



# 教材语料库的建构与应用

## ——以新加坡小学华文教材为例

罗庆铭 著

The Construction and Application of Textbook Corpus  
—Using the Singapore Primary School Chinese Textbook as Example

中国社会科学出版社



# 教材语料库的建构与应用

——以新加坡小学华文教材为例

罗庆铭 著



The Construction and Application of Textbook Corpus  
—Using the Singapore Primary School Chinese Textbook as Example

中国社会科学出版社

## 图书在版编目(CIP)数据

教材语料库的建构与应用:以新加坡小学华文教材为例/罗庆铭著. —北京:  
中国社会科学出版社, 2017. 5

ISBN 978 - 7 - 5203 - 0101 - 5

I. ①教… II. ①罗… III. ①小学—华文教育—教材—研究—新加坡  
IV. ①G749. 339

中国版本图书馆 CIP 数据核字(2017)第 067765 号

---

出版人 赵剑英  
责任编辑 周晓慧  
责任校对 无介  
责任印制 戴宽

---

出版 中国社会科学出版社  
社址 北京鼓楼西大街甲 158 号  
邮编 100720  
网址 <http://www.csspw.cn>  
发行部 010 - 84083685  
门市部 010 - 84029450  
经销 新华书店及其他书店

---

印刷 北京明恒达印务有限公司  
装订 廊坊市广阳区广增装订厂  
版次 2017 年 5 月第 1 版  
印次 2017 年 5 月第 1 次印刷

---

开本 710 × 1000 1/16  
印张 16  
插页 2  
字数 252 千字  
定价 69.00 元

---

凡购买中国社会科学出版社图书,如有质量问题请与本社营销中心联系调换  
电话:010 - 84083683

版权所有 侵权必究

## Abstract

This study constitutes an exploration of the construction and application of the Singapore primary school Chinese textbook corpus. It focuses mainly on three major objectives: a) the construction of a pedagogically tagged Singapore primary school Chinese textbook corpus; b) the utilization of the constructed primary school Chinese textbook corpus to conduct relevant research on language elements, explore the content and methodology of textbook corpus-aided primary school Chinese teaching; c) further undertake research on second language vocabulary acquisition in conjunction with teaching.

Constituting an important teaching resource and aid, textbook corpora play a significant role in aiding language teaching experts in the compilation of teaching syllabus, extraction of vocabulary table for teaching, and the provision of teaching reference materials. Nevertheless, textbook corpora are used mostly for indirect applications in the field of second language teaching. The application of textbook corpora to directly serve classroom teaching has yet to receive sufficient attention, especially in the field of Chinese language teaching.

Based on usage objectives, data sources and annotation methods, textbook corpora can be further divided into pedagogical corpora, textbook corpora and textbook-information databases, all of which vary widely in terms of usage. The textbook corpus constructed under this study can be directly applied to classroom teaching and textbook language research. While the development and application of this type of corpora remain in the infancy stage for non-Chinese languages, this study constitutes a first of its kind in the realm of Chi-

nese language.

Textbook corpora that target classroom teaching and textbook language research are different from other corpora in terms of corpus selection and annotation, functional design and interface presentation. Complying with the basic principles and procedures of corpus development, the Singapore primary school Chinese textbook corpus further exhibited innovation and breakthroughs in the context of functional development and search software design. It achieved the extraction and computation of language information such as phonology, Chinese characters, vocabulary, collocation, etc and was able to address the demand for diverse, personalized and convenient teaching in relation to search method and interface design, thereby realizing the designated research objectives of this study.

This study makes reference to corpus-based and corpus-driven research models and utilizes the constructed textbook corpus to undertake a non-systemic statistical analysis of the language elements of Singapore primary school Chinese textbooks. Firstly, Claude Shannon's Information Theory (1951) is applied to undertake a more comprehensive analysis of the phonological load of Singapore primary school Chinese textbooks. Results indicate that the total load is unevenly distributed within the internal Chinese phonetic system-vowels have the highest phonological load, followed by consonants, with tones having the lowest phonological load. In terms of the average load of phonemes, tones command the highest phonological load, with a greater divergence in the phonological load of the internal subsystems of consonants, vowels and tones. The divergence in the usage frequency of phonetic elements and the internal configuration of phonetic systems is the main cause for the differences in Chinese phonological load. These preliminary conclusions are of reference value to the study and pedagogical research of Chinese phonetic acquisition. Secondly, this study uses statistical methods to undertake a comprehensive analysis of the vocabulary system of Singapore primary school Chinese textbooks, revealing a skewed distribution of Chinese word class in textbooks, including a significant difference in the recurrence rate of different

word class as well as the internal recurrence rate within the same word class, and an abnormal distribution phenomenon of few high frequency words and numerous single frequency words. Textbook concurrence is not common, accounting for approximately 6% - 10%. Concurrence occurs more between content words, with a close relationship between the number of syllables of multi-category words and word class.

In exploring the application of the Singapore primary school Chinese textbook corpus to the teaching of Chinese, reference is made in this study to Tim Johns' data-driven learning theory (1991). From the perspectives of the applicability of textbook corpus to different teaching stages and segments, and the combination of textbook corpus with internet multimedia technology, this study further sets out an in-depth discussion of the feasibility of the application of textbook corpus to Singapore primary school Chinese teaching and puts forward concrete methods and examples to further prove that based on pedagogical principles, textbook corpus can achieve an effect unmatched by printed textbooks in aiding classroom teaching.

This study uses the constructed textbook corpus to conduct research on vocabulary acquisition in conjunction with teaching. Through a small-scale comparative experiment on pedagogy, this study further illustrates that teaching pedagogy which utilizes textbook corpus-based vocabulary repetition has a positive effect in improving the effectiveness of vocabulary learning of students.

This study concludes with a summary of the research topic, including the research findings, and further points out existing inadequacies as well as the direction for further research.

**Keywords:** Singapore; primary school Chinese; textbook corpus; construction; application

# 目 录

第一章 绪论 .....	(1)
第一节 研究缘起 .....	(2)
一 语料库在语言教学中的应用 .....	(2)
二 语料库辅助语言教学的不足 .....	(5)
第二节 课题的提出 .....	(7)
第三节 预期研究成果、研究价值及可能面对的困难 .....	(8)
一 预期研究成果 .....	(8)
二 研究价值 .....	(11)
三 本研究可能面对的困难 .....	(13)
第四节 研究性质、研究思路、研究方法及总体框架 .....	(14)
一 研究性质、研究思路和研究方法 .....	(14)
二 总体框架 .....	(15)
第二章 教材语料库及相关研究述评 .....	(18)
第一节 与教材语料库相关的研究 .....	(18)
一 语料库及相关概念 .....	(18)
二 语料库的类型 .....	(19)
三 汉语通用语料库 .....	(23)
四 学习者语料库 .....	(29)
第二节 教材语料库的建构与应用 .....	(38)
一 教材语料库的名称与分类 .....	(38)
二 非汉语教材语料库的建构与应用 .....	(42)
三 汉语教材语料库的建构与应用 .....	(49)

## 教材语料库的建构与应用

第三节	对教材语料库应用的评价 .....	(55)
一	教材语料库的特点 .....	(55)
二	教材语料库对教学的贡献 .....	(59)
三	汉语教材语料库建构与研究的不足 .....	(61)
第四节	新加坡小学华文教材语料库建构与应用的 理论基础 .....	(62)
一	语料驱动的学习理论 .....	(62)
二	语料库语言学研究模式 .....	(65)
<b>第三章</b>	<b>新加坡小学华文教材语料库(PCTC)的建构</b> .....	(69)
第一节	PCTC教材语料库的总体设计 .....	(69)
一	PCTC教材语料库的设计理念 .....	(69)
二	PCTC教材语料库的基本架构 .....	(70)
三	PCTC教材语料库建构的基本流程 .....	(71)
四	PCTC教材语料库的基本统计数据 .....	(82)
第二节	PCTC教材语料库操作平台与数据库建构 .....	(83)
一	操作平台与运行环境 .....	(83)
二	语料的加工处理 .....	(83)
第三节	PCTC教材语料库的系统功能 .....	(87)
一	导入功能 .....	(87)
二	导出功能 .....	(87)
三	检索功能 .....	(88)
四	统计功能 .....	(89)
五	对比功能 .....	(90)
六	辅助工具 .....	(91)
七	数据库结构 .....	(93)
八	安装程序 .....	(97)
九	运行 .....	(103)
第四节	PCTC教材语料库功能演示 .....	(104)
一	语料导入 .....	(104)
二	语料导出 .....	(105)



三	语料检索 .....	(108)
四	语料统计 .....	(116)
五	语料对比 .....	(118)
六	辅助工具 .....	(119)
第五节	小结 .....	(122)
<b>第四章</b>	<b>基于 PCTC 教材语料库的语言要素研究</b> .....	(124)
第一节	基于 PCTC 教材语料库的语音负担研究 .....	(124)
一	引言 .....	(124)
二	语料来源和研究方法 .....	(126)
三	研究结果与讨论 .....	(128)
四	对汉语教学的启示 .....	(139)
五	小结 .....	(141)
第二节	基于 PCTC 教材语料库的词类分布考察 .....	(142)
一	引言 .....	(142)
二	研究对象、研究工具和研究方法 .....	(144)
三	研究结果分析 .....	(146)
四	教学建议 .....	(158)
<b>第五章</b>	<b>PCTC 教材语料库在华文教学中的应用</b> .....	(160)
第一节	PCTC 教材语料库在不同教学阶段的应用	
——	以语音教学为例 .....	(160)
一	引言 .....	(160)
二	新加坡小学语音教学的现状及存在的问题 .....	(161)
三	PCTC 教材语料库辅助语音教学的基本 原理和方法 .....	(164)
四	PCTC 教材语料库在不同教学阶段对 语音教学的辅助 .....	(166)
五	PCTC 教材语料库辅助语音教学的优势 .....	(174)
六	小结 .....	(176)

## 教材语料库的建构与应用

第二节	PCTC 教材语料库在不同教学环节的应用	
	——以词汇教学为例 .....	(176)
一	引言 .....	(176)
二	PCTC 教材语料库在词汇教学环节中的应用 .....	(177)
三	PCTC 教材语料库辅助词汇教学的优势 .....	(180)
四	小结 .....	(188)
第三节	教材语料库与多媒体技术的结合	
	——以“我的华文动漫书”为例 .....	(188)
一	引言 .....	(188)
二	PCTC 教材语料库与电子动漫教材结合的理论依据和现实意义 .....	(189)
三	PCTC 教材语料库与电子动漫教材“我的华文动漫书”的结合 .....	(191)
四	“我的华文动漫书”PCTC 教材语料库的主要功能 .....	(193)
五	小结 .....	(199)

## 第六章 基于 PCTC 教材语料库的词汇重现对

	二语习得的影响 .....	(202)
第一节	语言重现率与语言习得的关系 .....	(202)
一	引言 .....	(202)
二	教材中词汇重现率对词汇习得的影响 .....	(203)
三	教学中的重现策略对词汇习得的影响 .....	(205)
第二节	PCTC 教材语料库影响语言习得的实验研究 .....	(206)
一	研究问题 .....	(206)
二	研究对象与方法 .....	(207)
三	实验结果 .....	(208)
四	讨论 .....	(210)
五	小结 .....	(213)

## 第七章 总结 .....

第一节	主要研究成果 .....	(215)
一	修订了教材语料库的概念 .....	(215)

二	建构了新加坡小学华文教材语料库 .....	(215)
三	语料驱动的学习模式与教材语料库的成功结合 .....	(216)
四	发现了教材中一些重要的语言规律 .....	(217)
第二节	研究局限和不足 .....	(218)
第三节	对未来研究的建议 .....	(218)
参考文献	.....	(220)

# 第一章 绪论

本研究探讨新加坡小学华文教材语料库（以下简称“PCTC 教材语料库”）的设计开发以及如何应用在教材分析和课堂教学实践的相关课题。书中新加坡小学华文教材特指 2007 年由新加坡教育部课程规划与发展司与中国人民教育出版社合作编写的小学华文系列教材，包括《小学华文》《小学高级华文》和《小学基础华文》三套。从容量上看，PCTC 教材语料库属于小型单套教材语料库，或称为“教学语料库”。同广泛收集各类不同水平、不同技能教材为建库对象的教材语料库相比，除了容量大小有别外，PCTC 教材语料库在建库的目的、语料库功能方面也存在差异。

本研究主要目标有三个：一是建构能够满足实际教学需要，具有教学辅助功能、方便检索的新加坡小学华文教材语料库；二是利用该语料库获取语音、词汇等方面的数量及分布信息，开展相关语言要素的研究，并以此为依据，对新加坡小学华文教材各级语言要素配置的合理性作出客观准确的评价；三是探讨如何利用 PCTC 教材语料库辅助华文课堂教学，以及结合语料库开展与华语习得相关的研究。

本研究穿插使用“华语”和“汉语”两个概念。前者是狭义的，专用于新加坡；后者是广义的，泛指汉语普通话。本研究以新加坡小学华文教材为研究对象，因此所建构的教材语料库称为“新加坡小学华文教材语料库”。当探讨新加坡小学华文教材语料库的建构以及在华文教学中的应用问题时，采用“华语”进行称说，而基于新加坡小学华文教材语料库所进行的语言要素和语言习得的研究，具有广泛的适用性，因此采用“汉语”进行称说。

## 第一节 研究缘起

### 一 语料库在语言教学中的应用

语料库在语言教学中的应用主要体现在两个方面：一是语料库作为教学辅助工具的使用，其中包括开发具有语料收集功能的应用软件、为语料库检索设计功能更强大的软件包以及研制各种旨在服务教学的赋码标注软件等；二是语料库方法的应用，主要指的是利用语料库技术分析语料库中收集的语言数据，并将研究成果应用到语言教学的各个领域，如基于语料库的语言要素研究、教学法应用、教材开发和课程设计等。前者涉及语料库的研发和建构，后者则侧重语料库的应用。语料库在语言教学中的应用可分为间接应用（indirect applications）和直接应用（direct applications）两个方面（Römer, 2009）（如表 1.1 所示）。

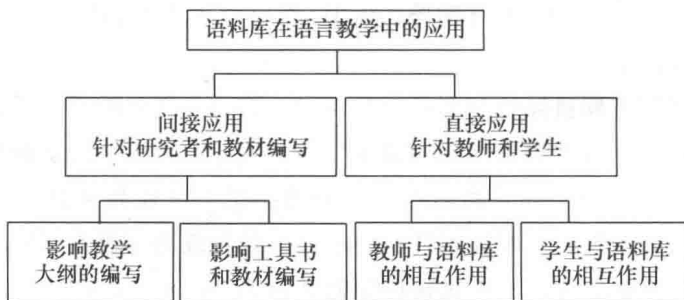


表 1.1 语料库教学应用模式

#### （一）语料库的间接应用

##### 1. 编制词汇大纲

语料库间接应用于语言教学的主要领域首先是帮助教材编写者制定词汇大纲。利用语料库强大的数据贮存和检索功能，语言教学专家可找到最常用的语言表达方式、最常使用的词汇、语法等语言要素，以此作为教材编写的依据。这方面的先行者当属 Dave Willis 和 Jane Willis

(1989)。他们编写的柯林斯 COBUILD<sup>①</sup> 英语课程 (CCEC) 是依据 2000 万词次的柯林斯伯明翰大学国际语料库 (COBUILD) 提取出的约 2500 个高频基本词编写而成的。该教材选取的词汇代表了英语在当代真实语境中最常用、最基本的词汇, 成为后继教材编写者遴选教材词汇最重要的参照对象。许多学者在评估英语教材词汇选择是否合理时也常以 CCEC 的词汇表作为目标参照词表 (Coniam, 2004; 梁健丽、何安平, 2009)。在汉语学术界, 较早运用语料库方法研究汉语语言规律的成果是 1985 年由北京语言学院语言教学研究所研制的《现代汉语频率词典》。它通过精选报刊政论、科普、剧本和文学作品、日常口语等领域的口语与书面语语料 180 万字, 运用频率、使用度等方法, 从词汇的构词、分布、频次、位次、音序、词性等方面筛选出领域词、通用词、常用词等, 统计出常用词 8548 个。这是第一份建立在坚实的语料库基础上, 最贴近对外汉语教学实际的词表。该词表的诞生对 20 世纪 90 年代汉语水平考试词汇与汉字等级大纲的研制有着重要的影响, 也成为编写对外汉语教材的基本参照词表之一。

2009 年, Routledge 出版了一部面向海外学习者使用的 A Frequency Dictionary of Mandarin Chinese (中文译名:《汉语频率词典》)。该词典由 Richard Xiao, Paul Rayson 和 Tony McEnery 合作编写, 它收集了当代汉语中最常用的 5000 个词和 2000 个汉字。这份常用字词表出自于容量为 5000 万字, 包括了当代口语和小说、新闻等书面语的大型语料库。词条除了列出拼音、词性、英文翻译、例句及例句翻译等内容外, 还特别在每个词条旁边标注了字词的使用频率。除此之外, 作者还按照 30 个主题对常用词进行了归类列表。该频率词典对海外汉语教学, 特别是汉语教材的编写提供了重要的参考资料。

近年来, 语料库方法的应用延伸到了专业汉语教学领域。季瑾 (2007) 通过收集中国 20 世纪 80 年代中期至今的所有商务类汉语教材, 利用语料库理论和信息处理技术建立了“商务汉语教材语料库”, 从中提取了一份商务汉语等级词语表, 并在此基础上尝试突破以往汉语作为

<sup>①</sup> COBUILD 是 COLLINS Birmingham University International Language Database 即“柯林斯伯明翰大学国际语料库”的缩写。1980 年建立, 由 Collins 出版社赞助。

第二语言学习词典的编纂方法，提出了编写基于语料库的商务汉语学习词典的理论依据和具体操作的设想。上述个案从不同角度展示了语料库语言学和语言教学的紧密关系。有关语料库应用在教学领域的详细内容将在第二章的文献综述中进行评述。

### 2. 编写词典或工具书

语料库可以为语言类工具书和教材编写提供丰富的语料素材。这些素材经过加工、处理后可以成为词典选择词条和例句的重要依据，也可以为教师编写辅助教材提供内容丰富且难易度受控的教学资源。许多出版商已经意识到语料库对语言教学的重要性，因此在推销词典时常以所编词典乃“基于语料库”作为卖点之一。例如，哈伯—柯林斯出版集团（Harper Collins）出版的《柯林斯 COBUILD 高阶英语学习者词典》（The Collins COBUILD Advanced Learner's English Dictionary）被认为是利用语料库编写词典最成功的案例之一。它的最大特点在于所有例句均来自于 COBUILD 语料库中的真实语料，与当代英语社会的实际生活贴近。该词典除了提供传统词典原有的拼写、发音、语法、意义、语体、用法、例证、短语、成语、同义词、反义词等信息外，还专门标注了词语的词频信息<sup>①</sup>以及在具有特殊语用义的词语旁用附加栏标出详尽的语用信息。

柯林斯词典的编写方法对后来的语言词典的编撰影响很大，朗文词典和牛津词典也都利用语料库编写各自的系列英语词典。

语料库间接运用于教学的工作主要是由语言教育专家完成的，教师或学生通过使用语料库的研究成果，如词典、参考书、教材等促进语言学习，他们并没有从语料库中直接受惠，因此，语料库的作用是间接的。与语料库间接运用不同，语料库还可以直接为语言教学服务。

#### （二）语料库的直接运用

语料库直接服务于语言教学最早出现在英语教学领域。Johns（1991）率先将语料库引入课堂教学。他利用语料库中的关键词检索功

---

<sup>①</sup> 《柯林斯 COBUILD 高阶英语学习者词典》的频率信息采用菱形符号标注，三个蓝色的菱形表示该词为常用词，两个菱形为次常用词，一个菱形为低频词，无菱形符号则表示该词为罕用词。

能收集特定词语的上下文搭配信息,将其中有代表性的例句提供给学生,学生通过合作学习寻求词语搭配规律,从而掌握词语的正确用法。这种学习方法被称为“语料驱动学习”(data-driven learning)模式。与教师和学生被动使用教材不同,语料驱动学习法直接利用语料库的真实语料进行教学,强调教师、学生与语料库的直接互动,通过教师从语料库中选取教学语料,引导学生分析语料,归纳语言规律,完成语言知识和技能的内化(第二章第四节将详细介绍)。

除 Johns 外,其他学者和教师也对利用语料库开展语言教学进行了大量的探索与实践。例如 Bernardini (2002) 利用英国国家语料库(BNC)对学生进行英语教学,发现语料库在辅助学生探索学习方面具有十分积极的作用; Bondi (2001) 发现采用特殊用途的专门语料库(language for special purpose corpus)进行语料驱动教学活动对有效提升学生的语言意识大有裨益。这些教学实践活动证明了语料库在辅助语言教学方面具有不可低估的潜能,它不仅改变了传统语言教学的手段,丰富了教学资源,也极大地改善了语言学习的环境(Aston, 1997)。

总之,语料库对语言教学所起的积极作用已经得到越来越多学者的认同。基于语料库的教学已成为语言教学的一种自然的、现代化的、必不可少的辅助方法(卢伟, 1999)。

## 二 语料库辅助语言教学的不足

### (一) 语料库与语言教学之间的鸿沟依然存在

尽管语料库已经日益显示出强大的功能和应用前景,但令人遗憾的是它在课堂语言教学中的运用范围还十分有限,无论是广度还是深度都不尽如人意。正如 Mukherjee (2004) 所指出的,语料库和语言教学之间的鸿沟依然存在,对语料库的种种非议依然不时出现,语料库在课堂教学中的实际应用还十分有限。语料库理论与语言教学之间的接口还不是很多。在中国,利用语料库进行教学的研究起步较晚。在英语教学界,已有学者利用通用语料库进行“语料驱动学习”的研究,取得了一些成果。例如孙静(2007)利用布朗语料库(BROWN)对68名高中二年级学生进行英语词汇教学效果的实验研究,研究结果显示,语料驱动学习方法对于高中学生学习词汇帮助明显,尤其对词汇深度知识的



学习和对词汇的长久记忆有促进作用。而在汉语作为第二语言教学领域，语料库直接应用于汉语教学的研究成果还十分有限。

语料库在语言教学中的潜能不能得到有效发挥的原因，除了教师和学生对话料库技术和检索工具缺乏了解，对话料库驱动的学习方法存在畏惧心理等主观原因外，还与大型通用语料库自身的缺陷有关。首先，大型语料库由于规模巨大，难免造成大量冗余信息，形成数据烟雾（data smog）。学生或教师在筛选例句时，必须披沙拣金，过滤掉重复或多余的语料，这无疑增加了学习负担。其次，大型语料库的设计开发通常不是或者不全是专门为语言教学服务的，语料库中语料的收集范围、难易度都不受控制。教师若要从中选取适合学生语言程度的教学材料，通常需要对语料库检索出来的语料进行二次人工筛选，剔除内容不合适或难度太大的语料，才能最终选定适合的教学材料。这一繁琐的筛选过程无疑使语料库快捷、准确满足教学需要的有效性大打折扣。因此，徐曼菲（2004）提出应建立小型、有针对性的专门语料库，以作为课堂教学使用教学材料的来源。

特别用途的小型语料库不具备大型通用语料库容量大、语料范围广、代表性强的特点，但由于专用语料库语料来源单一，建库目的性强，常常更符合教学需要，可以更有效地协助教师和学生完成特定的教学任务。例如，Bowker & Pearson（2002：46—47）利用包含10000词的小型专门语料库帮助学习者掌握机械专业的相关词汇和搭配，虽然用关键词检索软件从该语料库中检索到的例句比英国国家语料库中检索到的少，但是例句的专业相关度却很高，对机械专业学习者以及翻译人员掌握相关特殊文本的术语和搭配起到了很好的作用。

### （二）教材语料库的建构与应用研究依然薄弱

教材语料库是以教材为语料收集对象而建构的数据库。关于教材语料库的名称，至今尚存在分歧。有关教材语料库的操作定义，我们将在第二章的文献回顾中进行详细探讨。在此我们暂且概括为包含了各类教材信息的语料（含口语、书面语）集合体。

教材语料库在语言教学中的作用长期被忽视。Meunier & Gouverneur（2009：179—201）曾指出，当今出版英文教材的出版商在选材时大都把目光聚焦在母语或学习者语料库上，并据此编写词典和各种词