

外研社学术文库·当代国外语言学与应用语言学

An Introduction to Corpus Linguistics

语料库语言学入门

Graeme Kennedy

外语教学与研究出版社

Foreign Language Teaching and Research Press

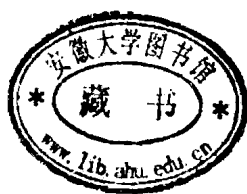
外研社学术文库·当代国外语言学与应用语言学

An Introduction to Corpus Linguistics

语料库语言学入门

Graeme Kennedy 著

顾曰国 导读



外语教学与研究出版社
北京

京权图字: 01-2000-1899

English edition © Addison Wesley Longman Limited 1998

All rights reserved. No part of this publication may be reproduced, stored or transmitted by any means without the prior permission of the publishers.

This edition of *An Introduction to Corpus Linguistics*, First Edition is published by arrangement with Pearson Education Limited. It is for sale in the mainland territory of the People's Republic of China only, excluding Hong Kong.

本书由培生教育出版集团授权外语教学与研究出版社出版

图书在版编目(CIP)数据

语料库语言学入门/(新)肯尼迪(Kennedy, G.)著;顾曰国导读. — 北京:外语教学与研究出版社, 2012. 8

(当代国外语言学与应用语言学文库)

ISBN 978-7-5600-8529-6

I. ①语… II. ①肯… ②顾… III. ①语料库语言学—基本知识
IV. H0

中国版本图书馆 CIP 数据核字(2000)第 66943 号

出 版 人: 蔡剑峰

责任编辑: 霍庆文

出版发行: 外语教学与研究出版社

社 址: 北京市西三环北路 19 号 (100089)

网 址: <http://www.fltrp.com>

印 刷: 北京京科印刷有限公司

开 本: 650×980 1/16

印 张: 21.5

版 次: 2012 年 8 月第 1 版 2012 年 8 月第 1 次印刷

书 号: ISBN 978-7-5600-8529-6

* * *

购书咨询: (010)88819929 **电子邮箱:** club@fltrp.com

如有印刷、装订质量问题, 请与出版社联系

联系电话: (010)61207896 **电子邮箱:** zhijian@fltrp.com

制售盗版必究 举报查实奖励

版权保护办公室举报电话: (010)88817519

物料号: 185290001

Preface by Halliday

Foreign Language Teaching & Research Press is to be congratulated on its initiative in making these publications in linguistics available to foreign language teachers and postgraduate students of linguistics in China.

The books are a representative selection of up-to-date writings on the most important branches of linguistic studies, by scholars who are recognized as leading authorities in their fields.

The availability of such a broad range of materials in linguistics will greatly help individual teachers and students to build up their own knowledge and understanding of the subject. At the same time, it will also contribute to the development of linguistics as a discipline in Chinese universities and colleges, helping to overcome the divisions into “English linguistics”, “Chinese linguistics” and so on which hinder the progress of linguistics as a unified science.

The series is to be highly commended for what it offers to all those wanting to gain insight into the nature of language, whether from a theoretical point of view or in application to their professional activities as language teachers. It is being launched at a time when there are increasing opportunities in China for pursuing linguistic studies, and I am confident that it will succeed in meeting these new requirements.

M. A. K. Halliday
Emeritus Professor
University of Sydney

王宗炎序

近年来，国际交往日益频繁，国际贸易急速发展，出现了一种前所未有的现象：学外语、教外语、用外语的人多了；研究语言学和应用语言学的人多了；开设这方面专业的高校也多了，语言学硕士生和博士生也多了。就是不以此为专业，学习语言学和应用语言学的也不乏其人。为了给从事这个专业的师生提供便利，同时又帮助一般外语教师、涉外工作者以及汉语研究者开阔思路，扩大视野，提高效率，我们献上这套内容崭新而丰富的丛书——英文版《当代国外语言学与应用语言学文库》。

文库首批推出 54 部外国英文原著，它覆盖了语言学与应用语言学 28 个分支学科。这批书是我们与各地有关专家教授反复研究之后精选出来的。出版这样大规模的语言学与应用语言学丛书，这在我国语言学界和外语教学界是破天荒第一次。

我们这样做，抱着什么希望呢？总的说来，是遵循教育部关于加强一级学科教育的指示，在世纪之交，推出一套书来给中国的外语教育领航，同时也给一般外语工作者和汉语研究者提供信息，拓宽思路。

我们希望这个文库能成为进一步带动外语教学改革和科研的发动机；我们希望它能成为运载当代外国语言学理论、语言研究方法和语言教学方法来到中国的特快列车；我们希望，有了这套书，语言学与应用语言学专业师生就能顺利地进行工作；我们希望，通过读这套书，青年外语教师和外语、汉语研究者能迅速把能力提高，把队伍不断扩大。

以上是我们的愿望，可是从广大读者看来，这个文库是否真的有出台的有必要呢？我们想，只要大家看一下今天

的客观情况，就知道这套书有填空补缺的作用，是让大家更上一层楼的扶梯。

我们跟许多人一样，认为国内的外语教学和语言学与应用语言学研究是成绩斐然的，但是某些不足之处也无庸讳言。

在语言研究方面，有大量工作还等着大家去做。汉语语法研究，过去由于结构主义的启示，已经成绩卓著，可是现在虽则引进了功能主义，还看不出什么出色的成果。语料语言学是新兴学科，在我国刚刚起步，机器翻译从 50 年代就有人搞，然而其进展至今不能令人满意。

在语言理论方面，我们不时听到一些片面的、所见不全的论调。有人说，1957 年前西方根本没有什么理论语言学，其创始者是 Chomsky；也有人说，语言纯属社会文化范畴；还有人说，搞语言研究只有量化方法才是科学方法，定性方法不值得一提。

谈到外语教学，某些看法做法是分明不值得赞许的。有人以为交际教学只管听说，不管读写，也有人以为教精读课就是教阅读，不管口语。在分析课文时老师满堂灌，学生开口不得，是常见的；教听力课时老师只管放录音，对学生不给半点提示点拨，也并非罕有现象。

上述这些缺点，我们早有所知，现在我们更加明白，必须力图改进，再也不能安于现状了。为了改进，我们就得参考国外的先进理论，借鉴国外的有效措施。眼前这个文库，就是我们上下求索的结果。

在编辑这个文库时，我们在两方面下了功夫。

一方面，在选书时，我们求全，求新，求有代表性和前瞻性。我们不偏爱一家之言，也不只收一家外国出版社之书。语言学与应用语言学的主干学科固然受到了应有的重视，分支学科可也不忽视。语料语言学、语言统计学是新兴学科，我们收入了专著；句法学、语义学久已有人研究，我们也找到了有关的最新著作。

另一方面，我们邀请了国内知名的博士生导师、硕士生导师为各书撰文导读，为读者铺平道路。语言学和应用语言学专著包罗宏富，初学者读起来可能觉得茫无头绪。为了助他们一臂之力，本文库中每一种书我们都请专家写了一万字左右的导读材料。哪怕书中内容比较陌生，谁只要在读书前看一下导读材料，读书后把材料再看一遍，一定能弄清脉络，掌握要点。

在结束本文时，我们想向爱好泛读的人们提个建议。语言和社会生活息息相关；我们靠语言与他人协作；通过语言继承传统文化，接受外国先进思想和科学知识；利用语言来教育下一代，帮助他们创造美好的未来；语言又反过来表达着我们的个性和我们充当的各种角色。学一点语言学和应用语言学，有助于增强我们的语言意识，对我们的工作和生活都是有利的。我们不妨把此事作为一个项目，列入自己的日程。持之以恒，必有所获。

王宗炎

中山大学教授
博士生导师

导 读

谈起读书，我们会想起英国哲学家 Francis Bacon (1561-1626) 说过的话：

Some books are to be tasted, others to be swallowed, and some few to be chewed and digested.

这句话一方面告诉我们读书的方法，不能随手拿到一本书就读起来；另一方面，是告诫我们择书阅读要慎重，值得仔细揣摩玩味的书是不多的。这两层意思在今天信息爆炸的时代尤其不失其指导价值。

那么，《语料库语言学入门》（以下简称《入门》）是一本什么样的书呢？属于翻翻而已，还是要浏览一番，还是值得揣摩玩味？这些问题不是片言只语就能回答的。这正是需要写这个导读的缘故。

一本书值不值得读，一是看书本身的质量，二是看读书人的兴趣。最终读不读还要看读书人有没有时间，把读书放在什么样的位置上。这些话听起来像是有点离题，其实不然。假如您从事语料库语言学研究已经多年，这本书对您来说就不值得深读，浏览一番足已。假如您对语料库语言学一无所知，那么这本书就值得您仔细揣摩。假如您已经知道一点，想作更深的了解，您就可以有选择地去读其中的章节。本导读的主要对象是尚未涉足但对语料库语言学有兴趣的读者以及已经具有初步语言学知识的读者。

下面我们首先介绍语料库语言学在当今语言学领域所处的位置，然后谈谈语料库语言学的几个核心问题，最后我们带着这些背景知识来读《入门》。

一、语料库语言学的位置

我们知道，二十世纪语言研究形成两个大的阵营，结构主义和功能主义。多年来结构主义一直占上风，西方学者称为“语言学主流”（mainstream linguistics）。功能主义相比之下处于劣势，守势。两阵对垒的天平近十几年来似乎在发生变化。如果我们把新兴学科如社会语言学、语用学、会话分析、语篇分析等都划入功能主义（因为这

些都是研究语言的各种功能的),那么功能主义大有取代结构主义成为语言学主流之势。

语料库和语料库语言学可以说是两阵对垒的天平上的一个举足轻重的砝码。语料库顾名思义就是放语言材料的仓库。现在人们谈起语料库不言而喻地指存放在计算机里的原始语料文本或经过加工后带有语言学信息标注的文本。在计算机出现之前,研究者特别是词典编纂者也有语料库,只是规模小范围窄而难以形成气候。索绪尔在世的年代,连录音机还闻所未闻,更谈不上计算机,对语言进行大规模的实际调查是十分困难的,想建立一个亿万字的语料库在当时是十足的天方夜谭!现在世界上已经有了不少规模较大的语料库,有些是国家级的,有些是大学和词典出版商联合的。另外,由于个人微机的迅猛发展和存贮数据的硬盘造价持续下降,研究者个人也开始建立适合于自己研究的小型语料库。

“语料库语言学”(corpus linguistics)这个术语其实有两层含义。一是利用语料库对语言的某个方面进行研究,也就是说“语料库语言学”不是一个新学科的名称,而仅仅反映了一个新的研究手段。二是依据语料库所反映出来的语言事实对现行语言学理论进行批判,提出新的观点或理论。只有在这个意义上“语料库语言学”才是一个新学科的名称。从现有的文献来看,属于后一类的研究还是极个别。所以,严格地说,我们现在不能把语料库语言学跟社会语言学、心理语言学、语用学等相提并论。

从实际语料出发,让实际语料说话,是我国语言学家研究语言历来的传统。由于科学技术的落后和科研经费的不足,我们的优秀传统未能发展成像西方人现在做的语料库和相应的语料库语言学。我们不得不从西方语言学的大环境下来审视语料库和语料库语言学。

为什么说语料库和语料库语言学(无论是上文说的哪一层意义)是结构主义和功能主义两阵对垒的天平上一个举足轻重的砝码呢?要回答这个问题,我们不得不把话题扯远点,回顾一下结构主义作为语言学的主流所走过的道路,以及功能主义与结构主义分歧的焦点。

我们知道,虽然有关语言的起源问题至今尚未有定论,但对语言发展的总体脉络的认识基本上是一致的。日常生活中自然而然的谈话是语言的最原始和最古老的形式,语言的其他功能以及书面语言都是后来发展起来的。现在世界上大约有 5000 到 6000 种语言(准确的数字要看如何界定语言跟方言的区别),而其中有文字的还

不到一半。儿童学母语也是从日常生活需要开始的，说话——从单字语到双字语再到多字语——是儿童达到某种实用目的的重要手段。

我们也知道，不仅语言之间有很大的差异，连同一语言内的方言之间也有很大的区别，许多甚至都无法进行交流。另外，操同一语言或方言的人说出来的话也是五花八门，各有各的腔调。我们如果再把时间的因素考虑进去，语言由于时间的变化而形成的差异同样也是惊人的。

以上这些可以说是语言的实际情况。现代语言学的鼻祖索绪尔深知语言的这些实际情况，以及研究这些差异的难度。他的做法是排除差异，找出共性，用他的话说，(共时)语言学研究的是 *langue*，即同质的、抽象的语言形式，而不是 *parole*，即异质的、五花八门的言语。这样，结构主义语言学家研究的语言是凌驾于个人和社会之上的一个具有高度独立性的抽象的符号系统。这些语言学家面临着一个非常棘手的难题：他们看得见听得着的是异质的、五花八门的言语，要研究的却是同质的抽象符号系统。早期的结构主义者（以美国 Bloomfield 为代表）不辞劳苦地对异质的、五花八门的言语进行调查，从中找出同质的、共性的形式结构。他们的方法无异于沙里淘金。由于他们采取了严格的筛选标准，在语音、音位和形态方面取得了举世瞩目的成就。在词组、句法层次上他们开始遇到了麻烦。首先，词组、句子不同于语音、音位和形态，它们不是一个有限的集，而是一个无限的集，不管他们调查了多少异质的、五花八门的言语，他们也不敢说已经找到了构成某个语言的全部词组结构和句法结构，语义层次上的问题就更多了。

Chomsky 提出转换生成语法理论，可以说帮助结构主义摆脱了困境。我们知道他区分了 *competence*（语言能力）和 *performance*（语言使用）。这可不是对索绪尔的 *langue* 和 *parole* 两个概念的简单术语替换。*Langue* 是凌驾于个人、属于语言社团的抽象系统，*competence* 是个人能生成无限言语的、根植于大脑之中的语言知识。Chomsky 把句法研究的主要任务视为找出那些能生成无限句子的有限规则。无论是索绪尔，还是 Bloomfield，都十分重视实际语言调查，而 Chomsky 则恰恰相反，对同一语言来说，他不需要做语言调查，只要问他自己大脑中的“语言直觉”就行了。在 Chomsky 的理论里，语言是生理和心理现象，人们说出来的话（即实际语料）不能完全反映语言的本质。他的最有力的证据就是人们能说出他们从

来没有听别人说过的句子。也就是说，不管语料库有多大，所收集到的都是已经说出来的，而没有说出的要更多。归根到底，语料库是有限的，语言是无限的。

Chomsky 的理论至少有两个明显的弱点。一是他的理论否定了语言的社会性。语言是为了人际间的交流，离开了交流，语言失去了存在的价值。二是虽然语料库是有限的，但比起某个人的语言直觉来无疑为研究者提供了更多而有效的线索去寻求语言的本质。

承认语言的社会性是功能主义区别 Chomsky 的心灵主义的一个重要方面，也直接关系到对语料库的态度和使用上。功能主义（以 Halliday 为代表）不同于转换生成语法，它最大的特点是强调语言的社会性，语言的使用功能为语言结构作出最终的解释。Halliday 在他的《功能语法导论》一书的序中写道（1985: Xiii）：

……是千万年来一代代人使用语言才使得语言形成今天这样一个系统。语言为了满足人类的需要在发展；它的组成结构之所以是功能性的正是因为这些需求。它不是任意性的。一部功能语法本质上是一部“自然”语法，即书中所说的一切都能在语言的使用方式中找到最终的解释。

功能主义注重语言实际使用情况，语料库是他们不可多得的帮手。语料库语言学对功能主义的发展会提供很多具有相当说服力的佐证。

前面说过，日常生活中自然而然的谈话是语言的最原始和最古老的形式，按理说，语言研究应该首先从这里入手。然而对自然谈话的研究可以说是刚刚起步，口语语料库与其它类语料库相比还少得可怜而且规模很小。多少年来，甚至可以说自古以来，书面语言是语言学家的唯一语料来源。由于这个缘故，大多数语言学家是先从句法入手研究语言，再从句法向其它层次如语义、语用等发展。早期的转换生成语法理论甚至坚持句法独立于其它层次的观点。这实际上是与语言发展的自然规律相悖的。我国学者研究古音古律也靠书面语（这是没有办法的办法），赵元任的 *A Grammar of Spoken Chinese* (1968) 在当时确实是非常超前的研究，因为涉及了不少口语现象。现在看来，他所分析的口语面较窄，研究方法局限于结构主义分析句法的方法，跟现在的要求相比，结果不够理想（就这一点可参阅 Shen and Gu, 1997）。我们在这里不是贬低前人的成就，

也不是求全责备，而是要站在这些巨人的肩头看到他们所处的时代给他们造成的局限，要更上一层楼。

二、语料库语言学的几个核心问题

上文说过，在现阶段，语料库语言学主要是关于语言研究的手段和方法。它的核心问题有以下五个方面：1. 建立语料库；2. 对语料进行转写；3. 对语料进行标注；4. 开发出管理和利用语料库的软件；5. 利用建立好的语料库对语言的某些方面进行研究。下面我们就这五个方面作简要说明。

1. 建立语料库

我们要考虑的因素很多，如建库的目的，语料品种，取样标准，规模，人力及资金来源等。另外还要考虑语料的存贮方式与格式。

2. 语料转写

如果要建立的语料库包括口语，把声音转写成文字和计算机可读的声音符号就非常重要，也是一项技术性非常强、人力资源投入巨大的工作。

3. 语料标注

所谓语料标注，就是把语料所具有的重要语言学信息用一组符号一一标注出来。以“吃饭”为例。其重要的语言学信息有（a）由“吃”+“饭”构成动宾结构，（b）可以做谓语；（c）主语是有生命的，等等。通俗地说，不带标注的语料库叫“生语料库”，标注过的叫“熟语料库”。生语料库给研究者的信息是有限的，标注得越是详细，研究者从中能得到的信息也就越大。

4. 管理与应用软件

开发语料库管理与应用软件跟建库同样重要。在现阶段，不少语料库开发者容易忽视这一点，导致语料库资源的闲置和浪费。

5. 语言研究

利用语料库对语言的某些方面进行大规模的研究是语料库语言学的最大潜力所在。目前这类研究最有成效的领域是词典学和句法学。

三、《入门》其书

《入门》共分五章，分别为引言、语料库的设计与开发、在语料

库基础上对英语作出的描写、语料分析和基于语料库的研究的实用意义及其前景。

引言是关于语料库语言学的一些常识介绍，其中包括语料库语言学的范围。已经有语料库语言学知识的读者简单浏览一下即可。

第二章“语料库的设计与开发”，是《入门》的核心章节之一。作者首先回顾语料库的历史，使读者对语料库有一个立体的认识。接着介绍几个较大的电子语料库，如 Brown Corpus, LOB Corpus, London Lund Corpus 等。再接着介绍的是几个特殊用途的语料库，以及第二代超大型语料库，如 COBUILD Corpus, Longman Corpus, British National Corpus 等。最后作者才回到语料库的设计上。读者先熟悉成果，再来看设计，这样安排有利于激发读者的兴趣，在读语料库设计这一节时做到有例有据。未涉足语料库语言学的读者和已经有一定了解的读者都应该仔细阅读这一章。

第三章“在语料库基础上对英语作出的描写”，介绍的是利用英语语料库对英语进行研究的一些成果。这一章对于从事英语研究和教学的读者来说也是很有参考价值的。如果读者正在准备建自己的语料库，对这一章则浏览一下即可。如果读者已经有了自己的语料库，则应仔细阅读，学习他人的研究方法和研究的着眼点。

第四章“语料分析”，介绍的是语料的标注和处理等研究，同时还介绍了处理英语语料库的几个常用软件。准备建库或正在建库的读者应当仔细阅读这一章。

最后一章“基于语料库的研究的实用意义及其前景”，作者主要就英语的研究成果发表自己的看法，实用意义则放在语言教学上，探讨语料库语言学对英语教学及其教学法带来的深刻影响。这跟作者自己的工作有关，也暴露了作者的视野的局限性。不过作者还是单辟一节谈语料库语言学跟计算语言学的关系，使读者看到语料库语言学对计算语言学所产生的深刻影响，也使读者看到了语料库语言学在现代信息高科技中所处的至关重要的地位。

至此，我们已经做了三件事：了解跟语料库语言学有关的一些背景知识，弄清语料库语言学的几个核心问题，对《入门》有了总体上的认识。下面请读者根据自己的实际情况订个读书计划，亲自品尝一下“梨子”的真实味道。

Author's acknowledgements

I owe a great deal to colleagues, students, friends and members of my family who have helped extend my involvement in corpus linguistics, and who, through their own research, comments and questions have encouraged, supported and enlightened me. The book owes a lot to Jane Dudley, Paul Nation and Robert Sigley, who each contributed in different ways.

Several graduate students who have worked with corpora during the last decade have also contributed to my own understanding of the nature, structure, acquisition and use of English. They include Morie Chan, Uthaivan Danvivath, Ann Devoy, Fang Xuelan, Susan Gardner, Mary Greenfield, He Anping, Alastair Ker, Jonathan Newton, Cucu Sutarsyah, Michelle Strand, Wang Sheng and Shunji Yamazaki.

I am grateful to the many authors who have given permission to reproduce data from their research. Every effort has been made to acknowledge data from published research wherever the source can be traced.

Corpus linguistics would not be the dynamic field of scholarship it has become without the work in corpus compilation and software design which many scholars have undertaken since the pioneering work of Randolph Quirk, Nelson Francis, Henry Kucera, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. Anyone working in the field of corpus linguistics today also owes a debt to a growing network of colleagues in many countries. In particular, I wish to acknowledge the work of Bengt Altenberg, who has supported many researchers with his bibliographical work; and also Stig Johansson and Knut Hofland, who, through ICAME and the Norwegian Computing Centre for the Humanities, have helped extend the availability of the major first-generation computer corpora on which corpus linguists have cut their teeth. I am especially grateful to Geoffrey Leech for his encouragement and painstaking editorial work, which did much to improve initial drafts of the book.

Publisher's acknowledgements

We are grateful to the following for permission to reproduce copyright material:

Association for Literary and Linguistic Computing for an extract from A Govindankutty, *ALLC Bulletin* (1973), 1, 3: 3; Guardian Syndication for the article "'Doomsday rocks" rattle Nasa' by Simon Tisdall © *The Guardian Weekly* 19.4.92.

Brown University Library and the authors W. N. Francis and H. Kucera for 'Structure of the *Brown Corpus*' (our Table 2.2) from *Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers* 1964 (revised 1979); Inso Corporation and the authors W. N. Francis and H. Kucera for 'A lemmatized entry for *collect* and related words' (our Figure 3.3), 'Perfect and progressive verb forms in the *Brown Corpus*' (our Table 3.16), 'The tagset used for tagging the *Brown Corpus*' (our Figure 4.1) all from *Frequency Analysis of English Language: Lexicon and Grammar* (1982); RELC (Regional Language Centre, Southeast Asian Ministers of Education Organization) for 'The 50 most frequent words used in an economics text compared with a corpus of the same size of general academic English (our Table 3.4) from Sutarsyah, Nation & Kennedy 'How useful is EAP vocabulary for ESP? – A Corpus-based Study' *RELC Journal* 25, 2: 35–40, 'Immediate right collocates of *at* occurring four or more times in the *LOB Corpus*, (our Table 3.27) from 'Collocations: Where grammar and vocabulary teaching meet' in Anivan, S. (ed.) *Language Teaching Methodology for the Nineties*. RELC Anthology Series No. 24, Regional Language Centre, 215–229, and 'Explicit devices marking causation in the *LOB Corpus*' (our Table 3.50) from Fang and Kennedy (1992: 68) in 'Expressing causation in written English' *RELC Journal* 23, 1: 62–60; Lund University Press and Bengt Altenberg for 'Relative proportions of major word classes in a sample (c. 50,000 words) from the *LLC*' (our Table 3.6) and 'Distribution of discourse items in a sample of 50,000 words from

the *LLC*' (our Table 3.53) both from 'Spoken English and the Dictionary' (1990: 185, 183) also Lund University Press and Anna-Brita Stenstrom for 'Categories of discourse items in the *LLC*' (our Table 3.54) and 'Categories of discourse items in dialogue and monologue' (our Table 3.55) both from 'Lexical items peculiar to spoken discourse' – all four tables in *The London-Lund Corpus of Spoken English: Description and Research* (1990) edited by Jan Svartvik; *ICAME News* and Stig Johansson for 'Most distinctive nouns, verbs, adjectives and adverbs in two genres of the *LOB Corpus*' (our Table 3.8) from 'Word frequencies in different types of English texts' *ICAME News* 5: 7–8 (1981); *ICAME News* and Eric Akkerman for 'Particles used with verbs in the *Brown Corpus*' (our Table 3.25) from 'Verb and particle combinations: Particle frequency ratings and idiomaticity' *ICAME News* 8: 63 (1984); *ICAME News* for 'A fragment of a tagged version of the *LOB Corpus*, horizontal format' (our Figure 4.2); *ICAME Journal* and Kyto and Voutilainen for 'A sample of the output of the *ENGCG Parser*' (our Figure 4.15) from 'Applying the constraint grammar parser of English to the Helsinki Corpus' *ICAME Journal* 19: 23–28 (1995); Rodopi and Henk Barkema for 'The occurrence of certain collocations in two corpora' (our Table 3.11) from 'Idiomaticity in English Nps' in Aarts, De Haan and Oostdijk (eds.) *English Language Corpora: Design, Analysis and Exploitation* (1993: 271); Rodopi and Karin Aijmer for 'Frequencies in spoken and written corpora of some individual hedges' (our Table 3.53) from 'Discourse variation and hedging' in Aarts and Meijs (eds.) *Corpus Linguistics II. New Studies in the Analysis and Exploitation of Computer Corpora* (1986: 10); Rodopi and Geoffrey Sampson for 'A fragment of the *SUSANNE Corpus*' (our Figure 4.13) from 'A domesday book of English grammar' in Oostdijk and De Haan (eds.) *Corpus-based Research into Language: In Honour of Jan Aarts* (1994: 175); Rodopi and Nellche Oostdijk for 'Format of a parsed sentence using the Nijmegen approach to the automatic analysis of English' (our Figure 4.14) from *Corpus Linguistics and the Automatic Analysis of English* (1991: 162); Mouton de Gruyter for 'Most frequent verbs in Ota's Corpus, and their tense and aspect use' (our Table 3.13) and 'Use of modals in the *London-Lund* and *LOB* corpora' (our Table 3.20) both from Kennedy (1992) 'Preferred ways of putting things with implications for language teaching in Svartvik (ed.) *Directions in Corpus Linguistics*'; Mouton de Gruyter and Bengt Altenberg for 'Amplifiers in recurrent combinations in the *LLC*' (our adapted Table 3.56) and 'Most frequent amplifier collocations in the *LLC*' (our Table 3.57) both from Altenberg (1991) 'Amplifier collocations in spoken English' in Johansson and Stenstrom (eds.) *English Computer Corpora*; Mouton de Gruyter and Matti Rissanen for 'Occurrence of periphrastic *do* in affirmative statements and of the progressive form in Early Modern English sub-periods in the *Helsinki Corpus* (excluding Bible transactions)' (our Table 3.62) from Rissanen (1992) 'The diachronic corpus as a window to the history of English' in

PUBLISHER'S ACKNOWLEDGEMENTS

Svartvik (ed.) *Directions in Corpus Linguistics*; Mouton de Gruyter for our Figures 4.9, 4.10 and 4.11 from Leech and Svartvik 'Running a grammar factory. The production of syntactically analyzed corpora or "treebanks"' in Johansson and Stenstrom (eds.) *English Computer Corpora*; Oxford University Press for 'Relative frequencies of use of categories of quantification' (our Table 3.39) in 'Quantification and the use of English: A case study of one aspect of the learner's task' *Applied Linguistics Volume 8*, No. 3, also 'Frequencies of different types of postnominal modifiers in three registers (per 1,000 words)' (our Table 3.45) and 'Linguistic characterization of ten spoken and written registers with respect to dimension 1 "Involved versus Informational Production", and dimension 5 "Non-Abstract versus Abstract Style"' (our Figure 3.11) both from Biber, Conrad & Reppon 'Corpus-based approaches to issues in applied linguistics' *Applied Linguistics* 51, 2: 173, and 'Outline of major grammatical and discoursal units and structures which might be studied in a corpus' from 'The corpus as a research domain' in Greenbaum (ed.) (1996) *Comparing English Worldwide: The International Corpus of English*; Cambridge University Press and the author Douglas Biber for 'Features used in Biber's multi-feature, multi-dimensional analysis of variation' (our Figure 3.9) and 'Factorial structure of two text type dimensions' (our Figure 3.10) from *Variation Across Speech and Writing* (1988); Addison Wesley Longman for 'Some habitually spoken sequences in Australian and New Zealand English' (our Figure 3.4) from Pawley and Snyder chapter 'Two puzzles for linguistic theory: Native-like selection and native-like fluency' in Richards & Schmidt (eds.) *Language and Communication* (1983), and 'Semantic classes of apposition' (our Table 3.43) and 'Total number of appositions per genre' (our Table 3.44) from Meyer chapter 'A corpus-based study of apposition in English', also 'Must in Australian, British and American English' (our Table 3.61) from Collins chapter 'The modals of obligation and necessity in Australian English', also 'Zero and *that* as object clause links in Late Middle English' (our Table 3.63) and 'Zero and *that* as object clause links in Early Modern English' (our Table 3.64) both from Rissanen chapter 'On the history of *that* / zero as object clause links in English', all five tables from Aijmer and Altenberg (1991) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*; Blackwell Publishers and the Editorial Board of *Studia Linguistica* for 'Distribution of the fifteen most frequent causal links in LLC and LOB' (our Table 3.49) from Altenberg article 'Causal linking in Spoken and Written English' from *Studia Linguistica* 38 (1984); Academic Press Inc. and G. Tottie for 'Distribution of negation types with *be*, *have* and lexical verbs' (our Table 3.51) from *Negation in English Speech and Writing* (1991).