

汉语方言同音字汇的计量比较模型

吴南开¹，侯兴泉¹，孔江平²

¹暨南大学中文系/汉语方言研究中心

²北京大学中文系/语言学实验室

汉语方言的计量比较一直是学界关注的重点问题，自上世纪七十年代以来，该领域已经发展出来多种计量比较方法。依据计算单位的类型，可以将这些方法分为以下类型(江荻，2022)：(a)语音特征计量法，包括郑锦全(1973, 1988)、陆致极(1987)、王士元&沈钟伟(1992)、肖双荣&吴道勤(2004)、谢建猷&张宗(2014)等；(b)词源统计法，包括徐通锵(1991)、邓晓华&王士元(2009)、林天送等(2010)等；(c)词汇相似度算法，包括邵慧君&秦绿叶(2008)、粟春兵&王文胜(2011)、郑伟娜(2017)等；(d)编辑距离算法(包括自动相似性算法)，包括江荻(2017)、赵志靖&江荻(2018)、冉启斌&索伦·维希曼(2018)、冉启斌(2019)等。这些研究的共同特点是在方言以及民族语的语音比较问题上，从不同侧面引入了量化比较方法，从而在传统定性研究手段之外，开拓出新的研究路径。

尽管如此，这些计量比较研究仍然各自存在一些突出的问题，例如音系处理的随机性会导致语音特征计量的失真，同源词或核心词的选取可能存在抽样不足的问题(尤其是汉语方言中，常用的字词往往都是同源的)。因此，各类方法既具备各自的特色，也有着各自的不足。

为了继续深入探索这一问题，本研究以汉语方言同音字汇作为研究对象，初步建立了一种面向汉语方言同音字汇的综合性计量比较模型。这一模型的特点是以同音字汇作为整体性的计量对象，从而尽可能囊括语音特征、同源词、同音词等各类对象

与现有的方法相比，本研究主要有两个优势：

(1) 材料方面的优势。与音系、音节表或有限的核心词相比，同音字汇是更有代表性的语言材料，它包含了汉语方言的音位、音节、词汇等多个维度的信息。可以说，任何对汉语方言语音的比较研究和对汉语方言语音特征的总结，都必然要在同音字汇的比较中找到依据。因此，同音字汇在汉语方言学中一直受到高度重视。

(2) 在方法上的优势。在传统的语音计量比较中，对音位处理的处理一直是一个难题，因为对一个音位的不同处理不可避免地会导致计量比较结果的差异。本研究则主要基于音节这一层级的同音关系，这就避免了音位处理的困难。最重要的是，这种方法不是简单地通过某种测量对方言进行某种参数的比较，而是对传统的音韵学-方言学方法进行系统的建模。虽然该方法仍需进一步发展，但它作为一种定量和定性相结合的研究范式是很有前途的。