

电脑辅助分析 语音与词汇

罗安源 廖乔婧 编著
叶典良 白丰霖 审订

DIANNAOFUZHU
FENXIYUYINYUCIHUI

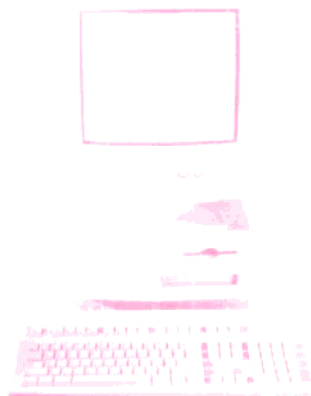


民族出版社

“211 工程”项目资助

电脑辅助分析 语音与词汇

罗云章 梁存清 编著
叶典俊 白丰霖 审订



民族出版社

责任编辑:孙 瑛

封面设计:金 一

图书在版编目(CIP)数据

电脑辅助分析语音与词汇/罗安源,廖乔婧编著.

北京:民族出版社,2001.12

ISBN 7-105-04764-X

I. 电… II. ①罗… ②廖… III. 元音识别-应用程序 IV. TN912.34

中国版本图书馆 CIP 数据核字(2001)第 087320 号

民族出版社出版发行

(北京市和平里北街 14 号 邮编 100013)

<http://www.e56.com.cn>

民族出版社微机照排 迪鑫印刷厂印刷

各地新华书店经销

2001 年 12 月第 1 版 2001 年 12 月北京第 1 次印刷

开本:850×1168 毫米 1/32 印张:9.875 字数:288 千字

印数:001—600 册 定价:22.00 元

该书如有印装质量问题,请与本社发行部联系退换

(总编室电话:64212794;发行部电话:64211734)

版权申明

本书的大部分重要内容，是解说世界少数民族语文研究院的两种语言分析应用程序的中文版和一部专著：其一是 Philip Brassett 开发的《WinCECIL》。这是《Computerised Extraction of Components of Intonation in Language》的缩称，中文名暂定为“语言音调提取器”；其二是 John Wimbish 和 David Coward 开发的《Shoebbox》。因为以往编词典要用手工记录大量的卡片，而卡片盒跟“鞋盒”的形状差不多，所以程序开发者就以“Shoebbox（鞋盒）”来称呼这个程序。本书没有将 Shoebbox 这个名称直译为“鞋盒”，而以“词典编纂器”作为这个程序的暂定中文名；其三是 David Coward 和 Charles Grimes 编著的《Making Dictionaries》（词典编纂）。这次选译了其中适用于中国用户的核心章节。上述程序和著作的中文版由廖乔婧翻译、罗安源校对。在本书出版之前，我们明确得到原作者及世界少数民族语文研究院计算机部的许可。

一、世界少数民族语文研究院计算机部来函之一

Professor Luo Anyuan

We are happy to grant you permission to include a number of sections of our publication, “Making Dictionaries” for the book you are now writing that covers the Chinese version of Shoebbox. Please include an acknowledgement that states: **Portions of the publication “Making Dictionaries” (1995) have been included by permission from SIL International** (使用《词典编纂》一书中文版的若干章节已经得到世界少数民族语文研究院的许可)。

Bill Mayes

VP Computing

JAARS

PO Box 248

Waxhaw, NC 28173

USA

(704)843-6151 bill_mayes@sil.org

二、世界少数民族语文研究院计算机部来函之二

Professor Luo Anyuan

We are happy to grant you permission to include a portion of our software publication, "WinCECIL" for the book you are now writing that covers the Chinese version of Shoebox. Please include an acknowledgement that states: **Portions of the software publication "WinCECIL" have been included by permission from SIL International** (使用《语言音调提取器》软件中文版的若干章节已经得到世界少数民族语文研究院的许可)。

Bill Mayes

VP Computing

JAARS

PO Box 248

Waxhaw, NC 28173

USA

(704)843-6151 bill_mayes@sil.org

今特在此郑重声明，上列两函明确了两种程序及著作中文版版权的授受关系，我们对世界少数民族语文研究院计算机部及有关开发者和作者表示衷心的感谢。我们相信，这两种程序和著作中文版的正式发表与推广，将大大有助于我国少数民族语文工作者的词典编纂工作。2001年5月底至6月初，世界少数民族语文研究院与中国社会科学院民族研究所合作举办的“计算机词典编纂研讨会暨培训班”，就是应用“词典编纂器《Shoebox》”的良好开端。

罗安源

2001年8月于

北京西郊魏公村

中央民族大学靖园



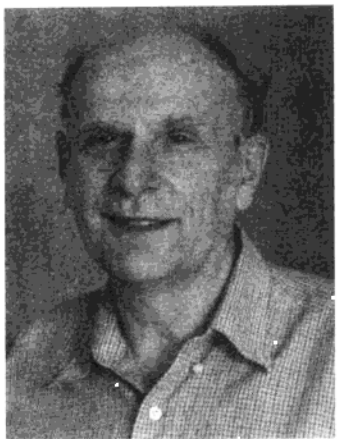
罗安源 湖南人，土家族，中央民族大学少数民族语言文学学院教授。曾任中央民族大学（原中央民族学院）汉语系副主任、《学报》主编、研究生部主任、中国教育学会少数民族教育研究会副秘书长；现任中国少数民族双语教学研究会副理事长。多年从事“语言学”、“语音学”、“语法学”、“少数民族语言文学”“汉语言文学”、“双语教育”、“电脑语言学”的教学与研究，享受政府特殊津贴。专业教材和科研著作有《语言学概论》

（合著，1981）、《语言调查研究讲座》（合著，1986）、《简明实用语音学》（合著，1990）、《现代湘西苗语语法》（1990）、《简明现代汉语语法》（1996）、《电脑语言学基础》（合著，1998）、《现代汉语简编》（2000）、《田野语音学》（2000）、《土家人和土家语》（合著，2001）等，并发表学术论文多篇。其中的个人重要论述如“音区论”，明确将人类各民族语言的元音划分为21个音区；“苗语句法成分的可移动性论”等，凸显了湘西苗语的语法特色；“白语声门混合紧擦音论”，阐释了白语的特殊音质；“多语一通、三语并重、汉语牵头论”等，提出了西部地区大开发中语言教育任务的具体建议；前此出版的《电脑语言学基础》和此次推出的《电脑辅助分析语音与词汇》，是新兴学科电脑语言学的基础建设工程。所录制的《国际音标发音练习》（1983）和所编制的《国际音标电脑教学片》（1998-2001），为语音的教学与研究演示了简明、实用、有效的方式方法。



廖乔婧 (Cecilia Brassett) 女, 出生于香港, 英国国籍。1981 年赴英国剑桥大学学习, 1984 年获优等学士学位, 1986 年获医学学士学位, 1988 年获硕士学位, 随后, 又在英国剑桥大学获专业硕士学位。1997 年曾经加入“世界少数民族语文研究院”。廖

乔婧女士与丈夫白丰霖 (Philip Brassett) 于 1999 年来中国, 受中国社会科学院民族研究所、中央民族大学、中国农业科学院之聘, 从事大量的汉英翻译、编辑和英语教学等工作, 并曾撰写中国境内语言简介和有关世界濒危语言问题的文章。成果有《妇女与发展》、《经济、文化、环境》(第二册), 中国社会科学院民族研究所的“中国新发现语言研究丛书”之《毕苏语研究》, 以及联合国教科文组织的《世界语言报告》的中国部分。2000 年 7 月参加在北京举行的中国少数民族双语教学研究会第九届学术研讨会暨首届国际双语教学研讨会。2001 年 5 月为中国社会科学院民族研究所和世界少数民族语文研究院举办的《中国少数民族语言词典编纂培训班暨研讨会》担任电脑培训的工作。2001 年 7 月受聘为湖南省吉首大学中文系教授, 准备到湘西地区从事土家族和其他民族语言文化的调查研究工作。



叶典良 (Daniel James Edwards)

美国北卡州人。在计算机领域已有 40 余年的工作经验。1957 年即在美国麻省理工学院 (MIT) 开始了电脑程序编制的学习与研究。1959 年大学毕业后, 参与了麻省理工学院人工智能项目的开发, 并担任至今仍广泛应用于人工智能领域的“LISP”程序语

言的主要程序员之一, 同时参加麻省理工学院的“CTSS”和“MAC 分时电脑系统”项目的工作。1967 年进入美国“国家电脑中心”(The National Computer Center), 担任项目主管人和几项“电脑与电脑通信”项目的首席工程师。在技术上的突出贡献有已经获得美国一项专利的将电话用做电脑输入装置的方法以及保护电脑免受黑客进攻的方法。他是一位进行个人电脑革命的“先锋”(early adopter)。1978 年他就用市场上的电脑零件, 装备了他自己的第一台个人电脑。从 1986 年起, 供职于“JAARS 电脑服务中心”, 担任应用程序员及软件部主任。在此期间, 帮助开发了供专业语言学专家使用的电脑程序。他的几个项目是: 改进《一致性转换程序》、编制输入复杂书写符号的专用软件、完善“语言学家数据库程序《Shoebox 2》”。1994 年来到中国, 受聘为中央民族大学的客座教授, 在中央民族大学与世界少数民族语文研究院合办的“康赛 CUNSL 电脑语言学研究”中心讲授电脑和语言学的课程, 并帮助师生们解决电脑硬件和软件的各种问题。他工作一贯勤勤恳恳, 有求必应, 不计名利, 不怕困难, 受到中央民族大学师生的热烈欢迎。



白丰霖 (Philip Brassett) 出生于英国伦敦。1979 年到剑桥大学学习, 1982 年获得优等学士学位, 随后获得该大学的硕士学位和博士学位 (1987 年), 博士论文涉及小麦的一于是种根病的电脑模拟。1987 年开始, 专门致力于为“世界少数民族语文研究院”

(SIL International) 开发了语音分析程序《WinCECIL》, 1997 年加入该院从事语言研究工作。白丰霖于 1999 年来中国北京, 受中央民族大学与中国农业科学院之聘, 从事电脑语言学、英语语言文学的教学研究工作。在中央民族大学少数民族语言文学学院语言学系亲自讲授他开发的《WinCECIL》的构成与应用。曾为联合国教科文组织的《世界语言报告》的中国部分设计语言分布地图。2000 年 7 月参加在北京举行的中国少数民族双语教学研究会第九届学术研讨会暨首届国际双语教学研讨会; 2001 年 5 月为中国社会科学院民族研究所和世界少数民族语文研究院举办的《中国少数民族语言词典编纂培训班暨研讨会》担任电脑培训工作, 同时开发了 POSH 排版打印软件。2001 年 7 月白丰霖受聘为湖南省湘西土家族苗族自治州吉首大学中文系教授, 准备到湘西地区从事土家族和其他民族语言文化的调查研究工作。

序 言

自从 1981 年世界上第一台个人电脑问世以来，个人电脑的迅猛发展和普及，不但改变和丰富了人们的社会生活，也在各门学科的教学与研究中起到了惊人的促进作用。“电脑语言学（或称‘计算语言学’）”的建立与发展，就是电脑科学与语言科学结合的崭新学科。

中央民族大学有鉴于此，于 20 世纪 90 年代初期与“世界少数民族语文研究院（SIL International）”共同创立了“康赛 CUNSL 电脑语言学研究中心”，为利用电脑开展中国的少数民族文研究开辟了一条新的途径。世界少数民族语文研究院拥有众多高水平的电脑科学家和语言学家，成功开发了一批适用于世界各民族语文研究的电脑软件。中央民族大学拥有众多高水平的民族语文学家，多年来取得了民族语文研究的丰硕成果。两个单位真诚合作，形成强强联手，为电脑语言学在中国民族语文研究领域播种、开花、结果，提供了肥沃的园地。从 1993 年至 1997 年，中央民族大学少数民族语言文学学院语言学系连续开办了几期电脑语言学研讨班，首先着眼于培养电脑语言学的教学研究骨干。从 1998 年起开办电脑语言学专业，电脑语言学正式进入本科和研究生的课堂。将近十年来，我校电脑语言学的教学人材不断成熟，电脑语言学的教学设备不断充实，电脑语言学教学研究的成果不断涌现。

中央民族大学少数民族语言文学学院为了适应教学研究的需要,于1998年出版了由罗安源教授主编的《电脑语言学基础》。该书受到业内学者的广泛欢迎,被认为是一部具有开创性,并具有理论意义和实用价值的教材。经过近三年的继续努力,在世界少数民族语文研究院几位专家的协助下,罗安源教授等又引进了世界少数民族语文研究院的电脑语言学的最新研究成果,结合中国少数民族语文研究的实际,编纂了新著《电脑辅助分析语音与词汇》一书。该书除了介绍利用电脑分析语音与词汇的有关理论方法之外,还详细解说了语音分析软件《WinCECIL》和词典编纂软件《Shoebox》的结构原理和操作流程。我们相信,该书的出版,将大大有利于少数民族语言的语音与词汇的教学研究,对于其他邻近学科的教学研究也会有普遍实用意义。

在祝贺《电脑辅助分析语音与词汇》正式出版的同时,我们也热诚地期待将有更多更好的电脑语言学教学研究成果问世。

中央民族大学
少数民族语言文学学院

二零零一年九月一日

前 言

让电脑语言学真正进入大学课堂，在我国是开展得比较晚的一项工作。20 世纪 90 年代初，世界少数民族语文研究院(SIL International)与中央民族大学少数民族语言文学系，在北京合作建立了一个专门机构“康赛电脑语言学研究(中心) (The CUNSL Computer Centre for Linguistic Research)”。该院先后派出白默翰(John Alsop)、叶典良(Daniel James Edwards)、狄乐伦(Lon Diehl)、白丰霖(Philip Brassett)、廖乔婧(Cecilia Brassett)等几位电脑语言学专家来协助工作，历时将近 10 年，培养了一批电脑语言学人材，获得了不少研究成果。前几年，中央民族大学少数民族语言文学系招收了几批电脑语言学研究方向的硕士生，近年又陆续开办了电脑语言学本科专业班。除了中央民族大学以外，云南民族学院等院校也正在大力开展电脑语言学的教学与研究。可以说，在民族高等院校中，电脑语言学的教学与研究正方兴未艾。为了适应教学的需要，我们于 1998 年出版了《电脑语言学基础》，如今又推出了《电脑辅助分析语音与词汇》。

《电脑辅助分析语音与词汇》包含以下内容：一部分是与分析语音、词汇相关的基本理论、方法或专著，一部分是与分析语音、词汇相关的电脑软件。前一部分主要是几位专家讲课的记录或专著的摘译。后一部分是有关软件的解说。这些软件和著作是：WinCECIL (Philip Brassett)、《Shoebbox》(John Wimbish & David Coward)、《Making Dictionaries》(David F. Coward & Charles E. Grimes)。现在征得上述各位专家的同意，

并经他们的主管机关“世界少数民族语文研究院计算机部”正式授权,我们将这些内容统编成一部教材,其中的基本内容在中央民族大学少数民族语言文学学院“教育部文科基地班”的“九八级电脑语言学专业班”讲授过。其中《Shoebbox》的主要内容于2001年6月在世界少数民族语文研究院与中国社会科学院民族研究所合作举办的“计算机词典编纂研讨会暨培训班”作了介绍。

全书的内容由叶典良教授、白丰霖博士审订。摘译工作是由廖乔婧负责的,译文校对、专家讲课记录、内容增补和书稿统编工作是由罗安源负责的。所谓“摘译、增补、统编”,并不是简单地照译照统,而是根据中文用户和教学对象的习惯和特点,在内容上有所选择,在材料上有所补充,在结构上有所调整。当然,译、增、编中的问题在所难免,欢迎业内外人士提出批评。

在译编过程中,也适当参考了中外时贤的一些论著,如:

《The Acoustic Analysis of Speech》(Ray D. Kent, Ph.D. and Charles Read, Ph.D.)、《Computers and Human Language》(George W. Smith)、《A Course in Phonetics》(Peter Ladefoged)、《实验语音学概要》(吴宗济、林茂灿等)、《计算语音技术》(朱民雄)、《普通语音学纲要》(罗常培、王均)等。在此衷心感谢这些论著的作者。

本书由中央民族大学少数民族语言文学学院出资付梓。院、系两级领导人给予了充分的重视,特此铭记。

编著者 谨识

二零零一年仲夏于

中央民族大学少数民族语言文学学院
康赛 CUNSL 电脑语言学研究中心

目录

版权申明

作者简介

序言

前言

上编 语音分析

第一章 电脑与语音

第一节 言语的传递 (3)

第二节 电脑语音学与其他学科的关系 (4)

第二章 声音的发生与传播

第一节 没有空气就没有声音 (6)

第二节 没有耳朵就没有声音 (7)

第三节 声波与空气粒子的关系 (8)

第四节 语音表达思想 (10)

第五节 独特的人类发音器官 (11)

第三章 声波的特点

第一节 频率、波长、音速 (12)

第二节 复合音与纯音 (13)

第三节 振动与位移 (15)

第四节 物体质量与音高的关系 (16)

第五节 频率、波长、音速的关系 (17)

第四章 声谱的模式

第一节 声波的品质 (19)

第二节 声谱的构成 (21)

第三节 节点与复合波声谱模式 (22)

第五章 重复声波与非重复声波

第一节 不同声音的和弦 (24)

第二节 重复声波与非重复声波的区别 (25)

第三节 声波的衰减率 (26)

第四节 白噪音 (27)

第六章 声音的感知与共鸣

第一节 人们感知音高与实际音高的差异 (29)

第二节 共鸣的形成 (30)

第三节 声音的频率与声音衰减的关系 (34)

第七章 一物与多物的共鸣

第一节 句子的音调频率高峰 (36)

第二节 共鸣体对发音体做出反应 (37)

第三节 物体产生共鸣响应所需的时间 (38)

第四节 一个物体对多个频率的共鸣 (39)

第五节 共振峰与倍音的关系 (40)

第六节 带宽与共鸣的关系 (41)

第八章 元音与共振峰

第一节 共振峰与元音舌位的关系 (43)

第二节 从声线谱看辅音向元音的过渡 (45)

第九章 声源与共鸣

第一节 声源 (47)

第二节 共鸣与共鸣器的质量 (49)

第十章 声道与音高

第一节 声道对声源的共鸣 (52)

第二节 声带的开合与空气压力的关系 (53)

第三节 非周期波 (55)

第四节 声调的实质 (57)

第十一章 辅音的声学特征

第一节 辅音的声学分类 (59)

第二节 噪音线 (60)

第十二章 共振峰与辅音

第一节 共振峰显示辅音特征 (63)

- 第二节 塞擦音的特征 (64)
- 第三节 带音辅音的特征 (65)
- 第四节 边音和半元音的特征 (67)
- 第十三章 响度与音强**
 - 第一节 响度的衡量 (69)
 - 第二节 音强的衡量 (70)
- 第十四章 声谱图样式**
 - 第一节 浊音条样式 (72)
 - 第二节 直切线样式 (72)
 - 第三节 间断区样式 (73)
 - 第四节 共振峰条带样式 (74)
 - 第五节 噪音样式 (74)
 - 第六节 噪音+共振峰样式 (75)
- 第十五章 WinCECIL 的应用**
 - 第一节 引言 (76)
 - 第二节 主要应用程序窗口 (77)
 - 第三节 声段 (79)
 - 第四节 声学数据: 查看 (80)
 - 第五节 声学数据: 解释 (83)
 - 第六节 振幅 (87)
 - 第七节 查看基频 (88)
 - 第八节 基频: 解释 (91)
 - 第九节 频谱图与声谱数据 (94)
 - 第十节 重放 (96)
 - 第十一节 录音 (98)
 - 第十二节 音标字符 (101)
 - 第十三节 屏幕布局 (102)
 - 第十四节 瞬像图框 (104)
 - 第十五节 输出: 图像 (105)
 - 第十六节 输出: 打印 (107)
 - 第十七节 其他项目 (108)

中编 词语分析

第一章 词语的选择

- 第一节 收集词语的方法 (113)
- 第二节 比较基于词根与基于词汇单位的数据库 (114)
- 第三节 基于词根的数据库 (116)
- 第四节 基于词汇单位的数据库 (117)
- 第五节 两种方法的比较 (118)
- 第六节 可行的综合方法 (119)
- 第七节 数据库结构 (120)

第二章 中心词和例句的选择

- 第一节 选择中心词的原则 (122)
- 第二节 选择例句的原则 (126)

第三章 词义与词汇功能

- 第一节 同音异义词与多义词 (129)
- 第二节 词义范畴(\sd, \th, \is) (132)
- 第三节 方言信息 (132)
- 第四节 词汇功能 (133)

第四章 有关特别词义范畴的条目

- 第一节 条目的分类法 (137)
- 第二节 植物条目 (139)
- 第三节 动物条目 (141)
- 第四节 关于人的条目 (145)
- 第五节 物质文化条目 (147)

第五章 句法类型及词类分配问题

- 第一节 活动与事件 (149)
- 第二节 状态与过程 (150)
- 第三节 借词与词源 (150)
- 第四节 处理用于仪式的词汇与其他特殊的语境 (151)
- 第五节 词类的分配 (152)
- 第六节 确定词类的基本原则 (152)
- 第七节 原则与实际上的差异 (153)
- 第八节 要特别注意的问题 (155)