

基于国际中文教学的语言点语料库构建方案初探

彭瑶, 刘人铭, 谭钰浠, 王飞华*

四川师范大学国际中文教育学院, 四川 成都

收稿日期: 2024年3月31日; 录用日期: 2024年5月17日; 发布日期: 2024年5月31日

摘要

随着全球汉语学习者的不断增加, 汉语教学的质量亟待提高, 适于汉语教学的语言点语料库的建设将有益于教学, 但目前这种语料库较少。本文拟对建设一个以汉语语法为切入点的用于语言对比和国际中文教学的语言点语料库的可行性进行研究。论文首先研究了汉语语料库建设的现状, 发现研究汉语的语料库较少, 并且缺少专用的语言点语料库。基于这一背景, 本文提出了构建基于国际中文教学的语言点语料库的设想, 并分析了构建语料库的可行性与意义。该语料库具有语言点检索、语言点注解、教学解释和练习四个板块, 共同组成了一个完备的系统, 为学生学习和教师备课带来了便捷。

关键词

国际中文教学, 语料库, 汉语语法

Research on the Construction of Language Point Corpus Based on International Chinese Teaching

Yao Peng, Renming Liu, Yuxi Tan, Feihua Wang*

School of International Chinese Education, Sichuan Normal University, Chengdu Sichuan

Received: Mar. 31st, 2024; accepted: May 17th, 2024; published: May 31st, 2024

Abstract

As the number of Chinese learners around the world increases continually, the quality of Chinese
*通讯作者。

文章引用: 彭瑶, 刘人铭, 谭钰浠, 王飞华. 基于国际中文教学的语言点语料库构建方案初探[J]. 现代语言学, 2024, 12(5): 381-388. DOI: 10.12677/ml.2024.125372

teaching needs to be improved. The construction of a language point corpus suitable for Chinese teaching will be beneficial to teaching. Currently, there are few such corpora. This paper intends to explore the possibility of the construction of a language point corpus for language comparison and international Chinese education, taking Chinese grammar as the starting point. The paper first studies the current situation of Chinese corpus construction and finds that there are few literatures on the construction of Chinese corpora and language point corpora are rare, too. Based on this background, this article puts forward the idea of constructing a language point corpus based on international Chinese teaching, and analyzes the feasibility and significance of constructing the corpus. The corpus has four sections: language point retrieval, language point annotation, teaching explanation and practice, which together form a complete system that brings convenience to students' learning and teachers' lesson preparation.

Keywords

International Chinese Teaching, Corpus, Chinese Grammar

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在经济全球化背景下，汉语学习者日益增加，各国汉语教学正在由普及型、兴趣型向正规化、专业化方向发展；组织机构、课程内容越来越多样化，适应教学需求的教学资料开发也在迅速增长。但目前适于汉语教学的语料库存在数量和质量上的不足，远没有达到汉语学习者和教师期望的高度。本文试图在分析现有汉语语言点语料库的不足的基础上，对其开发建设的可能性进行一定的探索。

1.1. 构建基于国际中文教学的语言点语料库的意义

1.1.1. 填补市场空缺

现有的面向对外汉语教学开设的语料库数量极少，尤其是以收集语法为主的语料库。当前对外汉语学习的语料库有语音教学语料库、动态作文语料库等，而语法汇总的语料库较少。因此，构建一个基于国际中文教学的语言点语料库能够有效填补市场空缺，满足学习者的学习需求。

1.1.2. 满足教学需求

母语为非汉语的学习者的汉语基础较差，缺少目的语交际环境，他们缺少一个自学、练习及复习的汉语语言点语料库。在汉语教学中，面对规则复杂、条目较多的语言点，国际中文教师需要尽量避免使用罗列理论的教学方式。虽然大量举例和示范能够降低学习难度，但这也成为教师备课的一大难点。构建语言点语料库能够满足各国汉语学习者自学语言点或提高汉语水平的需求，同时，为教师提供大量真实客观的语料，有效地缓解教师备课难的问题。

1.1.3. 促进文化传播

网络通讯技术、线上语言教育平台蓬勃发展，为汉语学习者提供了更加便利的平台。而面向各个阶段的汉语学习者提供的系统性的汉语语言点语料库正是顺应时代发展的产物。在帮助外国学习者自主学习汉语语法的同时，增强学习者对中华文化的兴趣，以此达到信息传递和文化交际的目的。

1.2. 构建基于国际中文教学的语言点语料库的可行性

1.2.1. 技术要求难度并不高

构建语言点语料库并非高技术门槛的任务，其所需技术要求基本上限于对常见办公软件的熟悉和掌握，尤其是像 Excel 和 Access 这样的数据库管理软件。这些软件以其直观友好的界面和简单易懂的操作步骤，使得即使没有专业编程或数据处理经验的人员也能够轻松上手。

这种低技术门槛的特性大大降低了构建语言点语料库的难度，使得更多的教育工作者和语言研究者能够参与到语料库的建设中来。与此同时，这也意味着构建语言点语料库不需要大量的人力、物力和财力投入。因此，技术要求的低难度保证了语言点语料库的构建可行性，使得更多人能够参与其中，共同促进国际中文教学和研究的的发展。

1.2.2. 可供参考的语料库较多

在国际中文教学领域，已经出现了多个可供参考的语料库，其中包括北京语言大学 HSK 动态作文语料库、中央民族大学国际汉语教学数据库以及 BCC 语料库。这些语料库汇集了大量真实的语言数据，覆盖了各种语言点和语境，为教学提供了丰富的素材和实践经验。例如，北京语言大学 HSK 动态作文语料库提供了丰富的写作范例和语言素材，有利于学习者模仿借鉴，提高写作能力。同时，语料库的分类和标注为教师提供了针对不同水平学生的教学参考。

以上三个语料库丰富的素材和经验对于构建一个国际中文教学语言点语料库具有重要的意义和价值。

1.2.3. 教学参考资料较多

国际中文教材的不断发展为本土语的构建提供了丰富的参考资料。一方面，国际中文教材的出版数量不断增长。自中华人民共和国成立至 2020 年底，全球共出版国际中文教材 19,530 种。其中，1981~1990 年共 1211 种，1991~2000 年共 1673 种。2000 年后出版数量迅速增长，2001~2010 年共 7278 种，2011~2020 年共 8039 种[1]。另一方面，国际中文教材的分类不断细化。根据不同的适用人群、专用目的以及课型设置等因素，国际中文教材呈现出通用教材、专用教材(涉及商务、旅游、科技、医学、媒体、公务、交通、工业、体育、法律、国防、政治、军事、外交、航空、工程、金融等十几个专业和职业方向)、文化教材、考试教材等多种分类，满足不同群体的中文学习需求。

因此，国际中文教材的不断发展为语料库的构建提供了大量的可用数据，同时也为语料库的应用提供了丰富的场景和应用场合，为国际中文教学和研究提供了更为丰富的资源和支持。

1.2.4. 具备专业知识的建库团队容易组建

语言点语料库建库人员在语言学、教育学以及国际中文教育专业方面应具备较广泛的知识 and 经验。一般来说，汉语国际教育方面的专业人才都能达到这个要求，这保证了建库团队组建的可能性。

2. 现有汉语语料库建设现状及文献综述

2.1. 现有汉语语料库建设现状

2.1.1. 研究汉语的语料库比例小

语料库有多种类型。现存较多的语料库是专用的，即只收集于某一特定用途的语料。一方面，按照语料的语种，语料库分为单语、双语、多语三类。另一方面，按照语料的组织形式分为平行语料库和比较语料库。我国汉语语料库的建设开始于 20 世纪 80 年代，虽然取得了不少成果，但起步时间晚，语料库数量和质量都存在不足。由于受国外语料库语言学研究 and 发展的影响，加上近年来我国对英语教学的重视程度越来越高，多数国内语料库的建设仍以英语为主，只有少数语料库是完全用来研究汉

语的。

2.1.2. 缺少专用的语言点语料库

目前,国内建设的对外汉语学习语料库有语音教学语料库、动态作文语料库等,而语法汇总的语料库较少。目前为止,国内没有一个以汉语语法为切入点的用于语言对比研究和学习的语言点语料库。

以北京语言大学语料库中心 BCC 语料库为例, BCC 语料库(<http://bcc.blcu.edu.cn/>)是以汉语为主,兼有英语、西班牙语、法语、德语、土耳其语等语言的语料库,其中汉语语料规模约 150 亿字,涵盖了报刊、文学、微博、科技、综合和古汉语等多领域语料。BCC 语料库包括了生语料、分词语料、词性标注语料和句法树,目前已对现代汉语、英语、法语的语料进行词性标注,中、英文句法树则是引自美国宾州大学的中文和英文树库[2]。但使用过后可发现,检索仅能发现它的多种出现场景,对于想要了解语法点的搜索者来说没有直接性的有效性。

2.1.3. 语料库操作运用缺乏科学性

目前国内的语料库建设大多数满足自用需求,缺少共享共用。语料库建设的规范性不强、缺少系统的理论指导、缺少建立标准和统一管理、缺少评测标准等。当前汉语语料库建设的质量良莠不齐、杂乱无序。语料库中的内容得到标注和重新加工的寥寥无几,这使得语料库的使用效率不高,存在着重复性建设[3]。为研究者提供功能丰富、使用方便灵活、效率高的语料库分析工具,是汉语语料库未来的发展方向。

2.2. 有关汉语语料库建设研究的文献概述

柏晓静等人(2002)对双语语料库的构建和加工提出了建设性意见和构想[4]。孙东云(2018)探讨了 BCC 汉语语料库在英汉翻译教学中的应用,为语料库翻译教学提供了新的思路[5]。黄水清和王东波(2021)探讨了国内语料库的构建和标注的方法、流程和策略,并阐述了语料库在语言教学等领域的应用现状,梳理了目前国内具有代表性的各类语料库,并对其建设和发展特点进行了总结和概括,包括单语语料库、双语语料库、特色语料库三种类型[6]。

单语语料库主要是以汉语为主,兼有其他语言的语料库,典型的有北京语言大学语料库中心 BCC 语料库。北京大学 CCL 语料库包含现代汉语语料、古代汉语语料两类单语语料,另外,还收录了二十五史、诸子百家等杂类语料,留学生汉语作文语料、中文学术文献语料库、海外华文网络语料等专题语料[6]。人民日报标注语料库则完成了词语切分、词性标注、专有名词标注、动词和形容词特殊用法标注等加工工作,形成了一个大规模、高质量的现代汉语词义标注语料库。清华汉语树库是有完整句法结构树的汉语句法树语料,清华 TH 语料库(已更名为 THCHS-30 语料库)包括生语料和熟语料,涵盖多种子库语料素材[6]。

双语语料库有以汉英语为主的,包括中国科学院汉英平行语料库、南京大学双语词典研究中心英汉双语平行语料库、清华大学中英平行语料库等。另有其他语种与汉语的双语语料库,例如北京大学计算语言研究所双语平行语料库、北京外国语大学双语平行语料库、南京农业大学古代汉语典籍平行语料库等。其他特色语料库有汉语中介语语料库、北京语言大学 HSK 动态作文语料库、中国传媒大学有声媒体文本语料库等。

综上所述,现有针对语料库的研究文献大多集中在语料库是如何建设、加工及应用的,并对国内已有语料库进行分类,列出其代表的大型语料库。我们发现国内现有的汉语语料库并没有从语言点出发,以不同级别的语法点为检索规则的语料库,而语法对于汉语学习者来说又是一大难点。这产生了语言点语料库建设这一亟待解决的问题。

3. 构建基于国际中文教学的语言点语料库框架设计设想

3.1. 总体框架说明

国际中文教学语言点语料库的设计旨在为学习者和教师提供一个全面且实用的语言点查询与练习平台。该语料库包含四个核心模块：语言点索引、语言点注解、教学解释和练习板块(见图 1)。

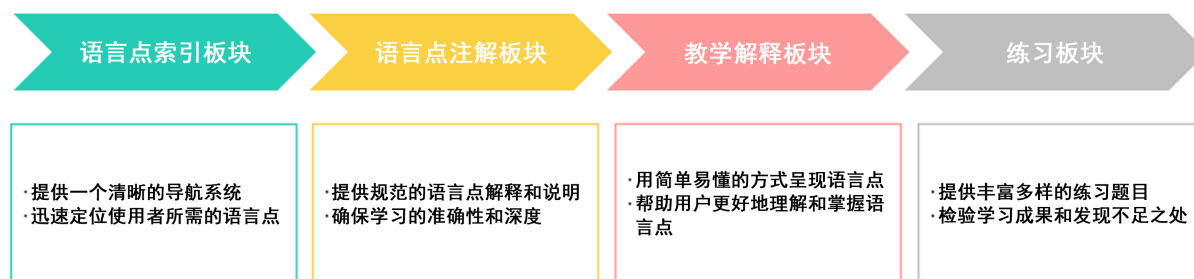


Figure 1. Flowchart of the international Chinese teaching language point corpus framework
图 1. 国际中文教学语言点语料库框架流程图

整体而言，这一框架的设计不仅致力于提高汉语学习的效果，还旨在为学生和教师提供一个互动性强、资源丰富的学习和教学平台，从而促进汉语教学的全面发展。

3.2. 语言点索引板块

以《国际中文教育中文等级水平标准》为依据，以目前主流的几套汉语教材为参考，确立初中高等级汉语语言点的主要内容，科学地、系统地编排索引页面。形成一个条理清晰、框架合理、内容全面的语言点索引板块，帮助汉语学习者或汉语教师对所需要查询的语言点进行快速检索，及时获得客观完整的语料。

3.3. 语言点注解板块

该板块旨在为用户提供客观真实的语料中心。在建立注解内容时，语料库充分考虑了主流汉语教材的使用情况，并进行了大量考证分析，确立了每个语言点的规范注解。

首先，注解板块采用了现实汉语教学中最广泛使用的几套汉语教材作为主要参考来源。这些教材包括了国际中文教育中文等级水平标准所推荐的教材，以及其他在国际中文教学领域具有重要影响力的教材。通过综合考量这些教材的教学理念、教学方法和语言点解释，注解板块确立了每个语言点的规范注解，使得注解内容更加贴近实际教学需求，具有较高的可信度和可用性。

其次，注解板块综合考虑了不同教材对于同一语言点的不同解释方法，进行了归纳总结，形成了一套风格统一、内容一致的注解方式。这种做法有助于用户理解不同教材之间的差异性，帮助他们更好地把握语言点的使用规范，提高了语料库的实用性和适用性。

此外，注解板块还对于每个语言点进行了大量的考证分析，以确保注解内容的客观性和真实性。通过收集和整理大量的语料数据，注解板块形成了丰富的语言点注解，为用户提供了全面而可靠的语料中心。

综上所述，注解板块在语料库中起着重要的作用，为用户提供了客观真实的语言点注解，帮助他们更好地理解和掌握汉语语言知识，提高了语言学习和教学的效率和质量。

3.4. 教学解释板块

该板块以简洁明了的方式将注解内容可视化、简易化，通过例证的方式进行教学展示，为教师和学生

生提供便捷易懂的教学辅助。在教学实施的角度出发,该板块深入分析和归纳了每个语言点,确保了教学解释的科学性和准确性。采用了简洁明了的语言,避免了过多的专业术语和复杂的解释,帮助教师和学生更好地掌握语言点的教学方法和技巧,提高了教学效果和学习效率。因此,教学解释板块为用户提供了全面而有效的语言点教学辅助,促进了汉语学习和教学的进步和提高。

3.5. 练习板块

该板块为用户提供丰富多样、有针对性的语言点练习,以促进学生的自主学习和教师的教学备课。该板块设计的内容根据主流汉语教材和国际汉语教学标准,提供不同难度和类型的练习题,以满足不同水平学生的学习需求。练习内容设计具有针对性和实用性,涵盖填空、改错、选择等题型,帮助学生巩固和提高语言点的应用能力。此外,练习板块还提供详细的练习解析和答案解析,帮助学生理解题目要求和答案逻辑,及时发现和纠正错误,提高学习效率。

4. 构建基于国际中文教学的语言点语料库的使用说明

4.1. 语言点索引板块

该板块旨在为用户提供便捷的语言点查询和检索功能。用户可以通过清晰的页面结构,轻松浏览和选择所需的语言点,每个语言点都有相应的注解和解释页面。此外,语言点索引板块提供了用户友好的界面设计和搜索功能,使用户能够快速准确地定位所需的语言点,确保了检索的高效性和精准性,为用户提供了全面、框架合理且用户友好的语言点查询与检索服务。

4.2. 语言点注解板块

语言点注解材料来源具有准确性与可靠性。每个语法点的规范注解经过深入研究和考证,确保其符合语法规则和实际语言使用的共识。这种一致性使教师和学生在学习过程中不会受到来自不同教材或资源的差异性干扰,从而避免了可能的混淆和误导。

例如,教师在搜索板块搜索“能愿动词:会”,语料库则能迅速跳转到“能愿动词:会”的注解页面,给出客观、规范的注解。既解释了“能愿动词:会”的语法意义,也给出了它的常用用法,让教师不用再花费大量时间翻阅纸质语法书,亦不用在网上搜索客观性、准确性无法保证的语言点注解。再如,学生在搜索板块搜索“能愿动词:想”,语料库则能迅速跳转到“能愿动词想”的注解页面,同时能发现它与“能愿动词:要”是联系在一起的,给出的注解是:句法上,“要”和“想”与其后面的谓词性成分构成述词的“想”宾结构,且“要”具有肯定和否定不对称的现象。

语料库通过连续性的框架为学习者提供了一个完整而连贯的语法学习体验。这一连续性的设计有助于学习者更好地理解语法体系的全貌,而不仅仅是孤立地学习一些零散的语法规则。通过逐步深入学习,学习者能够建立起对语法体系结构的整体认知。这种整体理解有助于避免对语法知识的片面掌握,提高学习者对语法点之间相互影响和衔接的敏感性。此外,适应性的学习特点使得语料库更具针对性,尤其适用于 HSK 初级中级学习者。学习者可以根据自身的水平和需求选择适当难度的语法点进行学习,从而更好地适应个体化的学习进程。这种适应性学习模式有助于提高学习者的学习动机,因为他们能够在相对轻松的环境中逐渐拓展知识。

4.3. 教学解释板块

依据语言点注释板块的内容,从教学实施的角度出发,用简洁明了的方式将注释可视化、简易化,以例证的方式进行教学展示,以供教师和学生自由使用。语料库被设计成一个多功能性学习工具,以满

足广泛的学习和教学需求。这意味着它不仅仅提供简单的语法点解释，更兼具其他丰富的功能，使其成为一个灵活而全面的语言学习平台。

例如，教师在搜索语言点“能愿动词：会”时，得到注解：能愿动词，也叫“助动词”，能用在动词、形容词前面表示客观的可能性、必要性和人的主观意愿，有评议作用。“会”表示通过学习具有了某种能力，不具有某种能力用“不会”。同时，得到例句：我会说中文。我不会说中文。由此得到对比，以例证的方式进行教学展示，为教师提供了对比教学的思路。学生再搜索语言点“能愿动词：想”时，也能得到例句：我想学中文。将教学解释简洁化、可视化，并以例证的方式进行展示，能让学生在自学时也能获得完整的学习体验。

4.4. 练习板块

该板块的设计是为了达到两方面的目的：

一方面，帮助教师进行语言点的练习设计，从对外汉语教材中获取资源，设计有针对性的练习，配合教学过程，为国际汉语教师教学备课提供便利。同时，语料库进行综合检索，形成的练习题客观度更高，逻辑上更为严谨。另一方面，帮助学生进行自主学习。学生可以自由下载或在线浏览语料库中的练习题。练习的设计遵循汉语国际教育的教学理论，实现机械练习、有意义的练习、交际性的练习的梯度排列，符合阶段性教学的要求。

除此之外，在这个有组织的学习路径上，教师和学习者还能清晰地了解每个语法点与其他点之间的关联，这能帮助他们建立起一个完整的语法脉络。例如，搜索“能愿动词：会”，则能出现相应的语言点练习，但练习部分则会和另一个与它关联性极高的“能愿动词：能”一起出现。具体表现为：

选词填空：会，能

- 1、我在图书馆，不()打电话。
- 2、他嗓子还没好，不()说话。
- 3、爷爷()走路了，可以出院了

答案：1、能；2、能；3、会

对于教师来说，这样的练习设计板块汇聚了丰富多样的汉语学习材料，包括各类句型、词汇、对话及实际应用场景等，为教师提供了充足的教学资源。在备课过程中，教师可以根据教学目标和学生水平，从语料库中挑选合适的练习材料，进行有针对性的教学设计。同时，练习板块还提供了多种形式的练习题目，如填空、选择、判断、造句等，有助于教师全面评估学生的汉语水平，并据此调整教学策略。

对于学生来说，语料库提供的练习板块，为学生自主练习汉语提供了极大的便利。这一板块不仅包含了丰富多样的练习材料，还具备灵活多样的练习形式，满足了学生个性化学习的需求。首先，语料库的练习板块涵盖了从基础词汇、语法句型到实际应用场景的各个层次和方面，使得学生能够根据自己的学习水平和目标，自主选择适合的练习内容。无论是初学者还是高级学习者，都能在这里找到适合自己的练习题目，从而进行有针对性的练习。其次，练习板块还提供了多种形式的练习题目，如选择题、填空题、判断题、造句题等，既有巩固基础知识的作用，也有拓展应用能力的功能。学生可以根据自己的喜好和学习需求，选择不同形式的练习，练习完成后能迅速获得正确答案，以便及时纠错，从而激发学习兴趣，提高学习效果。

5. 结语

随着“一带一路”进一步建设发展，对外开放持续深化，国际中文教育在传播中华优秀传统文化、讲好中国故事、传播好中国声音等方面的重要性日益增长。基于国内汉语语料库研究现状、国际中文教

育现状以及师生的教学需求,本文以期建设一个以语言点为核心的语料库。一方面,该语料库通过整体性的框架和适应性的学习设计,为学习者提供了更丰富、更有层次感的语法学习体验,使他们能够更全面、更深入地理解和掌握汉语的语法体系。另一方面,便捷的检索和生动的例子提高了教学者的备课效率和备课质量。国际中文教学语言点语料库的建设能够有效填补市场空缺,满足师生教学需求,传播中国优秀文化,为国际中文教育事业做出贡献。

基金项目

本论文为大学生创新创业训练计划《国际中文教学语言点语料库建设运行项目》(编号:S202310636112)的成果,论文由四川师范大学国际中文教育学院学生彭瑶、刘人铭、谭钰浠在通讯作者四川师范大学王飞华老师指导下完成,特此致谢。

参考文献

- [1] 马箭飞,梁宇,吴应辉,等.国际中文教育教学资源建设70年:成就与展望[J].天津师范大学学报(社会科学版),2021(6):15-22.
- [2] 荀恩东,饶高琦,肖晓悦,等.大数据背景下BCC语料库的研制[J].语料库语言学,2016,3(1):93-109+118.
- [3] 郑萌.国内汉语语料库概况[J].科技视界,2014(27):163+219.
- [4] 柏晓静,常宝宝,詹卫东,等.构建大规模的汉英双语平行语料库[C]//中国中文信息学会,中国人工智能学会.机器翻译研究进展——2002年全国机器翻译研讨会论文集.北京:北京大学计算语言学研究所,2002:8.
- [5] 孙东云.BCC汉语语料库在英汉翻译教学中的应用[J].外语教学理论与实践,2018(3):71-78+98.
- [6] 黄水清,王东波.国内语料库研究综述[J].信息资源管理学报,2021,11(3):4-17+87.