

—语言保护研究丛书—

戴庆厦 主编

# 汉藏语复合元音的类型及渊源

田阡子◎著



科学出版社

语言保护研究丛书 | 戴庆厦 主编

# 汉藏语复合元音的类型及渊源

田阡子 著

科学出版社

北 京

## 内 容 简 介

汉藏语包括汉语群、藏缅语群、侗台语群、苗瑶语群、南亚语群、南岛语群，每个语群又包括若干个次语群。本书采集了汉藏语六个语群 684 种语言和方言的数据，使用数据库技术对复合元音的概念、性质、分类、语言类型、历史起源及历史音变规律进行综合性研究，挖掘复合元音在共时层面的语言共性及语言差异，寻找划分语言类型的语音参项，讨论复合元音的起源及演化的几种途径，建立复合元音在汉藏语中的历史音变模式。

本书采用现代语音学、现代音系学、传统音韵学、类型学、历史语言学等多种研究方法，从多个视角剖析和阐述复合元音的特征、类型和历史，目的是尝试解决复合元音在语言类型学和历史语言学中存在的某些理论问题。

### 图书在版编目(CIP)数据

汉藏语复合元音的类型及渊源 / 田阡子著. —北京: 科学出版社, 2016.12  
(语言保护研究丛书/戴庆厦主编)

ISBN 978-7-03-051227-7

I. ①汉… II. ①田… III. ①汉藏语系—元音—研究 IV. ①H4

中国版本图书馆 CIP 数据核字 (2016) 第 321281 号

责任编辑: 王洪秀 / 责任校对: 何艳萍

责任印制: 张 伟 / 封面设计: 铭轩堂

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司

科学出版社发行 各地新华书店经销

\*

2016 年 12 月第 一 版 开本: 720×1000 1/16

2016 年 12 月第一次印刷 印张: 12

字数: 220 000

定价: 78.00 元

(如有印装质量问题, 我社负责调换)



本书受到云南省哲学社会科学学术著作出版专项经费资助

本书受到云南省哲学社会科学创新团队项目“云南少数民族语言研究”经费资助（项目编号 2014CX01）

# 序 言

研究东亚区域语言的复合元音及其类型特征涉及多个基本概念。其一，复合元音之不同于单元音的性质及界定；其二，复合元音的类型与数量；其三，复合元音的源与流。此外，复合元音在不同区域语言或不同系属语言中的分布类型，包括共性和差异、复合元音与辅音声母或者辅音韵尾之间的关系、复合元音音变的模式及演变途径等也是可深度探讨的问题。

阡子博士这部书稿在她的博士论文基础上几经修订而成。从数年前她的论文纲要设计到现在的修订稿，可以说不同程度涉及和解答了上述几方面问题，概括性较高，部分论述具有领域先行性。所以她提出让我为这部书稿撰写序言我也就答应了。阡子博士毕业后，先后跟随中央民族大学戴庆厦教授和“台湾中研院”语言学所孙天心教授从事博士后研究，学术大有精进，又熟悉了田野调查，为她今后的语言研究铺垫了重要基础。

这部书稿原题为“东亚语言复合元音的类型及渊源”。显然，全方位研究区域语言复合元音最好的办法是收集和构建大规模语言数据。就目前国内外的研究来看，阡子这部著作收集的复合元音语言数据量相当可观，涵盖了东亚区域的藏缅语、苗瑶语、侗台语、南岛语、南亚语和汉语及其方言，共6大语群684种语言和方言。因之，在这么大的一个范围，建立了这么广泛语言的数据库，她的讨论真是可以纵横驰骋，各类数据斟酌采用，描写和论证，结论的可信度是很高的。

近年来，我研究的部分内容涉及汉语方言。可是，对于复合元音这个命题，具体考察方言材料的时候还真是吓了一跳。从论文到专著，甚至词典，各地方言普遍呈现大量的复合元音，从二合元音到三合元音，数量极为惊人。为此，我查看了 Ian Maddieson (Patterns of Sounds. Cambridge University Press, 1984) 和加利福尼亚大学洛杉矶分校“音系音段总藏数据库”(UCLA Phonological Segment Inventory Database, UPSID)，在317种语言中，二合元音8项，分布在22种语

言中, 仅占全部语言的 6.9%, 完全没有三合元音。加之我经常接触的藏语等藏缅语言仅有少量复合元音的事实, 我怀疑其中有什么不对劲的地方。进一步深究, 感觉这个问题似乎跟汉语拼音方案的应用有关。我也跟学界好些位知名方言学家讨论和交流了这个问题, 大家的看法基本一致, 均认为目前汉语方言描写在记音技术上相当程度地受到汉语拼音模式的影响和制约。为此, 我想就此在这部书稿之外谈谈这个问题。

《汉语拼音方案》实施已有近 60 年之久, 由于其实用性和应用的广泛性, 目前已为社会各界普遍接受, 并在中国社会生活中发挥着巨大作用。按照方案, 汉语普通话的元音和韵母呈现出一个很独特的多符号和多音素面貌, 我们不妨回顾一下这个体系。

单元音: a o e i u ü(yu)

后滑复合元音(开口呼): ai ei ao ou

前滑复合元音(齐齿呼): ia ie iao iou

前滑复合元音(合口呼): ua uo uai uei

前滑复合元音(撮口呼): üe(yüe)

其中单元音 6 个, 二合元音 9 个, 三合元音 4 个。从符号所反映的语音性质看, 又分为带元音介音的和不带元音介音的两类, 恰与音韵学上的非开口呼(齐齿呼、合口呼、撮口呼)与开口呼对应。按照这本书稿的分类, 开口呼都是后滑复合元音(4 个), 其他则是前滑复合元音(9 个)。在一个总数为 19 个元音或韵母(不包括带其他辅音韵尾的韵母)的系统中, 单元音仅占约 1/4, 而复合元音约占 3/4, 这样的现象在世界语言中不敢说绝无仅有也是极其罕见的。

为什么会如此呢? 实际上, 汉语拼音是一套表音方案, 但又不完全标音方案, 本质上是书面汉字的语音象征符号系统, 带有很强的应用性质。正是因为作为符号系统, 受到符号应用目的和规则的制约(如采纳拉丁字符, 便于教学等), 可以一定程度上偏离表音真实, 即所谓表音的“神似”。而且, 这样的“象征表音”也符合语言学(语音学)的音位(变体)归纳原则和“标音多能性”传统。为此, 我查看了一下有关汉语拼音的一些相关论述。例如, 王理嘉先生就汉语拼音跟语音学的关系说: 《方案》是在语音学和音系学的基础上制订的。从音位与语境变体的角度去梳理字母和语音的联系, 则两者之间关系是非常清楚而又有条理的。但用于注音识字和拼写话语的字母系统跟语音学、音系学中的标音系统,

性质并不完全相同。因此,拼音设计只要不违反音位归纳的基本原则,主要考虑的是字母与语音的配置关系。(《语言文字应用》2008年第3期)

显然,我们不可把汉语拼音看作语言描写和语言研究的范式 and 工具。目前的方言乃至少数民族语言复合元音的实际记录或描写状况很可能是受到汉语拼音方案的影响而逐步形成的,结果造成大量前滑式复合元音。如果真是如此,这个倾向应该加以纠正。我曾把受汉语拼音影响的描写方法称为“汉语拼音式方言描写体系”,为了纠偏,这里我针对前滑式复合元音作一点补充讨论。

前滑式复合元音的要点是带元音性介音。但是根据我们对多种方言的听辨,包括普通话,总体感觉很难听出真正的所谓元音性介音。孙景涛教授(2006)在“介音在音节中的地位”一文明确指出,除了闽语,其他汉语方言的介音不具有元音性。而针对粤语是否有元音性介音的争议,黄家敦(1964, 1994)、麦耘(1999)等教授均提出否定意见。不妨举个实例,“假”、“家”等字,吴、粤、赣、客、闽等南部方言一般记作 ka/ko,多数北方方言记作 tɕia,后者这个-i介音似乎就是受到拼音方案影响制作音系而产生的(当然还有历史音韵“等”的观念制约)。有意思的是,汉语“假”借到部分藏缅语之后,多数记音材料并不带-i借音,例如:撒尼彝语 tɕa<sup>11</sup>;嘎卓语 tɕa<sup>55</sup>;白语 tɕa<sup>21</sup>;少数材料带-i介音,例如土家语 tɕia<sup>55</sup>。前者显然遵循了共时语音记音操作,后者是真有元音性-i介音还是调查人受汉语拼音影响所导致尚不得而知。

从经验上看,世界语言复合元音一般是后滑类型为主(Carrie E. Lang & John J. Ohala, 1996),汉语那些所谓前滑元音大多相应于其他语言处理系统中的声母特征变体,譬如唇化和腭化。前些年,我曾布置一次作业,要求研究生按照实际音质填写普通话声韵表,结果音系中增加了部分唇化和腭化声母,无需设立任何前滑二合元音,去除了所有三合复合元音。这样的处理结果跟前述黄家敦和麦耘教授对无需前滑复合元音的广州话音系的精准描述相当一致。由此可见,汉语南北方言记音不添加无元音性介音的复合元音描写未尝不可。

阡子博士是个勤奋的学者,她在著作中提出了一个很重要的实践命题,即寻找语言的共时类型和历史发生类型之间的内在关系。她通过现代藏缅、苗瑶、侗台、南岛、南亚和汉语及其方言六大语群全面阐述了各语群复合元音主体模式以及各种复合元音的来源,包括单元音裂变主模式和韵尾向滑音转变形成复合元音、接触借用而出现的复合元音等次要模式。可以设想,这其中的工作量之大,处理

数据强度之高，非努力勤奋而难以实现。可以说，这部分内容凝结了作者参透和解析复合元音奥秘的精华，同时也包孕了可观的大数据，我相信这部分内容将给读者提供特别丰富的复合元音类型信息，值得细读。

是为序。

江 荻

2016年12月

# 前 言

截至 2016 年,国内和国际上系统、深入地研究复合元音的成果寥寥,以大规模语言数据为语料,从传统语言学和现代语言学的理论视角综合研究复合元音的成果比较缺乏,对复合元音的基本定义、语音性质、语音特征的精准描写及以复合元音的基本性质为参项划分语言类型及复合元音的起源和历史演变规律还有待深入挖掘。

笔者尝试系统解决关于复合元音在历史语言学及语音类型学中存在的某些理论问题,借鉴和采用多种前沿理论,包括现代语音学、现代音系学、传统音韵学、类型学和历史语言学,并结合数据库技术,从多个研究视角剖析和阐述复合元音的基本原理、类型和历史。

在研究方法上,本书以汉语群、藏缅语群、侗台语群、苗瑶语群、南亚语群、南岛语群 684 种语言和方言大数据为基础,建立语音和音系数据库,梳理关于复合元音研究的文献资料,从中发现存有争议的重要问题,把语言学基础研究与现代计算机技术和统计方法相结合。

本书是系统研究复合元音的一本综合性著作,其研究价值和研究意义在于通过分析大量的语言数据,以求找到各种语言类型,揭示语言在历史发展过程中每个阶段所经历的演化形式,克服个别语言研究个体问题的局限,从局部过渡到整体,希望对语言学的理论构建有所贡献。

本书的一项重要创新在于:发现了划分汉藏语语言类型的一个重要参项——复合元音的单音节特征。音节一向是历史语音学非常关注的焦点问题,学术界长久以来就在争论单音节和多音节对于语言历史演变的作用。本书找出复合元音的单音节特征是造成不同语群产生复合语言多少的重要根源:单音节特征强,如汉语,复合元音的数量多;反之,多音节特征强,如南岛语,复合元音的数量少。笔者以单音节特征为主,并结合复合元音的长短、松紧、鼻化、有标记等次要特

征，构建汉藏语语言类型的参项，为汉藏语言划分出有标记——无标记语言类型等级序列，为汉藏语言增加一项类型特征。另外，本书充分揭示了复合元音起源和历史演变的多种途径，包括辅音声母和辅音韵尾变化、单元音裂变、音节合并、发声态的影响、语言接触、单元音化等，构成若干历史音变模式，以此弥补了前人仅在个别语言中寻找线索而造成类型不足的缺陷。

限于笔者的研究时间和研究能力，本书还有不足之处。关于复合元音的类型和历时演变有很多问题需要深入讨论，比如汉语和民族语接触产生复合元音，笔者在思考如何采用汉语音韵学和历史语言学方法划分出历史层次；南亚语群中由于发声态的影响产生复合元音，笔者计划采集更多的第一手语言数据进行历史比较。这些重要的理论问题都要在未来继续研究解决。

田阡子

2016年8月

# 目 录

序言 (江荻)

前言

|                                |     |
|--------------------------------|-----|
| 绪论                             | 1   |
| 1 复合元音的基本原理                    | 11  |
| 1.1 复合元音的概念                    | 20  |
| 1.2 复合元音的性质                    | 23  |
| 1.3 真性复合元音的性质                  | 35  |
| 1.4 复合元音的分类                    | 37  |
| 2 复合元音在汉藏语中的语言类型               | 41  |
| 2.1 /ua/、/ia/、/ui/在汉藏语语群中的语言类型 | 42  |
| 2.2 /au/、/ai/、/iu/在汉藏语语群中的语言类型 | 55  |
| 2.3 汉藏语语群复合元音的语言类型             | 64  |
| 2.4 语言类型与地理分布                  | 66  |
| 3 复合元音的起源及发展                   | 69  |
| 3.1 复合元音与辅音之间的转化               | 71  |
| 3.2 单元音裂变构成复合元音                | 78  |
| 3.3 音节的演化与复合元音的产生              | 92  |
| 3.4 发声态影响产生复合元音                | 98  |
| 3.5 复合元音单元音化                   | 100 |
| 3.6 在汉语借词中产生的复合元音              | 102 |
| 3.7 语言类型与历史演变                  | 104 |
| 4 复合元音在汉藏语语言类型中的音变模式           | 105 |
| 4.1 南岛语群中的复合元音                 | 105 |
| 4.2 汉语群复合元音的音变模式               | 107 |

|                   |     |
|-------------------|-----|
| 4.3 藏缅语群复合元音的音变模式 | 112 |
| 4.4 侗台语群复合元音的音变模式 | 119 |
| 4.5 苗瑶语群复合元音的音变模式 | 127 |
| 4.6 南亚语群复合元音的音变模式 | 134 |
| 4.7 汉藏语中复合元音的音变模式 | 139 |
| 5 汉藏语数据与东亚人类的渊源   | 140 |
| 5.1 计算机技术采集语言数据   | 140 |
| 5.2 汉藏语分类与人类史前迁移  | 141 |
| 参考文献              | 144 |
| 附录                | 150 |
| 附录 1              | 150 |
| 附录 2              | 157 |
| 后记                | 179 |

## 绪 论

在 20 世纪,国内汉藏语学界对复合元音进行比较详细而深入的研究著作有《藏缅语语音和词汇》、《藏语韵母研究》。在《藏缅语语音和词汇》中有专门一章“关于复元音问题”,全面描写了复合元音的类型,提出复合元音的来源问题。《藏语韵母研究》重点讨论了藏语方言中复元音韵母的分布特征和历史特征。其他关于汉藏语言复合元音的研究大多包含在综合性描写论著或者单篇论文之中,如民族出版社出版的《中国少数民族语言简志丛书》,孙宏开主编的“中国新发现语言研究丛书”,《民族语文》和《方言》等期刊刊载的语言简介和汉语方言同音字表等论文。单篇论文的代表作有马学良和罗季光(1962)的《我国汉藏语系语言元音的长短》、孙宏开(1982)的《藏缅语若干音变探源》、孙宏开(2001)的《原始汉藏语中的介音问题——关于原始汉藏语音节结构构拟的理论思考之三》、瞿霭堂(1991)的《藏语韵母的演变》,这几篇文章从不同研究视角对复合元音或复元音韵母的状况及历史构成原因进行了理论性的探讨。

随着描写材料的积累和增多,人们对复合元音的认识逐步加深,研究的领域也逐渐扩大,近些年来,有些学者正在尝试运用新的方法寻找复合元音的某些性质和变化规律,如江荻(2002b)的《缅甸语复合元音的来源》,朱晓农(2004)的《汉语元音的高顶出位》,覃晓航(2004)的《南丹话单元音复化的条件》,刘镇发(2006)的《温州方言在过去一世纪的元音推移》。这些成果的共同特点是充分占有某个民族语或者某个方言的语言资料,引进现代语音学关于元音空间链移理论,从一个新的研究视角来重新审视单元音和复合元音的历史演变过程,获得了新的理论认识。在 20 世纪所取得的成就基础之上,学者们把现代语言学理论与语言事实很好地融汇起来,构建起适合描写汉藏语复合元音特点的理论体系,从而推动复合元音的研究向前发展。

截至 2016 年,单语言中有关复合元音的研究虽然已经为数不少,但是还没

有一项涉及汉藏语各个语群的综合性研究,而有关民族语言和汉语方言的资料已经有相当程度的积累,越来越多的新语言和新方言被学者们发现进而描写出来。面对如此丰富、如此宝贵的大数据,开展汉藏语复合元音综合研究的条件趋于成熟。基于这种认识,本书收集了已经发表的语言资料,在汉语群、藏缅语群、苗瑶语群、侗台语群、南岛语群、南亚语群六个语群范围之内,建立汉藏语言复合元音数据库,并通过数据分析划分汉藏语复合元音的共时语言类型,挖掘复合元音的历史来源及演化模式,尝试解决复合元音研究中存在的某些理论问题。

陆俭明(2010)认为,从科学研究的角度说,无论哪个学科,对事实的考察和挖掘固然重要,但它毕竟只是研究的基础,还未达到真正意义上的科学研究。真正意义上的科学研究,必须对考察、挖掘所得的事实及其观察到的内在规律作出科学的解释,并进一步从中总结出具有解释力的原则,而且升华为理论,能用这些原则、理论来解释更多的事实,从而使学科得以自立,得以发展。

本书的研究基础建立在复合元音数据库建设之上。笔者认为,在计算机科学技术飞速发展的今天,专门从事汉藏语研究的专家学者不辞辛劳深入田间进行语言调查,他们积累下来的丰富的研究资料应该用计算机技术保存下来,把这些资料转化成电子数据的形式,为每一种语言现象建立数据库。数据库的作用在于把一些零散分布在各种资料中的语言现象聚合起来,构成一定规模的大数据,建立统一的体系,让研究者看到这种语言现象的整体面貌,方便研究者查询、检索所需要的语言信息,并提取所需要的语言数据。这样做弥补了过去手工制作卡片费时费力的不足,同时,数据库还可以执行统计计算,以获取所需要的统计数值,这样就为最后的研究结论提供实证性的证据,为证实或者证伪提供了有力的支持和保证。

本书的数据库包括共时数据库和历时数据库两个母库。共时数据库是一个汉藏语音系数据库,数据处理和数据库建设程序并不复杂,主要包括以下几个步骤:第一,筛选数据。从大量公开发表的语言调查资料中提取出语言、方言、土语数据,语言的传统系属关系数据,语言的音位系统数据,这三项数据构成数据库的主体资源。同时,还要提取出相应的描述性信息,包括每一种语言或者方言的地理分布、文献资料来源、作者等。第二,使用 Excel 电子表格软件,在电子表格中建设数据库。建设的程序包括:从第一列开始,依次向后排列,在每一列中创建字段,这些字段依次是语言系属、语言/方言、地理分布、资料来源、作者、音

位系统。在每一行的相应字段里录入描述信息及语言数据,这样就构成了一个由行与列构成的二维数据表格。提取及录入这些信息的目的是了解音系中的每一类音位,例如辅音音位、元音音位、声调音位属于哪一种语言或方言,每一种语言或者方言的语言系属情况、地理分布情况如何,文献资料来源与作者主要用来提供复核信息,在校对的时候可以方便、快捷地找到原始资料。

数据处理程序包括如下几项:第一,对语言数据进行分类。笔者依据传统汉藏语系属分类方法<sup>①②</sup>分成汉藏语系、南亚语系、南岛语系三个语系。每个语系下面分成若干个语群,如汉藏语系包括汉语群、藏缅语群、侗台语群、苗瑶语群;每个语群再分成若干个次语群,如藏缅语包括藏语次语群、缅语次语群、景颇语次语群、彝语次语群、羌语次语群;每个次语群下再分成若干个语言和方言,如藏语次语群包括中部卫藏方言拉萨话、东部康方言巴塘话、北部安多方言夏河话、西部方言达拉克话。第二,从这些语言和方言的音位系统中提取辅音音位、元音音位、声调音位,对每个系统进行次分类。辅音分成复辅音和单辅音两类,单辅音依据发音部位和发音方法进行归类,例如,双唇清塞音对应[p]类;双唇浊塞音对应[b]类;齿龈清擦音对应[s]类;齿龈浊擦音对应[z]类;软腭鼻音对应[ŋ]类,等等。复辅音(包括腭化与唇化辅音)依据基本辅音及其与前置辅音、后置辅音的组合关系进行归类,例如,[pl]、[pr]、[pj]、[pw]都是以[p]为基本辅音带后置辅音的复辅音,他们在数据库里平行排列;[nk]、[sk]、[lk]、[zk]都是以[k]为基本辅音带前置辅音的复辅音,他们在数据库里平行排列。元音分成单元音和复合元音两类,单元音依据发音时舌位的高低、前后、圆展等性质来进行归类,例如,前高非圆唇元音对应[i]类;后低圆唇元音对应[ɒ]类,等等。复合元音分成二合元音与三合元音,二合元音与三合元音都依据主元音的舌位及与滑音的组合关系进行归类。例如,[ia]、[ua]都是以[a]为主元音,主元音在后的复合元音,在数据库里他们平行排列;[ei]、[eu]都是以[e]为主元音,主元音在前的复合元音,他们在数据库里平行排列。在声调系统中,民族语和汉语方言的调类系统有

① 孙宏开. 1998. 二十世纪的中国少数民族语言文字研究//刘坚. 二十世纪的中国语言学. 北京:北京大学出版社: 658.

② 在实际操作过程中,把汉藏语语言划分成汉语群、藏缅语群、苗瑶语群、侗台语群、南岛语群、南亚语群六个语群,研究范围限定在传统的汉藏语系、南亚语群、南岛语系之内,不包括阿尔泰语系和印欧语系。这样分类的目的是要回避系属争议。

一定差异,民族语的调类有高、中、低三类,调形有平、升、降三类,调类与调形组合成一类。例如,彝语次语群怒苏语有高平调,对应的调值是 44; 高降调,对应的调值是 53; 高升调,对应的调值是 24; 低降调,对应的调值是 31。汉语方言声调分成平、上、去、入四类,每一类分阴阳。例如,潮汕方言的潮阳方言阴平调的调值是 33,阳平调的调值是 55,阴上调的调值是 53,阳上调的调值是 313,阴去调的调值是 31,阳去调的调值是 11,阴入调的调值是 11,阳入调的调值是 55。在进行这样的分类之后,再把每个语言或者方言的辅音、元音、声调归并到相应的类别中去。图 0-1 是声调-单辅音数据库取样图。

这个汉藏语语言音位数据库是一个母库,可以从中任意生成若干子库,复合元音音位数据库就是其中的一个子库,样图如图 0-2 所示。

|   | A                | B     | C    | D    | E      | F  | G    | H    | I    |
|---|------------------|-------|------|------|--------|----|------|------|------|
| 1 | 语系/语族/语支         | 语言    | 地理分布 | 文献出处 | 作者     | 声调 | 双唇清塞 | 齿龈清塞 | 软腭清塞 |
| 2 |                  |       |      |      |        |    | p    | t    | k    |
| 3 | 汉藏语系/藏缅语族/藏语中部卫藏 | 西藏最西部 | 语音史  | 江孜   | 52/低平  | p  | t    | k    |      |
| 4 | 汉藏语系/藏缅语族/藏语中部卫藏 | 西藏最西部 | 语音史  | 江孜   | 5/14短元 | p  | t    | k    |      |
| 5 | 汉藏语系/藏缅语族/藏语东部康方 | 巴塘话属  | 语音史  | 江孜   | 19;短元  | p  | t    | k    |      |

图 0-1 汉藏语语言声调-单辅音数据库

|   | A        | B    | C     | D            | E  | F              | G     | H              |
|---|----------|------|-------|--------------|----|----------------|-------|----------------|
| 1 | 语系/语族/语支 | 语言   | 地理分布  | 文献出处         | 作者 | 二合元音           | 二合元音  | 二合元音           |
| 2 |          |      |       |              |    | ia             | ai    | ua             |
| 3 | 汉藏语系/侗台  | 壮语田东 | 广西的田东 | 国少数民族语言均加/梁敏 |    | i <sup>a</sup> | ai/ai | u <sup>a</sup> |
| 4 | 汉藏语系/侗台  | 壮语田林 | 广西田林  | 国少数民族语言均加/梁敏 |    | i <sup>a</sup> | ai/ai | u <sup>a</sup> |
| 5 | 汉藏语系/侗台  | 壮语凌云 | 广西凌云  | 国少数民族语言均加/梁敏 |    | i <sup>a</sup> | ai/ai | u <sup>a</sup> |

图 0-2 汉藏语语言复合元音数据库

复合元音数据库具有分类功能、统计功能、排序功能、查询和检索功能。关于分类功能,前面已经有所提及,首先把复合元音分成二合元音与三合元音两类,然后再依据主要元音与次要元音的组合顺序及组合关系分类,[i]、[u]在前一组,例如[ia]、[ua];[i]、[u]在后一组,例如[ai]、[au],类与类之间可以任意调整。关于统计功能,主要运用 COUNTIF (range, criteria) 函数来计算每一个复合元音的出现比例。例如,计算[ai]在复合元音数据库中出现的比例:首先,在[ai]列最末一行的统计字段中输入公式——=COUNTIF (fa: fb, “\*”),这个公式是计算[ai]在语言和方言中分布总数,即有多少个语言中有[ai]这个复合元音音位,这个结果再比上语言总数,然后再乘以 100%就会得出[ai]的出现比例。例如,图 0-2 数据库样例中的[ai],首先统计[ai]的出现总数——=COUNTIF (f3: f10, “\*”) = 5,即在 5 种语言和方言里有[ai]这个音位出现,一共有 9 种语言,因此,[ai]的出现比例是  $5/9 \times 100\% = 55.6\%$ 。关于排序功能,可以以任意一个字段为主要关键字进行升序或者降序排列,依据使用目的来决定排序的主要关键字。例如,如果了解[ai]在语言中的分布情况,就以[ai]为主要关键字,扩展到所有的区域进行排序,结果就是所有有[ai]出现的语言排在前面,同时相关的地理分布信息、语言系属信息、资料来源、作者等描述性信息都会出现,还可以看到[ai]与其他音位的关系,如与[ia]的关系,与[au]的关系,复合元音与单元音的关系,与辅音的关系等。关于查询和检索功能,可以在编辑菜单栏中打开查找与替换对话框,对数据库里出现的信息任意检索。

运用统计功能计算每个复合元音在汉藏语语言中的出现比例,可以对复合元音进行定量分析。以 50%作为分界线,在汉藏语语言复合元音数据库里,出现比例在 40%~60%的复合元音一共有七组,这七组复合元音分别是/ua-au/、/ui-iu/、/ia-ai/、/ue-eu/、/ie-ei/、/io-oi/、/uo-ou/。Maddieson (1984: 124-125) 用统计结果证实世界语言出现比例最高的五个单元音分别是/a/、/i/、/u/、/e/、/o/, 笔者的统计结果恰好证实了汉藏语语言出现比例最高的复合元音正是/i/、/u/与其他三个元音的组合(也包括/i/、/u/自身组合),笔者又从中选择了出现比例最高的/ua/、/ia/、/ui/和/au/、/ai/、/iu/两组复合元音作为本书重点分析的案例。

在复合元音共时数据库的基础之上,我们又建立起历时数据库。历时数据库的数据采集方法是从每一种语言或方言里提取语音变异资料。例如,壮语横县北话有这样一些资料:/i:/、/u:/与韵尾之间有一个[ə]过渡音;/e:/与韵尾之间有一个