

语料库与语言教育

肖忠华 许家金

浙江大学 北京外国语大学

© 2008 中国外语教育(2), 48-58页

提 要: 随着语料库在语言学研究各领域的广泛应用, 语料库语言学作为一种研究方法已获得广泛认可。本文综述先前开展过的相关研究, 梳理语料库在语言教育中的用途。本文将涉及如下一些方面: 大纲设计、教材开发、数据驱动学习、专门用途英语、语言测试、教师培训及学习者语料库研究和中介语分析。

关键词: 语料库; 语言教学; 数据驱动学习; 学习者语料库; 中介语分析

1. 引言

过去40年中, 特别是上世纪80年代以来, 基于语料库的语言学和语言教育研究愈来愈受到重视。这是因为语料库研究差不多对语言学的各个分支都有所助益 (Leech 1997: 9; Biber *et al.* 1998: 11)。语料库的优势之一在于它的实证性, 它将大量语言使用者的语言直觉汇集起来, 从而使得语言分析更为客观 (McEnery & Wilson 2001: 103)。目前看来, 语料库对工具书的出版影响最大 (参见 Sinclair *et al.* 1990; Francis *et al.* 1996; 1998; Biber *et al.* 1999), 以至于“从未听说过语料库的人也在使用语料库的研究成果” (Hunston 2002: 96)。Hunston把语料库对词典编纂和工具书出版带来的变化概括为5个“关注”: 关注频率, 关注搭配和组块性, 关注语体变异, 关注词汇在语法体系中的作用, 和关注真实性 (authenticity)。除了工具书的出版, 语料库还被广泛用于几乎所有语言学分支, 例如: 词

学、词汇研究、语法研究、语言变异研究、对比与翻译研究、历时语言研究、语义学、语用学、文体学、社会语言学、话语分析、法律语言学及语言教学。虽然基于语料库的研究偶尔会招致责难, 但它还是具有广泛的影响。本文不想过多关注有关语料库是否适用于语言研究与语言教育这一争论。对语料库使用的争论, 可参看 Sinclair (1991), Widdowson (2000), de Beaugrande (2001) 以及 Stubbs (2001)。有关语料库在应用语言学中的应用的深入探讨, 读者可参看 Hunston (2002) 及 McEnery *et al.* (2006)。本文主要以英语教学为例讨论语料库在语言教育中的应用。

语料库研究应用于语言教学的热潮可从已举办了七届的“教学与语料库 (TaLC - Teaching and Language Corpora)”研讨会得到印证。从最近出版的文献来看, 这一趋势也显而易见。近来, 除了大量期刊杂志文章外, 至少有9本专著和论文集专门研究教学与语料库的问题: Wichmann *et al.* (1997), Burnard & McEnery (2000),

Aston (2001), Granger *et al.* (2002), Hunston (2002), Kettemann & Marko (2002), Tan (2002), Aston *et al.* (2004) 及 Sinclair (2004)。这些著作涵盖了语言教学与语料库领域的诸多方面。比如：基于语料库的语言描写、课堂语料分析及学习者语料研究。在国内，语料库与外语教学也是近年来语料库研究的一大热点，现已出版有关专著和论文集十多本，如杨达复 (2000)，濮建忠 (2003a)，何安平 (2004a, 2004b)，华南师范大学外国语学院 (2005)，卫乃兴、李文中、濮建忠 (2005)，杨惠中 (2005)，王立非、梁茂成等 (2007)。

2. 语料库与语言教学：学科共进

在 *Teaching and Language Corpora* (Wichmann *et al.* 1997) 一书的第一章里，Leech 指出教学与语料库研究之间显然已经达成了一种学科共进 (convergence)。Leech (1997) 特别指出这种共进有三个重点：第一是直接将语料用于教学（即“教语料库语言学”，“教如何利用语料库”，以及“用语料库来教”）；其二是语料间接用于教学（如：工具书的出版、教材开发和语言测试等）。另外，还有一些面向教学的语料库建设（专门用途语言语料库、母语习得语料库及二语学习者语料库）。

在这些重点之中，所谓“教语料库”指的是把语料库语言学作为像社会语言学和话语分析这样的语言学分支去教。语料库语言学目前已经进入语言和语言学相关本科和研究生学位课程。而“教如何利用语料库”是指教给学生研究语料库的知识和技能，这样学生就可以按照自己的意图去挖掘利用语料。一旦学生们掌握了语料库研究的必要知识和技能，相应的学习过程就会变成以学习者为中心了。如果“教语料库语言学”可以看作主要是针对语言和语言学专业学生的，那么“教如何利用语料库”则囊括了涉及语言研究和语言学习各个学科的学生。“用语料库来教”指的是用基于语料库的手段来教语言和语言学课程，而这些课程的教学本来也可以不采用基于语

料库的方法。至于说语言教学中对语料库的间接使用，如前所述，主要指语料库对工具书的出版带来的革命性变化。接下来，我们将看到语料库在大纲设计、教材开发、课堂检索练习和语言测试等诸多方面都有用武之地。最后，面向教学的语料库在专门用途语言、母语及二语习得方面也有很多用途。

3. 大纲设计与教材开发

在语料库被广泛用来对语言使用进行描述的同时，许多学者也利用语料库直接去考察现有的英语教学大纲和教材。比如 Mindt (1996) 发现英语作为外语教材中语法结构的使用同英语母语中的用法有明显不同。他认为英语教材的一个通病就是这些教材所教授的“是一种看来并不会出现在外语课堂之外的学校英语” (Mindt 1996: 232)。这样教出来的学生很难与英语母语者顺畅交流。语料库在语言教育中的一个通常的作用就是提供更加真实的语言实例。此外，语料库还可提供一些其他信息，特别是频率信息，这些都有助于进一步调整我们所教的内容。例如，Mindt 将情态动词、将来时表达法和条件句在英语母语语料库中的分布频率与这些结构在德国热门英语教材中的教学次序作了比较，结果发现：凡是没有基于语料库的教材中，这些语言结构的出现顺序“往往与人们预想的口语和书面语中的出现情况不一致”。他指出教学大纲的编写应当以实证为依据，而不是以直觉为基础，应当以实际语言使用的频率数据为指导来安排教学的先后顺序 (Mindt 1996: 245-246)。

Hunston (2002: 189) 的观点与 Mindt 相似。她指出“语料库的使用会引发大纲设计方面完全不一样的一些思路”。她这里提到的大纲即所谓的“词汇大纲”。这个大纲最早由 Sinclair & Renouf (1988) 提出并在 Willis (1990) 书中得到全面阐释。Sinclair & Renouf (1988: 148) 指出：词汇大纲一是关注语言中最常见的词形；二是关注语言用法中的核心结构；三是由前两者经常在一起构成的组合形式。然而“词汇大纲”这一术

语时常会被误解为是只涵盖词汇形式的大纲。事实上, 词汇大纲涵盖“语言的所有层面, 它与传统大纲的不同之处仅仅在于它在大纲组织中的核心概念是词汇”(Hunston 2002: 189)。Sinclair (2000: 191) 指出词汇大纲所涉及的语法是“lexical grammar”而不是“lexico-grammar”。这种 lexical grammar 试图将“语法和词汇等量齐观”。诚然, 正如 Murison-Bowie (1996: 185) 所评论的那样, “在教学中使用语料库时, 经常很难区分什么是对词汇的考察, 什么是对句法的考察。两者往往不分彼此, 研究词汇导致研究语法, 反之亦然。这在教学中也是有利的。” Sinclair 等人提出的词汇大纲得到了 Lewis (1993, 1997a, 1997b, 2000) 的积极支持。

语言教学中运用词汇大纲的一个主要方面就是搭配和预制词块的教学。许多学者认为(如: Bahns 1993; Zhang 1993; Cowie 1994; Herbst 1996: 389-391; Kita & Ogata 1997: 230-231; Partington 1998: 23-25; Hoey 2000, 2003; Shei & Pain 2000: 167-170; Sripicharn 2000: 169-170; Altenberg & Granger 2001; McEnery & Wilson 2001; McAlpine & Myles 2003: 71-75; Nesselhauf 2003; Xiao & McEnery 2006), 无论是对母语还是对二语语言技能的培养, 搭配知识都很重要。比如 Hoey (2003) 指出: “词汇学习包含与之共现以及相应的语法型式所包含的内容。” Cowie (1994: 3168) 提到, “接近母语水平的语言很大程度上依赖于对预制词块的掌握程度。” Aston (1995) 也认为: 预制词块的使用可以加速语言理解和产出的过程, 进而达到近似母语的流利程度。正如 Partington (1998: 20) 提出的那样, 搭配运用的一个强有力的理由在于“它促进了作为听者一方对交际信息的处理, 因为语言中包含有相当数量的固定词组, 这些固定词组较之其他语言结构更易于预测, 而在实时语言编码过程中, 听话人需要借助于他们所能获得的所有帮助。”因此, 一个人的语言能力无疑应包括搭配知识(参看 Herbst 1996: 389)。搭配知识指的是哪些词汇形式“频繁与其他词汇形式共现, 以及在句中怎样组合”。这样的知识显然比单个词汇形式的知

识更重要(参看 Kita & Ogata 1997: 230), 而且有效生成句子也需要这种知识(参看 Smadja & McKeown 1990)。例如, Zhang (1993) 发现与低水平学习者相比, 高水平学习者写作中明显较多使用搭配形式, 且正确程度更高, 变化形式也更多。搭配错误也是学习者常犯的一种错误(参看 McAlpine & Myles 2003: 75)。桂诗春和杨惠中(2002: 48) 对中国学习者英语语料库的分析表明, 搭配错误的确也是中国英语学习者一种主要的错误类型。Altenberg & Granger (2001) 以及 Nesselhauf (2003) 发现, 即使是高级英语学习者使用搭配也会存在相当困难。一种可能的解释是学习者在“自动处理搭配”方面存在欠缺(Kjellmer 1991)。“因此, 学习者需要对常用搭配和习语有所了解, 包括固定的、半固定的词汇搭配形式以及这些形式的变异程度; 某些特定搭配形式的相对频率和常用程度以及它们的正式程度”(McAlpine & Myles 2003: 75)。语料库在这方面无疑大有用武之地。这不仅是因为只有通过量化才能可靠地衡量搭配, 而且还因为语料库可以通过语境共现(KWIC)检索方式让学习者接触到大量排列整齐的真实语料。我们的观点与 Kennedy (2003) 是一致的。他在讨论语料库同语言教学本质之间的关系时尤其强调了搭配教学。他认为, 二语或外语学习是一个有意识地学习“显性知识”的过程, 而这一过程需要大量接触语言材料。

4. 课堂检索: 数据驱动学习

大纲设计和教材开发与“教什么”密切相关, 而语料库还对我们“如何教”提供了有益启示。关于如何在语言课堂上运用语料库这一话题已有广泛的讨论。就基于语料库的语言教学法而言, 传统的 3P 教学模式(即 Presentation-Practice-Production)未必完全适切。而更具探索性的 3I 模式(即 Illustration-Interaction-Induction)可能更为适用。Illustration 指的是观察真实语料, Interaction 指的是讨论和分享在语料中的发现, 而 Induction 指的是对某个语言点归纳出自己的规

则(参见 Carter & McCarthy 1995: 155)。这样得出的规则在通过观察更多语料的过程中逐步修正完善。Murison-Bowie (1996: 191) 将这种循序渐进式的归纳法总结为: 最初从有限数据中概括出不完整的规则, 并逐步完善。虽然 Carter 和 McCarthy 最初提出 3I 教学法主要是针对口语语法教学, 但我们认为, 3I 教学法也适用于整个语言教学。

这种 3I 教学模式显然与 Johns (1991) 的“数据驱动学习 (DDL)”有异曲同工之效。Johns 应是最早意识到语料库对语言学习者的潜在价值的语言学家(参见 Higgins & Johns 1984)。在他看来, “研究……不能只留给研究人员去做”(Johns 1991: 2), 因此, 他主张应鼓励语言学习者成为“研究者”, “其学习需要通过接触语料来驱动”。为适应不同层面学习者的学习需求, 数据驱动学习可以是教师指导的, 也可以是由学生主导的(即发现式学习), 但基本上来说, 数据驱动学习是以学习者为中心的。

Johns (1991) 将数据驱动学习中借助语料库进行的归纳式推理过程分为三个阶段: (对检索结果的) 观察, (对显著特征的) 分类, 以及(对规则的) 概括。这三个阶段大体与 Carter & McCarthy (1995) 提出的 3I 是一致的。数据驱动学习法和 3P 教学法本质区别在于前者是一种自下而上的归纳过程而后者是一种自上而下的演绎过程。有关在语言课堂中直接运用语料库和索引所涉及的理论、方法、技巧、问题及解决方法等文献中已多有论述(例如: Tribble 1991, 1997a, 1997b, 2000, 2003; Tribble & Jones 1990, 1997; Flowerdew 1993; Karpati 1995; Kettemann 1995, 1996; Wichmann 1995; Woolls 1998; Aston 2001; Osborne 2001), 在此不再赘述。

5. 专门用途与专业交流用途语言

除了在外语教学中的运用之外, 语料库在特定领域语言使用、专业交流(如专门用途英语和学术用途英语)方面的应用也受到了极大的关

注。比如 Thurstun & Candlin (1997, 1998) 探讨了将语料库检索用于学术英语写作和词汇教学。Hyland (1999) 比较了特定学科的导理性教材同研究性论文中元话语手段的异同。Upton & Connor (2001) 利用学生商务英语语料库开展了一项商务英语的“话步分析”。两位作者在这篇文章中通过分析来自不同文化背景的学习者的“礼貌策略”, 从文化视角剖析了专业交流话语的特点。Thompson & Tribble (2001) 在分析语料的基础上考察了学术文章中的参引(citation)的情况。Koster (2002) 通过对“工作坊(workshop)”会话中言语行为的分析, 提出一种教授英语口语交际功能的话语视角。Yang & Allison (2003) 研究了应用语言学研究性文章中的篇章组织结构。Carter & McCarthy (2004) 在“剑桥—诺丁汉英语话语语料库(CANCODE)”基础上探究了一系列体现创造性语言使用的社会情境。Hinkel (2004) 比较了母语和二语学术文章中的时、体、被动语态的使用情况。肖忠华(2003)评述了一系列运用特定语域的多语种语料库进行特定语域翻译的个案研究。所有这些研究表明, 专门用途语料库对专门用途语言和专业交流语言的教学是大有裨益的。

6. 语言测试

语言测试是语言教学中的另一个已经开始运用语料库研究方法的新兴领域。早在 1996 年, Alderson 就指出语料库研究方法可能运用于语言测试领域, 如, 试题的编写、汇编、筛选、试题的推出、测试反应的获取、评分、测试结果的计算和成绩报告等。他总结道: “以真实语料为基础来测试, 以语料数据为基础来判断考生的能力、知识和实际表现, 这样做的好处是显而易见的。问题的关键是这些好处能否付诸于实践”(Alderson 1996: 258-259)。Alderson 在结论中所提到的有关实际应用的顾虑似乎已经得到了解答。例如 Choi, Kim & Boo (2003) 发现机考与笔考效果相当。再如, Coniam (1997) 报告了如

何利用从语料库中提取的词频信息自动生成完型填空的试题。Kaszubski & Wojnowska (2003) 也报告了他们基于语料库的小程序 Testbuilder, 可以用来生成句级的英语试题。这个软件既可以用生语料生成测试练习, 也可以利用内嵌的词性赋码器对文本进行赋码后再利用词性赋码文本生成练习。而梁茂成 (2007) 推出的基于语料库的试题生成软件 CoolTomatoes, 无论从功能还是界面友好程度来看, 都是语料库与语言测试结合的一个成功范例。该软件可根据试题编写者的需要自动生成完型填空、多项选择等试题, 同时实现自动评分, 并可通过网络将评分结果报告给教师或教学管理人员。另外, 近来主要的测试机构都在将语料库用于各种目的, 比如: 1) 作为测试文本库; 2) 用于开发试题材料; 3) 优化测试流程; 4) 改善评分质量; 5) 提高测试的效度; 以及 6) 使测试标准化和规范化 (参见 Ball 2001; Hunston 2002: 205)。英国剑桥大学考试委员会 (即 UCLES) 即是一例。该中心历来重视语料库的开发 (如: Cambridge Learner Corpus, Cambridge Corpus of Spoken English, Business English Text Corpus 及 Corpus YLE Speaking Tests), 同时也致力于英语母语和学习者英语的分析研究。比如, 他们利用英国国家语料库 BNC 研究了英语中的搭配情况、真实的题干和合理的干扰项分析, 这就使试卷编写者能以真实文本为基础来编写试题 (Ball 2001: 7)。他们还用语料库方法来考察二语、外语学习者或者参加各类剑桥考试的考生写作中的明显特征, 并研究如何将这些特征“归入单一的等级体系, 即用同一体系来描述不同等级的二语写作水平” (Hawkey 2001: 9)。UCLES 还将语料库用于口语测试 (Ball & Wilson 2002; Taylor 2003), 并制定特定语域 (如商务英语) 的测试词表 (Ball 2002; Horner & Strutt 2004)。

7. 教师发展

要想让学生从语料库中受益, 语言教师首先应当掌握语料库研究方法。因此, 将语料库用于教师培训也就不足为奇了 (参见 Allan

1999, 2002; Conrad 1999; Seidlhofer 2000, 2002; O’Keeffe & Farr 2003)。譬如, Allan (1999) 展示了如何利用语料库来提高香港中学英语教师的语言意识。Conrad (1999) 介绍了一项基于语料库的连接状语 (如: therefore 和 in other words) 的研究。她通过这项研究指出, 如果教学中要充分利用语料库方法的话, 语言教师使用语料库时只是在课堂上进行语料检索和词汇分析是不够的。Conrad 对教师培训的建议得到了 O’Keeffe & Farr (2003) 的赞同, 她们认为要提高教师的研究技能和语言意识, 语言教师的培训应当包括语料库语言学的内容。在这方面, 外语教学与研究出版社和中国外语教育研究中心联合举办了“全国高等学校英语教师教育与发展系列研修班”, 该系列研修班至 2007 年底已经举办了 5 期“语料库在外语教学中的应用”及“语料库与外语研究”语料库专题研修班, 接受培训的全国高等学校外语教师累计近千人, 对推动语料库方法在外语教学和研究中的应用发挥了积极的作用。

8. 学习者语料库和中介语分析

近年来, 语料库研究中最令人振奋的进展主要体现在学习者语料库的创建和使用。除了国际上一些著名的学习者语料库 (如“英语学习者国际语料库” (ICLE)) 之外, 目前国内对该类语料库的研制也已取得骄人的进展 (如: 桂诗春、杨惠中 2003; 文秋芳、王立非、梁茂成 2005; 杨惠中、卫乃兴 2005)。在 2003 年英国兰卡斯特大学举办的第二届国际语料库研讨会 (Corpus Linguistics 2003) 的有关“学习者语料库”的会前专题研讨期间, 组织者 Tono 和 Meunier 认为学习者语料库研究已经走过婴儿期而步入了青少年期: 这方面的研究前景十分看好, 但潜力还没有得到充分挖掘。在语言教学领域, 学习者语料库已应用于课程设计、教材开发以及教学法研究等 (参见 Keck 2004: 99)。母语和二语教材之间的接口也有所研究。比如 Meunier (2002) 提出单靠从母语语料中得出的频率信息去指导课程和教材设计是不够的, “而应当在频率信息、难

易度以及与教学的相关性方面取得平衡。衡量这些因素各自的重要程度,这正是学习者语料库研究所能有所贡献之处”(Meunier 2002: 123)。Meunier 同时提倡在语言课堂中使用学习者语料。她提出通过比较学习者语料和目标语母语语料,并分析学习者语料中出现的错误,可以帮助学生意识到他们的中介语和目标语之间的差异。迄今为止的基于学习者语料库的中介语研究基本上都属于 Granger (2002) 称之为“中介语对比分析(CIA)”。这种方法将学习者语料同目标语母语语料或者不同母语背景的学习者的语料进行对比分析。前一种对比旨在发现学习者语言中某些语言特征过多使用(overuse)或使用不足(underuse)的情况,而第二种对比旨在揭示二语学习者的二语产出中是否存在母语迁移。运用“中介语对比分析”方法,国内今年出现了一大批研究成果,涵盖学习者语料中词汇、搭配、类联结、话语、语用等多个方面。如:马广惠(2002),李文中(2003),濮建忠(2003b),文秋芳、丁言仁、王文宇(2003),何安平、徐曼菲(2003),俞洪亮(2005),王立非、祝卫华(2005),许家金、许宗瑞(2007),胡学文(2007),陈新仁、任育新(2007),梁茂成(2008)。

9. 频率和语料真实性:是赞成还是反对?

在结束有关利用语料库进行语言教育的讨论之前,最后探讨一下有关利用语料库进行语言教学的一些反对意见。一般认为,频率信息和语言真实性是使用语料库带来的两大优势,而这两方面恰恰也是研究学者们批评的主要方面。例如, Cook (1998: 61) 认为,在语言学习中使用语料库会不恰当地关注了高频语言现象而舍弃那些不太常用却又十分重要的表达法。Widdowson (1990, 2000) 也指出,语料库中语料所谓的真实性是有局限性的,因为语料库中的语言已经脱离了语境。要在语言教学中使用这种语料必须将他们重新植于语境。针对有关频率的批评意见, Osborne (2001: 486) 反驳道:

“相反,运用语料库不仅增加了学习者接触相对低频语例的机会,而且还可以帮助学习者了解为什么这些语例是不典型的,它们通常用在什么语境中,并对它们与相应语言形式的典型用法之间如何交替使用有所认识。”

这一观点得到了 Goethals (2003: 424) 的支持,他指出“频率等级可以作为对学习材料排序和定级的一个指标”,因为“频率是衡量语言结构可用性的概率特征”,而“高频词构成的核心词汇比教师或教材中随机选取的词汇应当更有价值”。Hunston (2002: 194-195) 也认为,那些重要但又不常用的语言结构似乎与具有较高文化内涵的文本有关。但在许多情况下,文化中的某些显著性特征与频率信息并不矛盾。虽然,频率信息很容易从语料库中得到,但没有哪一位语料库语言学家断言过最高频的就是最重要的。相反, Kennedy (1998: 290) 指出“频率只应该是影响我们教学的标准之一”。他还指出,“从语料分析中产生出来的关于语言和语言使用的一些事实不应当成为语言教学的负担”。因此,如 Renouf (1987: 168) 所述的那样,从语料库中得出的原始频率在进入大纲时往往需要加以调整。当然,对语言教师、大纲设计者、教材编写者而言,无视频率这一“现成的、强有力的证据”也是不合适的。正如 Leech (1997: 16) 所言:

“‘最高频’=‘最难学’这个简单化的公式无论有多不完美,我们都很难否认从语料库中得到的频率信息对语言学习材料来说是一个重要的实证依据。”

现在再来重新思考一下 Widdowson (1990, 2000) 对在语言教学中使用真实性文本提出的质疑。在他看来,语料库中语料的真实性在语言教学中只不过是一种“幻觉”(1990: 44),因为即便该语料从某种意义上来说或许是真实的,但当把它用于教学时,由于语言学习者并非当初设定的听众或读者,该语料目的这一真实性即已遭到了损毁(参见 Murison-Bowie 1996: 189)。Widdowson 言下之意,只有为课堂想象出来的情景而产生的语言才是“真实的”语言。然而,

如 Fox (1987) 所言, 杜撰出来的语例通常不能反映语言使用的微妙之处。这也许就是为什么 Mindt (1996: 232) 认为, 学生运用所学到的“学校英语”很难马上就在真实生活中直接与讲英语母语的人用英语顺畅交流的原因。为此, Wichmann (1997: xvi) 指出, 在语言教学中, “对语言材料‘真实性’的倚重需要老师和学生共同努力来学好课本所不能预见的语言。”

10. 结语

本文概要地综述了语料库在语言教育中的主要用途。Conrad (2001) 曾预言, 语料库会给 21 世纪的语法等课程教学带来革命性的变化。而我们上述所介绍的这些研究也表明, 语料库还会从根本上改变我们进行语言教育的方方面面, 包括“教什么”和“怎么教”。正如 Gavioli & Aston (2001) 所言, 我们不仅仅应该把语料库看作是一种有助于教师决定教什么的资源, 而且还应该把它看作是一种学生直接可以从中学习的资源。

参考文献

- Alderson, C. 1996. Do corpora have a role in language assessment? [A]. In J. Thomas & M. Short (eds.). *Using Corpora for Language Research* [C]. London: Longman, 248-259.
- Allan, Q. 1999. Enhancing the language awareness of Hong Kong teachers through corpus data [J]. *Journal of Technology and Teacher Education* 7(1): 57-74.
- Allan, Q. 2002. The TELEC secondary learner corpus: a resource for teacher development [A]. In S. Granger, J. Hung & S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [C]. Philadelphia: John Benjamins, 195-212.
- Altenberg, B. & S. Granger. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing [J]. *Applied Linguistics* 22(2): 173-195.
- Aston, G., S. Bernardini, & D. Stewart (eds.). 2004. *Corpora and Language Learners* [C]. Amsterdam: John Benjamins.
- Aston, G. (ed.) 2001. *Learning with Corpora* [C]. Houston, TX: Athelstan.
- Aston, G. 1995. Corpora in language pedagogy: Matching theory and practice [A]. In G. Cook & B. Seidlhofer (eds.). *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson* [C]. Oxford: Oxford University Press, 257-270.
- Bahns, J. 1993. Lexical collocations: A contrastive view [J]. *ELT Journal* 47(1): 56-63.
- Ball, F. 2001. Using corpora in language testing [J]. *Research Notes* 6: 6-8.
- Ball, F. 2002. Developing wordlists for BEC [J]. *Research Notes* 8: 10-13.
- Ball, F. & J. Wilson. 2002. Research projects relating to YLE Speaking Tests [J]. *Research Notes* 7: 8-10.
- Biber, D., S. Conrad & R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use* [M]. Cambridge: Cambridge University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English* [M]. London: Longman.
- Burnard, L. & A. McEnery (eds.). 2000. *Rethinking Language Pedagogy from a Corpus Perspective* [C]. New York: Peter Lang.
- Carter, R. & M. McCarthy. 1995. Grammar and the spoken language [J]. *Applied Linguistics* 16(2): 141-158.
- Carter, R. & M. McCarthy. 2004. Talking, creating: interactional language, creativity, and context [J]. *Applied Linguistics* 25(1): 62-88.
- Choi, I., K. Kim & J. Boo. 2003. Comparability of a paper-based language test and a computer-based language test [J]. *Language Testing* 20(3): 295-320.
- Coniam, D. 1997. A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests [J]. *CALICO Journal* 16(2-4): 15-33.
- Conrad, S. 1999. The importance of corpus-based research for language teachers [J]. *System* 27: 1-18.
- Conrad, S. 2000. Will corpus linguistics revolutionize grammar teaching in the 21st century? [J] *TESOL Quarterly* 34: 548-560.
- Cook, G. 1998. The uses of reality: A reply to Ronald Carter [J]. *ELT Journal* 52(1): 57-64.
- Cowie, A. 1994. Phraseology [A]. In R. Asher (ed.). *The Encyclopedia of Language and Linguistics Vol. 6* [C]. Oxford: Pergamon Press, 3168-3171.
- de Beaugrande, R. 2001. Interpreting the discourse of H.

- G. Widdowson: A corpus-based critical discourse analysis [J]. *Applied Linguistics* 22(1): 104-121.
- Flowerdew, J. 1993. Concordancing as a tool in course design [J]. *System* 21(3): 231-243.
- Fox, G. 1987. The case for examples [A]. In J. Sinclair (ed.). *Looking Up: An Account of the COBUILD Project* [C]. London: HarperCollins, 137-149.
- Francis, G., S. Hunston & E. Manning. 1996. *Collins COBUILD Grammar Patterns 1: Verbs* [M]. London: HarperCollins.
- Francis, G., S. Hunston & E. Manning. 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives* [M]. London: HarperCollins.
- Gavioli, L. & G. Aston. 2001. Enriching reality: Language corpora in language pedagogy [J]. *ELT Journal* 55(3): 238-246.
- Goethals, M. 2003. E.E.T.: The European English Teaching vocabulary-list [A]. In B. Lewandowska-Tomaszczyk (ed.). *Practical Applications in Language and Computers* [C]. Frankfurt: Peter Lang, 417-427.
- Granger, S. 2002. A bird's-eye view of learner corpus research [A]. In S. Granger, J. Hung & S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [C]. Philadelphia: John Benjamins, 3-33.
- Granger, S., J. Hung & S. Petch-Tyson (eds.). 2002. *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching* [C]. Philadelphia: John Benjamins.
- Herbst, T. 1996. What are collocations: Sandy beaches or false teeth [J]? *English Studies* 4: 379-393.
- Higgins, J. & T. Johns. 1984. *Computers in Language Learning* [M]. Oxford: Oxford University Press.
- Hinkel, E. 2004. Tense, aspect the passive voice in L1 and L2 academic texts [J]. *Language Teaching Research* 8(1): 5-29.
- Hoey, M. 2000. A world beyond collocation: New perspectives on vocabulary teaching [A]. In M. Lewis (ed.). *Teaching Collocations* [C]. Hove: Language Teaching Publications, 224-245.
- Hoey, M. 2003. Lexical priming and the properties of text [OL]. www.monabaker.com/tsresources/LexicalPriming&thePropertiesofText.htm (accessed 08/05/2005).
- Horner, D. & P. Strutt. 2004. Analyzing domain-specific lexical categories: Evidence from the BEC written corpus [J]. *Research Notes* 15: 6-8.
- Hunston, S. 2002. *Corpora in Applied Linguistics* [M]. Cambridge: Cambridge University Press.
- Hyland, K. 1999. Talking to students: Metadiscourse in introductory coursebooks [J]. *English for Specific Purposes* 18(1): 3-26.
- Johns, T. 1991. "Should you be persuaded": Two samples of data-driven learning materials [A]. In T. Johns & P. King (eds.). *Classroom Concordancing ELR Journal* 4 [C]. Birmingham: University of Birmingham.
- Karpati, I. 1995. *Concordance in Language Learning and Teaching* [M]. Pecs: University of Pecs.
- Kaszubski, P. & A. Wojnowska. 2003. Corpus-informed exercises for learners of English: The TestBuilder program [A]. In E. Oleksy & B. Lewandowska-Tomaszczyk (eds.). *Research and Scholarship in Integration Processes: Poland-USA-EU* [C]. Łódź: Łódź University Press, 337-354.
- Keck, C. 2004. Corpus linguistics and language teaching research: Bridging the gap [J]. *Language Teaching Research* 8(1): 83-109.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics* [M]. London: Longman.
- Kennedy, G. 2003. Amplifier collocations in the British National Corpus: Implications for English language teaching [J]. *TESOL Quarterly* 37(3): 467-487.
- Kettemann, B. 1995. On the use of concordancing in ELT. *TELL & CALL* 4: 4-15.
- Kettemann, B. 1996. Concordancing in English Language Teaching [A]. In S. Botley, J. Glass, A. McEnery & A. Wilson (eds.). *Proceedings of Teaching and Language Corpora* [C]. Lancaster: Lancaster University, 4-16.
- Kettemann, B. & G. Marko. 2002. *Teaching and Learning by Doing Corpus Analysis* [C]. Amsterdam: Rodopi.
- Kita, K. & H. Ogata. 1997. Collocations in language learning: Corpus-based automatic compilation of collocations and bilingual collocation concordancer [J]. *Computer Assisted Language Learning* 10(3): 229-238.
- Kjellmer, G. 1991. A mint of phrases [A]. In K. Aijmer & B. Altenberg (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik* [C]. London: Longman.
- Koester, A. 2002. The performance of speech acts in workplace conversations and the teaching of communicative functions [J]. *System* 30: 167-184.
- Leech, G. 1997. Teaching and language corpora: A convergence [A]. In A. Wichmann, S. Fligelstone, A. McEnery & G. Knowles (eds.). *Teaching and Language Corpora* [C]. London: Longman, 1-23.
- Lewis, M. 1993. *The Lexical Approach: The State of ELT*

- and the Way Forward* [M]. Hove: Language Teaching Publications.
- Lewis, M. 1997a. *Implementing the Lexical Approach: Putting Theory into Practice* [M]. Hove: Language Teaching Publications.
- Lewis, M. 1997b. Pedagogical implications of the lexical approach [A]. In J. Coady & T. Huckin (eds.). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* [C]. Cambridge: Cambridge University Press, 255-270.
- Lewis, M. (ed.). 2000. *Teaching Collocation: Further Developments in the Lexical Approach* [C]. Hove: Language Teaching Publications.
- McAlpine, J. & J. Myles. 2003. Capturing phraseology in an online dictionary for advanced users of English as a second language: A response to user needs [J]. *System* 31: 71-84.
- McEnery, A. & A. Wilson. 2001. *Corpus Linguistics* [M]. Edinburgh: Edinburgh University Press.
- McEnery, A., Z. Xiao & Y. Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book* [M]. London: Routledge.
- Meunier, F. 2002. The pedagogical value of native and learner corpora in EFL grammar teaching [A]. In S. Granger, J. Hung & S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [C]. Philadelphia: John Benjamins, 119-142.
- Mindt, D. 1996. English corpus linguistics and the foreign language teaching syllabus [A]. In J. Thomas & M. Short (eds.). *Using Corpora for Language Research* [C]. London: Longman, 232-247.
- Murison-Bowie, S. 1996. Linguistic corpora and language teaching [J]. *Annual Review of Applied Linguistics* 16: 182-199.
- Nelson, M. 2000. *A Corpus-based Study of Business English and Business English Teaching Materials*. Unpublished PhD thesis, the University of Manchester, Manchester [OL]. <http://www.kielikanava.com/thesis.html> (accessed 03/04/2006).
- Nesselhauf, N. 2003. The use of collocations by advanced learners of English and some implications for teaching [J]. *Applied Linguistics* 24(2): 223-242.
- O'Keeffe, A. & F. Farr. 2003. Using language corpora in initial teacher education: Pedagogic issues and practical applications [J]. *TESOL Quarterly* 37(3): 389-418.
- Osborne, O. 2001. Integrating corpora into a language-learning syllabus [A]. In B. Lewandowska-Tomaszczyk (ed.). *PALC 2001: Practical Applications in Language Corpora* [C]. Frankfurt: Peter Lang, 479-492.
- Partington, A. 1998. *Patterns and Meanings* [M]. Amsterdam: John Benjamins.
- Renouf, A. 1987. Moving on [A]. In J. Sinclair (ed.). *Looking Up: An Account of the COBUILD Project* [C]. London: HarperCollins.
- Seidlhofer, B. 2000. Operationalizing intertextuality: Using learner corpora for learning [A]. In L. Burnard & A. McEnery (eds.). *Rethinking Language Pedagogy from a Corpus Perspective* [C]. New York: Peter Lang, 207-224.
- Seidlhofer, B. 2002. Pedagogy and local learner corpora: working with learning driven data [A]. In S. Granger, J. Hung & S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [C]. Philadelphia: John Benjamins, 213-234.
- Shei, C. & H. Pain. 2000. An ESL writer's collocational aid [J]. *Computer Assisted Language Learning* 13(2): 167-182.
- Sinclair, J. 2000. Lexical grammar [J]. *Naujoji Metodologija* 24: 191-203.
- Sinclair, J. (ed.). 2004. *How to Use Corpora in Language Teaching* [C]. Amsterdam: John Benjamins.
- Sinclair, J., S. Bullon, R. Krishnamurthy, E. Manning, & J. Todd. 1990. *Collins COBUILD English Grammar* [M]. London: HarperCollins.
- Sinclair, J. & A. Renouf. 1988. A lexical syllabus for language learning [A]. In R. Carter & M. McCarthy (eds.). *Vocabulary and Language Teaching* [C]. London: Longman.
- Smadja, F. & K. McKeown. 1990. Automatically extracting and representing collocations for language generation [A]. In *Proceedings of the 28th Annual Meeting of Association for Computational Linguistics* [C]. Pittsburgh: University of Pittsburgh, 252-259.
- Sripicharn, P. 2000. Data-driven learning materials as a way to teach lexis in context [A]. In C. Heffer, H. Sauntson & G. Fox (eds.). *Words in Context: A Tribute to John Sinclair on His Retirement* [C]. Birmingham: University of Birmingham.
- Stubbs, M. 2001. Texts, corpora, and problems of interpretation: A response to Widdowson [J]. *Applied Linguistics* 22(2): 149-172.
- Tan, M. 2002. *Corpus Studies in Language Education* [M].

- Bangkok: IELE Press.
- Taylor, L. 2003. The Cambridge approach to speaking assessment [J]. *Research Notes* 13: 2-4.
- Thompson, P. & C. Tribble. 2001. Looking at citations: Using corpora in English for academic purposes [J]. *Language Learning and Technology* 5(3): 91-105.
- Thurstun, J. & C. Candlin. 1997. *Exploring Academic English: A Workbook for Student Essay Writing* [M]. Sydney: NCELTR.
- Thurstun, J. & C. Candlin. 1998. Concordancing and the teaching of the vocabulary of academic English [J]. *English for Specific Purposes* 17: 267-280.
- Tribble, C. & G. Jones. 1990. *Concordances in the Classroom: A Resource Book for Teachers* [M]. London: Longman.
- Tribble, C. & G. Jones. 1997. *Concordances in the Classroom: Using Corpora in Language Education* [M]. Houston TX: Athelstan.
- Tribble, C. 1991. Concordancing and an EAP writing program [J]. *CAELL Journal* 1(2): 10-15.
- Tribble, C. 1997a. Corpora, concordances and ELT [A]. In T. Boswood (ed.). *New Ways of Using Computers in Language Teaching* [C]. Alexandria VA: TESOL.
- Tribble, C. 1997b. Improving corpora for ELT: Quick and dirty ways of developing corpora for language teaching [A]. In B. Lewandowska-Tomaszczyk, P. Melia (eds.) *Practical Applications in Language Corpora—Proceedings of PALC '97* [C]. Łódź: Łódź University Press, 107-117.
- Tribble, C. 2000. Practical uses for language corpora in ELT [A]. In P. Brett & G. Motteram (eds.). *A Special Interest in Computers: Learning and Teaching with Information and Communications Technologies* [C]. Kent: IATEFL, 31-41.
- Tribble, C. 2003. The text, the whole text ... or why large published corpora aren't much use to language learners and teachers [A]. In B. Lewandowska-Tomaszczyk (ed.). *Practical Applications in Language and Computers* [C]. Frankfurt: Peter Lang, 303-318.
- Upton, T. & U. Connor. 2001. Using computerized corpus analysis to investigate the textlinguistic discourse move of a genre [J]. *English for Specific Purposes* 20: 313-329.
- Wichmann, A. 1995. Using concordances for the teaching of modern languages in higher education [J]. *Language Learning Journal* 11(1): 61-63.
- Wichmann, A. 1997. General introduction [A]. In A. Wichmann, S. Fligelstone, A. McEnery & G. Knowles (eds.). *Teaching and Language Corpora* [C]. London: Longman, xvi-xvii.
- Wichmann, A., S. Fligelstone, A. McEnery & G. Knowles (eds.). 1997. *Teaching and Language Corpora* [C]. London: Longman.
- Widdowson, H. 1990. *Aspects of Language Teaching* [M]. Oxford: Oxford University Press.
- Widdowson, H. 2000. The limitations of linguistics applied [J]. *Applied Linguistics* 21(1): 3-25.
- Willis, D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching* [M]. London: HarperCollins.
- Woolfs, D. 1998. Multilingual parallel concordancing for pedagogical use [R]. Paper presented at *Teaching and Language Corpora*. Keble College, Oxford, 24-27 July 1998, 222-227.
- Xiao, Z. & A. McEnery. 2006. Collocation, semantic prosody and near synonymy: A cross-linguistic perspective [J]. *Applied Linguistics* 27(1): 103-129.
- Yang, Y. & D. Allison. 2003. Research articles in applied linguistics: Moving from results to conclusions [J]. *English for Specific Purposes* 22(4): 365-385.
- Zhang, X. 1993. English Collocations and Their Effect on the Writing of Native and Non-native College Freshmen [D]. Unpublished PhD thesis. Indiana University of Pennsylvania.
- 陈新仁、任育新, 2007, 中国高水平英语学习者重述标记语使用考察 [J], 《外语教学与研究》(4): 294-300。
- 桂诗春、杨惠中, 2003, 《中国学习者英语语料库》[M]。上海: 上海外语教育出版社。
- 何安平 (编), 2004a, 《语料库在外语教育中的应用: 理论与实践》[C]。广州: 广东高等教育出版社。
- 何安平, 2004b, 《语料库语言学与英语教学》[M]。北京: 外语教学与研究出版社。
- 何安平、徐曼菲, 2003, 中国大学生英语口语 Small Words 的研究 [J], 《外语教学与研究》(6): 446-452。
- 胡学文, 2007, 中国学生英语双宾构式的习得——一项基于语料库的对比研究 [J], 《外语研究》(5): 48-53。
- 华南师范大学外国语学院 (编), 2005, 《语料库语言学的研究与应用》[C]。长春: 东北师范大学出版社。
- 李文中, 2003, 基于英语学习者语料库的主题词研究 [J], 《现代外语》(3): 283-293。
- 梁茂成, 2007, Bridging the gap between corpora and language testing [R]. 2007 语言测试国际研讨会

- (北京, 2007年11月3-4日)宣读论文。
- 梁茂成, 2008, 中国大学生英语笔语中的情态序列研究 [J], 《外语教学与研究》(1): 51-58。
- 马广惠, 2002, 中美大学生英语作文的语言特征分析 [J], 《外语教学与研究》(5): 345-349。
- 濮建忠, 2003a, 《学习者动词行为: 类联接、搭配及词块》[M]。开封: 河南大学出版社。
- 濮建忠, 2003b, 英语词汇教学中的类联接、搭配及词块 [J], 《外语教学与研究》(6): 438-445。
- 王立非、梁茂成等(编), 2007, 《计算机辅助第二语言研究方法与实践》[C]。北京: 外语教学与研究出版社。
- 王立非、祝卫华, 2005, 中国学生英语口语中话语标记语的使用研究 [J], 《外语研究》(3): 40-45。
- 卫乃兴、李文中、濮建忠(编), 2005, 《语料库应用研究》[C]。上海: 上海外语教育出版社。
- 文秋芳、丁言仁、王文宇, 2003, 中国大学生英语书面语中的口语化倾向 [J], 《外语教学与研究》(4): 268-274。
- 文秋芳、王立非、梁茂成, 2005, 《中国学生英语口语笔语语料库》[M]。北京: 外语教学与研究出版社。
- 肖忠华, 2003, 平行语料库与可比语料库在语言研究中的应用 [J], 《中国英语教育》(1)。
- 许家金、许宗瑞, 2007, 中国大学生英语口语中的互动话语词块研究 [J], 《外语教学与研究》(6): 438-445。

- 杨达复, 2000, 《英语错误型式分析》[M]。西安: 陕西人民出版社。
- 杨惠中(编), 2005, 《基于CLEC语料库的中国学习者英语分析》[C]。上海: 上海外语教育出版社。
- 杨惠中、卫乃兴(编), 2005, 《中国学习者英语口语语料库建设与研究》[C]。上海: 上海外语教育出版社。
- 俞洪亮, 2005, 《中国大学生口头叙事话语的结构范式研究》[M]。北京: 外语教学与研究出版社。

作者简介

肖忠华(1966—), 浙江大学外国语言文化与国际交流学院教授、博士生导师, 英国兰卡斯特大学语言学专业博士。主要研究领域: 语料库语言学、语言对比与翻译、英语语言学、应用语言学。电子邮箱: lancsxiao@googlemail.com。

许家金(1977—), 北京外国语大学中国外语教育研究中心专职研究员。主要研究领域: 话语分析、语用学、语料库语言学。电子邮箱: ustcxujj@sina.com。