



北京市社会科学理论著作出版基金资助

汉藏语言比较的方法与实践

——汉、白、彝语比较研究

汪 锋◎著



北京大学出版社
PEKING UNIVERSITY PRESS



北京市社会科学理论著作出版基金资助

汉藏语言比较的方法与实践

——汉、白、彝语比较研究

汪 锋 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

汉藏语言比较的方法与实践: 汉、白、彝语比较研究 / 汪锋著. —北京: 北京大学出版社, 2013.8

ISBN 978-7-301-22905-7

I. 汉… II. 汪… III. 汉藏语系—对比语言学—汉语、白语、彝语 IV. H4

中国版本图书馆 CIP 数据核字 (2013) 第 173199 号

书 名: 汉藏语言比较的方法与实践——汉、白、彝语比较研究

著作责任者: 汪 锋 著

责任编辑: 杜若明

标准书号: ISBN 978-7-301-22905-7/H·3350

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn> 新浪官方微博: @北京大学出版社

电 话: 邮购部 62752015 发行部 62750672 编辑部 62753374 出版部 62754962

电子信箱: zpup@pup.pku.edu.cn

印 刷 者: 北京富生印刷厂

经 销 者: 新华书店

787毫米×1092毫米 16开本 17.5印张 441千字

2013年8月第1版 2013年8月第1次印刷

定 价: 40.00 元

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究 举报电话: 010-62752024

电子信箱: fdp@up.pku.edu.cn

本书的研究先后得到以下项目资助：

国家社科基金青年项目（#07CYY025）《白彝关系语素研究》

霍英东教育基金会资助项目（#131102）《勒墨白族的语言及口传文化的数字化保护及基础理论研究》

国家社会科学重大项目（#10&ZD125）《中国有声语言及口传文化保护与传承的数字化方法研究和基础理论研究》

教育部人文社会科学重点研究基地重大项目（#11JJD740004）《基于系统语音对应的核心词分阶及建模研究》

本书的出版得到北京市社科理论著作出版基金的资助。

序

王士元

语言总是在变的，最早把这句话讲得很清楚的，应该是明末的陈第。众所皆知，《诗经》里的语言与我们相隔两三千年，所以历经大量的演变。清朝学者在陈第的理论基础上，开始把古汉语的音韵逐步拟构出来，让我们体会到音变的来龙去脉，这确实是功不可没的贡献，值得中国学术界引以为傲。

同样的道理，两个人群虽然原本说的是同一个语言，可是因为迁居而分开后，久而久之，他们的语言就会因变迁的方向不同而变得越来越不相像。18 世纪末的 William Jones 就是观察到这一点，才把欧洲的语言和远在印度的梵文归纳为同一个印欧语系，接着又有一些非常杰出的语言学家，用所谓的比较方法，历时一百多年，把这个语系里几百个语言的谱系关系细心地研究出来，这些成就不愧为现代语言学的开端。直到现在，印欧语言学的研究仍旧在语言学界享有崇高的地位。

中国境内的语言材料非常丰富，可是系统的研究起步得比较晚，也有人认为，因为东亚的一些语言没有复杂的构词屈折，所以根本谈不上用比较方法来研究它们的谱系。

李方桂先生 1937 年的文章，是一次成功的开始，首次为汉藏语系提出了一个有用的假设，把中国境内汉藏语系分为四大支。可是因为问题的确非常复杂，争议很多，甚至有人指出“汉藏”一词根本就不恰当，如果改用“藏缅”来代表这个语系，反而更符合语言历史。

这个领域的研究没有能够达成共识的原因，主要可从两方面来谈，一方面是学者们系统整理出来可供运用的材料，不是不够丰富就是有欠严谨，因为有些关键的少数民族语言，分布在深山茂林里，不容易亲身去研究记录；另一方面是，某些语言并不使用构词屈折，所以我们只能透过其他方法来研究，如利用词汇上的对应。

本书的作者汪锋教授，之所以在少数民族的语言研究上能达到如此高水平的成就，正是因为他同时克服了这两方面的困难。他并不是只依赖别人所搜集的第二、三手材料，而是亲自走遍了云南各地，多次造访不同的村落以收录语言材料，他所收集的语料，绝大部分都清晰有序地呈现在本书的图表里，作为他分析问题的依据。这些材料也一部分出现在他 2006 年讨论白语的那本书里。

另一方面更重要的是，他充分利用一些新的语言理论，以发挥词汇的比较功能，这些理论的源头，来自 Morris Swadesh 于 1950 年代提出的一些关于基本词汇的假设。基本词汇

在时代上的消失,当然不可能像碳 14 定年那样完全规则,因为语言毕竟不是一种化学元素,可是 Swadesh 的假设,尤其是加上陈保亚教授的建议,把基本词汇分阶来帮助了解历史,是个很得力的理论工具。而本书的成功,也基于汪锋很合时宜地应用了这些新理论。

本书的中心贡献,当然是它替我们澄清了汉、白、彝三者之间的谱系关系,这是个争论已久的课题,如今有了这么一个科学的解答,无疑是件可喜的事。同时书里也提供了很多宝贵的语言材料,将来其他有兴趣的志同道合者可以慢慢继续琢磨、研究。

最后,作者引用了 van Driem 的比喻,把目前的一些语料看作是一棵老树下一堆堆的落叶。作者说:“在本研究中,我们将一片原始白语叶、一片汉语叶和一片原始彝语叶重新缀起,我们深知,这仅仅是复原了根深叶茂的汉藏大树的微小一枝。不过,我们相信,随着越来越多同道者的加入,在不久的将来,这棵大树终将在众人辛勤汗水的浇灌下露出原本遒劲的枝桠。”我觉得这段话讲得很正确,也很富有诗意,我也企盼在不久的将来,作者的期待会如愿实现。

2011.08.11 于香港中文大学

序：汉语藏缅语同源的 一个重要证据

陈保亚

汪锋这部著作的研究成果再次为汉语和藏缅语同源提供了重要证据。

汉语和藏缅语有没有同源关系，并没有得到严格的证明。19 世纪末西方有些学者认为汉语和藏缅语有同源关系，主要是以类型相似为标准。但是，近几十年的调查显示，语言接触会造成同一地域的语言在类型上相似，所以根据汉语和藏缅语之间的类型相似来断定汉语和藏缅语有同源关系是不能成立的，寻找同源词才可能是一个有效的路子。

20 世纪上半叶，Laufer、Simon、Shafer、Benedict 陆续在汉语和藏语之间找出了几百个语音上有对应的词，他们认为这些对应词就是汉藏同源词。但是他们建立的语音对应规则并不严格，择词的标准也比较随意。从 20 世纪下半叶开始，Bodman、龚煌城、Coblin、俞敏等陆续在汉语和藏语、缅语之间找出了一批语音对应比较严格的对应词，其中龚煌城的对应标准更为严格。龚煌城、Bodman、Coblin、俞敏等学者认为这些汉藏对应词就是汉藏同源词。事实上语言的深刻接触也可以造成成批的对应词。严格的语音对应是语源研究的必要条件，并不是充分条件。这是历史比较语言学没有深入考虑过的问题。龚煌城等的研究给汉藏同源研究提供了一个必要条件，即这些对应语素由于是依据上古汉语和古藏语（或古缅语）的对应建立起来的，很可能是最早时间层面的对应语素。但是，如何判定这些最早时间层面的对应语素是同源词而不是借词，仍然是一个没有解决的问题。

我曾经汇集龚煌城等学者所找到的汉藏对应词，在此基础上作了一些调整，把对应规则控制在和龚煌城所建立的对应规则相一致的范围内，以便排除可能是借贷或偶然对应的情况，然后对龚煌城等的汉藏对应词做了有阶分析。有阶分析就是把核心词分成最核心的词集和次核心的词集，即高阶词集和低阶词集，然后观察对应语素在不同词集中的分布。从已经观察到的材料看，语言接触产生的对应词是高阶比例低于低阶比例，即核心对应词呈外聚式有阶分布，语言分化保留的对应词是高阶比例高于低阶比例，即核心对应词呈内聚式有阶分布。如果严格控制语音对应规则，分化和接触的有阶对立是明显的。从龚煌城等的汉藏对应词表中我们发现，越是核心的词集，汉藏对应词比例越高。后来我自己又根据我所调查的云南中甸藏语做了一个初步的汉藏对应规则表，结果也是越核心的词集汉藏

对应比例越高（陈保亚,1993）。根据我们当时的田野调查结果，这种对应词的分布关系是同源语言的分布关系。由于龚煌城等的对应规则在声母上还留下一些问题，声调所赖以产生的条件也没有加以考虑，所以龚煌城等的汉藏对应词经过我们的筛选后仍然有可能有些是后来汉藏接触借用的结果，如果有办法剔除这些借词，汉藏对应词最核心部分的比例会更高。我初步认为龚煌城等所给出的汉藏对应语素应该是汉藏同源分化的结果。由于藏语是藏缅语中最有代表性的语言之一，如果汉语和藏语有同源关系，根据同源关系的可传递性，则汉语和藏缅语也就应该有同源关系。后来国内有的学者陆续找出的汉藏对应语素，其对应规则和龚煌城等建立的对应规则一致的部分，也应该是汉语和藏语同源分化的结果。

由于当时核心词分阶的检验只限于部分语言（陈保亚,1996），根据汉藏早期对应语素的内聚有阶分布来断定汉藏同源还只是一种相当初步的结论。随着近20年来我们对核心词集检验工作的广泛展开，有阶分布的可信度一直在增加。我们利用我们在 Swadesh 200 核心词基础上调整的两阶核心词集，对几十种民族语言和上百种汉语次方言专项调查的结果是：越核心的词集，同源保留率越高，借贷的比例越低。汉藏同源的可能性在增加。

除了汉语和藏语的语源比较，能否有其他藏缅语言的证据？

彝语是藏缅语中的重要一支，内部分歧大，显示出有久远的发展历史。如果汉语和彝语的语源关系能够得到充分研究，汉语和藏缅语的语源关系会有更多的讨论机会。

汪锋这部著作就是关于汉语、白语、彝语关系的研究。

汪锋从我这儿硕士毕业以后，去香港跟随王士元先生攻读博士，博士论文就是研究白语和汉语的关系。他先根据严格的语音对应重构原始白语，然后根据严格的语音对应，建立汉语和白语的对应语素，再找出最早层面的汉白对应语素，然后根据有阶分析，发现越是核心的词集，汉白对应词比例越高，这是语言同源的分布方式，于是汪锋断定最早时间层面的汉白对应语素是汉语和白语同源的结果，不是接触的结果。我认为他的方法严密，对应规则表全部给出，核心词的有阶分布情况也穷尽列出，具有可检验性。在汪锋以前，Greenberg (1953)、Starostin (1994a)、Benedict (1982)、Lee & Sagart (1998)、郑张尚芳(1999)等也主张白语和汉语有同源关系，但汪锋的这项研究比以前的汉白关系研究更具有实证性，也正是他的这项研究使我接受了汉白有同源关系的结论。

学界都认为白语和彝语有很密切的联系，不过证明汉语和白语有同源关系还不能证明白语和彝语的语源关系。白语和彝语的密切关系是接触的结果还是同源的结果，当时并不清楚。李方桂 (Li 1937)、闻宥 (1940)、赵衍荪 (1982)、戴庆厦 (1994)、Matisoff (2000a) 都认为白语和彝语有同源关系，但都没有给出系统的对应规则和核心词的分布情况。我认为汪锋关于汉语和白语关系的研究再往深处走一定要追问彝语处于一个什么地位。2004年，汪锋重返北大跟我做博士后，开始了彝语和汉语、白语的比较研究，并申请了国家社科基金，出博士后站留校任教，同时跟随戴庆厦先生做博士后，继续补充这项研究。后来又和

我一起主持教育部基地重大项目“基于严格语音对应同源语素的汉藏语系数理谱系分类”和四川凉山州资助项目“彝语文本解读和华夏文明起源研究”，继续完善汉白彝的比较工作。可见汪锋的这项成果是多年连续研究的结果。

从前面汉藏语比较的讨论看，要确定语言的亲属关系，有两个前后相继的难关不能绕过。这两个难关是：

- 1. 全面比较两个语言的语素，建立系统的语音对应规则，确定最早时间层面的对应语素。
- 2. 判定最早时间层面的对应语素是借用的结果还是同源的结果。

面对这两大难关，汪锋多次深入彝族地区调查。他先建立大规模语音对应数据库，根据几个大的彝语方言构拟了原始彝语，又在彝语、白语、汉语之间建立最早时间层面的对应规则，最后通过有阶分析发现，彝语和白语、汉语之间越是核心的词集，最早时间层面的对应词比例越高。这就是说，不仅彝语和白语有同源关系，彝语和汉语也有同源关系。下面是汉语和彝语最早时间层面对应词在两阶核心词集中的一种分布：

词阶	数目	实例
高阶 (第 100 词)	6+15	吃、蛇、犄角、二、树、死、 飞、耳、鱼、石、手、来、女、眼睛、黑、你、 我、长、肝、虱子、新
低阶 (第 200 词)	3+4	母亲、五、洗、 蛇、打、近、天(day)

当然，由于汉语和彝语分化时间较长，最早时间层面的对应语素相对较少，落在两阶核心词集中的词更少（以上共 28 个），要求所有的汉彝对应语素满足声、韵、调的完全对应是不可能的，因此汉语和彝语最早时间层面的对应也可能有少数是偶然对应或借贷的结果，核心词集中也不排除有偶然对应或借贷。不过，正如我们前面讨论汉语和藏语关系时提到的，从核心对应词中排除偶然对应，最核心的对应词比例高这一性质一般不会变化，而从核心对应词中排除借贷，最核心的词的比例会升高，因此在更深入的研究基础上如果能进一步排除偶然对应和借贷，汉彝同源会得到更进一步的确认。

另外，汉语、白语、彝语之间对应语素有阶分布呈现出的结果，还构成证据链的相互支持。不仅汉语和彝语的最早时间对应词呈现出同源分布，汉语和白语以及白语和彝语最早时间对应词也呈现出同源分布，由于汉语和白语同源，白语和彝语同源，根据同源关系的可传递性，汉语和彝语也应该同源。这些证据链没有出现矛盾，相互支持，这种结果不是偶然的。

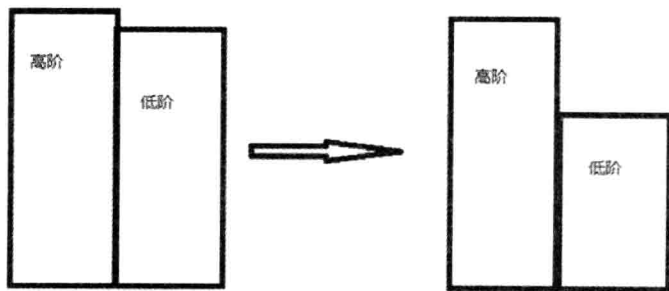
汪锋的这项研究在严格的对应规则上找出早期汉彝对应语素，再从核心词的分布上判

定这些对应的语源性质，直接面对确定亲属关系的两道难关，所得出的结论具有实证性。

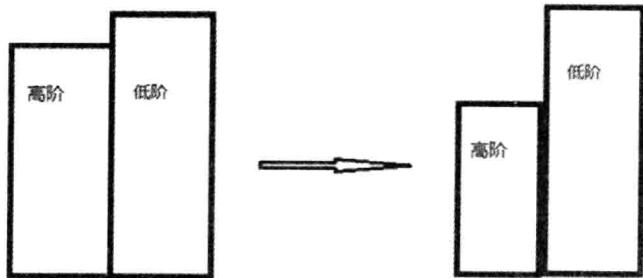
如果汉彝同源，而彝语又属于藏缅语，根据同源关系的可传递性，汉语和藏缅语同源。汪锋的这项研究再次为汉语和藏缅语同源提供了重要证据。考虑到前面提到的汉语和藏语同源的证据，现在我们可以比较有把握地说，汉语和藏缅语同源。

如果汉语和藏缅语同源，汉语的宾语语序就需要加以解释。藏缅语是“主语+宾语+述语”(SOV)型，汉语是“主语+述语+宾语”(SVO)型。原始汉藏语应该是相同的语序。有两种可能。一种可能是：原始汉藏语是SVO，后来藏缅语变成SOV，原始汉语在一定条件下变成SOV，如否定句中的代词宾语等，这些有条件的SOV又变回到SVO。还有一种可能是：原始汉藏语是SOV，其中一支即原始汉语在中原和其他语言的接触下变成SVO型，形成汉语今天的SVO面貌。古汉语宾语在一定条件下呈现出SOV的格式，过去多用焦点、强调等语用的原因加以解释，现在看来历史变化的解释也是一种可能。

回到一个方法论的问题上来。有阶分布分析还需要完善的地方，已经有不少学者提出这个问题，那就是如何使高阶词集和低阶词集更能反映语言分化和语言接触的对立。黄布凡、郑张尚芳、黄行、江荻等已经陆续提出了新的词集，都有一定的经验基础。目前调整两阶词集最迫切的工作是核心词在语言间传递情况的大规模专项调查。最近我们正在做一个基于大规模语音对应语素库的核心词调整，希望通过大量的调查和统计，使保留率高的词进入高阶词集，保留率低的词进入低阶词集，调整的结果是使得两阶核心词更能反映接触和同源的对立：



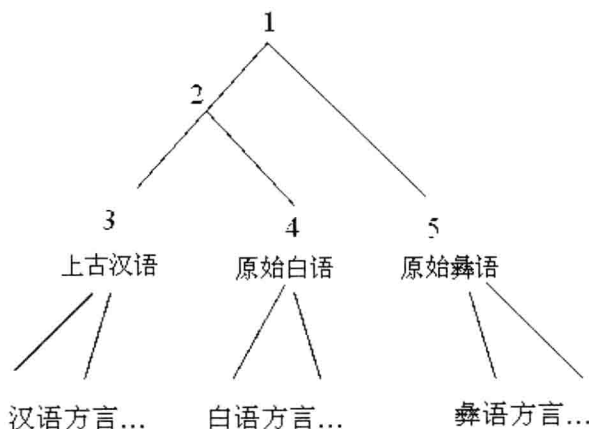
语言同源核心词分布方式



语言接触核心词分布方式

随着词阶准确性的提高，判定最早时间层面的对应语素的语源性质就会更准确。

再来看汉白彝谱系关系或亲缘关系的远近。汪锋最后给出的谱系关系是：



这个分类结果有些出乎意料，和目前把白语归入藏缅语族的结论不一致。

早些时候，徐琳、赵衍荪等学者根据白语的一些类型特征，比如有元音的松紧对立、闭音节少、复元音少等，再参考部分白语和彝语的共同词汇，把白语归入彝语支。这种分类代表了多数学者的意见。与此不同，Starostin 和 Benedict 认为白语是一种古汉语，后来郑张尚芳将白语和汉语作为一个汉白语族，理由是白语中含有大量的古汉语基本词汇。

对同源语言作谱系分类可以从语音、语法特征的共同性入手，这种方法可以称为结构特征分类法。但在我们所进行的具体分类工作中，这种分类方法实际操作起来会遇到很多问题，因为两个亲属语言分化以后，其语音、语法可以相互扩散，尤其是可以通过母语干扰形成同构，所以现代亲属语言的共同特征并不一定是两个亲属语言分化以前的共同特征。共同的语音、语法特征有可能是按照语言演变的普遍趋势发展出来的，也可能是亲属语言相互接触形成的，还可能是亲属语言共同受其他语言影响形成的，因此语音、语法特征常常会形成交叉现象，比如根据数量词的位置，仡佬语和侗语相近，数量词都在名词前，但根据动物性别词的位置，仡佬语和傣语相近，性别词都在动物名词后，拿这两个标准中的任何一个标准来确定仡佬语和侗语、傣语亲属关系的远近都会有争议。我们也尝试过这样的做法：列出存在于这三个语言中的任何两个语言之间的结构特征，然后比较哪两个语言享有更多的特征。但是两个语言享有的结构特征多并不总能证明亲属关系近，仍然可能是接触的结果，比如金沙江地区的傣语深受汉语影响，很多结构特征和深受汉语影响的侗水语一致，如果因此把金沙江傣语作为侗水语的下属语言，就不能反映金沙江傣语的历史。

我们也考虑过用历史上一些重大的历史音变规律来确定亲属关系的远近。比如汉语的浊音清化、平分阴阳、入派三声等重大音变规律，就有人用来作为汉语方言历史分类的依据。但是，重大音变规律在当代亲属语言或方言中的体现仍然是结构特征，可能通过接触

扩散,所以仍然不能解决用结构特征确定亲属关系远近所遇到的困难。另一方面,语音特征是错综复杂的,根据入声字归阳平的特点,可以把云南话归入汉语北方方言西南官话,根据是否区分平卷舌的特点,又可以把云南话归入汉语北方方言北方官话或江淮官话。最近也有人尝试用共同创新来作谱系分类的标准,问题在于共同创新也可以通过接触扩散,如何鉴定两个语言的某个共同特征是共同创新的结果而不是扩散的结果,目前还没有严格的判定标准。比如云南话入声归阳平是西南官话共同创新的结果,还是扩散的结果,现在还难以判定。

由于结构特征在划分谱系树、确定亲属关系远近时会遇到上述困难,用结构特征进行谱系树分类的做法目前还没有可行性。当然,由于共同存古和共同创新所导致的同构现象对认识语言历史有重要价值,不过这个价值可能不在谱系分类中。

汪锋处理汉、白、彝谱系关系所用的方法是对 100 核心词进行聚敛分类的方法。这一方法的基本思想是,两个同源语言 100 核心词中同源词数量越多,它们的亲缘关系就越近。Swadesh (1955) 首先用 100 核心词来确定同源语言的分化年代。王士元 (1995) 首先用生物谱系分类中的聚敛分析方法,根据陈保亚 (1996) 给出的侗台语 100 核心词保留表,画出了侗台语谱系树。当时还不清楚这种划分是否能真正反映同源语言亲缘关系的远近,因为有两个问题还没有得到解决:

1. 如果 100 核心词的衰变率在不同的亲属语言中差异较大,100 核心词中同源词的保留率就不能反映同源语言亲缘关系的远近。

2. 如果核心词也容易借用,聚敛分类的结果就不一定完全反映亲缘关系的远近,在某些方面反映的可能只是接触的结果。

关于第 1 个问题,徐通锵 (1991) 将汉语方言 100 核心词中同源词的保留率和文献中所记录的汉语方言分化年代做了比较,同源词保留率基本能反映方言的分化顺序。我和木霁弘曾经研究过小乘佛教传入傣族地区的年代,也分析过德宏、西双版纳傣语和未受佛教影响的金沙江傣语 100 核心词中同源词的保留率,该保留率也基本能反映金沙江傣语从原始傣语分化出来的年代。因此我们认为不同的同源语言 100 核心词中,同源词的衰变率基本上是一致的。

再来看第 2 个问题。为了弄清核心词借用率,我们对西南官话和台语的接触进行了追踪调查,我们发现在保持语音对应的前提下,100 核心词作为自由语素借用的情况相当少。比如,在各地傣语方言中,除了元江傣语有一个“杀”字是西南官话词,西双版纳、德宏、临沧、金平、马关、孟连、武定、元阳、元江、绿春、禄劝等地的傣语均未发现 100 核心词中有西南官话借词,只有少数语素是以黏着方式借入的,如“星期一、星期二”中的“一”和“二”。在我们所调查分析的几十种侗台语或侗台语方言中,100 核心词中借入西南官话的词也很少。1990 年,我们曾对滇藏川三角地带的十几种藏缅语族语言进行过核心词调查,

没有发现西南官话借词。1999年以来，我们又在云南、西藏对藏缅语几十种语言的关系词作了专门调查，均无西南官话词借入藏缅语核心词。西南官话在侗台语、藏缅语地区有很大的势力，很多民族都会说西南官话，从民族语言到西南官话的母语转用也频繁发生。有些地区还有汉族说民族语言，使当地民族语言的结构面貌发生了很大的变化。即使在这样一种深刻接触的背景下，100核心词仍然很稳定。

由于100核心词的借用率很低，并且衰变率基本上是均匀的，用100核心词中同源词比例的高低来确定同源语言亲属关系的远近比其他方法似乎更可靠。当然，100核心词集也有需要改进的地方，这是目前我们正在做的一项工作。至于Swadesh的200核心词，其保留率稳定程度不如100核心词。在我们所做的调查中，200核心词都有不同程度的借用，不同地区差别比较大，所以现在不宜用200核心词中关系词的比例作为亲属关系远近的根据。基本词汇也不太合适。现在有的学者提出了新的核心词词集，这样的词集若能够得到田野工作的检验，一定会有可利用的价值。

回到白语、汉语、彝语的谱系关系上来。在100核心语素中，由于白语和上古汉语的核心对应语素有39个（汪锋），数量很大，远远超过白语和彝语的核心对应语素数量（最严格对应要求下只有4个），白语和上古汉语（原始汉语）的亲属关系要近一些，和彝语亲属关系要远一些。

根据前面提到的理由，我同意汪锋的这种划分以及他所划分的理由。现在看来，Starostin和Benedict认为白语是一种古汉语的说法，郑张尚芳认为存在一个汉白语族的说法，都有一定的道理，只是证据不够严密、充分，比如，根据汪锋所建立的对应规则表，郑张尚芳所依据的汉白共同词汇，其中有大部分是后来的借词，需要从同源词表中排除。至于说白语是汉语方言还是存在一个汉白语族，尚可进一步讨论。

那么如何解释白语和彝语有不少同构现象？从上面的谱系树看，白语很可能是汉语受彝语强烈影响而形成的汉语方言。目前在连接云南和四川的两条主要古道有零关道和五尺道，这两条通道网络上分布有不少汉墓，说明汉语进入云南不会晚于汉代。而这两条重要通道网络上分布有大量彝族，汉语作为一种强势语言，在语言接触的过程中从结构上受到彝语的影响是可能的。

汉语和彝语先分化、后接触的细节还有待深入研究。彝语有自源文字，有相当丰富的语言文本，包括古彝文书写的文本和口语文本，这些文本中有大量关于彝族起源、分化、迁移、祖先崇拜的记录，这在其他藏缅民族中不多见。由于肯定了彝族和汉族的同源关系，这些彝语文本对进一步解释汉藏语的起源及其来龙去脉有了新的价值。若再结合三星堆及巴蜀考古发现的成果，这方面的研究应该很有前景。最近我和汪锋正做这方面的工作，希望能有一些实质性的结果。

无论将来的研究结果如何，从汪锋对白语与汉语以及白语与彝语所做的扎实而严密的

比较研究和统计数据看，我们可以比较有信心地说，白语和汉语、彝语都有亲属关系，并且在白语、汉语、藏缅语之间，白语和汉语在语源上有更直接的亲属关系。早些时候学者们关于白语和汉语同源、白语和彝语同源的争论，在汪锋的研究工作中都得到了协调，原来这两种观点并不矛盾，这里的关键是，汉语、白语、彝语都同源。汪锋的工作给了我们一个汉、白、彝语源关系的更为细致的蓝图。

汪锋的这项研究成果是建立在解决两道难关的基础上的，即建立严格的对应规则，并给出核心词的有阶分布结果。正是在这种意义上，汪锋的这项研究成果为汉语和藏缅语同源提供了又一个具有实证性的重要证据。

2011年10月2日于北街家园静山斋

内容摘要

本书以汉藏语言中的汉语、白语和彝语为对象,探讨汉藏语言比较的方法。汉语、白语和彝语三者之间纷繁复杂的历史关系是汉藏系语言关系的典型代表,要研究三者的关系,必须考虑三者之间横向传递和纵向传递交织的复杂背景。为此,我们坚持以下严格的步骤来进行三个语言之间的两两比较:1. 在内部比较的基础上重构原始语言;2. 两两比较各原始语言,构建关系语素数据库;3. 区分关系语素的层次,尽量离析出最早层次;4. 根据不可释原则、词阶法等推断最早层次的关系语素体现的语源性质;5. 如果三者之间是同源关系,则根据独特的共享创新、共同核心语素的比例等进一步比较三者之间亲缘关系的远近。

本书主要重构了原始彝语。根据《彝语简志》的划分,彝语至少有六个方言群:北部方言、东部方言、南部方言、西部方言、东南部方言和中部方言。我们从每一个方言群中选择一个代表点,分别是喜德彝语、武定彝语、江城彝语、巍山彝语、石林撒尼彝语、南华彝语。在比较这些彝语方言点的基础上,我们重构了原始彝语的声韵调系统和词汇。同时,根据创新特征和核心词汇同源保留率分别探讨了这六个方言的谱系树分化图景。我们对原始彝语词汇进行了分级,不同的级别可以显示其时间深度和证据支持的强度:在所有的彝语方言材料中都有分布支持的原始彝语形式几乎可以肯定是从原始彝语中遗传下来的成分,在研究中标为1级;2级是指在彝语东支(撒尼、武定)和西支(巍山、南华、江城、喜德)都有分布。

原始白语的材料主要根据Wang(2004; 2006b),在本研究中根据最新调查作了少许修订。上古汉语韵部研究相对成熟,在王力三十韵部系统(1937)建立之后,基本没有大的变动,本文采用李方桂(1971)的系统,其韵部基本与王力系统一致;至于声母系统的研究,李方桂系统后,龚煌城(2002)根据新的材料,做出一些修订,因此本文也随之更新。在认定语言之间的关系语素时,只有声、韵、调三方面都构成语音对应,才列入关系语素集。两两比较原始白语、原始彝语和上古汉语,分别得出汉白关系语素集、彝白关系语素集、彝汉关系语素集。我们分析了各语素集中体现的一对多的对应,将能判定为借用造成的晚近层次排除,剩余的关系语素集作为判定三者语源关系的依据。根据不可释原则和词阶法,三个语言都是同源关系。因此,三者的亲缘关系问题可以具体化为同源语言间亲缘关系远近的问题,也就是亲缘语言分群(Subgrouping)的问题。

我们同时运用两方面的证据来探讨汉、白、彝三者之间的远近:一是比较共享的核心同源语素;二是比较共享的创新。从基本词汇、语音、语义和语法各方面来看,原始白语

都和上古汉语关系更近，三者的关系应该写作：((原始白语，上古汉语)原始彝语)。

在原始白语和原始彝语的基础上，通过原始语言分级，探讨不同层级的白彝关系语素，并在最严格的基础上，引入汉语，分析三者之间的关系语素及其中的分布，探讨其中体现的语源判断方法与材料之间的互动。

在方法论方面，本研究还探讨了重构形式分级在语言比较中的必要性，分析了在系统匹配概念下层次对应的形成机制，以及词汇同质性对于层次校准的重要参考价值。根据白语方言和彝语方言的数据对词阶法进行了检验，此外，利用重构形式的不同级别和完全对应的不同情况，观察词阶在白彝关系语素集和彝汉关系语素集中的反映，发现重构形式的放宽会导致更多晚近借用成分的干扰。本研究还探讨了语言亲缘分类的特征法和词汇统计学对语言亲缘结构的不同表述。

关键词：原始白语，原始彝语，上古汉语，语言比较，词阶

Abstract

In the arguments for the evolution of the Bai language, the two frequently involved languages are Chinese and Yi. The crucial point to solve the controversies on the Bai language is to find out the genetic structure between the three. Due to the complex interactions of horizontal transmission and vertical transmission between the three, the Distillation Method proposed in Wang (2004) has been implemented in this case.

The Proto-Yi is reconstructed based on six Yi dialects in China, which are representatives of the six Yi subgroups according to the classification in Yiyu Jianzhi. They are Xide Yi, Wuding Yi, Jiangcheng Yi, Weishan Yi, Shilin Yi and Nanhua Yi. Both character-based method and lexicostatistics are used to draw the genetic tree of the six dialects. Meanwhile, the proto-forms are ranked to show their time depth and how strong they are supported. Proto-forms reflected in all these Yi dialects should be inherited from Proto-Yi, which are marked as rank 1. Proto-forms reflected both in eastern branch (Shilin, Wuding) and western branch (Weishan, Nanhua, Jiangcheng, Xide) are marked as rank2, which are not so sure to be inherited from Proto-Yi.

The data of Proto-Bai are extracted from Wang (2004; 2006b) with some modifications based on recent fieldwork. The reconstruction of Old Chinese follows Li (1971) with partial modifications in Gong (2002). Only forms corresponding in initial, final and tone between languages will be counted as related forms. In this spirit, the Sino-Bai, Bai-Yi, Sino-Yi related forms are collected. After stratification of these related forms, the oldest layers are distilled as the evidence to identify their genetic status. According to the Inexplicability Principle and Rank Analysis, the three languages are recognized to be genetically related. Therefore, the following question is how to subgroup them.

Two kinds of evidence are used to find out the genetic distance between Bai, Chinese and Yi. Firstly, the shared innovations between them are compared. Secondly, the shared kernel cognates between them are counted. All the evidence suggests the genetic structure to be ((Proto-Bai, Old Chinese), Proto-Yi).

In this study, some methodological issues in comparative studies are explored. The necessity of ranking proto-forms is discussed. In terms of systematic match, the mechanism of the formation of strata is analyzed and the homogeneity of lexicon is argued to be importance reference to identify the same layer. The rank theory is examined based on data from Bai and Yi