



A Corpus-Based Study of the Differences between Translations of Famous Translators and Translation Learners

Shilin Ding, Yuanfei Tang, Hui Jia

East China University of Science and Technology, Shanghai, China
Email: Dingslllove@163.com, 1748257771@qq.com, jhflora@163.com

How to cite this paper: Ding, S.L., Tang, Y.F. and Jia, H. (2022) A Corpus-Based Study of the Differences between Translations of Famous Translators and Translation Learners. *Open Access Library Journal*, **9**: e8865.
<https://doi.org/10.4236/oalib.1108865>

Received: May 8, 2022

Accepted: June 25, 2022

Published: June 28, 2022

Copyright © 2022 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Under the background of the construction of new liberal arts, corpus-based translation studies fit the development trend. In order to investigate the differences between translations of translation learners and famous translators, this paper collects learners' translations from the English-Chinese Translation Course at a university in Shanghai and various kinds of famous translators' translations to form comparable corpora. Then it uses AntConc version 3.2.1 for data processing and Chi-Square and Log-likelihood Calculator to determine if there is a significant difference, and analyzes the differences with specific examples. The study finds that the translation learners use longer sentences and more "be" indicating the passive tense and "de" showing possession than famous translators. The differences are significant, resulting in a decrease in the readability of the translations of the learners. Through the quantitative study of corpus data, the learners can reflect on weakness and improve translation competence; teachers can optimize translation lesson plans in response to the obvious differences.

Subject Areas

Linguistics

Keywords

Corpus, Average Sentence Length, *bei*, *de*

1. 引言

2019年,教育部提出全面推进新工科、新医科、新农科和新文科建设,而其中对于新文科建设,专家们给出了不同的解释。吴岩(2019: 6) [1]在名为

《新使命 大格局 新文科 大外语》的文章中指出“结合社会发展新需求、学科交叉融合新趋势、科学研究新成果，加强传统文科专业的内涵建设，建设新兴的文科专业”，而语料库翻译研究正是融合了语料库语言学研究与翻译实践，并与大数据结合，从科学的角度分析译文，体现了学科交叉融合的新文科特点。

对于刚接触翻译的初学者而言，教师课上讲授的翻译理论只是指引方向的路标。如何让学生译者利用翻译练习，更好地提升翻译能力，一步步到达目的地才是关键。如果学生能从自身译文与名家译文的对比中认识不足，并从中学习，则更有助益。故本文旨在搜集上海某高校英语专业《英汉翻译》课程翻译练习，组成学习者译文语料库，与搜集的各类名家译文组成的名家译文语料库作为对比，分析两者存在的差异，并探究存在差异的原因，以此找到减少差异的办法，辅助学习者翻译实践。

2. 文献综述

纵观文献，基于大数据新文科的背景，语料库研究在近年来成为热门趋势，堪称文科与大数据结合的典范。并且基于中国知网(CNKI)文献库检索出的文献，总结分析研究热点，发现国内语料库翻译教学研究领域已经在理论和实践上取得一定进展，如王克非(2004) [2]根据北京外国语大学外语教育研究中心通用汉英平行语料库探讨平行语料库在翻译教学中的应用价值，于连江(2004) [3]认为基于语料库研究有利于拓展和深化翻译研究、促进翻译实践、提高翻译和翻译教学的质量，而肖维青(2007) [4]提出，语料库研究在应用翻译层面的巨大实用性和开发潜力：译员培训、翻译教学、翻译批评和机器翻译等都已经进入语料库研究的视野。可见语料库在翻译教学领域已经有诸多成果，但在译者培养等理论和实践方面有所欠缺，未来可以进行更多尝试。而本文正是从学习者角度切入，研制以学习者翻译文本和名家翻译文本为样本的两个单语语料库，在译者培养的实践上迈进一步。研究旨在研究英汉翻译过程中，初学者译文与名家译文之间存在何种差异，而进一步通过数据量化差异，以求指导学习者通过语料库提高翻译实践的效率和能力。

3. 研究方法步骤

3.1. 语料收集

长期的经验积累有助于增强英汉翻译能力。在此过程中学习者不免需要先对翻译名家的翻译技巧进行学习借鉴，以增强对翻译的理解，再将其内化成自己的翻译方式，从而提升个人翻译能力。为了能够更加清晰直观地体现出与名家相比，学习者的不足主要存在于哪些方面，从而可以在英汉翻译教学中对其进行更有针对性的指导帮助，笔者通过数据采集的方式，分别收集名家译文和学生译文，在控制两者字数大致等同的条件下，自建两个规模相当的语料库，分别命名为名家译文语料库以及学习者译文语料库。

其中所建两套语料库语料的选择基本遵循贴合英汉翻译课程教学适用性原则，并在此基础上尽可能丰富内容，广泛覆盖各类语料，兼具教学性和实

用性。名家译文语料由三部分组成，其中韩素音翻译大赛是中国翻译界影响最广的翻译比赛，其参考译文由名家探讨后给出；《中国翻译》自学之友板块收录了诸多名家的翻译片段；《西南联大英文课》[5]汇集中外名家名篇，且由清华大学、北京大学等多校英语名师联合翻译。而学习者译文选自上海某高校英语专业《英汉翻译》课程翻译练习。该课程面向英语专业三年级学生，学生译者在经过笔译课程学习后，对基本的翻译理论有初步了解，但并没有在实践当中熟练掌握、使用理论，符合本文对“初学者”的定义。两套语料库均涵盖了文学、科技、自然科学、社会科学等方面的语料内容。名家译文语料库收录超三十五万字，学习者译文语料库收录超三十七万字。（具体见表 1）

表 1. 语料库语料来源

	名家译文语料库			学习者译文语料库
	韩素音翻译大赛	《中国翻译》自学之友	《西南联大英文课》	《英汉翻译》课程课后作业
字数	65,687	143,891	142,466	372,991
总计		352,044		372,991

3.2. 研究步骤

接下来笔者对两个语料库进行对比分析。主要研究内容为比较名家译文与学习者译文之间句长差异以及两者译文中“的”字和“被”字的不同使用频率。首先将名家译文与学习者译文分别归入两个 Word 文档，得出两者总字数。接着将文档转化为 txt 格式，利用 Antconc3.2.1 版本[6]查找断句标点以及“的”字与“被”字的个数，并且通过计算得出平均句长以及“的”字与“被”字的使用频率。之后运用 Chi-Square and Log-likelihood Calculator 进行卡方检验，得出差异程度，印证其研究价值。为了将差异进一步具体化，下一步将在呈现数据的基础上，结合语料库中具有代表性的具体实例分析差异产生的主要原因，以便在日后的翻译教学中能够更加注重有效方法的教学，避免此类情况的重复发生，从而达到提升学习者个人翻译能力这一最终目的。

4. 结合实例具体分析

根据研究步骤，两个语料库的平均句长、“的”字和“被”字使用频率的数据呈现在表 2 中。根据统计学描述，P 值即概率，反映某一事件发生的可能性大小。通过卡方检验方法所得到的 P 值，一般 $P < 0.05$ 为有统计学差异， $P < 0.01$ 为有显著统计学差异， $P < 0.001$ 为有极其显著的统计学差异。将数据输入 Chi-Square and Log-likelihood Calculator [7]进行卡方检验，得到结果如下：

由此可见，名家译文和学习者译文在句长对比，以及“被”字和“的”字的使用频率方面，存在着极其显著的差异，具有极大研究价值。具体表现为：

- 1) 名家译文平均句长短于学习者译文

2) 名家译文“被”字频率低于学习者译文

3) 名家译文“的”字频率低于学习者译文

为了进一步理解差异，下一步将结合语料库中具有代表性的双语实例，分析差异产生的主要原因。

表 2. 名家语料库与学习者语料库差异

		名家译文语料库(352,044 字)	学习者译文语料库(372,991 字)	P 值
句长	断句标点(个)	10297	8559	P < 0.001
	平均句长	34.19 字/句	43.58 字/句	
“被”字 “ <i>bei</i> ”	“被”(个)	414	713	P < 0.001
	频率	0.12%	0.19%	
“的”字 “ <i>de</i> ”	“的”(个)	14,532	17,996	P < 0.001
	频率	4.13%	4.82%	

4.1. 平均句长

平均句长是指每句句子的平均字数，是影响可读性的重要因素之一。相较于短句，学习者译文与名家译文的句长差异主要体现在对英语长难句的翻译上。党争胜(2006: 64) [8]指出，“在翻译时，英语长句会造成理解困难，中国学生对于长句经常无从下手，翻译结果往往言不达意。”英语长句翻译的困难性在于句中通常包含多个从句或分句，这些从句或分句之间存在互相指代、归附等关系，使整句结构复杂，导致学习者在做翻译时难以理清其中存在的逻辑关系。程洪珍(2003) [9]指出如若无法很好地处理英语长句，译文必然诘屈聱牙，难以顺畅自然，必须按意群拆开长句，分开叙述，重组译文短句。

其中一个极具代表性的例子出自第 32 届韩素音翻译比赛英译汉练习，原文为：“Faced with what seems to be such overwhelming confirmation of aesthetic educators’ irrelevance to today’s understanding of national progress—namely, advancement in business, science and technology—aesthetic educators in the humanities and the arts are struggling to communicate to others outside our field, and to the public at large, our vital role.”

参考译文：“审美教育工作者与今天所谓的‘国家进步’(即商业、科学与技术的进步)毫不相干。面对这种近乎一边倒的认知，人文艺术工作者正在竭力向其他学科和公众阐明自己在国家进步中的关键作用。”

学生译文 1：“面对对于美学教育家与现在国家进步的理解毫无相关性的似乎是压倒性的证实——即商业、科学和技术的进步——人文学科和艺术方面的美学教育家正努力向我们领域外的其他人以及公众介绍我们所扮演的至关重要的角色。”

学生译文 2：“面对的这些有关审美教育者与国家发展——即工商，科学与技术学科的发展，毫无关联的言论看起来是如此无懈可击，艺术和人文学科领域的审美教育者正努力与领域外的人交流，对于大部分的公众，我们的重要角色。”

学生译文 3: “压倒性的多数人认为审美教育对国家进步并无助益,而商业、科学、技术教育则举足轻重。面对这类窘境,美育工作者难于与非专业领域人士进行交流,也无法在社会公众前施展拳脚、发挥效用。”

学生译文 4: “当今世界,在人们对于国家进步的理解中,商业和科学进步蒸蒸日上,而相比之下,美学教育者又是如此轻如鸿毛。面临这样的难题,人文艺术领域的美学教育者们,正努力向公众传达人文教育不可或缺的理念。”

原句属于长句,探讨审美教育与国家进步两者间的关系。参考译文以“such overwhelming confirmation”作为切入点,将此句分为两个部分,第一句指出具体的“一边倒的认知”是什么,第二句阐述面对这种境况相关人士的反应。这样处理使得整句句子脉络清晰,符合汉语用语逻辑,可读性强,便于读者理解。

而许多学习者在处理这句话的时候难以梳理出主干,导致译文要么显得复杂冗长,要么不符合汉语的用语习惯,甚至通过部分译文可以看出学习者在原句的理解上就出了问题,可读性也就逊色很多。

学生译文 1 完全遵照了原句的语序,连标点符号都与原文保持一致,这种做法确实可以做到不遗漏原句信息,保证内容的完整性,但也导致译句整体结构十分僵硬,读起来不符合汉语的用语习惯。

学生译文 2 暴露出学习者在翻译时不够细心的问题,许多标点符号都没有标在最恰当的位置上,造成译句断句混乱、语句不通的严重后果。

学生译文 3 可看出学习者已初步具有了梳理句子主干信息能力以及调整句子语序的意识,然而问题在于译句在一定程度上歪曲了原句所想要表达的意思。原句中所传递的是相关人士努力与外界沟通的积极含义,而译句则展现出了消极的定论,与原句情感意义不符。

当然也有较为优秀的学习者译文。学生译文 4 不仅能够以“such overwhelming confirmation”作为切入点,将原句分为两个部分,规避句长过长的的问题,而且译句不循规蹈矩,拥有自己的翻译风格。“蒸蒸日上”和“轻如鸿毛”这两个四字短语的对比运用,将美育的穷迫境地体现得淋漓尽致,也使读者能够更加容易地理解作者想要表达的意思。

4.2. “被”字的分析

英语以物称为主,因此多被动语态(the passive voice),尤其在科技类文本中这一特征更为显著,而汉语被动语态则少得多。在翻译的过程中,被动语态的翻译可以彰显译者的翻译策略。王克非、刘鼎甲(2018) [10]曾基于超大型英汉平行语料库,考察科技文本中英语被动结构在汉语中的对译形式,发现英语被动结构翻译形式可以灵活多样。但学生译者并未熟练掌握翻译技巧,导致译文中“被”字比例过高,可读性较差。

例如第 33 届韩素音翻译比赛英译汉练习中,利用 TreeTagger 对原文进行词性标记,并通过 VB(D/G/M/P/Z) + VVN 的格式搜寻,找到被动句 6 句。而在 33 届韩素音参考译文中,共有“被”字 4 处,而学生译文平均“被”字为 6.4 处。故可以看出,参考译文选用了一些策略规避“被”字出现。

差异最明显的是以下这句, “We had the sky, up there, all speckled with stars, and we used to lay on our backs and look up at them, and discuss about whether they was made, or only just happened—Jim allowed they was made, but I allowed they happened; I judged it would have took too long to make so many.”

参考译文: “头顶上就是低垂的天空, 挂满了星星。那个时候, 我们常常就这样躺着, 望着天上的星星, 琢磨这些星星是怎么出来的, 造出来的, 还是自己冒出来的? 吉姆说它们是造出来的, 可我觉得它们是自己冒出来的, 不然, 造这么多星星, 得花多少时间啊。”

学生译文 1: “星星曾点缀着我们头顶的天空。那时我们背靠着船只, 仰望着它们, 讨论它们到底是被创造出来的呢, 还是天然就存在——吉姆觉得它们是被创造出来的, 但我觉得它们天然就存在, 因为要创造出这么多的星星未免也太耗时间了。”

学生译文 2: “我们曾经常仰面朝天, 一边看着夜空中的繁星, 一边讨论它们究竟是被刻意创造出来的, 还是自然而然地诞生的: 吉姆认为一定是前者, 但我却坚持是后者, 因为我觉得没谁能花这么久创造出这么多的星星。”

学生译文 3: “我们曾头顶浩瀚星空, 也曾躺在木筏上, 仰望着星空, 共同探讨星星究竟从何而来, 是孕育而生还是浑然天成。吉姆认为它们并非天然形成, 而我持相反意见。我想要孕育出这么多星星, 一定得花不少时间。”

原文中有两个重复的被动形式“was made”。其中, 名家译者化被动为主动, 省略了主语“造物主”。并且考虑到文章背景, 行文流露出孩童的活泼天真。

学生译文 1 依照原文思路, 将被动语态顺势翻出, 而这也是大部分学生采取的策略, 搭配用法“被造出”、“被创造”居多。

学生译文 2 因前后出现相同结构, 故用“前者”代替, 也是一个很好的解决办法。但存在的问题是, 原文选自马克·吐温《哈克贝利费恩历险记》, 背景是哈克与吉姆两个没有接受正式教育的孩童之间的对话, 而“前者”过于正式, 与文体存在差异。

学生译文 3 注意到了被动结构, 有意地使用主动语态, 以及否定结构规避“被”字, 行文流畅。

4.3. “的”字的分析

翻译中“的”字出现频率过高, 为此余光中先生在 1996 年 4 月参加“翻译学术会议”时作了题为《论“的”的不休》的讲演反思“欧化”现象, 此后这也成为许多学者研究的对象, 如戴玉群(2003: 133) [11]指出: “英语定语可以很长且层次复杂, 出现在被修饰语之前或之后都可能; 而汉语在传统上是一种少用或基本不用定语的语言……两种语言在定语机制上的差异性给翻译工作带来了许多必须要谨慎对待的问题。”可见, “的”字在一定程度上可以反映译文“欧化”现象的程度。

表 2 数据清晰反映学生译文与名家译文之间“的”字使用频率存在显著差异, 学生译文使用频率高于名家译文, 说明学生受原文影响较大。再如在

32 届韩素音比赛英译汉练习中, 参考译文“的”个数为 77 个, 而学生平均“的”个数为 119 个。差异可见一斑。

其中最明显的是对 Germaine de Staël 思想的一句评价, 原文为“Her prescient philosophy turns the definition of progress on its head and could give aesthetic educators a powerful tool to fight for the increased relevance and vitality of the humanities and the arts in the broader notion of progress.”

参考译文为“她的哲学思想颇有远见, 颠覆了‘进步’的定义, 可以为审美教育者提供有力武器, 拓展‘进步’的概念内涵, 提升人文艺术的重要性和生命力。”

学生译文 1: “她的理论具有先见之明, 彻底颠覆了对进步的定义, 为美育工作者提供了强有力的工具, 使他们能够运用关于进步的更宏观的概念, 为人文艺术的重要性和生命力的增强而奋斗。”

学生译文 2: “她的预知性的哲学彻底颠覆了进步的定义, 为美学教育者提供了一个有力的工具, 让他们能够在更广泛的进步概念中, 为人文艺术的相关性和生命力的增强而奋斗。”

学生译文 3: “她富有先见的哲学眼光定义了进步本身, 给予美学教育者强有力的武器, 捍卫其学科对国家进步的相关性和重要性。”

原文为长句, 参考译文利用语义, 将整句切分成短句, 进行翻译。且中文多以过程为导向, 参考译文以“哲学思想”做主语, 统领四个动词, 使整句语义流畅, 这一做法值得借鉴。

而多数学生在处理这一句话时, 使用“的”字次数大于等于参考译文。虽也有学生使用次数少于参考译文, 却以牺牲关键信息为代价, 造成漏译的现象。

学生译文 1 前半部分与参考译文基本相符, 差异主要在于后半部分的分析。学生遵从原文“broader”, 将其翻成“更宏观的”用来修饰“notion”; 而参考译文将其动词化, 译为“扩展”, 前后连贯。最后学生译文两处“的”形成了嵌套, 不如学习参考译文, 将“增强”变为动词, 简洁明了。

学生译文 2 在定语部分翻译出现问题。虽然只有两个修饰词“her prescient”, 但是学生使用两处“的”用来修饰“哲学”, 导致定语过长。可借鉴参考译文处理办法, 将“预知性的”移至句末, 作为动词。

学生译文 3 有意识地使用动词以避免“的”, 这个策略值得学习。与参考译文对比发现, 因为“强有力”是三字词, 所以后面跟名词时加“的”更为符合语音习惯, 而如果换成“有力武器”, 就可以组成一个偏正短语, 从而省略“的”。但这一译文漏译了“in the broader notion of progress”, 故存在可改进的地方。

5. 结语

本文在收集名家译文和学生译文基础上, 建立了两个规模相当的单语语料库, 并对比分析了名家译文与学习者译文在句长、“的”字与“被”字使用频率上的差异。

1) 在句长方面, 学生译文平均长于名家译文。这是因为学生没有很好理

清句子之间的逻辑关系，以及受原文影响没有进行断句，导致译文可读性不强。

2) 被动语态是英文常见语态，而学生译文“被”字平均多于名家译文，译文略显生硬，这体现了学生没有很好地运用翻译策略，使译文符合中文习惯。

3) “的的不休”是学生译者的通病，主要原因在于没有将原文断句以及对长定语处理不当。

这些差异较为容易体现，并且可以通过一些简单的算法获得数据予以证明。而往往这种处于浅层易得的差异形式最能区分名家与学习者翻译水平的差距，同时也是在英汉翻译教学过程中，学习者能够在通过结合翻译理论以及大量训练后最容易得到改进和提升的地方。

对于学生译者而言，利用语料库对比数据，可以明确自己的不足所在，从而反思自己的翻译策略，进一步针对不足进行练习，提高自己的翻译水平。而对于教师而言，语料库数据从学生语料出发，贴合学生实际水平，并可以将其作为教学重难点的辅助，从而针对差异较大的方面进行专门讲解。

但本文仍存在不足之处。如本文建立的语料库依旧属于中小型自用语料库，因此未来将继续搜集名家译文以及学习者译文，继续扩充语料库规模，以求获得更多差异研究，并且可以开放使用。另外，本文从两个语料库整体差异入手，得到较为普遍的结论，未来可从语料库所含不同类别语料入手，分析名家译文与学习者译文，以期获得更多创新性成果。

项目

本研究是大学生创新创业项目(编号: S202110251126)的阶段性成果。

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] 吴岩. 新使命 大格局 新文科 大外语[J]. 外语教育研究前沿, 2019(2): 3-7+90.
- [2] 王克非. 双语平行语料库在翻译教学上的用途[J]. 外语电化教学, 2004(6): 27-32.
- [3] 于连江. 基于语料库的翻译教学研究[J]. 外语电化教学, 2004(2): 40-44.
- [4] 肖维青. 平行语料库与应用翻译研究[J]. 中国科技翻译, 2007, 20(3): 25-28.
- [5] 陈福田, 罗选民. 西南联大英文课(英汉双语版) [M]. 北京: 中译出版社, 2017.
- [6] Anthony, L. (2005) AntConc: Design and Development of a Free-Ware Corpus Analysis Toolkit for the Technical Writing Classroom. 2005 *IEEE International Professional Communication Conference Proceedings*, Limerick, 10-13 July 2005, 729-737. <https://doi.org/10.1109/IPCC.2005.1494244>
- [7] 许家金. 对数似然率计算器[M]. 北京: 中国外语与教育研究中心, 2009.
- [8] 党争胜. 结构分析翻译法初探——浅论英语长句的汉译[J]. 外语教学, 2006, 27(4): 64-66.
- [9] 程洪珍. 英汉语差异与英语长句的汉译[J]. 中国科技翻译, 2003, 16(4): 21-22.
- [10] 王克非, 刘鼎甲. 基于超大型英汉平行语料库的英语被动结构汉译考察与分析

- [J]. 外国语(上海外国语大学学报), 2018, 41(6): 79-90.
- [11] 戴玉群. 欧化有限度: 英汉定语对比与翻译[J]. 四川外语学院学报, 2003, 19(5): 133-136.

Appendix (Abstract & Keywords in Chinese)

基于语料库的名家译文与学习者译文差异研究

摘要: 在新文科建设的大背景下, 语料库翻译学契合发展趋势。为探究学习者与名家翻译之间存在的差异, 本文搜集上海某高校英语专业《英汉翻译》课程翻译练习组成学习者语料库, 并且收录各类名家语料组成规模相当的名家语料库, 运用 AntConc3.2.1 版本进行数据处理, 且进行卡方检验, 并结合实例具体分析。研究发现学习者平均句长、“被”字与“的”字均长于或多于名家译文, 且差异显著, 导致译文可读性下降。通过语料库数据量化研究, 学习者可以反思不足之处, 提升翻译能力; 教师可以针对明显差异, 优化翻译教学。

关键词: 语料库, 句长, “被”, “的”