

中国特色对外话语中日平行语料库的建设及应用研究

张 岩

珠海科技学院外国语学院 广东珠海

【摘要】为落实高校外语专业“三进”工作要求，推动课程思政改革与外语教学深度融合，本文以中国特色对外话语为研究对象，采用信息化技术与语料库建设方法，构建中日平行语料库，完成语料采集处理、分词标注、平行对齐及检索功能搭建，形成 50 万字规模的数字化教学资源。该语料库聚焦日语教学实际需求，通过辅助教材编写、融入不同课型日语课堂教学、支撑语料库驱动学习等应用路径，弥补当前“三进”工作中数字化资源不足、基础课程融入不够等短板。研究结果表明，该语料库不仅为新时代中国特色社会主义思想融入外语教学提供了科学工具，也助力学生提升中国话语日语表达能力与家国情怀，为外语专业课程思政改革与教学模式创新提供了实践参考。

【关键词】中国特色对外话语；中日平行语料库；语料库建设；教学应用

【基金项目】广东省哲学社会科学规划 2025 年度常规项目一般课题（项目编号：GD25CWY19）；2025 年度广东省教育科学规划课题（高等教育专项）（项目编号：2025GXJK0694）；教育部产学合作协同育人项目（2510095033）

【收稿日期】2025 年 9 月 15 日

【出刊日期】2025 年 10 月 15 日

【DOI】10.12208/j.ije.20250380

The Chinese-Japanese parallel corpus of International Discourse with Chinese Characteristics: compilation and applications

Yan Zhang

School of Foreign Languages, Zhuhai College of Science and Technology, Zhuhai, Guangdong

【Abstract】To meet the requirements of the "Three Entries" initiative for foreign language majors in colleges and universities and promote the in-depth integration of curriculum ideological and political reform with foreign language teaching, this study takes International Discourse with Chinese Characteristics as the research object, adopts information technology and corpus construction methods to build a Chinese-Japanese parallel corpus. It completes corpus collection and processing, word segmentation and annotation, parallel alignment, and the construction of retrieval functions, forming a digital teaching resource with a scale of 500,000 characters. Focusing on the actual needs of Japanese teaching, the corpus addresses the current shortcomings in the "Three Entries" work, such as insufficient digital resources and inadequate integration into basic courses, through application paths including assisting textbook compilation, integrating into Japanese classroom teaching of different course types, and supporting corpus-driven learning. The research results show that the corpus not only provides a scientific tool for integrating Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era into foreign language teaching but also helps students improve their ability to express Chinese discourse in Japanese and enhance their family and country feelings, offering practical reference for the ideological and political reform of foreign language courses and the innovation of teaching models.

【Keywords】International Discourse with Chinese Characteristics; Chinese-Japanese parallel corpus; Corpus construction; Teaching application

1 引言

新时代中国特色社会主义思想，是党和国家必须长

期坚持的指导思想，是全党全国人民为实现中华民族伟大复兴而奋斗的行动指南。中国特色对外话语相关文本，

收录了中国在治国理政新的实践中发表的一系列重要论述,提出的许多新思想、新观点、新论断,是对外介绍中国的百科全书。中国特色对外话语多语种译本进高校、进教材、进课堂(以下简称“三进”),“明确了外语类高校和高校外语专业以立德树人为导向的课程思政教改路径,激发了外语教学改革的内生动力”^[1]。

本文以高校外语专业贯彻落实“三进”工作为契机,将信息化技术与外语教育教学深度融合,尝试构建中国特色对外话语中日平行语料库。同时,探索该语料库在辅助教材编写、将新时代中国特色社会主义思想融入不同课型的日语课堂教学实践,通过语料库驱动学习改变学习方式等方面的应用。

2 文献综述

2.1 “三进”工作推进历程及现状

2019年年底,中共中央宣传部在北京外国语大学、上海外国语大学、四川外国语大学启动“三进”试点工作。2021年6月7日,光明日报刊登通讯文章,首次介绍了“三进”试点工作在三所院校的进展情况^[2]。同年,6月16日,光明日报再次发表通讯文章,报道了北京语言大学、天津外国语大学、大连外国语大学等高校围绕“三进”试点工作进行的讨论^[3]。2021年6月,“三进”试点高校增加到25所,“三进”工作开始在全国高校外语专业中得以推广和落实。

通过知网检索发现,部分教师在四川外国语大学“三进”教学改革实践的基础上,尝试将中国特色对外话语多语种版本融入英语、日语、俄语专业的翻译鉴赏课程。同时,亦有教师以课程思政为背景,将中国特色对外话语中的思政元素融入英语词汇学课程,齐金鑫(2022)^[4]探讨了大学英语通识课程中的课程思政设计。教材编写方面,2022年8月,外语教育教学研究出版社出版《理解当代中国》多语种系列教材。该系列教材包括读写教程、演讲教程、翻译教程、高级翻译教程等多本教材,涵盖英语、俄语、德语、法语、日语等9个语种和国际中文版本。

但整体来看,目前“三进”工作还处于启动初期,研究的深度和广度还有待进一步拓展。尚存在如下问题:1)中国特色对外话语多语种版本数字化电子匮乏,使用不便。2)将中国特色社会主义思想融入外语翻译课程的教学研究较多,外语基础课、阅读课、听力课程的相关研究较少,融入课堂教学的实践研究不足。3)目前,国内权威出版社已经出版了不同语种的系列教材,但各高校外语专业学生能力差距较大、专业特色各异,在使用过程中需要进行有针对性的修改和补充。

2.2 中日平行语料库的建设及应用

现代语料库语言学(modern corpus linguistics)是20世纪中后期兴起的一门语言研究科学。双语语料库的开发始于20世纪90年代,近年来国内已建成一批双语语料库,如北京外国语大学研发的“中国英汉平行语料库”、上海交通大学建设的“莎士比亚戏剧英汉平行语料库”等。日语方面,语料库建设主要有两种方式。一种是以外语类知名高校为主导依托国家级项目建设的双语语料库,一种是研究者根据研究需要自建的小型语料库。前者有2003年北京外国语大学建设的《中日对译语料库》,建成后一直供内部研究人员使用,现已停用。2022年6月,上外语料库研究院发布中国特色对外话语多语种数据库综合平台2.0版,但目前平台上免费版仅有中文和英文语料,高校购买后可以使用其他语种的第一卷语料。后者,与汉英平行语料库的研究相比,基于自建中日平行语料库的相关研究较少,代表性学者有孙成志(2020)^[5]、王华伟,曹亚辉(2021)^[6]等。孙成志^[5]基于自建日汉平行语料库,从汉译本的宏观语言特征、“视点人物”及“视点表达”的翻译等视角入手,对《起风了》这部作品的林少华(2017)和施小炜(2015)两译本的翻译风格进行对比研究。王华伟,曹亚辉^[6]基于自建的《政府工作报告》小型语料库,分析政治文献日译本的语言特征。

与英语相比,国内中日平行语料库的研制和应用相对滞后,存在的主要问题如下:1)已建成的语料库大多处于非公开状态,无法直接应用于科研、教学研究。2)现有语料库语料内容相对陈旧,不适用于“三进”工作的推进。3)由于翻译文本的特殊性,研究问题多局限在语言研究和翻译研究方面,拓展不够,鲜有教学应用研究。

3 中国特色对外话语中日平行语料库的建设

3.1 建库思路

语料库的建设目的在很大程度上决定了语料库的类型和库容、语料的属性及其加工程度^[7]。本文拟建设的中国特色对外话语中日平行语料库,主要服务于日语教材编写、各种课程的课堂教学实践和学生的语料库驱动学习。

语料库采用穷尽性原则,选取中国特色对外话语第一卷至第四卷中文原文和日文译文全部内容,预计库容达到200万字(以汉字字符计)。目前,已完成第一卷语料库的建设工作。语料库库容中文17.9万字,日文27.9万字,建成的语料库包含中日文双语句级层面平行对齐的EXCELL文件和TXT格式的翻译记忆

文件。使用语料库检索工具,可以实现字符串检索、句型检索、特定组合词语检索等适应各种语言研究和翻译研究要求的检索功能,提供词频、词簇、词语搭配、索引行等信息及可视化展示。

3.2 建库步骤

(1) 语料采集与处理

首先,使用智能扫描软件“扫描全能王”扫描中国特色对外话语中日文纸版书,保存为 PDF 文件。此后,使用“扫描全能王”OCR 文字提取功能,将图片转换成可以编辑的 Word 文档。此后,利用 EditPlus 文本处理软件,再次删除多余的空格,并将校对编辑后的文本保存为 txt 格式,统一为 UTF-8 编码。此时,需要统一文本的命名规则,且注意文件名不能出现汉字。

(2) 文本分词与标注

“分词是指将连在一起的字符分割成相互分离的形符”,“词性赋码是指根据文本的语境信息标注文本内词语的词性”^[8]。为便于对平行处理后的语料进行更为深入的分析,需要对语料进行分词和标注。首先,使用自动分词工具 SegmentAnt 对中文进行分词和词性标注。日语方面,毛文伟(2012)^[8]在对日语开源自动词性赋码器 JUMAN、ChaSen 和 MeCab 进行赋码精度和信度分析的基础上,指出“无论是在赋码质量还是抗干扰性方面,MeCab 都是最出色的”,因此我们也采用日语自动分词赋码器 Mecab 对日语进行分词和词性标

注。此外,还通过人工标注的形式,在每篇文章开头标注文章所属的语言、卷次、主题、小节,文章发表的时间、文本字数等。

(3) 语料平行对齐

为便于后期对双语文本的检索和对比分析,需要对中日文语料进行平行对齐。语料的平行对齐包括段落、句子、短语、单词等层面的对齐,这也是双语语料库区别于单语语料库的建设难点之一。目前常用的双语对齐软件有 Tmxmall、ABBYY Aligner 等。我们使用双语句对齐软件 LF Aligner 2.0 辅以人工调整的方式,实现汉语-日语的句级对齐,并导出 Excell 或 txt 格式文件。

在中日语料对齐过程中,汉语和日语都以“。”“!”等标点作为对齐的依据,整体对齐效率较高。但是,在日语文本中有很多间隔号“・”,在 LF Aligner 2.0 中,将其认定为句末结句符号。因此,需要在前期的文字校对和降噪环节对此进行预处理。

(4) 语料检索

语料库的使用效能,很大程度上取决于对语料库的检索程度。在此,已建成的中国特色对外话语中日平行语料库可以使用本地语料库检索工具 AntConc 进行单语检索,使用语料库检索软件 ParaConc 进行双语检索。上述检索软件均可以实现类符/形符比、句段长、句长、高频词汇检索等功能,并提供可视化的信息和展示界面,为科研分析和教学应用提供了电子化、智慧化的数据。

```
cn10404-gsy-ja10404-wg.tmx - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE tmx SYSTEM "tmx14.dtd">
<tmx version="1.4">
  <header
    creationtool="LF Aligner"
    creationtoolversion="4.21"
    datatype="unknown"
    segtype="sentence"
    adminlang="ZH"
    srclang="ZH"
    o-tmf="TW4Win 2.0 Format"
  >
</header>
<body>
<tu creationdate="20220911T003202Z" creationid="LF Aligner 4.21"><prop type="Txt::Note">cn10404-gsy</prop>
<tuv xml:lang="ZH"><seg>党的十八大作出了实施创新驱动发展战略的重大部署,强调科技创新是提高社会生产力和综合国力
<tuv xml:lang="JA"><seg>中国共産党第十八回全国代表大会は、革新駆動型発展戦略の実施という重要な布石を打ち、科学

<tu creationdate="20220911T003202Z" creationid="LF Aligner 4.21"><prop type="Txt::Note">cn10404-gsy</prop>
<tuv xml:lang="ZH"><seg>这是党中央综合分析国内外大势、立足我国发展全局作出的重大战略抉择。</seg></tuv>
<tuv xml:lang="JA"><seg>これは党中央が国内外の大勢を総合的に分析し、わが国の発展の全局に立脚して下した重要な戦

<tu creationdate="20220911T003202Z" creationid="LF Aligner 4.21"><prop type="Txt::Note">cn10404-gsy</prop>
<tuv xml:lang="ZH"><seg>进入21世纪以来,新一轮科技革命和产业变革正在孕育兴起,全球科技创新呈现出新的发展态势和
<tuv xml:lang="JA"><seg>二十一世紀に入り、新たな科学技術革命と産業革新を迎えつつあり、グローバルな科学技術イノ
```

图1 双语对齐语料的翻译记忆文件样例

A	B
让十三亿人民享有更好更公平的教育 (2013年9月25日)	十三億の人民に、より良いより公平な教育を (二〇一三年九月二十五日)
这是习近平在联合国“教育第一”全球倡议行动一周年纪念活动上发表的贺词要点。	国連「グローバル・エデュケーション・ファースト」イニシアチブ周年記念活動へのビデオメッセージの要旨
百年大计，教育为本。	「百年の大計は、教育にあり」。
教育是人类传承文明和知识、培养年轻一代、创造美好生活的根本途径。	教育は人類の文明や知識を伝承し、若い世代を育成し、より良い生活をつくり出すための根本的な手段である。
中国将继续响应联合国的倡议。	中国は引き続き国連の呼びかけに応える。
中国有2.6亿名在校学生和1500万名教师，发展教育任务繁重。	中国には二億六千万人の児童・生徒・学生、千五百万人の教師があり、教育の発展という任務は非常に重いものである。
中国将坚定实施科教兴国战略，始终把教育摆在优先发展的战略位置，不断扩大投入，努力发展全民教育、终身教育，建设学习型社会，努力让每个孩子享有受教育的机会，努力让13亿人民享有更好更公平的教育，获得发展自身、奉献社会、造福人民的能力。	中国は科学技術・教育による国家振興戦略を確固として実施し、常に教育を優先的に発展させるという戦略的位置に据え、資金投下を絶えず増やし、全民教育と生涯教育の発展、学習型社会の構築に努め、すべての子どもに教育を受ける機会を提供し、十三億の人民がより良いより公平な教育を受け、自己発展、社会貢献、人々に幸福をもたらす力を獲得できるよう努力する。
中国将加强同世界各国的教育交流，扩大教育对外开放，积极支持发展中国家教育事业，同各国人民一道努力，推动人类迈向更加美好的明天。	中国は世界各国と教育面の交流を強化し、教育の対外開放を拡大し、発展途上国の教育事業の発展を積極的にサポートし、各国人民と共に、人類がより素晴らしい明日へと向かう努める。

图2 双语对齐语料的 Excell 文件样例

4 中国对外话语中日平行语料库的教学应用

4.1 辅助教材编写

多年以来，外语教材追求原汁原味的目的语教学素材，多采用引进的日语原版教材，或在中国编写的教材中加入体现日本传统文化、日本人思维方式的相关文章。在日语教材中，很少提及中国文化、中国近现代经济、科技变化，更鲜见新时代中国特色社会主义思想相关内容。因此，日语专业的学生对日本的世情了解较多，甚至对日本文化如数家珍，但却对中国的国情了解不足，不能用日语讲述中国文化和近年来中国在各项改革中取得的成就。近年来，越来越多的教师认识到教材作为教学的重要媒介和载体，必须坚守国家立场，体现价值引领。上海外国语大学德语系教师张帆老师指出，“中国特色对外话语多语种版本作为中国政治话语外译的范本，是打造培根铸魂、启智增慧的精品外语教材的丰富思想宝库”^[2]。

普通院校外语专业，虽然无法直接编写专业基础教材，但可以在现有教材的基础上，编写教材辅助资料和辅助讲义。例如，可以通过自建语料库检索中国特色对外话语中的高频词汇、常用词语搭配信息、重点单词、语法例句、日语翻译例句、阅读文本素材等，编制中日对译的重点词汇库、经典语句库、经典文章诵读库等补充教学资源。让学生在学习语言知识的同时，学习中国特色对外话语中所蕴含的理论、理念，引导学生用中国视角看待中国、观察世界。

4.2 融入专业课程课堂教学实践

将新时代中国特色社会主义思想融入外语专业教学是个系统工程，需要从学院层面进行顶层设计。融入专业课程的路径可采用自上而下、各有侧重、逐级推进的方式进行。首先，从学院层面制定“三进”工作的总体目标和规划，此后将目标细化到不同的课程，根据学生所处的年级，课程的特色分层次、有侧重的将中国特色对外话语中日文本内容纳入“基础日语”“日语视听说”“日语写作”“日语翻译”等专业主干课程的知识目标、能力目标和素质目标。同时，修改课程教学大纲、修改对应的教案和课件，基于语料库构建科学、系统的“三进”教学资源。

其次，发挥课程教学主战场的优势，基于各个课程的课程特色融入不同的教学内容，设置适宜的教学活动。例如，“基础日语”“日语精读”课程可以以教材内容为依托，使用语料库正文库检索合适的素材，从词汇学习、语法讲解、习题演练、话题讨论、课外实践等环节设计设计不同的教学活动。“日语视听说”课程，可依据语料库术语库中的文化关键词，搜集相关的影音资料，通过视听相结合的形式辅助学生对这些词汇和历史事件等背景的理解。还可以设计日语影子跟读、视频配音等任务，让学生在综合锻炼学生日语听、说能力的同时，提升讲述中国故事的能力。“日语泛读”“批判性阅读”课程，可以将中国特色对外话语的内容以专题的方式嵌入到教学内容中，亦可以结合国内外热点

话题,引导学生了解历史和国情,提升思辨能力。同时,中国特色对外话语是典型的政治文本,且译文质量较高,自建的语料库还可以应用于翻译语言特征教学、翻译策略与技巧教学等。

4.3 语料库驱动学习

现在大学生,从小就接触智能手机、平板电脑,是计算机和网络的原住民。基于语料库的驱动学习,更适合这些能够熟练使用计算机,适应线上线下混合学习,崇尚自主学习方式的学生们。针对低年级的学生,教师可以在课上通过语料库现场展示某一词汇的常用搭配、某一语法的语境共现,引导学生通过观察语料库展示的文本,总结单词和语法的细微差异及常见用法。针对高年级的学生,可以引导学生开展线上线下混合式学习。还可以将已建成语料库搭载在校园网的学习平台上,由教师布置学习任务,如在语料库中找出一对近义词的常用搭配,分析近义词词义的细微差别等。学生通过小组合作的方式在线上检索、查阅语料,通过小组讨论的方式总结规律,实现语料库驱动学习。

5 结语

本文在分析外语类高校“三进”工作推进背景和历程,梳理国内中日双语语料库建设和研究现状的基础上,简要阐述了中国特色对外话语中日平行语料库的建库思路、建库步骤。此后,从语料库辅助教材编写、融入专业课堂教学实践、语料库驱动学习三个方面探讨了语料库的教学应用。中国特色对外话语中日平行语料库的构建,数字化资源在外语教学中的应用,对外语专业课程思政改革、对学生了解中国特色对外话语体系的构建、对学生学习方式的转变都具有重要意义。

参考文献

- [1] 冀文亚,张颖天.新时代国际传播人才培养的有益探索[N].光明日报,2021-06-16(01).
- [2] 冀文亚,王璘.心怀“国之大者”担当复兴大任[N].光明日报,2021-06-07(01).
- [3] 冀文亚,张颖天.新时代国际传播人才培养的有益探索[N].光明日报,2021-06-16(01).
- [4] 齐金鑫.大学英语通识课中的课程思政设计——以翻译理论与实践为例[J].山东高等教育,2022,10(04):43-49.
- [5] 孙成志.基于语料库的《起风了》两译本的翻译风格研究——以视点人物和视点表达为中心[J].大连理工大学学报:社会科学版,2020,(5):121-128.王华伟,曹亚辉.基于翻译语料库的政治文献日译本语言特征研究——以《政府工作报告》(2002-2017)日译本为例[J].日语学习与研究,2021(06):11-20.
- [6] 王华伟,曹亚辉.基于翻译语料库的政治文献日译本语言特征研究——以《政府工作报告》(2002-2017)日译本为例[J].日语学习与研究,2021(06):11-20.
- [7] 王克非,黄立波.语料库翻译学十五年[J].中国外语,2008(06):9-14.
- [8] 毛文伟.日语自动词性赋码器的信度研究[J].外语电化教学,2012(03):10-14.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



OPEN ACCESS