基于语料库的浙江民俗文化英语外宣话语策略的研究

闫 玉*，田 颖

绍兴文理学院 浙江绍兴

【摘要】浙江民俗文化源远流长，代表了浙江的文化符号。然而，目前研究缺乏基于语料库对浙江民俗文化英语外宣话语的研究。浙江民俗文化英语外宣不仅能提升浙江国际影响力、推动本土发展，而且还顺应了文化全球化的趋势，助力中华文化更好地走向世界。本研究聚焦浙江民俗文化英语外宣，在文化“走出去”及相关国家倡议背景下，借语料库技术剖析其话语特征，以期梳理浙江民俗文化外宣特征，提高外宣质量。研究发现：（1）词汇层面，整体词汇丰富度中等，不同民俗种类差异显著，民俗建筑词汇丰富度高，民间传说较低；实词分布不均，名词主导而动词匮乏，整体词汇密度有限。（2）句子层面，平均句长 23.41 个单词属中长句，文本风格明了，句法复杂度为 1.54，结构简单，利于信息传递与理解。

【关键词】浙江民俗文化；语料库；话语特征；英语外宣

【收稿日期】2025 年 5 月 10 日 【出刊日期】2025 年 6 月 9 日 【DOI】10.12208/j.ssr.20250225

A corpus-based study on discourse strategies of Zhejiang folk culture in English publicity

Yu Yan, Ying Tian*

Shaoxing University, Shaoxing, Zhejiang

【**Abstract**】Zhejiang folk culture has a long history and represents the cultural symbol of Zhejiang. However, there is a lack of corpus-based research on the discourse of Zhejiang folk culture in English publicity. Zhejiang folk culture in English publicity can not only enhance Zhejiang's international influence and promote local development, but also conform to the trend of cultural globalization and help Chinese culture go global better. This study focuses on Zhejiang folk culture in English publicity. Against the background of cultural "going out" and related national initiatives, this study uses corpus technology to analyze its discourse characteristics in order to sort out the characteristics of Zhejiang folk culture in publicity and improve the quality of publicity. The study found that: (1) At the lexical level, the overall lexical richness is moderate, with significant differences among different folk customs. The lexical richness of folk architecture is high, while that of folk legends is low; the distribution of content words is uneven, with nouns dominating and verbs scarce, and the overall lexical density is limited. (2) At the sentence level, the average sentence length is 23.41 words, which is a medium-long sentence. The text style is clear, the syntactic complexity is 1.54, and the structure is simple, which is conducive to information transmission and understanding.

【**Keywords**】Zhejiang folk culture; Corpus; Discourse characteristics; English propaganda

引言

在当今时代，推动中国文化走向世界，促进世界对中国的了解，是提升文化自信、全力推进社会主义文化强国建设的必由之路。21 世纪以来，随着中国文化“走出去”以及“一带一路”两大国家倡议的相继提出，对外宣传话语策略研究愈发重要。习总书记[1]（2022，p18）在党的二十大报告中指出：“加快构建中国话语和中国叙事体系，讲好中国故事、传播好中国声音，展现可信、可爱、可敬的中国形象。”由此可见，民俗文化“走出去”成为中国发展的必然道路。

浙江历史悠久，是河姆渡文化和良渚文化的发源地，其民俗文化资源丰富，并且在国家级非物质文化遗产（国遗）项目中连续多年位居全国第一。各种文化相互交融，展现出鲜明的地域特色和深厚的文化底蕴。可见，浙江民俗文化具有很强的地域性、民间性和多样性等特点。正如陈华文[2]（2008）所言，其囊括了不同社

*通讯作者：闫玉

会层次的不同文化形态，如戏曲、社戏、茶文化、民间艺术、木版年画、饮食、服饰等，品种众多、独具韵味。

在数字化和跨学科研究趋势的推动下，语料库技术为民俗文化外宣的策略研究提供了新的方法论支撑。目前国内对于浙江民俗文化的研究相对较少，且尚有借助语料库技术对此进行研究。基于此，本论文主要通过搭建语料库对浙江民俗外宣话语特征进行研究。通过对语料库的词语丰富度、词长、词汇密度等方面进行研究，探析浙江民俗文化外宣话语规律，提高外宣质量。

1 文献综述

1.1 民俗文化外宣话语研究

在文化全球化的发展及在相关国家倡议的背景下，民俗文化外宣是“文化走出去”进程中不可或缺的内容。民俗文化是各族人民在长期生产生活的实践中形成的集体生活文化，是群体内大多数成员都认同的思维方式、精神情感、价值取向和道德规范的总和，其蕴含的价值潜在性、实践性、情感性等特点。[3]国内众多学者将民俗文化外宣话语的研究置于翻译学视域下，例如，徐志伟[4]（2015）以朝阳地方民俗文化研究为基础，从归化和异化、中式外语及多元翻译策略的综合运用等方面分析了地方民俗文化外宣的话语翻译策略；冯欢&潘春静[5]（2018）提出了“二元共存”外宣翻译策略，灵活运用“归化”和“异化”，实现“归化”和“异化”二元共存。其次，不同领域学者从各自专业视角对民俗文化外宣话语策略进行了多维度研究，扩大了民俗文化外宣话语内容。例如，在传播学领域，赵佳佳[6]（2024）从文化交流、全民参与度、社交媒体传播方面，分析外宣话语演变并提出新时代转型策略。在文化研究领域，逯凌虹[7]（2023）从宏观、中观、微观三个层面提出生态翻译学视阈下构建齐文化外宣话语体系的策略，以期提升齐文化外宣传播力。

1.2 基于语料库的民俗文化研究

语料库语言学作为以实证数据驱动的研究范式，通过大规模真实语料的采集、标注与量化分析，为揭示语言使用规律提供了科学路径。Bowker 和 Pearson[8]（2002）强调语料库的真实性和动态可更新性，胡开宝[9]（2011）进一步界定其为依据一定抽样方法收集的自然出现语料构成的电子数据库，是按照研究目的和语料选择方法选择并有序排列的语言运用材料的汇集。

语料库语言学理论对于民俗文化的研究发挥着独特价值，该理论可用于剖析民俗文化语言文本，挖掘其中的文化内涵和语言特色，辅助民俗文化的传承与对外传播，如分析民俗文化外宣文本的话语特征，优化传播策略。它突破了传统语言学依赖直觉或少量例句的研究模式，强调实证数据的重要性，致力于从宏观和微观层面揭示语言的本质特征。高莹等[10]（2020）分析了梳理近年来国内民俗翻译研究成果，提出建立山东省民俗翻译语料库的必要性。王佳等[11]（2023）通过调查江西非遗民俗翻译现状，发现诸多问题，采用语料库研究方法，提升翻译的效率和精确性。罗飞[12]（2024）基于陕西民俗语料库分析了陕西民俗文化英语外宣文本词语特征和句子特征，发现外宣文本多为中长句和长句，整体语言较为平缓，符合文化类文本的语言特征。

综上所述，基于语料库语言学理论来研究民俗文化外宣已经有一定的积累，这为浙江民俗文化外宣提供了丰富的理论基础与实践经验。但基于语料库对浙江民俗文化研究的内容却鲜有涉及。建立浙江民俗文化语料库，分析其外宣特征，对于促进浙江民俗文化的外宣事务发展，加速浙江民俗文化“走出去”，进一步推动浙江民俗文化的国际传播有着重要意义。

2 方法论、语料库设计和汇编

本文通过自建浙江民俗语料库，运用话语分析的定量分析对浙江民俗文化语料库的词汇、句子特征等方面进行整体分析，以期发现浙江民俗文化外宣的话语规律，总结出浙江民俗文化外宣特征从而提高外宣质量。

在深入探究浙江民俗文化英语外宣相关语篇的进程中，搭建专业语料库是关键一环。以下详述具体搭建方法：

首先，语料收集。秉持多元性原则采集语料，力求涵盖丰富且权威的素材。一方面，深度挖掘浙江当地旅游官方网站、官方公众号、论文文献等所发布的英文文本。另一方面，突破传统文本局限，广泛涉猎 B 站、YouTube、抖音等热门 APP 中的视频资源，从中筛选出以浙江民俗文化外宣为主题的视频内容，为语料注入新活力。

其次，语料清洗。为确保语料质量，制定严苛筛选标准。针对收集到的文本，剔除存在格式错乱、内容残缺不全的部分，同时坚决排除与浙江民俗文化主题关联度低或毫无关联的冗余信息。整个筛选流程通过多轮人工精细审核，逐字逐句甄别，力保录入语料库的数据具备较高质量。

最后，语料整理。将收集好的语料进行录入以 AntConc 分析软件适配的电子文档格式进行保存，保存建立浙江民俗文化语料库。所有文件都存储在名为 ZhejiangFolkCultureCorpus 的文件夹中，依据浙江民俗

文化七个类别命名子文件夹，符合了分类检索及跨类对比分析的要求。

3 浙江民俗文化英语外宣话语特征分析

3.1 词汇特征

（1）词汇丰富度

本文通过运用 AntConc 软件对语料库进行分析，Baker[13]（1995）对词汇的类符/形符比（Type-TokenRatio，TTR）进行定义，类符是指语料库中不同单词的数量，形符则是语料库中所有单词的总数。TTR 值越高，表明词汇丰富度越高。通过统计，浙江民俗外宣的词汇丰富度见表 1。

表 1 说明不同民俗种类在词汇丰富度上存在差异，民俗建筑的词汇丰富度最高，其次是民俗戏剧、民俗舞蹈、民间艺术、民俗饮食，最后是民间传说，其词汇丰富度最低。

总的词汇丰富度为 39.07%，反映出浙江民俗英语外宣文本整体的词汇丰富度处于中等水平，这意味着在整体的外宣资料中，既包含了一定数量的不同词汇来准确传达各种民俗文化信息，但也存在一定的重复性，可能有一些常用词汇在不同的民俗类别描述中被频繁使用。

表 1　浙江民俗英语外宣词汇丰富度

序号	种类	形符数	类符数	类符/形符比（TTR）（%）
1	民俗舞蹈	1236	537	43.45%
2	民俗戏剧	1248	581	46.55%
3	民间艺术	1002	433	43.21%
4	民间传说	3469	1156	33.32%
5	民间手工艺	1896	746	39.35%
6	民俗饮食	2889	1050	36.34%
7	民俗建筑	526	289	54.94%
	总计	12266	4792	39.07%

（2）词汇密度

词汇密度即是一种测量文本中含有的信息量的方法，可以衡量词汇多样性[17]。通过计算文本的词汇密度，可以判断文本是否传达了更多实质信息。其中较高的词汇密度通常意味着文本信息含量更丰富，反之亦然。词汇密度的计算公式为：实词数量/总词数×100%。本文使用 FreeCLAWSWWWtagger 对语料库进行词性编码，接着再使用 AntConc 对编码后的语料库各类实词（名词、动词、形容词、副词等具有实际意义的词）进行统计，见表 2。

表 2　浙江民俗文化外宣语料库的实词分布和总体情况

序号	词类	数量	词汇密度（%）
1	名词	3995	15.36
2	动词	433	1.66
3	形容词	1307	5.03
4	副词	532	2.05
5	总计	6267	24.10

表格 2 中统计的名词、动词、形容词和副词这四类实词的总数为 6267 个，词汇密度为 24.10%。在浙江民俗文化外宣语料中名词总体占比最高，其次是形容词、副词，最后是动词。整体而言，当前的比例在大多数情况下是符合民俗文化外宣的语言表达要求的，说明文本在信息传达上具有一定的实质内容。

3.2 句子特征

外宣话语的句子特征包括平均句长及句法复杂度方面的特征。平均句长能反映文本在句子层面的简洁或复杂程度。基于 Butler[14]（1985）适用于外宣等非文学文本的信息型文本句长分类标准，本研究将句长为 1—9 个单词的句子称为短句，句长为 10—25 个单词的句子称为中长句，句长超过 25 个单词的句子称为长句。句法复杂度用于描述句子在语法结构上的复杂程度。在本文通过 NeoSCA 软件可以测量出平均句长和句法复杂度来研究英语外宣句子特征，从而揭示文本难易程度。浙江民俗文化外宣语料库的平均句长及句法复杂度的情况见表 3。

表 3　浙江民俗文化外宣的平均句长及句法复杂度

句子特征	长度/复杂度
平均句长（MLS）	23.41
句法复杂度（C/S）	1.54

数据结果显示，浙江民俗文化外宣文本的平均句长为 23.41 个单词，没有超过 25 个单词，在中长句的范围之内。这意味着浙江民俗文化外宣的文本风格较为明了，在表述上通俗易懂，传达信息较为直接。句法复杂度为 1.54，表明浙江民俗文化外宣文本中，大多数句子的结构较为简单。句法复杂度越低，文本更易于读者理解和接受，更有助于民俗文化信息的有效传播。

4 结语

从文本数据的细读和分析，可以发现浙江民俗文化英语外宣在话语层面呈现以下特点：

第一，整体上的词汇丰富度处于中等水平，但不同民俗种类在词汇丰富度差异显著。民俗建筑词汇丰富度最高（TTR=54.94%），可能是因为其涉及众多建筑专业术语、独特的建筑风格等方面，需要较多不同词汇来描述；而民间传说词汇丰富度最低（TTR=33.32%），或许是由于情节相对固定、角色类型有限，用词相对集中。

第二，在词汇密度方面，词汇密度为 24.10%，说明文本在信息传达上具有一定的实质内容，符合民俗文化外宣的语言表达要求。但在实词分布中，名词数量众多，反映出民俗文化外宣需大量名词描述实体、概念和现象。而动词数量少，表明对民俗文化动态过程描述不足，更多侧重于静态事物的呈现。

第三，浙江民俗文化外宣文本多为中长句，传达信息直接，没有使用过于冗长复杂的句子符合外宣文本高效传递信息的需求。同时，其句法复杂度数值较低，这意味着大多数句子结构简单，这样的文本更易被读者理解和接受，有效助力民俗文化信息的传播，规避了因句子结构复杂给读者造成理解阻碍的情况。

参考文献

[1] 习近平. 高举中国特色社会主义伟大旗帜 全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告[M]. 北京：人民出版社,2022.17-18 .

[2] 陈华文. 论浙江民俗的演变轨迹及其特点[J]. 民俗研究, 2008, (02): 198-211.

[3] 郑恩 & 刘旭东. (2025). 民俗文化赋能铸牢中华民族共同体意识研究. 湖北民族大学学报(哲学社会科学版), (02), 22-32.

[4] 徐志伟. (2015). 民俗文化的外宣及其翻译策略研究. 长春师范大学学报, 34(03), 84-87.

[5] 冯欢,潘春静. (2018). 黑龙江省民俗文化外宣翻译策略研究——“一带一路”战略背景下. 边疆经济与文化, (01), 12-13.

[6] 赵佳佳. 国际传播背景下外宣话语新趋向研究[J]. 采写编, 2024, (03): 71-73.

[7] 逯凌虹.生态翻译学视阈下齐文化外宣话语体系的构建[J].海外英语, 2023, (19): 50-52.

[8] Bowker, L. & Pearson, J. (2002). Working with specialized language. A practical guide to using corpora. London: Routledge. Retrieved from http://dx.doi. org/10.4324/9780203469255

[9] 胡开宝. 主持人语[J]. 当代外语研究, 2011, (01) : 12.

[10] 高莹, 宋玲枝. 山东省民俗翻译语料库建设的必要性分析[J]. 海外英语, 2020, (06): 86-88.

[11] 王佳, 杜习举. 江西非遗民俗平行语料库的建设意义研究[J]. 吉林广播电视大学学报, 2023, (02): 151-153.

[12] 罗飞. 基于语料库的陕西民俗文化英语外宣话语策略研究[J]. 品位 • 经典, 2024, (02) : 72-75.

[13] BAKER M. Corpora in translation studies: An overview and some suggestions for future research[J]. Target. International Journal of Translation Studies, 1995, 7 (2): 223-243.

[14] Butler,Christopher. Statistics in Linguistics [M] Oxford: Basil Blackwell, 1985.

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