

外向型融媒汉语学习词典与大语言模型释义对比研究

——以汉语方言词为例

李湘怡

鲁东大学人文学院, 山东 烟台

收稿日期: 2026年6月2日; 录用日期: 2026年7月14日; 发布日期: 2026年7月24日

摘要

外向型汉语学习词典是国际中文教育的重要工具,但其方言词的释义长期缺乏统一标准。以《JUZI汉语》为代表的融媒词典正在尝试突破传统编纂模式,而大语言模型作为新兴的汉语检索学习途径,其语言处理能力同样受到关注。两者均为学习者可通过网络直接获取的语言资源,但二者在具体方言词的释义差异尚缺乏系统的实证比较。本研究以《JUZI汉语》为融媒词典案例,以DeepSeek和豆包为两个大语言模型代表,将三者进行对比:首先对《JUZI汉语》收录的方言词进行难度标注,继而再依据语义透明度理论,在每个难度层级内部选取典型样本词;最后从释义原则与释义用词等维度,对三者所展示的方言词释义进行对比分析。本研究将专用学习词典与大语言模型进行对比,初步揭示了二者在方言词释义处理上的异同。相关发现可为学习者在网络环境下选择方言词查询途径提供参考,也可为融媒词典的后续优化及大语言模型在教育领域的应用探索提供实证依据。

关键词

融媒词典, 大语言模型, 国际中文教育

A Comparative Study on Definitions between Outward-Oriented Media-Integrated Chinese Learning Dictionaries and Large Language Models —A Case Study of Chinese Dialect Words

Xiangyi Li

School of Humanities, Ludong University, Yantai Shandong

Received: June 2, 2026; accepted: July 14, 2026; published: July 24, 2026

Abstract

Outward-oriented Chinese learning dictionaries are important tools for international Chinese language education, yet the definition of dialectal words in such dictionaries has long lacked a unified standard. The media-integrated dictionary represented by “JUZI Chinese” is attempting to break through traditional compilation models, while Large Language Models, as emerging channels for Chinese language retrieval and learning, have also attracted attention for their language processing capabilities. Both are language resources directly accessible to learners via the internet, but there is still a lack of systematic empirical comparison regarding their differences in defining specific dialectal words. This study takes “JUZI Chinese” as a case of media-integrated dictionary, and DeepSeek and Doubao as two representatives of Large Language Models, conducting a three-way comparison. First, the dialectal words included in “JUZI Chinese” are graded by difficulty. Then, based on the theory of semantic transparency, typical sample words are selected from each difficulty level. Finally, a comparative analysis of the dialect word definitions presented by the three resources is carried out from the dimensions of definition principles and definition wording. By comparing a specialized learner’s dictionary with Large Language Models, this study preliminarily reveals the similarities and differences in their handling of dialect word definitions. The findings may serve as a reference for learners in choosing query channels for dialectal words in an online environment, and also provide empirical evidence for the future optimization of media-integrated dictionaries and the exploration of Large Language Models applications in educational contexts.

Keywords

Media-Integrated Dictionary, Large Language Models, International Chinese Language Education

Copyright © 2026 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

外向型汉语学习词典是国际中文教育领域的重要辅助工具，其释义质量直接关系到学习者的词汇习得效果。然而，长期以来，方言词作为兼具地域特征与文化内涵的特殊词汇类别，在现有学习词典中缺乏统一、规范的释义标准，给学习者的理解和运用带来困难。近年来，以《JUZI 汉语》为代表的融媒词典尝试突破传统编纂模式，借助多模态技术提升释义的可理解性与教学适配性；与此同时，以大语言模型为代表的新兴语言处理工具，凭借其强大的统计学习与动态生成能力，正逐步成为学习者可便捷获取的语言资源。两类资源虽均面向网络环境下的汉语学习者，但二者在方言词释义处理上的差异尚未得到系统的实证比较。

本研究以《JUZI 汉语》为融媒词典案例，以 DeepSeek 和豆包为两个大语言模型代表，从释义原则、释义语言种类、语言难度及释义方式等维度进行对比分析。通过对十个方言词样本的考察，揭示三者各自的优势与不足，并在此基础上提出融媒词典优化及大语言模型辅助二语教学与词典编纂的可行路径。研究旨在为学习者在网络环境下选择方言词查询途径提供参考，也为融媒词典的后续发展与大语言模型在教育领域的深度应用提供实证依据。

2. 文献综述

2.1. 方言词本体研究

方言词研究历来是汉语方言学与词汇学的重要交叉领域。近年方言词研究成果主要集中在四个方面。第一，方言词调查与描写研究。吴沁茗(2020)等分别对金坛方言、修水方言、淮滨方言等进行系统调查，积累了丰富的第一手材料[1]。第二，方言词个案考释研究。许洁红(2025)等分别展开本字考释与词义演变分析[2]。第三，方言词比较研究。这类研究多数是对方言词汇研究进行系统梳理，部分成果涉及方言与普通话的对应关系考察。第四，方言词文化阐释研究。众多学者分别从丧葬词、饮食词等角度探讨方言词的地域文化内涵。

2.2. 外向型汉语学习词典编纂研究

从外向型汉语学习词典编纂研究来看，微观结构研究主要关注词条内部的释义、配例、语法信息及词条结构设计等问题，是学习词典研究中成果最为丰富的领域。其中，释义研究始终是关注重点。众多学者从释义元语言及语义场理论、释义用词难度以及释义原则等方面进行了讨论与研究。配例研究方面，付娜(2010)强调例句的用法示范功能[3]，杨玉玲、段彤彤(2023)提出从难度控制角度优化例句设计[4]。

宏观结构研究主要涉及词典的收词原则、收词范围及整体框架设计。李禄兴(2006)指出学习词典应以学习者需求为中心[5]，杨金华(2016)提出针对性、科学性、系统性和实用性四项编纂原则[6]。然而，现有外向型汉语学习词典研究主要集中于普通词汇及部分特殊词类，对方言词的关注十分有限。

3. 外向型融媒学习词典与大语言模型方言词释义对比分析

陈贤德等(2025)指出释义质量是评价一部词典水准的核心维度之一，一部优质的外向型汉语学习词典，需在释义上同时满足准确性、可理解性、系统性等多元标准[7]。然而对于编者而言，要实现高质量释义是一项极具挑战性的任务。这不仅要求编者需兼具深厚的语言学理论素养和广博的跨学科知识储备，还需编者熟练掌握并灵活应用多种释义策略，整个释义过程耗时、耗力，而在大语言模型的辅助下，这一过程可以得到优化。

3.1. 融媒词典与大语言模型的词汇释义原则与理论

3.1.1. 《JUZI 汉语》释义理论依据

释义原则是词典编纂理念的核心体现，决定了释义内容的组织方式与呈现效果。对于外向型汉语学习词典而言，释义原则的制定必须立足于二语学习者的认知特点与习得规律。鲁健骥和吕文华(2006)系统提出了四项释义原则——可读性、熟知性、区别性与提示性，这四项原则被学界视为外向型汉语学习词典编纂的重要理论依据[8]。

• 可读性原则：从“文字易懂”到“多模态易懂”

可读性原则要求释义语言必须控制在学习者的现有语言水平之内，确保其能够独立理解。这一原则的学理依据在于克拉申的“可理解输入”假说——只有输入内容略高于学习者现有水平且具备可理解性，语言习得才能发生。鲁健骥和吕文华(2006)将这一理念引入词典编纂，明确提出释义用词不应超出被释词难度等级[8]。

《JUZI 汉语》在继承这一原则的同时，借助融媒技术实现了从“文字易懂”向“多模态易懂”的拓展。杨玉玲(2024)指出，多模态释义(图文匹配、视频辅助)可显著降低用户的认知难度[9]。陈贤德等(2024)的眼动研究进一步证实，图文匹配的释义方式能够减少学习者的总注视时间，提高词义理解的效率[10]。这意味着，在融媒环境下，可读性原则的内涵已从“控制文字难度”扩展为“优化多模态认知路径”。

- 熟知性原则：释义用词的系统化控制

熟知性原则要求释义所用词汇必须是学习者已经掌握或熟悉的，避免用陌生词汇解释目标词。蔡永强(2018)指出，当前外向型汉语学习词典释义的突出问题在于“内汉痕迹”过重——即外向型词典仍沿用内向型词典的释义方式，使用超过学习者认知范围的词汇[11]。熟知性原则正是对这一问题的针对性回应。

《JUZI 汉语》将熟知性原则落实为可操作的技术路径：杨玉玲(2024)说明了《JUZI 汉语》依据《国际中文教育中文水平等级标准》对所有词条进行难度标注，在释义中优先选用难度不高于被释词的词汇。这种“分级控制”策略，使熟知性原则从理念倡导走向了系统化的编纂实践[9]。对于方言词而言，这一原则尤为关键——释义需借助学习者已掌握的普通话词汇来解释方言表达，而非引入更多生僻词语。

- 区别性原则：从“定义”到“辨析”的功能转向

区别性原则要求释义能够清晰地区分近义词、多义词的不同义项，避免同义循环。

《JUZI 汉语》在微观结构中专设“辨析”模块，将区别性原则从释义文本内部延伸到独立的对比分析区域。杨玉玲(2024)指出，这一设计体现了外向型学习词典从“解码功能”向“编码功能”的转向——词典不仅要帮助学习者理解词义，更要指导其准确运用[9]。

- 提示性原则：语法语用信息的显性化呈现

提示性原则要求在释义中标注词语的语法限制、搭配关系、语用条件等信息。章宜华(2010)指出，外向型学习词典应将句法结构、搭配结构等分布信息纳入释义，而非仅关注词义界定。这一原则的核心在于承认：对于二语学习者而言，知道“用什么词”远不如知道“怎么用这个词”重要[12]。

《JUZI 汉语》通过多种方式落实提示性原则：设置“搭配”模块呈现典型组合，以“注意”项标注易错用法，通过“会话”视频将词语置于真实交际语境中。这些设计使词典从静态的知识载体转变为动态的学习工具。在方言词释义中，提示性原则尤其重要——许多方言词的使用受到地域、语体、交际场景的严格限制，若缺乏明确的语用提示，学习者极易发生偏误。

3.1.2. 大语言模型释义理论依据

与传统词典的释义范式不同，大语言模型的词汇释义行为根植于其技术架构与训练机制，呈现出一种“数据-统计-生成”驱动的全新逻辑。当前学界尚未为大语言模型建立起类似研究趋于成熟的传统词典的系统性释义原则，但学者们从技术性能、教学适配性及词典学应用等角度，揭示了大语言模型在处理词汇释义时所隐含的若干规律性特征。本文将这些特征归纳为以下三个核心原则，以期为本研究的对比分析建立理论框架。

- 以高频用法为基准的统计显著性原则

大语言模型对词义的理解与输出，本质上依赖于训练语料中词汇共现模式的概率分布。这一技术特性决定了大语言模型在释义时天然倾向于优先呈现统计上显著的高频义项与典型搭配，即遵循“统计显著性优先”的原则。陈贤德等(2025)对 DeepSeek 的实证研究表明，大语言模型对名动兼类词义频的统计结果与基于 BCC 语料库的人工统计结果高度趋近，这意味着大语言模型能够自动将高频义项排在前列，客观上契合了二语习得中“高频优先”的学习规律。[7]然而，该原则在处理方言词时存在明显局限：方言词的高频用法往往局限于特定地域，在通用语料中呈现低频甚至缺失状态，大语言模型极有可能优先输出其普通话对应词的字面义，而掩盖其真正具有地域文化特征的方言义。因此，在面向二语学习者的方言词释义中，必须对大语言模型的统计输出进行人为干预与校准。

- 上下文感知的动态适切性原则

大语言模型区别于传统词典的显著特征在于其对上下文的动态敏感性。借助词嵌入与注意力机制，大语言模型能够将词汇映射为高维向量，并根据查询语境实时调整其输出的语义解释。这种“上下文动

态適切性”原则使得大语言模型可以为一个词在褒义、贬义或中性语境中生成侧重点不同的释义,展现出传统静态词典难以具备的灵活性。章宜华(2024)进一步指出,大语言模型的多模态信息处理能力可以显著增强词典的释义表达能力,使释义更贴近真实实际场景[13]。但是,这种动态性也带来了释义“确定性”的丧失——同一方言词在不同对话轮次中可能得到不一致的解释,这对于依赖权威、稳定知识输入的二语学习者而言是一个显著风险。因此,大语言模型的动态释义必须与词典的系统性要求相协调,而非简单取代。

• 人机协同的可校验与可干预原则

鉴于大语言模型的“幻觉”现象以及统计偏差的客观存在,众多学者普遍认为,大语言模型难以独立承担面向学习者的权威释义任务,必须纳入人类专家的校验与干预机制。这一“人机协同的可校验原则”是大语言模型应用于外向型词典编纂的前提条件。陈贤德等(2025)的研究证实,通过向 DeepSeek 提供《JUZI 汉语》中的释义元词与用法型式作为“垂域知识库”,其生成的释义质量得到显著提升[7]。这表明,大语言模型的释义输出并非不可控,而是可以通过专家知识的注入实现有效引导。

3.2. 释义语言种类对比

3.2.1. JUZI 汉语的释义语言种类

在外向型融媒汉语学习词典中,释义语言种类的选择直接影响到学习者的理解效率与词典的实用性。

《JUZI 汉语》作为融媒词典的代表性成果,在释义语言的设置上呈现出“汉语主导、英语辅助”的双语协同特征。这一模式既区别于传统内向型词典以汉语单语释义为主的编纂传统,也与部分外向型电子词典(如 Pleco、TrainChinese 等)过度依赖英语对译、忽视汉语释义的做法形成差异。

具体而言,《JUZI 汉语》的释义语言主要包含以下两种类型。

汉语释义。这是该词典释义语言的核心组成部分。杨玉玲(2024)指出,《JUZI 汉语》在编纂过程中贯彻了“释义元语言”理念,即严格控制释义所用词汇的数量与难度,优先选用《国际中文教育中文水平等级标准》中的初、中级词汇进行释义,以确保释义文本的可理解性。在具体实践中,该词典对大部分词目提供了完整的汉语解释,释义语言力求浅显、简短,避免使用超出学习者认知范围的复杂句式[9]。

英语对译。在汉语释义的基础上,《JUZI 汉语》为绝大多数词目提供了对应的英文翻译。杨玉玲(2022)将这一模式概括为“目的语 + 英语”释义模式,即主要用汉语来解释汉语,辅之以必要的英语进行释义[14]。英语对译通常以短语或单词形式呈现,置于汉语释义之后,起到语义锚定和辅助理解的作用。这种设计对于初级水平学习者或英语母语者而言具有明显的认知支持价值。然而,部分学者在研究中也指出,《JUZI 汉语》在部分词目的处理上存在释义语言不一致的现象,这一不足在一定程度上削弱了词典“用目的语理解目的语”的编纂初衷。

综上所述,《JUZI 汉语》在释义语言种类上形成了以“汉语释义 + 英语对译”为核心的双语模式。该模式兼顾了语义理解的准确性与学习者的认知可及性,体现了外向型词典“用户中心”的编纂理念。然而,其在多语种覆盖、释义语言一致性以及非英语母语者适配性等方面仍存在一定不足,这也为后续融媒词典的优化升级提供了明确方向。

3.2.2. 大语言模型的释义语言种类

与《JUZI 汉语》采用相对固定的“中文释义 + 英文对译”模式不同,大语言模型在释义语言种类上表现出显著的动态性、混合性与用户适配性。冯志伟、张灯柯(2024)指出这种差异根植于大语言模型的技术架构:其生成过程并非基于预设的双语词表,而是通过对海量多语言语料的统计学习,形成了一种“语境敏感”的语言选择机制[15]。

从技术潜能上看, DeepSeek 与豆包均具备多语言输出能力, 能够根据训练语料中的语言覆盖范围, 生成包括英语、日语、韩语、法语、德语等多种语言的释义文本。而大语言模型相较于传统词典的优势之一, 在于其能够根据用户的提示语(prompt)或输入语言, 动态调整输出语言的种类与风格。这种“用户适配”能力是大语言模型释义语言灵活性的集中体现。

DeepSeek 对用户指令的响应较为敏感。当用户在提问中明确要求“请用英语解释”或“Please explain in English”时, DeepSeek 能够切换为纯英文输出; 当用户使用日语、韩语等语言提问时, DeepSeek 也倾向于用相同语言回复, 但其小语种输出的准确性和自然度有限。豆包在动态适配方面表现更为积极。它不仅能够响应显式的语言指令, 还表现出一定的“输入语言跟随”能力: 当用户用英文提问时, 豆包更倾向于用英文回复; 当用户用中文提问时, 则以中文为主、英文为辅。蔡薇(2025)指出, DeepSeek 等工具可以通过精心设计的提示语实现语言输出控制, 但不同模型对提示语的敏感度存在差异。豆包的这种适配能力使其在人机交互中更有优势, 降低了用户的学习成本[16]。

然而, 动态适配能力也带来了新的问题: 如同一方言词在不同交互语境下可能获得截然不同的语言呈现, 这对追求稳定、权威知识输入的二语学习者而言, 可能造成认知困惑。此外, 目前两者均未提供显式的“语言偏好设置”功能, 用户只能通过反复调整提示语来间接控制输出语言, 这增加了使用和查询的技术门槛。

3.3. 释义语言难度对比

“释义用词”是解释词目时所用的词语, 为了使第一语言为非汉语的学习者能理解词语的释义, 一般释义用词应当比较平白、通俗。

本文分别把选定的十个方言词(俺、云吞、心水、撮火、摆平、歇菜、忽悠、发小、巴巴结结、弄堂)的融媒词典与大语言模型释义文本输入到“汉语阅读分级指难针”中进行查询, 对它们的方言词释义用词进行考查。查询结果如下:《JUZI 汉语》的选定方言词释义词汇中处于 HSK 六级及六级水平以下的词汇占 64.33%, DeepSeek 的选定方言词释义词汇中处于 HSK 六级及六级水平以下的词汇占 78.46%, 豆包的选定方言词释义词汇中处于 HSK 六级及六级水平以下的词汇占 67.11%。由此可见, 三者都较好地大部分释义用词控制在了中级以内, 释义较为简单, 容易理解; 其中用词最为简单的是 DeepSeek。

4. 外向型融媒汉语学习词典方言词释义编写建议

前文对《JUZI 汉语》、DeepSeek 与豆包三者方言词释义的对比分析表明, 大语言模型在释义语言难度控制、义频统计及语境构建方面具有显著优势, 但在释义准确性、系统性与教学适配性上仍存在不足。基于上述发现, 本章从提升系统性、提升释义效能、增强教学适配性三个维度提出具体建议。

4.1. 利用大语言模型提升方言词释义系统性

释义的系统性要求同一语义场内词语的释义模式保持一致, 义项划分与排序合乎逻辑。然而, 传统编纂中多编者协作易导致释义行文不一致, 而方言词因其地域性与文化特殊性, 这一问题更为突出。陈贤德等(2025)的研究表明, DeepSeek 能够通过语义场聚类分析自动识别相关词群, 并为编者提供结构化的对比参数[7]。据此, 建议编纂团队利用大语言模型的语义聚类与模板生成能力, 对方言词进行语义场归属分析, 提取共性语义特征与个性差异, 并据此制定统一的释义模板。在此基础上, 将释义元词表作为约束条件, 要求大语言模型按照模板批量生成初稿释义, 再由编者进行一致性校验与修订。此外, 在义项排序方面, 可借助大语言模型对大规模语料的义频统计功能, 获取各义项的真实使用频率, 以此作为义项排序的主要依据, 实现“高频优先”的教学原则。这一方法可有效弥补传统编纂中编者语感差

异及汉语义频语料库缺失的局限。

4.2. 利用大语言模型提升方言词的释义效能

释义效能涵盖可理解性、准确性及信息丰富性等多个方面。前文对比发现,大语言模型在释义语言难度控制上优于《JUZI 汉语》,但在准确性上仍有折损。为扬长避短,建议采取以下措施:首先,将外向型汉语学习词典释义元词表,嵌入大语言模型的生成过程,要求其对已有释义进行“降级改写”,将超纲词替换为元词表中的初级词汇,同时将复杂句式拆解为短句。其次,对于具有显著地域文化特征的方言词,应利用大语言模型的多模态生成能力,同步创建图像、音频或短视频等视觉模态资源,实现文字释义与多模态呈现的协同。章宜华(2024)指出,大语言模型为多模态词典释义提供了技术支撑,通过图文匹配可显著提升学习者的理解效率。需要注意的是,大语言模型生成的释义初稿必须经过专业编者的准确性校验,不可因其“浅显易懂”而直接采信[9]。

4.3. 利用大语言模型增强方言词释义的教学适配性

外向型词典的最终目标是服务教学,而方言词的教学适配性体现在其能够根据不同学习者的水平提供差异化的释义与例证。梁霞(2025)指出,DeepSeek 能够依据《国际中文教育中文水平等级标准》生成指定难度的内容[16]。据此,建议编纂团队利用大语言模型的分级生成能力,为同一方言词配置初、中、高三个难度版本的释义与例句。具体操作上,编者先编写一个供中级学习者使用的标准释义,然后要求大语言模型分别向上和向下改写,并补充文化背景与近义词辨析。三个版本一并存入词典数据库,用户可根据自身 HSK 等级选择查看。此外,蔡薇(2025)强调,有效的语言学习需要有意义的社会互动[16]。方言词的使用具有强烈的场景依赖性,建议利用大语言模型的情景生成能力,为方言词创设虚拟交际场景,并以对话示例或短剧形式呈现。进一步,可开发词典内置的对话机器人功能,模拟方言区的真实交际场景,引导学习者在人机互动中运用目标词汇。通过这些措施,词典将从静态的知识查询工具转型为动态的教学辅助平台。

5. 完善大语言模型辅助二语教学与词典编纂的发展建议

5.1. 加强词典编纂垂域知识库建设与应用

通用大语言模型擅长泛化任务,但在完成词典编纂这类专业性任务时,常常输出看似合理而实际不够准确的结果,或是输出结果无法直接达到词典编纂的标准要求。建议从以下三个方面推进垂域知识库建设:第一,明确知识库的内容构成,至少应包括分级词表(如《等级标准》词汇表及各层级释义元词表)、用法型式库(词语的典型搭配与句法框架)、释义模板库(不同词类、不同语义场的标准化释义句式)以及方言词文化注释库(地域分布、使用人群、文化背景)。第二,依托现有词典编纂成果进行数字化转化与结构化处理,将《JUZI 汉语》《商务馆学汉语词典》等权威学习词典的释义、例证、用法说明等数据进行清洗、标注和格式化存储。第三,采用检索增强生成架构,将垂域知识库作为外部知识源,大语言模型在生成释义或例证时首先检索相关知识,然后将其融入生成过程。长远来看,可探索开发词典编纂专用大语言模型,在通用模型基础上使用垂域知识库进行持续预训练或微调,最大程度优化辅助编纂效果。

5.2. 发挥大规模语料库与大语言模型的协同作用

学习者所需的典型语言知识要求词典编者对汉语词语规律性复现的特征和模式进行提取与描写。经过海量语料训练的大语言模型有着较强的语言规律挖掘能力,能够高效识别词汇的语义、句法、语用等特征,但在数据溯源性、准确性以及词语用法频率的准确计量分析方面较为薄弱,而这些恰恰是大规模

语料库的优势所在。因此,大规模语料库与大语言模型之间并非迭代关系,而是互补关系。

建议在词典编纂流程中建立明确的分工协作机制:语料库负责提供真实、可追溯的语言实例,以及基于大规模统计的频次信息(如搭配强度、义项分布比例);大语言模型则负责从语料库提供的实例中抽象出规律性模式,生成初稿释义和例证,并模拟不同语境下的用法。同时,利用大语言模型的语义理解能力对语料库检索进行智能增强,例如将关键词检索扩展为语义相关检索,提高语料采集的全面性。此外,应建立人机协同的校验与反馈闭环:编者使用语料库检索结果校验大语言模型生成的释义,将修改后的最终版本作为参考数据反馈给系统,用于后续模型的优化训练。这种协同路径可兼顾词典编纂的效率与精度。

5.3. 运用大语言模型提升汉语语料库性能

相较于国外语料库及其分析工具,国内汉语语料库及其分析工具在频次统计、搭配查询和对比分析等功能上显示出明显不足,并未很好地满足学习词典编纂的需要。大语言模型具有动态语义分析、上下文感知和智能对比等能力,将其与传统语料库工具相融合,有望提升汉语语料库的性能。

建议从以下两个方向推进:第一,利用大语言模型的多义词判别能力,对语料库中的词语进行自动义项标注。具体方法为:对语料库中每个目标词及其上下文窗口提交给大语言模型,要求其判别当前义项,经过人工抽检验证后,将标注结果回写至语料库。这可大幅扩充带义项标注的语料规模,为义频统计提供数据基础。第二,利用大语言模型进行跨语体、跨时段语料的对比分析。方言词的用法往往随语体和时代变迁而发生变化,大语言模型可辅助研究者快速对比不同语体或不同年代语料中同一方言词的分布特征,为词典修订中义项的新增与合并提供参考依据。通过这些措施,汉语语料库的性能将得到系统性提升,进而更好地服务于学习词典编纂。

6. 结语

本研究以《JUZI 汉语》为融媒词典代表、以 DeepSeek 和豆包为大语言模型代表,对三者方言词释义进行对比分析。但本研究仅选取上述三个平台作为分析对象,所选方言词样本数量有限,因此研究结论主要反映这三个平台在选定样本词范围内的处理情况,其普遍适用性有待后续通过扩大词典类型与模型数量、增加样本词量加以验证。基于上述有限样本的对比,本文尝试初步揭示融媒词典与大语言模型在方言词释义处理上的异同,以为学习者在网络环境下选择方言词查询途径提供参考,也为融媒词典的后续优化及大语言模型在教育领域的应用探索提供实证参考。

参考文献

- [1] 吴沁茗. 金坛方言词汇专题研究[D]: [硕士学位论文]. 喀什: 喀什大学, 2020.
- [2] 许洁红. 浅析粤方言词“揼”的本字[J]. 汉字文化, 2025(21): 96-98.
- [3] 付娜. 外向型汉语学习词典配例中搭配信息的呈现原则及实现条件[J]. 辞书研究, 2010(5): 23-30.
- [4] 杨玉玲, 段彤彤. 外向型汉语学习词典配例难度控制原则及编写建议——以《当代汉语学习词典》为例[J]. 国际中文教育(中英文), 2023, 8(3): 66-75.
- [5] 李禄兴. 突出外向型特点编写实用的对外汉语学习词典——第二届对外汉语学习词典学国际研讨会综述[J]. 辞书研究, 2006(3): 173-176.
- [6] 杨金华. 论外向型汉语学习词典编纂的四项基本原则[J]. 辞书研究, 2016(1): 45-51+93-94.
- [7] 陈贤德, 潘宇华, 杨玉玲. 大语言模型辅助外向型汉语学习词典编纂的研究——以 DeepSeek 为例[J]. 云南师范大学学报(对外汉语教学与研究版), 2025, 23(2): 11-20.
- [8] 鲁健骥, 吕文华. 编写对外汉语单语学习词典的尝试与思考——《商务馆学汉语词典》编后[J]. 世界汉语教学, 2006(1): 59-69+3.

- [9] 杨玉玲. 融媒词典《JUZI 汉语》的研发理念与实现路径[J]. 语言战略研究, 2024, 9(3): 61-72.
- [10] 陈贤德, 陈松林, 段彤彤, 等. 基于眼动实验的英语学习者词典图文释义阅读中的图文整合研究[J]. 外语导刊, 2024, 47(4): 101-110+160.
- [11] 蔡永强. 外向型汉语学习词典的释义用词[J]. 辞书研究, 2018(4): 58-68.
- [12] 章宜华. 内向型普通词典与外向型学习词典的对比研究——对外汉语词典及其蓝本词典的问题探讨[J]. 广东外语外贸大学学报, 2010, 21(5): 5-9.
- [13] 章宜华. 融媒框架下多模态词典及其释义的理论模式研究[J]. 外语教学与研究, 2024, 56(4): 519-531+639.
- [14] 杨玉玲. 融媒时代外向型汉语学习词典编纂理念与实践[J]. 首都师范大学学报(社会科学版), 2022(2): 167-177.
- [15] 冯志伟, 张灯柯. 语言模型与人工智能[J]. 外语研究, 2024, 41(1): 1-19+112.
- [16] 蔡薇, 储诚志, 崔希亮, 等. DeepSeek 赋能国际中文教育创新与发展[J]. 世界汉语教学, 2025, 39(3): 291-315.