



“十二五”普通高等教育本科国家级规划教材

新经典
ENGLISH MAJOR

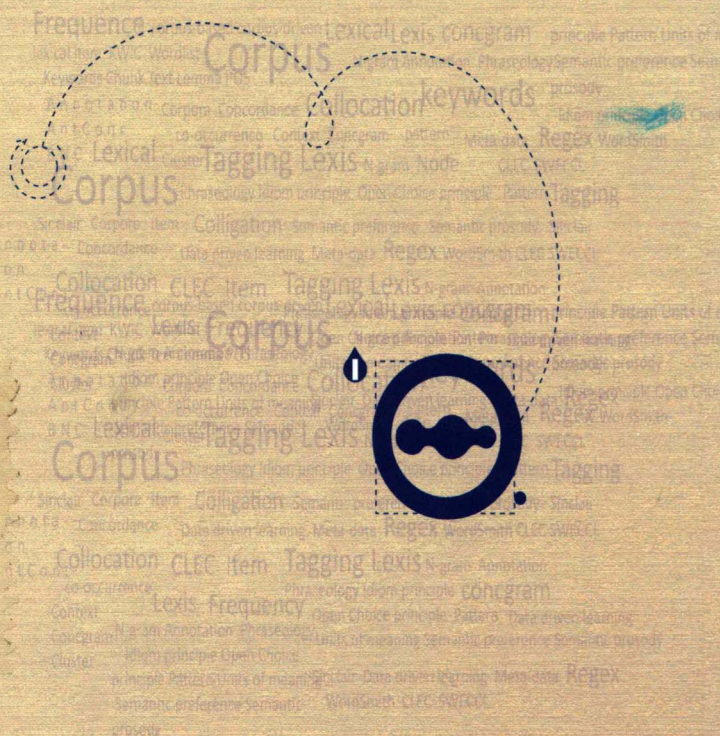
高等学校英语专业系列教材

语料库辅助英语 教学入门

(修订版)

主 编：何安平

副主编：梁红梅 唐洁仪



外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS



新经典
ENGLISH MAJOR

“十二五”普通高等教育本科国家级规划教材

高等学校英语专业系列教材

语料库辅助英语 教学入门

(修订版)

主 编：何安平

副主编：梁红梅 唐洁仪

外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

北京 BEIJING

图书在版编目 (CIP) 数据

语料库辅助英语教学入门 / 何安平主编. — 修订本. — 北京: 外语教学与研究出版社, 2017.2 (2017.11 重印)

新经典高等学校英语专业系列教材

ISBN 978-7-5135-8569-9

I. ①语… II. ①何… III. ①语料库—应用—英语—教学研究—高等学校
IV. ①H319.3

中国版本图书馆 CIP 数据核字 (2017) 第 046914 号

出版人 蔡剑峰
责任编辑 任 佼
执行编辑 黄 骥
封面设计 彩奇风
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 北京九州迅驰传媒文化有限公司
开 本 650×980 1/16
印 张 14
版 次 2017 年 3 月第 1 版 2017 年 11 月第 3 次印刷
书 号 ISBN 978-7-5135-8569-9
定 价 35.90 元

购书咨询: (010) 88819926 电子邮箱: club@fltrp.com

外研书店: <https://waiyants.tmall.com>

凡印刷、装订质量问题, 请联系我社印制部

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

凡侵权、盗版书籍线索, 请联系我社法律事务部

举报电话: (010) 88817519 电子邮箱: banquan@fltrp.com

法律顾问: 立方律师事务所 刘旭东律师

中咨律师事务所 殷 斌律师

物料号: 285690001

再版前言

本书是《语料库辅助英语教学入门》(何安平著 北京:外语教学与研究出版社 2010)的修订版,属实验实践类教材。本书于 2014 年 10 月入选为教育部“十二五”普通高等教育本科国家级规划教材。读者对象是大、中、小学的英语教师以及高等院校相关专业的本科生和研究生。

现代信息技术对教育发展具有革命性影响,具有大数据理念和技术特征的语料库是其中的重要组成部分。语料库语言学作为现代语言学、计算机科学与教育学跨学科互动的新兴学派,已经进入了当今科研的主流。它以其前所未有的巨量语言信息储备、高速精准的计算机检索方式和直观的批量化语言观察界面取胜,既为语言教育提供大量优质教学资源,又能创设大信息量、多维演示的立体教学和人机互动的优质学习平台。

目前我国高等教育和基础教育的计算机普及率增长很快,但在将语料库新资源、新技术和新理念应用于促进外语教育教学方面还相对滞后。作者自 20 世纪 90 年代起一直致力于语料库语言学在中国的传播和普及,尤其关注如何让语料库的理论及应用成果走进一线教学课堂。故本书自 2010 年问世以来已经成为越来越多高校语料库课程的教材和中小学教师国家级培训等各级各类英语教师培训项目的特色资源。本次修订有以下创新点:

(1) 在编写理念上,更加聚焦 21 世纪国际人才“能够使用语言知识和科技讯息互动沟通”的核心素养,将“问题导向→研究介入→理论提升”的科学研究精神贯穿全书。

(2) 在内容选材上,瞄准语料库语言学在国内外发展的最新趋势和

应用成果，同时结合语料库最基本的理念与核心技术，融入 44 个可操作性教学活动。尤其在第五章和第六章，新增了语料库信息技术融入多媒体教学课件制作和运用互联网免费语料资源制作教学微型语料库等案例展示与分析。

- (3) 在编写体例上，改为全书以系列教学活动展开。各章节遵循“问题导入→要点指引→学习目标→系列活动→本章小结”的思路，在每个教学活动中体现“明示探究目标→提供语料样品→展示操作步骤→指引观察分析→归纳要点和理论提升”的知识构建过程。尤其在每个活动的“述评”部分增补了对语料库语言学和外语教育教学相关核心理念的归纳性陈述。

- (4) 在教学资源整合上，实践“纸质、网络 and 软件实操等多维资源和手段互动”的立体化教材理念。其中教学活动涉及的语料库检索工具均有详细指引，可从网上免费获取；全书的语料样品（即 Minifiles）可在外语教学与研究出版社“高等英语教学网”（heap.unipus.cn）中的“教材支持中心”查找本书——《语料库辅助英语教学入门（修订版）》，即可免费下载。

本书由主编何安平负责上述的全部修订，两位副主编负责补充和订正。编者对初版之中杨素香（河南理工大学）和杨文坤（华南师范大学）撰写的 2010 版第二章和第六章初稿进行了全面改版及内容更新。我们衷心感谢初版和修订版中所有参与编写的各位老师，并感谢华南师范大学外国语言文化学院师生及业内同行多年来对该书的支持和反馈的意见。在今后的教学中我们将继续听取各方的批评和建议，不断更新和完善教材。

何安平

2016年5月5日于广州

致 谢

首先感谢 Laurence Anthony 同意在本书中使用他研发的语料库检索工具 AntConc (2007) 和 AntProfile (2009) 来演示本书的活动设计。同时感谢 Mark Davies、James Thomas、Matt Gee、Pascual Pérez-Paredes、David Willis 分别同意本书从以下网站下载语料库辅助语言教学的截图或使用其编写教材中的语料 (依次见带 * 号的资源)。其余的教学活动微本均为国内外英语语料库检索的语例, 其中有些语例的作者由于种种原因无法联系到, 除在书中标注出处之外, 也在此一并致谢!

语料库检索工具

*AntConc (Version 3.2.1w): Anthony, L. 2007. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>

*AntWordProfiler (Version 1.200w): Anthony, L. 2009. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>

国内外语料库及网站

BROWN: The Brown University Standard Corpus of Present-Day American English. Kucera H., Francis. 1960. Brown University, USA.

* COCA: Corpus of Contemporary American English. Davies M. 1990-2015. Brigham Young University, USA. <http://corpus.byu.edu/coca/>

English-Chinese Parallel Concordancer (香港教育学院在线中英平行语料库). Wang Lixun. 2008-. The Hong Kong Institute of Education. <http://ec-concord.ied.edu.hk/paraconc/>

- FLOB: The Freiburg-LOB Corpus of British English. Christian Mair C. 1999. Freiburg University, Germany. <http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html>
- LLC: The London-Lund Corpus of Spoken English. Quirk, R., Greenbaum, S., Svartvik, J. 1980. University College London, UK. & Lund University, Sweden.
- LINDSEI: Louvain International Database of Spoken English Interlanguage. Gilquin, G., D Coke, S., Granger, S. 2010. UCL Press Universitaires De Louvain. Belgium.
- LOB: The Lancaster-Oslo/Bergen Corpus. Johansson, S., Leech, G., Goodluck H. 1978. Lancaster University. UK/University of Oslo & Bergen University, Norway.
- *SkELL: Sketch Engine for Language Learning. Thomas J. 2012-. Masaryk University, Czech. <http://skell.sketchengine.co.uk/run.cgi/skell>
- *WebCorp: WebCorp Linguist's Search Engine. 1999-. Birmingham City University, UK. <http://wse1.webcorp.org.uk/or> <http://www.webcorp.org.uk/lse>
- *Sacodeyl: 欧洲七国多模态语料库学习平台. Pascual Pérez-Paredes. 2008-. University of Cambridge, UK. <http://www.um.es/sacodeyl/>

教材、书籍

- English Classics 3000 (英文名著 3000 部) [CD]. 2000. 北京: 北京大学出版社.
- Touchstone: Students' Book [S]. McCarthy, M. McCarten, J., Sandiford, H. 2006. Cambridge: Cambridge University Press.
- *Collins COBUILD English Course [S]. Willis, J., Willis, D. 1988. London: Collins.
- 普通高中课程标准实验教科书英语 (New Standard English) [S]. 陈琳等. 2006. 北京: 外语教学与研究出版社.

- 中国学习者英语语料库 (CLEC) [M]. 桂诗春、杨惠中. 2003. 上海:
上海外语教育出版社.
- 综合教程 [S]. 何兆熊. 2010. 上海: 上海外语教育出版社.
- 新编英语教程 (第五册) [S]. 李观仪. 2008. 上海: 上海外语教育出版社.
- 普通高中课程标准实验教科书英语 (NSEC) [S]. 刘道义等. 2007. 北京:
人民教育出版社.
- 普通高中课程标准实验教科书英语 [S]. 王蔷等. 2009. 北京: 北京师
范大学出版社.
- 21 世纪大学英语 [S]. 翟象俊、郑树棠、张增健. 1999. 上海: 复旦大
学出版社.

目 录

第一章 语料库语言学简介	1
第一节 学科定义.....	2
第二节 发展简史.....	7
第三节 语料库类型.....	10
第二章 语料库检索工具操作及教学应用	18
第一节 检索工具简介.....	19
第二节 提取分析检索项的语境共现.....	21
第三节 提取与分析各类词频表.....	45
第三章 语料库辅助英语教学的理论基础	55
第一节 语料库与语言学.....	55
第二节 语料库与认知心理学.....	59
第三节 语料库与现代语言教育.....	64
第四章 语料库辅助的教学设计	71
第一节 语料库教学加工的背景和理念.....	72
第二节 语料库辅助的语音教学.....	74
第三节 语料库辅助的语法教学.....	83
第四节 语料库辅助的词汇教学.....	93
第五节 语料库辅助的阅读教学.....	108

第五章 语料库辅助的教学实施	127
第一节 语料库的“教学微本”建设	128
第二节 语料的局部语境凸显	136
第三节 语料库与多媒体课件的协同	150
第六章 小型教学语料库的建设与应用	161
第一节 教学语料库的定义及内涵	162
第二节 网上语料库的教学应用	164
第三节 自建微型教学语料库	184
参考书目	205
附录：44个教学活动目录表（按章排序）	
【活动1.1】什么是语料库？	2
【活动1.2】了解语料库语言学发展史	7
【活动1.3】用语料库检索工具展示各类语料库文本	10
【活动2.1】熟悉AntConc主页界面	19
【活动2.2】什么是语境共现？	21
【活动2.3】语境共现行的提取准备	22
【活动2.4】设置检索项	27
【活动2.5】深度观察语境共现行	38
【活动2.6】提取单词词频表	45
【活动2.7】提取词串频次表	48
【活动2.8】提取关键词表	51
【活动3.1】了解并体验语言的频数理念	55
【活动3.2】了解并体验语言认知图式理念	59

【活动3.3】 了解并体验语言教育理念	64
【活动4.1】 什么是语料库的教学加工?	72
【活动4.2】 探究单词的发音和拼写规律.....	74
【活动4.3】 探究单词的构词法及重音规律.....	78
【活动4.4】 训练口语交际中的重音及语调.....	80
【活动4.5】 被动语态一定带有by短语吗?	83
【活动4.6】 情态动词must只表示“必须”吗?	88
【活动4.7】 辨析同形异义词LIKE	93
【活动4.8】 探讨文学作品中的HEART.....	98
【活动4.9】 探讨动词CAUSE的语用意义	102
【活动4.10】 巧用“语境共现行”做阅读猜词.....	108
【活动4.11】 巧用词频表做速读.....	111
【活动4.12】 语料库辅助深层阅读教学.....	115
【活动4.13】 巧用词串表探究文体特征.....	121
【活动5.1】 基于批量语料制作“教学微本”	128
【活动5.2】 制作语法填空微本.....	131
【活动5.3】 制作语境意义配对微本.....	134
【活动5.4】 简化语境共现行的阅读界面.....	136
【活动5.5】 屏蔽阅读界面的局部信息.....	139
【活动5.6】 凸显语料局部信息的多种手段.....	142
【活动5.7】 善用教学设问凸显语言特征.....	148
【活动5.8】 教学微本与幻灯片融合.....	150
【活动5.9】 教学微本与多媒体协同.....	154
【活动5.10】 国外多媒体语料库学习平台展示.....	157
【活动6.1】 阅读并归纳教学语料库的定义、内涵及功能.....	162

【活动6.2】 运用在线免费语料库辅助教学.....	164
【活动6.3】 免费下载在线语料制作微本.....	172
【活动6.4】 使用词汇检测工具分析语料难度.....	177
【活动6.5】 自建基于话题的阅读教学语料库.....	184
【活动6.6】 基于学习者口笔语料的微型语库建设与使用.....	191
【活动6.7】 语料的录入、标注及储存的案例展示.....	196

第一章 语料库语言学简介

【问题导入】

以下是一线英语教师和学生初学语料库时的反馈语截图。

- | | |
|----|----------------------------------|
| 1 | 本来对语料库的使用还比较陌生 |
| 2 | 对语料库的了解不透彻 |
| 3 | 语料库并不是自己想象的那样，很多骨干老师都知道 |
| 4 | 首次接触了语料库资源和技术这一概念 |
| 5 | 之前有接触过语料库这个概念 |
| 6 | 上这样的课一定要先对语料库有个简明的整体介绍 |
| 7 | 我第一次解了语料库 |
| 8 | 我们曾经很好奇的视频音频处理及语料库的应用 |
| 9 | 我是第一次这么直观地了解语料库的运用 |
| 10 | 以前只对语料库有最最基本的了解 |
| 11 | 一直觉得语料库好复杂 |
| 12 | 语料库并不是第一次接触 |
| 13 | 语料库是什么 |
| 14 | 语料库对于我是一个新事物 |
| 15 | 这是我第一次接触到语料库 |
| 16 | 以前一直认为语料库就是写论文的时候才用得上的“高大上”的一个工具 |
| 17 | 对语料库产生了兴趣 |
| 18 | 对语料库感到很高兴 |

【要点指引】

以上是一批中学及高校教师在初涉语料库时的想法和困惑。多数人认为语料库是一个新生事物，对其既有好奇心、感兴趣，又觉得太“高大上”，难以把握。可见在当今信息社会，计算机科学和教育技术虽然为外语教师带来前所未有的教学理念与手段更新，但是作为最早体现大数据时代理念的语料库，其在我国英语教育教学领域的应用却相对滞后。所以本章将通过一系列学习和体验活动，达成以下学习目标。

【学习目标】

1. 初步了解语料库语言学的定义及学科发展沿革；
2. 了解和观察各种类型语料库样品及其体例特征。

第一节 学科定义

【活动 1.1】什么是语料库?

步骤一: 阅读以下多位语料库语言学家对语料库和语料库语言学定义的阐述。注意定义中反复出现的词和短语, 看能否捕捉到语料库以及语料库语言学的定义, 包括语料的特点、规模、来源、储存和提取手段以及用途等等。

1. 关于语料库

- [A corpus is] a collection of naturally-occurring language text, chosen to characterize a state or variety of a language.

The whole point of assembling a corpus is to gather data in quantity.

(Sinclair, 1991: 171, 1995: 21)

- A corpus is a body of texts assembled according to explicit design criteria for a specific purpose.

(Atkins & Clear, 1992: 5)

- [A corpus is] a body of naturally-occurring (authentic) language data which can be used as a basis for linguistic research.

In the past thirty-five years, the term corpus has been increasingly applied to a body of language material which exists in electronic form, and which may be processed by computers for various purposes.

(Leech, 1997: 1)

- [A corpus is] a collection of texts, especially if complete and self-contained: the corpus of Anglo-Saxon verse.
- [A corpus is] a body of texts, utterances, or other specimens considered more or less representative of a language, and usually stored as an electronic database.

Currently, computer corpora may store many millions of running words, whose features can be analysed by means of tagging (the addition of identifying and classifying tags to words and other

formations) and the use of concordancing programs. Corpus linguistics studies data in any such corpus.

Bodies of natural language material (whole texts, samples from texts, or sometimes just unconnected sentences), which are stored in machine-readable form. Computer corpora are rarely haphazard collections of textual material: they are generally assembled with particular purposes in mind, and are often assumed to be (informally speaking) representative of some language or text type.

(McArthur & McArthur, 1992)

2. 关于语料库语言学

- Corpus linguistics is not an end in itself but is one source of evidence for improving descriptions of the structure and use of languages, and for various applications, including the processing of natural language by machine and understanding how to learn or teach a language.

(Kennedy, 2000: 1)

- What, then, is a “corpus linguist”? I would like to think that it is a linguist who tries to understand language, and behind language the mind, by carefully observing extensive natural samples of it and then, with insight and imagination, constructing plausible understandings that encompass and explain those observations. Anyone who is not a corpus linguist in this sense is, in my opinion, missing much that is relevant to the linguistic enterprise.

(Chafe, 1992: 96)

步骤二：将上述两部分文献合成可由语料库检索工具查阅的微本语料（见本书语料样品 Minifiles\corpus + definition.txt），可获得图 1-1 检索界面截图。观察图中各行 corpus 或 corpus linguistics 后面的文字内容，看是否能归纳出语料库的基本定义。反思这种阅读方式与步骤一有何不同、二者分别有何优势。

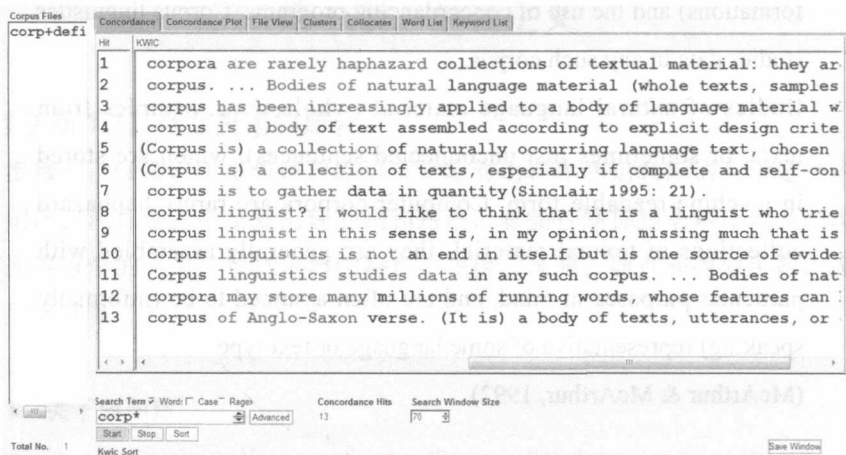


图 1-1 源自文献 1 和 2 的微本语料检索截图

步骤三：使用 Adobe Acrobat Reader 9 的高级检索工具（Search，如图 1-2 所示）可以聚焦性阅读有关语料库书籍的 PDF 版文档，如本例中 *From Corpus to Classroom* (O’Keeffe 等人，2007) 一书部分章节中所有带 corpus is 的语境共现行（见图 1-3），也可以点击其中任意一行直接进入该行出现在书中的上下文，进行更大语境范围内的阅读（见图 1-4），看能否找到关于语料库定义的表达。

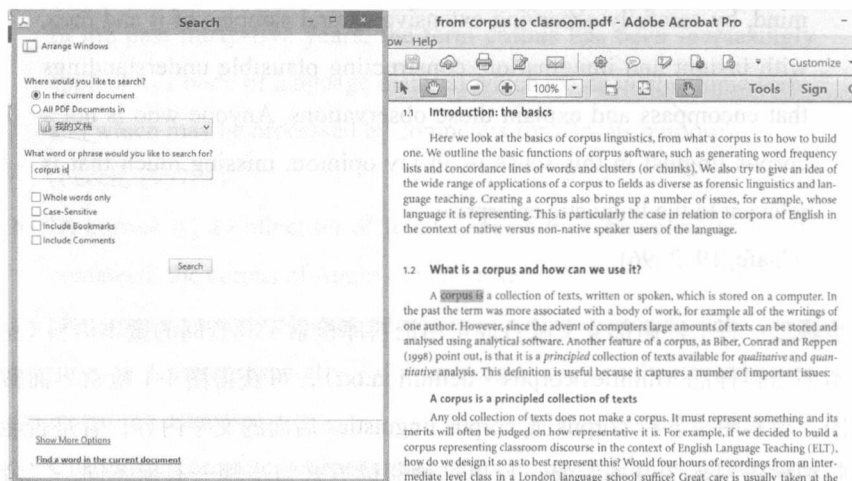


图 1-2 Adobe Acrobat 的高级搜索框

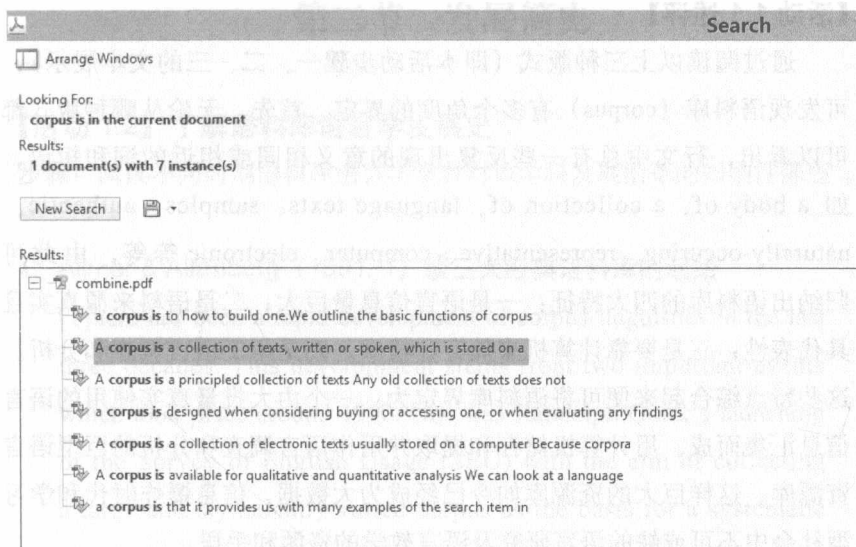


图 1-3 PDF 展示的批量 corpus is 语境共现行

1.1 Introduction: the basics

Here we look at the basics of corpus linguistics, from what a corpus is to how to build one. We outline the basic functions of corpus software, such as generating word frequency lists and concordance lines of words and clusters (or chunks). We also try to give an idea of the wide range of applications of a corpus to fields as diverse as forensic linguistics and language teaching. Creating a corpus also brings up a number of issues, for example, whose language it is representing. This is particularly the case in relation to corpora of English in the context of native versus non-native speaker users of the language.

1.2 What is a corpus and how can we use it?

A **corpus is** a collection of texts, written or spoken, which is stored on a computer. In the past the term was more associated with a body of work, for example all of the writings of one author. However, since the advent of computers large amounts of texts can be stored and analysed using analytical software. Another feature of a corpus, as Biber, Conrad and Reppen (1998) point out, is that it is a *principled* collection of texts available for *qualitative* and *quantitative* analysis. This definition is useful because it captures a number of important issues:

A corpus is a principled collection of texts

Any old collection of texts does not make a corpus. It must represent something and its merits will often be judged on how representative it is. For example, if we decided to build a corpus representing classroom discourse in the context of English Language Teaching (ELT), how do we design it so as to best represent this? Would four hours of recordings from an intermediate level class in a London language school suffice? Great care is usually taken at the

图 1-4 点击图 1-3 第 2 行打开该书章节的相关语段