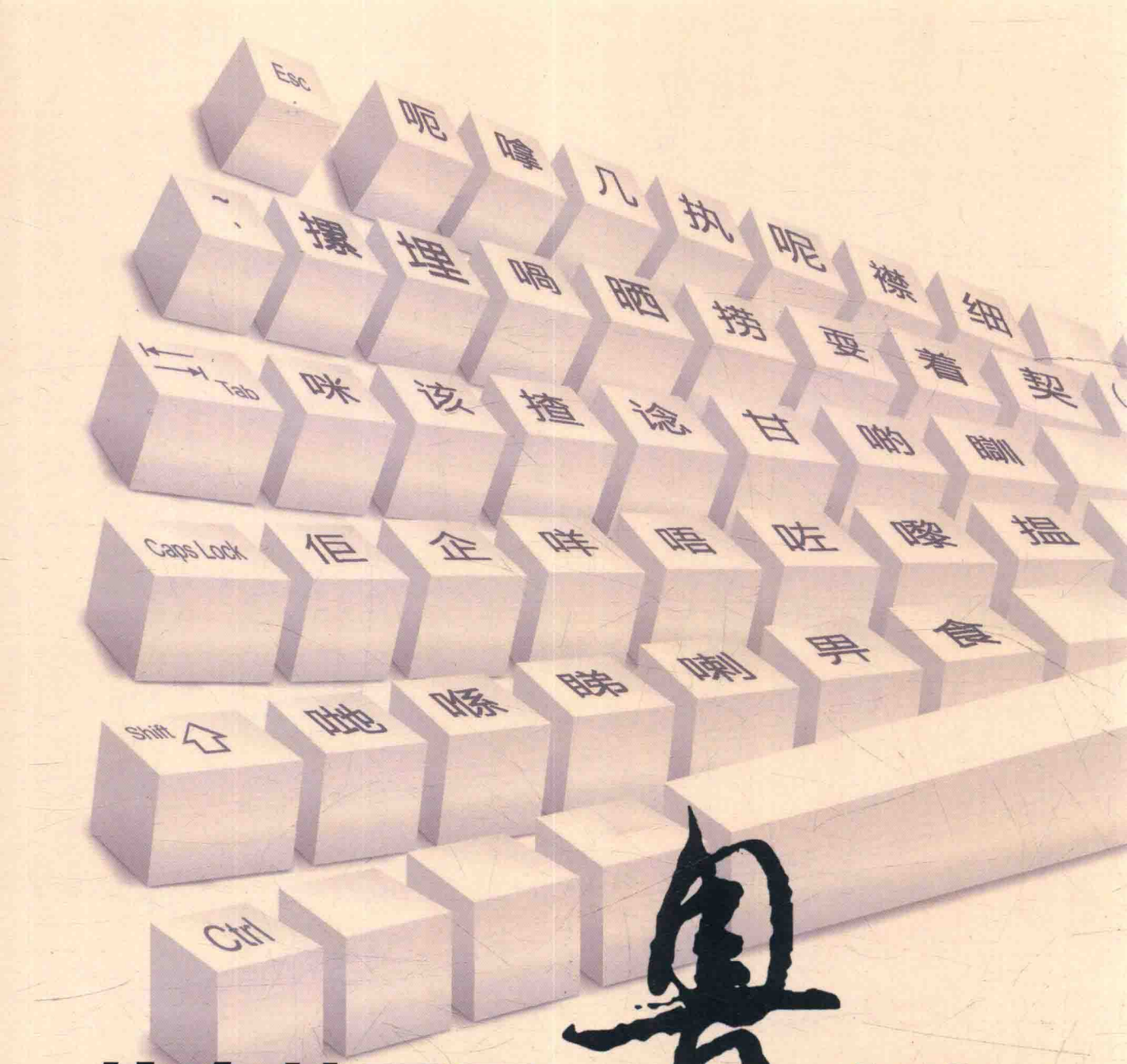


# 信息处理用粤方言字词规范研究

侯兴泉　吴南开　著

SPM
南方出版传媒
广东人民出版社

# 信息处理用粤方言字词规范研究

侯兴泉　吴南开　著

本书获暨南大学广东省重点学科“中国语言文学”建设经费资助出版

SPM
南方出版传媒
广东人民出版社

·广州·

**图书在版编目（CIP）数据**

信息处理用粤方言字词规范研究 / 侯兴泉，吴南开著. —广州 ：广东人民出版社，2017.5

ISBN 978-7-218-11766-9

Ⅰ. ①信… Ⅱ. ①侯… ②吴… Ⅲ. ①粤语—方言研究 Ⅳ. ①H178

中国版本图书馆 CIP 数据核字（2017）第 102929 号

XINXI CHULI YONG YUEFANGYAN ZICI GUIFAN YANJIU

**信息处理用粤方言字词规范研究**

侯兴泉 吴南开 著



出 版 人：肖风华

责任编辑：林小玲 骆 妮 刘 奎

责任技编：周 杰 易志华 吴彦斌

设计排版：广州六宇文化传播有限公司
Guangzhou Liuyu Culture Communication Co., Ltd.

出版发行：广东人民出版社

地 址：广州市大沙头四马路 10 号（邮政编码：510102）

电 话：（020）83798714（总编室）

传 真：（020）83780199

网 址：http：//www. gdpph. com

印 刷：虎彩印艺股份有限公司

开 本：787mm × 1092mm 1/16

印 张：24.5 字 数：493 千

版 次：2017 年 5 月第 1 版 2017 年 5 月第 1 次印刷

定 价：45.00 元

如发现印装质量问题，影响阅读，请与出版社（020-83795749）联系调换。

售书热线：020-83780517 83781479

# 序一

## 一本服务于语言应用的好书

在迎春接福、万象更新的喜庆日子里，兴泉给我送来了特殊的“红包”：一部题为《信息处理用粤方言字词规范研究》的书稿。这是他和他的研究生吴南开的研究成果，是一份别具新意的新春礼物。兴泉在书稿中附上几句话，要我为这本新著作序。

随着各地“语保”工程的启动，各种语言资源的开发、利用和保护日益受到重视。作为我国丰富语言资源的重要组成部分，汉语方言的社会应用也深受社会各界，特别是语言学界的关注。近期各地出现了方言节目纷纷进入传媒，方言读物、方言辞书不断面世，方言活动日渐活跃的情况，反映出方言资源的运用日渐广泛，方言所承载的地域文化繁荣发展。

汉语方言千差万别，其中被认为是“强势方言”的粤方言，一直是应用范围较广、海内外影响较大的一大方言。正因为应用范围广，粤方言在应用中难免有许多现象需要认真面对，有许多问题需要研究解决。就拿最引人注目的正音问题来说，鉴于粤语中某些字、词存在读音分歧的现象，给其社会应用带来困扰，我们在上个世纪九十年代曾经组织粤港澳的粤语学者进行旷日持久的逐字审音，结果好不容易才在十年后（2002 年）出版了一本《广州话正音字典》。然而，时至今日，粤语的正音仍然存在不同的看法，应该说，还是议论纷纷，各执一词，没能完全取得共识及得到彻底解决。举此一例，足见粤语在应用中还需面对现实，还有大量的工作要做。除正音问题外，在粤语的应用中，另一个需要面对的现实问题，就是方言字（词）的使用问题。众所周知，粤语是汉语方言中唯一一个“方言入文”现象比较普遍的方言，在一般通俗性的书面语中，形形色色的方言词语登堂入室，司空见惯。这些出现在书面语中的方言字词，写法花样百出，不是看惯用惯的人，简直莫名其妙，无法看懂。可见粤语用字问题同样是一个应该大做文章、应该开研讨会讨论解决的课题。十五年前（2002 年）香港理工大学曾经举办过一次汉语方言书写国际研讨会，在那次会上，香港理工大学的张群显博士和包睿舜（Bauer）博士报告了他们联合撰写的英文著作《以汉字写粤

语》（*The Representation of Cantonese with Chinese Characters*），就粤语的方言用字进行了详尽的论述，深受与会学者的瞩目。我在会上也发表了《关于方言词的用字问题》，以粤方言为例，就粤方言应用中这一突出的问题略抒管见。打那以后，陆续有粤语学者就此进行研究，发表论述。然而，迄今也还没有能够取到共识，未能采取可行的办法来解决。可喜的是：这次兴泉博士给我送来的书稿，正是一部有关粤方言用字问题的专论。作者涉猎这一课题多年，2014 年他和彭志峰、钟奇和彭小川在《语言教学与研究》上联合发表《面向中文信息处理的粤方言字规范刍议》一文，为这本著作的编写定下了基本的框架和主要的研究内容。现在我们看到的这本新著，开门见山就表明本书要着重探讨的是粤方言字词的规范问题，特别是为信息处理用的用字用词规范问题，紧紧抓住粤方言应用中这个亟待解决的突出问题展开深入的剖析探讨，可谓切中要害，对症下药。

粤方言用字的规范多年来常有人在相关的学术会议上呼吁，例如澳门大学的邓景滨博士就曾三度撰文论及此事，资深的粤语辞书编纂学者周无忌也屡屡呼吁粤方言用字应予规范，并且提出过一些具体的建议。如此等等。这些粤语学者有关粤语用字规范问题的论述，都被吸收到兴泉这本新著中来了，从这一点看来，兴泉这本新著可说是建立在博采前人相关论述的基础上，通览前人的见解而进一步加以阐述，加以延伸，多少带有一点总结性色彩的专著。尤其值得称道的是：作者能够与时俱进，在运用现代科技手段的基础上，刻意把粤方言用字的规范和信息处理这一现代化需求挂起钩来，这就使粤语用字规范的探讨更具时代精神与现实意义，充分体现出作者的前瞻意识。

打开这本书稿的目录，全书包括“绪篇”“字篇”“词篇”三大部分，外加 5 个附录。“绪篇”的内容相当丰富，在第一节“研究概况”中除对粤方言字词规范研究进行综论外，还对历来粤方言字词工具书的编纂进行详细的介绍；“字篇”和“词篇”是本书的中心部分，分别对粤方言用字和用词的标注规范进行论述；附录部分提供以下资料：常见粤方言异体字音形义对照表、常见粤方言多音字表、粤方言字（含异体字）字码对照表、常见粤方言异形词（多音节）音形义对照表、信息处理用粤方言词性标注集等。这些相关资料，对于粤方言用字的研究都是很有参考价值的。总体来说，这本书章节不多，内容集中而又安排紧凑，突出主干而又避开不必要的枝叶，堪称是一本服务于语言应用的好书。

詹伯慧

丁酉元宵于羊城暨南园苏州苑

# 序二

新春时节，兴泉送来他与他的开门弟子吴南开合著的书稿《信息处理用粤方言字词规范研究》请我写序。我一直关注着他们师生俩的这项研究，深知书稿倾注了他们大量的心血，自然欣然应允。

## 一

对信息处理用粤方言字词进行规范，这是粤语理论研究与应用研究相结合的非常重要的基础性研究。

众所周知，汉语方言是国家语言资源的重要组成部分，而粤语又是在海内外具有巨大影响力且极有活力的一种汉语方言。多年来，方言学界对粤语的方方面面作了大量的调查与研究，成果丰硕，但这些成果大都是零散的；另外，不少乡音也随着时间的推移在不断地消失。人们越来越深切地意识到，要充分发挥方言资源的作用，要满足粤语语音识别、文语转换、数字出版、机器翻译、现代化教学、通讯、公共服务以及抢救乡音、高质量采录记载粤语实态语料等方面的需要，迫不容缓的是必须走信息化和现代化的道路。

然而，存在的现实问题是，粤方言的字词跟其他方言一样，一直以来都缺乏相应的规范，而缺乏标准和规范，粤语信息处理的有效性势必大打折扣。毫不夸张地说，在大数据的年代，制定信息处理用粤语字、词以及拼音标准和规范，这是粤语研究走信息化和现代化道路必不可少的基础和重要前提。

值得点赞的是，兴泉和他的研究生们迎难而上，做了深入的开创性的工作。整本著作从信息处理的角度全面讨论了粤语字词在字形、读音、编码和分类标注等方面的问题，提出不少颇有见地的可供规范操作的理念和方法；文末还附录了大量的实用性表格（如异体字表、多音字表、字码对照表、异形词表、词性标注表、规范词形表等），这些表格从2011年开始设计和制定，花费了兴泉和他的学生们大量的时间和心血，可供各行业参考和广泛应用，初步实现了制定出一套完善的信息处理用粤方言字词的规范和标准的目标，为今后粤语研究与应用的信息化和现代化，为提高粤方言信息处理的效率和质量，从而更好地发挥粤方言字词的应用价值，奠定了坚实的基础。

同时，也为其他方言的信息化提供了很好的示范。可以说，这是该书的一大亮点，意义非凡。

## 二

粤方言字词规范和标准要制定得好并非随心所欲的事情，它有赖于开阔的学术视野、扎实深厚的专业功底，以及严谨的科研态度。这几方面，兴泉都做到了，颇值得赞许。

学术视野开阔。

粤方言字词的规范主要包括粤方言字词的定性、读音规范、字形规范、繁简对应、无码字的规范、词类和词性标注规范、字符集的研制等内容，涉及面很广。细读该书，不难发现，这些方面前贤的种种研究成果，不论是中国内地的，还是港澳地区甚至国际上的，也不论是今的，还是古的，作者均广泛地一一涉猎；在充分肯定前人成果的积极意义的同时，也中肯地指出，粤方言字词规范所涉及到的许多基础性问题尚未得到解决，譬如粤方言字词的定量研究、异体字和异形词的收集与整理、多音字词的收集与整理、繁简字的收集与整理、无码粤方言字的收集与整理、词类和词性标注规范的制定等等，从而使自己的这项研究起点更高，针对性更强。

专业功底扎实深厚。

通观全书，可以看到，粤方言字词规范涉及到的语言本体知识涵盖面很广。其一，有异体字、繁简字、多音字、异形词、词类划分与标注、分词及标注等等。其中异形词与异体字、繁简字等还有着复杂的联系。其二，众所周知，粤方言跟其他汉语方言相比，有一个明显的不同之处，那就是具有一套完善的记录方言的文字系统，其中狭义的粤方言字主要包括大量的自造字，另还有部分沿用古汉字、假借字和训读字。其三，尽管作者指出，鉴于字音规范方面已有詹伯慧先生主编的《广州话正音词典》，该书主要对字词进行规范，但对多音字、异体字、异形词等进行规范，必然涉及到音韵、音系的问题。此外，除了语言本体，该书还涉及到语言信息处理的许多知识，包括字符集、汉字编码、操作系统等等。值得称道的是，该书中无论是涉及语言本体方面的字、词、语法和语音知识，还是语言信息处理方面，兴泉都有着扎实的功底，驾驭得很好。

科研态度非常严谨。

该书制定了粤方言字词规范的主要原则，即："通用性原则""区别性原则""系统性原则""实用性原则""稳定性原则"。具体到每一项，又有该项的规范原则，如"粤方言异体字整理的原则""粤方言繁体方言字的简化原则""粤方言多音字的规范原则""粤方言异形词整理原则""粤方言词类划分及词性标注的原则""粤方言文本分词的总体原则"等等。

原则制定后，还分别提出具体的操作方法，并配有例释，每一项都考虑得非常细

致、周到。比如粤方言异体字的规范，作者在吸收前人研究成果的基础上提出了“通用性”“示意性”“示音性”“传承性”“简易性”“系统性”“区别性”“操作性”八项原则。考虑到运用这些原则对具体的异体字组进行操作时还会产生不少问题，为了尽可能地减少研究者主观因素的加入，该书提出了先平行处理后加权的操作方法，并对可能发生的种种情况提出几点权衡意见。随后以对几组较为典型的异体字组进行规范整理选出推荐字作为示范，所作的例释都很严谨、精彩，结论令人信服。

上面我们只是分开进行评价，其实，这三方面常常是有机地统一在整本书稿的分析研究中的。例如，该书4.4对粤方言几组较为典型的异体字进行规范整理，作者在考察方言字通用性的时候，主要考察目前市面上通用的粤方言字词典、粤方言教科书、作者自建的100万字的广州话方言语料库，以及对应方言字在谷歌或百度上的出现频率；评价其理据性时，则主要参考《说文解字》《广韵》《集韵》和《康熙字典》等古代字书或韵书，尽量符合音义皆通的原则。几方面结合，最终确定推荐用字。整个过程，不难看到，既体现了作者视野的开阔、态度的严谨，又足见其具有扎实的学术功底。

## 三

兴泉之所以能在这部书稿中游刃有余地对信息处理用粤方言字词规范问题作出高质量的研究，一切皆源自于他对语言学的深切的热爱，以及扎实深厚的语言学功底。

兴泉早在大学阶段就已对语言学表现出浓厚的兴趣，并有志于从事方言研究。当年他就读于深圳大学师范学院，除了本专业开设的课程之外，他刻苦求学，主动到文学院系统修读该院所开设的一系列语言学选修课。这样的本科生，我想，无论是过去还是现在都是凤毛麟角的。

2003年，他以优异的成绩考上暨南大学应用语言学专业对外汉语教学方向的研究生，本人成为他的硕士导师。读研阶段，他依然保持好学勤思的好品质。当时，对外汉语方向的课程由本人所在的华文学院开出，我支持他在学好本方向课程的同时，坚持到文学院中文系修读名师伍巍教授的“音韵学”和“国际音标”、甘于恩教授的“广东汉语方言研究”等课程，以使知识结构搭建得更加全面。他思维活跃敏捷，创新意识和能力都很强，在读期间，以硕士研究生的身份在核心期刊《方言》《暨南学报》（社科版）上独立发表论文两篇，内容分别涉及方言语法、方言音韵与实验语音。此外，还先后四次在国际性或全国性的学术会议上宣读颇有见地的论文四篇，并积极参与讨论，较广泛地引起了方言学界前辈学者们的注意，可谓崭露头角。因表现突出，毕业那年被评为广东省“南粤优秀研究生”。

2006年，兴泉报考北京大学方言学专业的博士生，成绩名列前茅，有幸成为著名方言学家李小凡教授的弟子。北大四年（北大规定博士生至少读够四年）期间，他跟

随李小凡教授和项梦冰教授系统学习了汉语方言调查及研究的系列课程，并两次参加暑期的方言调查实践。兴泉对实验语音学一直都有浓厚的兴趣，硕士期间就做过大量的粤语语音标注及切分工作，还自费参加过社科院举办的实验语音学培训班。读博期间，他跟国内著名的语音学家孔江平教授学习了实验语音学和汉藏语研究的相关课程，跟著名历史语言学家王洪君教授系统学习了历史语言学、理论语言学及生成音系学等课程。兴泉还有幸跟随国际著名音系学家端木三教授学习了西方的音系学理论和研究方法，开拓了研究视野与眼界。得益于北大四年系统专业的学习与训练，兴泉的博士毕业论文送审时被五位匿名评审一致评为优秀。

三个阶段的求学经历，兴泉打下了非常扎实而全面的专业功底。博士尚未毕业，暨南大学就已盯上了他这个好苗子。毕业后，他来到暨大文学院中文系和汉语方言研究中心从事教学科研工作，很快就成为教学和科研骨干。他先后成功获得教育部人文社科青年基金项目、国家社科基金青年项目，并以子项目负责人身份参加了詹伯慧教授主持的国家社科基金重大项目“汉语方言学大型辞书编纂的理论研究与数字化建设”；论文《论粤语和平话的从邪不分及其类型》（《中国语文》2012 年第 3 期）和《勾漏片粤语的两字连读变调》（《方言》2011 年第 2 期）荣获第二届中国语言学会“罗常培语言学奖”三等奖；专著《粤语勾漏片封开开建话语音研究——兼与勾漏片粤语及桂南平话的比较》也入选“清华语言学博士丛书”第二辑（2016 年出版）。

难能可贵的是，兴泉对传承母语方言和方言文化有着很高的使命感。在从事方言学理论研究的同时，他还致力于方言的应用研究，除了对信息处理用粤方言字词规范进行研究之外，还无怨无悔地积极从事广东汉语方言语言生态和语言资源保护工作，以及粤方言文化的调查、抢救、保育和推广工作。他建设了“方言文化网”（dac. jnu. edu. cn），开设“粤学堂”微信公众号；担任过中央电视台纪录片频道主办的“微方言大赛”、南方卫视二台主办的“谁语争锋”以及广东卫视“汉字密码”等大赛或方言文化综艺节目的学术顾问，担任过广州图书馆和暨南大学汉语方言研究中心联合举办的“知粤讲堂”以及中央人民广播电台“华夏之声”和羊城网合办之栏目《粤讲越过瘾》的主讲嘉宾；进行方言吟诵的收集和整理；帮助东莞市档案局建立方言档案管理平台……。

2016 年 6 月，兴泉被任命为暨南大学语言资源保护暨协同创研中心副主任，这带给他的更多的是一份责任。深信他会对方言和方言文化的研究与传承始终保持一颗热爱之心，借着国家正式启动“中国语言资源保护工程”的东风，在方言理论研究和应用研究相结合的学术道路上插翅飞翔，飞得更高，飞得更远！

彭小川

2017 年 2 月 20 日于翡翠绿洲寓所

# 目　录

## 绪　篇

## 字 篇

## 词 篇


# 绪　篇