

基于语料库的中美政治文本国家叙事对比分析

——以 2022 年《政府工作报告》英译本和美国《国情咨文》为例

阎 瑾, 邱俊珺

长沙理工大学, 湖南 长沙

收稿日期: 2022年9月18日; 录用日期: 2022年10月13日; 发布日期: 2022年10月20日

摘 要

基于语料库, 以LancsBox等语料库辅助软件为工具, 以2022年《政府工作报告》英译本和美国《国情咨文》为语料, 采用语料库语言学的方法, 从类符/形符比、平均词长、词汇密度、高频词以及关键词等多角度在文本层面进行对比, 分析两国国家叙事的差异及其原因, 总结语言特征, 并对政治文本的翻译提出建议。

关键词

语料库, 政治文本, 国家叙事

A Corpus-Based Comparative Analysis of National Narratives in Chinese and American Political Texts

—Taking the English Version of *the 2022 Government Work Report* and *the US State of the Union* as Examples

Jin Yan, Junjun Qiu

Changsha University of Science & Technology, Changsha Hunan

Received: Sep. 18th, 2022; accepted: Oct. 13th, 2022; published: Oct. 20th, 2022

Abstract

This paper is based on the corpus and adopted a corpus linguistic approach such as LancsBox.

With the English translation of *the 2022 Government Report and the US State of the Union Address* as a corpus, this paper uses a corpus linguistic approach. By comparing the type/token ratio, average word length, lexical density, high-frequency words, and keywords on the textual level, try to analyze the differences between the national narratives of the two countries and their causes, summarize linguistic features, and make suggestions for the translation of political texts.

Keywords

Corpus, Political Texts, National Narratives

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

2022年《政府工作报告》和美国《国情咨文》是体现中美两国国家叙事的重要文本,国家叙事是建构国家形象的重要途径。本文以语料库为工具,通过对2022年《政府工作报告》英译本和美国《国情咨文》两份语料的类符/形符比、平均词长、词汇密度、高频词、关键词以及句法等多角度比较研究,分析两国政府工作文本国家叙事的异同,并对政治文本的翻译提出建议,以期更好地向世界讲述中国故事、传递中国声音。

2. 2022年《政府工作报告》与美国《国情咨文》的概述

近年来,中国国际交流越来越频繁,在世界政治经济生活中的影响日益扩大,在国际舞台上也越来越受到瞩目。政府工作报告是中国政府向人民代表所作的工作总结和工作部署,对国家的发展具有重大意义,同时,也是对外宣传的重要途径,是外界了解中国政治经济文化的权威来源,在一定的程度上代表着中国形象,因此,政府工作报告的翻译对中国与外界的交流和发展起着重要作用。2022年3月5日,国务院总理李克强在第十三届全国人民代表大会第五次会议上报告政府工作,其公文形式也随之在中国政府网上进行发布。2022年《政府工作报告》的主要内容包括一年内工作回顾、当年工作任务、政府自身建设和其他内容,文本具有全面客观、重点突出、富有新意、文风朴实的特点。政府工作报告的英文版是世界了解中国政府工作的重要窗口,也是中国政府向世界进行政治宣传的重要媒介。

2022年3月1日,美国总统拜登在美国国会上对参众两院发表了(*The State of the Union Address*)演说。其根本目的在于最大限度地说服听众接受并支持总统的观点和政治主张,并按照政府预期的方式付诸行动。[1]

3. 语料库研究方法及步骤

本文利用定量和定性结合的研究方法,利用LancsBox对自建的语料库进行分析和探讨,详细步骤如下:第一步,分别从中国政府网和美国白宫政府网获取2022年《政府工作报告》英文版与美国《国情咨文》的全文,以txt格式保存;第二步,利用软件EditPlus对文本中多余的空行与符号进行删除处理;第三步,利用LancsBox软件分别对两份文件进行分析,包括高频词统计及关键词统计。第四步,利用软件LancsBox分析文本可读性,包括平均词长、平均句长、困难词占比等数据。

基于语料库对2022年《政府工作报告》英译本与美国《国情咨文》的文本特征进行对比分析,着重

探究前者的文本特征, 借助量化分析更直观地判断二者在国家叙事方面的异同, 并分析其原因, 进而不断完善政治文本的翻译策略, 彰显中国人民当家作主、自强自信、改革创新、协调发展、实干兴邦、合作共赢的国家形象。

4. 语料分析和讨论

(一) 词汇层面

通过语料库对 2022 年《政府工作报告》英译本与美国《国情咨文》进行分析, 二者的字符数量分别为 13,102 与 7677, 在此基础上结合其他工具对两个文本的类符/形符比、平均词长、词汇密度、高频词等逐一对比分析, 从而得出中美两国国家叙事的异同。

1) 类符/形符比与平均词长。类符(type)指的是文章中所出现的不同形符, 不包含重复计算的形符; 形符(token)指的是所有词, 也称词次。例如, 在 “I wish to wish the wish you wish to wish.” 这一句子中, 类符数为 5 (I, wish, to, the, you), 形符数为 10。通过计算文本的类符/形符比(TTR)可以衡量词汇的丰富程度, TTR 值越高, 说明文本用词越丰富。平均词长(mean word length)是指文本中单词所含字母的平均数量, 一般而言, 这个平均值越大, 说明文本更复杂难懂, 反之则更浅显易懂。在实际的操作中, 可通过 LancsBox 得到文本的类符与形符数, 再进行计算, 结果如见表 1 所示。

Table 1. Type/token and mean word length

表 1. 类符/形符比与平均词长

指标	2022 年《政府工作报告》英译本	2022 年美国《国情咨文》
形符	13102	7677
类符	2530	2195
类符/形符比(%)	19.31	28.60
平均词长	5.81	3.71

见表 1 可以看出, 2022 年《政府工作报告》英译本与美国《国情咨文》的 TTR 指分别为 19.31% 与 28.60%, 说明在词汇丰富程度上前者不如后者, 但平均词长却要高于后者。这些数据表明, 虽然 2022 年《政府工作报告》英译本的词汇丰富程度不如美国《国情咨文》, 但整体的阅读难度却更高。

国家叙事在词汇丰富程度差异主要原因有如下几点: 一是两种语言的结构差异。英语句子像葡萄, 注重文本简洁性, 结构紧凑; 汉语句子像竹子, 注重内容传递, 循序渐进, 因此, 一般情况下一个汉语句子会比一个英语句子要承载更多的内容, 尤其是对于政府工作报告这一类文本, 其中包含大量的长句, 译者会不可避免地译出更多的长难句, 例如, 2022 年《政府工作报告》英译本与美国《国情咨文》中句子平均单词数分别为 21.71 与 15.14。二是撰写人所处的语言环境不同。政府工作报告英译本由母语为中文的人士所撰写, 在英译过程中会不自觉地保留一些中文特征, 传达出中国特色, 而撰写美国《国情咨文》的作者更熟悉英语语言的运用, 用词自然就更丰富, 同时表达更简洁流畅。因此, 在政府工作报告英译中, 译者在进行国家叙事时可以考虑在保留一定中国特色的同时, 适度提高文本的整体可读性。

2) 词汇密度。词汇密度(Lexical Diversity)是计算语言学(Computational Linguistics)中的一个概念, 用于衡量人类交流过程中所使用语言的结构与复杂度(Michael 1985: pp. 41-64)。[2]它可以通过计算语料中词项(Lexical Item)与总词数的比率来获得。简而言之, 词汇密度是指文本实词在总词数中所占的比例。英语实义词指具有稳定词汇意义的词语, 与之相对的是功能词, 指不具备稳定词义或意义模糊而主要起语法功能作用的词语。词汇密度被认为可以反映篇章的信息容量和难度。词汇密度偏高说明该篇章的实

词比例较大, 因而信息容量大, 难度也相应增加。相反, 词汇密度偏低就说明篇章难度相对较低, 容易理解。在实际操作过程中, 分别将 2022 年《政府工作报告》与美国《国情咨文》的全文上传到 LanksBox 软件, 然后在 KWIC 一栏搜索名词、动词、形容词、副词, 所得数据如下。

Table 2. Lexical diversity
表 2. 词汇密度

Lexical Item	2022 年《政府工作报告》英译本	2022 年美国《国情咨文》
名词	4236/13,996 (32.12%)	2432/7677 (31.68%)
动词	2407/13,996 (18.69%)	1419/7677 (18.48%)
形容词	1673/13,996 (12.77%)	653/7677 (8.51%)
副词	395/13,996 (2.62%)	384/7677 (5.00%)

见表 2 可以看出, 就词汇密度而言, 除形容词外, 2022 年《政府工作报告》英译本与美国《国情咨文》内含实词的比重各项差异都不大, 总的实词词汇密度分别为 66.20% 与 63.67%。形容词密度相差较大是由于汉语注重意合, 句法形式区别的手段相对比较单调, 除了少数名词和数量词之外, 大多数的定语都是形容词。相比之下, 英语的句法形式和修饰限定性手段要复杂得多, 除了形容词之外, 名词、介词词组和从句都可以担当定语来修饰限定定语。因此, 形容词的使用在 2022 年《政府工作报告》英译本中呈现较高频率。

2022 年《政府工作报告》与美国《国情咨文》都属于政治类文本, 是一种国家叙事, 是塑造良好大国形象的尝试。2022 年《政府工作报告》会注重传递更多实际信息, 而美国《国情咨文》的演讲人也希望通过此文本达到获取支持的目的, 因此, 不难理解这两份文本都具有较高的词汇密度。同时, 从表 2 数据也可看出两份文本中的名词占比极高, 动词、形容词的名词化现象在正式文本中较为常见, 更多地运用名词可以让文本读来更加客观详实, 而这一特征也十分符合文本的国家叙事要求。

3) 词汇频率。词频统计是语料库研究的一个基本统计手段, Sinclair 指出, 语料库中频率最高的那些词的分布是稳定的, 因此, 如果在排列顺序上出现任何显著的变化都是有意义的。一般情况下, 一个语料库词表中的高频词绝大多数是功能词, 实词的出现则表明该实词所指的人、事物或行动等是语料关注的焦点。[3]

本文通过 LanksBox 的 Words 一栏的功能分别统计了 2022 年《政府工作报告》英译本与美国《国情咨文》中出现频率前 20 的高频词, 通过 LanksBox 的 Ngrams 功能统计了 2022 年《政府工作报告》英译本相对于美国《国情咨文》中出现频率更高的关键词。通过高频词可以看出文本在语言应用上的特点, 而通过关键词可进一步判断出某一文本相对于另一文本在内容方面更突出的特征。统计结果分别如表 3、表 4 所示。

Table 3. Comparison of high-frequency words
表 3. 高频词比较

Rank	2022 年《政府工作报告》英译本	频次	美国《国情咨文》	频次
1	and	944	the	355
2	the	668	and	290
3	of	478	to	296
4	to	409	of	186
5	will	352	a	138

Continued

6	we	316	in	134
7	in	261	we	104
8	for	169	that	78
9	a	143	for	72
10	development	111	our	65
11	on	97	i	60
12	with	75	is	59
13	that	69	you	58
14	more	64	with	57
15	new	58	it	55
16	government	57	have	50
17	our	57	more	45
18	improve	55	on	45
19	ensure	54	this	41
20	support	53	their	40

见表 3 可知, 两份语料中排在前四位的高频词皆是定冠词 *the*、连词 *and* 与介词 *of* 和 *to* 之类起连接句子作用的成分。两份高频词中都包含较大比重的主语 *we* 与其主格形式 *our*, 从国家叙事角度看, 两份文本从主体身份出发更能拉近政府与民众之间的距离, 建立和谐关系, 构建出始终与人民站在一起来的国家形象。美国《国情咨文》中出现了较高频次的主语 *I*, 使用第一人称能够充分体现集体观和团结观, 这样很容易让听众对演讲者产生认同感, 从而达到支持演讲者的目的, 这说明美国总统在国家叙事过程中常从个人出发。同时, 2022 年《政府工作报告》英译本中 *will* 出现了多次, 语料库中统计显示 *wewill* 搭配的频次为 223 次, 从一定程度上中国政府在进行国家叙事中注重表达国家领导人愿意与人民共同进步, 构建出关注民生、人民团结的大国形象。两份文本中都是虚词居多, 2022 年《政府工作报告》英译本中的高频实词有 *development*、*government*、*services*、*improve*、*ensure*, 这说明中国政府一直以来关注的焦点是以为人民服务为中心和国家未来的发展建设。

因此, 要想从根本上改变西方国家对我国形象的偏见, 构建良好的国际形象, 必须从词汇选用以及语义表达方面谨慎对待政府工作报告等相关政府话语的翻译工作, 在进行国家叙事时, 对政治文本的翻译在准确的前提下, 必须塑造符合国家利益, 有利于我国发展的国家形象, 更好地向世界展现出立体全面的中国。

(二) 句子层面

通过语料库对 2022 年《政府工作报告》英译本与美国《国情咨文》从平均句长、句子数量和被动式展开分析, 从而得出中美两国国家叙事的异同。

1) 平均句长。平均句长是指翻译文本语句的平均长度, 可以反映翻译文本语句的复杂程度以及翻译文本的可读性, 其计算方式以句号、问号和感叹号为标记。[4]表 4 数据显示, 2022 年《政府工作报告》英译本的平均句长高于同年度发布的美国《国情咨文》, 表明前者使用长句更为频繁, 文本句法更为复杂。然而, 对于一种代表中国形象的政治文本而言, 要想更好地塑造中国形象, 让其他国家更加了解中国的文化和国家大政方针, 在翻译政治文本时, 应当遵循简明英语的原则。关于“简明英语”, 简明英语运动组织给出的定义是: 作者在考虑读者并使用恰当语气情况下写出的言简意赅的信息, 英语句子长度不应超出 20 词, 且长句不超过 3 项信息。

Table 4. Sentence level data characteristics of the two texts

表 4. 两文本句子层面数据特征

	2022 年《政府工作报告》英译本	2022 年美国《国情咨文》
句子数量	605	521
平均句长	21.71	15.14
句子标准差	14.61	10.76

2022 年《政府工作报告》平均句长为 21.71, 说明译本中含有大量的复杂句而且承载的信息量较大, 这在一定的程度上降低了文本的可读性, 增大了国家叙事的难度。该现象的产生是由于政治文本涉及国家大政方针、基本政策, 在处理 2022 年《政府工作报告》的长难句时, 译者把忠实原文放在第一位置, 为了稳妥而选择对照原句型进行翻译。在翻译政治文本时, 在保持文本的准确性前提下, 可以适当简化句子结构, 增强文本的可读性, 这样有利于国家形象的传播。

2) 被动句式使用情况。被动句式的含义是指说明主语所表示的人或者事物是被动者的语法结构。汉语的被动句式有时没有形式上的标志, 如: 稻子收割了; 有时会在动词前边加“被”“给”“受”“挨”, 如: 对手被我们打败了。主语是动作的承受者, 或者说动作不是主语而是由其他人完成的, 则用被动式。根据对 2022 年《政府工作报告》的中文版进行分析可得知其中文本中均未是用被动语态。本研究对 2022 年《政府工作报告》英译本和美国《国情咨文》被动句式使用情况进行统计, 具体结果见表 5。

Table 5. Use of passive sentences

表 5. 被动句式使用情况

	2022 年《政府工作报告》英译本	2022 年美国《国情咨文》
句子数量	605	521
被动句式	112	35
被动句式/句子数量	18.51%	6.72%

在 2022 年《政府工作报告》英译本中有五类被动句式, 其中包括原文隐含被动语意、主动语意译作被动语意、无主句的翻译、部分动作执行者不必强调和译者意图强调宾语部分, 使用情况灵活多样。此外, 姚雪梅指出, 被动句作为英语中一个重要句式, 能够对信息进行重新编码和解码, 在句子和语篇建构中具有信息重组的语用功能, 其中包括主题确立、主题连接和焦点凸显三个方面。[5]见表 5 可知, 2022 年《政府工作报告》英译本被动句式占比高于美国《国情咨文》, 从国家叙事角度而言, 被动句式的频繁使用呈现出我国政府更多地以人民为中心。鉴于 2022 年《政府工作报告》的行文特点, 在翻译过程中, 为有效标记句子主题和聚集话题信息中心, 构建出更加开放、更加具有亲和力、充满希望、充满活力的社会主义大国形象, 译者多使用被动句式来实现信息重组, 利用被动语态应对多种文本语境下的翻译需求。

5. 结语

基于语料库的中美政治文本中国家叙事的对比分析, 结果显示, 2022 年《政府工作报告》英译本的类型符/形符比低于美国《国情咨文》, 说明其词汇丰富程度不如后者, 但由于 2022 年《政府工作报告》英译本的平均句长、平均词长以及困难词占比都要高于美国《国情咨文》, 从而在文本可读性上给读者造成了一定的理解困难, 因此, 在今后的英译中, 进行国家叙事时可适当减少长难句的运用, 增加文本可读性。就词汇密度而言, 两份语料的实词占比都达到了 60%左右。从 2022 年《政府工作报告》英译本

中的高频词及其相对于美国《国情咨文》的关键词可以看出一定的中国特色, 以及中国政治话语运用的差异, 即中国强调发展和以人民为中心。

政府工作报告属于一种特殊的文体, 反映着国家意志和人民意愿, 译者在翻译时力求与原文文体特征保持一致, 结构严谨。同时, 译者为了使原文信息更加准确地传达, 采用了归化和异化相结合的翻译策略, 对一些中国的特定名词进行解释, 也增加了文本的信息量。在今后的政府工作报告英译过程中, 译者在进行国家叙事传递中国特色的同时, 2022年《政府工作报告》英译本总体词汇应接近于本族语中同题材的文本词汇特征, 被目的语读者所接受。因此, 根据与美国《国情咨文》对比所得结论, 在进行国家叙事时, 政府工作报告的翻译应在保持原文意思的基础上, 遵循简洁的原则, 尽可能在词汇使用上靠近本族语者的阅读习惯, 可多使用定语后置修饰名词, 也可使用不同英语词汇或表达方式来翻译原文中的同一词语, 以提高用词多样性, 但也要避免使用太多的词汇而减弱文本的可读性。总而言之, 在国家叙事过程中, 注重政治文本的翻译策略是更好地向世界传达我国政府发展理念和外交理念, 讲好中国共产党治国理政的故事、中国人民奋斗圆梦的故事和中国坚持和平发展合作共赢的故事, 是构建良好国家形象的重要因素。

基金项目

2021年湖南省哲学社会科学基金项目“英语国家的中国叙事话语研究”(21WLH12)。

参考文献

- [1] 张森, 刘风光. 政治演讲语篇中人际意义的政治语言学研究——以特朗普任内首份国情咨文文本为例[J]. 东北亚外语研究, 2018, 6(3): 27-34.
- [2] Halliday, M.A.K. (1985) *Spoken and Written Language*. Deakin University, Geelong, 61-64.
- [3] Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford University Press, New York, 32-34.
- [4] 胡开宝. 语料库翻译学概论[M]. 上海: 上海交通大学出版社, 2011: 99-100.
- [5] 姚雪梅. 英语被动语态的语篇功能[J]. 河南师范大学学报(哲学社会科学版), 2010, 37(4): 189-191.