分出瑶语支各语言的较小的聚类 and 关系程度，我们的图示一显示，瑶语支内部勉瑶跟藻敏的距离最近，然后是标敏。就整个图示一来看，苗瑶语言距离最远的两个语言是优诺跟标敏，虽然优诺跟标敏同是瑶族，但他们的关系却最为疏远，这说明语言分类跟民族分类的差异性。

就苗语支跟瑶语支比较来说，跟勉瑶的树枝距离最近的是苗族的川黔滇和滇东北，而并非跟勉瑶同属瑶族的优诺、巴哼和炯奈。

#### 4.3 同其他计量方法的分类结果的比较

我们在前面提到黄行用计量的方法作苗瑶语的分类研究，但这只是苗瑶语的亲疏关系的研究而不是苗瑶语的亲缘关系的研究。黄行的论文进行的是聚类分析，采用的是平均连结的方法以得到语言之间的分类。他得到的树形图的末端到根部的距离都是等距离的，不能反映不同语言的不同变化速度。在这样的树图中，一个明显的特征就是每一个方言和这些方言的母语的遗传距离相等。这种方法用来分析方言间的亲疏关系无可厚非，但用来表示方言间的亲缘关系便不合适，因为我们没有理由假定语言变化的速度是一直均衡的，从而，每一个方言和其母语的距离一致。我们认为更容易接受的假设是，方言是以不同的速度发生变化的。从这个假设出发，用平均系连法分析出来的树图就无法表述方言间的亲缘关系。

对比我们和黄树的树图，可发现同异之处；由于黄文只比较苗语（湘西、黔东、川黔滇三大方言及其次方言）、布努语、畲语和瑶语支的勉和标敏语，而没有比较巴哼、炯奈、优诺和藻敏语，所以，我们只比较相同的语言。

相同点:

- 1) 树图的大簇分类基本相同,即如果两分,则可以分苗、瑶两大聚类,畬语归苗语支。
- 2) 布努语插入苗语三大方言内部,这是跟传统分类不同的共同点。而且,黄的数据表明畬与布努最近,这跟我们的树图完全一致。

不同点:

- 1) 在苗语支内,黄的布努跟黔东南关系最近。而我们的树图则表示布努跟川黔滇和滇东北关系最近。应该说,我们的分类更接近历史上族群的分布和接触关系,即自称为“布努”的瑶人跟川黔滇区域的苗人关系特别密切。此外,斯特雷克用定性研究方法得出的结论是:布努为川黔滇下属的方言。这也证明了我们的分类更合理。

- 2) 我们的树图树枝长度可以表示个体语言变化的速度,树枝长度的加和表示语言之间的亲缘距离。黄文的树图就没有这些方面的信息,反映的只是语言亲疏的关系。

#### 4.4 畬语的系属和地位问题的讨论

分类和来源问题的研究一直是语言学最关注、最具有突破性意义的研究课题之一。有关畬语的系属和地位问题的讨论,语言学界颇多争论。这主要有两种意见:

第一,陈其光认为根据畬语的语音和语法特点,畬语应划归瑶语支。

第二,毛宗武、蒙朝吉认为根据畬语的基本词汇特点,畬语应划归到苗语支。毛选取了150个同源词来作百分比统计。但陈批评这种方法有选词方面的主观随意性,即倾向性。

根据我们的研究,畬语的表层结构确实较接近瑶语支,借用的成分较多,这暗示畬的“汉化”过程跟瑶相似,反映语音和语法变化速度较快;但深层结构如最常用的基本词汇的形式则较接近苗语支。所以这两个特点应该分别看待,这反映了发生学与类型学上的差异。在我们的树图中,畬是在苗的大簇内相对独立的一枝。但有一点很值得注意,数理统计出来的树枝的长度表明,畬并不是距离瑶语支最近的语言,距离瑶语支最近的分别是川黔滇、黔东南和布努,距离瑶语支最远的则是优诺和巴哼语。这个结论跟凭直觉判断的传统观点很不一致,传统认为畬应跟瑶最近。这就是整体比较跟每对语言单独比较的区别。

### 五 结语:语史观和方法论上的重要意义

我们根据一种语言中的不同的词汇范畴具有不同的变化程度,即使是基本词汇的变化程度也不能假定为一种相同的值的理论,认为 Swadesh 的 100 词可以用作苗瑶语分类的标准,同时,我们根据苗瑶语的人文特点,在 Swadesh 的 100 词的基础上,排除借词等文化接触方面的因素,补充到 111 词,最后,画出语言数理种系树形图,得出苗瑶语的分类以及计算出每对语言之间的亲缘距离程度。这种方法可以用公式来推导,并可反复使用和验证。我们比较传统的分类和其他用计量方法研究苗瑶语的分类,认为语言的亲缘关系分类不同于语言的亲疏关系的研究,词源统计法并不同于词汇统计法,前者研究历时,而后者研究共时。从分类的结果看,我们跟传统的分类还是有重要的分歧之处:我们把苗瑶两分,而传统苗瑶畬三分;把布努划入苗语三大方言内,由原来的方言之间分类改变为语言之间的分类,而传统则苗、布努分立。传统无法精确计算每对语言的距离程度,凭经验认为畬跟瑶亲缘距离最近,而我们的树图则表明川黔滇跟瑶亲缘距离最近。我们的数理树图可以从整体上把握全部 12 种语言的相互关系,并可发现一些有意义的问题。而传统分类则只能作每对语言的定性比较,无法作数理分析。

长期以来,传统的历史语言学总是希望通过同源词的定性分析,从而重建“想像”中的古代

滇东北)

诺)

的。因为  
于瑶族不  
东的距离  
优诺当作  
西的距离  
一个独立

图一显  
距离最远  
说明语言

北,而并非

疏关系的  
均连结对  
不能反映  
些方言的  
六方言间的  
一个方言  
变化的。从

黔滇三大方  
和藻敏语,

语 文

音韵体系,虽然也强调建构各个语言自己的结构系统,但不作量化分析,凭经验作定性分类,更无法作亲缘距离关系程度的数量描述。其结果问题很多,难以完善。运用词源统计分析法对语言作数理分类,并描述出语言之间亲缘距离的程度,这种研究在国内仍属开创阶段。但毫无疑问,这种方法能够很好地跟计算机和生物学领域的最新发展的理论和方法相结合,对中国历史语言学的重新发展具有突破性的意义。

#### 附 注

① 语言学数理分析表明历史上确实存在一个苗瑶共同体。人文学科的研究亦表明苗瑶语民族集团历史演化过程:传说时代的“三苗”→商、周时代的“荆蛮”集团→秦汉、六朝时代洞庭“武陵蛮”、“五溪蛮”→隋唐时代苗、瑶民族→唐宋时代苗、瑶、畬民族。

② 关于语言与方言的分类:由于我国民族语言和方言的分类理论和方法还不完善,还未建立起完整的分类体系,导致语言和方言的界定不清。有些语言学家可能受到传统民族学理论影响,在作语言分类时,往往把独立的语言当作方言来看待,按照传统民族学分类理论(例如斯大林民族理论),一个民族必须具有共同的语言,所以将瑶族讲的“巴哼”话、“炯奈”话、“优诺”话当作同是瑶族讲的“布努”语的方言来分类。实际上,民族学的分类同语言学分类的标准不同,关系错综复杂。词源统计法理论上并没有什么特别对区分语言和方言有效的方法,不过在诠释得到的树图结论时,我们发现苗瑶语族 12 支语言,放在“语言”层面处理比起把有的视作“语言”,有的视作“方言”来处理更为客观。从我们的数理树图看,“布努”等的分布打破了传统的语言和方言的格局。

③ 有关语言亲疏关系与亲缘关系的不同的研究方法的比较可参考王、沈(1992)的文章。

④ 限于篇幅,本文的词汇附录省略。有关同源词的确认工作得到陈其光、毛宗武两先生的大力协助,深表感谢。词汇引自中央民族学院苗瑶语研究室(1987),王辅世、毛宗武(1995),毛宗武、李云兵(1997),毛宗武、蒙朝吉(1986)。

⑤ 这两个程序被收进了由 Felsenstein 编制的一个可在个人计算机上运行的软件包 PHYLIP 里。PHYLIP 是个被广泛使用的种系发生学关系分析的软件包。

#### 参考文献

- 本尼迪克特(Benedict, P. K) 1992 《苗瑶语之迷:那峨语》,乐赛月译《民族语文研究情报资料集》第 14 期。
- 陈其光 1984 《畬语在苗瑶语族中的地位》,《语言研究》第 1 期。
- 邓晓华 1994 《南方汉语中的古南岛语成分》,《民族语文》第 3 期。
- 1999 《客家话与苗瑶语、壮侬语的关系》,《民族语文》第 3 期。
- 1999 《福建境内的闽、客族群及畬族的语言文化关系比较》,《国立民族学博物馆研究报告》24 卷 1 号,日本。
- 邓晓华 王士元 2003 《藏缅语族语言的数理分类及其形成过程的分析》,《民族语文》即刊。
- 董同龢 1953 《中国语言》,《中国文化论集》第一集,又见于丁邦新编《董同龢先生语言学论文选集》,食货出版社,台北,1974。
- 黄 行 1999 《苗瑶语方言亲疏关系的计量分析》,《民族语文》第 3 期。
- 蒋炳钊 1988 《畬族史稿》,厦门大学出版社。
- 罗常培 傅懋勣 1954 《国内少数民族语言文字的概况》,《中国语文》第 3 期。
- 罗美珍 邓晓华 1995 《客家方言》,福建教育出版社。
- 毛宗武 蒙朝吉 郑宗泽 1982 《瑶族语言简志》,民族出版社。
- 毛宗武 蒙朝吉 1985 《试论畬语的系属问题》,《中国语言学报》第 2 期。
- 1986 《畬语简志》,民族出版社。
- 毛宗武 李云兵 1997 《巴哼语》,上海远东出版社。

定性分类,更  
充分析法对  
介段。但毫无  
合,对中国历

吾民族集团历史  
“蛮”→隋唐时

建立起完整的分  
分类时,往往把  
具有共同的语  
。实际上,民族  
分语言和方言有  
且比起把有的视  
系统的语言和方

的大力协助,深  
兵(1997),毛宗

包 PHYLIP 里。

斗集》第 14 期。

究报告》24 卷 1

文选集》,食货出

国 语 文

- 蒙朝吉 1993 《畲语属苗语支补证》,《民族语文》第 2 期。
- 斯特雷克(David Strecker) 1992 《简评白保罗的“苗瑶语之迷:那峨语”》,乐赛月译《民族语文研究情报资料》第 14 期。
- 唐 纳(Gordon B. Downer) 1986 《原始苗瑶语构拟中的一些问题》,《民族语文研究情报资料集》第 7 期。
- 王辅世 毛宗武 1995 《苗瑶语古音构拟》,中国社会科学出版社。
- 王辅世 1986 《苗瑶语的系属问题初探》,《民族语文》第 1 期。
- 王士元 沈钟伟 1992 《方言关系的计量表述》,《中国语文》第 2 期。
- 王士元 2000 《语言的探索——王士元语言学论文选译》,北京语言文化大学出版社。
- 王文光 1999 《中国南方民族史》,民族出版社。
- 吴永章 1993 《瑶族史》,四川民族出版社。
- 伍新福 龙伯亚 1992 《苗族史》,四川民族出版社。
- 张 琨 1947 《苗瑶语声调问题》,《中研院历史语言研究所集刊》第 16 本。
- 1983 《苗瑶语声调的构拟》,《民族语文研究情报资料集》第 1 期。
- 郑锦全 1988 《汉语方言亲疏关系的计量统计》,《中国语文》第 2 期。
- 中国少数民族语言研究所(主编) 1959 《中国少数民族语言简志——苗瑶语族部分》,科学出版社。
- 中国社会科学院、澳大利亚国立大学 1988 《中国语言地图集》,朗文出版社。
- 中央民族学院苗瑶语研究室 1987 《苗瑶语方言词汇集》,中央民族学院出版社。
- 竹村卓二 1983 《瑶族的历史和文化》,日本国立民族学博物馆调查报告。
- Cavalli-Sforza, L. L. & W. S. Wang 1986 Spatial distance and lexical replacement. *Language*. 62. pp38 - 55.
- Chang, Kun 1953 On the Tone System of the Miao-Yao Language, *Language*, 29.
- 1976 The Prenasalized stop initials of Miao-Yao, Tibeto-Burman, and Chinese, 《中研院历史语言研究所集刊》第 47 本第 3 分,王辅世译文载《汉藏语系语言学论文选译》。
- Chao, Yuen-ren 1943 Languages and dialects in China, *Geographical Journal*. 102.
- Dyen, Isidore, Joseph B Kruskal, & Paul Black 1992 An Indoeuropean Classification: A Lexicostatistical Experiment. *Transactions of the American philosophical Society*, Vol. 82, Part 5.
- Ellegard, A. 1959 Statistical measurement of linguistic relationship. *Language*. 35.
- Li, Fang-gui 1937 Languages and dialects of China. In *Chinese Yearbook*, Shanghai. Reprinted in *Journal of Chinese Linguistics* 1, 1973.
- Qiao, Sanzheng & Wang, Willian S-Y 1998 Evaluating Phylogenetic Trees By Matrix Decomposition. *Anthropological Sciences*. 106 (1), pp1 - 22.
- Swadesh, M. 1952 Lexico-statistic dating of prehistoric ethnic contacts, *Proceedings of the American Philosophical Society* 96.
- 1955 Time depths of American linguistic groupings, *American Anthropologist* 56.
- Teeter, K. V 1963 Lexicostatistics and genetic relationship. *Language* 39.
- Wang, Willian S-Y. 1994 Glottochronology, Lexicostatistics, and Other Numerical Methods. *Encyclopedia of Language and Linguistics*. Pergamon Press.
- (ed.) 1995 *The Ancestry of the Chinese Language*. *Journal of Chinese Linguistics* Monograph No. 8.
- 1995 Linguistics diversity and language relationships. In *New Horizons in Chinese Linguistics*. James Huang and Audrey Li, editors. Kluwer Academic Publishers.
- 1995 A quantitative study of Zhuang-Dong Languages, 余霭芹、远藤光晓编《桥本万太郎纪念——中国语言学论集》,内山书店。

(邓晓华 厦门大学人类博物馆 360000; 王士元 香港城市大学电子工程学系)

2003 年第 3 期

· 263 ·