

跨境电商丝绸实务领域多模态语料库建设与应用研究

茹盈盈

浙江机电职业技术学院, 浙江 杭州 310053

DOI: 10.61369/ETR.12310

摘 要： 随着互联网技术的快速发展，在“一带一路”国家战略和数字经济的推动下，跨境电商网络平台的迅速发展对语言人才培养提出了新的要求。建设多模态语料库，可进一步提升对特定领域相关语汇的认知，从而使学习者提升语言应用能力和自主学习能力，同时为基于机器学习的言语活动研究提供数据材料。与跨境电商相关的语料库中，丝绸文化相关语料库具有一定的代表性，其原因在于丝绸作为中国突出的文化符号，具有巨大的出口量，并衍生出了完善的产业链。本研究基于与丝绸企业的合作，探索语料库在教学和商业领域的应用前景。

关 键 词： 丝绸文化；多模态语料库；应用研究

Research on the Construction and Application of Multimodal Corpus in the Field of Cross-border E-commerce Silk Practice

Ru Yingying

Zhejiang Institute of Mechanical and Electrical Technology, Hangzhou, Zhejiang 310053

Abstract： With the rapid development of Internet technology and driven by the national "Belt and Road" strategy and the digital economy, the rapid growth of cross-border e-commerce platforms has raised new requirements for language talent cultivation. The construction of a multimodal corpus can further enhance the cognition of relevant vocabulary in specific fields, thereby improving learners' language application and self-learning abilities. Additionally, it provides data materials for research on speech activities based on machine learning. Among corpora related to cross-border e-commerce, the silk culture-related corpus is representative. This is because silk, as a prominent Chinese cultural symbol, has a huge export volume and has spawned a well-developed industrial chain. Based on cooperation with silk enterprises, this study explores the application prospects of the corpus in teaching and business fields.

Keywords： silk culture; multimodal corpus; applied research

引言

语料库作为“大量的言语活动以及以文字形式记录的语言的抄录集合”（McEnery et.al, 2011），其内容形式经过发展已经得到了极大的扩充，目前“多模态语料库”作为语料库的“4.0”版本，其内容包含文字、音频及视频等多种材料的集成，对其进行的研究与分析在深入剖析如何进行跨境电商运营模式优化，并完善跨境电商相关网站建设方面，具有一定指导意义，同时，大量语言材料的集成为相关领域的语言学习者提供了集中化，针对化的学习材料，在高职教育的应用中体现了其重要价值。跨境电商实务类别众多，其中丝绸产业国内外市场广阔，素有“丝绸之府”之称的浙江，更是中国丝绸的重点出口基地，迫切需要大量拥有较强语言能力的跨境电商双语人才，本研究以多模态理论与语料库语言学为指导，探讨多模态丝绸实务领域语料库的研制方法与路径，以便跨境电商平台提供基础资源，助力跨境电商英语教学。

一、丝绸跨境电商与多模态语料库

目前，国内外学者越来越多地关注到多模态语料库的研究价值。在理论层面：多模态语料库语言学与其上位范畴语料库语言

学一样；在研究范式层面：分为语料库驱动研究和基于语料库的应用研究，主要包括三个方面的研究：一是多模态语料库与语言研究及人文社科研究。二是多模态语料库与教学研究。三是多模态语料库与商业实用研究。

课题1：2023SCG297 跨境电商丝绸实务领域多模态汉英平行语料库建设与应用研究；

课题2：《教育数字化转型战略下职业本科院校教师数字胜任力模型构建及提升路径研究》，中国职业技术教育学会2024年度分支机构科研课题重点项目，课题编号：SZ24A32。

国内多模态语料库建设虽方兴未艾,但关于跨境电商实务方面的多模态汉英平行语料库建设甚少。一方面,国内已建成的跨境电商语料库中多模态语料库十分少见。高媛和郭立甫指出我国跨境电子商务英文网页设置应用中存在的问题,文中列举了词汇、语法、句法、语言习惯等差异问题,但均未将其语料升级至多模态层面,也未做出进一步实证性应用研究。另一方面,多模态语料库在建设过程中尚存诸多问题,相关标注体系较为简单,语料库的类型较为单一,多模态语料数据挖掘的深度和广度尚需增强。

二、跨境电商丝绸实务领域多模态汉英平行语料库建设

多模态丝绸语料库的建设和研究目的就能突破传统文本传播叙事的局限性,形成多维度对外传播叙事空间,更大程度地还原非遗“活态化”传播的直观度、真实度和全面性,为非遗对外传播研究探索新思路、开启新视角提供有价值的参考。根据 Bowker Jennifer(2002),LSP (Language for Special Purpose) 的运用在促进特定领域的交流与沟通方面具有重要作用。同时,语料库的应用有助于 LSP 的习得,因此,该项目语料库的构建以电子商务作为目标领域,根据跨境电商语言服务内容从相关企业收集与整理实务案例语料,包括丝绸文化、丝绸科技、消费特征、品牌宣传、产品介绍、价格咨询、渠道沟通、售后服务等等,作为一级分类,以适应不同场景的分析需要。搜集包括包含跨境电子商务领域的术语 (terminology), 固定表达结构 (Collocations), 以及特殊语法结构 (Style&Grammar), 并基于此对数据进行分类整理,作为该语料库的二级分类。最后,根据语料的分层属性,以语言的表层结构,深层结构,媒介,情景作为三级分类,以便于更为精细的研究。

(一) 语料选取

该语料库的构建,旨在选取具有代表性的,使用频率高,且来源清晰,形式多样的语料。同时,由于跨境电商是一个系统,连续性强的产业结构,选取上下文相对完整清晰的语料作为分析对象更为合理。值得注意的是,由于语料的来源有限,该语料库可能出现部分术语及语言模式缺失,无法进行可靠性强的泛用化 (generalization) 等问题,在运用时需谨慎对待该影响。

(二) 语料加工

构建语料库的流程主要包括语料筛选、处理、多层次标注和最终的数据库构建,旨在创建一个组织良好、易于检索的语料库,以便于后续的研究和分析。具体如图 1 所示。

1. 语料数据筛选

为了确保语料选材的真实性,根据跨境电商语言服务内容从相关企业收集与整理实务案例语料,包括丝绸文化、丝绸科技、消费特征、品牌宣传、产品介绍、价格咨询、渠道沟通、售后服务等等各种密切相关的多模态语料。

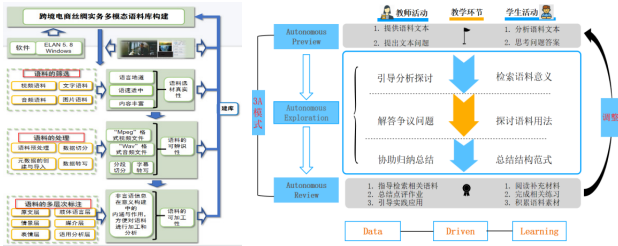
2. 语料数据预处理

语料的预处理包括视频软件的格式转换、元数据创建与导

入、语料的切分和转写。

3. 语料多层次标注

ELAN 软件能够对包含多种媒介的文件进行多层次标注,用户可以根据自己的研究需求定义不同的标注层级和方案。这种多层次标注有助于明确揭示非言语信息在构建意义时的内涵和作用,使得语料的处理和分析更加精确和高效。通过这种方式,研究人员能够深入理解语言与其他模态如何交互作用,从而丰富了语料的研究价值。



>图 1 跨境电商丝绸实务多模态语料库构建

>图 2 基于语料库的数据驱动跨境电商英语教学“3A”模式

三、跨境电商丝绸实务多模态汉英平行语料库应用

(一) 教学应用

本文提出了一套基于语料库的数据驱动跨境电商英语教学“3A”模式设计。如图 2 所示。该模式包括课前自主预习 (Autonomous Preview)、课中自主探索 (AutoExploration)、课后自主巩固 (Autonomous Review) 三个阶段 (如图 2)。这种教学模式涵盖了课前自主预习、课中自主探讨和课后自主巩固三个阶段,旨在促进学生的自主学习。

1. 课前自主预习

在具体实施时,可基于 Bowker 和 Jennifer(2002)提出的分类法 (Term; Collocation; Style; Grammar) 作为每课时的学习方向,引导学生进行语料选取。例如,在“丝绸科技”的主题背景下(一级分类),分别从 Term, Collocation, Grammar&Style 中选取高频语料作为教学材料。以 Term 为例,假定选出与丝绸工艺有关的术语“weaving”,学生应自主利用语料库查找其语境 (context) 并整理其常见用法。

2. 课中自主讨论

课中自主探讨是一种学生中心的教学法,它通过语料库技术让学生在教师的引导下主动分析文本、解读意义和用法,并进行互动讨论。在此过程中,教师的角色是促进者,负责引导学生观察、分析并解决讨论中的争议问题,同时对讨论内容进行补充和深化。最终,教师帮助学生进行归纳性总结,以巩固学习成果,这种方法旨在提高学生的参与度、批判性思维和自主学习能力。多模态的引入使更多的副语言信息加入到语料库中,尤其是在基于 CMC (Computer Mediated Communication) 时,副语言信息的产生尤为庞大,如何对这些信息进行解读,并应用于教学中,也是一大难点。根据 Scott 和 Tribble,语境 (context) 具有不同的范围 (Contextual Scope),即指语境具有延展性,其范围最大可延伸到文化层面。基于该语料库数据源自与企业合作的事实,该语

料库拥有更加完整的语境构建,学生不仅需关注语料的表层结构,更需理解其在特定商务情境下的应用。教师应深入研究以语料为中心的不同情景下的语境范畴变化,并强化学生的语境意识。同时教师应引导学生更加注重语料学习与实践工作的结合,以语料为中心归纳类似工作环境中的常用语汇,以及高效的沟通方式。

3. 课后自主回顾

课后自主练习巩固是教学模式的一个重要环节,学生在课后通过软件工具复习和加深对课堂内容的理解,同时进行自我监督和评价,以提升学习参与度。教师则根据学生的练习结果和反馈,对教学计划和方法进行调整,以优化教学效率和成果。

(二) 跨境电商平台应用

对于小型个体跨境电商而言,借助大型企业的语料库能快速实现产品及宣传的本地化翻译,快速理解目标市场的用语习惯,提高沟通效率。以企业对“真丝羊毛双面披肩——玲珑花影”(Double Silk Wool Shawl-Exquisite Flower Shadow)的宣传文案为例,该文案中对真丝缎面(silk satin),对比色调(contrast tones),质地(texture)等丝绸领域的常用语汇的翻译提供了可靠参考。以成熟语料库为基准的翻译模式能使文案工作者将更多的精力转移到“意译”色彩较重以及受文化背景影响较大的文字上,如“玲珑”“生生不息”等源于古汉语较难直译的用语。值得注意的是,跨境电商中的翻译不仅是语言的转换,还涉及到文化的适应。通过语料库,企业可以更好地理解不同市场中的文化差异,并据此调整翻译策略,确保宣传文本在目标市场中的有效性。某些具有文化背景的表达或符号可能在不同市场中具有不同的意义或影响。通过语料库,翻译人员可以识别这些差异,并做出适当的调整,确保文本的文化适应性。

语料库还可以与机器翻译系统集成,提供更高效的自动化翻译服务。通过结合语料库中的数据和机器学习算法,企业可以实现高效、准确的自动化翻译。该技术能大大降低客服人员的招聘门槛,令业务能力优秀而语言能力有所欠缺的从业人员更好地发挥自身能力。针对实践中出现的翻译漏洞,则由语料库的开发团队进行处理分析并不断完善语料库建设,由于副语言信息和情感

色彩的关联紧密,结合副语言信息在上述不同语料类别中的密度分布和情感趋向,同样能使该语料库在商业方面发挥一定价值。情感分析还可以用于个性化营销策略的实施。通过了解不同消费群体情感倾向和需求,企业可以设计个性化的营销方案,提高客户的参与度和忠诚度。基于对不同客户群体的情感分析,企业可以为不同的客户推送定制化的产品推荐或服务,优化客户体验。

基于科学,合理的分类系统下的语料库能较为便捷地做到对特定产品的反馈进行定位和数据分析,通过与客户对话记录的追踪分析,明确市场动向。数据的导出可利用 Wechat, Exporter 等开源工具进行操作,对于构建自身的实时通信平台的企业而言,利用 Webscokt 建立连接进行实时数据获取将更加便利。

四、总结

在“一带一路”背景下,本研究探讨了跨境电商语料库建设及其在高等职业教育中的应用的可能性。随着跨境电商的快速发展,高校面临培养具有双语技能的高素质人才的挑战。本研究提出构建跨境电商丝绸实务领域多模态汉英平行语料库,旨在提升学习者对 LSP 的认知,增强语言应用能力和自主学习能力,并初步探讨其在商业领域可能的应用前景。

在教育应用方面,研究建议将语料库融入教学,通过自主学习和教师引导相结合的方式,帮助学生掌握跨境电商相关的语言技能。特别是,语料库中的丰富语境信息可以让学生理解如何在实际工作环境中应用所学知识。

在跨境电商领域,语料库的应用不仅限于教学,还包括情感分析与营销策略调整。例如,通过分析客户反馈和对话数据,企业可以更好地把握市场动向。此外,语料库还能支持小型电商企业进行本地化翻译,提升与国际市场的沟通效率。

总体而言,本文研究了语料库在跨境电商和职业教育中的双重作用,展示了其在提升教育质量和企业竞争力方面的巨大潜力。研究以丝绸文化为切入点,通过与实际企业的合作,探索了语料库在商业和教育领域的广泛应用前景。

参考文献

- [1] 刘剑. 基于多模态语料库的口译教学模式研究 [J]. 外语电化教学, 2017(02).
- [2] 刘剑. 国外多模态语料库建设及相关研究述评 [J]. 外语教学, 2017(07): 40-44.
- [3] 许家金. 多因素语境共选: 语料库语言学新进展 [J]. 外语与外语教学, 2020, (03): 1-10+21+146.
- [4] 李威. 语料库语言学在英语教育中的应用 [J]. 中国教育学报, 2020, (S1): 81-82.
- [5] 苏妮娜. 语料库语言学与英语教学方法研究——评《语料库语言学与外语教学》[J]. 人民长江, 2020, 51(08): 221.
- [6] 黄立鹤. 多模态范式与后疫情时代的外语教学 [J]. 当代外语研究, 2021(01).
- [7] 赵小晶. 语料库语言学运用于商务英语课堂教学可行性研究——评《语料库辅助的商务英语短语教学研究》[J]. 外语电化教学, 2021, (03): 124.
- [8] 胡富茂, 宋江文, 王文静. 多模态旅游翻译语料库建设与应用研究 [J]. 上海翻译, 2022, (05): 26-31.
- [9] 王大鹏. 基于语料库的大学英语移动学习探究——评《语料库语言学: 研究实践》[J]. 外语电化教学, 2022, (05): 104.
- [10] 张丽霞. 基于自建语料库的商务英语翻译课程教学研究 [J]. 江西电力职业技术学院学报, 2023, 36(09): 139-141.