

语义创新与方言的亲缘关系*

汪 锋¹ 王士元²

(1, 北京大学 北京 100871; 2, 香港中文大学)

提要 本文提出语义创新可以作为衡量亲缘关系(genetic relationship)的特征。根据基本词的语义创新,我们尝试以汉语方言为例探讨语言的亲缘关系,运用一个俭省(parsimonious)程序——PENNY,来进行汉语方言亲缘关系的量化分析。本文主要希望在研究方法上作一个新的探索。

关键词 语义创新 亲缘关系 俭省 量化

壹 语义创新和衡量亲缘关系的标准

在历史语言学中,亲缘分类(genetic classification)常用来表述一个语系的亲缘演化关系。普遍接受的亲缘分类标准是“独特的共享创新(unique shared innovation)”,这个标准背后的基本假设是这类创新不会在两个已经分离的语言群中独立地平行发展出来。因此,一组共享有此类创新的语言(或方言)可能来源于一个共同的祖先,在那个祖语状态,此类创新已经发生。

衡量语言亲缘关系时,一些干扰应该尽量清除。首先是语言借用造成的干扰,一旦遗传成分和借用成分混淆起来,语言之间的亲缘关联多少要模糊一些。其次,来自祖语的遗存应该排除,既然所有的子孙语言都继承了这些特征,遗存就不能为衡量亲缘关系远近提供什么信息。最后,值得注意的是,即使是创新,也有不同的权重。有些创新,从跨语言的角度来看,很普遍,比如说腭化,就很可能是语言各自独立的平行发展,也就不太适合用来计量语言的亲缘关系(Campbell 1998:115-122)。

在语言亲缘分类的实践中,语音创新最常用,其次是词法创新,句法创新。但语义创新较少用到^①。这可能是由于语义演变的序列比较不规则。如 Hock(1991:308)所说,“语义变化是‘模糊’的,非常不规则,并且极其难以预测。”^②然而,语义变化的这些特征正好符合上文提到的语言亲缘分类标准的要求。

Wilkins(1996)对语义变化的界定有详细的探讨。根据他的意见,语义变化的模式可以图

* 本研究得到香港城市大学[#9010001],香港 RGC[#9040781]和台湾中研院项目资助。台湾中研院语言研究所允许我们使用他们的汉语数据库。Laurent Sagart 教授和 James W. Minett 博士为本文提供了宝贵意见。郑锦全教授和张秀英女士教我们使用 CCLANG 软件。本文曾在香港城市大学语言工程实验室(2003.2.11)报告过,Christophe coupé,柯津云,Marco Caboaro,陈家豪,区靖邦等同事的讨论对本文多有启发。匿名审稿专家提供了许多宝贵的修改意见。谨此致谢。

① Laurent Sagart 教授提醒我们,Blust(1977),Burling(1983)和 Sagart(2001a, 2001b, 2002)曾以“词汇创新(lexical innovations)”来作为亲缘分类的依据。但本文中的“语义创新(semantic innovations)”限定得更加严格,与之有所不同。

② 本文英文引用的中文翻译均出自笔者,下同。



示如下：

时间(T): T1 T2 T3
 形式(F): F1 F1 F1
 意义(M): M1 → M1 & M2 → M2

图 1 语义变化图(改编自 Wilkins 1996)

图 1 显示了 F1 的语义演变,从 T1 时的 M1 转变到 T3 时的 M2。通常,在 T1 期间,意义 M2 与另外一个形式(F2)相连。这样,在 F1 取代 F2 获得意义 M2 之前,两个形式 F2 和 F1 之间的竞争要持续一段时间。例如,汉语的“树”从“种植”义演变到“树木”义,但“树木”这个意义却并不是创新,它早先由“木”这个字承担。同其它演变一样,“树”并不是突变,一下子在所有的语境中替换掉“木”来做为“树木”的表现形式,而是渐变。这一渐变过程可以仿照词汇扩散(Wang, 1969)的图示表达出来:(U=未变,V=变异,C=已变)

表 1 语义创新的扩散

	M1	M2	
T1	F1	F2	U
T2	F1	F1/F2	V
T3		F1	C

如果借用成分混入,情形就不同了,可以图示如下:

表 2 借用的扩散

	M1	
T1	F1	U
T2	F1/F2[借用]	V
T3	F2	C

比较表 1 和表 2,如果竞争的形式突然出现,而不是通过语义演变转化而来,就很可能源自借用。因而,仔细观察比较词汇形式的状态可能提供线索排除借用成分,而留下语义创新。换个角度说,语义变化属于语言的内部演变。

如果完整的语义变化过程可以从一个语言的历史纪录中获取,有两种方式来表述这个变化:(1)从 F1[M1] 到 F1[M2];(2)从 F2[M2]到 F1[M2]。无论哪种方式,F1[M2]都是变化中的创新成分。

尝试以汉语方言为对象、用语义创新来计量语言亲缘关系,有很多便利。汉语历时材料丰富,为获取近 3000 年来的语义演变提供了良好的基础。一般都承认,现代汉语方言来自一个共同的祖先——原始汉语。因此,从汉语历史文献中发现的语义变化对了解现代汉语方言的形成十分有用。如前文所述,语义变化“极其难以预测”(Hock 1991:308),两个语言群独立发展出相同的语义变化的几率就相当小。也就是说,共同的语义创新作为衡量语言亲缘关系的特征的价值就高。因此,语义创新在方言中的表现能很好地体现方言之间亲缘关系的远近。为了尽可能减少借用的干扰,并鉴于 Swadesh 的 100 基本词被接受为最稳固和最难被借用的成分(Swadesh 1955, Wang 2004),我们只考虑涉及这个词表的语义变化。

贰 汉语史中的语义创新

得益于汉语史研究的积累和汉语文献电子数据库^①的建立,一些词语相对完整的基本语义演变史可以建立起来。下面列出 10 个词条^②:

1 树^③

	“树”	“种植(v.)”	
- 211B. C. (战国)	木	树	U
206B. C. - 23 (西汉)	木/树 ^④	树	V
25 - 100 (东汉)	树/木	V	
100 - 220 (东汉)	树		C

2 狗

	“狗”	“小狗”	
- 25 (东汉前)	犬	狗	U
220 - 479 (魏晋, 刘宋)	狗/犬		V
618 - 907 (唐)	狗		C

3 肚子

	“肚子”		
25 (东汉前)	腹	肚 ^⑤	U
220 - 420 (魏晋)	腹/肚		V
420 - 479 (刘宋)	肚/腹		V
618 - 960 (唐, 五代)	肚		C

一些方言现在用“颌”来表达“脖子”的意思。从文献中可以知道“颌”曾有两个常见义项“下巴”和“点头”,以及一个不常见义项:下巴和喉头之间的部位。还没有文献材料说明“颌”有“脖子”的意思。我们推测:这个少见义项可以作为从“下巴”到“脖子”的转变过渡。为此,我们将方言中以“颌”表“脖子”的变化作为一项语义创新。

① 不特别注明的材料都是根据台湾中研院语言研究所的汉语文献数据库,主要包括:诸子(先秦),史记(西汉),论衡(东汉),世说新语(刘宋),六祖坛经(唐),游仙窟(唐),祖堂集(五代),朱子语类(宋),元刊杂剧三十种(元),金瓶梅(明),红楼梦(清),儿女英雄传(清)。

② 受汉语史研究的限制,我们并不清楚每一个有变化的基本词的语义演变脉络。正如我们在第壹节所说,如果我们不能确定其语义演变过程,就很难区分演变源自于内部还是外部借用,因此,我们只选取了目前视野范围内我们认为语义演变框架比较清楚的 10 个词条。

③ 参见汪维辉(2000:81-87)。

④ 更常用的列在斜线左边,括弧则表示很少用。

⑤ 在佛经译文中出现,主要用作贬义。

4 脖子^①

	“脖子”	“前脖”	“后脖”	
1121B. C. - 771B. C.	领			U
770B. C - 207B. C.		颈	项	V1
206B. C. - 25(战国 - 西汉)	颈/项			V1
25 - 618(汉 - 隋)	颈			C1
1271 - 1368(元)	颈/脖项/(项)			V2
1368 - (明清)	脖子/脖项/(颈)			C2

5 脚

	“脚”	“小腿”	“下肢”	
- 25(东汉前)	足	脚	足	U
25 - 618(东汉 - 隋)	脚/足	脚	脚	V
618 - 907(唐)	脚			C

6 眼睛^②

	“眼睛”	“眼球”	
- 475B. C. (战国前)	目		
206B. C. - 25(战国 - 西汉)	目	眼	U
25 - 220(东汉)	目/眼		V
220 - 420(魏晋)	眼		C

7 站

	“站”	“独自站立”	
960 - 1368(宋元)	立	站 ^③	U
1368 - 1644(明)	立/站		V
1616 - 1911(清)	站/立		C

① 管锡华(2000:207 - 217)讨论了“脖子”在汉代以前的情况,我们补出了余下部分。

② 材料来自汪维辉(2000:24 - 31)。

③ 《广韵》陷韵:“站,俗言独立”。

	“跑”			“走”	
-206B. C. (先秦)		走	U	行	U
420 - 907(刘宋 - 唐)	走	跑②	U	行/走	V
960 - 1368(宋元)	走		U	走/行	V
1368 - 1644(明)③	奔/跑		V	走	C
1616 - 1911(清)	跑		C	走	

9 听→闻

	“闻”	“知晓”	“听”	
-206B. C. (先秦)	嗅	闻④		U
206B. C. - 1644(汉 - 明)			闻	
1644 - 1911(清)	闻		闻	V
1911 -	闻			C

10 鼻子→闻

在客家话和闽语中有一个有趣的语义延伸,它们以“鼻”字来表示动词“闻”。在文献中,“鼻”字用作动词“闻”的例子还没有发现。这样,我们把这个语义延伸作为这些方言的创新。

从上述语义变化中,13个语义创新可以概括出来:1)木→树[“树”];2)目→眼[“眼睛”];3)犬→狗[“狗”];4)腹→肚[“肚子”];5)足→脚[“脚”];6)领→颈[“脖子”];7)脖[“脖子”];8)行→走[“走”];9)走→跑[“跑”];10)立→站[“站”];11)嗅→闻[“闻”];12)颌[“下巴”]→颌[“脖子”];13)鼻[“鼻子”]→鼻[“闻”]。注意1-11项创新和12-13项创新分别是按照第壹节提到的两种不同方式来界定的。

叁 以语义创新为据的方言亲缘关系的计量表述

通过俭省方法,我们可以把语义创新体现的语言关系转化为树形图。俭省方法的基本要点与Krishnamurti(1983)的精神相似。最俭省意味着在这个树图上需要设置的变化总数最少。但要找出一个最俭省的树图需要大量的计算。就19个方言而言,手工计算几乎不可能完成任务(参见Wang 1987, 1996)。因此,我们利用一个俭省方法程序——PENNY^⑤(Felsenstein 1986-1993),根据第2节总结的13项语义创新来计量汉语方言的亲缘关系。

① 参见蒋绍愚(2002)。

② 本义指动物以爪刨土。

③ 蒋绍愚(2002)把《元曲选》、《金瓶梅》和《西游记》作为明朝的材料,但不能肯定三者的先后顺序。但他确信“跑”这个字在《西游记》中已经替代了“奔”。

④ 参见洪成玉(1996)。

⑤ 就我们所知,William Baxter教授最早尝试引入这一程序到语言研究中。我们感谢他让我们参阅他尚未发表的论文Mandarin dialect phylogeny。

这个程序应用了一个算法叫 Camin-Sokal 俭省法,允许 0→1 的变化,但禁止相反的变化 1→0(参见 Camin and Sokal 1965)。我们以 0 代表遗存状态,而 1 代表创新状态。这个俭省法的蕴涵就是:我们假定遗存状态可以转化为创新状态,但创新状态不能变回遗存状态。至于多级状态的语义变化可以利用 Sokal and Sneath(1963)提出并由 Kluge and Farris(1969)发展的办法把它转化为一系列的(0,1)二级编码。例如,“脖子”义在文献中构成的演变一个三级状态链:领→颈→脖,可以线性编码为 0→1→2。0 是遗存状态,1 和 2 都是创新状态。可以将这个三级链改编为一个二级编码来符合 Camin-Sokal 俭省法的输入数据要求,如下所示:

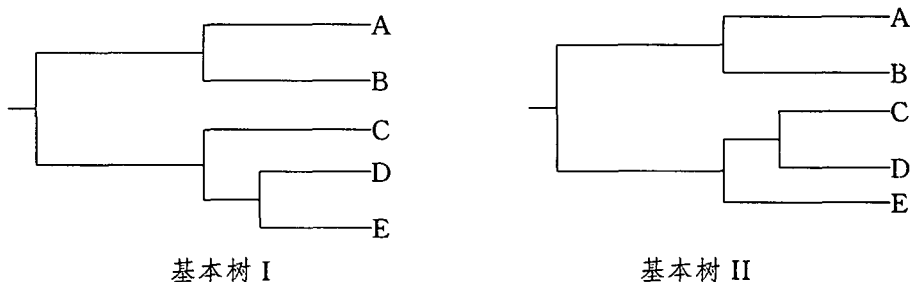
旧编码	新编码
0	00
1	10
2	11

语言中未知的状态编码为“?”。例如:福州“□pieŋ[‘跑’]”可以标为“?”,因为我们还不清楚它在语义演化链中的位置。按照上述编码方法,我们根据《汉语方言词汇》中的数据,将 13 项语义创新在 19 个现代汉语方言(各方言的代号参见附录 I)^① 的对应项编码为(0,1)的矩阵,作为 PENNY 程序的输入数据(参见附录 II)。

PENNY 程序可以找出适合输入数据的所有最俭省的树形图。比较理想的结果是树形图的数量不大。但根据上述语义创新数据运行 PENNY 程序得出了大量(21 360)的最俭省(17 个变化)的树形图。看来对于 19 个方言的二分结构而言,13 项特征的信息量不够充分。我们期待以后有新的材料增添进来,以更准确地澄清这些方言的亲缘结构。

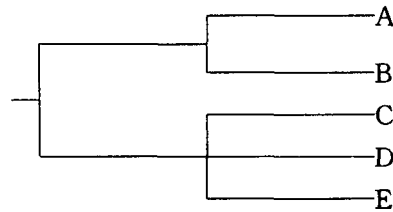
从附录 II 可看出,WH, HF 和 YZ 三个方言的编码完全相同,我们就选择 YZ 作为他们的代表,就可以去掉 WH 和 HF。同理, JN, XA, YJ 和 CZ 也可以被去掉。这样,再运行程序,就得到较少量(720)的最俭省(18 个变化)的树形图。

每一个最俭省的树形图叫做基本树(fundamental trees),可以绘制一个综合树(consensus tree)来概括 720 个基本树的结构。根据不同的目的,可以绘制不同的综合树(参见 Swofford 1993)。我们选择最严格综合树(strict consensus tree)来概括 720 个基本树体现的方言关系。在最严格综合树上,只有在所有基本树上都是单源的分叉才会被选择到。例如,这里有两个基本树, I 和 II, 如下图:



那么它们的最严格综合树 III 就如下图所示:

^① 《汉语方言词汇》中成都方言的数据同北京大学陈保亚教授提供的材料有相当大的差异,参见下列网址上的 BCD(Basic-words of Chinese Dialects)计划:<http://chinese.pku.edu.cn/wangf/wangf.htm>。我们觉得比较稳妥的做法是先暂时不考虑成都方言。



综合树 III

最严格综合树显示出对 A, B, C, D 和 E 关系的最小解决方案。在树 III 上, 两个语言群 (A, B) 和 (C, D, E) 可以建立起来。至于说对 (C, D, E) 的进一步分化状况, 树 III 表明没有可接受的方案。

根据 720 个基本树, 一个表现汉语方言关系的最严格综合树可以通过 Tree-Explorer^① 程序 (Tamura 1997 - 1999) 绘制出来:

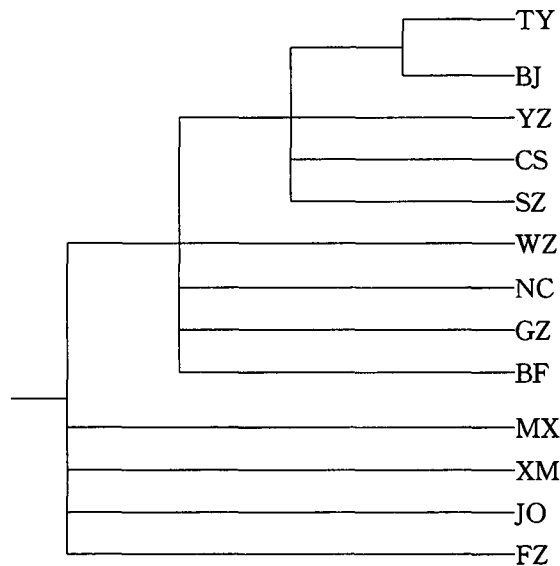


图 2 汉语方言关系的最严格综合树

图 2 中至少有两点对考虑汉语方言的亲缘关系具有启发性。一是太原 (TY) 在树形图上的地位, 与北京 (BJ) 构成姐妹方言关系, 而另一官话方言扬州 (YZ) 在它们外围。晋语地位问题长期以来众说纷纭 (参见李荣 1985, 丁邦新 1996, 王福堂 1998, 侯精一 2002)。同样应用最俭省法, Wang (1996) 选择了 4 个语音特征来绘制包括 BJ, CD, TY, YZ, SZ, GZ 和 XM 在内的 7 个汉语方言的树形图, 所有的最俭省的树形图都要求太原 (TY) 和扬州 (YZ) 相邻。换言之, 把晋语独立为一个大方言区的理由并不十分充分。图 2 的汉语方言关系的最严格综合树也指向同样的观点。图 2 上官话方言构成 [(TY, BJ), YZ] 的亲缘结构, 扬州 (YZ) 作为官话方言的地位是普遍接受的, 而它处在 (TY, BJ) 的外围。在这种情形下, 太原 (TY) 很难独立到官话之外。不过, 本计量所考察的词条有限, 晋语方言也只选取了太原一处, 因此, 这个结果并非对晋语分立问题的确定性答案。

另一个问题是闽方言 (FZ, JO 和 XM) 的关系。在这个综合树中, 这几个闽方言并没有构

^① 此程序可以通过如下网址下载: <http://evolgen.biol.metro-u.ac.jp/TE/TE-man.html>。



成一个一致对外的结构。历史上曾有三次汉人入闽的浪潮(参见周振鹤、游汝杰 1986, 游汝杰 2000)。第一次入闽大约在西汉末, 体现为福州(FZ)方言; 第二次入闽大约在东汉末, 主要分布在闽西北和沿海地区, 建瓯(JO)方言恰好符合; 第三次浪潮发生在隋唐期间, 造就了“闽南话”, 厦门(XM)可以作为代表。从这个角度看, 一个只有众闽方言共享的原始闽语是否存在, 还可以再加考察。

如果图 2 揭示的汉语方言的亲缘关系图是正确的话, 它可能对传统的认识提出了一些疑问, 比如, 作为湘语的长沙(CS)和双峰(SF); 作为吴语的苏州(SZ)和温州(WZ)。在图 2 上, CS 和 SZ 更近于官话, 而不是分别与 SF 和 WZ 更近。

肆 余论

可能由于词汇标准容易应用, 它们常常充当汉语方言亲缘分类的条件。如 Norman(1988: 182)所说: “我选择了十种分类特征, 它们可以直接应用, 并且可以导致一个相对清楚的结果。”但根据我们第壹节所作的讨论, 作为遗存或者创新的词汇应该区别对待, 因为只有创新成分才有资格作为衡量亲缘关系的标准。

在一些汉语方言亲缘关系的计量研究中, 遗存成分与创新成分的区分被忽略了。王士元、沈钟伟(1992)以语素为基本计量单位来测算汉语方言之间的关系, 但他们的编码没有区别遗存和创新, 例如:

["走"]	BJ	JN	MX	GZ
行	0	0	1	1
走	1	1	0	0

按照这样的编码, BJ 和 JN 之间的亲缘距离就与 MX 和 GZ 之间的距离相同了, 二者都是 2。但“行[‘走’]”是古汉语的遗存, 而“走[‘走’]”是后来的创新。前者在汉语方言亲缘关系的计量上就没有信息量, 后者才有。我们的编码将得到 BJ 与 JN 之间的距离是 1, 而 MX 与 GZ 之间的距离是 0。因此我们认为, 把不同性质的语言成分放在一起计算语言之间的距离, 可能导致相当的偏离。

语义创新作为衡量亲缘关系远近的标准还可能遇到另外两个难题: 语义的平行独立发展(parallel development)问题和如何设定不同语义演变的权重(weight)问题。这两个难题的解决都期待着大型语义演变数据库的建立。我们可以从大量的语料中获得对特定语义演变的概率的把握, 同时了解特定语义演变对整个语言的重要程度, 进而对这些了解作量化处理, 这或许能解决这些难题。

语言演进只有一个真实的过去。由于汉语方言的复杂性和我们所能见到的汉语历史语料和现代方言语料的限制, 以及其他种种制约, 我们只能窥测其中的些许部分。尽管我们坚持用创新成分作为标准, 方言之间复杂的借用情形还是使我们不能确定对方言亲缘关系的计量到底有多准。即使是用 Swadesh 的 100 基本核心词, 仍然不能完全排除借用的可能。不过我们相信, 随着越来越多干扰信息的排除, 一个越来越清晰的语言演变图景将随着研究的深入而逐渐获得。一方面, 汉语史的研究将提供越来越多的语义创新的资料; 另一方面, 语义变化普遍性的研究将帮助我们建立起更多更可靠的语义链(Wilkins 1996), 为文献并不丰富的语言也提供应用语义创新的基础。总之, 我们期待这些方面的深入研究能给语言演进的描述提供更多的启发。



附录 I: 13 项语义创新在 19 个方言中的对应(据《汉语方言词汇》)

	树	眼睛	狗	肚子	脚	脖子	走	跑	站	闻
北京(BJ)	树	眼睛	狗	肚子	脚	脖子	走	跑	站	闻
济南(JN)	树	眼	狗	肚子	脚	脖子/脖儿颈	走	跑	站	闻
西安(XA)	树	眼窝	狗	肚子	脚	脖项/脖子	走	跑	站/立	闻
太原(TY)	树	眼[睛]	狗	肚子	脚	脖子	走	跑	站	闻
武汉(WH)	树	眼睛	狗	肚子	脚	颈子	走	跑	站	闻
合肥(HF)	树	眼	狗	肚子	脚	脑颈把子	走	跑	站	闻
扬州(YZ)	树	眼睛	狗	肚子	脚	颈项	走	跑/奔	站	闻
苏州(SZ)	树	眼睛	狗	肚皮	脚	头颈	走/跑	奔/跑	立	闻
温州(WZ)	树	眼乌珠	狗	肚	脚	头颈	走/行	射	倚	嗅 ^①
长沙(CS)	树	眼睛	狗	肚子	脚	颈根	走	跑/打飞脚	站/倚	闻
双峰(SF)	树	眼珠	狗	肚子	脚	颈壳	行	打飞脚/跑	倚	嗅
南昌(NC)	树	眼睛	狗	肚子	脚	颈	走	跑	倚/站	嗅
梅县(MX)	树	目珠/眼珠	狗	肚屎	脚	颈茎	行	ts ^h iak ₁	倚	鼻
广州(GZ)	树	眼	狗	肚	脚	颈	行	走	倚	闻
阳江(YJ)	树	眼	狗	肚	脚	颈[柄]	行	走	倚	闻
厦门(XM)	树	目珠/目	狗	腹肚	骹	颌管	行	走	倚	鼻
潮州(CZ)	树	目	狗	肚	骹	颌	行	走	倚	鼻
福州(FZ)	树	目珠/目	犬	腹老 ^②	骹	脰骨	行	pie ₁	倚	鼻
建瓯(JO)	树	目珠	狗	腹	骹	□□toeyŋ-ttoeyŋ-t/脰	行	走	倚	鼻

① 温州方言中，“听”意为“远闻”。

② 原为“腹肚”。我们根据福建师范大学陈泽平教授的建议(通信交流)改为“腹老”。



附录 II: PENNY 程序的输入数据

19 13 A

ANCESTOR	000000000000	CS	111110111100
BJ	111111111100	SF	11111001?000
JN	111111111100	NC	111110111000
XA	111111111100	MX	10111100??001
TY	111111111100	GZ	11111001?100
WH	111110111100	YJ	11111001?100
HF	111110111100	XM	1011?0001?011
YZ	111110111100	CZ	1011?0001?011
SZ	111110111100	FZ	1000???0??001
WZ	1111101??000	JO	1010???00?001

参考文献

北京大学中文系语言学教研室 1995 《汉语方言词汇》, (北京) 语文出版社

丁邦新 1996 书评:《中国语言地图集》, 《国际中国语言学评论》第 1 期

管锡华 2000 《史记单音词研究》, (成都) 巴蜀书社

洪成玉 1996 闻的初义及其具体用法, 《古汉语研究》第 1 期, (北京) 中华书局: 94 - 104 页

侯精一 2002 《现代汉语方言概论》, 上海教育出版社

蒋绍愚 2002 从{走}到{跑}的历史更替. Proceeding of the international symposium on the historical aspects of Chinese language-Commemorating the centennial birthday of the late professor Li Fang-Kuei, University of Washington, August 15 - 17, 2002

李 荣 1985 官话方言的分区, 《方言》第 1 期

汪维辉 2000 《东汉 - 隋常用词演变研究》, 南京大学出版社

王福堂 1998 《汉语方言语音的演变和层次》, (北京) 语文出版社

王士元、沈钟伟 1992 方言关系的计量表述, 《中国语文》第 227 期: 81 - 92 页

游汝杰 2000 《汉语方言学导论》, 上海教育出版社

周振鹤、游汝杰 1986 《方言与中国文化》, 上海人民出版社

Baxter, W. H. 1998 *Mandarin dialect phylogeny*. M. S

Blust, Robert A. 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. University of Hawai'i working papers in Linguistics 9.2: 1 - 15

Burling, R. 1983 The Sal languages. *Linguistics of the Tibeto-Burman Area* 7.2:1 - 32

Camín, J. H. and R. R. Sokal 1965 A method for deducing branching sequences in Phylogeny. *Evolution* 19: 311 - 326

Campbell, Lyle 1998 *Historical linguistics an introduction*. Edinburgh University Press

Felsenstein, J. 1986 - 1993 *PENNY-Branch and bound to find all most parsimonious trees*. University of Washington

Hock, H. H. 1991 *Principles of historical linguistics*. Berlin; Mouton de Gruyter. Second edition

Kluge, A. G. and Farris, J. S. 1969 Quantitative phyletics and the evolution of anurans. *Systematic Zoology*. 18: 1 - 32

Krishnamurti, Bh., Moses, L., and Danforth, D: 1983 Unchanged cognates as a criterion in linguistic



- subgrouping. *Language* 59: 541 – 568
- Norman, J. 1988 *Chinese*. Cambridge University Press
- Nei, M. and S. Kumar 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press
- Sagart, L. 2001a Nanxiong and Hakka. *Fangyan* 23, 2: 142 – 151
- Sagart, L. 2001b Comment: Malayo-Polynesian features in the AN-related vocabulary in Kadai. Paper presented at the workshop “Perspectives on the Phylogeny of East Asian Languages”, August 28 – 31, 2001, Prigueux
- Sagart, L. 2002 Sino-Tibeto-Austronesian: an updated and improved argument. Paper presented at the 9th International Conference on Austronesian Linguistics, Canberra, January 8 – 11, 2002
- Sokal, R. R. and Sneath, P. H. A. 1963 *Principles of numerical taxonomy*. San Francisco: W. H. Freeman
- Swadesh, M. 1955 Time depths of American linguistic groupings. *American Anthropologist* 56
- Swofford, D. L. 1993 PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer program distributed by the Smithsonian Institution, Washington, D. C.
- Tamura, K. 1997 – 1999 Tree-Explorer, Version 2.12, Computer program in MEGA (Molecular Evolutionary Genetics Analysis) package, Version 1.02, Copyright by S. Kumar, K. Tamura, M. Nei and the Pennsylvania State University
- Wang, Feng and Wang, William S.-Y. 2004 Basic Words and Language Evolution. *Language and linguistics* 5.3: 643 – 662
- Wang, William S.-Y. 1969 Competing changes as a cause of residue. *Language* 45: 9 – 25
- Wang, William S.-Y. 1987 Representing language relationships, in H. Hoenigswald and L. Wiener (eds.) *Biological metaphor and Cladistic classification*, Pittsburgh: University of Pennsylvania Press: 243 – 256
- Wang, William S.-Y. 1996 Linguistic diversity and language relationships. In Huang, C.-T. James and Li, Y.-H. Audrey (eds.) *New horizons in Chinese linguistics*. Dordrecht: Kluwer Academic Publishers: 235 – 268.
- Wilkins, David P. 1996 Natural tendencies of semantic change and the search for cognates. In M. Durie and M. Ross (eds.) *The comparative method reviewed*. Oxford University Press: 264 – 304

Semantic Innovations as a Criterion for Linguistic Subgrouping

Feng WANG and William S.-Y. WANG

Abstract In this paper, we argue that semantic innovations can be the characters for linguistic subgrouping. Chinese dialects as the case are classified by semantic innovations in basic semantic fields. A parsimonious algorithm, PENNY, is employed to do the quantitative analysis. The result of subgrouping is discussed in connection with other quantitative researches and studies on migrations.

Key words Semantic innovation, genetic relationship, parsimonious, quantitative