

# 翻译汉语和原创汉语句法复杂度 对比研究\*

首都师范大学 吴继峰 香港理工大学 刘康龙  
北京师范大学 胡韧奋 首都师范大学 周 蔚

**提要:**本文基于 LCMC 和 ZCTC 两个可比语料库,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。研究发现:1)翻译汉语在 8 个指标上显著高于原创汉语,在 4 个指标上显著低于原创汉语;2)在 8 个指标上文本类型与体裁类型有显著的交互作用,在 7 个指标上翻译汉语和原创汉语不同体裁的句法复杂度有不同表现。研究表明翻译汉语同时存在“简化”和“繁化”的翻译共性。本文采用对比分析、认知翻译学的“引力假说”和心理语言学的“翻译的水平模型”对此研究结果进行了解释。

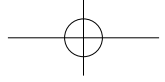
**关键词:**翻译汉语、原创汉语、句法复杂度、对比

[ 中图分类号 ] H315.9 [ 文献标识码 ] A [ 文章编号 ] 1000-0429 ( 2023 ) 02-0264-12

## 1. 引言

翻译汉语和原创汉语的语言特征对比研究是近年来国内翻译研究的热门话题。相关研究成果对于揭示翻译汉语的语言特征及验证各种翻译理论如翻译共性 ( Baker 1993, 1996 )、引力假说 ( Halverson 2003 )、翻译的水平模型 ( Maier, Pickering & Hartsuiker 2017 ) 等具有重要价值 ( Liu, Liu & Lei 2022; 王克非、胡显耀 2008 )。以往语料库研究发现翻译共性主要包括简化、显化、规范化和匀质化四类 ( 王克非、胡显耀 2008; 胡显耀 2021 )。近年来翻译共性研究又出现了新的发展趋势,聚焦在译入语独特项减少、语言干扰和异常搭配上 ( 胡显耀 2021 )。

\* 本文为教育部中外语言交流合作中心国际中文教育中外联合研究专项课题“基于新标准的智能化语言分析技术研究” ( 22YH04ZW ) 和北京市社科基金青年项目“面向汉语国际教育的学术汉语资源库建设研究” ( 21YYC011 ) 的阶段性成果。衷心感谢《外语教学与研究》编辑部、匿名审稿专家和金檀教授的宝贵意见。



国内对翻译汉语的研究在宏观和微观两个层面,尤其是在词汇和具体语法点层面,都取得了丰硕成果(徐佐浩、蒋跃 2021),但对翻译汉语句法复杂度的挖掘还不够,缺少多层面成体系的句法复杂度测量。

## 2. 汉语书面语句法复杂度研究述评

句法复杂度是指语言产出中句法结构的多样度、精细度和复杂程度(Ortega 2003; Lu & Wu 2022)。根据汉语的句法特点,学界开发出了一系列专门测量汉语句法复杂度的指标,从话题链指标(Jin 2007)到复杂名词短语(Lu & Wu 2022; 吴继峰、陆小飞 2021)和基于搭配的句法复杂度指标(Hu, Wu & Lu 2022; 胡韧奋 2021)。基于以上研究,吴继峰、胡韧奋和陆小飞(2022)首次建立了汉语句法复杂度测量体系,包括句子、搭配、短语和语法点层面的 60 项指标。这些指标可分为两大类型:一是绝对复杂度,其测量基于句法结构,计算句法结构的数量、长度、多样度、密度等;二是相对复杂度,其测量基于用法和统计,计算频率值和互信息值。

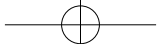
目前对翻译汉语句法复杂度的研究较少,常用的测量指标仅限于平均句长,无法全面反映翻译汉语句法复杂度的特征。而汉语句法复杂度指标的发展为文本研究提供了更方便的工具。秦洪武、孔蕾(2018)提出了计算汉语结构复杂度的方法,对词数、内嵌小句、从属小句、总小句数和中心语修饰数量这五个变量赋值进行计算。Liu, Liu & Lei (2022)使用词性熵值对比了翻译汉语和原创汉语的句法复杂度,发现翻译汉语的句法复杂度高于原创汉语,并据此认为翻译汉语的句法结构更多样和复杂。然而,此两项研究仍有待更深入的探讨。第一项研究主要关注结构复杂度,但对句法复杂度还可从多个角度衡量;第二项研究仅采用了词性熵值作为指标,该指标表明翻译汉语的句法结构更加多样,但不能证明其句法结构更加复杂。

综上所述,目前翻译汉语句法复杂度的研究还比较薄弱,存在较大研究空间。本研究基于已建立的汉语句法复杂度测量体系,利用汉语文本自动分析工具,考察翻译汉语与原创汉语的句法复杂度差异。

## 3. 研究设计

### 3.1 研究问题及实验设计

本文旨在考察文本类型和体裁类型对句法复杂度的影响,具体包含以下两个研究问题:1)从总体上看,翻译汉语和原创汉语的句法复杂度是否存在差异?2)在不同体裁中,翻译汉语和原创汉语的句法复杂度是否存在显著差异?本研究采用  $2(\text{文本类型:翻译汉语、原创汉语}) \times 4(\text{体裁类型:新闻类、通用类、学术$



类、小说类)两因素完全随机实验设计,因变量是 14 个句法复杂度指标测量值。

3.2 语料来源

原创汉语语料来自兰卡斯特现代汉语语料库(Lancaster Corpus of Mandarin Chinese, LCMC),该语料库由 T. McEnery 和 肖 忠 华 于 2003 年 建 立,共 1,023,387 词,包含四大类 15 小类体裁,共 500 份样本(具体分布见表 1)<sup>1</sup>,每份样本约 2000 词。翻译汉语语料来自浙江大学汉语译文语料库(ZJU Corpus of Translational Chinese, ZCTC),该语料库完全按照 LCMC 模式等比例建立,共 1,017,395 词,同样包含四大类 15 小类体裁,500 份样本,每份样本约 2000 词,且各体裁样本的分布比例也完全一致。二者构成可比语料库。

表 1. LCMC 和 ZCTC 两个语料库的体裁大类分布

| 体裁大类 | 新闻类   | 通用类   | 学术类 | 小说类   | 共计   |
|------|-------|-------|-----|-------|------|
| 样本数  | 88    | 176   | 110 | 126   | 500  |
| 比例   | 17.6% | 35.2% | 22% | 25.2% | 100% |

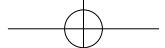
3.3 测量工具

本研究使用句子、短语和搭配三个层面的 14 个指标对原创和翻译汉语的句法复杂度进行测量(见表 2)。

表 2. 汉语句法复杂度测量指标

| 复杂度类型 | 测量层面 | 测量指标           |
|-------|------|----------------|
| 绝对复杂度 | 句子层面 | 1. 平均大句长       |
|       |      | 2. 平均小句长       |
|       |      | 3. 平均小句数       |
|       | 短语层面 | 4. 名词短语平均长度    |
|       |      | 5. 介词短语平均长度    |
|       |      | 6. 简单子句平均并列短语数 |
|       |      | 7. 简单子句平均名词短语数 |
|       |      | 8. 简单子句平均动词短语数 |
|       |      | 9. 句均介词短语数     |
| 相对复杂度 | 搭配层面 | 10. 整体搭配多样性    |
|       |      | 11. 通用搭配多样性    |
|       |      | 12. 特殊搭配多样性    |
|       |      | 13. 特殊搭配比例     |
|       | 搭配层面 | 14. 低频搭配比例     |

1 由于有的体裁小类的样本数不足 10 篇或 20 篇,为保证统计效力,本研究只考察四种体裁大类句法复杂度的差异。



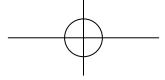
测量工具包括北京师范大学团队开发的“L2C-Rater”(Wang & Hu 2021)及德国图宾根大学与北京语言大学团队合作开发的“汉语文本复杂度分析平台——中文CTAP”(Cui *et al.* 2022)。14个指标来自以上两个自动测量工具,可以涵盖多个维度、多种测量方式并体现句法复杂度多维立体的内涵。其中,整体搭配包含跨语言通用搭配(动宾、主谓、形名、状中搭配)和汉语特色搭配(量名、框式介词、介动、述补搭配)。低频搭配的界定基于胡韧奋、肖航(2019)构建的大规模中文搭配库中搭配的频次和互信息。具体操作定义和计算方法可参见胡韧奋(2021)或Hu, Wu & Lu(2022)。

#### 4. 结果

原创汉语和翻译汉语14项句法复杂度指标的测量结果见表3。

表3. 原创汉语和翻译汉语句法复杂度测量结果

| 测量指标        | 文本类型 | 体裁类型  |      |       |      |       |      |       |      |
|-------------|------|-------|------|-------|------|-------|------|-------|------|
|             |      | 新闻类   |      | 通用类   |      | 学术类   |      | 小说类   |      |
|             |      | 平均数   | 标准差  | 平均数   | 标准差  | 平均数   | 标准差  | 平均数   | 标准差  |
| 平均大句长       | 原创   | 35.78 | 6.61 | 31.31 | 6.47 | 40.93 | 8.57 | 25.58 | 7.05 |
|             | 翻译   | 36.78 | 5.79 | 32.07 | 8.61 | 38.82 | 6.84 | 22.30 | 6.46 |
| 平均小句长       | 原创   | 12.39 | 1.82 | 10.70 | 1.55 | 14.62 | 2.29 | 9.29  | 1.46 |
|             | 翻译   | 15.99 | 1.92 | 13.36 | 3.19 | 16.54 | 2.92 | 9.84  | 1.54 |
| 平均小句数       | 原创   | 2.89  | 0.40 | 2.93  | 0.48 | 2.83  | 0.58 | 2.76  | 0.67 |
|             | 翻译   | 2.31  | 0.33 | 2.40  | 0.36 | 2.37  | 0.35 | 2.23  | 0.43 |
| 名词短语平均长度    | 原创   | 4.55  | 0.52 | 3.98  | 0.61 | 5.50  | 0.87 | 3.11  | 0.49 |
|             | 翻译   | 4.91  | 0.57 | 4.32  | 0.97 | 5.64  | 0.76 | 3.22  | 0.56 |
| 介词短语平均长度    | 原创   | 8.96  | 1.72 | 7.65  | 1.63 | 10.45 | 2.53 | 6.26  | 1.38 |
|             | 翻译   | 10.26 | 1.95 | 9.55  | 3.26 | 12.82 | 2.36 | 6.80  | 1.19 |
| 简单子句平均并列短语数 | 原创   | 0.16  | 0.09 | 0.12  | 0.08 | 0.31  | 0.17 | 0.05  | 0.03 |
|             | 翻译   | 0.19  | 0.09 | 0.17  | 0.12 | 0.36  | 0.25 | 0.06  | 0.04 |
| 简单子句平均名词短语数 | 原创   | 4.26  | 0.77 | 3.53  | 0.65 | 4.92  | 1.07 | 2.99  | 0.49 |
|             | 翻译   | 4.95  | 0.77 | 4.25  | 1.19 | 5.70  | 1.70 | 3.24  | 0.61 |
| 简单子句平均动词短语数 | 原创   | 3.27  | 0.32 | 3.25  | 0.40 | 3.52  | 0.54 | 3.09  | 0.36 |
|             | 翻译   | 3.41  | 0.44 | 3.43  | 0.58 | 3.80  | 0.85 | 3.03  | 0.36 |

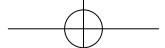


(续表)

| 测量指标    | 文本类型 | 体裁类型  |      |       |      |       |      |       |      |
|---------|------|-------|------|-------|------|-------|------|-------|------|
|         |      | 新闻类   |      | 通用类   |      | 学术类   |      | 小说类   |      |
|         |      | 平均数   | 标准差  | 平均数   | 标准差  | 平均数   | 标准差  | 平均数   | 标准差  |
| 句均介词短语数 | 原创   | 0.90  | 0.26 | 0.84  | 0.29 | 1.18  | 0.39 | 0.55  | 0.22 |
|         | 翻译   | 1.03  | 0.23 | 0.92  | 0.35 | 1.32  | 0.76 | 0.57  | 0.27 |
| 整体搭配多样性 | 原创   | 20.51 | 1.24 | 20.65 | 1.50 | 18.60 | 1.94 | 21.21 | 1.03 |
|         | 翻译   | 20.48 | 1.71 | 20.50 | 1.47 | 18.24 | 2.81 | 20.75 | 1.19 |
| 特殊搭配多样性 | 原创   | 7.18  | 1.41 | 7.53  | 1.24 | 5.64  | 0.94 | 8.60  | 0.90 |
|         | 翻译   | 6.89  | 0.95 | 7.38  | 1.20 | 5.93  | 1.06 | 8.72  | 0.86 |
| 通用搭配多样性 | 原创   | 18.89 | 1.10 | 18.90 | 1.37 | 17.52 | 1.83 | 19.00 | 0.99 |
|         | 翻译   | 19.00 | 1.68 | 18.81 | 1.46 | 17.01 | 2.69 | 18.43 | 1.20 |
| 特殊搭配比例  | 原创   | 0.18  | 0.05 | 0.19  | 0.05 | 0.15  | 0.03 | 0.22  | 0.04 |
|         | 翻译   | 0.17  | 0.04 | 0.20  | 0.05 | 0.17  | 0.04 | 0.24  | 0.04 |
| 低频搭配比例  | 原创   | 0.17  | 0.03 | 0.14  | 0.03 | 0.17  | 0.04 | 0.13  | 0.02 |
|         | 翻译   | 0.16  | 0.02 | 0.14  | 0.02 | 0.15  | 0.03 | 0.14  | 0.02 |

为考察文本类型和体裁类型的主效应以及二者之间的交互效应,我们采用双因素方差分析对测量结果进行检验,结果见表 4。从表 4 可知,在文本类型的主效应上,除了平均大句长和特殊搭配多样性不显著外,其他 12 个测量指标均显著。其中,在平均小句长、名词短语平均长度、介词短语平均长度、简单子句平均并列短语数、简单子句平均名词短语数、简单子句平均动词短语数、句均介词短语数、特殊搭配比例 8 个指标上,翻译汉语显著高于原创汉语。在平均小句数、整体搭配多样性、通用搭配多样性、低频搭配比例 4 个指标上,翻译汉语显著低于原创汉语。

在体裁类型的主效应上,14 个测量指标均显著,说明体裁类型之间存在显著差异。为进一步检验哪些体裁之间存在显著差异,我们进行了 Bonferroni 事后多重比较,结果显示:1)在平均大句长、平均小句长、名词短语平均长度、介词短语平均长度、简单子句平均名词短语数、特殊搭配多样性、特殊搭配比例 7 个指标上,四种体裁之间差异均显著;2)在简单子句平均并列短语数、简单子句平均动词短语数、句均介词短语数、整体搭配多样性 4 个指标上,除新闻类与通用类差异不显著外,各大类体裁两两之间差异均显著;3)在通用搭配多样性指标上,新闻类与学术类、通用类与学术类、学术类与小说类之间差异均显著;



4)在低频搭配比例指标上,除新闻类与学术类差异不显著外,各大类体裁两两之间差异均显著;5)在平均小句数指标上,只有通用类和小说类差异显著,其他各组之间差异均不显著。

表 4. 翻译汉语和原创汉语句法复杂度指标双因素方差分析结果

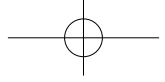
| 测量指标        | 主效应     |         |         |         | 交互效应   |         |
|-------------|---------|---------|---------|---------|--------|---------|
|             | 文本类型    |         | 体裁类型    |         |        |         |
|             | F 值     | p 值     | F 值     | p 值     | F 值    | p 值     |
| 平均大句长       | 3.717   | > 0.05  | 212.691 | < 0.001 | 5.391  | < 0.01  |
| 平均小句长       | 227.325 | < 0.001 | 327.212 | < 0.001 | 18.894 | < 0.001 |
| 平均小句数       | 298.664 | < 0.001 | 6.637   | < 0.001 | 0.675  | > 0.05  |
| 名词短语平均长度    | 26.231  | < 0.001 | 480.260 | < 0.001 | 2.185  | > 0.05  |
| 介词短语平均长度    | 117.356 | < 0.001 | 226.910 | < 0.001 | 8.239  | < 0.001 |
| 简单子句平均并列短语数 | 18.612  | < 0.001 | 201.778 | < 0.001 | 1.203  | > 0.05  |
| 简单子句平均名词短语数 | 91.969  | < 0.001 | 219.781 | < 0.001 | 4.003  | < 0.01  |
| 简单子句平均动词短语数 | 17.909  | < 0.001 | 54.111  | < 0.001 | 4.666  | < 0.01  |
| 句均介词短语数     | 13.685  | < 0.001 | 134.915 | < 0.001 | 1.246  | > 0.05  |
| 整体搭配多样性     | 5.286   | < 0.05  | 110.689 | < 0.001 | 0.786  | > 0.05  |
| 特殊搭配多样性     | 0.015   | > 0.05  | 277.148 | < 0.001 | 3.184  | < 0.05  |
| 通用搭配多样性     | 6.417   | < 0.05  | 56.284  | < 0.001 | 2.384  | > 0.05  |
| 特殊搭配比例      | 9.384   | < 0.01  | 104.887 | < 0.001 | 7.500  | < 0.001 |
| 低频搭配比例      | 14.644  | < 0.001 | 66.993  | < 0.001 | 15.789 | < 0.001 |

从均值排序来看,呈现出两个特点:第一,在平均大句长、平均小句长、名词短语平均长度、介词短语平均长度、简单子句平均并列短语数、简单子句平均名词短语数、简单子句平均动词短语数、句均介词短语数、通用搭配多样性 9 个指标上,学术类 > 新闻类 > 通用类 > 小说类;第二,在整体搭配多样性、特殊搭配多样性、特殊搭配比例 3 个指标上,小说类 > 通用类 > 新闻类 > 学术类。其他两个指标的具体表现为:在平均小句数上,通用类 > 新闻类 > 学术类 > 小说类;在低频搭配比例上,新闻类 > 学术类 > 通用类 > 小说类。

文本类型与体裁类型的交互效应在平均大句长、平均小句长、介词短语平均长度、简单子句平均名词短语数、简单子句平均动词短语数、特殊搭配多样性、特殊搭配比例、低频搭配比例 8 个指标上均显著,说明原创汉语和翻译汉语中四大类体裁在这 8 个指标上均可能存在差异。为探究不同文本类型在同一体裁上是否存在显著差异,我们进行了简单效应检验,结果如下:

1)平均小句长、介词短语平均长度、简单子句平均名词短语数 3 个指标,在新闻类、通用类和学术类体裁上,文本类型的效应均显著,说明翻译汉语在这 3





个指标上均显著高于原创汉语;但在小说类体裁上,文本类型的效应不显著,说明翻译汉语与原创汉语无显著差异。

2) 平均大句长指标在小说类体裁上,文本类型的效应显著,说明翻译汉语的平均大句长显著低于原创汉语;但在新闻类、通用类和学术体裁上,文本类型的效应均不显著,说明翻译汉语与原创汉语的平均大句长无显著差异。

3) 简单子句平均动词短语数指标在通用类和学术类两种体裁上,文本类型的效应均显著,但在新闻类和小说类体裁上,文本类型的效应均不显著。这说明在通用类和学术类体裁上,翻译汉语的简单子句平均动词短语数显著高于原创汉语,但在新闻类和小说类体裁上均无显著差异。

4) 特殊搭配比例指标在学术类和小说类体裁上,文本类型的效应均显著,但在新闻类和通用类体裁上,文本类型的效应均不显著。这说明在学术类和小说类体裁上,翻译汉语的特殊搭配比例显著高于原创汉语,但在新闻类和通用类体裁上均无显著差异。至于特殊搭配多样性指标,在四大类体裁上,文本类型的效应均不显著。

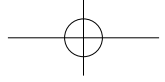
5) 低频搭配比例指标在新闻类、学术类和小说类体裁上,文本类型的效应均显著,但在通用类体裁上,文本类型的效应不显著。这说明在新闻类和学术类体裁上,翻译汉语的低频搭配比例显著低于原创汉语,但在小说类体裁上显著高于原创汉语;在通用类体裁上二者无显著差异。

## 5. 讨论

### 5.1 结果讨论

从以上结果可知,在平均小句长、名词短语平均长度、介词短语平均长度、简单子句平均并列短语数、简单子句平均名词短语数、简单子句平均动词短语数、句均介词短语数、特殊搭配比例 8 个指标上,翻译汉语显著高于原创汉语;在平均小句数、整体搭配多样性、通用搭配多样性、低频搭配比例 4 个指标上,翻译汉语显著低于原创汉语。可见,与原创汉语相比,翻译汉语同时存在“繁化”和“简化”两种翻译共性。

Liu, Liu & Lei (2022) 发现翻译汉语的句法复杂度(词性熵值)高于原创汉语,本研究则发现在句子、短语和搭配层面的 8 个指标上翻译汉语均高于原创汉语。这两项研究都发现翻译汉语的句法更加复杂,即翻译汉语具有“繁化”特征。但本研究还发现翻译汉语在另外一些指标上(如低频搭配比例)句法复杂度更低,具有“简化”特征。造成差异的原因是基于熵的方法主要衡量特征分布的多样性,因此 Liu, Liu & Lei (同上)发现的句法“繁化”主要反映了句法多



样性更高;而本研究采用了句子、短语和搭配三个维度的多种计算方法——绝对复杂度和相对复杂度,测量指标的全面性和数量均高于前者。这进一步证明句法复杂度是一个多维立体的复杂系统(Ortega 2003; Lu 2011; Kyle & Crossley 2018),如要全面考察其特征,需包含多个维度、多种测量指标。

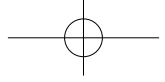
在句子层面和短语层面,与原创汉语相比,翻译汉语小句内部的构成更加复杂,具体体现在小句内短语数量增多,如并列短语、名词短语、动词短语等;小句内短语长度增长,如名词短语和介词短语。这可能是受“源语干扰”的影响,即“与原文结构相关的现象倾向于被转移到译文中”(Toury 1995: 275-276)。ZCTC 语料库中大部分语言材料译自英语,在翻译过程中,英语的复杂结构会迁移到翻译汉语中,导致结构扩容,句法结构更加复杂,甚至现代汉语在发展过程中也受其影响,这已经得到实证研究的证实(王克非、秦洪武 2009, 2017; 秦洪武、孔蕾 2018)。另外,本研究发现翻译汉语的平均小句长显著高于原创汉语,但在平均大句长上二者却无显著差异,原因是原创汉语在句子中使用的小句数更多,这也体现了原创汉语的句法特点,即汉语中经常使用话题链和流水句的形式,通过多个简单的小句来表达完整的意义(Chao 1968; 胡明扬、劲松 1989)。

在搭配层面,与原创汉语相比,翻译汉语的特殊搭配比例更高,但低频搭配比例、整体搭配多样性、通用搭配多样性更低。本研究的特殊搭配是指具有汉语特色的搭配,包括量名搭配、框式介词搭配、介动搭配和述补搭配。进一步分析发现,在介动搭配比例和框式介词搭配上,翻译汉语显著高于原创汉语,原因可能是介动搭配包括“把/被/由/对/为……V”等结构,以往研究发现翻译汉语的被动句数量远高于原创汉语(Xiao & Dai 2014),导致翻译汉语的介动搭配比例显著高于原创汉语。关于框式介词搭配,英语中有多种形式在翻译汉语中被译为框式介词搭配,例如吴继峰(2019)基于英汉平行语料库发现,在翻译汉语中,不仅“英语介词+X”可译为“在X上/下”,部分英语名词、动词、形容词、副词等词类和短语以及从句等其他形式也可译为“在X上/下”。该原因可能导致翻译汉语的框式介词搭配比例显著高于原创汉语。

低频搭配比例是本研究中唯一衡量相对复杂度的指标,基于外部语料库的频次和互信息统计来判定,主要是找到词语关联比较紧密且不是很高频的搭配,它们通常是比较地道的搭配,对于二语学习者来说难度较高(胡韧奋、肖航 2019)。为进一步考察原创汉语中有哪些高于翻译汉语的低频搭配,我们提取了全部语料中的低频搭配,发现原创汉语中更容易出现如例(1)所示的低频搭配:

- (1) 中国特色搭配:亲笔题词、先进党支部、基本路线、优秀基层干部  
汉语特有搭配:沉重的包袱、思想成熟、清醒的头脑、陶冶情操





例(1)中的两种低频搭配,前者在英语中可能很少甚至不出现,因此翻译汉语中也会非常有限甚至不存在;后者是汉语特有表达,英语中可能有同义词组,但通常采用其他的表达方式翻译,如例(2):

(2) 原创汉语: 他有沉重的思想包袱。

原创英语: He is very stressed. / He is under a lot of stress.

翻译汉语: 他压力很大。

这样一来,只能从翻译汉语中提取到“压力大”这个高频普通搭配。我们推测这种现象在原创和翻译汉语中都比较普遍,使得原创汉语中的低频搭配比例高于翻译汉语。

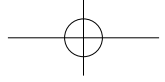
翻译汉语整体搭配多样性、通用搭配多样性更低的原因可能是,翻译汉语的词汇多样性低于原创汉语(Liu, Liu & Lei 2022),从而导致搭配多样性也低于原创汉语。因为搭配多样性计算的依据是词汇多样性,词汇多样性越低,搭配多样性也会越低。

另外,需要注意的是,以上结论表明翻译汉语中“繁化”和“简化”现象同时存在,但“繁化”和“简化”的程度在不同体裁上并不相同。例如,在平均小句长、介词短语平均长度、简单子句平均名词短语数 3 个指标上,新闻类、通用类和学术类体裁的翻译汉语显著高于原创汉语,但小说类体裁无显著差异。在低频搭配比例上,新闻类和学术类体裁的翻译汉语显著低于原创汉语,但在小说体裁上显著高于原创汉语,在通用类体裁上无显著差异。

## 5.2 理论解释

以上研究结果可用认知翻译学的“引力假说”(the hypothesis of gravitational pull)和心理语言学的“翻译的水平模型”(the horizontal model of translation)进行解释。

Halverson (2017) 将最初的“引力假说”(Halverson 2003) 分解为三个影响因素:源语的突出性、目标语的突出性和连接强度,译语同时受此三因素的影响。根据 Halverson (2017: 14) 的定义,源语的突出性,即“引力”,是指“一种认知力量,使译者难以摆脱源语中高度突出的表征元素的认知牵引”;目标语的突出性,即“磁力”,是指“在对目标语言项目的认知搜索中,译者更可能被吸引到具有高突出性/频率的目标语言项目”,与“引力”相排斥;连接强度“反映了一个翻译对高频共现的影响, [……] 通过频繁激活翻译对中的一个成员,不同语言之间的联系也会得到加强”。本研究发现,“磁力效应”和“引力效应”在翻译汉语句法复杂度的多个层面上都有所体现,具体表现为: 1) 在平均小句长度和介动搭配比例等指标上,翻译汉语高于原创汉语,这体现了“引力效应”,导



致翻译汉语的句法结构更加复杂; 2) 翻译汉语的框式介词搭配比例显著高于原创汉语, 这体现了“磁力效应”, 即译者在译文中再现目标语典型突出的语言特征, 甚至比目标语本身使用的典型特征更多。

“翻译的水平模型”认为“源语的语言特征会影响翻译, 因为目标语在译者理解源语的过程中已被激活, 或者在生成目标语过程中仍然活跃”(Maier, Pickering & Hartsuiker 2017: 1576)。因此, 它预测译者会在译语中重现源语的句法和其他语言形式。Liu, Liu & Lei (2022: 11) 提出“在水平翻译中, 源语和目标语的词汇和句法属性通过共享的特征相连, 意味着源文本的解码与目标文本的赋码是并行的”。该模型认为译语是源语与目标语竞争的结果。本研究发现, 在低频搭配比例和平均小句数等指标上, 翻译汉语显著低于原创汉语, 表明目标语的一些搭配和句法特征, 由于没有通过共享特征与源语相连导致未在译语中被激活, 因此低频搭配比例和平均小句数等更低。在平均小句长、特殊搭配比例等 8 个指标上, 翻译汉语高于原创汉语, 这可能是由于跨语言之间的结构启动对句子翻译的影响, 使译者在翻译句子时倾向于在译语中使用与源语句法结构相似的结构 (Maier, Pickering & Hartsuiker 2017; Liu, Liu & Lei 2022)。

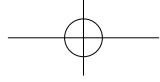
## 6. 结语

本文基于 LCMC 和 ZCTC 两个可比语料库, 对比了翻译汉语和原创汉语的句法复杂度差异。研究发现, 在平均小句长、名词短语平均长度、介词短语平均长度、简单子句平均并列短语数、简单子句平均名词短语数、简单子句平均动词短语数、句均介词短语数、特殊搭配比例 8 个指标上, 翻译汉语的句法复杂度显著高于原创汉语; 在平均小句数、整体搭配多样性、通用搭配多样性、低频搭配比例 4 个指标上, 翻译汉语显著低于原创汉语。研究还发现, 在 8 个指标上文本类型与体裁类型有显著的交互作用, 其中在 7 个指标上翻译汉语和原创汉语不同体裁的句法复杂度有不同表现。研究表明翻译汉语同时存在“简化”和“繁化”的现象。本文采用对比分析、“引力假说”和“翻译的水平模型”对研究结果进行了解释。

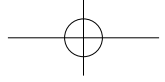
未来研究将进一步扩充不同层面的句法复杂度测量指标, 以更加细致、全面地考察翻译汉语和原创汉语的句法复杂度差异。另外, 受语料库容量的影响, 本文未考察 15 小类体裁之间的差异, 未来可建设更大规模的可比语料库, 实现小类体裁之间的对比。

## 参考文献

- Baker, M. 1993. Corpus linguistics and translation studies: Implications and applications [A]. In M. Baker, G. Francis & E. Tognini-Bonelli (eds.). *Text and Technology: In Honour of John*



- Sinclair [C]. Amsterdam: John Benjamins. 233-250.
- Baker, M. 1996. Corpus-based translation studies: The challenges that lie ahead [A]. In H. Somers (ed.), *Terminology, LSP and Translation* [C]. Amsterdam: John Benjamins. 175-186.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese* [M]. Berkeley: University of California Press.
- Cui, Yue, et al. 2022. CTAP for Chinese: A linguistic complexity feature automatic calculation platform [R]. Paper presented at the Thirteenth Language Resources and Evaluation Conference, Marseille, France, June 2022.
- Halverson, S. 2003. The cognitive basis of translation universals [J]. *Target* 15(2): 197-241.
- Halverson, S. 2017. Gravitational pull in translation: Testing a revised model [A]. In G. De Sutter, M. Lefer & I. Delaere (eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions* [C]. Berlin: Mouton de Gruyter. 9-46.
- Hu, Mingyang & Song Jin [胡明扬、劲松]. 1989. The first exploration of flowing sentences [J]. *Language Teaching and Linguistic Studies* (4): 42-54. [流水句初探,《语言教学与研究》4]
- Hu, Renfen [胡韧奋]. 2021. On the relationship between collocation-based syntactic complexity and Chinese second language writing [J]. *Applied Linguistics* (1): 132-144. [基于搭配的句法复杂度指标及其与汉语二语写作质量关系研究,《语言文字应用》1]
- Hu, Renfen, Jifeng Wu & Xiaofei Lu. 2022. Word-combination-based measures of phraseological diversity, sophistication, and complexity and their relationship to second language Chinese proficiency and writing quality [J]. *Language Learning* 72(4): 1128-1169.
- Hu, Renfen & Hang Xiao [胡韧奋、肖航]. 2019. The construction of Chinese collocation knowledge bases and their application in second language acquisition [J]. *Applied Linguistics* (1): 135-144. [面向二语教学的汉语搭配知识库构建及其应用研究,《语言文字应用》1]
- Hu, Xian Yao [胡显耀]. 2021. *A Corpus-based Study of Chinese-English Translation Style* [M]. Beijing: Science Press. [《基于语料库的汉英翻译语体研究》。北京:科学出版社]
- Jin, Honggang. 2007. Syntactic maturity in second language writings: A case of Chinese as a foreign language (CFL) [J]. *Journal of the Chinese Language Teachers Association* 42(1): 27-54.
- Kyle, K. & S. Crossley. 2018. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices [J]. *The Modern Language Journal* 102(2): 333-349.
- Liu, Kanglong, Zhongzhu Liu & Lei Lei. 2022. Simplification in translated Chinese: An entropy-based approach [J]. *Lingua* 275: Article No. 103364.
- Lu, Xiaofei. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development [J]. *TESOL Quarterly* 45(1): 36-62.
- Lu, Xiaofei & Jifeng Wu. 2022. Noun phrase complexity measures in Chinese and their relationship to L2 Chinese writing quality: A comparison with topic-comment-unit-based measures [J]. *The Modern Language Journal* 106(1): 267-283.
- Maier, R., M. Pickering & R. Hartsuiker. 2017. Does translation involve structural priming? [J]. *Quarterly Journal of Experimental Psychology* 70(8): 1575-1589.
- Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing [J]. *Applied Linguistics* 24(4): 492-518.
- Qin, Hongwu & Lei Kong [秦洪武、孔蕾]. 2018. The impact of translational Chinese on original



- language: A syntactic complexity perspective [J]. *Journal of Foreign Languages* (5): 15-26. [ 翻译语言影响原创语言的途径和方式——基于汉语结构复杂度的分析,《外国语》5]
- Toury, G. 1995. *Descriptive Translation Studies and Beyond* [M]. Shanghai: Shanghai Foreign Language Education Press.
- Wang, Kefei & Xianyao Hu [ 王克非、胡显耀 ]. 2008. A parallel corpus-based study on lexical features of translated Chinese [J]. *Chinese Translators Journal* (6): 16-21. [ 基于语料库的翻译汉语词汇特征研究,《中国翻译》6]
- Wang, Kefei & Hongwu Qin [ 王克非、秦洪武 ]. 2009. A parallel corpus-based study of general features of translated Chinese [J]. *Foreign Language Research* (1): 102-105. [ 英译汉语言特征探讨——基于对应语料库的宏观分析,《外语学刊》1]
- Wang, Kefei & Hongwu Qin [ 王克非、秦洪武 ]. 2017. A diachronic multiple corpus-based approach to the role of translational Chinese in the evolution of Chinese [J]. *Foreign Language Teaching and Research* (1): 37-50. [ 基于历时复合语料库的翻译与现代汉语变化考察,《外语教学与研究》1]
- Wang, Yupei & Renfen Hu. 2021. A prompt-independent and interpretable automated essay scoring method for Chinese second language writing [A]. In Sheng Li *et al.* (eds.). *Chinese Computational Linguistics* [C]. Switzerland: Springer. 450-470.
- Wu, Jifeng [ 吴继峰 ]. 2019. *A Study of the Prepositional Frame Structure "Zai X Shang/Xia" for Teaching Chinese to English Speakers* [M]. Beijing: China Social Sciences Press. [《面向对英汉语教学的介词性框式结构“在 X 上 / 下”研究》。北京: 中国社会科学出版社]
- Wu, Jifeng, Renfen Hu & Xiaofei Lu [ 吴继峰、胡韧奋、陆小飞 ]. 2022. Effects of production modalities on syntactic complexity in intermediate level CSL learners [J]. *Chinese Teaching in the World* (3): 399-415. [ 产出方式对中级水平汉语二语者句法复杂度的影响,《世界汉语教学》3]
- Wu, Jifeng & Xiaofei Lu [ 吴继峰、陆小飞 ]. 2021. The relationship between syntactic complexity and L2 Chinese writing quality: Large-grained vs. fine-grained indices [J]. *Applied Linguistics* (1): 121-131. [ 不同颗粒度句法复杂度指标与写作质量关系对比研究,《语言文字应用》1]
- Xiao, R. & Guangrong Dai. 2014. Lexical and grammatical properties of translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective [J]. *Corpus Linguistics and Linguistic Theory* 10(1): 11-55.
- Xu, Zuohao & Yue Jiang [ 徐佐浩、蒋跃 ]. 2021. Activity of translational Chinese: A study based on three online corpora [J]. *Foreign Language Teaching and Research* (1): 113-123. [ 翻译汉语的活动度——基于在线语料库的研究,《外语教学与研究》1]

收稿日期: 2022-10-06;修改稿 2023-01-31;本刊修订 2023-02-03

通讯地址: 100048 北京市 首都师范大学国际文化学院(吴)  
999077 香港 香港理工大学中文及双语学系(刘)  
100875 北京市 北京师范大学国际中文教育学院(胡)  
100048 北京市 首都师范大学心理学院(周)