

面向教材文本智能改编的 教师语料库素养及实践应用^{*}

金 檀 李芷莹 徐曼菲 唐洁仪

摘要：教材文本智能改编主要包括机器定级后教师改编、机器标注后教师改编、机器替换后教师改编三种范式，每种范式都有各自的适用方法与应用工具。本文介绍了智能改编的三种范式，并结合语料库知识、语料库技术操作技能与语料库教学应用技能三个维度，阐释每种范式所需的教师语料库素养。文章进而通过在职外语教师培训案例，探讨了提升教师语料库素养及智能改编能力的可行路径。

关键词：教材文本；智能改编；语料库素养；在职教师培训

Abstract: Texts in teaching materials can often be adapted by three data-driven paradigms, namely readability assessment for adaptation, complex word and sentence identification for adaptation, and complex sentence replacement for adaptation, each of which has its typical methods and novel tools. This paper introduces the three paradigms and elaborates on the three-dimensional corpus literacy needed for teachers' data-driven text adaptation, including corpus linguistics knowledge, technical skills and pedagogical skills. Through a case of in-service foreign language teacher training, it further discusses a feasible path to developing teachers' corpus literacy and competence for data-driven text adaptation.

Key words: texts in teaching materials; data-driven adaptation; corpus literacy; in-service teacher training

中图分类号：H319 文献标识码：A 文章编号：1004-5112(2023)03-0023-08

1. 引言

英语教材是英语语言和文化知识的载体，是学生英语学习和教师教学的重要素材来源。由于学生的语言能力和发展需求存在差异，教师可对教材文本进行适当调整 (Tomlinson & Masuhara 2017)，以满足学生的多样化学习需求。就语言难度而言，教师可对教材文本进行改编，如调整文本的语言复杂度，使其符合特定语言水平 (Jin & Lu 2018; Jin et al. 2020)。早期的文本改编主要依赖教师或专家的自身知识与经验 (Green & Hawkey 2012; Jin & Lu 2018)。随着人工智能、自然语言处理、机器学习等技术的发展，文本改编逐渐转向智能改编，即借助技术在保留原有文本内容与意义的同时调整语言复杂度，以提高文本可理解性与可读性 (Al-Thanyyan & Azmi 2022)。国内外相关研究主要关注智能改编的工具开发与应用 (Bingel et al. 2018; Jin & Lu 2018; 金檀等 2019)，以及智能改编文本对学生学习的影响 (Odo 2022)。但是，教师借助改编技术与工具对教材文本进行难度调控并开展教学应用的研究仍然少见。

语料库语言学理论与技术工具为教材文本智能改编带来了新思路与新方法。例如，教师可智能选取学术英语阅读素材 (金檀等 2021)，还可通过有效解决基准语料问题提高改编效率

* 本文为广东省本科高校教学质量与教学改革工程建设项目(粤教高函〔2023〕4号，项目序号508)的成果之一。徐曼菲为通讯作者(xumanfei@m.scnu.edu.cn)。感谢《外语界》编辑部及匿名审稿专家的修改意见。

与质量(*Guo et al. 2021*)。教师具有语料库理论与技术知识并将其应用于教学的能力被视作教师利用技术辅助教学的一种专业能力,即教师语料库素养(*Leńko-Szymańska 2017*)。这种素养的外显特征包括对教材文本进行针对性智能改编,设计、开展并评价相应的教学活动。本文介绍智能改编三种范式的概念原理与技术工具,结合教师语料库素养的知识、技能与教学应用能力三个构成维度,阐述教师智能改编工作及其所需的语料库素养,并探讨提升教师语料库素养及智能改编能力的可行路径。

2. 相关概念界定与阐释

2.1 智能改编

智能改编包括难度定级、难点标注与难句替换三种主要技术(*Saggion 2017; Jin & Lu 2018; Jin et al. 2020*),并借助相关应用工具实现。三种技术提供不同层面的改编参考信息,教师需相应采取不同的改编方法与策略,因此对智能改编的应用可分为三种范式:机器定级后教师改编、机器标注后教师改编和机器替换后教师改编(见图 1)。具体而言,在机器定级后教师改编中,教师通过机器获取难度定级信息,确定难度值代表的意义,然后结合教学实际提高或降低文本难度,再将其应用于教学,可总结为“难度意义理解、难度改编参考与难度改编应用”。在机器标注后教师改编中,教师获取难点标注信息,确定各难点(难词或难句)对学生构成的学习难度,然后进行相应改编调整,再在教学中应用文本,可总结为“难点意义理解、难点改编调整与难点改编应用”。在机器替换后教师改编中,教师获得自动改编文本,对其进行效果评估,接着对需要修改的语句进行相应调整,再将文本应用于教学,可总结为“替换效果评估、难句修改调整与自动改编应用”。

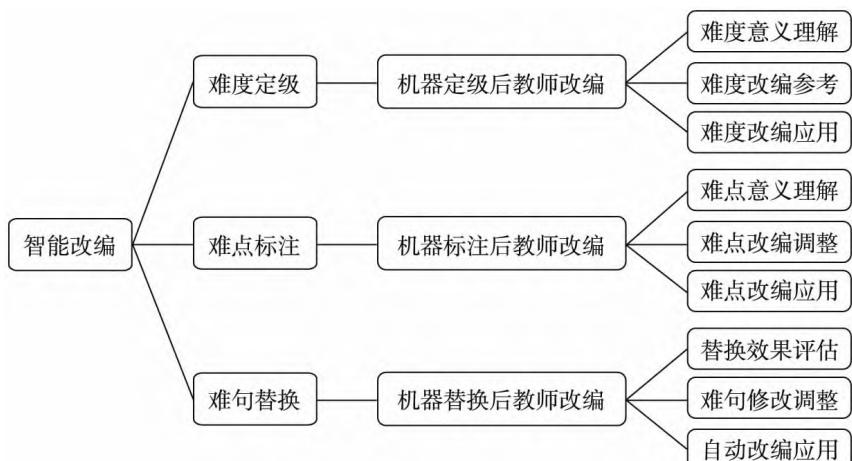


图 1 智能改编的主要技术与范式

2.2 教师语料库素养

教师语料库素养是指教师使用语料库语言学技术和方法开展相关研究,以促进学习者语言习得与发展的素养(*Heather & Helt 2012*)。它是一个多维构念,体现为不同子技能的组合。*Mukherjee (2006)*、*Dalton-Puffer(2014)*认为教师语料库素养包含四个子技能,即理解语料库基本知识、应用语料库工具检索与分析语料库数据、归纳语言使用趋势与解释数据、应用数据分析与解释结果设计教学方案。*Leńko-Szymańska(2017)*则将教师语料库素养总结为语料库知识、语料库技术操作技能和语料库教学应用技能三个维度:第一,教师了解通过语料库能够获

取何种信息、如何进行检索、如何解释检索结果;第二,教师较为熟悉语料库工具与资源,并掌握操作与使用方法;第三,教师能够基于语料库数据设计适宜的教学材料,并运用教学技能将其与教学情境相融合。教师语料库素养的两种界定在本质上相似,只是概括程度不同。两者均提到首先要掌握“语料库知识”,而后者总结的第二种技能“语料库技术操作技能”是前者概括的第二、三种技能;两者均注重教师教学应用视角。本文主要讨论教师智能改编所需的语料库素养,Leńko-Szymańska(2017)更强调语料库素养三个维度在教师培训中的有效培养,因此我们采纳其三维构念开展进一步探讨。

3. 智能改编与语料库素养应用

智能改编三种范式的概念原理、技术工具、教师工作各有差异,所需教师语料库素养的具体构成也有所不同。总体而言,语料库知识可为智能改编应用提供概念知识、技术原理的支持,促进改编信息的理解、评估与调整;语料库技术操作技能可为智能改编应用提供工具使用、结果获取的支持,优化改编工具的操作;语料库教学应用技能可为智能改编应用提供教学设计、教学实践的支持,助力改编文本的教学应用(见图2)。以下分别对智能改编三种范式中教师语料库素养的应用进行阐述。

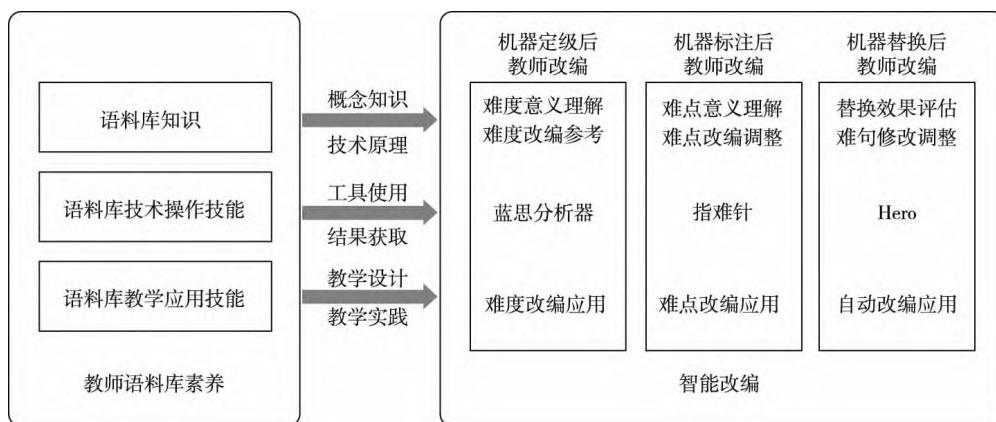


图2 智能改编中教师语料库素养的应用

3.1 机器定级后教师改编

难度定级是指确定文本的复杂程度和难度等级,主要关涉文本可读性。随着语料库技术的发展,文本可读性自动分析逐渐兴起。难度定级的实现一般分为三步。第一步是基准语料建设,即收集本身具有分级性质的典型文本,如不同等级考试真题等(Jin & Lu 2018),构建基准语料库。第二步是核心特征提取,选取对文本复杂度影响较大的指标(如词频、句长等)(Sung *et al.* 2015; Jin *et al.* 2020),以准确预测文本难度。第三步是难度等级匹配,将输入文本经测量所得的指标参数与基准语料文本的等级指标参数范围进行匹配和衔接,从而判断难度等级(Battisti & Ebling 2019)。

难度定级的典型工具之一是“蓝思分析器”(Lexile Analyzer)(MetaMetrics 2022)。该分析器的研究者以美国幼儿园至十二年级的阅读材料作为基准语料,构建了总词汇量达1.4亿的基准语料库。然后,选取难度预测力较强的词频与句长作为核心指标。最后,基于选取的指标,将输入文本的指标参数与基准语料文本的年级指标参数范围进行匹配和衔接,获取输入文本的难度系数与难度等级。比如,输入目标文本的难度系数为1100L,则可匹配至美国八年级

难度水平(难度参数范围为 1010L—1185L)。虽然蓝思分析器主要面向英语为母语的文本,但是整体难度定级实现原理也适用于二语教学文本。Fan *et al.*(2022) 使用蓝思分析器与阅读分级指难针(以下简称“指难针”)测量了中国中考英语、高考英语、大学英语四级与六级考试的难度级别及文本复杂度,发现蓝思分析器能够有效区分四种英语考试的水平等级,并与“指难针”判定的水平等级具有较强的相关性。

机器难度定级可为教师提供文本难度信息。教师应熟悉难度定级工具的功能及其可获取的维度难度系数或等级。例如,蓝思分析器通过文本难度分析后可获取整体难度系数和型式、解码、词汇、句子维度的难易等级。机器定级后,教师改编主要包括三方面工作。第一,难度意义理解。通过蓝思分析器获取难度值与难度等级后,教师理解难度意义,判断文本是否需要改编。教师应通过掌握难度定级、文本复杂度指标等相关概念知识,以及难度定级实现原理“基准语料建设、核心特征提取与难度等级匹配”,理解该工具如何获取文本难度值及其蕴含的数据意义,比如其对应的难度等级。第二,难度改编参考。教师通过比较文本当前难度和改编目标难度,把握难度差距范围。例如,教师参考蓝思分析器衡量文本难度的主要指标,找准难度调整的方向,从词频与句长两个维度改编文本,使其符合目标难度。第三,难度改编应用。教师可在教学设计中灵活运用难度调整后的改编文本,如将难度稍高于学生水平的文本用于拓展阅读,将不同难度的文本用于不同水平学生的分层阅读等;还可充分考虑文本难度因素,基于文本设计相关活动或任务,增强文本的教学应用性。

3.2 机器标注后教师改编

难点标注主要是指对文本中的难词与难句进行识别和标注。难词识别通常是基于词表的词汇识别(Watanabe *et al.* 2010),即使用预制的单词列表区分输入文本中复杂和非复杂单词。难词标注过程主要可分为三步。第一步是目标词表选取,确定学习者在某个水平阶段应掌握的核心词汇。第二步是词族词表生成,将核心词汇同一词族的其他词汇合并生成词族词表。第三步是超纲词汇标注,基于目标词表与词族词表对目标文本进行词汇覆盖,将未在两个词表中出现的词汇标注为难词。

难句识别通常与句法复杂度相关联,而不同指标对句法复杂度的影响程度不同。Jin *et al.*(2020)对教材文本句法复杂度的研究发现,每个 T 单位中从属子句数量、每个子句中复杂名词短语数量和每个子句中非限定性成分数量等指标对句法复杂度的影响程度较高,可用于判断句子难度。难句标注可分三步实现。第一步是较难句子标注,主要依据句子或子句层面的句法指标,如每个 T 单位中从属子句数量等。第二步是较难短语标注,主要参考短语层面的句法指标,如每个子句中复杂名词短语数量、每个子句中非限定性动词短语数量等。第三步是句子长度标注,句子长度是一个能够有效衡量句法难度且便于教师理解与操作的指标。

难点标注的典型工具之一是“指难针”(金檀,陆小飞等 2023)。“指难针”包含三个核心模块,分别为“难度定级”、“词汇分析”和“句法分析”,其中“词汇分析”与“句法分析”主要对目标文本中的超纲词汇与复杂句法结构进行标注。在“词汇分析”模块,研制者首先以不同学习阶段(初中、高中、大学英语四级等)课程标准或考试大纲的词表作为参考核心词汇(金檀等 2019),如《全国大学英语四、六级考试大纲》词表(全国大学英语四、六级考试委员会 2016)。然后,基于 Nation(2017) 的词族列表,拓展所选词表中核心词汇的词族成员,生成词族词表。最后,根据所选的学习阶段,标注目标文本对应阶段的超纲词汇。在“句法分析”模块,对文本的限定性从属子句、复杂名词短语与非限定性动词短语进行标注,以此衡量较难句子与较难短

语。此外,长句通常是学生理解的难点,也需标注。

机器难点标注可为教师提供文本难点信息。教师可了解难点标注相关工具,熟悉其标注功能以及需获取的维度标注结果。例如,使用“指难针”的教师应了解该工具通过文本分析后可得到超纲词汇与复杂句法结构的标注。机器标注后,教师改编主要包括三方面工作。第一,难点意义理解。通过“指难针”获得难词与难句的标注信息后,教师理解难点难在何处。教师应掌握词族、词汇覆盖率、句法复杂度等概念知识,以及难词标注实现原理“目标词表选取、词族词表生成与超纲词汇标注”、难句标注实现原理“较难句子标注、较难短语标注与句子长度反馈”,以掌握难词与难句的界定。第二,难点改编调整。基于“指难针”的难词难句标注及教师的难点理解,调整词汇或句法结构,以降低难词难句的比例。调整难点时亦可参考“指难针”与难点标注相关的拓展功能,如超纲词汇在“指难针”内嵌基准语料库中的出现频次与例句等。此外,还可参考专家改编方法,如形式上改变词频、拆分句子,内容上省略信息、补充解释,衔接上增加连接词、还原代词等方法(Rets et al. 2022)。第三,难点改编应用。教师根据教学目标和教学重难点,结合具体教学情境调试难点比例,以便将改编文本有效应用于教学。比如,某些超纲词对特定情境下的语篇理解较为重要,教师可在改编文本中予以保留,并提供学习语境;某单元的语言学习重点为定语从句,则可相应保留语篇材料中的定语从句,或者可将其他复杂句法结构转换为定语从句,便于开展定语从句教学。

3.3 机器替换后教师改编

难句替换是指以句子为单位,将其自动替换为更简单的同义词或句法结构(Al-Thanayan & Azmi 2022)。难句替换主要可分为三步。第一步是文本难点识别。文本难点主要为难词或难句:难词识别可参考词汇频率等词汇特征,以及母语者或学习者的难度评分;难句识别可参考句法结构规则,如从属、关系从句等。根据词汇特征与句法结构规则,通过相关算法模型可对难词与难句进行识别。第二步是文本难点标注,机器标注识别的难点,突显难点所在位置。第三步是难句自动替换,机器在句子层面进行自动替换,实现自动改编。难句替换通常基于句法结构,通过规则重写实现简化操作,如句子拆分、子句重建等。

难句替换的典型工具之一是Hero自动改编工具(Bingel et al. 2018)。Hero首先对导入文本进行难点识别。难词识别参考“Complex Word Identification”的词汇难度评分,难句识别参考句法结构规则。然后,机器将包含难点并可进一步自动改编的句子标为紫色。最后,使用者选择并点击需要改编的标紫句子,机器对句子进行自动替换,改编完成后句子变为绿色。以一篇关于冬奥会志愿者的文章(“指难针”测量难度为7.00,对应考研难度)(Nizeyimana 2022)为例,Hero将“I was excited to see it as I am a sports enthusiast, and in my home country this kind of competitions have not taken place yet.”标注为难句,将其改编为“I was excited to see it because I love sports. In my home country, such competitions are not held yet.”。这一改编主要是使用更简单的词汇表达和句法结构,如将“a sports enthusiast”替换为“love sports”。

机器难句替换可为教师自动改编提供参考。教师应了解难句替换相关工具,熟悉其替换功能与特点。比如,Hero通过文本分析可标注并自动替换难句。机器替换后,教师改编主要包括三方面工作。第一,替换效果评估。Hero自动生成改编文本后,教师评估改编文本的质量与效果。教师应掌握自动替换等相关概念知识,以及难句替换实现原理“文本难点识别、文本难点标注与难句自动替换”。在此基础上,教师结合自身语言知识与教学经验进一步评估改编文本,如衡量改编文本在词汇层面是否过易、在语篇层面有无信息缺失等。第二,难句修

改调整。依据替换效果评估,教师确定改编文本中需要修改之处并进行调整,以达到目标改编效果。比如,改编文本在词汇层面过易时可将其替换为水平相符的近义词汇,在语篇层面信息缺失时可根据其在文中的重要程度进行适当还原等。第三,自动改编应用。教师将机器自动改编文本应用于教学。比如,将机器自动改编文本修改调整后应用于阅读教学,或给学生同时呈现原始文本与改编文本,让其通过比较文本某一层面的差异来归纳或推断相关语言特点,促进语言习得和学习策略运用。

4. 语料库素养培训与发展

前文阐述了智能改编三种范式的概念原理、技术工具及教师智能改编工作所需的语料库素养。然而,由于相关理论知识或技术能力水平的差异性,教师在教材文本的智能改编中可能会面临不同困难。语料库素养的提升可为教师有效应用智能改编奠定基础,主要通过教师自主学习和专业培训课程实现(Bennett & Dhonnchadha 2023)。从个体层面而言,教师须掌握智能改编相关知识,学习使用智能改编工具,在教学中应用,在实践中学习,并反思智能改编过程,不断提升智能改编应用效果与质量。从培训机构层面而言,培训课程须整合梳理应教授的知识与技能,合理设计各个培训环节的要点与内容,有效实施培训并开展评价。培训课程可增强教师对语料库技术的应用意识,以及将语料库技术与教学融合的能力(Zareva 2017; Li & Xu 2022; Ma et al. 2022),由此发展教师语料库素养。

围绕语料库知识、语料库技术操作技能与语料库教学应用技能,可设置相应的培训环节与内容。培训以教学任务驱动,从实际教情出发,通过概念原理讲解、技术工具操作示范为教学任务提供实施方案,旨在使教师培训后能在实际教学设计过程中参考并应用培训所学内容。各培训环节促进不同维度语料库素养的发展:概念原理讲解、技术工具操作示范主要针对语料库知识水平与语料库技术操作技能提升;教学任务实施与教学设计参考主要针对语料库教学应用技能提升。

以素材改编专题培训中的难点标注主题为例,首先采取任务驱动法,要求学员针对提供的教材文本(如上海外语教育出版社出版的《大学学术英语读写教程》下册第二单元课文),指出需改编的语言难度较高之处,并提出改编建议,使其适合大学英语四级水平学生的阅读。接着,培训者结合相关文献介绍文本改编方法,并通过具体例子讲解改编方法,如形式维度的改编方法之一是“改变词频”,可用高频词替换低频词等。最后,培训者讲解“指难针”的难词难句标注实现原理,并进一步演示与引导如何参考“指难针”的标注结果简化文本难度,具体步骤分为难词难句标注、改编方法使用与目标难度确认:

(1) 基于“指难针”,标注出难词“aerobic, ambiguity, alleviate, meditation, psychiatry”等,以及复杂句法结构,如文章最长句“However, they also wonder whether the benefits of exercise are a side effect of the sound sleep that people experience after exercise, or perhaps a sense of accomplishment or the fact that one’s improved physique may influence one’s emotional state and self-esteem.”。(2) 采用相关改编方法对难词难句进行调整,如改变词频、拆分句子、补充解释等。(3) 文本难度由 6.24 降至 5.47,“指难针”设定难度值 5.00—5.99 之间为大学英语四级难度水平(参见金檀等 2023),因此文本适合大学英语四级水平学生。

至此,培训任务完成,改编文本也为参与培训的教师提供了难度简化版文本参考。

5. 结语

本文围绕智能改编的机器定级后教师改编、机器标注后教师改编、机器替换后教师改编三

种范式,介绍其概念原理与技术工具,并从教师语料库素养的语料库知识、语料库技术操作技能与语料库教学应用技能三个维度切入,阐述教师智能改编工作及其所需的语料库素养。文章进一步以在职外语教师培训中的智能改编专题为例,说明技术赋能下教师语料库素养及智能改编能力的提升路径。

教材文本智能改编为外语教师灵活运用教材开拓了更广阔的空间。教师一方面可以调控教材文本难度,对难度过高或过低的语言进行针对性处理,为实现个性化教学提供基础。另一方面,通过了解和掌握智能改编的原理和技术,教师可实施拓展教学。比如,选择同一主题下的多篇文本开展教学,可以是控制语言难度的跨学科文本,也可以是利于培养比较、分析等能力的不同难度文本。技术赋能的外语教学为教师带来更多课堂创新的可能性,体现教师专业发展的动向,即教师能够融合应用信息技术和教学资源开展教学活动,促进教学的数据化、可视化和科学化,实现智慧教学(胡开宝 2022)。在未来研究与实践中,可进一步关注智能改编在教与学过程中的应用情况,如教师智能改编方法的使用效果、学生学习改编文本的学业表现,以及教师智能改编的能力发展等。相关研究成果也将不断丰富教师语料库素养培训方式与内容。

参 考 文 献

- [1] Al-Thanyan S S & Azmi A M. Automated text simplification: A survey [J]. *ACM Computing Surveys* , 2022 , 54(2) : 1–36.
- [2] Battisti A & Ebling S. A corpus for automatic readability assessment and text simplification of German [J]. *arXiv* , 2019. <https://doi.org/10.48550/arXiv.1909.09067>.
- [3] Bennett C & Dhonnchadha E U. Becoming corpus literate: An in-service EFL teacher education framework for integrating corpora into EFL teaching [J]. *Applied Corpus Linguistics* , 2023 , 3(1) . <https://doi.org/10.1016/j.acorp.2023.100048>.
- [4] Bingel J , Paetzold G & Søgaard A. Lexi: A tool for adaptive , personalized text simplification [A]. In Bender E M , Derczynski L & Isabelle P (eds.) . *Proceedings of the 27th International Conference on Computational Linguistics* [C]. Santa Fe , NM: Association for Computational Linguistics , 2018. 245–258.
- [5] Dalton-Puffer C. *Corpus Linguistics in Language Teacher Education* [R]. Klagenfurt: The 14th Klagenfurt Conference on Corpus-Based Applied Linguistics , 2014.
- [6] Fan J , Jin T & Chen I. *Applying the Lexile Framework to Evaluate Text Complexity in the Chinese Context: A Corpus-based Study* [EB/OL]. https://www.britishcouncil.org/sites/default/files/rirg_fan_jin_chen_layout_1.pdf , 2022.
- [7] Green A & Hawkey R. Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading [J]. *Language Testing* , 2012 , 29(1) : 109–129.
- [8] Guo K , Chen J , Lei J & Jin T. Teacher engagement with technology-enhanced text adaptation for reading assessment: A case study [J]. *International Journal of Computer-Assisted Language Learning and Teaching* , 2021 , 11(4) : 100–112.
- [9] Heather J & Helt M. Evaluating corpus literacy training for pre-service language teachers: Six case studies [J]. *Journal of Technology and Teacher Education* , 2012 , 20(4) : 415–440.
- [10] Jin T & Lu X. A data-driven approach to text adaptation in teaching material preparation: Design , implementation , and teacher professional development [J]. *TESOL Quarterly* , 2018 , 52(2) : 457–467.
- [11] Jin T , Lu X & Ni J. Syntactic complexity in adapted teaching materials: Differences among grade levels and

- implications for benchmarking [J]. *The Modern Language Journal* , 2020 , 104(1) : 192–208.
- [12] Leńska-Szymańska A. Training teachers in data-driven learning: Tackling the challenge [J]. *Language Learning & Technology* , 2017 , 21(3) : 217–241.
- [13] Li S & Xu M. Corpus literacy empowerment: Taking stock of research to look forward for practice [J]. *Journal of China Computer-Assisted Language Learning* , 2022 , 2(1) : 126–155.
- [14] Ma Q , Tang J & Lin S. The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration [J]. *Computer Assisted Language Learning* , 2022 , 35(9) : 2731–2760.
- [15] MetaMetrics. *The Lexile® Framework for Reading Development and Validity Evidence* [EB/OL]. https://partnerhelp.metametricsinc.com/image/Lexile%20Framework%20for%20Reading%20Validity%20Evidence_2022.pdf , 2022.
- [16] Mukherjee J. Corpus linguistics and language pedagogy: The state of the art – And beyond [A]. In Braun S , Kohn K & Mukherjee J (eds.) . *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods* [C]. Frankfurt: Peter Lang , 2006. 5–24.
- [17] Nation I S P. *The BNC/COCA Level 6 Word Family Lists* (Version 1.0.0) [EB/OL]. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists> , 2017.
- [18] Nizeyimana P. *Volunteering at Winter Olympics* [EB/OL]. http://www.china.org.cn/sports/2022-04/08/content_78154617.htm , 2022.
- [19] Odo D M. The effect of automatic text simplification on L2 readers' text comprehension [J]. *Applied Linguistics* , 2022. <https://doi.org/10.1093/applin/amac057>.
- [20] Rets I , Astruc L , Coughlan T & Stickler U. Approaches to simplifying academic texts in English: English teachers' views and practices [J]. *English for Specific Purposes* , 2022 , 68: 31–46.
- [21] Saggion H. *Automatic Text Simplification* [M]. Cham: Springer , 2017.
- [22] Sung Y-T , Lin W-C , Dyson S B , Chang K-E & Chen Y-C. Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR [J]. *The Modern Language Journal* , 2015 , 99(2) : 371–391.
- [23] Tomlinson B & Masuhara H. *The Complete Guide to the Theory and Practice of Materials Development for Language Learning* [M]. Hoboken , NJ: John Wiley & Sons , 2017.
- [24] Watanabe M W , Candido Jr A , Amâncio M A , De Oliveira M , Pardo T A S , Fortes R P M & Aluísio S M. Adapting web content for low-literacy readers by using lexical elaboration and named entities labeling [J]. *New Review of Hypermedia and Multimedia* , 2010 , 16(3) : 303–327.
- [25] Zareva A. Incorporating corpus literacy skills into TESOL teacher training [J]. *ELT Journal* , 2017 , 71(1) : 69–79.
- [26] 胡开宝. 语言数据科学与应用学科: 特征、领域与方法 [J]. 外语界 2022 (3) : 37–44.
- [27] 金檀 李世炜 刘丰恺. 学术英语阅读素材的智能选取与教学应用研究 [J]. 外语界 2021 (1) : 20–27.
- [28] 金檀 刘康龙 吴金城. 学术英语教材词表的研制范式与实践应用 [J]. 外语界 2019 (5) : 21–29.
- [29] 金檀 陆小飞等. 阅读分级指难针 [DB/OL]. <https://www.languagedata.net/tester> , 2023.
- [30] 全国大学英语四、六级考试委员会. 全国大学英语四、六级考试大纲(2016 年修订版) [Z]. 上海: 上海交通大学出版社 2016.