Modern Linguistics 现代语言学, 2024, 12(6), 28-36
Published Online June 2024 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2024.126430

Hans汉斯

原创汉语与翻译汉语中小说语言特征对比研究

姚明玉

浙江工商大学外国语学院，浙江 杭州

收稿日期：2024年4月19日；录用日期：2024年6月4日；发布日期：2024年6月12日

摘　要

“第三语码”用以描述一种既不同于源语，也与目的语母语有所区别的翻译语言。本文基于兰卡斯特现代汉语语料库(LCMC)和浙大翻译汉语语料库(ZCTC)中的小说文本建立子库，运用PowerGREP 5.0、AntConc3.5.9和Python等研究工具对小说文本的语言特征进行对比分析。在词汇、句子和语篇三个层面上，我们进一步细分了研究维度。在词汇层面，通过分析词汇的平均长度，词汇复杂度、词汇密度以及高频词的使用情况；在句子层面，比较分析平均句长；在语篇层面，研究了两种文本在连词的使用情况的差异。研究发现，除高频词外，其他词汇层面、句子层面和语篇层面都存在差异。在以往的研究中，发现原创汉语与翻译汉语的差异可能受到不同文体的影响，但在小说的对比研究中，也可能受到不同小说体裁的影响。

关键词

原创汉语，翻译汉语，语言特征，语料库

A Comparative Study of Linguistic Features in Original Chinese and Translated Chinese Fiction

Mingyu Yao

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Apr. 19th, 2024; accepted: Jun. 4th, 2024; published: Jun. 12th, 2024

Abstract

“Third Code” is used to describe a translated language that is distinct from the source language and also differs from the target language’s native language. This paper establishes a sub-corpus

文章引用: 姚明玉. 原创汉语与翻译汉语中小说语言特征对比研究[J]. 现代语言学, 2024, 12(6): 28-36.
DOI: 10.12677/ml.2024.126430

based on novel texts from the Lancaster Modern Chinese Corpus (LCMC) and the Zhejiang University Chinese Translation Corpus (ZCTC), utilizing research tools such as PowerGREP 5.0, AntConc 3.5.9, and Python to conduct a comparative analysis of the language features of novel texts. We further divide the research dimensions into vocabulary, sentence, and discourse levels. At the vocabulary level, we analyze the average length of words, vocabulary complexity, vocabulary density, and the usage of high-frequency words. At the sentence level, we compare and analyze the average sentence length. At the discourse level, we study the differences in the usage of conjunctions between the two types of texts. The study reveals differences at the vocabulary, sentence, and discourse levels, apart from high-frequency words. Previous research has shown that the differences between original Chinese and translated Chinese may be influenced by different literary styles. However, in the comparative study of novels, they may also be influenced by different novel genres.

Keywords

Original Chinese, Translated Chinese, Linguistic Features, Corpus

1. 引言

Mona Baker 于 1993 年首次论述了语料库语言学与翻译的结合、语料库技术应用于翻译研究的可行性[1]，成为语料库翻译学的开端，许多学者开始基于语料库探索翻译研究的实践。学界基于语料库多探讨翻译共性包括翻译的显化、简化和范化[2] [3]，译者风格对比[4]等相关研究。语料库语言学的工具、技术和方法对大量真实的翻译现象进行描述并从所描述的翻译“自身”的语言特征中寻找到翻译现象所固有的普遍特征[5]，也叫翻译共性。翻译共性是指翻译语言作为一种客观存在的语言变体，相对于源语或目标原创语言从整体上表现出的规律性语言特征被称为“第三语码”，既不同于源语，又不同于母语。Laviosa [6]考察译文与母语在词汇使用上的不同发现译文比母语使用更多的高频词；秦宏武和王克非[7]基于汉英双向对应语料库描写和分析英译汉语言的词汇特征，分别从实词、虚词的使用情况探讨英译汉语言的特征。吴俊峰等[8]使用 14 项绝对复杂度和相对复杂度指标对比翻译汉语和原创汉语的句法复杂度差异。这些研究主要考察了翻译语言与原创语言在词汇或句法单一层面的特征，本研究将从词汇、句子层面探讨原创汉语和翻译汉语的语言特征。这些研究从总体上探究了原创汉语和翻译汉语的差异性，不管是词汇还是句法，没有区分不同话题或者文体的影响作用。

细化方向、题材类型能够较少对数值统计的影响。朱一凡和李鑫[9]探讨了翻译与原创汉语新闻语料库各词类使用的差异情况，发现翻译汉语相对于原创汉语的使用过多和使用不足现象。小说作为一种文学体裁，因其美学意义和主题意义而与一般的翻译在词汇和句子表达层面发生变化。胡显耀[10]探讨了翻译小说的词语特征，基于 CCTFC“当代汉语翻译小说语料库”、LCMC“兰卡斯特现代汉语语料库”以及 LCMC(N) (LCMC 小说体裁部分)三个语料库，研究发现，与同一语言的非翻译文本相比，汉语翻译小说存在倾向于使用更少的词汇、重复使用一定数量的常用词、实词减少等特征。本研究将基于 ZCTC 原创汉语语料库和 LCMC 兰卡斯特翻译汉语语料库，选取 K-P 小说分类建立子库，从词汇、句子和语篇层面展开语言特征的比较分析。

2. 研究方法

2.1. 研究问题

1) LCMC 和 ZCTC 在词汇特征方面是否存在差异？

2) LCMC 和 ZCTC 在句子和语篇层面的语言特征是否存在差异？

2.2. 语料来源

本研究选择兰卡斯特现代汉语语料库(Lancaster Corpus of Mandarin Chinese, LCMC)、浙大汉语译文语料库(ZJU Corpus of Translational Chinese, ZCTC)，来对比翻译英语与对应的非译文汉语的语言特征。LCMC 是英国兰卡斯特大学的 Tony McEnery 和肖忠华在 2003 年建成的原创汉语语料库，约一百万词，其中有 500 篇字数约为 2000 词的文本，文本内容包含 15 种类别，语料为 1989 到 1993 年中国大陆的出版物，语料库采用 ICTCLAS 2008 进行分词和词性标注，分词精度达 98.13%，词性标注精度达 94.63% [11]。ZCTC 是按照 LCMC 的模式建立的，其在文本分类、数量、词数方面基本一致。LCMC 和 ZCTC 可以为研究提供样本量大致均衡的语料库，使对比研究更具可信度。从两个语料库中分别挑选编码为 K-P (一般小说、侦探小说、科幻小说、武侠传奇小说和爱情小说)小说类文体，分别有 117 篇。

2.3. 研究工具

该研究主要使用 PowerGrep5.0、AntConc 3.5.9 和 Python 进行数据统计和分析。

3. 结果

3.1. 词汇层面

在词汇层面的分析主要从平均词长、词汇复杂度，词汇密度和高频词几方面展开。

3.1.1. 平均词长

词长是指语料库中各种长度词的频数，平均词长指文本中词汇的平均长度，可以体现文本的正式程度，具体表现为正式程度越高的文本平均词长的数值越大，而口语化程度越高的文本平均词长越小[12]。根据“表 1”统计结果，我们可以看出 ZCTC 语料库小说题材的平均词长略高于 LCMC 语料库中小说体裁平均词长，翻译汉语的正式化程度较高，原创汉语的口语化程度偏高(Python 部分运行代码见“图 1”)。

```
import osimport pandas as pdfrom lxml import etree
def extract_text_from_xml(file_path):
    """从 XML 文件中提取文本内容"""
    with open(file_path, 'r', encoding='utf-16') as file:
        xml_content = file.read()
        root = etree.XML(xml_content.encode('utf-16'))
        texts = root.xpath('//s/text()')
        return ' '.join(texts)
def calculate_average_word_length(text):
    """计算文本的平均词长"""
    words = text.split()
    total_length = sum(len(word.split('_')[0]) for word in words)  # 只计算单
词长度
    return total_length / len(words) if words else 0
def process_corpus(folder_path):
```

Figure 1. Python code for computing average word length
图 1. Python 计算平均词长部分代码

Table 1. Average word length table of LCMC and ZCTC
表 1. LCMC 和 ZCTC 平均词长表

LCMC	平均词长	1.35	ZCTC	平均词长	1.39
	K 一般小说	1.33		K 一般小说	1.39
	L 侦探小说	1.40		L 侦探小说	1.40
	M 科幻小说	1.43		M 科幻小说	1.37
	N 武侠小说	1.30		N 武侠小说	1.43
	P 爱情小说	1.35		P 爱情小说	1.35

3.1.2. 词汇复杂度

TTR (类符/形符比)在语料库学和文本分析中被广泛应用，是了解文本特征的一个重要参数。类符是指语料中所存在的不同的词汇，型符是指语料的总数量。“类符/形符比(TTR)”表示相同长度的句子中含有的不同的词汇的数量，一般被用来衡量语料的难易程度。比值越大，说明语料中的词汇种类就越多，表达越丰富，词汇的复杂度就越高；比值越低，则说明语料中可能语言模式比较简单，多使用一些重复的词汇。词汇多样性越少，复杂性就越低。根据“表 2”，可以发现 LCMC 小说题材文本的 TTR 数值高于 ZCTC 小说体裁的 TTR 数值。由此可见，原创汉语的词汇表达更加丰富，词型变化多，可见原创汉语的词汇复杂度更高。而在小说文本中，不同的小说类型 TTR 数值也不相同。在 LCMC 语料库中，词汇复杂度整体排名为侦探小说 > 爱情小说 > 武侠小说 > 一般小说 > 科幻小说；在 ZCTC 语料库中，词汇复杂度整体排名为武侠小说 > 侦探小说 > 科幻小说 > 一般小说 > 爱情小说。

STTR (标准化类符/型符比)也是用来反映文本词汇丰富度和信息含量的度量标准，这个比值反映了文本的词汇多样性，即文本中词汇的丰富程度。STTR 能够减弱比文本长度差异的影响，该研究中计算了每 1000 词的类符/型符比，统计结果如表一所示。我们可以发现，忽略文本长度的差异，LCMC 的标准化类符/型符比(52.41)却低于 ZCTC (54.29)，却与之前的研究结果出现不一致。原创汉语语料库中，侦探小说 STTR 最高，而爱情小说 STTR 最低；而翻译汉语语料库中武侠小说 STTR 最高，一般小说 STTR 最低。这可能是小说类型的影响，一些小说类型保留文学特征的原因，用词更加丰富，追求多样化的表达方式，一切类型比较接近生活，表达比较简单(Python 计算词汇复杂度部分运行代码见“图 2”)。

Table 2. Statistics of genre tokens, type tokens, TTR, and STTR of LCMC and ZCTC novels
表 2. LCMC 和 ZCTC 小说体裁类符、型符、TTR、STTR 统计表

语料库	类别	类符	型符	TTR	STTR
LCMC	整体	92485	990460	35.01	52.41
	K 一般小说	20693	60190	34.49	51.26
	L 侦探小说	18036	49268	36.60	54.73
	M 科幻小说	4225	12374	34.18	53.96
	N 武侠小说	20856	60232	34.62	52.85
	P 爱情小说	20806	59686	34.86	50.90
ZCTC	整体	83676	239795	34.90	54.29
	K 一般小说	20607	60540	34.04	50.66

续表

L 侦探小说	17216	48920	35.19	55.38
M 科幻小说	4227	12265	34,46	53.42
N 武侠小说	21666	59039	36.70	57.02
P 爱情小说	19960	59031	33.82	54.47

```
def calculate_sttr(text, segment_length=100):
    """计算文本的 STTR，将文本分割成等长的片段"""
    tokens = text.split()
    if len(tokens) == 0:
        return 0, 0, 0

    segment_sttrs = []
    total_types = set()
    total_tokens = 0

    for i in range(0, len(tokens), segment_length):
        segment = tokens[i:i + segment_length]
        types = set(word.split('_')[0] for word in segment)
        total_types.update(types)
        total_tokens += len(segment)
        segment_sttr = len(types) / len(segment) if segment else 0
        segment_sttrs.append(segment_sttr)
```

Figure 2. Python process for computing genre tokens, type tokens, and STTR
图 2. Python 计算类符、型符和 STTR 部分过程

3.1.3. 词汇密度

词汇密度一般用公式 Lexical Density(LD) = (content words/total token) * 100 来表示，即实义词占型符的比重。Ure [13]认为词汇密度应该是由单位文本中实词数量在文本总词数的比例来确定，实词数量越多，文本的词汇密度就越高。文本密度越高，意味着文本中含的信息量就越高。在统计中把名词、动词、形容词、副词、数、量词六个具有稳定词义的词类称为实义词[14]。本研究所采用的两个语料库均使用 ICTCLAS 2008 标注词性，根据词性赋码，运用 PowerGREP 软件通过运用正则表达式统计实义词的频率和总次数来计算词汇密度，统计结果如下表所示(见“表 3”)。

Table 3. Vocabulary density table of genre in LCMC and ZCTC novels
表 3. LCMC 和 ZCTC 小说体裁词汇密度表

LCMC				ZCTC			
	实词数量	总型符	词汇密度		实词数量	总型符	词汇密度
总	241750	241750	49.5%	总	112061	239795	46.7%
K	29744	60190	49.4%	K	27722	60540	45.8%
L	25219	49268	52.2%	L	23039	48920	47.1%
M	6063	12374	49.0%	M	5764	12265	47.0%
N	30697	60232	51.0%	N	28703	59039	48.6%
P	27836	59686	46.6%	P	26833	59031	45.5%

经过词汇密度统计，可发现 1) LCMC 原创汉语小说体裁中的总体词汇密度(49.5%)要大于 ZCTC 翻译汉语的小说词汇密度(46.7%)，说明原创汉语小说使用更多的实义词，信息量比较大，难度也会稍微增加；2) 通过小说题材中不同主题的小说类别词汇密度对比，不管是一般小说、侦探、科幻、武大还是爱情小说，LCMC 原创汉语的词汇密度都比 ZCTC 翻译汉语中的词汇密度要大。3) 爱情小说(P)在两个语料库中的词汇密度都是最低。

3.1.4. 高频词

本研究分别利用 AntConc 3.5.9 和 PowerGREP 统计了 LCMC 和 ZCTC 中小说文本排名前 20 的高频词(见“表 4”)，如下表所示，我们可以发现前 20 的高频词都是单字词，虽然排序有略微差异，但是基本都是“的”、“着”、“了”、“你”、“我”、“她”等人称代词和助词，以及少量的动词(如“是”、“有”)和介词(如“在”)。通过分析其高频词的占比，我们发现翻译汉语小说中高频词的占比更高，说明翻译汉语小说中高频词重复使用的次数更多。

Table 4. Top 20 high-frequency words in LCMC and ZCTC
表 4. LCMC 和 ZCTC 高频词表(Top 20)

LCMC			ZCTC		
排名	频率	占比%	排名	频率	占比%
1. 的	10,107	4.18	1. 的	12,123	5.06
2. 了	4,693	1.94	2. 他	4,553	1.90
3. 是	3,320	2.37	3. 了	4,388	1.82
4. 一	2,866	1.19	4. 我	3,613	1.51
5. 不	2,706	1.12	5. 是	3,396	1.40
6. 他	2,611	1.08	6. 她	3,278	1.38
7. 在	2,369	0.98	7. 在	3,020	1.26
8. 人	2,277	0.94	8. 一	2,649	1.10
9. 我	2,260	0.93	9. 不	2,439	1.01
10. 说	1,760	0.73	10. 你	2,104	0.88
11. 她	1,735	0.72	11. 着	1,544	0.64
12. 你	1,722	0.71	12. 说	1,533	0.64
13. 着	1,610	0.67	13. 这	1,345	0.56
14. 有	1,433	0.59	14. 人	1,296	0.54
15. 这	1,421	0.59	15. 地	1,283	0.54
16. 就	1,344	0.56	16. 有	1,184	0.49
17. 地	1,203	0.50	17. 就	1,179	0.49
18. 也	1,183	0.49	18. 上	1,143	0.48
19. 上	1,132	0.47	19. 到	1,077	0.46
20. 到	1,008	0.42	20. 那	1,021	0.43
总计	4,8760	21.28	总计	5,4168	22.59

3.2. 句子层面

在句子层面，该研究主要讲两个语料库的平均句长做比较分析。平均句长与文体差异有很大的关系[15]，所以我们很难从整体上把握原创汉语和翻译汉语在平均句长上的差异性。所以本研究通过确定小说这一体裁，探究同一文体中原创汉语和翻译汉语在平均句长上的差异，将其他问题对研究的差异降低。本研究选取 LCMC 和 ZCTC 语料库运用 Python 运行结果如下(见“表 5”) (Python 计算平均句长部分代码见“图 3”)。

肖忠华、戴光荣[14]在之前的研究中计算出 LCMC 和 ZCTC 在 15~20 之间，但本研究确定了文体，发现小说体裁的平均句长高于整个语料库中总体的句长，也验证了平均句长受文体影响的假设。汉语原创小说的平均句长高于翻译小说的平均句长。再从不同的小说类别观察，两个库中武侠小说平均句长最长，爱情小说平均句长最短，在汉语原创小说中，科幻小说平均句长与武侠小说平均句长一同居于最高。从平均句长的研究结果来看，平均句长可能受到小说体裁、小说类型、难度的影响。

Table 5. Top 20 high-frequency words in LCMC and ZCTC
表 5. 平均句长表

LCMC	平均长度	21	ZCTC	平均长度	18
	K 一般小说	19		K 一般小说	18
	L 侦探小说	21		L 侦探小说	18
	M 科幻小说	24		M 科幻小说	17
	N 武侠小说	24		N 武侠小说	20
	P 爱情小说	20		P 爱情小说	16

```
# 计算
average_sentence_length_by_label =
df.groupby('label')['average_sentence_length'].mean()
average_sentence_length_overall = pd.Series({
    'ZCTC': df[df['label'].str.startswith('ZCTC')]['average_sentence_length'].mean(),
    'LCMC': df[df['label'].str.startswith('LCMC')]['average_sentence_length'].mean()
}, name='Average Sentence Length')
# 保存到 Excelwith pd.ExcelWriter('sentence_length_results.xlsx') as writer:
    df.to_excel(writer, sheet_name='Sentence Length Data', index=False)
    average_sentence_length_by_label.to_excel(writer, sheet_name='Average
Sentence Length by Label')
    average_sentence_length_overall.to_excel(writer, sheet_name='Overall
Average Sentence Length')
```

Figure 3. Process of calculating average sentence length
图 3. 平均句长计算过程

3.3. 语篇层面

表达各种逻辑关系的连接成分可看作语言形式化和显化程度的标志之一[16]。已经有学者基于语料库研究出翻译汉语文本比原创汉语文本中使用更多的连词[6] [14]，支撑了翻译显化共性的假说。但是连词的使用和文体也有一定的关系，不同文体之间连词使用可能存在差异，本研究在语篇层面主要探究了原创汉语与翻译汉语小说中连接词使用的差异。

经统计发现(见“图 4”)，在小说题材翻译汉语连词使用了 5133 次，原创汉语连词使用 3839 次，ZCTC 小说文体连词使用高于 LCMC 小说中连词使用，这也和之前学者研究的结果一致，这可能也证明了小说

体裁中翻译连词的显化。该研究统计了各种类型小说连词平均使用情况(见“图 5”)，可以观察到，除了在科幻小说中，原创汉语的连词使用高于翻译汉语，而其他几种类别中，翻译汉语的连词使用都高于原创汉语(Python 计算连词数量部分代码见“图 6”)。

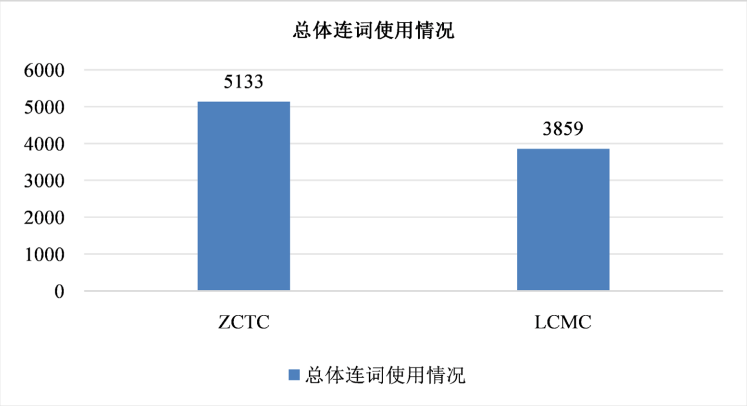

Figure 4. Usage of conjunctions in genre of ZCTC and LCMC novels
图 4. ZCTC 和 LCMC 小说题材中连词使用情况图

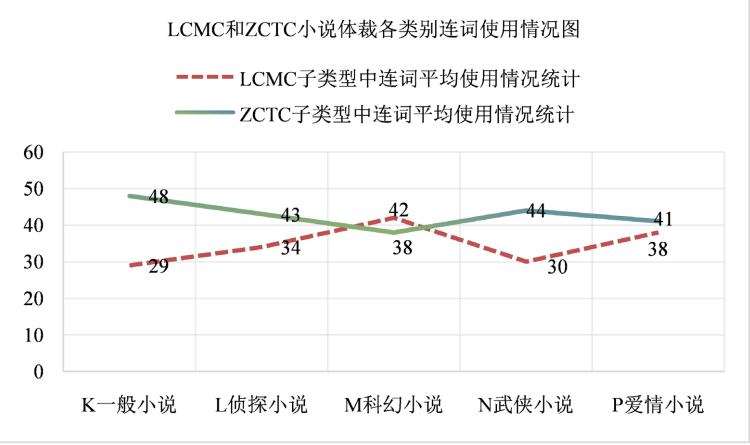

Figure 5. Average usage of conjunctions in different genre of LCMC and ZCTC novels
图 5. LCMC 和 ZCTC 不同小说类型中连词平均使用情况图

```
import osimport pandas as pdfrom lxml import etree
def extract_text_from_xml(file_path):
    """从 XML 文件中提取文本内容"""
    with open(file_path, 'r', encoding='utf-16') as file:
        xml_content = file.read()
        root = etree.XML(xml_content.encode('utf-16'))
        texts = root.xpath('//s/text()')
        return ' '.join(texts)
def count_conjunctions(text):
    """计算文本中连接词的数量"""
    tokens = text.split()
    return sum(1 for word in tokens if word.split('_')[-1] in ['c', 'cc'])
```

Figure 6. Python process for calculating conjunctions
图 6. Python 计算连词的部分过程

4. 结论

通过对原创汉语与翻译汉语小说文体从词汇、句子和语篇层面的统计分析，结果发现，单从小说来看原创汉语与翻译汉语的语言特征可能与基于整体语料库考察的结果并不一致，这可以说明其语言特征受到文体的影响。在词汇层面，该研究从平均词长、词汇复杂度、词汇密度、和高频词使用来分析：1) 翻译汉语小说平均词长 > 原创汉语小说的平均词长，说明翻译汉语小说的正式化程度更高；2) LCMC 小说题材文本的 TTR 数值高于 ZCTC 小说体裁的 TTR 数值，原创汉语小说词汇更加丰富、多样，词汇复杂度高；3) 但经过标准化的数值(STTR)结果出现不一致，LCMC 的标准化类符/型符比(52.41)却低于 ZCTC (54.29)；4) LCMC 原创汉语小说体裁中的总体词汇密度(49.5%)要大于 ZCTC 翻译汉语的小说词汇密度(46.7%)，原创汉语小说的实词使用相对较多，信息量较大；5) 通过分析其高频词的占比，我们发现翻译汉语小说中高频词的占比更高，说明翻译汉语小说中高频词重复使用的次数更多。但主要都是助词、人称代词，动词主要有“是”、“有”、“到”等，量词“一”等。在句子层面上，研究发现：1) 汉语原创小说的平均句长高于翻译小说的平均句长；2) 武侠小说在两个库中平均句长最长，爱情小说平均句长最短。在语篇层面上主要考察的连词在原创汉语小说和翻译汉语小说的使用情况，发现：1) 总体上看，ZCTC 小说文体连词使用高于 LCMC 小说中连词使用。这也和之前学者研究的结果一致，这可能也证明了小说体裁中翻译连词的显化。2) 只有在科幻小说中，原创汉语的连词使用高于翻译汉语。总的来说，原创汉语和翻译汉语的比较在同一文体中进行，可比性才能提高，其语言特征更聚焦。

参考文献

[1] Baker, M. (1993) Corpus Linguistics and Translation Studies: Implications and Applications. In Baker, M., Francis, G., and Tognini-Bonelli, E., Eds., *Text and Technology*: *In Honour of John Sinclair*. John Benjamins, Amsterdam, 233-250. https://doi.org/10.1075/z.64.15bak

[2] 柯飞. 翻译中的隐和显[J]. 外语教学与研究, 2005, 37(4): 303-307.

[3] 胡加圣, 郭鸿杰, 戚亚娟. 翻译共性之范化假设的短语学考察[J]. 中国翻译, 2021, 42(4): 141-149.

[4] 刘泽权, 刘超朋, 朱虹.《红楼梦》四个英译本的译者风格初探——基于语料库的统计与分析[J]. 中国翻译, 2011, 32(1): 60-64.

[5] 胡显耀. 用语料库研究翻译普遍性[J]. 解放军外国语学院学报, 2005, 28(3): 45-48.

[6] Laviosa, S. (1998) Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta*, **43**, 557-570.

[7] 秦洪武, 王克非. 基于对应语料库的英译汉语言特征分析[J]. 外语教学与研究: 外国语文双月刊, 2009, 41(2): 131-136.

[8] 吴继峰, 刘康龙, 胡韧奋, 等. 翻译汉语和原创汉语句法复杂度对比研究[J]. 外语教学与研究, 2023, 55(2): 264-275+320-321.

[9] 朱一凡, 李鑫. 对翻译汉语语言特征的量化分析——基于翻译与原创汉语新闻语料库的对比研究[J]. 中国外语, 2019, 16(2): 81-90.

[10] 胡显耀. 基于语料库的汉语翻译小说词语特征研究[J]. 外语教学与研究, 2007, 39(3): 214-220.

[11] McEnery, A. and Xiao, Z. (2004) The Lancaster Corpus of Mandarin Chinese: A Corpus for Monolingual and Contrastive Language Study. *Religion*, **17**, 3-4.

[12] 张旭冉, 杏永乐, 张盼, 等. 《道德经》四个英译本的翻译风格对比研究——基于语料库的统计与分析[J]. 上海翻译, 2022(3): 33-38.

[13] Ure, J. (1971) Lexical Density and Register Differentiation. *Contemporary Educational Psychology*, **5**, 96-104.

[14] 王克非, 胡显耀. 基于语料库的翻译汉语词汇特征研究[J]. 中国翻译, 2008, 29(6): 16-21.

[15] 肖忠华, 戴光荣. 寻求“第三语码”——基于汉语译文语料库的翻译共性研究[J]. 外语教学与研究, 2010, 42(1): 52-58+81.

[16] 董敏, 冯德正. 英汉科技翻译逻辑关系显化策略的语料库研究[J]. 外语教学, 2015, 36(2): 93-96.