Modern Linguistics 现代语言学, 2022, 10(11), 2698-2705
Published Online November 2022 in Hans. http://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2022.1011367

Hans汉斯

基于语料库的《自然》高频动词汉译研究

王祎玮

郑州轻工业大学外国语学院，河南 郑州

收稿日期：2022年10月19日；录用日期：2022年11月16日；发布日期：2022年11月28日

摘　要

该文以全球顶级科研期刊《自然》及其平行文本《〈自然〉百年科学经典》制作而成的双语平行语料库为基础，从译文中的高频动词“产生”入手，对英汉词语的翻译策略进行初步考察。研究方法系机器辅助翻译、词频统计、句对检索等组合，旨在以词语表现出的共性与个性，通过定性与定量分析来描述译文的翻译特色，总结翻译策略。

关键词

自然，双语平行语料库，翻译特征

A Corpus-Based Study of Translation Strategies of High-Frequency Verbs in *Nature*

Yiwei Wang

School of Foreign Language, Zhengzhou University of Light Industry, Zhengzhou Henan

Received: Oct. 19th, 2022; accepted: Nov. 16th, 2022; published: Nov. 28th, 2022

Abstract

Based on the bilingual parallel corpus produced by the world's top scientific journal *Nature* and its parallel text *Nature*: *The Living Record of Science*, this paper makes a preliminary investigation on the translation strategies of English and Chinese words from the perspective of the high-frequency verb “产生”. The research method contains computer-aided translation, word frequency statistics and sentence-pairs retrieval etc. Concerning the common and individual characters of its phrases and collocations, the translation features are described and translation strategies are concluded through qualitative and quantitative analysis.

文章引用: 王祎玮. 基于语料库的《自然》高频动词汉译研究[J]. 现代语言学, 2022, 10(11): 2698-2705.
DOI: 10.12677/ml.2022.1011367

Keywords

***Nature*, Bilingual Parallel Corpus, Translational Features**

1. 引言

在世界疫情冲击下，百年变局加速演进，围绕科技创新的国际战略博弈更加激烈。中国目前处于加快推进《国家中长期科学和技术发展规划(2021~2035)》全面实施，实现高水平科技自立自强实现的重要阶段。如何进一步引介世界学术精华，促进中外科技深度交流成为这一阶段的重要课题。科技期刊能够及时报道学术进展，交流科学思想和撰写方法，探讨未来发展方向，而大部分科学文献都没有中文版本，为了实现良好的科技交流，优秀的科技英语翻译必不可少。《自然》作为一本发表所有科学领域开创性研究的期刊，几乎是独一无二。研究其汉译有利于中国学者更好地去学习和研究国外先进的学术期刊论文发表的成果，带动学术争鸣与繁荣，促进公众对科学的理解。本文基于自建双语平行语料库，从《自然》译文中的高频动词“产生”入手，通过定性与定量分析来描述译文的翻译特色，总结翻译策略，以期为提高科技文献翻译质量略尽绵薄之力。

2. 学术期刊翻译研究概述

国内外针对英文学术期刊论文的汉译研究并不热门。截至 2022 年 10 月，在中国知网上搜索“学术期刊翻译”,搜索结果只有 114 条,且大部分为硕士论文。《JOURNAL OF SPECIALISED TRANSLATION》、《中国翻译》等翻译类核心期刊上与之相关的文章更是屈指可数。总的来说，学术期刊翻译研究主要集中在以下几个方向：1) 学术翻译质量批判，如孙张支南，肖亦天的“中国社会科学英文学术期刊高质量发展探究”[1]；2) 中国学术期刊论文特点及翻译策略探究，如杨言，胡翠娥的“学术研究型深度翻译：陈荣捷《老子》英译研究”[2]；3) 探讨专业学术期刊如何在国家语言政策的指导下发挥窗口与桥梁之用，如韩子满，钱虹的“当代中国翻译理论国际传播：现状与展望”[3]；但有关学术期刊论文翻译的研究大多数是汉译英角度出发，英译汉的相关翻译研究比较少。随着中国与西方国家对学术期刊的研究力度日益增强，英文原版学术期刊论文汉译需求日益增加，学术期刊论文翻译的研究仍有不少空间。

3. 《自然》双语平行语料库研究设计

3.1. 研究方法

应用语料库方法来研究翻译文本，可以采用定性和定量分析相结合的方法。翻译实践中，定性分析指某个概念可以译成某个英语词汇；定量分析指某个概念可以译成若干个英文词汇，通常先定量后定性。具体而言，在建成《自然》语料库后，首先通过语料处理软件 WordSmith Tools 列出词表，通过观察词频筛选出高频关键词作为研究对象(本文中为“产生”)；然后通过语料库检索软件 CUC_ParaConc 确定该关键词的原文对应项及每种译文的词频是多少，进而筛选出具体语境中某个概念的正确译文。

3.2. 研究问题

采取定量与定性研究结合的方法，借助机助翻译工具、词频统计、句对检索等组合工具对相关语料

数据进行对比分析，探讨以下三个问题：

A) 《自然》英语原文文本词汇有哪些典型特征？

B) 高水准译文有哪些具体表现？

C) 如何更好地处理《自然》文本汉译，从而更好地服务于中外科技交流？

3.3. 研究语料

采取定量与定性研究结合的方法，借助机助翻译工具、词频统计、句对检索等组合工具对相关语料数据进行对比分析，探讨以下三个问题：本研究语料来自自然出版集团出版的周刊杂志《自然》与2016年北京外语教学与研究出版社出版的《〈自然〉百年科学经典》第四卷[4]。《自然》是全球最具影响力、最有名望的科学期刊。它报道过现代科学中一些最重要的发现，并刊登过如艾萨克·牛顿、阿尔伯特·爱因斯坦、伽利略·伽利雷、弗朗西斯·克里克等世界级科学家的顶级文章。1953年4月25日，詹姆斯·沃森和弗朗西斯·克里克在《自然》杂志发表仅1000余字的论文A Structure for Deoxyribose Nucleic Acid，是20世纪最有名的、也可能最具影响力的文章。文章描述了DNA这一携带有机体基因的分子的结构，阐释了这种结构对于遗传学及遗传的意义。正是基于对这篇划时代文章的理解，才解开“生命之谜”，有了人类基因组解码和克隆多莉羊等一系列进展。研究语料围绕该文章，选择了同时期(1953~1965)生物学的十九篇文章，共计二十篇文章。本语料库词汇总量约为116,553字词，其中汉语文本为74,959字、英语文本约为41,577词。

3.4. 《自然》语料库的创建

《自然》语料库创建主要分三个步骤：1) 语料转写；2) 语料分词和标注；3) 语料的平行对齐。

3.4.1. 语料的转写

语料转写工作繁琐，其完成情况直接决定建库质量。首先，从网络上下载上述文本的PDF版本，利用OCR文字识别工具ABBYY Finereader将识别后的PDF文档转为可编辑的Word文档，以便后续校对地进行。其次，将扫描识别后的文本材料进行人工降噪及校对。在《自然》语料库建库过程中，降噪主要包括不符合建库规范的内容或格式的清洗，如页眉、页码、注释、图片等。校对涉及错别字和乱码纠正等。最后，校对后的文本将保存为TXT格式，并统一编码为UTF-8，为后续存储及加工做准备。

3.4.2. 语料的分词

汉语以字为单位，汉字之间无需空格隔开。英语以词为单位，词与词之间以空格隔开。由于英汉语言之间的这一差异，难以词汇为单位对汉语语料进行统计分析，如类形符比，词汇密度和词频等。为此，选用中科院计算研究所研制的汉语词法分析系统ICTCLA进行分词处理。

3.4.3. 语料的句级对齐

《自然》语料库利用ABBYY Aligner软件进行对齐，对齐语料输出为RTF格式。检查后，利用Word将RTF格式的语料输出为TXT格式，中英文分别存储。

4. 检索结果与讨论

采用WordSmith Tools工具从词汇和句法层面对语料进行描述和数字统计。《自然》双语平行语料库中共1773句对；字词数比，英语:汉语 = 1:1.803；平均句子长度，英语约为23个词，汉语约为42个字。王克非2003年的研究结果，“通常将汉语转换成英语时，一句拆译成两句的较多，反过来，英语转换成汉语时，两句合并成一句的较多[5]”。进行对比，考虑到科技类文本的翻译风格、句子切分等因素，得出数据基本相符。

4.1. 筛选高频关键词

采用 WordSmith 软件对汉语译本进行统计分析。统计结果显示，整个语篇共使用词语 1294 个，最高词频为虚词“的”，词频达 3697 次，最低词频 5 次。实词最高词频(是)为 783 次。词语与词频范围关系见表 1。

Table 1. Word frequency range
表 1. 词频范围

词数(个)	出现频数(次)
1	3000 以上
3	500~1000
8	300~500
18	200~299
45	100~199
16	90~99
8	80~89
19	70~79
25	60~69
33	50~59
58	40~49
73	30~39
138	20~29
362	10~19
487	5~9

有很多学者将高频词的数量确定为 10 的倍数。例如，吕兆杰将高频词确定为 20 个[6]；这种方法虽然缺乏科学依据，只是利用主观经验的判断，但便于后续分析。本次研究假定 50 次词频以下为低频词，共计 1118 个，约占总词语的 86.4%；高频词共计 176 个，约占总词语的 13.6%。

通过划分高低频词语，初步实现了词语的取舍，接着通过观察高频词语，删去专业名词，虚词，系动词，以及情态动词等翻译过程中表现力甚弱的词，保留表现活跃的词语(保留率约为 9.1%)，制作词云图(见图 1)，并选择频数最高的动词“产生”(词频为 139 次，频数位列第 49)作为研究对象。如果译文词语对应的原文词语丰富、表现力相对较强、译文词语搭配多样化，这样的译文词语将被作为“活跃词”。

4.2. 选词的归一性(显化与简化)

在所选词语“产生”的基础上，进行句对检索。这里所应用的是 CUC_ParaConc (中国传媒大学平行语料检索)的双语正则式检索功能，通过检索，在《自然》语料库中得到 129 条包含“产生”的索引，逐条检查发现，有些索引所包含的检索词不只一个，最多的一条索引包含 3 个“产生”。从词语的原文与译文的对比来看，仅“产生”一词对应的原文就有 23 个词语之多，另外有 13 例增译，具体统计见表 2 (同根词归为一个对应项；且如果是动词，其基本形式，-ing 形式和-ed 形式归入同一个对应项，下同) [7]。

Figure 1. High frequency keyword cloud map
图 1. 高频关键词词云图

Table 2. Corresponding meanings and frequency of “产生” from the original texts
表 2. “产生”原文对应项及频次

汉语检索词	原文对应项	频数
产生	produce	56
	result	13
	无	16
	occur	9
	arise from	7
	induce	6
	yield	5
	give	4
	give rise to	3
	affect	3
	appear	2
	be formed by	2
	there be	2
	set	2
	a consequence of	1
	be made	1
	introduce	1
	radiating from	1
	draw through	1
	diffract	1
	present in	1
	the source of	1
	spontaneous	1

“产生”一词词频为 139 次，对应原文“produce”的有 56 处。另有其他 22 处与原文表述不同，但隐含之意多少还是与“produce”之意有关。此外，也有增词翻译(例 6)，原文并未包含“产生”之意或与

其相关的含义，这是一种显化趋势，翻译过程中给译文添加或明示了原文中隐含的语言成分[7]。其生成的原文译文对比如下：

1) 这通常被解释为细胞内产生正常多肽链和异常多肽链的速率不同所致。

This has usually been interpreted as the outcome of a difference in the rate of production of normal and abnormal polypeptide chains within the cell.

2) 由此产生的“G-3b ox”可以通过层析法进行分离与鉴定。

The resultant peptide “G-3b ox” is definitely identifiable, since it is separated in chromatography.

3) 由此可见，通过孤雌生殖产生四倍体是完全不可能的。

Thus the occurrence of a parthenogenetic tetraploid seedling of demissum is extremely unlikely.

4) 野谷等则发现 RNA 噬菌体 f 2 的一个琥珀突变体是由谷氨酰胺突变产生的。

And Notani et al. have found an amber mutant to arise from glutamine in the RNA phage f 2.

5) *β*Tp V 水解产生 3 个天冬氨酸残基。

*β*Tp V yields on hydrolysis three aspartic acid residues.

6) 在总共的 115 个回复突变体中，有 62 个是色氨酸突变产生的回复突变体。

Among a total of 115 revertants, 62 are to tryptophan.

7) 双减数常被用来解释许多异常杂种的产生。

Double reduction has been invoked on a number of occasions to account for the appearance of unusual hybrids.

从上表“产生”及其对应的原文词语可以看出，原文词语较为丰富，译文用词却略显单一。结合前文高频词和低频词所占百分比，再次验证翻译文本具有简化趋势[7]。

4.3. 主导含义的非对称性

“产生”一词所对应的原文主要以“produce”为主。如果反过来，用“produce”一词来检索对应译文的译文翻译，那么结果又会如何呢？检索发现，原文共计有 90 个“produce”，包含有“produce”一词的句对共计 83 对，其中有 7 个句对包含 2 个“produce”。除了翻译成“产生”的 56 次 produce 外，具体汉译统计见表 3。

Table 3. Chinese translation and frequency of “produc*”
表 3. “produc*”汉译及频数

原文检索词	译文	频数
produc*	得到	7
	产物	5
	决定	2
	制备	2
	引起	2
	省译	2
	合成	2
	编码	2
	存在	1

Continued

produc*	注入	1
	长出	1
	诱发	1
	诱导	1
	引入	1
	形成	1
	是	1
	发生	1

这里的 produce 的汉译与《英汉大词典》中所出现的义项并非完全相符，除了主要含义(产生)相符之外，还存在并未出现的义项。此外，也有属于省词翻译的(例 7)。即从译文语境传达的隐含信息中，可明显识别出“产生”的含义，这是一种隐化趋势。具体原文译文对比如下：

1) We have, however, produced one set (as suppressors of FC 7) using acridine yellow as a mutagen.

不过，我们用吖啶黄作为诱变剂得到了一组突变体(作为 FC 7 突变体的抑制子)。

2) The repeat distance of this form has been found to vary with the method of preparation and the different varieties produced have been further designated FLS—I, II, etc., depending on the spacing observed.

人们发现这种形态的重复距离随着制备的方法不同而改变，根据观察到的间隔，将这些不同的产物种类进一步定义为纤维长间距 I 型、II 型等。

3) Thus, the *α*A gene produces *α*A chains, the *α*G gene produces *α*G chains with lysine in place of asparagine in peptide A-3, and so on.

因此，基因 *α*A 决定 *α*A 链，基因 *α*G 决定 *α*G 链以及肽段 A-3 上的赖氨酸残基取代天冬酰胺残基等。

4) As we shall see, this involves the collection of many more observations and the production of three or four different isomorphous replacements of the same unit cell, a requirement which presents great technical difficulties in most proteins.

就如同我们即将看到的，这将涉及到更多观测结果的收集与同一晶胞的 3~4 种同晶置换晶体的制备支持，上述需求对于大多数蛋白质来说意味着巨大的技术困难。

5) Unusual plant hybrids, thought to be produced by double chromosome reduction, had been reported before but definitive proof of their origin was lacking.

曾有报道认为一些异常杂种植株可能是由染色体双减数引起的，但是缺乏关于其起源的确切证据。

6) The RNA polymerase which produces the new plus strands can rotate around this cyclic primer producing a long chain of plus strand, which is repetitive in sequence.

合成新正链的 RNA 聚合酶能够绕着这种环状的引物链合成一条在序列上重复的很长的正链。

7) Apparently, heating single-stranded virus RNA in this manner produces structural or configurational changes in the molecule which are sufficient to greatly reduce its sedimentation coefficient.

很明显，用这种方法加热单链病毒 RNA 可以改变其原来的分子结构或者构型，并足以降低其沉降系数。

“产生”一词在原文中主要对应 produce，原文检索 produce 有 90 次，译文检索产生共 139 次，可以得出，produce 的译文词语有缩减的趋势，且明显少于“产生”所对应的原文词语。汉译时，译文词语变

化程度低于英文原文，词语重复率更高，表达方式更单一。这与 Olohan 和 Baker [8]、Laviosa [9]的研究成果相符合，体现了文本简化的特征。此外，省译也是隐含意义或关联含义的表现之一，需要充分考虑词语所处的句子环境。翻译的灵活性能使译文具有非凡的表现力，科技译文的可欣赏性也得到大幅提升。

4.4. 词语搭配的有限多样性

词语搭配通常有三种情形：一是关键词或节点词和意义相关的有限范围的词项搭配，二是与几个意义不同的词项搭配，三是和某个特定的词项搭配[10]。

动词 produce 可以分别与意义不同的词项搭配，构成不同的含义，属于上述情况中的第二种情形。但其搭配有一定范围限制，搭配对象一般为名词，较为单一。词语搭配存在的有限性使得译文语言更加自然地道，并直接表现为词语搭配的有限多样性。

4.5. 意义相近词语的译文

根据语料统计，produce 的同义词 reproduce，clone，breed，bear 所对应的译文词语如下：

- clone 在原文中出现 11 次，对应的译文有 7 处“克隆”，3 处为“产生”，1 处省略。
- reproduce 在原文中出现 3 处，对应的译文 2 处为“复制”，1 处为“展示”。
- breed 在原文中出现 2 次，对应的译文均为“繁殖”。
- bear 在原文中出现 1 次，对应的译文为“联系”。

统计结果显示，produce 与 reproduce，clone，breed，bear 虽互为同义词，但译文选词各不相同。而高质量的译文正是因为原文中相似的词语在译文中的良好运用和处理，表现出原文的精髓。

5. 结语

本研究借助于机器辅助翻译工具、语言分析软件等进行双语平行语料库的制作、句对检索，从英汉和汉英对应词语两方面着手，通过定性和定量分析，对词语翻译特色进行研究。旨在发现高水准科技译文的具体表现。但所建的语料库文本容量不够大，涵盖领域有限。研究发现译文高水准在于其翻译选词的显化、隐化及简化特点，词语搭配存在有限性和多样性以及对意义相近词语的灵活运用和处理等，可以从以上方面入手处理文本的汉译以促进中外科技交流。

参考文献

[1] 张支南, 肖亦天. 中国社会科学英文学术期刊高质量发展探究[J]. 科技与出版, 2022(7): 68-73.

[2] 杨言, 胡翠娥. 学术研究型深度翻译: 陈荣捷《老子》英译研究[J]. 外语学刊, 2022(4): 72-77.

[3] 韩子满, 钱虹. 当代中国翻译理论国际传播: 现状与展望[J]. 中国翻译, 2021, 42(6): 103-110+192.

[4] 路甬祥. 《自然》百年科学经典第四卷[M]. 北京: 外语教学与研究出版社, 2016.

[5] 庞双子, 王克非. 翻译文本特征和语言接触研究的进展[J]. 外语与外语教学, 2021(6): 100-108+149-150.

[6] 吕兆杰. 《洛阳伽蓝记(汉英对照)》高频词译介对比——基于语料库的双语实证分析[J]. 中国科技翻译, 2022, 35(2): 57-60.

[7] 石秀文, 管新潮. 基于语料库的汉英词语的翻译特色研究[J]. 上海翻译, 2015(4): 80-84.

[8] Olohan, M. and Baker, M. (2000) Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation. *Across Languages & Cultures*, **1**, 141-158. https://doi.org/10.1556/Acr.1.2000.2.1

[9] Laviosa, S. (2002) Corpus-Based Translation Studies. Rodopi, Amsterdam. https://doi.org/10.1163/9789004485907

[10] 卫乃兴. 词语搭配的界定与研究体系[M]. 上海: 上海交通大学出版社, 2002.