

梧州粤语和周边勾漏粤语词汇相似度的计量分析

粟春兵¹, 王文胜²

(1. 杭州师范大学 人文学院, 浙江 杭州 310036 ;

2. 杭州师范大学 应用语言研究中心, 浙江 杭州 310036)

[摘 要] 以梧州话、倒水话为例, 对两种方言的 1300 多个日常所用词项进行词汇相似度的计量分析, 并以粤方言的标准方言广州话与梧州话的相关系数作为参照, 进而说明梧州粤语和周边勾漏粤语的相关性及其差异。

[关键词] 梧州粤语; 勾漏粤语; 词汇相似度; 计量分析

[中图分类号] H178

[文献标识码] A

[文章编号] 1673- 8535(2011)05- 0007- 07

Quantitative Analysis of Lexical Similarities between Wuzhou Cantonese and Goulou Cantonese of Wuzhou Neighboring Areas

Su Chunbing¹, Wang Wensheng²

(1.College of Liberal Arts, Hangzhou Normal University, Hangzhou 310036, China ;

2.Applied Linguistics Research of Hangzhou Normal University, Hangzhou 310036, China)

Abstract : The paper makes a quantitative analysis of the lexical similarities in more than 1300 daily lexical items between Wuzhou and Daoshui dialects so as to illustrate the correlation and differences between Wuzhou Cantonese and Goulou Cantonese of Wuzhou neighboring areas with reference to the correlation coefficient between the Guangzhou dialect, i.e. standard Cantonese, and the Wuzhou dialect.

Key words : Wuzhou Cantonese; Goulou Cantonese; lexical similarity; quantitative analysis

一、概述

广西梧州市现通用的语言为粤语, 市区为广府粤语口音, 郊区为勾漏粤语口音。

广府粤语以广州话为代表, 是粤语中最具影响、使用人数最多的一种方言, 主要分布在广东珠江三角洲一带及广西西江流域上游的部分地区, 共 36 个县市, 还有香港、澳门两个地区, 使用人数约 2834 多万。勾漏粤语主要分布于广东西部的清远、肇庆两市所辖的部分地区及广西东部地带,

共 41 个县市, 使用人数约 1030 万。其中广东省 11 个区县市, 使用人口 350 余万, 广西壮族自治区 35 个县市, 使用人口 680 余万。在地理分布上, 该种方言基本连成一个整片^[1]。

众所周知, 广府粤语和勾漏粤语在语音上差异较大, 但两者在词汇方面的关系如何? 这是笔者关注的问题。笔者出生在广西苍梧县倒水镇, 5 岁时随家人搬迁到梧州市区, 开始兼用梧州话(指梧州城区话, 属广府粤语)和倒水话(属勾漏粤语)。根据《苍梧县志》记载, 苍梧县的粤方言大体划分为: 南五乡片, 包括龙圩、林水、广平、

大坡、新地，以龙圩为代表点；东安片包括沙头、石桥、木双、梨埠、六堡，以沙头为代表点；抚河片包括倒水、旺铺、夏郢、岭脚、人和、师寨、京南、长发，以倒水为代表点^[2]。为了能够细致、精确地研究梧州粤语和周边勾漏粤语间词汇的相似度，本文以梧州话和倒水话为例，对这两种方言的 1300 多个日常所用词项进行词汇相似度的计量分析。笔者使用 Spss19.0 统计软件计算梧州话和倒水话的相似度，并以粤方言的标准方言广州话与梧州话的相关系数作为参照，进而说明梧州话和倒水话的相关性及其差异。

二、计量分析的方法

(一) 计量方法的回顾

对方言间或不同语言间词汇相似度的计量研究方法有：语言年代学 (glottochronology)、相关系数统计法、算术统计法、概率法、矩阵分解 (matrix decomposition) 和加权平均法等。本文选用的相关系数统计法，最早是由郑锦全于 1973 年提出的，他用“皮尔逊相关”和“非加权平均系联法”计算不同方言的字音和词汇文字表现形式的亲疏程度。用于比较的词目有 905 条，在获得它们的相关系数的基础上，做聚类分析，用树形图对 18 种方言间的亲疏程度作出直观而细密的描写。但此方法有两个主要缺点：一是没有考虑词频对计量结果的影响；二是比较词汇的异同时，只考虑词形异同，并不顾及词内部词根或中心语素的异同。针对该方法的不合理性，王士元和沈钟伟于 1992 年共同提出在汉语方言的分类上应该以语素作为计算的基本单位，而不是词。他们进一步完善了相关系数统计法和聚类分析法 (简称王沈计量法)，给人们提供了一种切实可行的计量方言词汇相关系数的途径。

(二) 王沈计量法的概述^[3]

1. 特征选择

选择语素作为计算的基本单位，同时把构词方法也纳入计量的范围。因为词与词的关系是由

语素和构词方法表达出来的，如果把这两个要素排除在外，对词和词之间的部分相似关系弃之不顾的话，以词为计算单位的计算必然会增大方言间的实际距离。

2. 计量公式

就一对方言 (甲方言和乙方言) 而言，它们之间某一个语素或构词方法出现的情况只有“双有”、“有无”、“无有”、“双无”四种，分别用 a、b、c、d 代表^[4]。其中 d“双无”的情况对于分析两方言间关系没有实际意义。所以得出计算词汇相关系数的公式：

$$S_{ij}^k = a / (a+b+c) \quad \text{公式 1}$$

式中：k 表示用于比较的某个词项；

i、j 表示 k 词项在两种方言中的反映形式^[2]；

S_{ij}^k 表示 i 和 j 两种反应形式的相关系数；

$a / (a+b+c)$ 表示两种方言词汇语素和构词法的“双有”关系和所有关系^[3]的比值。

3. 计量原则

一对方言的总相关系数为单个词项相关系数总和的平均值。如果甲、乙两种方言共有 n 条词项参与比较，可以得到 n 个 S_{ij}^k 。其总体相关系数应该是这 n 个相关系数的平均值：

$$S_{AB} = \sum_{h=1}^n S_{ij}^k / n \quad \text{公式 2}$$

式中： S_{ij}^k 表示在两个方言中某个词项 k 的相关系数，h 表示它是 n 个相关系数中的任意一个，n 表示两个方言中用于比较的词项的总数。

(三) 方言词汇计量细节补充说明^[4]

计算两地方言词汇的相关系数时，本文需要补充几点王沈计量法中没有具体说明的细节。

1. 语素和构词法同时作为分析方言词汇关系的要素

其中，语素的比较方法是：首先分别找出用于比较的方言单个词项的反映形式 i 和 j 的所有语素。然后分析，某语素若是属于上文中提到的

“双有”情况 a，便将其作为比较项，标记为“1”（“1”表示反映形式存在，“0”表示反映形式不存在。“反映形式”包括语素和构词法两方面）。若某语素只在 i 或 j 中出现，则属于“有无”情况 b 或“无有”情况 c，亦将它们所属的不同语素作为比较项，对具有某语素的反映形式标记为“1”，不具有此语素的反映形式标记为“0”。最后，按照公式 1 计算语素项的相关系数。

在进行构词法的比较时，其依据是：词的构成形式的三个层次，如图 1。

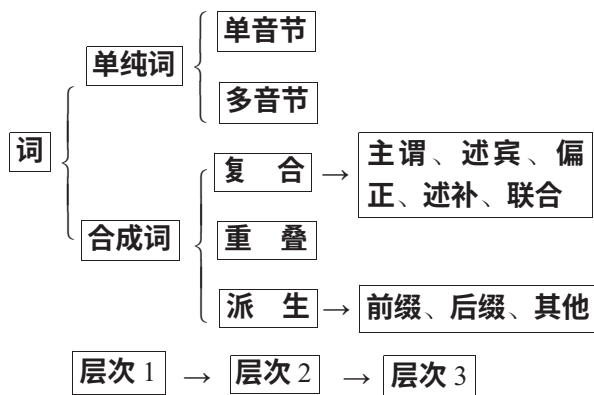


图 1 词构成的层次结构

说明：派生中的“其他”项包括中缀、叠缀等形式

比较词构成的层次结构，其次序为层次 3→层次 2→层次 1，分析时以最小区分特征为区分点。如果在某一层次结构上，反映形式 i 和 j 都具有某种结构，便将此结构作为比较项，两者都标记为“1”，属于上文中提到的“双有”情况 a。如果在某一层次结构上反映形式 i 或 j 具有某种结构，亦将此结构作为比较项，把具有此结构的反映形式标记为“1”；不具有此结构的反映形式标记为“0”，属于上文提到的“有无”或者“无有”情况。具体实例参见下文（四）。

2. 进行方言词汇计量时不需考虑音节的数目

因为在计量过程中，分析语素项的多少实际上已经体现了音节数量的差异。如果再将音节数作为一个要素进行分析，就会产生重复比较的结

果，使得词汇比较没有较好的区分度，因此在计量时，不需要将音节作为分析的要素。

3. 词和词组之间也可以进行比较

在比较词和词组的时候，语素之间的比较可按照词与词的比较方式进行。本文不将词组排除在统计对象之外，是因为词组在方言词汇中占有较大的比例，若将这部分词项删去的话，最后计量结果的准确性将会受到影响。而现有的方法其实完全可以对词和词组进行计量分析。

4. 两种方言同一词项的多种反映形式如何比较

在方言词汇的调查结果中，单个词项在某种方言里往往具有若干个反映形式。在计算这个词项的相关系数时，不能只将其中的某一个反映形式和另一个反映形式进行简单的比较分析。根据王士元和沈钟伟先生的意见，在同一词项有多种反映形式的情况下，可按照以下的方法单独求出该词项在两种方言中的相关系数。他们在文章中提到一个假设的例子。该例子的表述方式如表 1。

表 1 词项和反映形式关系

词项	反映形式	方言甲	方言乙
k	A	1	1
	B	1	0
	C	0	1

说明：1 表示有这个反映形式；0 表示无这个反映形式。

方言甲中有反映形式 A，而方言乙中也有，即“双有”情况 a=1。方言甲中有反映形式 B，而方言乙无，即“有无”情况 b=1。方言甲中无反映形式 C，而方言乙中有，即“无有”情况 c=1。再根据公式 1 计算 $S_{ij}^k = a / (a+b+c) = 1 / (1+1+1) = 1/3$ 。

可见，在计算具有多种反映形式的同一词项时，应先找出该词项在两种方言中所有的反映形式，然后根据这些反映形式分别考察它们的语素和构词法在这两种方言中的存在形式，确定“双

有”、“有无”和“无有”的情况之后，可得到 a、b、c 三个系数的值，最后根据公式 $S_{ij}^k=a/(a+b+c)$ 求出两种方言中该词项的相关系数。

(四) 汉语方言词汇计量举例

例一：“下雨”这个词项在梧州话中反映形式是“落雨”，在倒水话中的反映形式是“落水”，如表 2。

表 2 计量例一

语素及构词		反应形式	梧州话	倒水话
语素	1	落	1	1
	2	雨	1	0
	3	水	0	1
构词法	4	述宾	1	1

说明：1 表示这个反映形式存在，0 表示这个反映形式不存在。按层次 3 向层次 1 的顺序进行比较分析（以最小区分特征为区分点），表 3 同。

例二：“小孩”这个词项在梧州话中有“细蚊仔”“细佬仔”“细佬哥”三个反映形式，倒水话有“细子”“细子儿”两种反映形式，如表 3。

表 3 计量例二

语素及构词		反应形式	梧州话	倒水话
语素	1	细	1	1
	2	蚊	1	0
	3	仔	1	0
	4	佬	1	0
	5	哥	1	0
	6	子	0	1
	7	儿	0	1
构词法	8	后缀	1	1

“双有”即“1—1”的情况为 a，“有无”即“1—0”的情况为 b，“无有”即“0—1”的情况为 c。根据分析结果确定 a、b、c 的值后，由公式 1 计算相关系数数值。

例一	例二	平均相关系数
$S_{ij}^k=2/4=0.5=S_a^1$	$S_{ij}^k=2/8=0.25=S_a^2$	$(S_a^1+S_a^2)/2=0.375$

三、梧州话和倒水话相关系数的计量分析

(一) 梧州话、倒水话词汇相关系数的计量分析材料

本文所用方言词汇材料来源：

广州粤语词汇——白宛如《广州方言词典（现代汉语方言大词典·分卷）》，江苏教育出版社 1998 年版。广州市地方志编纂委员会编《广州市志（卷 17 社会卷）》，广州出版社 1999 年版。饶秉才，欧阳觉亚，周无忌《广州话方言词典》，商务印书馆香港分馆 1981 年版。

梧州粤语词汇——笔者 2011 年暑假在梧州市区做的方言调查材料。辅以梧州市地方志编纂委员会编《梧州市志》，《梧州史志》编辑部 1996 年版。

倒水粤语词汇——笔者 2010 年暑假在倒水镇做的田野调查材料。辅以苍梧县志编纂委员会编《苍梧县志》，广西人民出版社 1997 年版。

(二) 计量所用词汇表的范围及分类^[5]

本文选定的词量适中，有 1300 多个。所选用语词来源于中国社会科学院语言研究所方言研究室资料室在《方言》2003 年第一期刊出的《汉语方言词语调查条目表》，并参照詹伯慧主编《汉语方言及方言调查》中的第九章“汉语方言词汇语法调查表”进行增删。确定后的调查词表共有 27 个词类，1300 多个词项，如下页表 4。

表 4 分类号和分类项目对应表

1	2	3	4	5	6	7	8	9
天文	地理	时令时间	农业	植物	动物	房舍	器具用品	称谓
10	11	12	13	14	15	16	17	18
亲属	身体	疾病医疗	衣服	饮食	红白大事	日常生活	讼事	交际
19	20	21	22	23	24	25	26	27
商业交通	文化教育	文体活动	动作	位置	代词	形容词	副 / 介词	量词

(三) 词汇相关系数的计量分析步骤

首先, 将广州话、梧州话、倒水话按照方言调查词表的词项一一对应, 以 Excel 表的形式列出。然后根据每个词项的反映形式进行语素和构词法的比较, 并赋以相应的值。最后, 按照公式 1 计算单个词项的相关系数, 以公式 2 计算所有词项的平均相关系数。得到三地方言词汇的相关系数值, 可以利用 Spss 统计工具来进行更深一步的统计分析。

(四) 计量结果比较分析

1. 相关系数值的比较

据统计, 广州话与梧州话的词汇平均相关系数是 0.954, 梧州话和倒水话的平均相关系数是 0.757。广州话和梧州话的平均相关系数高于梧州话和倒水话的相关系数, 说明广州话和梧州话的词汇相似度高于梧州话和倒水话的相似度。

2. 相关系数分布比较分析

笔者使用 Spss19.0 统计软件处理广州话和梧州话的数据材料, 得出在 1321 个调查条目中, 梧州话和广州话相关系数为 1 (即完全相同) 的条目有 1210 个, 占总条目的 91.6%, 相关系数为 0 (即完全不相同) 的条目有 6 个, 占同条目的 0.5%, 剩余 105 个词项的相关系数大于 0 小于 1 (即部分相同), 它们占总条目的 7.9% (如图 2)。

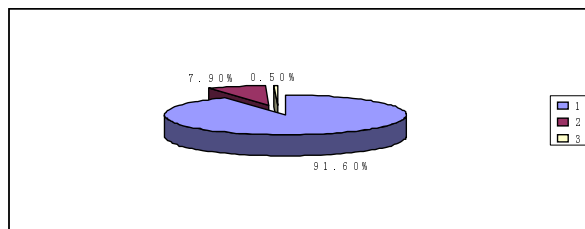


图 2 广州话和梧州话词汇相关系数比例分布

说明: 相关系数为 1 的占 91.6%, 相关系数为 0 的占 0.5%, 相关系数大于 0 小于 1 的占 7.9%。

在 1321 个调查条目中, 梧州话和倒水话相关系数为 1 (即完全相同) 的条目有 803 个, 占总条目的 60.8%, 相关系数为 0 (即完全不相同) 的条目有 52 个, 占同条目的 3.9%, 剩余 466 个词项的相关系数大于 0 小于 1 (即部分相同), 它们占总条目的 35.3%, 如图 3。

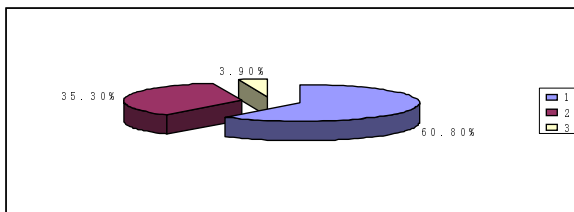


图 3 梧州话和倒水话词汇相关系数比例分布

说明: 相关系数为 1 的占 60.8%, 相关系数为 0 的占 3.9%, 相关系数大于 0 小于 1 的占 35.3%。

3. 相关系数的词类分布比较分析

为了更细致地揭示梧州话和倒水话的关系，笔者对不同词类相关系数的分布情况进行统计分析，分别求出了“广州—梧州”、“梧州—倒水”各词类相关系数的平均值，如图 4。

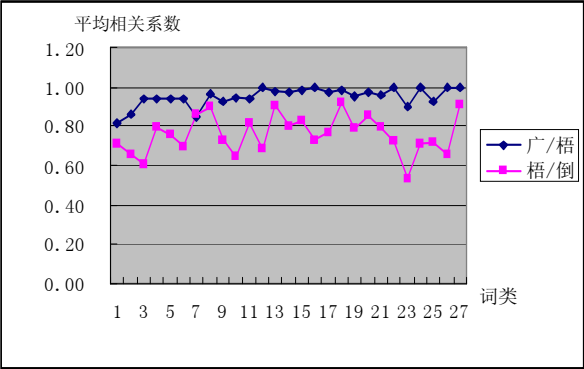


图 4 “广州—梧州”、“梧州—倒水”
各词类的平均相关系数分布

说明：图中有 27 个词类，上面那条折线为“广州—梧州”、下面那条为“梧州—倒水”。

如图 4 所示，广州话和梧州话的各个词类的平均相关系数值普遍较高，说明它们高度相关。而梧州话和倒水话的各个词类的平均相关系数值起伏较大，且均低于广州话和梧州话的平均相关系数，说明梧州话和倒水话的词汇存在着一定的差异。若将梧州话和倒水话各词类的平均相关系数进行分组，大致情况见表 5 和下页图 5。

表 5 “梧州—倒水”各词类的平均相关系数分组

平均相关系数	$1 > X \geq 0.9$	$0.9 > X \geq 0.8$	$0.8 > X \geq 0.7$	$0.7 > X \geq 0.6$	$0.6 > X > 0.5$
词类	8 器具/用品、13 衣服穿戴、18 交际、27 量词	7 房舍、11 身体、14 饮食、15 红白大事、20 文化教育、21 文体活动、	1 天文、4 农业、5 植物、6 动物、9 称谓、16 日常生活、17 讼事、19 商业交通、22 动作、24 代词、25 形容词	2 地理、3 时令时间、10 亲属、12 疾病医疗、26 副词介词	23、位置

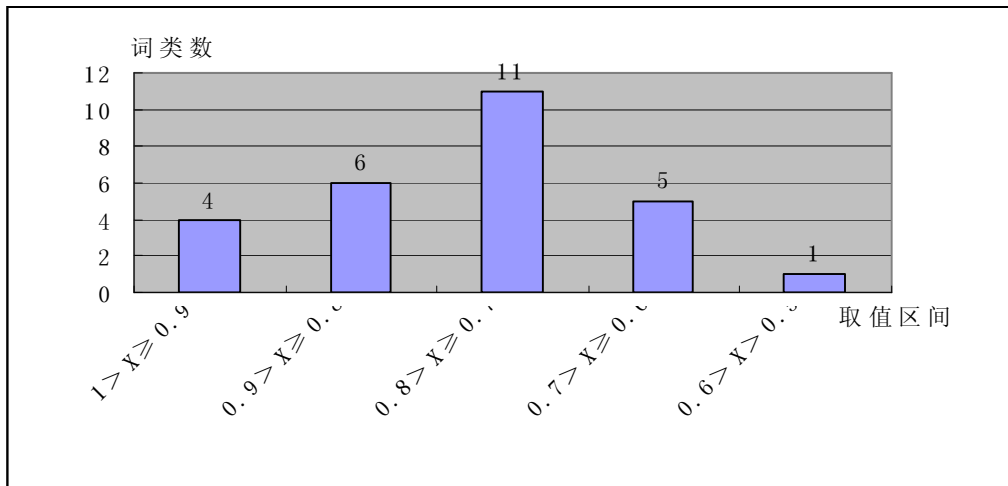


图5 “梧州—倒水”各词类的平均相关系数分布

如表5所示,在梧州话、倒水话中已有相当一部分词类具有较高的相似度,它们大部分是实词,虚词类词语的相似度不及实词类的高。这说明梧州粤语和周边的勾漏粤语有较为频繁的生活接触,因此,两种方言的实词类词语更加容易发生趋同。而图5中,“梧州—倒水”各词类的平均相关系数分布呈现较为明显的正态分布,这说明在梧州话和倒水话词汇计量比较中,相关系数值的分布有一定的普遍性和规律性。

四、结语

梧州粤语和周边勾漏粤语的接触研究成果较少,即便有也是局限于对现象的描写,很少涉及定性、定量的分析。梧州粤语和周边勾漏粤语的相似度程度,它们的整体关系如何,至今还没有人用统计分析的方言加以验证。本文尝试通过使用细化的方言词汇相关度的计量方法,并借助 Excel 和 Spss19.0 统计软件,对梧州话和倒水话中1300多个词项进行定量统计分析,借此精确地刻画出梧州粤语和周边勾漏粤语的相似度,为今后这两种方言的接触研究提供佐证。由于方言词汇的计量比较研究尚处于起步和探索阶段,文中难免会有疏漏之处,有待进一步的推敲。

注释:

- (1) a 表示在两个方言中都出现; b 表示在甲方言出现而在乙方言不出现; c 表示在乙方言出现而在甲方言不出现; d 表示在两个方言中都不出现,是以称之为“双有”,“有无”,“无有”和“双无”。
- (2) “反应形式”就是某个词项在方言中对应的词条。
- (3) “所有关系”不包括双无关系。

参考文献:

- [1] 伍巍. 粤语[J]. 方言, 2007(2): 167-176.
- [2] 苍梧县志编纂委员会. 苍梧县志[M]. 南宁市: 广西人民出版社, 1997.
- [3] 王士元, 沈钟伟. 方言关系的计量表述[J]. 中国语文, 1992(2): 81-92.
- [4] 邵慧君, 秦绿叶. 廉江市粤客词汇相似度的计量分析[J]. 中国语文, 2008(2): 154-167.
- [5] 秦绿叶. 粤西三地粤语客话方言词汇分析计量研究[D]. 华南师范大学硕士学位论文, 2007.

[作者简介] 粟春兵 (1986-), 女, 广西梧州人, 杭州师范大学人文学院汉语言文字学专业硕士研究生, 研究方向: 语言学。

王文胜 (1968-), 男, 杭州师范大学应用语言学研究

中心研究员, 研究方向: 汉语方言学。

(责任编辑: 高 坚)