

## 《基于语料库的语言研究：高级资料读本》简介

T McEnery, R. Xiao, Y. Tono 著 许家金简介

T McEnery, R. Xiao (肖忠华) 和 Y. Tono 合著的《基于语料库的语言研究：高级资料读本》(*Corpus-based Language Studies: An Advanced Resource Book*) 一书 2006 年由 Routledge 出版社出版。该书既是语料库语言学入门读物, 又可作为进阶读本。

### 1 内容简述

第一部分“导论”共 10 个小单元, 简要介绍了语料库语言学的基本概念、术语和观点, 主要内容是 A2 ~ A9 这 8 个单元 (代表性、均衡性和取样, 语料库标注 (mark-up), 语料库赋码 (annotation), 多语言语料库, 提出统计学结论, 使用已有的语料库, 自行构建语料库, 著作权)。其作用好比烹制大餐前的准备, 在用语料库做语言研究之前预备必要的素材。

语料库研究中首要的是用于分析的语料。语料的有无、多寡、质量好坏将直接影响最终的研究结果。因此, 所收语料是否均衡、是否具代表性都是应当考虑的因素 (A2)。拥有了大量语料, 通常便可以开展很多与词汇相关的研究了。然而, 由纯文本汇集而成的生语料所能提供的信息终究有限, 因为语言形式之外的语法、语义、语用等信息很难从生语料中直接获取。这好比缺了佐料和搭配用的小菜, 做出来的菜品再好, 终究只解决温饱, 满足不了色香味的全面需求。因此, A3 和 A4 单元专门讨论了如何在语料文本中标注或赋码。有了赋码语料后, 就可以从语料中提取更多复杂的语言学信息。此外, 不同的研究目的往往需要不同性质和类型的语料。目前通常是利用已建成的大规模语料库 (A7) 来进行研究, 因为自建大规模语料库相当耗费资金、时间和人力, 很难完成。当然, 某些情况下, 也可根据研究需要自建小规模语料库 (A8)。对于希望从事诸如英汉对比研究的人来说, 则需借助平行或对应语料库 (A5)。有关语料库创建和使用中的相关问题, 在导论中基本都可找到相应解答。导论还特别强调了语料收集和出版中著作权问题的处理 (A9), 这一点在国内语料库建设中关注不够, 值得注意。本书介绍了国际上处理此类问题的常见对策。有了语料, 借助合理的检索方法和统计方法 (A6), 就能得到想要考察的语言形式或结构在语料中的频率分布情况。在第一部分结尾处, 作者用 40 页左右的篇幅介绍了语料库方法在词典编纂、翻译、对比语言学、社会语言学、语法、语义、语用、话语分析等 13 个领域中的应用情况。

第二部分“扩展研究”节选了 21 篇比较有代表性的已发表的论文, 分别归在 6 大热点主题之下 (语料库的代表性和均衡性, 反对语料库的观点: 目前的学术争议, 词汇和语法研究, 语言变异研究, 对比和历时研究, 语言教学和语言学习)。这一部分浓缩了学界成果, 对文献资料匮乏的中国读者尤为相宜。该部分节选了很多经典文章, 比如: Biber (1993: 125 - 7); Widdowson 和 Sinclair, Stubbs 等人关于应用语言学中是否应提倡使用语料库的著名论战

(Pp. 140 - 4); 用语料库进行语法研究的代表作——Carter和 McCarthy(1999)关于 get引导的遭受式被动结构的研究 (Pp. 152 - 5); Biber的语域变异研究以及他开创的“多维度多特征 研究法 (MD/MF或 MDA) (Pp. 160 - 5)。

第三部分“深入研究 围绕 6个案例分析展开 (词语搭配和教学词典学; 对 help和 help to的语料库对比研究; 语素的二语习得; 现代英语中的骂詈语; 美式英语会话和言语; 英汉翻译与语域、文本类型和体标记)。案例分析引导读者从寻找研究问题入手, 进入语料检索, 再就检索结果进行统计和语言学分析, 进而讨论并解决所提出的研究问题。这 6个研究问题涉及词汇分析、语法结构研究、口语会话特征考察、语域变异探究, 对初学者和已有语料库研究经验的学者都有帮助。所有案例分析都包含语料库软件操作步骤截图配合文字说明, 初学者很容易掌握。而对有经验的学者来说, 也可学到一些新的方法和思路, 比如如何将语料库软件得到的统计数据导入统计软件深入处理 (如频率标准化、两组频率数据的卡方检验等)。作者还把 Biber(1988)中介绍但并未给出操作方法的 MDA方法一步一步地详细解析, 十分利于读者掌握 (Pp. 287 - 320)。

## 2 简评

该书具有以下特色:

(1) 内容新。这本书虽没有大的理论创新, 但相比 Stubbs(1996), McEnery(2001 [1996]), Kennedy(1998), Meyer(2002)而言, 内容上有了很大的更新, 较能代表语料库语言学的研究现状, 对新语料库、软件以及最新研究成果的介绍尤其如此。

(2) 理论与实践结合。在众多语料库语言学专著中, 该书是理论与实践结合最好的一部。作者安排了大量操作性内容, 很好地反映了语料库语言学研究的实际——较多依赖语料库和检索软件, 如果缺少相应指导, 往往不容易开展实际研究。该书第三部分的实例分析图文并茂, 使得计算机操作的技术问题迎刃而解。而且, 作者在介绍软件时不局限于其中一款, 而是对 WordSmith Tools, MonoConc Pro, BNCWeb以及 ParaConc等常用软件分别解析, 特别有利于读者认识不同语料库工具的优缺点, 可以根据研究需要自由选择。此外, 该书配套的网站 (<http://www.routledge.com/textbooks/0415286239/Resources/default.html> 点击日: 2008年 10月 6日)上提供了 6个实例分析中软件、语料、标注方案 (如词性赋码集 CLAWS tagset 7)和检索方式 (比如用于案例 6的 58个语言特征的检索表达式)的链接和说明。

(3) 经典文献选读。作者将相关经典文献片段汇集在一起, 可以帮助读者了解该领域的学者对于同一研究问题的不同切入点以及一些观点上的交锋。据此, 读者能够更好地设计和选择研究方向和研究问题。读者也可先将所选的研究作为学习和仿效的样板, 从模仿开始, 逐步开展原创研究。

该书也存在一些缺憾:

(1) 全书结构安排上有一些问题。比如, 第一部分的 A6单元“提出统计学结论”放在 A8单元“自行构建语料库”后面, 会显得更合乎逻辑。第二部分的 B1单元“语料库的代表性和均衡性”和 B2单元“反对语料库的观点: 目前的学术争议”与该部分其他几个单元的具体研究并不在同一层面上。而在第三部分, 可能由于作者研究兴趣所限, 所介绍的研究课题仍

然未能摆脱目前该学科的通病——以词汇研究为中心，只面向书面语和转写文本。

(2) 从宏观研究方法和视野来看，书中的研究课题仍然没能做到多种研究方法并举 (triangulation)，尤其是定量定性研究方法的充分结合。而只有两种方法结合，才能挖掘频率信息、语域变异等形式特征背后的深层动因，才能回应有些学者 (如 Widdowson, Pp. 140 - 4) 对语料库方法的诟病，走出语料库研究目前重形轻义、重定量轻定性的困境。

(3) 与上面两条相呼应，从语料库语言学的研究层面和研究思路看，我们应重新审视语言研究的主要任务，应更多地关注互动话语，特别是自然口语。先前注重文本 (书面语及转写后的文本) 或许是一个研究方向上的谬误。Saussure 的《普通语言学教程》(1916: 6) 中就曾指出，实际使用中的话语才是语言学真正需要的。基于文本而开展的语言形式和结构的研究原本只是权宜之计——鉴于语言学家限于时空，不可能总是出现在话语发生的第一现场，所以传统上只能借助文本记录开展研究。

近年来，随着多媒体技术、言语工程研究方面的迅猛发展，从技术和方法上已经使得自然话语的研究特别是多模态 (multimodal) 研究成为可能。人们在具体社会情境中，总是运用多种符号资源来建构意义，本身就具有多模态性，而多模态的研究方法就是将语言及其相关的资源整合起来 (Gibbon, et al 2000)，尽可能保持实际话语的原貌 (Kress and Leeuwen 2001: Preface)。因此，语料库语言学的研究应该结合其他定性分析方法，更多地采用多模态研究方法考察自然互动话语，更多关注语言本体特征和语言使用者及其与社会环境之间的互动因素。

#### 参考文献

- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press
- Saussure, F. de. 1916. *Cours de Linguistique Générale*. Lausanne / Paris: Payot. R. Harris, trans. 1983. *Course in General Linguistics*. London: Gerald Duckworth
- Gibbon, D., I. Mertins, and R. Moore. 2000. *Handbook of Multimodal and Spoken Dialogue Systems*. Boston: Kluwer Academic
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Pearson Education
- Kress, G. and T. van Leeuwen. 2001. *Multimodal Discourse*. London: Edward Arnold
- McEnery, A. and A. Wilson. 2001 [1996]. *Corpus Linguistics*. Edinburgh: Edinburgh University Press
- Meyer, C. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press
- Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell

作者通讯地址：100089 北京 北京外国语大学中国外语教育研究中心

E-mail: xujiajin@bfsu.edu.cn