

基于语料库的政治文本英译特征分析

——以2013~2022年《政府工作报告》和《国情咨文》为例

李 婷

上海海事大学外国语学院, 上海

收稿日期: 2023年5月10日; 录用日期: 2023年6月20日; 发布日期: 2023年6月30日

摘 要

基于语料库的翻译研究已成为翻译研究领域一种新的研究范式。政治文本作为外界了解中国的一个重要手段, 其翻译工作有着举足轻重的地位。近年来, 国内关于政治文本的英译研究大多都是针对某一年的单个文本, 对其系统的总体特征关注不足。因此, 本文以2013~2022年《政府工作报告》及其英译本为研究对象, 以同时期美国《国情咨文》为参照, 利用AntConc和Wordsmith两种语料库工具, 从类符-形符比、词频、句长、关键词四个方面分析政治文本英译的语言特征, 并从齐普夫定律来验证《报告》英译本的用词是否符合自然语言规律, 以期探索政治文本翻译的特征及规律, 从而为今后的政治文本翻译提出建议。

关键词

语料库, 政治文本, 《政府工作报告》

A Corpus-Based Analysis of the Characteristics of English Translation of Political Texts

—Taking Report on the Work of the Government and State of the Union Address from 2013 to 2022 as Examples

Ting Li

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: May 10th, 2023; accepted: Jun. 20th, 2023; published: Jun. 30th, 2023

Abstract

Corpus-based translation research has become a new research paradigm in the field of translation research. As an important means for the outside world to understand China, the translation of political texts has a pivotal position. In recent years, most of the domestic studies on English translation of political texts mostly focused on individual texts of a certain year, while the overall and systematic characteristics have not received adequate attention. Therefore, this paper takes the *Report on the Work of the Government* from 2013 to 2022 and its English translation version as the object of study, and takes the *U.S. State of the Union Address* of the same period as a comparison, uses the corpus tools, AntConc and Wordsmith to analyze the linguistic features of the English translation of political texts in terms of four aspects, namely, the type-token ratio, word frequency, sentence length, and keywords, and verifies whether the use of words in the English translation of the Report conform with the principle of natural language by the Zipf's law to explore the characteristics and rules of political text translation, so as to make suggestions for the future translation of political texts.

Keywords

Corpus, Political Texts, *Report on the Work of the Government*

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

自 20 世纪 90 年代以来,随着计算机技术空前迅速的发展,语料库语言学已经成为语言研究的主流,其发展也对应用语言学等其他语言研究领域产生愈来愈大的影响。与此同时,以 Baker 为代表的翻译理论家开始利用语料库对翻译的性质和特征进行描述[1]。目前,越来越多的学者开始关注政治文本这一独特领域的翻译研究。《政府工作报告》(以下简称《报告》)作为最重要的一项政治文件,不仅是对于国计民生的全民总结,也正是外界了解中国形势和政治工作的最权威来源之一。因此,此类问题用词讲究,考量严谨,准确规范。近年来,国内关于《报告》的英译研究大多是针对某一年的单个文本,对其词汇特点和句子翻译做定性分析,然而对《报告》英译本词汇系统的总体特征研究较少。本文以 AntConc 和 Wordsmith 两种语料库工具作为辅助,从类符-形符比、词频、句长、关键词四个方面具体分析政治文本英译的语言特征,探索政治文本翻译的特征及规律,从而为今后的政治文本翻译提出建议,以期为中国外交话语及外宣工作做出贡献。

2. 文献综述

在语料库语言学蓬勃发展的背景下,语料库文体学(corpuslinguistics)的概念应运而生。语料库文体学通常使用语料库检索软件(wordsmith, antconc 等),从词频(frequency)、索引(concordance)、搭配(collocation)、词丛(clusters)、情节发展(plot)、主题词(keyness)等方面定量、定性分析文学文本,特别是小说和翻译文本,从而对文学作品中的人物塑造、情节发展、主题、作家风格等方面进行更深层次的研究[2]。

陈建生、崔亚妮(2010) [3]基于语料库方法,对中国《政府工作报告》英译文和美国《国情咨文》的

词汇特征进行了对比分析,并建议,《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。胡开宝(2018) [4]以中国外交话语 1.0 版为研究平台,从高频词、关键词、情态动词的应用和“we”的搭配角度,对中国外交话语英译文本和美国外交话语的语言特征进行比较分析,并以此为基础探讨中国外交形象成因。李耀(2021) [5]以 2020 年《政府工作报告》的中英文版本和 2020 年美国《国情咨文》为语料库,利用 AntConc, WordSmith 6.0 等语料库检索与分析工具从语篇、句长、词汇难度、人称代词和连接词等方面开展研究,研究发现《报告》英译文在语篇、句长、人称代词和连接词等方面体现出显化策略,而在词汇难度方面并未体现。金锦(2022) [6]采用计量的文法探究了《红楼梦》《生死疲劳》两部小说的文体特征,结果表明,两部作品在词汇方面、小句和词的关系方面符合奇普夫定律和门策拉-阿尔特曼定律,但是在句子和小句层面的拟合效果欠佳。董计华(2022) [7]采用评价理论框架,以英文网络媒体中的中国共产党话语为研究对象,从主题特征以及情感态度视角探究英文网络媒体对中国共产党话语宣传的动态发展特征,研究发现中国共产党网络媒体话语具有明显的主题特征,英文网络媒体在中国共产党的相关报道中总体呈现中立态度,在经济类热点话题的报道多呈现出积极态度。

由上述研究可以看出,关于政治文本英译研究缺乏系统的总体英译特征的分析。本文将回答以下 3 个问题:

1) 与美国《国情咨文》相比,中国《报告》英译文所呈现的特征有哪些?

2) 中国《报告》英译文及美国《国情咨文》是否具有明显的主题特征?这一特征与两国向来所关注的焦点是否有相似之处?

3) 中国《报告》英译文的用词是否符合自然语言规律?

3. 研究设计

3.1. 研究语料

原文语料库的时间跨度为 10 年,收录了从 2013 到 2022 年的《报告》的原文 10 篇及其英译本。中文原文来源于中华人民共和国中央人民政府网站,英文译本来源于中国网双语文件库。可比语料库选自美国政府文件《国情咨文》(State of the Union Address),为美国政府公报,时间跨度也是 10 年(2013~2022)。

《国情咨文》的语料来源于加州大学圣塔芭芭拉分校(University of California, Santa Barbara)毕业生所创建的 APP (the American Presidency Project)。拜登(Joseph R. Biden) 2021, 2022, 共两篇;特朗普(Donald J. Trump) 2017~2021, 共四篇;奥巴马(Barack Obama) 2013~2017, 共四篇。其中,拜登 2021, 特朗普 2017 为在国会联席会议上发表的类似的非国情咨文。出于研究目的,将这些信息归类为国情咨文信息可能是无害的。这样的演讲对公众、媒体和国会对总统领导能力和权力的看法的影响应该与正式的国情咨文演说一样。之所以选择美国《国情咨文》,是因为其也是总统向国会发表的年度报告,阐述政府的施政方针,与我国的政府工作报告在内容和形式上极为相似。

3.2. 理论框架

本文以批评话语分析为理论依据,对《报告》英译文语言结构进行分析。批评话语分析是现代语言学的一个新分支,适用于语言学、社会学等多个学科领域。批评话语分析主张对语言结构进行分析,阐明语言所表达的社会意义,并通过语言的运用建立和维护社会组织之间的权力关系,强调通过文本的语言学分析揭示蕴含于其中的意识形态,阐明统治阶级如何运用语言来实施意识形态控制,维护自己的权利地位[4]。在对语篇进行批评分析时“要对语篇的语境、功能和相关的社会关系具备丰富的直觉知识和准确理解”的基础上,重点分析其中可能具有重要社会意义的某些语言结构和过程,然后就整个语篇的

意识形态作出有意义的概括[8]。

3.3. 研究步骤和方法

选定汉英语料后,将语料分别保存为 word 和 txt 格式,便于后续操作。接着使用文本整理器 V5.0 对中英两种语料进行去噪。由于中文字与字之间没有空格,要使用语料库处理就必须先对中文语料进行分词。笔者使用 ROST CM6 进行中文语料分词,转码。本文所用语料均以 UTF-8 编码纯文档格式存储。语料检索工具为 WordSmith 4.0 和 Antconc 3.4.3。最后,用齐普夫定律验证《报告》英译文的用词是否符合自然语言规律。

4. 结果与讨论

4.1. 标准形符 - 类符比

在语料库语言学中,类符(type)指的是不重复计算的形符(token)。例如, To be or not to be, that is a question.这句话中,有 10 个词,即有 10 个形符,但是由于 to 和 be 重复出现,只能算为出现一次,因此只有 8 个形符。类符 - 形符比(TTR, type token ratio)可以用来研究文体特征,反映文本中词汇的丰富程度。一般情况下,超过 1000 词的语料要使用标准化类符形符比,即按一定的长度(通常是 1000 个形符)分段计算文本的类符形符比[3]。利用检索软件 wordsmith 对《报告》原文、译文及《国情咨文》原文分别进行统计后,数据如下图 1~3 所示。

N	Overall	1
text file	Overall	分词后.txt
file size	456,524	456,524
tokens (running words) in text	126,628	126,628
tokens used for word list	107,743	107,743
types (distinct words)	2,469	2,469
type/token ratio (TTR)	2	2
standardised TTR	27.33	27.33
standardised TTR std. dev.	71.04	71.04
standardised TTR basis	1000.00	1000.00

Figure 1. Statistical chart of the word list in the Chinese version of the Report

图 1. 《报告》中文版词表统计图

N	Overall	1
text file	Overall	English.txt
file size	956,827	956,827
tokens (running words) in text	148,243	148,243
tokens used for word list	146,401	146,401
types (distinct words)	5,931	5,931
type/token ratio (TTR)	4	4
standardised TTR	42.71	42.71
standardised TTR std. dev.	55.89	55.89
standardised TTR basis	1000.00	1000.00

Figure 2. Statistical chart of the word list in the English version of the Report

图 2. 《报告》英文版词表统计图

N	Overall	1
text file	Overall	state of the union.txt
file size	380,037	380,037
tokens (running words) in text	66,690	66,690
tokens used for word list	65,887	65,887
types (distinct words)	6,372	6,372
type/token ratio (TTR)	10	10
standardised TTR	44.80	44.80
standardised TTR std. dev.	53.92	53.92
standardised TTR basis	1000.00	1000.00

Figure 3. Statistical chart of the word list in the *State of the Union Address*

图 3. 《国情咨文》词表统计图

表中第一行是文本名称，第二行是文本大小，第三行是文本中的形符数，第四行是用于词表的形符数，第五行是形符数，第六行是类符 - 形符比，第七行是标准类符 - 形符比，第八行是标准化的类符 - 形符比标准方差，第九行是标准类符 - 形符比基数。中文版《报告》中用于构建词表的词的形符数是 107,743，英文版《报告》中用于构建词表的词的形符数是 146,401，形符数增加了 38,658，应证了王克非(2003, 2004) [9] [10]的研究发现：无论英译汉还是汉译英，都呈现目标文本扩增的特点。从标准形符比来看，中文版《报告》的标准形符比为 27.33，低于英文版《报告》的标准形符比 42.71，这表明汉语文本的词汇密度要低于英语文本，词汇丰富度也较低。而《政府工作报告》中英文版标准形符比均低于《国情咨文》的比值 44.88，说明《国情咨文》的词汇密度较高，用词较丰富程度较高，说明美国读者更倾向于阅读词汇丰富多样的材料。这也表明，翻译工作者在进行政府公文外译时，可以适当丰富词汇选择以进一步提高译本的 STTR 值，以符合英语国家阅读习惯，达到更好地外宣目的。

4.2. 平均句长比较

根据 wordsmith 统计结果可知，《报告》英文版平均句长为 24，高于《国情咨文》可比语料库平均句长 16，这验证了 Wang & Qin (2008)的结论：译文句子的长度高于可比语料库中的英语原文，说明了翻译过程中添加了相关的信息，从而出现了明晰化的特征[1]。

例如，综合研判国内外形势，今年我国发展面临的风险挑战明显增多，必须爬坡过坎。

译文：A comprehensive analysis of evolving dynamics at home and abroad indicates that this year our country will encounter many more risks and challenges, and we must keep pushing to overcome them.

此句话中，译文增译了 indicate，意为这种形势表明了我国面临的风险挑战增多，将原文中所隐含的“表明”给翻译出来了，并且在必须爬坡过坎一句前增译主语“we”，将无灵主语句巧译为有灵主语句。时态方面，中文里没有明显的时态标志，但是通过句意可以判断出这是一个将来时，因此，译文中通过添加“will”对翻译进行明晰化。需要注意的是，译文句子长度的增加也势必会增加阅读的难度，影响阅读速度。因此，好的翻译作品需要反复打磨，译者要能够做到词尽其用，充分发挥每个词的意义。

4.3. 主题词表征分析

主题词表，是指对比某一完整、连续的文本和一个更大的参照语料库把文本中差异显著的词语提取出来所生成的词语列表。主题词表可以客观地揭示语篇的大致内容[2]。笔者以 BNC 语料库为参照语料库，分别制作了《报告》英译本和《国情咨文》的关键词表，并提取了居前 50 的关键词，具体情况如图 4 所示：

序号	《报告》英文版			《国情咨文》			
N	Key word	RC. Freq.	Keyness	N	Key word	RC. Freq.	Keyness
1	WE	300833	10771.22656	1	OUR	93455	3847.82251
2	WILL	251179	9097.519531	2	WE	300833	2295.016602
3	AND	2624341	6304.788086	3	AMERICA	8934	1697.675781
4	DEVELOPMENT	32010	5525.211426	4	AMERICANS	2745	1511.954346
5	IMPROVE	6171	3771.438477	5	AMERICAN	16100	1366.31897
6	REFORM	5848	3430.107178	6	JOBS	9596	829.6456299
7	RURAL	6192	2865.156006	7	I'VE	0	818.5631104
8	STRENGTHEN	1193	2659.186279	8	TONIGHT	6172	818.1407471
9	PROMOTE	3143	2531.034668	9	CONGRESS	5414	667.8257446
10	CHINA	4912	2456.594727	10	NATION	3567	504.7866211
11	ENSURE	10225	2160.381836	11	THANK	11714	495.5765686
12	CHINA'S	564	1977.815063	12	LET'S	8057	476.7904663
13	CHINESE	4153	1836.350464	13	COUNTRY	27959	398.943512
14	ECONOMIC	23376	1762.969116	14	THEY'RE	0	394.6526184
15	INNOVATION	1687	1663.937744	15	FAMILIES	8222	389.7004089
16	URBAN	5371	1647.01709	16	AMERICA'S	1277	380.6244812
17	AREAS	23258	1625.173462	17	APPLAUSE	541	379.2957153
18	YUAN	545	1563.630493	18	WE'RE	18644	354.7539978
19	GROWTH	12800	1542.643066	19	PERCENT	2926	334.6766052
20	ENTERPRISES	1740	1536.697266	20	STATES	17873	313.6548462
21	GOVERNMENT	56343	1528.943726	21	LAUGHTER	2068	303.6236572
22	IMPLEMENT	1548	1479.266235	22	ECONOMY	10365	299.4622803
23	EFFORTS	5516	1398.654297	23	BIPARTISAN	52	293.2540588
24	DEEPEN	171	1349.814575	24	EVERY	39156	278.6866455
25	DEVELOP	8544	1157.847656	25	PRESIDENT	15747	270.3843079
26	OVERSIGHT	167	1150.37085	26	FOLKS	429	263.9572449
27	POLICIES	8778	1137.881714	27	UNITED	19030	257.8166199
28	GOVERNANCE	180	1081.200073	28	IMMIGRATION	1081	248.8986511
29	CONTINUE	11642	1054.418457	29	COVID	0	248.4824219
30	OUR	93455	1051.703369	30	WORLD	53806	237.741745
31	COOPERATION	1288	1046.375366	31	PROTECT	5047	231.2201233
32	JINPING	0	1041.675171	32	WILL	251179	227.7261658
33	SERVICES	24866	1033.14502	33	AND	2624341	219.0813141
34	PERCENT	2926	985.237793	34	US	80226	217.3435059
35	BASIC	10992	983.939209	35	MILLIONS	2638	217.1062622
36	STABLE	3384	977.9755249	36	VETERANS	440	215.9586792
37	SYSTEM	44076	949.411499	37	PANDEMIC	17	214.6303711
38	BUSINESSES	3657	908.8424072	38	ALLIES	1578	211.7842102
39	INVESTMENT	10858	891.067749	39	REPUBLICANS	523	206.0131378
40	WORK	89381	883.4902954	40	BUSINESSES	3657	205.8360291
41	MAJOR	28146	883.168396	41	THAT'S	66435	204.4584961
42	INDUSTRIES	4322	845.0880127	42	ADMINISTRATION	6408	199.3022308
43	TRILLION	122	838.0805054	43	PEOPLE	116196	193.9389648
44	INCREASE	16808	833.6784668	44	INFRASTRUCTURE	993	191.9231873
45	SUPPORT	30668	830.2711792	45	QAIDA	0	190.0151978
46	GOVERNMENTS	4691	823.6982422	46	AFGHANISTAN	613	188.3955688
47	ENCOURAGE	5073	822.2379761	47	YEAR	68714	188.2736969

Figure 4. Keyword list for the English translation of the *Report* and the *State of the Union Address* using BNC as the reference corpus

图 4. 以 BNC 为参照语料库的《报告》英译本和《国情咨文》的关键词表

图4中, key word 一列为以 BNC 为参照语料库的《报告》英译本和《国情咨文》关键词, RC.freq. 列为该词在参照语料库即 BNC 语料库中出现的频率。关键性(keyness)可以测量词语的搭配或共现强度。统计《报告》英译本前 50 个关键词, 我们可大致了解中国政府近十年间的工作重点, 如 development (发展)、reform (改革)、rural (农村的)、government (政府)、system (制度)、economic (经济的)、urban (城市)、cooperation (合作)、enterprises (企业)等词汇, 还有诸如 improve (提高、改善)、strength (加强)、promote (促进)、implement (贯彻)、等具有积极色彩的动词, 以彰显政府新的一年开展工作的决心和魄力。《国情咨文》前 50 个关键词也能反映出美国政府所关注的焦点, 如 America (美国)、congress (国会)、country (国家)、pandemic (疫情)、citizen (公民)等。根据图 4, 在《报告》英译本和《国情咨文》的关键词表中, 共有的关键词较少, 如“businesses”, “economy”, “we”等。显然, 两国都重视经济发展。不过两者的关键词表存在显著差异。在《报告》英译本中, 可以看出内容符合中国国情, 如关键词“rural”, “poverty”和“urban”, 过去十年, 为中国脱贫攻坚的关键十年, 可以看出过去十年中国始终将减贫摆在治国理政突出位置, 持续增强脱贫攻坚投入力度, 共同努力以实现全面脱贫和共同富裕。而在美国外交话语中, 关键词表包括“bipartisan”, “allies”, “Qaida”, “Afghanistan”。其中, allies 表示同盟国, Qaida 为基地组织, 该组织主要在中亚、中东地区活动。有必要指出的是, 在《报告》英译本出现频率较高的关键词如“development”, “cooperation”, “innovation”, “stable”等词在《国情咨文》关键词表中并未出现。

4.4. 高频词比较

通过词频检索, 可以发现语料库中使用频率最高的那些词。一般情况下, 一个语料库词频表中的高频词绝大多数是功能词(或称虚词), 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, 因此才会被反复提到[3]。笔者通过 antconc 检索出了《报告》英译本和《国情咨文》中按词频从高到低排列的前 20 个高频词及其频序(rank)和频次(frequency)。

《报告》英译本				《国情咨文》			
频序	频次	词语	Z	频序	频次	词语	Z
1	9721	and	9721	1	2942	the	2942
2	7520	the	15040	2	2368	and	4736
3	5026	of	15078	3	2168	to	6504
4	4381	to	17524	4	1603	of	6412
5	4077	we	20385	5	1437	we	7185
6	3427	will	20562	6	1208	a	7248
7	2985	in	20895	7	1163	in	8141
8	1810	for	14480	8	1065	that	8520
9	1665	a	14985	9	1029	our	9261
10	1219	development	12190	10	830	is	8300
11	987	on	10857	11	669	for	7359
12	938	with	11256	12	639	it	7668
13	841	be	10933	13	626	have	8138
14	818	that	11452	14	621	you	8694
15	766	people	11490	15	491	this	7365
16	697	china	11152	16	478	are	7648
17	673	our	11441	17	459	on	7803
18	668	government	12024	18	441	will	7938
19	651	all	12369	19	417	with	7923
20	615	new	12300	20	413	america	8260

Figure 5. A comparison of the first 20 high-frequency words in the English translation of the Report and State of the Union Address

图 5. 《报告》英译本及《国情咨文》前 20 个高频词的比较

图5数据显示,《报告》英译本词频表前20个高频词中只有5个实词(25%),虚词15个(75%),而《国情咨文》的词频表中前20个高频词中只有2个实词(10%),虚词18个(90%)。这两个高频词表的差异在一定程度上反映出《报告》所关注的焦点比《咨文》更为集中和明确。

通过在 Antconc 中 Clusters/N-Grams 一栏检索“we”,将词丛的大小调整为2,可得出图6。

《报告》英译本				《国情咨文》			
rank	freq	range	cluster	rank	freq	range	cluster
1	2594	2	we will	1	2678	2	we will
2	215	2	we have	2	357	2	we have
3	211	2	we should	3	245	2	we must
4	193	2	we must	4	242	2	we should
5	109	2	we need	5	163	2	we are
6	38	2	we are	6	153	2	we need

Figure 6. The first six word clusters of “we” in the English translation of the Report and the State of the Union Address

图6.《报告》英译本及《国情咨文》的词表中“we”的前6个词丛

由图6可以看出,两文本出现频率前6的“we”的词丛均为“we will”,“we have”,“we should”,“we must”,“we need”,“we are”,只是《报告》英译本中“we should”和“we need”的频数比《国情咨文》稍前,但是搭配词无变化。频繁使用“we will”,“we should”,“we must”等词语,实际上表明了两国政府新一年的政策方针,以及政府决心为人民服务、为人民做好事的态度,政府带领人民创造美好未来的决心和行动。通过使用 we,演讲者在展示和强调自己的立场和态度,这使得演讲更加有说服力,更加亲民,有利于他将自己的意志强加于公众。

4.5. 齐普夫定律验证

1932年,哈佛大学的语言学家齐普夫(G. K. Zipf)在研究英文单词的出现频率时,发现如果把单词频率从高到低的次序排列,每个单词出现频率和它名次的常数次幂存在简单反比关系。齐普夫定律表明,在自然语言文本中,文档中词汇的频次与其排序等级呈现反比,且两者乘积为一常数[12]。根据图4,Z列为《报告》英译本及《国情咨文》的词频表中前20个高频词频数及频次的乘积,由Z列数据可知,该乘积趋近于一常数。参照齐普夫公式对语料库所有词语的频数和频次数据进行拟合,得到下图7。

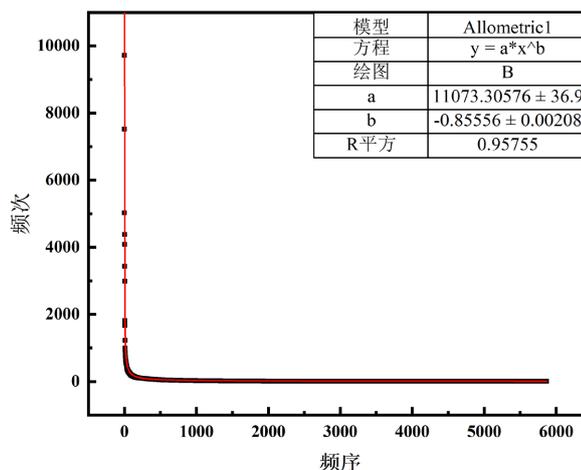


Figure 7. A Zipf's law fit of rank and frequency of words in the English translation of the Report

图7.《报告》英译本中词的频数与频次的齐普夫定律拟合

齐普夫定律符合幂律关系 $y = a \cdot x^{-b}$ 。如图 7 所示，散点是实际统计的数据，红色线为拟合的效果。拟合效果根据拟合优度指标 R^2 的值来判断， $R^2 > 0.90$ 为拟合效果优秀； $R^2 > 0.80$ 为良好； $R^2 > 0.75$ 为可接受； $R^2 < 0.75$ 为不可接受。由图 7 可知，拟合优度指标 $R^2 = 0.95755$ ，说明拟合效果很好。《报告》英译本中词的频序与频次的关系与幂函数模型 $y = ax^{-b}$ 之间的拟合效果好，也就是说词的频率与频次的关系符合齐普夫定律，也说明《报告》英译本的用词符合自然语言规律。

笔者将《报告》英译本中词的频序与频次取对数后，那么图 7 中的散点数据将会如图 8 所示。

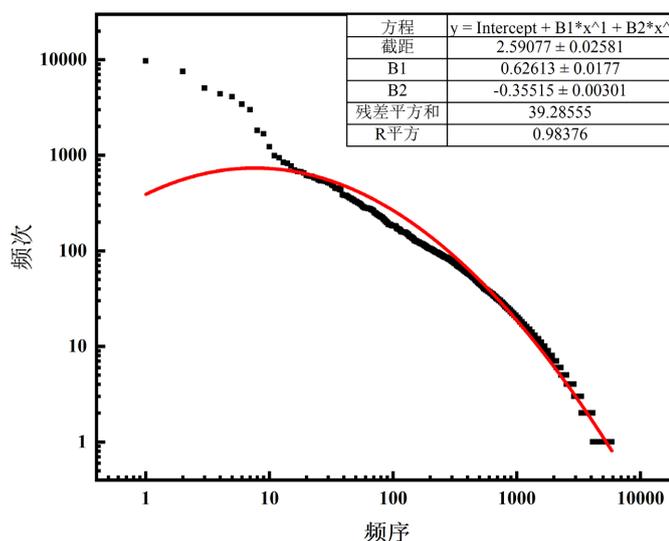


Figure 8. Linear fitting of rank and frequency (taking log values) of words in the English translation of the Report

图 8. 《报告》英译本中词的频序与频次(取 log 值)的线性拟合

从图 8 中可以看到拟合的上半部分出现了较大的偏差，这进一步说明齐普夫定律比较充分地解释了中频词的分布规律。图线结果 $R^2 = 0.98376 > 0.90$ ，拟合效果非常好，进一步说明《报告》英译本的用词符合自然语言规律。此外，由图 4 前 20 个出现频率最高的词大部分是虚词，也就是功能词，这样符合齐普夫提出的省力原则。我们在交流时会尽可能地使用比较少的字表达比较充分的信息。

5. 结语

本文以 2013~2022 年《政府工作报告》及其英译本为研究对象，以同时期美国《国情咨文》为参照，利用 AntConc 和 Wordsmith 两种语料库工具，从类符-形符比、词频、句长、关键词四个方面分析政治文本英译的语言特征，并从齐普夫定律来验证《报告》英译本的用词是否符合自然语言规律，以期探索政治文本翻译的特征及规律，从而为今后的政治文本翻译提出建议。语料库方法为话语分析研究提供了大量真实可靠的数据支撑。

分析显示，与美国《国情咨文》相比，中国《报告》英译文的词汇密度较高，用词较丰富程度较低，译文句子的长度要高于《国情咨文》英语原文，说明了翻译过程中添加了相关的信息，从而出现了明晰化的特征，以使目的语受众更好地理解文本。由检索关键词可知，中国政府除关注脱贫攻坚、经济发展等热点领域外，使用较多具有积极色彩的动词和名词，以彰显政府新的一年开展工作的决心和魄力。美国政府虽也重视经济发展，但话语中多关注党派、中东及中亚问题。通过文本中出现频率最高的前 20 词表也在一定程度上反映出《报告》所关注的焦点比《咨文》更为集中和明确。并且，两文本中都多次使

用“we”及“we + 情态动词”结构,演讲者意在展示和强调自己的立场和态度,这使得演讲更加有说服力和号召力,更加亲民。最后,通过齐普夫定律的验证,可知《报告》英译本的用词符合自然语言规律,也符合齐普夫提出的省力原则。

因此,翻译工作者在进行政府公文外译时,可以适当丰富词汇选择以进一步提高译本的 STTR 值,以符合英语国家阅读习惯,达到更好地外宣目的。译者需要对译文进行反复打磨,要能够做到词尽其用,充分发挥每个词的意义,秉持省力原则。在热点话题领域方面,侧重经济建设话题和民生话题,较少报道教育、文化等领域建设,因而有必要在对外宣传中加强对中国共产党的思想文化等相关领域的报道和宣传工作,以帮助国际社会更全面真实地了解中国特色社会主义路线、我国社会主义教育、文化等领域建设中所取得的成就。

参考文献

- [1] 朱晓敏. 基于自建语料库的政治文本英译特点研究[J]. 解放军外国语学院学报, 2011, 34(3): 73-77+128.
- [2] 姜晓艳. 基于语料库的《简·爱》语言特点及主题词表征分析[J]. 江苏科技大学学报(社会科学版), 2016, 16(2): 77-84.
- [3] 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.
- [4] 胡开宝, 田绪军. 中国外交话语英译中的中国外交形象研究——一项基于语料库的研究[J]. 中国外语, 2018, 15(6): 79-88.
- [5] 李耀. 基于语料库的英译显化研究——以 2020 年《政府工作报告》为例[J]. 海外英语, 2021(16): 35-36+54.
- [6] 金锦. 《红楼梦》与《生死疲劳》计量文体特征对比分析[J]. 开封文化艺术职业学院学报, 2022, 42(4): 22-27.
- [7] 董记华, 董帅. 基于语料库的中国共产党网络媒体话语的历时研究[J]. 外语研究, 2022, 39(5): 39-44.
- [8] 严世清, 赵霞. 政治语篇中的情态表达及其批评性话语分析[J]. 苏州大学学报(哲学社会科学版), 2009, 30(2): 117-120.
- [9] 王克非. 英汉/汉英语句对应的语料库考察[J]. 外语教学与研究, 2003, 35(6): 410-416+481.
- [10] 王克非, 刘鼎甲. 大规模英汉平行语料库的检索与应用: 大数据视角[J]. 外语电化教学, 2017(6): 3-11.
- [11] 吴冰. 齐普夫定律在中文字频测定的推广[J]. 佳木斯大学学报(自然科学版), 2014, 32(3): 478-480.
- [12] 谢靖, 陈静, 王东波. 齐普夫定律在中文短语知识中的呈现[J]. 情报学报, 2014, 33(1): 11-22.