

2022年《政府工作报告》的英译研究

邹 徽

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2023年8月4日; 录用日期: 2023年9月13日; 发布日期: 2023年9月28日

摘 要

中国政治文本的英译是构建和传播中国形象的重要途径,其他国可以凭此正确了解中国国情及发展状况。本文借助语料库的方法,通过将2022年《政府工作报告》英译文与其源语文本及同时期的美国《国情咨文》进行对比,从词汇、句子层面分析英译文的语言特征及其原因,旨在为政治文本英译提供些许参考。通过对比研究发现,英译文的词汇丰富度较低,词汇密度较高,平均词长和平均句长较长,名词数及总词数增多,而造成以上结果主要是由于英译中的名词化现象和显化现象。

关键词

语料库, 《政府工作报告》, 语言特征

A Study on the English Translation of the 2022 *Government Work Report*

Wei Zou

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Aug. 4th, 2023; accepted: Sep. 13th, 2023; published: Sep. 28th, 2023

Abstract

The English translation of Chinese political texts is an important way to construct and disseminate the image of China, and it is also one of the ways for the international community to understand China correctly. By comparing the English translation of the 2022 *Government Work Report* with its source language and *The State of the Union* of the United States in the same period, this paper analyzes the linguistic features and reasons of the English translation from the lexical and sentence levels, aiming to provide some references for the English translation of political texts. Through comparative study, it is found that the English translation has lower lexical richness, higher lexical density, longer average word length and sentence length, and increased noun num-

ber and total word number, which are mainly due to the phenomenon of nominalization and manifestation in English translation.

Keywords

Corpus, Government Work Report, Linguistic Feature

Copyright © 2023 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

随着改革开放和全球一体化的加速,中国越来越积极地参与国际事务。同时,中国与其他国家的关系也日益紧密。政治文本的翻译对于提升国家的国际形象和维持国家主权至关重要,它为外国人了解中国的指导方针、政策以及在政治和民主等领域取得的巨大进步提供了一个窗口。正因如此,政论文翻译研究成为研究的热点。不仅中央政府,地方政府也意识到了政论文翻译的重要性和必要性。然而,目前大多数政论文翻译研究主要基于翻译理论和译者经验,缺乏结合客观数据的全面、有说服力的结论。本文选取2022年《政府工作报告》及其英译文为研究对象,借助语料库软件,参照2022年美国《国情咨文》,以得出当下政治文本英译的语言特征。《政府工作报告》和《国情咨文》分别是中美政府的施政纲要,这两份报告均覆盖经济、政治等多个领域的内容[1]。因此,本文选择同一时期的《国情咨文》作为参照。

2. 《政府工作报告》英译的相关研究

政府工作报告,作为最为重要的政治文件之一,在每年的人民代表大会和中国政治协商会议上由各级政府发表。总体来看,它包含三个主要部分:对过去一年政府工作的综述;下一年的工作目标和计划;以及政府职能、民主建设和法制建设的完善。政府工作报告象征着民主执政,使人们了解政府过去所做的努力和未来的目标。

考虑到政论文在国与国交往中的关键作用,有不少学者从多个角度对《政府工作报告》的英文翻译进行了研究。在翻译理论方面,李红霞以2010年中英文版《政府工作报告》为语料,从功能目的论出发,探讨政论文的文本功能及翻译目的[2]。胡峰笙等人以《2010年政府工作报告》为实例,从语篇衔接理论出发,分析了语篇衔接理论在政治文献翻译中的实用性和有效性[3]。在翻译问题方面,武光军从句子和语篇两个层面分析了2010年政府工作报告英译本中的迁移性冗余[4]。陆仲飞和王大伟研究《2016政府工作报告》英译文用词的灵活性,译文在多种情况下小幅度地偏离原文,以便更好地被目标语读者所接受[5]。此外,还有一些学者对《政府工作报告中》某类词的英译进行了研究,如田绪军研究了情态动词[6]、汪晓莉和谭苗研究了虚化动词[7]。

3. 研究问题

本文采用基于语料库的方法,结合客观数据,旨在探索以下问题。

- 1) 2022年《政府工作报告》英译本有何语言特征?
- 2) 构成这些特征的原因是什么?

4. 研究设计

本文选取了 2022 年《政府工作报告》的原文、英译文以及 2022 年美国《国情咨文》作为研究材料，采用基于语料库的方法，在定量分析的基础上，结合汉语和英语的特点以及汉英翻译技巧做一些定性分析。在语料库语言学的帮助下，对《政府工作报告》译文语言特征的描述将会更加科学和客观。

本文选用的研究工具有 SegmentAnt 软件、TreeTagger3.0 软件以及 WordSmith6.0 软件。SegmentAnt 软件可用于中文文本的分词及词性标注。TreeTagger3.0 软件可用于对英文文本进行词性标注。WordSmith6.0 是由 MikeScott 开发的一款语料库软件，通过该软件我们可以获得类符/形符比、词长分布、句子数量、平均句长、高频词等语言参数。

研究步骤可分为数据处理和数据分析两个阶段。首先，数据处理阶段。1) 在新华网(<http://www.xinhuanet.com>)获取 2022 年《政府工作报告》原文(以下简称 RWGC)及英译文(以下简称 RWGE)，在美国白宫官网(<https://www.whitehouse.gov/>)获取《国情咨文》(以下简称 SUA)。2) 对文本进行清洗，再以 txt 形式保存在 D 盘的文件夹中。3) 使用 TreeTagger3.0 软件对英文文本进行标注，注意确认编码为 UTF-8。4) 使用 Segment Ant 软件对中文文本进行分词及标注。5) 因为 WordSmith6.0 只认 Unicode 和 UTF-16LE 编码，故将文本编码先都改为 UTF-16LE，再通过 WordSmith6.0 中的 Wordlist 功能获取三个语料库(RWGC/RWGE/SUA)的语言参数。其次，数据分析阶段。对呈现的语言参数进行分析，总结出 2022 年《政府工作报告》的语言特征，挖掘构成这些特征的原因。

5. 发现与讨论

5.1. 词汇丰富度(类符/形符比)

词语丰富度是指相同长度的语料中不同的词语数量，可通过语料库的类符/形符比(TTR)来衡量[8]。形符是指出现在一个文本中的单词的数量，类符是指不同单词类型的数量。如果一个单词在同一文本中出现两次或两次以上，只能算作一个类符。TTR 的数值越高，表明文本的词汇丰富度越高，文本的阅读难度越大；反之，词汇丰富度越低，文本的阅读难度越小[9]。但是 TTR 值可能会受到文本长度或语料库容量的影响[10]。当扩充一个较短的文本时，文本的形符数会明显增加，但类符数可能不会增加。因此，Scott 由此提出“标准类符/形符比”(STTR)的概念，STTR 计算的是每 1000 个单词的平均 TTR 值[11]。本文的三个语料库的标准类符/形符比如表 1 示。

Table 1. Lexical richness

表 1. 词汇丰富度

	RWGC 语料库	RWGE 语料库	SUA 语料库
形符	7404	14062	8066
类符	2433	2359	1839
形符/类符比	33.17	16.92	23.04
标准类符/形符比	57.97	42.76	43.5

就篇幅而言，通过观察形符数和类符数，我们可以得知 RWGE 语料库的文本长度远高于 RWGC 语料库和 SUA 语料库。RWGE 语料库中类符数也远高于 SUA 语料库。而类符数更多并不意味着该文本的词汇更丰富，因为文本长度是导致类符数更多的重要原因。反观标准类符/形符比，前者的标准类符/形符比要略小于后者。这些表明，2022《政府工作报告》英译本中的词汇丰富度低于 2022 美国《国情咨文》。这是由于外国人用词更为灵活，常用不同词汇表达同一意思。其次，我国政治文本的英译要求严谨、忠

实原文，因而造成许多词汇及表达的重复使用，降低了词汇的丰富度。

5.2. 词汇密度

词汇密度是一个文本中实词数的占比。词汇密度可以用来衡量文本信息量和阅读难度。词汇密度越大，文章越难理解[12]。在英语中，单词可以根据它们所表达的意思分为实义词和功能词。表达语法意义的词，如连词、介词、代词和冠词等都是功能词。那些有词汇意义的，也就是那些指物质、行为和性质的，如名词、动词、形容词和副词都是实义词。在本研究中，在用词性赋码软件标注后，根据赋码集，在 Word 中检索三个语料库中的所有实义词，最后计算两个语料库的词汇密度，结果如表 2 所示。

Table 2. Lexical density

表 2. 词汇密度

	RWGC 语料库	RWGE 语料库	SUA 语料库
名词	2771	4431	2423
动词	3271	2113	1410
形容词	756	1776	668
副词	229	359	410
实义词	7027	8679	4911
总词数	7404	14062	8066
词汇密度	94.91%	61.72%	60.89%

从表 2 可以看出，2022 年《政府工作报告》中文版的词汇密度大于英文版，而《政府报告》英文版略大于《国情咨文》。这是由于汉语和英语的语言结构及形式的差异。中文重“意合”，英文重“形合”，故而在英文中会频繁出现一些连接词、介词等功能词。因此，中文文本的词汇密度会高于英文文本。《政府报告》英文版与《国情咨文》的词汇密度相差不大，说明该英译符合英文的使用习惯。

通过对表 2 的观察，发现在《政府报告》中动词数要大于名词数。而在英译本中，名词要远多于动词。这主要是由于显化现象和动词的名词化现象。在此我们先探讨名词化现象。在英语中，动词名词化是常见的，并且是词汇丰富和表达灵活的重要手段。动词的名词化通常涉及特定的词缀，如：-tion, -sion, -al, -ment, -y, -ing 等。通过 WordSmith 的 Concord 功能检索到《政府报告》英译文中大量词缀为-tion, -ment、-sion 的动词名词化词汇，如 development/employment/construction/implementation/expansion/cooperation 等。动词名词化可以使语篇正式、客观和简洁。此外，它还可以作为语篇中的衔接手段。具体来说，名词化在语篇中使用得越多，语篇就越抽象，越正式。名词化有利于表达客观事实。它使信息更加密集，从而使文本更加简洁，更有凝聚力[13]。因此，在《政府工作报告》中我们能发现大量名词化词汇。

5.3. 词长分布

平均词长是指一段文本中，每个词的平均字母字数，是通过一篇文本的总字母数与总词数相除得出。研究平均词长可以帮助我们了解文本的语言风格和文本特征。较短的平均词长通常反映着简洁、流畅的表达，适合快速传递信息；而较长的平均词长则可能表明复杂、深入的探讨，语体更为正式，适合学术等领域。例如，学术类文本的平均词长会大于报刊类文本，报刊文本平均词长又会大于各类小说；而语体正式程度也在依次递减。因此，语体正式程度与平均词长也有着一定的联系[14]。表 3 显示了 2022 年《政府工作报告》英文版和美国《国情咨文》的平均词长和词长标准差，图 1、图 2 显示了各自的词长分布。

Table 3. Average word length
表 3. 平均词长

	RWGE 语料库	SUA 语料库
平均词长	5.47	4.6
词长标准差	2.94	2.45

2022年政府工作报告(英文版).lst

File	Edit	View	Compute	Settings	Windows	Help														
N	1-letter words	2-letter words	3-letter words	4-letter words	5-letter words	6-letter words	7-letter words	8-letter words	9-letter words	10-letter words	11-letter words	12-letter words	13-letter words	14-letter words	15-letter words	16-letter words	17-letter words			
1	191	2134	2530	1708	1248	1258	1396	1173	773	751	509	203	113	51	13	9	2			

Figure 1. Word length distribution in RWGE

图 1. RWGE 词长分布

美国国情咨文.lst

File	Edit	View	Compute	Settings	Windows	Help														
N	1-letter words	2-letter words	3-letter words	4-letter words	5-letter words	6-letter words	7-letter words	8-letter words	9-letter words	10-letter words	11-letter words	12-letter words	13-letter words	14-letter words	15-letter words	16-letter words	17-letter words			
1	224	1492	1558	1426	876	644	745	470	296	176	81	30	25	18	2	2	1			

Figure 2. Word length distribution in SUA

图 2. SUA 词长分布

在书面语言中，尤其是正式或官方文件倾向于使用复杂而长的词语。较长的平均词长反映了文本的正式程度。表 3 中的数据显示 RWGE 语料库的平均词长要明显高于 SUA 语料库，这表明 RWGE 比 SUA 使用的长词更多。通过图 1、图 2 可以发现，SUA 中长词数很少，这也意味着《政府工作报告》要比《国情咨文》的语体更为正式。因为《政府工作报告》本身是一个专业性强的政治文本，所以该译文也保留了原文的特色。在上节中讲到了《政府工作报告》中的名词化现象。名词化的词通常比其他形式的名词长，因此这也是 RWGE 平均词长高于 SUA 的原因之一。

5.4. 高频词分布

词频是语料库研究的基本方法。高频词是在语料库中出现频率较高的单词，它们在自然语言中扮演着重要角色。通过分析高频词的使用，我们可以揭示语言的基本结构和模式。利用 WordSmith6.0 软件的 WordList 功能，很容易得到按降序排列的词频列表。表 4 显示了这三个语料库中出现频率最高的前 20 个单词及其频率，以及分别在全文中的百分比。

Table 4. High-frequency words

表 4. 高频词

	RWGC	频率	占比%	RWGE	频率	占比	SUA	频率	占比
1	和	191	2.58	and	1000	7.11	the	367	4.55
2	的	133	1.80	the	702	4.99	and	314	3.89
3	发展	127	1.72	of	506	3.60	to	272	3.37
4	加强	68	0.92	to	430	3.06	of	187	2.32
5	建设	68	0.92	we	410	2.92	a	138	1.71
6	推进	67	0.90	will	376	2.67	in	138	1.71

Continued

7	要	65	0.88	in	276	1.96	we	115	1.43
8	支持	59	0.80	for	175	1.24	that	82	1.02
9	等	55	0.74	a	147	1.05	for	76	0.94
10	政策	45	0.61	development	126	0.90	I	72	0.89
11	经济	44	0.59	on	99	0.70	you	69	0.86
12	企业	43	0.58	with	74	0.53	our	68	0.84
13	推动	43	0.57	that	71	0.50	it	67	0.83
14	加快	38	0.51	be	67	0.48	is	61	0.76
15	促进	36	0.49	all	65	0.46	with	57	0.71
16	新	36	0.49	as	65	0.46	have	51	0.63
17	改革	35	0.47	new	64	0.46	more	49	0.61
18	实施	34	0.46	government	63	0.45	on	46	0.57
19	就业	34	0.46	more	63	0.45	this	43	0.53
20	服务	34	0.46	our	62	0.44	America	41	0.51
合计		1255	16.95		4841	34.43		2313	28.68

通过观察表 4 的数据可以得出以下发现：(1) RWGC 语料库的前 20 个词中仅有 2 个是虚词，实词占 80%，而 RWGE 和 SUA 中都仅有 2 到 3 个实词，故而可以应验前文所说的中文文本中的虚词要远少于英文文本，中文文本词汇密度往往更高。(2) 源语文本中“和”字的占比为 2.58%，而其英译本中“and”占比高达 7.11%，之所以这数值翻倍增长，正是由于英文更注重语篇的衔接。因此也印证了汉语重“意合”，英语重“形合”这一说法。(3) 在 RWGE 中出现频率第 3 的实词“government”在对应的源语文本中出现的频率并不算很高，这是由于中文无主句句式英译时可能需要添加主语。(4) 在 SUA 的前 20 个高频词中出现“I”和“You”，而 RWGE 前 20 个高频词中并未出现，表明《政府工作报告》比《国情咨文》的语体正式。

5.5. 平均句长

平均句长是指一个句子中的平均词数，通常是用文本中的总词数/该文本的总句数来计算。标准差反映了数据的集中程度和离散程度，标准差越大，文本的句子长度变化越大。语料库中的平均句长可以在一定程度上反映句子的复杂程度。一般来说，句子越长，就越难理解，文本也就越正式。表 5 列出了三个语料库的句子数量、平均句长和句长标准差。

Table 5. Average sentence length

表 5. 平均句长

	RWGC 语料库	RWGE 语料库	SUA 语料库
句数	525	607	516
平均句长	13.97	22.96	15.47
标准差	11.27	15.27	10.86

通过表 5 可知，(1) RWGE 的句数要多于 RWGC，这是因为汉语句子的特点是连续句子结构；而英

语句子的特点是树型结构，它有一个独立的句子来表达重要信息，有许多从属分句和短语来表达次要信息。在翻译 RWGC 中的长句时，需要分析整个句子所表达的所有意义，确定它们之间的主次关系，然后将原句分离成几个句子。(2) RWGE 的平均句长要多于 RWGC，这是由于汉语强调“意合”，英语强调“形合”，英文文本常用一些连词来反映句子之间的逻辑关系。因此译者在翻译具有逻辑关系的句子时经常需要使用连接词将两个句子合并成一个句子，这也使得 RWGE 的平均句长更长。(3) RWGE 语料库的平均句长大于 SUA 语料库。导致这一结果的原因是因为受到源语文本的影响。在《政府工作报告》中经常会出现一些专业术语、中国特色词汇以及缩略词，因此要在英译时需要对这些词进行解释，也就是下文要讲的英译中的显化现象。

5.6. 显化

显化现象是一种语言的表达手段，旨在将抽象的概念具象化，使其可以通过语言传达给他人。显化现象分为形式显化和意义显化，可以体现在连词和解释性词汇的使用上[15]。通过表 4 可以看到，“和”与“and”分别是政府工作报告及其英译本中出现频率最高的词，但值得注意的是，英译文本中“and”的出现频率(7.11%)要远远高于源语文本(2.58%)。而且，表 1 的结果显示 RWGE 中的总字数多于 RWGC，表 2 显示译文的名词数量远高于源语文本，表 5 显示 RWGE 中的平均句长多于 RWGC。以上结果都表明在 2022 年《政府工作报告》的英译上存在显化现象。因为汉语不需要连词、代词等也能理解句子的内在关系。而英语中的逻辑关系必须通过连词、代词和介词来表达完整的意思，因此便出现了形式显化。在 2022《政府工作报告》的英译过程中，译者通过增加连词和人称代词，使得翻译句子的意思更完整，更容易理解。例如：提升科技创新能力。实施基础研究十年规划，加强长期稳定支持，提高基础研究经费占全社会研发经费比重。

We will raise our capacity for scientific and technological innovation. We will press ahead with the 10-year action plan on basic research to ensure stable support over the long term, and increase the share of basic research spending in the country's total R&D expenditures.

意义显化是指汉语中的一些成语、谚语等词，在英语中需要用更多的词来解释其真实意义。在 2022 年《政府工作报告》必然会有许多中国特色词及缩略词等，因此在英译过程中也固会出现意义显化现象。如：“放管服”改革——the reform to streamline administration and delegate power, improve regulation, and upgrade services; 纠治“四风”——tackle pointless formalities, bureaucratism, hedonism and extravagance; “十四五”规划——14th Five-Year Plan; 坚持“两个毫不动摇”——work unswervingly both to consolidate and develop the public sector and to encourage, support and guide development of the non-public sector 等。

6. 结论

本文以语料库为辅助工具，从词汇、句子两方面对 2022 年《政府工作报告》原文(RWGC)、英译文(RWGE)和同期的美国《国情咨文》(SUA)进行了全面的定量和定性分析。在词汇层面讨论了词汇丰富度、词汇密度、词长、高频词和名词化现象；在句子层面对句数、平均句长、显化现象进行了讨论。经过对比研究，发现 2022 年《政府工作报告》英译本有以下几个特征：第一，译本词汇丰富度较低，这是由于政治文本翻译要求较高的准确度，使得译者翻译原文中的一些词时倾向于重复使用一些英语单词，而本族语者则在语篇中使用不同的单词或短语来表达相同的意思。第二，因为 RWGC 中有许多缩写，译者需要使用更多的词汇来传达更多的信息，因此译文相对于 SUA 具有较高的词汇密度。这结果也表明词汇密度高不一定意味着词汇更丰富。第三，由于名词化现象使得译文中的名词数显著增高，平均词长更长，文本更为正式。第四，由于汉语是分析性语言，不受句子结构的限制，所以 RWGC 中有大量的无主语句

子, 在英译时大部分是通过加主语“we”和“government”翻译出来的, 这也导致了 RWGE 中这两词使用频率的增加。第五, 由于政治文本的特性及英译的显化现象, 使得译文中使用“and”的频率相较于原文显著增高, 总词数增多, 平均句长更长。

参考文献

- [1] 李晓倩, 胡开宝. 中国政府工作报告英译文中主题词及其搭配研究[J]. 中国外语, 2017, 14(6): 81-89.
- [2] 李红霞. 目的论视域下的政论文英译策略研究——以 2010 年《政府工作报告》为例[J]. 外国语文, 2010, 26(5): 85-88.
- [3] 胡峰笙, 荆博, 李欣. 语篇衔接理论在政治文献翻译中的应用——以《2010 年政府工作报告》为例[J]. 外语学刊, 2012(2): 89-91.
- [4] 武光军. 2010 年政府工作报告英译本中的迁移性冗余: 分析与对策[J]. 中国翻译, 2010, 31(6): 64-68.
- [5] 陆仲飞, 王大伟. 汉译英中的灵活用词——《2016 政府工作报告》英译文研究[J]. 上海翻译, 2016(4): 21-27, 93.
- [6] 田绪军. 基于语料库的《政府工作报告》中情态动词英译操作规范研究[J]. 上海翻译, 2022(5): 20-25.
- [7] 汪晓莉, 谭苗. 基于语料库的《政府工作报告》英译本虚化动词应用研究[J]. 邢台学院学报, 2022, 37(3): 116-123, 135.
- [8] 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.
- [9] 于红. 基于语料库的政府公文翻译“简化”趋势考察——以白皮书《2010 年中国的国防》英译文为例[J]. 外语研究, 2016, 33(3): 79-86.
- [10] Baker, M. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. *International Journal of Translation Studies*, 12, 241-266. <https://doi.org/10.1075/target.12.2.04bak>
- [11] Scott, M. (2004) *The Word Smith Tools*. Oxford University Press, Oxford.
- [12] Swann, J., Deumert, A., Lillis, T. and Mesthrie, R. (2004) *A Dictionary of Sociolinguistics*. Edinburgh University Press, Edinburgh.
- [13] 曾庆敏. 英语中的名词化隐喻及其语篇功能[J]. 重庆工学院学报, 2006, 20(3): 125-128.
- [14] 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, 62.
- [15] Baker, M. (1996) Corpus-Based Translation Studies: The Challenges That Lie Ahead. In: Somers, H., Ed., *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, John Benjamins, Amsterdam, 175-186. <https://doi.org/10.1075/btl.18.17bak>