

AI 赋能区域语言资源开发与社会影响研究——以英语语言发展为例

付鹏

三亚学院，海南三亚，572000；

摘要：本文探讨人工智能技术(AI)在区域语言的资源开发，AI 技术可以为区域语言资源的开发通过跨语言知识图谱、多模态数据采集、智能翻译生态构建等模式，重塑“语言保护-发展-传播”的语言生态的发展模式，用以保护文化的多样性以及社会影响。通过剖析英语语言发展史下产生的自适应翻译系统的开发上，结合 AI 赋能濒危语言保护的实践案例，阐述人工智能技术驱动的全球化进程，论证 AI 技术赋能发展下如何重塑语言生态，并更加有效的平衡语言发展标准化同时保护文化的多样性。

关键词：区域语言资源；濒危语言保护；生态构建；文化多样性

DOI:10.69979/3041-0673.25.05.009

引言

联合国教科文组织(UNESCO)2024 年发布的数据中，显示全球约 43%的濒危语言面临消亡危机，然而，人工智能技术的出现为语言发展的保护提供新的路径。^[1]例如：澳大利亚利用语音识别技术记录本土原住民语言、中国学者通过智能机器翻译系统，重新构建少数民族语言数据库，维护区域语言生态。本文以英语语言发展作为典型案例，呈现技术革新从《牛津英语辞典》到全球化的 AI 英语词典，如何推动语言标准化与全球化，通过自适应翻译系统开发，构建混合模型构建智能翻译生态，为当前濒危语言保护提供重要的启示和 AI 赋能下区域语言研究提供社会价值。

1 人工智能赋能区域国别学发展

人工智能正在利用各项技术对区域国别研究提供帮助，进包括数据的提取、翻译和整合，这大大降低了各项研究的语言门槛，实现准确、实时、多维、高效的分析模式，AI 赋能区域国别研究领域提供有效的技术支持。^[2]例如，人工智能利用网络爬虫和自然语言处理(NLP)技术，能够实时抓取并整合多语种新闻、社交媒体、学术文献和政府报告等信息，构建动态数据库，并通过数据清洗与标注，提取关键信息，结合政策变化和舆情热点，为区域研究提供实时支持。

AI 的普及是否会取代区域国别学者？对于这个命题，本文的立场是否定的。^[3]AI 不仅不会取代，反而会为区域国别研究的发展提供更加有力的支持，因为该项研究不仅要具备“与域外之人打交道”的独特优势，同时也是跨文化沟通的使者强有力的工具。AI 的信息采集、

语言翻译、资源整合提供了一定的技术支持，对于特殊的群体，特别是区域研究领域，AI 无法获取小众且私密的信息，也无法替代人类进行富有情感温度的跨文化沟通和交流。

2 AI 技术在语言资源开发中的应用

本文阐述人工智能技术在语言资源的开发中的具体应用模式，通过四种不同的应用模式，分析区域语言资源的开发和方向。

2.1 跨语言知识图谱

跨语言知识图谱是将不同语言的知识内容和资源，通过跨语言信息检索、翻译技术和文本识别等软件技术，整合在一个统一的多语言知识体系内。构建知识图谱可以为翻译系统和软件提供比较丰富的语义信息，帮助其更好地理解文本的上下文和含义。例如，当机器翻译涉及到特定领域的术语时，跨语言知识图谱可以提供相关的专业知识和文化背景信息，提高机器翻译的准确性。

夸文化知识图谱在众多领域发挥着不可或缺的作用，在 AI 系统翻译下利用机器学习和深度学习的方法，实现跨语言的对齐与融合。在智能推荐系统中，跨语言知识图谱提供重要的语言应用价值，它能够提供较为准确的输出和语义的理解，保护濒危语言资源的同时发展区域文化多样性。特别是目前在 Deepseek 软件的运用下，结合跨语言知识图谱，开发语言资源的同时，保留语言翻译过程中的文化和区域属性，在不同语言之间的形成更加准确的信息检索和贴合文化输出形式的语言加工，更加多样化的做到区域文化的传播和跨文化的有

效沟通。

2.2 多模态数据采集系统

多模态数据采集系统,除了处理文本信息,结合音频、视频、图片、文字等多模态信息下进行机器翻译。例如,在图书馆和旅游景点的介绍中,就可以同时做到音频、视频的输出生、多语言的文字切换等,帮助用户更加形象的理解翻译的内容。同时,多模态数据采集系统中,对全球事件数据库的技术进行解析,通过 NLP 技术实时抓取 110 种语言的媒体数据,实时监控近全球 65,000+个新闻源,构建动态语言图谱。^[3]相较传统的田野调查方式,效率提升 80%,同时该系统还能识别隐性语言变迁规律。通过四维动态语言图谱构建技术,词汇层、语法层、语用层、社会网络层总共四层,逐层分析和处理,监控语言传播的准确性和实用性。

以英语语言在太平洋岛国的变体监测为例, GDELT 系统捕捉到皮钦英语 (Tok Pisin) 的三种比较大的隐性变迁:第一种是词汇创新机制的变迁,比方说传统的借词像 "lotu" 这个词的使用率有所下降,再比如技术复合词像 "wokabout-selphone" 这个词的使用率年增长 217%。第二种是语法趋同现象的变迁,比方说双重否定结构, "Me no see nothing" 这个句式被标准英语句式替代的速度提升了 3.2 倍。第三种是社会功能转移的变迁,目前已经从家庭交流主导向商业合同等正式场景转变,特别是法律文本中的出现率突破了近 40%。这些隐形变迁的发现,为制定 "AI 辅助语言复兴计划" 提供了有利的依据,包括皮钦英语 (Tok Pisin) 的标准英语双向神经机器翻译系统的开发,为基于地域特征的文化适配算法提供帮助。

2.3 智能翻译生态构建

智能翻译生态的构建是诸多翻译软件开发下非常重要的一个环节,对于区域语言资源的开发而言,良好的翻译系统有助于保护和传承濒危语言。将这些语言所承载的文化、知识、历史等内容更广泛地传播出去,促进区域文化的多样性发展。翻译软件的构建中,对数据资源的开发和语料库建设都起着重要的作用。智能翻译生态构建通过收集源语言和目标语言相互对应的文本数据以实现生态构建保护。例如,在构建英汉互译的系统时,需要收集多样的英语和中文的新闻报道、小说、期刊、学术论文等资源,同时还可以利用网络资源库、专业的文献资料库、翻译记忆库等多种渠道来扩充语料库。

智能翻译生态构建可以融合跨语言知识图谱、多模

态信息和个性化的翻译来满足区域语言资源的发展,根据用户的语言习惯和使用场景等多方因素,为用户提供一种个性化的翻译服务。同时,与人工智能技术深度融合,生成一种处理翻译系统的语言和自然语言的技术,加上逻辑推理的深度融合,实现高水平的语言交互环境。

3 AI 赋能区域语言资源开发-以英语语言发展为例

3.1 从《牛津英语辞典》到全球化的 AI 英语词典

从第一版的《牛津英语辞典》看英语语言发展的历史,这部词典的编纂从 1884 年到 2024 年,历史已有近七十年每年的更新才确立了比较完整的应用拼写规范。^[4]然而 AI 的智能性可以自动追踪全球各类媒体信息,实现每年大量的新词收入,编纂效率大幅提升,高达两百倍之多。其标准化进程加速发展,覆盖面也大幅提升,当前人工智能下的 MOOC 平台,英语教学覆盖面多达 190 个国家,内容还涵盖各类方言英语的地方变体词汇。其显示了英语语言发展下最基础的词汇发展也深受 AI 的冲击,需要在人工智能的赋能下得到更好的发展。

3.2 自适应翻译系统开发

AI 赋能语言资源开发的技术路径,实现多样化的区域语言技术资源。^[5]例如自适应翻译系统的开发和当下流行的社交媒体的传播方式,在英语语言发展中,翻译软件的开发模式在 AI 赋能的构建下,形成较好的生态构建的同时,需要开发自适应的软件开发模式。例如:基于 Transformer 架构开发 "Māori Translate Pro" 软件,这是一款集成文化特异性语料,也就是本土的部落语言,此软件在实现英语和毛利语的双向翻译的 BLEU (双语评估替换指标) 值达 38.7,超越传统统计模型 26% (项目技术白皮书, 2023), 目前很中国的部分教育机构通过此软件来开发 AR 双语教材,用以提升学生的学习效率。

3.3 混合模型构建智能翻译生态

目前有很多翻译软件可以做到比较好的翻译生态构建,下面列举三类不同的算法模型。^[6]1. 统计机器翻译 (SMT) 是一种基于统计模型,通过大了的平行语言资源进行统计分析,计算出位于源语言和目标翻译语言之间的大概率算法。其缺点就是翻译上下文时理解能力受限制,对比较复杂的语法结构处理受限。2. 神经网络机器翻译 (NMT) 是一种利用神经网络模型,直接学习源语言到目标语言的一种映射关系算法,其功能能更好地处理上下文信息,在翻译的质量上较有优势,成为

目前比较主流的算法模型,其缺点就是处理速度比较受限。3. 结合统计机器翻译(SMT)和神经网络机器翻译(NMT)结合起来的混合模型,它可以充分利用SMT的统计模型和NMT的深度学习能力,来提高机器翻译的性能和输出效果。在英语语言发展中,能够做到混合模型构建,在实现翻译的准确性和接近源语言的处理上起着非常重要的作用。

4 AI 技术赋能区域语言发展的社会影响分析

4.1 社交媒体资源对区域语言资源的影响

社交媒体通过AI驱动下的自适应翻译系统和多模态翻译工具,打破了地理位置对语言传播的限制,AI方言语言库和大数据的算法,可以将方言通过社交媒体软件的多语言字幕功能技术实现跨地域的传播。^[7]例如:TikTok毛利语挑战赛,通过运用AI语音合成技术,生成方言模块,此视频累计播放超2亿次,吸引近十万青少年参与语言创作,大数据算法推荐机制,可以使趋于边缘化的语言,得到语言资源的生态保护,打破英语语言霸权的困境。

4.2 语言生态系统的多维影响

Facebook的语言多样性计划在NLP技术下已经收录了全球四百多种濒危语料。同时,越来越多的软件兼容多语言和小众语言的操作系统,甚至出现了混合语体的发展,区域性语言的发展获得了数字生存空间。但是,大数据算法下的语言传播出现代际传播断层现象,使得青少年过度依赖翻译工具,对青少年的主动学习动机有所削弱,但始终在维护语言生态,呈现多维共振的影响。

4.3 AI 赋能濒危语言保护的社会价值

AI智能软件在语言发展中,使得传统语言与现代的语言产生联接,通过语言数据库的构建,为不通过国家和区域的语言发展产生交互性的作用,例如混元大模型的交互式学习系统,可以让学习者通过语言对话,智能聊天沉浸式的进行创作和创新。运用AI适配文化,保护区域文化多样性,已实现保护濒危语言保护的目的。

4.4 AI 技术赋能语言资源的未来发展方向

AI技术正在重塑“语言保护-发展-传播”的语言生态全链条。英语发展史的影响下,AI技术的应用需平衡标准化和多样性共同发展。在未来研究中,应关注AI伦理,例如语言数据主权和技术适老化,针对语言传承人高龄化的问题,开发自适应软件,用以实现濒危语言扩展计划。同时利用跨学科教育平台进行整合和联动,

建立全球语言与文化数据库,允许人文历史、语言发展、计算机科学和AI智能操作技能等多学科的研究者协同发展,构建开放型的智慧语言生态系统。

5 结论

综上所述,助力AI时代下的区域语言资源开发,促进不同区域之间的文化交流,推进语言学的发展,保护好区域语言的生态环境,进一步加强探索AI在此技术应用开发中的监管,保障数据安全性,提高区域语言的开发价值和社会场景影响力。本文从探讨人工智能赋能区域国别学发展中,研究AI技术在语言资源开发中的应用,从跨语言知识图谱、多模态数据采集系统、智能翻译生态构建三方面阐述区域语言的智能开发。再以英语语言发展为例,从《牛津英语辞典》到全球化的AI英语词典进行探索,结合自适应翻译系统开发和混合模型构建智能翻译生态,提出更加比较具体的开发路径。最后,从AI技术赋能区域语言发展的社会影响进行分析,探讨社交媒体资源对区域语言资源的影响和语言生态系统的多维影响,最终从AI赋能濒危语言保护的社会价值衍生到AI技术赋能语言资源的未来发展方向。为构建多元且包容的语言文化环境做出贡献。

参考文献

- [1]Brunnbauer U.Why we study Southeastern Europe Area Studies in a Globalized World[J].Universität Regensburg Press,2024:15-28.
- [2]杨永芳.智能AI赋能本科职业教育英语教学的研究[J].中国多媒体与网络教学学报,2024(9):14.
- [3]郭英剑.外语学科的国别与区域研究:概念、内涵、定位与内容[J].上海海洋大学学报,2024(12):45-58.
- [4]王华树.GenAI时代大语言模型技术发展及其在翻译实践中的应用[C].天津大学春晖大讲堂,2024
- [5]Crystal D.English as a Global Language[M].Cambridge University Press,2023:112-135.
- [6]钟智翔,王戎.论外语学科的国别与区域研究方向及其人才培养[J].国别和区域研究,2020(4):168-188.
- [7]张宗海.西方主要国家的高校学生责任教育与启示[J].高教探索,2002,(3):37.

作者信息:付鹏,女,1984年1月出生,汉族,籍贯河北,悉尼科技大学商科硕士,现任三亚学院教师,研究方向为国别与区域研究。