

基于语料库工具的《2020中国航运白皮书》 英译本语言特征研究

蒋文静

上海海事大学外国语学院, 上海

收稿日期: 2022年9月19日; 录用日期: 2022年10月14日; 发布日期: 2022年10月25日

摘 要

中国作为航运大国, 近年来在航运建设各方面取得了瞩目的成就, 而《中国航运白皮书》是对我国航运业发展及成就的年度系统总结, 其英译本的重要程度不言而喻。因此, 文章尝试运用语料库工具Antconc和ABBY Aligner, 对2020年《中国航运白皮书》的英译本进行翻译评估, 分别从词汇和句子层面具体分析中国航运白皮书的语言特征。通过分析对比, 发现英译本具有以下特征: 词汇密度较低, 高频词中实义词占比较大, 动词名词化现象显著, 汉语中的长句在英译本中常被分为较短的简单句, 汉语无主句在翻译时多添加对应成分作为主语等。

关键词

《中国航运白皮书》, 语料库工具, 英译本, 语言特征

A Corpus-Based Study on Linguistic Features of the English Version of the *China Shipping White Paper* in 2020

Wenjing Jiang

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Sep. 19th, 2022; accepted: Oct. 14th, 2022; published: Oct. 25th, 2022

Abstract

As an influential shipping country, China has made remarkable achievements in various aspects

of shipping construction in recent years. *China Shipping White Paper* is an annual systematic summary of the development and achievements of China's shipping industry, so it is self-evident that its English translation is quite significant. Therefore, it attempts to use the corpus tools Antconc and ABBYY Aligner to evaluate the English version of *China Shipping White Paper* in 2020, and analyzes the linguistic features of *China Shipping White Paper* from the lexical and sentence levels respectively. Through comparison and analysis, it is founded that the English version has the following characteristics: low lexical density, big proportion of content words in high-frequency words, numerous nominalization of verbs, long sentences in Chinese are often divided into short simple sentences in English, and the addition of subjects in the Chinese non-subject sentences.

Keywords

China Shipping White Paper, Corpus Tools, English Version, Linguistic Features

Copyright © 2022 by author(s) and Hans Publishers 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

1. 引言

《中国航运白皮书》(又称《中国航运发展报告》)是由交通运输部水运局和上海航运交易所组成的编委会编写,全面涵盖了国际和国内航运、港口服务、集装箱运输、水运基础设施建设与维护、航运服务业、海事监管与搜救等内容,是中国关于航运业发展及成就的年度系统总结,是对部分现行法律、法规 and 政策的解释。《航运白皮书》已连续出版发行 23 年,是中国向国内外宣传航运政策、交流行业信息的重要文件资料,已成为相关的政府部门和企业了解中国航运的主要信息渠道。如今,中国已经成为世界上具有影响力的航运大国,正在稳步开启建设航运强国的新征程,《航运白皮书》英译本的重要性不言而喻。

本文将尝试采用语料库的研究方法,分别从词汇、句子两方面分析 2020 年《航运白皮书》英译本的语言特征。首先运用语料库工具 AntConc 从单词检索、词频分析以及关键词三方面发现英译本在词汇使用上的特征,同时继续利用平行语料库软件 ABBYY Aligner 实现中英双语句对齐,从而发现该英译本在句子方面的特征,挖掘此类英译本的翻译规律和特点。

2. 文献综述

鉴于与传统人工研究方法相比,语料库具有明显优势,可以快速对大量语料进行分析,且借助语料库开展翻译研究可信度较高,自上世纪 90 年代以来,很多学者专心于翻译语料库的研究。Mona Baker [1] 是主张运用语料库来探究翻译现象的第一人,倡导对翻译文本本质的实证研究,开辟了基于语料库的翻译研究之路,旨在描述翻译的性质,揭示翻译中可能的模式或规律,为翻译操作提供范式[2]。

在语料库翻译学这一全新译学研究范式确立后,国内外学者越来越关注翻译共性和具体语言对翻译语言特征在内的研究。如 Laviosa [3] 的研究表明,与未译文本相比,译文的词汇范围较窄,语法虚词和实词所占比例较低,高频词所占比例高于低频词,平均句子长度较短,且译文表现出简化倾向。Olohan [4] 考察了翻译文本中的非强制性显化现象,主要以翻译英语语料库和英国国家语料库为研究平台。结果表明,句法成分和非强制性词汇在翻译文本中常被用来表达原文的隐含信息和句法关系,所以译文呈现出

较为明显的显化趋势。在语料库翻译研究领域,国内也有一些重要研究成果,如肖忠华、戴光荣[5]运用中国英语学习者语料库,对被动句型的习得情况进行了分析,能够对第二语言习得的研究产生启发。秦洪武、王克非[6]基于英汉平行语料库,描述分析了“so...that”结构和它的汉语对应结构。语料库方法在政治语篇翻译评估中的应用性研究也产生一些成果,例如陈建生和崔亚妮[7]将美国《国情咨文》的词汇特征与《政府工作报告》的英译本进行对比,从而发现了两者的差异,并对未来《政府工作报告》的翻译工作提出了一些建议。

《航运白皮书》是中国在航运领域的重要文件,对中国在国际社会上的贸易往来、外交合作甚至国家形象都有不可小觑的影响,英译本的主要作用是向国际社会介绍中国在航运各领域的发展建设情况。然而,通过文献搜集发现,很少有对《航运白皮书》的翻译进行评估性研究,且目前尚未有学者基于语料库来研究《航运白皮书》的翻译。

3. 语料库工具与研究方法

研究中使用了两种语料库工具, AntConc 和 ABBYY Aligner。

AntConc 是一种语料库检索软件,可以通过词语检索、生成词表、主题词、搭配和词频研究等功能,对文本进行分析和研究。

ABBYY Aligner 是一款计算机翻译辅助软件,通过建立原文本与目标文本在句子层面对齐的平行语料库,实现双语文本的对齐,进而通过比较分析,可以达到研究语言现象,发现语言规律的目的。

4. 研究结果与讨论

文章将从词汇和句子两个层面展开分析,通过 AntConc 软件统计,得出 2020 年《航运白皮书》正文部分中文原文共 52,200 字,英文译本共 35,236 词。

4.1. 词汇层面

4.1.1. 词频

Sinclair [8]表明,语料库中出现频率最高的词呈稳定分布,所以,在排列顺序上的任何显著变化都可能有意义。在语料库研究方法中,词频统计是基本手段之一。通常在一个语料库中,绝大部分的高频词都是虚词(或称作功能词),如果出现实词,则说明该实词所代表的人、动作或事物等是语料关注的重点,因而会被反复提及(见表 1)。

通过观察,2020 年中国航运的重点可以从这些出现频率最高的前 20 个实义词中发现,如“port”(港口)、“transport”(运输)、“container”(集装箱)和“construction”(建设)等。

4.1.2. 以“Transport”为例进行词语检索

检索发现“transport”在文本中总共出现了 150 次,在语料中所占比例为 0.43%。以下为“transport”在文本中的部分原文出处:

- a. making new progress in the water transport “chapter”
- b. in the building of a strong transport nation
- c. the water transport infrastructure kept improving
- d. passenger transport volume dwindled rapidly owing to the coronavirus epidemic
- e. China has forged ahead with water transport construction projects in poor areas
- f. foster a system for high-quality maritime transport growth

Table 1. Top 20 Content Words in the English version of *Shipping White Paper* in 2020
表1. 2020年《航运白皮书》英译本出现频率最高的20个实词

Rank	Frequency	Word
9	265	China
12	236	shipping
13	227	ships
14	219	million
16	212	port
18	203	transport
19	202	year
20	197	ports
21	191	ship
22	187	MOT
23	185	river
29	121	total
30	118	water
31	114	container
32	113	domestic
34	113	trade
35	109	market
36	104	international
37	102	capacity
38	101	construction

g. Coordination between regular pandemic control and water transport guarantee

“transport”后面多与“construction”、“guarantee”、“growth”等词紧密搭配，再观察其前后更大语境，可以看出纵使前路有万般艰难，但中国航运发展仍然稳中求进，统筹做好各项水路运输保障，在各方面取得了令世界瞩目的成绩，继续加快推进交通强国水运篇建设。这些搭配同样可以看出《航运白皮书》英译本用词的精确。

4.1.3. 类符形符比

类符指文本中不同的词形，形符指文本中的词汇总数，而类符形符比(Type/Token Ratio)是指特定语料中类符数与形符数的比值，它可以在一定程度上反映文本中词汇的丰富程度。比值越大，说明该文本使用的不同词汇越多，反之则越少[9]。《航运白皮书》英译本类符形符比如表2所示。

可以发现，《航运白皮书》英译本所使用的不同词汇数量相对较少，词汇密度较低，易于理解。由此可见，在翻译时，主要的翻译准则是简洁、准确，做到可读性强。

4.1.4. 关键词排序

词语的选择能反映语篇的主题。通过提取和分析语篇中频率极高的词或短语，可以明确语篇主题词和表达该主题的词集[9]。要提取出主题词，首先需要找到一个更大的参照语料库，然后将要考察的语篇同这个参照语料库进行比较，提取出语篇中词频差异显著的词，即可生成主题词表[10]。

Table 2. The Type/Token Ratio in the English version of *Shipping White Paper* in 2020**表 2.** 2020 年《航运白皮书》英译本类符形符比

类符	形符	类符形符比
3553	34,404	10.33%

本文选用了由美国布朗大学(Brown University)于 20 世纪 60 年代初建立的 Brown 语料库作为参照语料库,因为它是美国英语的首个文本语料库,取自不同主题的报纸文本、书籍以及政府文件,库容达 100 多万个单词,具有较高的参考价值。以下是对《航运白皮书》进行关键词排序的结果,如表 3。

Table 3. The keyword ranking results in the English version of *Shipping White Paper* in 2020**表 3.** 2020 年《航运白皮书》英译本关键词排序结果

Rank	Frequency	Keyness	Word
1	236	1535.31	shipping
2	265	1526.36	China
3	227	1367.88	ships
4	212	1354.91	port
5	197	1354.31	ports
6	203	1312.08	transport
7	187	1310.14	MOT
8	191	1005.97	ship
9	219	969.65	million
10	185	832.62	river
11	114	737.09	container
12	96	645.35	coastal
13	96	645.35	maritime
14	95	621.03	cargo
15	87	615.06	dwt
16	86	607.98	yangtze
17	113	573.08	domestic
18	96	547.89	tons
19	1648	546.52	and
20	100	534.66	transportation

经关键词检索和分析,可基本了解中国 2020 年航运建设的工作重点,如“ships”(船舶)、“port”(港口)、“transport”(运输)、“river”(内河)以及“container”(集装箱)等。这些关键词能帮助国际社会更直观地了解中国航运建设的重点。

除此之外,从关键词检索中还发现《航运白皮书》英译本中有较多的名词,可能有两方面原因,一是中英之间的双语差异,二是译者所使用的翻译策略。《航运白皮书》中有许多针对中国国情的方针和政策,需要译者对这些具有中国特色的词汇进行翻译,造成英译本中名词数量相对较多,如表 4 所示。

Table 4. The noun vocabulary in the English version of *Shipping White Paper* in 2020**表 4.** 2020 年《航运白皮书》英译本名词词汇

原文	译文
乡村振兴战略	the rural revitalization strategy
新发展格局	a new pattern of development
“五位一体”总体布局	the five-front overall strategy of balanced economic, political, cultural, social, and ecological progress
“四个全面”战略布局	the four-pronged strategy of building a moderately prosperous society in all respects, deepening reform, advancing the law-based governance of the country, and strengthening Party self-governance.

在翻译如表 4 所列的具有中国特色的词汇时，译者需要使用解释性的翻译方法，对特有名词进行适当解释和拓展，使得《航运白皮书》英译本中的名词数量增多。

另外，通过观察发现，动词名词化现象在《航运白皮书》英译本中经常出现，即由动词转化为名词或起名词作用的词。之所以使用这种表达手段，是因为它具有使语篇简洁、正式、客观的功能，能起到衔接作用使信息密集，从而使语篇更简洁紧凑。如表 5 所举示例。

Table 5. The verb nominalization in the English version of *Shipping White Paper* in 2020**表 5.** 2020 年《航运白皮书》英译本动词名词化现象

原文	译文
统筹做好疫情防控和水路运输保障	Coordination between Covid-19 response and waterway transportation
如期完成长江经济带船舶和港口污染突出问题整治	Scheduled completion of rectifications of serious ship/port pollution in the Yangtze River Economic Belt
修订港口经营管理规定	Amendment of the administrative provisions on port operations
不断完善水运工程标准体系	Constant improvement of waterway transport engineering standards

以上句子的英译中，比起字面翻译，将动词“统筹”、“完成”、“修订”、“完善”等词名词化之后，整个句子的翻译显得更加地道、简洁、通顺。

4.2. 句子层面

运用 ABBYY Aligner 软件，以句子为单位，对《航运白皮书》进行中英文对齐。对齐后发现，英文全文有 1520 句，中文全文有 1303 个句子。在句子层面，中英文翻译的一对多现象共有 145 句，占句子总数的 11.13%，由于汉语原文较长，这类句子在翻译过程中采取了拆分翻译的方法，将一个较长的汉语句子用两个及两个以上的英文句子翻译出来。另外，全文有 3 处多对一的译文，占 0.23%，主要是排比句或句群，多句被结合翻译成了一个英文单句。

4.2.1. 中译英一对多的现象

在组织句子时，汉英两种语言的基本结构和机制不同，汉语更注重语义而英语更注重结构。具体而言，汉语是主题突出型机制，多为竹式结构，即以零聚整的开放性结构，而英语是主语突出型机制，多为树式结构，即以整驭零的封闭式结构[11]。因此，在汉译英时，应先确定主语，把中文零散的小句融合为英文中的整句，形成清晰的主谓结构，并在必要时补充主语，如表 6 所示。

Table 6. One Chinese sentence to several English sentences in the English version of *Shipping White Paper* in 2020
表 6. 2020 年《航运白皮书》中译英一对多

原文	译文
2020 年, 全国港口完成货物吞吐量 145.50 亿吨, 比 2019 年增长 4.3%, 其中沿海港口完成 94.80 亿吨, 内河港口完成 50.70 亿吨, 分别比 2019 年增长 3.2% 和 6.4%。	In 2020, China's ports handled a total cargo volume of 14.55 billion tons, up 4.3% from 2019. Within the volume, 9.48 billion tons were achieved by coastal ports, and 5.07 billion tons by river ports, up 3.2% and 6.4% respectively.
据海关统计, 2020 年, 我国煤及褐煤进口 3.04 亿吨, 比 2019 年增长 1.5%, 增速比 2019 年下降 4.9 个百分点; 出口 319 万吨, 比 2019 年下降 47.1%。	Customs statistics indicate that imported coal and lignite amounted to 304 million tons in 2020, up 1.5% from the 2019 volume, but down 4.9 percentage points from the 2019 growth rate. Coal exports recorded 3.19 million tons, a downturn of 47.1% from 2019.
全年新增运力 22 艘、14.25 万载重吨; 提前退出市场 23 艘、4.62 万载重吨; 没有强制报废的船舶。	A total of 22 ships with 142,500 DWT were delivered in the year. 23 ships with 46,200 DWT were withdrawn from the market ahead of time. There was no mandatory scrapping of chemical carriers.

在政论文的汉英翻译中, 由于英语注重句子结构, 所以为了符合英语读者的阅读习惯, 应注意句子的长度, 在适当的地方断句, 同时加入相应的主语进行翻译。上述 145 处一句中文对应多个英文句子的情况, 占未达到一一对应句子总数的 98%, 这表明, 在翻译此类政论文时, 避免根据惯有的中文思维翻译, 应摆脱中文框架的桎梏。

4.2.2. 汉语无主句的处理

“无主句”指没有主语而只有谓语的句子, 在《航运白皮书》中经常可以见到, 这并不影响读者在中文语境下的理解。与中文中的无主句相比, 英文中的无主句要少得多。所以, 中译英时, 要使句子结构完整就务必补充缺少的主语, 而且在对外宣传政策法规时, 加入主语可以强化说话人的语气和态度, 因而在将《航运白皮书》进行双语对齐时, 可以发现, 许多中文原句在英译时都额外添加了句子主语, 如表 7 所示。

Table 7. The processing of Chinese sentences without subject in the English version of *Shipping White Paper* in 2020
表 7. 2020 年《航运白皮书》英译本汉语无主句处理

原文	译文
推进贫困地区水运项目建设, 提升内河基础设施供给能力。	China has forged ahead with water transport construction projects in poor areas, and elevated the capacities of inland river infrastructure.
编制发布水运高质量发展指标, 推进水运领域交通强国试点。	MOT will draw up and publish the development indices for high-quality water transport development, and advance the pilot water-transport projects to contribute to the building of a strong transport country.

统计发现, 《航运白皮书》英译本中添加“交通运输部”或其他相应成分作为无主句主语成分的句子共有 233 处, 在所有英文句子中所占比例为 15.3%, 这表明在英译时应注意处理汉语的无主句, 是否需要添加“中国”、“交通运输部”(MOT)或其他相应成分作为句子主语, 应根据具体语境和文体特征来确定, 从而加强语言表达的态度和立场。

5. 结语

基于语料库的检索与对比, 2020 年《航运白皮书》英译本在词汇和句子层面具有以下特征: 1) 实义词在高频词中所占比重较大, 文本关注的焦点较为明显; 2) 词汇密度较低, 易于理解, 说明译文追求可

读性强；3) 动词名词化现象显著，使语篇更客观、简洁、紧凑、地道；4) 中文中的长句在英译本中通常被分割成较短的单句，强调译文的句子结构；5) 无主句的英译往往会添加上“中国”、“交通运输部”或其他对应成分作主语。

通过观察和分析，发现译者在翻译时力求与原文文体特征保持一致，用词严谨，注重表达的精确性；翻译时在保持原文意思的基础上，遵循简洁原则；注意译文句子结构，多使用简单句表达汉语中的长句；注重名词化运用，提高翻译地道性；尽可能贴近英语读者的阅读习惯，如在修饰名词时多使用定语后置，达到对外宣传的目的。

在建设海洋强国、航运强国的新征程上，未来中国将受到世界上越来越多的关注，《航运白皮书》英译本起着举足轻重的作用。因此，译者必须考虑到目标读者的思维模式和语言表达习惯，同时力求忠实于原文，采用恰当的翻译策略，以达到翻译文本的预期目的和交际功能。

参考文献

- [1] Baker, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *TARGET*, 7, 223-243. <https://doi.org/10.1075/target.7.2.03bak>
- [2] 赵晴. 国外基于语料库翻译研究述评[J]. 重庆交通大学学报, 2008, 8(3): 100-104.
- [3] Laviosa, S. (2002) Corpus-Based Translation Studies: Theory, Findings, Applications. Brill Rodopi, Amsterdam. <https://doi.org/10.1163/9789004485907>
- [4] Olohan, M. (2001) Spelling out the Optionals in Translation: A Corpus Study. *UCREL Technical Papers*, No. 13, 423-432.
- [5] 肖忠华, 戴光荣. 语料库在语言教学中的运用——中国英语学习者被动句式习得个案研究[J]. 浙江大学学报, 2010(4): 189-200.
- [6] 秦洪武, 王克非. 基于语料库的翻译语言分析——以“so...that”的汉语对应结构为例[J]. 现代外语, 2004(1): 40-48.
- [7] 陈建生, 崔亚妮. 基于语料库的中国政府工作报告英译本词汇特征研究[J]. 当代外语研究, 2010(6): 39-43.
- [8] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press, New York.
- [9] 杨惠中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2002.
- [10] 张海云, 谢群芳. 基于语料库的文学作品检索分析——以越南中篇小说《志飘》为例[J]. 解放军外国语学院学报, 2010, 33(3): 57-61.
- [11] 潘文国. 汉英语对比纲要[M]. 北京: 北京语言文化大学出版社, 1997.