

● 中国的语言接触与语言关系研究丛书 ●

总主编◎戴庆厦

汪 锋◎著

语言接触与语言比较 ——以白语为例

YUYAN JIECHU YU YUYAN BIJIAO



商务印书馆
THE COMMERCIAL PRESS

语言接触与语言比较

——以白语为例

Language Contact and Language Comparison

——The Case of Bai

汪 锋 著

Wang Feng



商务印书馆
The Commercial Press

2012 年 · 北京

图书在版编目(CIP)数据

语言接触与语言比较:以白语为例/汪锋著. —北京:商务印书馆, 2012

(语言接触研究丛书)

ISBN 978-7-100-07652-4

I. ①语… II. ①汪… III. ①汉语史②白语—语言史

IV. ①H1-09②H252-09

中国版本图书馆 CIP 数据核字(2011)第 017226 号

所有权利保留。

未经许可,不得以任何方式使用。

YǔYÁN JIÈCHÙ Yǔ YǔYÁN BǐJiào

语言接触与语言比较

——以白语为例

汪锋 著

商 务 印 书 馆 出 版

(北京王府井大街36号 邮政编码 100710)

商 务 印 书 馆 发 行

北京瑞古冠中印刷厂印刷

ISBN 978-7-100-07652-4

2012年3月第1版

开本 787×1092 1/16

2012年3月北京第1次印刷

印张 14½

定价: 36.00 元

知之为知之,不知为不知,是知也。

——孔子《论语》

We can imagine that this complicated array of moving things which constitutes “the world” is something like a great chess game being played by the gods, and we are observers of the game. We do not know what the rules of the game are; all we are allowed to do is to watch the playing. Of course, if we watch long enough, we may eventually catch on to a few of the rules.

——R. P. Feynman 1963:2-1

术语简称

简写	中文全称	对应英文
C	辅音	Consonant
Demo.	指示	Demonstrative
DS	大石白语	Dashi Bai
EG	俄嘎白语	Ega Bai
EPB	早期古白语	Early Proto-Bai
EQ	恩棋白语	Enqi Bai
G	滑音	Glide
GX	共兴白语	Gongxing Bai
JM	金满白语	Jinman Bai
JX	金星白语	Jinxing Bai
MC	中古汉语	Middle Chinese
MN	定中结构	Modifier + Noun
MZL	马者龙白语	Mazhelong Bai
Neg	否定标记	Negative marker
NNum	数量结构	Noun + Numeral + Classifier
OC	上古汉语	Old Chinese
OV	宾动结构	Object + Verb
PAUP	软件名	Phylogenetic Analysis Using Parsimony (Swofford 1993)
PB	古白语	Proto-Bai
PENNY	软件语	A Parsimony Algorithm (Felsenstein, J. 1986-1993)
PL	古彝语	Proto-Loloish
P. P.	人称代词	Personal Pronoun
Pron.	代词	Pronoun
PST	原始汉藏语	Proto-Sino-Tibetan
RelN	关系句	Relative + Noun
Res.	敬称	Respectful address
ST	汉藏语	Sino-Tibetan

SVO	主动宾	Subject + Verb + Object
SWM	西南官话	South-western Mandarin
TB	藏缅语	Tibeto-Burman
TL	妥洛白语	Tuoluo Bai
V	元音	Vowel
WT	藏文	Written Tibetan
XS	西山白语	Xishan Bai
ZC	周城白语	Zhoucheng Bai
<>	对应	corresponding to
→	演变为	change into
>	演变为(仅用于连接同源词)	change into (only for two states of a cognate)

其他说明:

- * 星号标记重构形式
- * * 双星号标记重构的更早形式
- 黑体** 黑体突出的音为不规则变化
- 阴影** 阴影突出的音为尚待深入调查研究之处
- { } 附录中用以标记晚期借用形式

序

陈保亚

1997年下半年,我给北大中文系4年级同学开理论语言学课。课堂上我注意到教室最后一排有一位同学容貌消瘦,在凳子上挪来挪去,看上去被问题折磨得坐立不安,时常嘀咕出或冒出一阵不同的意见和想法,并且常常能打中要害。这位同学就是汪锋。面临同样的知识结构,汪锋最大的长处就是有良好的判断力,这是做科学研究难得的资质。这就好比两个数学家面临相同的已知条件,一个能得出正确的结论而另一个不能或只能得出错误的结论。后来我指导了他的本科学位论文、硕士学位论文。他的兴趣在语言演变,后来就推荐他去香港王士元教授那里读博士,研究历史比较语言学。获得博士学位后又回到北大先跟我做了两年博士后,然后又去中央民族大学跟戴庆厦教授做了两年博士后,最后留在北大工作。

汪锋对白语和汉语关系的兴趣始于2002年。那年云南大学木霁弘教授邀请王士元教授和我去考察茶马古道上的语言和文化。茶马古道是以滇藏川三角地带为中心,跨越世界屋脊带喜马拉雅山脉、帕米尔山系、横断山脉和青藏高原的远征古道,也是沟通欧亚大陆的重要古道网络。茶马古道上的民族语言包括了几乎整个汉藏语系和南亚语系的全部语言种类和南岛语系的很多语言种类。研究茶马古道以及古道上的语言接触、汉藏语系的起源、汉藏民族在横断山的移动过程、各种文化模式的形成过程,都具有很高的价值。考虑到汪锋的兴趣和已经具备的良好知识结构,我建议木霁弘教授也邀请汪锋以王士元教授助手的身份参加这次考察。汪锋在这次考察过程中的最大收获是看准了方向,确定了自己博士论文的选题:白语和汉语关系研究。

白语中既有大量汉语的成分,也有大量藏缅语的成分,另外还有一些其他语系语言的成分。和侗台语的情况类似,白语的系属问题成为汉藏语中的另一个难点问题。只有解决了这类的难点问题,历史比较语言学的方法论才能得到合理的评价。所以研究白汉关系是极具难度而又极有学术价值的选题。

要确定语言的亲属关系,有两个前后相继的关口是不能绕过的。一个是要确定最早时间层面的对应语素,另一个是要确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。要确定汉语和白语的关系,也必须越过这两道关口。在汪锋以前,有不少学者对白语的系属问题进行过深入研究,积累了很多材料和研究经验,但是,由于此前对语言接触规律性认识不够,这两道关口没有从根本上被越过。

要越过第一道关口,首先要建立系统的语音对应,然后在对应语素中排除借贷语素,确定

最早时间层面的对应语素。系统的对应首先应该是满足完全对应,即所比较的词在每个音段或音段组合上都应该得到对应。Lacouperie (1887)首先讨论过白语的系属问题,他认为白语和汉语、摩梭语、彝语、藏语、孟语都有一些相同的词,也有一些结构上的相似,他最后认为白语和孟语有同源关系,显然他的最终标准是语序,即白语和孟语在语序上有更多的一致性。语序标准是一种结构标准,就我们所收集和调查的情况看,语言接触到一定深度,会有结构上的传播。最典型的情况是母语干扰,比如傣族学习汉语时,会把傣语的一些结构带入汉语。当有些傣族村寨发生母语转移,即他们的母语不再是傣语而是汉语,这时的汉语就带有很多傣语结构。近几十年来,国内外大量的调查材料证明结构的横向传递广泛存在。过去语言学概论课程中讲语言结构不容易变化,也不容易被借用,那是从语言借贷的角度看语言接触。如果从母语干扰的角度看语言接触,结构的传递广泛存在。所以,要从结构类型的相似来确定系属关系是很困难的。近些年来仍然有学者从结构的角度来断定白语的系属关系,会遇到同样的问题。

Davies (1909)进一步提出白语和孟语有同源关系,他认为白语和孟语不仅有一些相同的词汇,而且白语和孟语在地域上不相邻。Davies 不用结构标准而用对应词,在方法上是一个进步。但是,Davies 建立的对应规则很不严格,明显带有偶然相似(见汪锋所引材料)。后来闻宥(1940)通过对更多材料的考察,确实证明 Davies 所找到的白语和孟语的共同词汇其实并没有关系。这从一个侧面说明,如果不建立完全对应,偶然对应的危险总是存在的。

闻宥寻找同源词比 Davies 更严格。最后他认为白语和藏缅语有一些最古老的词是同源词,主张白语和藏缅语有同源关系。

赵韬父(1949)找出了 500 多个词,认为是白语和汉语的同源词。截至汪锋这项研究以前,关于白语的系属研究基本上分成两派,一派认为白语和藏缅语有同源关系,这一派的代表有李方桂(Li 1937)、闻宥(1940)、赵衍荪(1982)、戴庆厦(1994)、Matisoff (2000a);另一派认为白语和汉语有同源关系,这一派代表有 Greenberg (1953)、Benedict (1982)、Starostin (1994a)、Lee & Sagart (1998)、郑张尚芳(1999)。两派都力图找出最古老的同源词,为什么会存在如此大的区别?让我们回到第一道难关上来,即如何解决语音对应问题。

我认为完备或系统的对应必须涉及下面四个相关问题:

- 1,对应的普遍性问题
- 2,对应的充分性问题
- 3,对应的完全性问题
- 4,对应的一致性问题

对应的普遍性是指,一个满足对应的语素在方言或亲属语言中分布越广,普遍对应程度就越高。从普遍对应角度看,一个在白语诸方言中分布很广的语素和一个分布很窄的语素,其比较性质是完全不同的。闻宥(1940)已经注意到白语中一义多音的现象,所以他尽量选择在白语多个方言中都有分布的读音形式进行比较,并且尽量选用老一辈人的读音形式。闻宥希望今后的比较应该采取原始白语的构拟形式。闻宥的历史比较已经涉及普遍对应问题,但是他

并没有在白语诸方言中事先建立对应规则,因此没有进行原始白语的构拟,对应的普遍性问题也无法观察到。后来的学者也都没有考虑普遍性问题,这是各家同源词表产生重大分歧的主要原因之一。汪锋在本项研究中做了这项工作,并且在构拟中有意识地考虑普遍对应情况。

对应的充分性要求充分比较亲属语言或方言中全部的语素,而不仅仅是找核心语素进行比较,这样对应规则、对应语素、音变条件才不会漏掉。而要完成白语各个方言间的充分对应,必须要充分调查各个方言,建立语素对应数据库。据我的初步调查,汪锋以前的白语研究者并没有做白语各方言语素对应数据库,这是对应语素或对应词数量差距很大的一个原因。汪锋在前人词表的基础上,多次深入茶马古道上的白族村寨,调查并核实材料,花大力气建立了白汉语素对应数据库。

完全对应要求在所比较的语素音形各个音段之间都形成对应。这样做的目的是最大限度地减少偶然对应。闻宥(1940)认为 Davies (1909)在白语和孟语之间找到的词基本上都属于偶然对应,现在看来就是因为没有建立完全对应。后来的白语研究学者在对应的完全性程度上要求是不一样的,这也导致对应词数量的不一致。从以往的研究看,由于没有充分对应做基础,反映完全对应的有序规则表并没有建立起来。汪锋首先建立了白汉语素对应数据库,反映完全对应的有序对应规则表也就能建立起来。

完全对应之所以重要,因为完全对应还有另一个作用是排除择词困难。我们知道亲属语言或方言分化时间越长,同源词的语义差别就越大。比如古汉语的“走”,本来指跑,现在指步行。Benedict (1939)已经注意到汉藏语比较中语义对当的复杂性,提出语音对应整齐而语义不同的同源词候选项要优于语义等同而语音不规则对应的候选项。这就是著名的语音对应优于语义对应原则。这个原则考虑到了语义对应的复杂性,但也降低了语义对应的实证性。比如,李方桂认为汉语中的“髑髅[* duk-lug]”对应台语中的* ?dl/r-D1L “骨头”(Li 1976, P235),即认为“头颅”和“骨头”有语义对当。前面提到古汉语的“跑”(走)和现代汉语的“走”语义对当,这是有大量文献证据的。怎样证明汉语“头颅”和台语“骨头”语义对当?这一直是一个未解决的问题,因为没有文献或其他证据。Benedict 所说的对应整齐往往只是声母或韵母或某个音段对应整齐,语义不对当的情况并没有得到解决。如果我们能够坚持完全对应,就有可能排除大量语义不对当的情况。

坚持完全对应还有更重要的理由。完全对应不仅可以最大限度地排除偶然对应,还进一步为一致对应提供了条件。一致对应是指满足相同对应规则的一套对应。不同的对应规则属于不同的一致对应,不同的一致对应或几套对应反映了对应的不同时空层次。如果没有完全对应有序规则表,就不可能很好地判定一致对应,就不能有效地区分对应层次。比如:

语素	北京	成都
院	yan51	uan214
院	yan51	yan214

如果只在声调和韵之间建立对应,不管韵头,就分不开两个层次。成都人很容易识别 uan214 是白读层,yan214 是文读层,根据的就是韵头的区别。

就以上北京话和成都话的关系看,在完全对应的基础上才有可能分析时空层次,这里的道理和原则是容易理解的,但是,涉及时空关系比较远的亲属语言时,这里的原则就没有得到遵守。

也有人可能会反对完全对应,认为这样的要求太严格了,太苛刻了。我认为,如果通过充分对应找出了大量的对应语素,就应该做完全对应的工作。事实上白语和汉语之间是存在大量对应语素的,有必要做完全对应的工作。建立完全对应的条件是语言之间还保留了足够的对应语素。当两个语言之间的对应语素太少,要建立完全对应就比较困难。这时我们就必须承认我们的材料有限,对应语素中存在偶然对应,判定层次存在困难,这时我们需要文献或其他材料来进一步分析不完全对应的性质,而不能贸然说某些不完全对应的语素是最古老的层次。

由于汪锋依次展开了普遍对应、充分对应、完全对应、一致对应的工作,所以他最终给出了汉白最早时间层次的对应语素。郑张尚芳认为汉语和白语 100 核心词全部是对应的,而且全部是同源词。根据汪锋的比较,郑张的词表中很多对应不是最早时间层次的对应。当然,由于后来的接触所产生的对应偶尔可能和早期对应规则一致,不排除汪锋所找出的最早对应层中有少数是接触的结果,但汪锋找出的汉白最早时间层是目前最严格的。由于汪锋给出了有序对应规则表,他的工作也是最透明的。当然,必须承认自闻宥以来很多学者在寻找同源词上所做的努力,他们的工作极具洞察力,为汪锋寻找语义对当语素起到了很大的引导作用。

以上是说汪锋在寻找汉白最早对应语素层上,在前人基础上高出一筹。我认为汪锋跨越了第一道难关。但汪锋并没有就此止步。前人在比较白语和汉语或白语和藏缅语时,先不论寻找对应规则是否严格,是否考虑了我们前面提到的几个步骤,总是把基本词汇或核心词中有对应的词都看成同源词,没有或很少考虑语言深度接触时基本词汇甚至核心词也会借用的情况。汪锋有长期田野调查的经验,对语言接触的深度有体会,所以他并不急于断言汉白最早时间层次一定是同源词,而是进一步从不可释原则和词阶分布两个方面论证白语的语源性质。他的认识论深度比以往的白语研究学者又上了一个台阶。

汪锋先从不可释原则出发来认识汉白最早对应语素层的语源性质。这是一种可能性确认思路,即通过一系列分析,排除借用的可能性,于是汉白最早层次是同源证据的可能性就进一步加强。接下来汪锋又进一步展开词阶分析,发现高阶(最核心的)比例高于低阶(次核心的),显示出同源语言分化的迹象,所以他认为他所找出的最早时间层次的对应是白语和汉语同源的证据。

汉白同源的最后的结论正确与否,需要进一步追问词阶法的有效性。

词阶法是我们 20 世纪 90 年代提出的判定同一时空对应层语源性质的一种方法,目前学术界还有不同的看法。词阶法根据接触和分化的有阶性把 Swadesh 核心词(Swadesh 称为基

本词)分成第 100 词集和第 200 词集两个不同的阶。通过对大量材料的分析统计,我们发现在语言接触和语言分化中,有语音对应的关系词的分布是对立的。在语言接触中,第 100 词集的关系词比例低于第 200 词集的关系词比例,在语言分化中情况正好相反。我们又进一步考察了汉语和侗台语之间一批对应比较严格的古代关系词的分布,发现第 100 词集的关系词远远低于第 200 词集的关系词,由此我们认为这一部分关系词不是侗台语和汉语有同源关系的证据,而是接触的证据(陈保亚 1996)。

丁邦新(2000)提出,如果把第 100 词和第 200 词中的词互换几个,就会影响统计的结果。我们对这个问题的回答可以概括成几个方面:

1. 统计样本要有较大的认同性,不能根据研究者的倾向有意选择符合研究者目的的样本。我们认为 Swadesh(1952)的 200 词是为了研究语言年代学设立的,没有分阶的思想,我们借用 Swadesh 的 200 词作样本,基本上能满足样本选择和研究目的相互独立这个条件。

2. 下一步就需要给 200 词分阶。Swadesh 提出 200 词是想通过词汇变化率的统计来确定语言分化的年代,最初他认为 200 词可以作为人类语言最稳定的核心词,但在实际工作中他发现 200 词也可以借用,他又从 200 词中筛选出 100 词(Swadesh 1955),认为 100 词是人类语言最稳定的核心词。Swadesh 的语言接触观念是有界的,他想在借用和不借用之间划出绝对的界线。在我们所做的接触研究中,100 核心词也可以借用。如果我们从无界有阶性的角度看,尽管 100 词也可以借用,但比起 200 词的借用率要小得多。由此出发,我们从 Swadesh 的 200 词减去 100 词,并经过一些调整(陈保亚 1994)就得到两个阶:第 100 核心词和第 200 核心词。尽管第 100 核心词不是绝对稳定的,但从概率上讲比第 200 核心词稳定。

3. 更重要的是,两阶词集核心程度差异已经经过一定的检验。在我们进行专项调查的几十种民族语言中,西南官话的借词总是第 100 核心词中的比例低于第 200 核心词中的比例。同样,在我们进行专项调查的几十种明确有亲属关系的民族语言或方言之间(比如傣语内部之间、壮语内部之间,等等),语音对应关系很容易建立起来,所保留的同源词总是第 100 核心词比例高于第 200 核心词比例。在汉语方言和汉语普通话之间,同源词总是第 100 核心词比例高于第 200 核心词比例。如果排除因为普通话的强烈影响带来的借词,汉语方言和汉语普通话之间第 100 核心词比例高于第 200 核心词比例的情况更显著。其实 Swadesh 从 200 词到 100 词的筛选工作本身也说明,在实际研究过程中,他发现 100 词更稳定。

第 100 核心词比第 200 核心词稳定是从整体上看的,这正是概率统计思想的核心。当然,正如丁邦新(2000)想到的,不排除第 100 核心词中少数词比第 200 核心词核心程度差。这正是我们现在正在做的一项工作,即尽可能使核心程度高的词进入高阶词,核心程度低的进入低阶词,这项工作做得好,有阶分析就会更准确。新的分阶词集必须经过田野考察的检验:在已知的同源词集中,高阶比例高于低阶比例,在已知的借词词集中,高阶比例低于低阶比例。但在新的经过检验的词集出现以前,我们认为第 100 核心词集和第 200 核心词集仍然是目前最有效的样本。

丁邦新(2000)提到的另一个问题是可比性问题。Swadesh 的 100 词表中有 bark(树皮)和 skin(皮肤)两个词项,并不是所有的语言都这样区分,有的语言可能只有一个。关于这个问题,我们有过讨论(陈保亚 1994, P186)。比如第一百词中“虱子”一词,侗台语普遍分为“衣虱”和“头虱”两词,这时的样本不再是 100,而是处理成 101($n+1$)。对应的计算最后是看比例而不是看词数。

概括地说,给 200 核心词分阶主要依据两个观察事实:

1. 在已知有同源关系的语言之间(文献材料或其他证据),第 100 词集的同源词高于第 200 词集的同源词。如日耳曼语族诸语言之间,台语诸语言之间,汉语北方方言诸次方言之间,傣语诸方言之间,同源词都满足这一条件。相反,在已知为接触关系的语言之间,第 100 词集的借词低于第 200 词集的借词。如傣语诸方言和汉语西南官话之间,核心词分布都满足这一条件。这可能不是偶然现象。

2. 在傣语和西南官话的接触中,构词能力越强的语素,在台语诸语言中分布越广的语素,越不容易被西南官话的借词取代(陈保亚 1996, P94-124)。台语第 100 词集中的词充当构词语素时,构词能力普遍比第 200 词集中的词要强,在台语诸语言中的分布也比第 200 词集中的词普遍要广。拿分布来说,在第 100 词中,分布在台语西南、中部、北部三个语群中的关系词占 61%;在第 200 词中,分布在台语三个语群中的关系词占 53%。这也可能不是偶然现象。

正是依据以上两个观察事实,我们把第 100 词集作为高阶核心词,把第 200 词集作为低阶核心词。我们始终认为,这种分阶的实质不在于这 200 核心词是否完全反映了汉语和侗台语的特点,而在于第 100 词和第 200 词这两个词集被作为分析样本时,能够反映同源词和借词在分布上的对立。

总之,有阶分析尽管还有需要进一步改进的地方,尤其是高阶词集和低阶词集的调整,但仍然有可行性。有阶分析给我们提供了一个观察核心语素分布的机会,展开有阶分析比不展开有阶分析多出一个证据,从这个意义上讲,汪锋的汉白关系研究比以前的研究可信度更强,证据更强。

有阶分析必须有一个前提,先确定对应语素的时空层次。孙宏开、江荻(2000)、丁邦新(2000)等不少学者都谈到关系词选择不同,会影响到有阶分析的结果,其实这正是我们一贯的主张,所以在这一点上我们没有矛盾。我们认为,严格的对应规则不是确定语源关系的充分条件,但却是确定语源关系的必要条件,要使有阶分析的结果可靠,必须要有严格的语音对应语素做基础,必须区分语音对应语素的时空层次(陈保亚 1996, P201-231)。这个道理就好比音位分析方法和记音的关系。如果语音都记错了,当然音位系统的归纳也会出现问题。这时我们不说是音位分析方法有问题,而是说记音有问题。当然,音位分析方法也有一些问题,但这些问题不是靠准确记音能够解决的。因此,记音不是归纳音位的充分条件,但却是归纳音位的必要条件。同样,在通过有阶分析确定对应语素的语源关系的时候,严格的对应规则是必要条件。对应规则越严格,有阶分析就越准确。目前有人用词阶法分析某些汉语方言,认为它们和

汉语没有同源关系,和我们的分析结果不一致,原因就在于这些学者并没有建立系统的语音对应,没有做或没有公布有序对应规则表,他们把早期对应和后来的接触对应混在一起了。同源关系的阶曲线是一种走向,接触关系的阶曲线是相反的走向,两种不同的阶曲线叠加在一起,同源关系是看不出来的。目前还有一种最显著的错误就是用核心语素来建立对应规则,这就违反了充分对应原则,少量的语素是不能建立对应规则的,更不能展开完全对应,也不能进一步确定对应语素的时空层次。合理的做法是先在全部语素中建立对应,区分层次,找出最早时间层次的对应,然后看有多少最早时间层次对应语素落在核心语素的两个阶中,这正是汪锋采取的方法。当我们拿汪锋所建立的汉白最早时间层和其他学者所建立的汉白同源词进行比较,可发现其他学者的很多同源词实际上是较晚期的借词。

由于汪锋建立了截至目前为止最为严格的汉白语对应关系语素,最早时间层面的对应得到了尽可能严格的限制,在此基础上再做分阶研究就有了条件,于是他跨越了第二道难关。所以我认为他得出的汉白同源的结论是成立的。他的这一结论支持赵韬父、Greenberg (1953)、Benedict (1982)、Starostin (1994a)、郑张尚芳(1999)汉白同源的假说,但证据更充分,论证更严格。汪锋对语源关系的鉴别有相当的信心,我认为他的信心有扎实材料和严密分析作依据。前面说过,不排除这个最早汉白层次中有少数对应语素是接触造成的,这些语素的对应规则和同源语素正好相同,无法分离出来,这会不会影响汉白同源的结论?我认为不会。因为接触对应是高阶比例低于低阶比例,汪锋所找出的汉白核心词阶曲线走向并没有受到影响,说明最早汉白层即使有接触因素,也不是很显著,并没有改变汉白最早对应层语素同源分布的主要趋势。

无论将来的研究结果如何,从汪锋对白语与汉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,白语和汉语有亲属关系。

陈保亚

2010年9月16日

目 录

术语简称	iii
序	陈保亚 1
1 语言接触与语言比较——以白语为例	1
1.1 还原比较法 (The Distillation Method)	1
1.2 为什么选白语 (Bai)	2
1.3 白族和白语	3
1.3.1 历史背景	3
1.3.2 白族的遗传系联	6
1.4 以往关于白语亲缘地位的研究	8
1.4.1 初始期	9
1.4.2 过渡期	11
1.4.3 现代期	13
1.4.4 语言亲缘关系的澄清	15
2 研究目标及方法	18
2.1 内部比较 (Intra-comparison)	18
2.1.1 重构 (Reconstruction)	18
2.1.2 亲缘分群 (Subgrouping)	22
2.1.3 年代测定 (Dating)	26
2.2 外部比较 (Inter-comparison)	27
2.2.1 什么是分层的基本单位	27
2.2.2 互补制约 (Complementarity Constraint)	29
2.2.3 层次识别	29
2.2.4 排除晚期层次的工作程序	32
2.3 语源鉴别 (Recognition)	32
2.3.1 不可释原则 (Inexplicability Principle)	32
2.3.2 词阶法 (Rank Theory)	33
2.4 还原比较法与其他历史语言学方法	33
3 内部比较 (Intra-comparison)	35

3.1	现代白语方言	35
3.2	重构原始白语(Proto-Bai)	45
3.2.1	音系重构	45
3.2.2	词句法重构(Morphosyntactic Reconstruction)	70
3.2.3	部分基本词汇重构	82
3.3	亲缘分群(Subgrouping)	88
3.4	年代测定(Dating)	94
4	外部比较(Inter-comparison)	97
4.1	汉白比较	97
4.2	汉白关系语素的最早层次	98
4.2.1	声调对应的最早层次	98
4.2.2	声母对应的最早层次	99
4.2.3	韵母对应的最早层次	111
5	语源鉴别(Recognition)	124
5.1	不可释原则及应用	124
5.2	词阶法及其应用	128
6	还原比较法结果的应用	130
6.1	了解上古汉语的新角度	130
6.1.1	上古汉语去声的两种类别	130
6.1.2	介音*-l-与中古汉语重组四等的来源	140
6.1.3	上古汉语部分韵部的重构	141
6.2	原始白语送气擦音的来源	145
7	余论	150
	附录1: 原始白语词汇	152
	附录2: 汉白关系语素	179
	附录3: 汉白关系词中核心词的层次	186
	参考文献	192
	图表索引	204
	表格索引	205
	后记	213

1 语言接触与语言比较——以白语为例

1.1 还原比较法 (The Distillation Method)

历史语言学的一个中心问题是如何区分横向传递与纵向传递。“祖语的特征是通过纵向传递延续下来的,而与其他语言接触造成的特征则是由横向传递带来的 (Cavalli-Sforza & Feldman 1981, Wang 1999a: 250-251)”。只有沿纵向传递而来的成分才能为语言的亲缘关系提供证据。如果上述两种特征在一种语言中混淆起来了,这个语言的亲缘线索也就会变得模糊不清。因此,在语言比较中,应尽早离析出横向传递而来的成分。

本文提出“还原比较法”来系统剔除语言中横向传递而来的成分,进而获得相对可靠的纵向传递成分以作为语言亲缘关系的证据。还原比较法的框架界定如下:

还原比较法 (The Distillation Method)

一、定义 (Definition):

通过以下步骤逐步实现向语言早期状态的还原:(1) 内部比较 (Intra-comparison); (2) 外部比较 (Inter-comparison); (3) 语源鉴别。

二、目标 (Effect):

剔除后起创新的成分以及横向传递而来的成分。

三、相关界定 (Corollaries):

(1) 内部比较 (Intra-comparison) 力图建立一个语言下众方言之间的语音对应,并在此基础上重构出该语言的原始形式。

(2) 外部比较 (Inter-comparison) 在互补制约 (Complementarity Constraint) 下建立原始语言之间的对应,并离析关系语素的层次,进一步还原到语言的早期状态。

(2a) 互补制约 (Complementarity Constraint): 如果所比较的原始语言之间的一些对应呈互补分布状态,这些对应就应该剔除出去,因为这种分布必定是横向传递造成的。

(3) 语源鉴别 (Recognition) 部分以“不可释原则 (Inexplicability Principle)”和陈保亚 (1996) 提出的词阶法 (Rank Theory) 来分析通过内部比较和外部比较两次还原得到的成分,鉴别它们源自遗传还是借贷。

(3a) 不可释原则 (Inexplicability Principle) 指无法通过施借语言 (Donor language) 的音韵系统来解释受借语言 (Recipient language) 中的对应表现形式。在没有其他证据的情况, 不可释的这些成分更像是遗传自共同祖语, 而不是源自横向借贷。

(3b) 词阶法 (Rank Theory): 如果在最后还原出的关系语素中, 落在高阶的比率高于低阶, 那么这些材料表明了遗传关系; 反之, 则显示借贷关系。

第二章将详细讨论还原比较法的理论基础和操作细节。

1.2 为什么选白语(Bai)

就语言亲缘关系的研究而言, 白语十分有挑战性。白语主要分布在云南省, 一些与之临近的民族(傈僳、彝、纳西等)也能讲白语。同样, 一些白族也可以讲临近的其他民族语言。而且很早以前白族就开始接受汉文化。许多白族都可以说很流利的汉语(常常是当地的西南官话)。因此白语发展出了一种“混合(mixed)”的面貌。一个世纪以来, 白语这种复杂的性质引起了广泛的关注, 许多学者都提出了他们各自的白语系属分类方案(参见 § 1.4)。但遗憾的是, 自 Lacouperie (1887) 以来, 一直没有人尝试以系统周密的比较研究来理清严重受语言接触影响的白语中的遗传成分。

在 § 1.1 中提出的还原比较法正可以用来研究白语这样有复杂接触史的语言。还原比较法的第一步可以通过白语方言之间的内部比较来揭示从原始白语到白语方言的纵向传递。这一步涉及的理论问题在历史语言学中已经有很坚实的基础, 包括三项基本任务: 1. 重构原始白语 (Proto-Bai); 2. 白语方言的亲缘分类 (subgrouping); 3. 白语演化中的一些相关时间测定(dating)。

还原比较法的第二步是通过外部比较来进一步分析原始白语中的成分, 也就是通过同影响过白语的汉藏语言(主要是汉语)进行比较, 剔除早期借入原始白语的成分。在严格对应的基础上重构原始语对排除借词十分有利, 但并不足以剔除所有阶段的借词。王士元(Wang 1996a: 253–255)已经指出, 大量的汉借词借入日语后, 就造成了与其汉语来源形式的系统对应。在白语中, 各个时代借入的汉语词汇造成了不同层次的语音对应, 因此, 区分原始白语与汉语之间的对应的时间层次就十分必要(详见 § 2.2)。根据音变的先后次序以及其他方面的一些知识, 可以分辨出一些晚近层次。不过, 或许仍有部分横向传递的成分在这一步还原之后仍能混入。

不可释原则 (Inexplicability Principle) 和词阶法 (Rank Theory) 可以用来鉴别还原后的汉白对应语素, 看它们是应该归因于共同祖先的遗传还是由于汉语的影响。

这样, 白语与汉语及其他汉藏语之间的比较不仅能澄清白语的演化历史, 还能揭示这些比较中涉及的语言的一些历时演变, 尤其是汉语的历史音韵演变。如果本研究可以实现这些目标, 就可以为汉语的演化研究开辟一个新的角度, 拓展汉语研究的材料范围。