

苗瑶语族语言的亲缘关系的计量研究*

-----词源统计分析方法

王士元 邓晓华

提要 运用词源统计分析法对苗瑶语族语言作数理分类, 比较传统的分类和其他计量方法的分类结果, 认为 Swadesh 的 100 词可以用作苗瑶语族分类的标准, 并描述出苗瑶语族语言之间亲缘距离的程度, 同时也说明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。

关键词 苗瑶语族 词源统计分析法 树形图 亲缘距离 历史语言学

A Quantitative Study on the Genetic Relationship of Miao-Yao Languages—the Lexicostatics Approach

William S-Y. Wang Deng Xiaohua

This study applies the lexicostatics method to quantitatively classify a group of Miao-Yao languages. The resultant classification is compared with those obtained from traditional methods and other computational methods. It confirms that the first 100 basic words proposed by Swadesh provide a valid basis for classifying Miao-Yao languages. The lexical distance between languages are described in the classification derived from the lexicostatics analysis. It has been observed that words in different lexical categories have different replacement rates. Even words in the basic vocabulary cannot always be assumed to have a similar replacement rate.

一、苗瑶语族语言的关系的研究概况

1、传统的苗瑶语分类

这里只讨论苗瑶语族内部的分类, 不涉及有关苗瑶语的系属分类问题。

李方桂(1937、1973), 赵元任(1943), 董同和(1953), 罗常培、傅懋绩(1954)对苗瑶语族内部的划分十分相似, 都只分为苗语支和瑶语支。苗语支只有苗语, 瑶语支只有瑶语, 仅两种语言。

1954年, 法国学者欧德里古尔指出越南的那峨语(又称巴腾语, 中国学者称巴哼语)属于苗瑶语族。1959年, 《中国少数民族语言简志苗瑶语族部分》将苗瑶语族分为两个语支, 苗语支包括苗、布努两个语言, 瑶语支仅有勉一个语言。1962年, 毛宗武等将“巴哼”话、“炯奈”话、“优诺”话归为布努语的方言。1982年的《瑶族语言简志》仍作此分类。1982年, 毛宗武、蒙朝吉认为畲语属苗瑶语族的苗语支。1984年, 陈其光反对将畲语划归苗语支, 认为应划归瑶语支。1987年, 美国学者斯特雷克将苗瑶语族分为7个语支: 苗语支包括黔东南苗语、湘西苗语、川黔滇苗语和未定的9个语群; 瑶语支包括勉金门语、标交语和藻敏语, 巴哼语支包括巴哼语和那峨语, 另外还有唔奈语支, 炯奈语支、优诺语支、畲语支。但他把布努语划归川黔滇下属的方言。

* 本课题的研究得到香港政府“中国南方濒危语言研究”(项目批准号: 9040237)及台湾中研院“中国各民族的起源及形成过程”的专项研究基金的资助。课题由王士元教授主持并指导。本文是本课题的阶段成果之一。课题组成员南开大学石锋教授、北京大学陈保亚教授、香港城市大学向倩先后为本课题的资料收集整理做了大量的工作。课题组成员香港城市大学柯津云小姐为本文的数据计算付出许多辛勤的劳动, 香港城市大学 James Minett 博士亦为本文的计算工作提供了许多帮助。苗瑶语专家陈其光教授、毛宗武教授为本文的同源词识别提出很好的意见。在此一并致以深深的感谢。

从族属看,苗语为苗族语言,有3大方言:湘西、黔东、川黔滇;而巴哼、炯奈、优诺、布努都是瑶族的支系;畲语则为畲族语言。

1995年,王辅世、毛宗武的《苗瑶语古音构拟》修改他们以前的观点,认为苗瑶语族分为3个语支:苗语支包括苗语、布努语、巴哼语(优诺在内)和炯奈语,瑶语支只有勉语,畲语支仅畲语。

最近,毛宗武再次修改他原先的观点,他认为苗瑶语族可分为语支、语丛、语言、方言4个层次,苗瑶语族分为苗语支和瑶语支,苗语支下面分为苗语(湘西方言、黔东方言、川黔滇方言),布努语,(布努方言、瑙格劳方言、努茂方言),巴哼语(巴哼方言、唔奈方言),优诺语,炯奈语,畲语,其中巴哼语和优诺语为一语丛,炯奈语和畲语为一语丛。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》(以下简称“地图”)将苗瑶语族分为苗语支:苗语(黔东方言、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),畲语,瑶语支:勉语(勉金方言、标交方言、藻敏方言);这代表政府权威机构的看法。

2、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、畲语元音不分长短;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。

3、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。黄主要借鉴郑锦全的方法,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。

(2)、使用的方法不同,黄用平均距离法。

(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。王士元曾指出:“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。”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亲缘关系。

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的共享程度来判断。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,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。

1、优选核心同源词的原则和问题

如何优选核心同源词, 即设计一种通用的基本词汇表是词源统计分析法的最重要步骤之一。这个问题一直存在较大争论, 例如蒂特(Teeter)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)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、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, 如何把通过时间的纵向传播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

数理分类学分类的主要步骤有: 1、特征选择, 2、特征量化, 3、计算相关系数, 4、聚类分析。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亲缘程度。

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, 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(Similarity Matrix)，如表 1。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 1:

	湘西	黔东	川黔滇	滇东北	布努	巴哼	优诺	炯奈	畚	勉瑶	标敏	藻敏
湘西		68	68	69	66	64	59	63	63	50	47	46
黔东			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
畚										53	50	50
勉瑶											64	67
标敏												61
藻敏												

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 2。为了视觉上的好看，我们将每个数目乘以 100。

表 2:

	湘西	黔东	川黔滇	滇东北	布努	巴哼	优诺	炯奈	畚	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、从无根树到有根树

计算出距离矩阵后，就可以作各种聚类分析了。测量时间深度基本上有两种方法：距离法和最小法。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，生物学家为生物种系发生分类设计出很好的计算程序。最有影响的是，1967年由 Fitch 和 Margoliash 发明的以及 1987年由 Saitou 和 Nei 发明的程序，1990年由 Felsenstein 将此两种合成为称 PHYLIP 的软件。Qiao 和王士元 1998 年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 三个分枝的任一分枝上，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，五种语言正好有十五棵无根树。

由于每个节点连接的分枝恰好限制在三条，一棵表示 L 种语言的无根树正好有 2L-3 条分枝，如下面 B 行所示。为了计算表示 L+1 种无根树的数目 U(L+1)，我们用 B(L)乘以 U(L)。计算 U(L)的等值方式是连乘从 1 开始的 L-2 个的奇整数。例如，如果 L=12，那么 U(12)的值为 $1 \times 3 \times 5 \times 7 \times 9 \times 11 \times 13 \times 15 \times 17 \times 19$ ，即 654729075。这就是说，用无根树表示 12 种语言有 654729075 种方式。

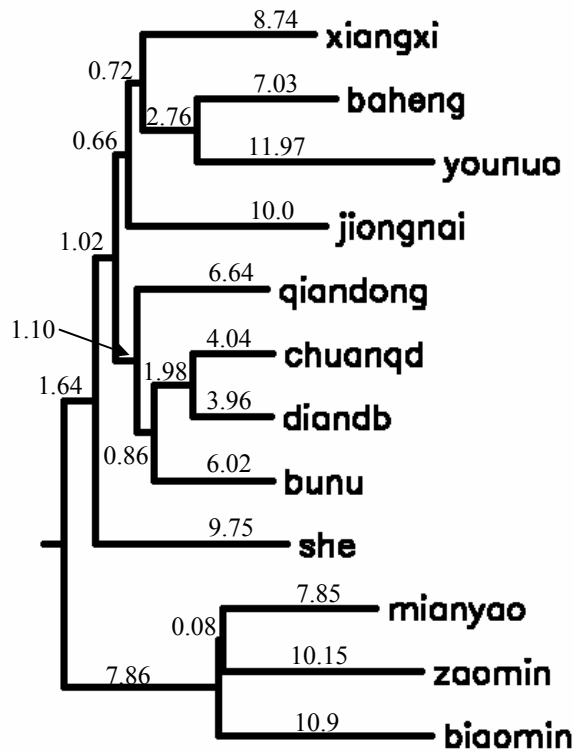
L, 语言数量	3	4	5	6	7	8	9	12
B, 分枝数量	3	5	7	9	11	13	15	21
U, 无根树	1	3	15	105	945	10395	135135	654729075
R, 有根树	3	15	105	945	10395	135135	2027025	13749310575

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，用上表 $R(L)=B(L) * U(L)$ 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，表 3：



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 fetch 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，则畲语归入瑶语大簇内，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

(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。见附表。

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 词有，而 swadesh 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、我们可以发现这么一个有特点，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，可以说是基本同源。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。而苗语支里面族属属于瑶族的，除了布努外，借词比例略高些。例如优洛有 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，其次是优洛、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。

(2)、同传统分类的比较

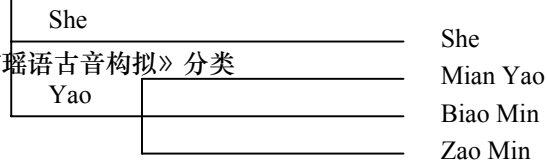
本文前面已列举迄今为至各种主要的苗瑶语分类，为了比较的方便，我们选择了两种最有代表性的苗瑶语分类，即 88 年的《中国语言地图》和 95 年的王辅世、毛宗武的《苗瑶语古音构拟》的分类来作比较。我们将这两种分类转换为树形图，但这种树形图不同于我们的数理树形图，它的树枝没有如何意义，不表示距离远近，仅仅反映语言聚类而已。

下面具体比较我们跟传统分类的异同：

图示 1: 《中国语言地图》分类



图示 2: 《苗瑶语古音构拟》分类



比较分析如下：

1、在我们的树形图显示，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。

2、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。

3、《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，甚至跟黔东南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。

4、《地图》和《构拟》都没有细分出瑶语支各语言的较小的聚类和关系程度，我们的树图显示，瑶语支内部勉语跟藻敏的距离最近，然后是标敏。就我们的整个树图来看，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。

5、就整个苗语支诸语言来说，川黔滇和滇东北跟勉瑶距离最近。

(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。黄用平均距离法，使用平均系连法是根据方言间的相互系数决定的，其结果是一个树图。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。黄文的方法，仍是使用一种承认树的所有末端离根部是等距离的，假定变化率是恒定的平均值联系和不加权每对集团的算术平均值的方法。

对比我们和黄的树图，可发现同异之处：

由于黄文只比较苗语(湘西、黔东南、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，畲语归苗语支。
- 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，黄的数据表明畲与

布努最近，这跟我们树图完全一致。

不同点：

在苗语支内，黄的布努跟黔东关系最近。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。

3、我们的树图的树枝表示语言之间的距离的亲缘关系，而黄文的树枝仅仅表述等距离。

(4)、畲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。有关畲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：

1，陈其光认为根据畲语的语音和语法特点，畲语应划归瑶语支。

2，毛宗武、蒙朝吉认为根据畲语的基本词汇特点，畲语应划归到苗语支。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，畲语的表层结构确实较接近瑶语支，借用的成分较多，这暗示畲的“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，畲是在苗的大簇内，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，畲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，传统认为畲应跟瑶最近。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的。我们比较传统的分类和其他的用计量方法来研究苗瑶语的分类，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，而传统苗瑶畲三分；我们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，凭经验认为畲跟瑶亲缘距离最近，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，并可以发现一些有意义的问题。而传统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”中的上古音体系，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极具突破性的意义。

参考文献

- Cavalli-Sforza, L.L. and W.S.-Wang. 1986. Spatial distance and lexical replacement. *Language*. 62. pp38-55
- Dyen. Kruskal. and Black. 1992. An Indo-European 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, F.K. 1973. Languages and dialects of China. *Journal of Chinese Linguistics* 1.1
- Swadesh, M., 1952, Lexico-statistic dating of prehistoric ethnic contacts, with special reference, *Philosophical Society* 96
- Swadesh, M., 1955, Time depths of American linguistic groupings, *American Anthropologist* 56.
- Teeter, k.v. 1963. Lexicostatistics and genetic relationship. *Language* 39
- Sanzheng .Qiao and Wang, Willian S-Y ,1998 Evaluating Phylogenetic Trees By Matrix Decomposition. *Anthropological Sciences*. 106 (1), pp1-22
- Wang, Willian S-Y., 1994. Glottochronology, Lexicostatistics, and other Numerical Methods. *Encyclopedia Of Language and Linguistics*. Pergamon Press.
- 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, pp81-96, 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。
- 王士元、沈锺伟，1992，方言关系的计量表述，中国语文
- 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。
- 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
- 郑锦全，1988，汉语方言亲疏关系的计量统计，中国语文第2期
- 黄行，1999，苗瑶语方言亲疏关系的计量分析，民族语文第3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，1986.1，《苗瑶语的系属问题初探》，《民族语文》
- 陈其光，1984，《畲语在苗瑶语族中的地位》，《语言研究》
- 毛宗武、蒙朝吉，1985，《试论畲语的系属问题》《中国语言学报》
- 毛宗武、蒙朝吉，1986，《畲语简志》民族出版社
- 蒙朝吉，1993，《畲语属苗语支补证》，《民族语文》
- 毛宗武、李云兵，1997《巴哼语》上海远东出版社
- 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
- 邓晓华，1994，《南方汉语中的古南岛语成分》，《民族语文》
- 1999，《客家话与苗瑶语、壮侗语的关系》民族语文 第三期
- 1999，《福建境内的闽、客族群及畲族的语言文化关系比较》《国立民族学博物馆研究报告》99年24卷1号，日本
- 罗美珍、邓晓华，1995，《客家方言》，福建教育出版社
- Benedict, P.K., 本尼迪克特，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
- David Strecker, 斯特雷克，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集》第14期
- Chang, Kun, 张琨，1947，《苗瑶语声调问题》，中央研究院历史语言研究所集刊》第16本
- Chang, Kun, 1953. On the Tone System of the Miao-Yao Language, *Language*, 29.
- Chang, Kun, 张琨，1976, The Prenasalized stop initials of Miao-Yao, Tibeto-Burman, and Chinese, 《中央研究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》
- Chang, Kun, 张琨，1983，《苗瑶语声调的构拟》《民族语文研究情报资料集》第1期
- Gordon B. Downer, 1986, 唐纳《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

竹村桌二，1983，《瑶族的历史和文化》日本国立民族学博物馆调查报告
王文光，1999，《中国南方民族史》，民族出版社
伍新福、龙伯亚，1992《苗族史》，四川民族出版社
蒋炳钊，1988，《畲族史稿》，厦门大学出版社
吴永章，1993，《瑶族史》，四川民族出版社

附录：苗瑶语族语言核心同源词对应表

词表说明：在苗瑶语中，声调形式，即调类反映形式是判断是否同源词的非常重要的标准，不管声、韵变化得如何面目全非，但声调的调类应相同，或至少可以对应。例如，“水”在苗语支各语言音韵形式差异很大，但调类和语义却一致，即都是第一调，所以，可以确认为有同源关系。为了便于观察同源词的形式，我们根据在《构拟》书中能找到的同源词调类形式，在音标的右上角用单位数标识出来，而双位数则表示调值。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。

