

章宜华 著

计算词典学 与新型词典

*Computational Lexicography
and New-Typed Dictionaries*

本书展现和评析

计算机对词典的支持
各种电子词典的特点

上海辞书出版社

计算词典与新型词典

*Computational Lexicography
and New-Typed Dictionaries*

上海辞书出版社

计算词典学概略

语料库与语料库词典学

词库的研究与开发

新型的英语学习词典

机器词典介绍

光盘电子词典评析

掌上电子词典评析

网络词典介绍

图书在版编目(CIP)数据

计算词典学与新型词典/章宜华著. —上海:上海辞书出版社,
2004. 6

ISBN 7-5326-1493-X

I. 计... II. 章... III. 电子词典—研究 IV. H085-5

中国版本图书馆 CIP 数据核字(2004)第 010564 号

计算词典学与新型词典

世纪出版集团 出版、发行
上海辞书出版社

(上海陕西北路 457 号 邮政编码 200040)

上海华成印刷装帧有限公司印刷

开本 850 × 1168 1/32 印张 14.75 插页 1 字数 383 000

2004 年 6 月第 1 版 2004 年 6 月第 1 次印刷

印数 1—3 100

ISBN 7-5326-1493-X/G · 514

定价: 33.00 元

序

黄建华

别人索序的,我得先通读原书才能执笔,多少感到肩上多一个额外的任务,因为那种翻翻标题或看看目录就能下笔成序的本事,我是不具备的。而章宜华教授这本《计算词典学与新型词典》是我必须优先阅读的著作,阅后为文,似乎顺理成章。不过章教授的书,在“词典学”类的著作中,分量可不算轻,开卷时,不免有点“硬着头皮”之感,但读着读着,一股欣悦的心情却油然而生。花了几天功夫读毕,“序”已成竹在胸,随即坐到计算机旁边,个人的印象或感想便随手指对键盘的敲击而变成了文字。

本书给我的第一个印象是新。这“新”既反映在观念上也表现在材料上。作者正确地指出:“毫不夸张地说,计算机给词典学带来了一场革命。”他敏锐地把这场“革命”的脉搏摸准了,而且为我们描画了一个令人向往的图景。他所用的资料近至2003年上半年,这正反映出作者紧紧跟上了信息时代的快速步伐。当然,写进书中的某些内容不久也会过时而归于无用,但始终力求站在当代学科前沿的精神,那是永存的。

我的第二个印象是全。读者只要略翻一下本书的章节标题便可知道,在书名所涉及领域的范围里,作者几乎是“面面俱到”(撇开此成语可能包含的贬义)。如果进一步再读书中的内容,就会发现资料丰富而翔实。作者对许多电子词典都亲自作过考察,他所提供的不是人云亦云的“二手货”。例如,他对《牛津剑桥基础辞典》中“笨蛋”的“笨”错为“苯”,也指了出来。至于一些电子词

典的浮夸的收词数字,鉴于读者无法目测“部头大小”而易受迷惑,他更以内行人的坚实理据而加以辨析和提示。全面掌握相关的材料,这是开展一项研究工作的基本功,可以说,这一点作者是真正做到了。

第三点我觉得是尤为难得的:作者具备深厚的语言学 and 词典学的理论素养,这使得本书的阐述处处流露出引人思考的理论智慧。可以说,这是一本具有纵深度、富于“立体感”、理论与实际相结合的力作,而不是那种堆砌材料、平铺直叙、令人一览无余的就事论事的著述。

1996年我在谈及网络词典的时候这样说过:“一种新技术出现之后,以这种技术为依托的工艺或艺术必将随之而变化。例如,音响设备进步之后,同声传译便发展起来,于是就产生同声传译的理论和实践技巧。又如,电影的发明,并不仅仅意味着把舞台戏拍在胶片上,电视连续剧也不等于长篇电影的分段、分集。可以预期,互联网络词典出现之后,词典的编纂原则和方法乃至某些基础理论都将会发生若干重大变化。”(见《新词典 新观念——关于网络词典的思考》,载《辞书研究》1995年第6期)章宜华更把这种信息技术介入之后的“变化”视为“革命”。就目前国内所见,他的这本《计算词典学与新型词典》便是展示这场“革命”的最新、最全而又最有词典学理论深度的著作。希望经过他这一番的总结和展示,能引起读者和有关行家的注意,今后大家携起手来,共同努力,将这场“革命”推向一个新的高潮。

行文至此,我对本书都只是赞赏之词。为了消除“净说好话”的嫌疑,我还是指出一点“瑜中之瑕”吧:本书在谋篇布局方面对理论与实践部分的安排还可以更紧凑一些,在遣词造句方面还可以去掉“欧化”痕迹,使之更精练一点。

前 言

本书最初的设想源于广东省“九五”重点理论规划项目——90年代新型词典研究。该项目主要想通过阐述现代词典学的理论基础,探讨20世纪90年代出现的各种电子词典对词典学编纂理论和实践的影响,弄清电子词典的类别、结构特征和基本结构单元,包括词典的“内部”知识结构和“外部”操作界面的结构等内容。参加该项目部分研究的有赵彦春、源可乐、麦志强、王小海、李茜等同志。在对电子词典研究的过程中,笔者对计算词典学的概貌有了大致的了解,从而萌发了进一步研究的念头。

通过广泛调研,我们清楚地认识到计算词典学有三大任务:(1)实现词典编纂和修订的数字化或自动化;(2)开发供机器使用的词典数据库;(3)开发供人使用的词典数据库。围绕这三大任务又有大量的工作要做,比如语料的搜集、整理、标注处理和建库储存,机读词库对词汇数据的自动标注和理解,词典的计算机辅助编纂系统和自动生成系统的开发和利用。可见,计算词典学研究的内容是一项复杂的、跨学科的系统工程。我们词典界有责任对我国和西方主要国家的计算词典学的实践活动进行系统调查和梳理,弄清计算词典学得以形成的理论基础和社会基础以及其研究方向和研究任务,以期建立计算词典学的理论框架,促进我国词典研究和词典编纂实践尽快地与现代数字技术相结合。

在系统研究的同时,我们还想在计算词典学的某一方面有所突破。经过细致的调查研究和充分论证,我们申报的“基于语料库的WEB词典编纂&自动生成系统”又得到广东省社科规划项目的立项批准,并获全额资助。因此,本书的内容概括了两项科研

成果的相关内容,吸收了已结项的“90年代新型词典”的相关部分内容,并进行了增删、修改和完善,同时还大量采纳了新课题的研究成果,形成了现在这部较系统地论述计算词典学和新型词典的专著。全书共十二章,主要内容如下:

第一章 计算词典学概略 论述计算词典学产生和发展的社会基础及理论基础,包括计算词典学的起源、特色以及与计算语言学等相关学科的关系。

第二章 语料库与语料库词典学 系统介绍语料库的基本特征,语料库的类型及其基本功能,国内外语料库建设与发展的历史情况以及未来发展的趋势。

第三章 词库的研究与开发 首先对词库的理论进行简单的概说,然后简要介绍词库的相对关系、词库的基本内容和特色,最后对 WordNet、FrameNet 和 HowNet 等国内外具有代表性的词库作比较系统的说明。

第四章 现代化技术与新型词典的编纂 论述语料库与词典编纂的关系,探讨语料库的辅助词典编纂功能,并介绍朗文、牛津和柯林斯等出版社基于语料库编纂的各类新型词典。

第五章 电子词典的类型与功能设置 首先阐明电子词典的定义,区分电子词典的类别,然后概括说明电子词典的功能设置。

第六章 电子词典的内部结构 揭示一般电子词典和机用电子词典的结构特点,包括词典的整体结构、词汇知识库结构、词汇知识信息内容以及这些信息的表述方法和内部编码语言的应用。

第七章 电子词典的界面结构 总结并概括电子词典的界面功能设置、界面结构类型,包括界面与检索入口、界面与窗口形式、界面与窗口的多媒体功能,并分析了部分精品词典的界面。

第八章 光盘电子词典 简要介绍欧美和我国(包括内地和港台地区)出版的光盘词典的一般特点、个性特色、界面与查询以及不足之处。

第九章 网络电子词典 概括介绍“Onelook Dictionaries”和“YourDictionary”等专门网站词典的界面特点、功能设置、词典类别和检索系统,以及英国朗文、剑桥和法国拉鲁斯、哈歇特等著名出版社的网络版词典的主要特色。

第十章 微型芯片电子词典 从词典的知识功能、信息处理功能、附加功能、软件配置等几个方面,简要介绍“快译通”、“好易通”、“文曲星”等十多个品牌的掌上电子词典。

第十一章 电子词典编纂的一般原则 结合电子词典与印刷本词典的区别,论述电子词典编纂的一般原则,包括收词原则、拼写和注音原则、义项划分及释义原则、例证选用原则、多媒体原则和孳生原则等,突出了电子词典的优势特色。

第十二章 电子词典的发展趋势与结构模块 预示并阐述电子词典发展的趋势,并对词典界面结构进行了模块化切分,提出了几十种模块结构,以探索新型、理想的多媒体电子词典的结构框架。

通过上述十二章的阐述,本书大致界定了计算词典学的性质、任务及其研究范围,以利辞书工作者根据实际情况确定自己的研究方向,切实推动计算词典学在我国的发展。

本书的写作参阅了大量中外文献,主要参考文献在书后列出了目录,若有疏漏,在此谨表示歉意。此外,本书还参考了 WordNet、FrameNet、HowNet 网站的原始词库资料和相关文献,也参考了朗文、牛津、柯林斯、剑桥、拉鲁斯、哈歇特、阿特拉斯等出版社网站和一些掌上电子词典网站上未署名的数据资料,在此谨表谢意。

本书在撰写过程中得到国内同行和在读研究生的大力支持。源可乐教授和麦志强先生撰写了“光盘电子词典”的初稿,本书采用了其中的大部分内容;在职博士研究生王小海提供了一百八十多页有关网络词典和计算词典学的资料,本书第一章和第十章的个别段落采用了其中部分数据;夏立新老师和在读博士研究生

前 言

陈伟对全书作了认真的文字校对,还提出了一些修改意见;雍和明、亢世勇、曾用强等教授和本校计算机系副主任朱建华老师分别认真审读了本书的相关部分,并提出了宝贵意见;还有石亚丽同志也协助做了许多文字和数据整理工作。在此谨向他们表示衷心感谢。

章宜华

2003 年仲夏
于广州白云山下

目 录

序.....	1
前言.....	1
第一章 计算词典学概略.....	1
第一节 计算词典学的基础.....	1
一、计算词典学的起源.....	1
二、计算语言学.....	4
三、语料库语言学.....	7
四、计算词汇学.....	10
第二节 计算词典学的研究特色.....	12
一、计算词典学的特色.....	12
二、语料库词典学的特色.....	15
三、计算词典学与相关学科.....	17
第三节 计算词典学与机器词典.....	19
一、机器翻译的研究.....	19
二、机器词典的开发.....	24
三、机器词典与机读人用词典.....	28
第四节 计算词典学的国际环境.....	30
一、国际学术环境.....	30
二、语篇与词汇编码的国际标准.....	35
三、与计算词典学相关的国际学术组织.....	36
第二章 语料库与语料库词典学.....	42
第一节 语料库概述.....	42
第二节 语料库的基本特征.....	43

一、语料库的目的性	43
二、语料库语料的真实性	44
三、语料库语料的典型性	45
四、语料库语料的机读性	46
五、语料库语料的标准性	47
第三节 语料库的类型	50
一、按语言种类划分	51
二、按语言方式划分	51
三、按语料的载体划分	52
四、按语言的时效划分	53
五、按语言的应用划分	53
六、按语料的流通划分	54
七、按语料的分布划分	54
八、按语料的处理划分	54
第四节 语料库的基本功能	55
一、语料管理功能	55
二、语料索引功能	56
三、语料统计功能	57
四、语料标注功能	58
五、语音分析功能	60
第五节 语料库的建设与发展	61
一、语料库的历史背景	61
二、语料库的发展阶段	65
三、国内语料库的繁荣发展	68
四、语料库发展趋势	76
五、语料库发展的“轨迹”	79
第三章 词库的研究与开发	80
第一节 词库的理论概说	80
第二节 词库的相对关系	83

第三节 词库的基本内容和特色	85
第四节 主要词库的性能特征	88
一、WordNet 词库	89
二、FrameNet 词库	97
三、知网(HowNet)	113
第四章 现代化技术与新型词典的编纂	128
第一节 概述	128
第二节 语料库在词典编纂方面的功能	130
一、提供词典例证	132
二、辅助词典释义	133
三、辅助义项划分	133
四、辅助词典立目	134
五、辅助语用注释	134
六、提供句法信息	134
七、辅助语法注释	135
八、揭示文化信息	135
九、揭示同义词的分布与使用	136
十、词典自动生成	136
第三节 计算机辅助编纂的新型词典	137
一、“牛津系列”词典	138
二、“朗文系列”词典	143
三、“柯林斯 COBUILD 系列”词典	151
第五章 电子词典的类型与功能设置	158
第一节 电子词典概说	158
一、电子词典的发展简况	158
二、电子词典的定义	159
第二节 电子词典的类别	160
一、机器词典	160
二、微型芯片电子词典(掌上电子词典)	164

三、光盘电子词典	167
四、网络电子词典	170
第三节 电子词典的主要功能	172
一、f1:信息查询功能	172
二、f2:媒体浏览功能	174
三、f3:信息显示功能	175
四、f4:发音功能	176
五、f5:翻译功能	177
六、f6:写作助理功能	178
七、f7:汉化功能	178
八、f8:系统设置功能	179
九、f9:人机互动功能	181
十、f10—f15:其他功能	182
第六章 电子词典的内部结构	184
第一节 电子词典的内部结构特点	184
一、一般电子词典的结构特点	184
二、机用电子词典的结构特点	189
第二节 电子词典知识库的编码语言	191
一、标准通用标记语言(SGML)	192
二、可扩展标记语言(XML)	193
三、可扩展样式表语言及其转换	194
四、文件类型定义(DTD)	195
第三节 电子词典知识库的信息编码	196
第四节 电子词典词汇知识的形式化表述	204
第五节 电子词典的知识信息	209
一、词汇内容	211
二、基本信息	212
第七章 电子词典的界面结构	238
第一节 界面形式和功能区	239

一、界面形式·····	239
二、界面的功能区·····	240
第二节 界面结构类别·····	241
一、界面与检索入口·····	241
二、界面窗口形式·····	244
三、界面窗口性能·····	247
四、界面窗口的多媒体功能·····	248
第三节 精品词典界面赏析·····	249
第四节 对电子词典界面特征的归纳·····	254
第八章 光盘电子词典·····	257
第一节 欧美国家的光盘词典·····	257
一、《牛津高阶英汉双解词典》·····	257
二、《梅里亚姆-韦伯斯特大学电子词典》·····	261
三、《牛津剑桥基础词典》·····	265
四、《美国传统有声词典》·····	268
五、《朗文交互式英语词典》·····	271
六、《兰登书屋足本电子词典》·····	276
第二节 我国台湾地区的光盘词典·····	279
一、《21 世纪多媒体英汉双向辞典》·····	279
二、《远东英汉百科大辞典》·····	281
三、《超级莱思康光碟字典》·····	286
四、《译典通》·····	289
第三节 我国内地的光盘词典·····	294
一、《着迷词王 2001》·····	294
二、《金山词霸》·····	299
三、《汉语大词典》·····	308
四、《多媒体汉字字典》·····	313
第四节 各类光盘电子词典的比较·····	316
第九章 网络电子词典·····	318

第一节 概述	318
第二节 OneLook [®] 词典网站	320
一、概述	320
二、主界面与主要功能	321
三、网站词典的类别	322
四、检索系统	327
第三节 YourDictionary 词典网站	331
一、概述	331
二、主界面与主要功能	332
三、YDC 的各种索引	333
四、YDC 的查询	341
五、语际互译功能	342
第四节 英语词典网络版简介	343
一、《朗文网络词典》	344
二、《剑桥在线词典》	345
第五节 法语网络词典简介	347
一、拉鲁斯出版社的网络词典	347
二、阿歇特书社的网络词典	348
三、《阿特拉斯网络百科词典》	351
第六节 综合词典网站的性能特点	353
第十章 微型芯片电子词典	356
第一节 概述	356
第二节 “快译通”掌上电子词典	358
一、快译通 EB 系列	358
二、快译通的《牛津当代大词典》系列	363
三、快译通 EC 系列	364
四、快译通 LD 系列	366
五、快译通 PDA 系列	370
第三节 “好易通”掌上电子词典	372

一、好易通 CD 系列	373
二、好易通 MD 系列	374
三、好易通牛津系列	376
四、好易通 PDA 系列	379
第四节 “文曲星”掌上电子词典	382
一、文曲星 CC 系列	383
二、文曲星 PC 系列	385
三、文曲星 NC 系列	387
四、文曲星新东方系列	389
第五节 其他品牌的掌上电子词典	390
一、“名人”电子词典	390
二、“雷登智典”电子词典	392
三、“诺亚舟”电子词典	393
四、“联想”电子词典	394
五、“锦囊”电子词典	395
六、“步步高”电子词典	396
七、“佳能酷易典”电子词典	397
八、“信利”电子词典	398
九、“卡西欧”中英电子词典	399
十、“无敌”互动学习词典	399
第六节 掌上电子词典的知识结构分析	401
一、宏观结构信息分析	402
二、微观结构信息分析	404
三、改善知识内容是电子词典的努力方向	406
第十一章 电子词典编纂的一般原则	407
第一节 概述	407
第二节 收词、立目与注音的原则	408
一、收词原则	408
二、立目和注音原则	410

第三节 义项分立及释义的原则	411
一、义项划分的原则	411
二、义项排列的原则	414
三、释义的原则	414
第四节 括注和设例的原则	419
一、括注的原则	419
二、设例的原则	421
第五节 多媒体应用与附录设置的原则	422
一、多媒体应用的原则	422
二、附录设置的原则	424
第六节 综合、集合与孳生原则	425
第十二章 电子词典的发展趋势与结构模块	427
第一节 概述	427
第二节 电子词典的发展趋势	428
一、语言处理形式化	428
二、词典编纂自动化	429
三、小型词典情景化	429
四、专科词典集合化	430
五、大型词典综合化	430
六、综合词典网络化	431
七、常用词典芯片化	431
八、词典载体多元化	431
九、信息表述多媒体化	432
十、界面结构模块化	432
十一、界面窗口多样化	433
十二、窗口结构弹性化	433
十三、词典查询智能化	433
十四、信息显示任意化	434
十五、信息交流互动化	434