

# A Corpus-Based Study on the Translation of 2020 Chinese Government Work Report

Bingbing Song

School of Foreign Languages, China University of Petroleum (Beijing), Beijing  
Email: 13241294917@163.com

Received: Jul. 30<sup>th</sup>, 2020; accepted: Aug. 13<sup>th</sup>, 2020; published: Aug. 20<sup>th</sup>, 2020

---

## Abstract

Government work report is an important form of China's official documents, whose translation is the window for the world to understand China, as well as the road for China to reach world. Applying the theory and approaches of corpus linguistics, the comparative analysis was made on the 2020 Chinese government work report, its English translation and the remarks by president Trump in 2020 State of the Union address in four aspects: the proportion of high-frequency words, type token ratio, average sentence length and part of speech. The study also investigated the language features of the translation for government work report in simplification and explicitation, and some suggestions about the translation were given in the study.

## Keywords

Government Work Report, Translation Universals, Language Features, Corpus

---

# 基于语料库2020年政府工作报告的英译研究

宋冰冰

中国石油大学(北京)外国语学院, 北京  
Email: 13241294917@163.com

收稿日期: 2020年7月30日; 录用日期: 2020年8月13日; 发布日期: 2020年8月20日

---

## 摘要

政府工作报告是中国官方文书的重要形式之一, 其英译文是世界了解中国的窗口, 也是中国走向世界的通道。本文应用语料库语言学的理论及研究方法, 对2020年中国政府工作报告中文原文、英译文及美国2020年国情咨文进行了语际、语内比较。从高频词比例、词语变化性、平均句长以及词类分布四个角度

出发, 并从简化和显化的层面进一步考察译文的语言特征, 并针对政府工作报告译文的迁移性冗余问题及连词超用问题提出适当建议。

## 关键词

政府工作报告, 翻译共性, 语言特征, 语料库

Copyright © 2020 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

2020年5月22日, 国务院总理李克强在第十三届全国人民代表大会第三次会议上做了本年度的政府工作报告。2020年是中国发展的关键一年, 是中国第一个百年目标的收官之年, 国家目标在2020年全面建成惠及十几亿人口的更高水平的小康社会。此外伴随新冠肺炎病毒在全球范围内肆虐, 本年度也是中国政府打好疫情防卫战的关键之年。报告回顾了中国2019年在社会各方面取得的工作成果, 并为今年的政府工作设定了目标, 制定了计划。在国务院官方网站发布中文政府工作报告全文的同时, 中国官方新闻媒体新华网也于6月10日发布了相应的英文版。鉴于中国政府在国家发展以及抗击疫情中取得的重大成就, 英文版政府工作报告的发出有利于国外借鉴相关经验以及塑造良好的中国国际形象。

政府工作报告的英译版是翻译界的重点关注对象之一, 众多学者从理论应用的视角出发, 探讨了各理论在政府工作报告英译文中的及应用及作用, 例如, 童孝华(2014 [1])研究了不同翻译主体在译文中的体现及作用, 其他理论还有目的论(李红霞, 2010 [2])、语篇衔接理论(胡峰笙, 2012 [3])、权力话语理论(赵琦, 2016 [4])、以及评价理论(司炳月、高松, 2019 [5])等。也有研究探讨了报告中概念隐喻的翻译(王瑞昀, 2016 [6]; 田苗, 2016 [7]), 还有研究探讨了报告英译文中的翻译问题及对应策略, 例如宏观层面的迁移性冗余问题(武光军, 2010 [8])以及具体层面的词汇翻译问题(吴文子, 2006 [9]; 陆仲飞、王大伟, 2016 [10])。在研究方法层面, 有研究运用了语料库分析工具来研究英译文的语言特征(朱晓敏, 2011 [11]; 李晓倩、胡开宝, 2017 [12])。

中国政府工作报告英文版的主要受众是海外人群, 因此有必要将政府报告英译文同目标语原创文本进行比较, 以此分析政府工作报告的英译质量, 而国内现有针对政府工作报告英译的研究鲜有从翻译共性的视角出发, 采用语料库检索工具的研究也相对较少。鉴于当前对2020年政府工作报告的研究较少, 而其对国外借鉴中国发展及抗疫经验意义重大, 且关乎中国国际形象塑造, 本文将2020年政府工作报告及其英译文为研究语料, 应用语料库翻译学的理论的方法, 通过与美国2020年国情咨文进行比较, 尝试从多个维度考察报告英译文的语言特征。

## 2. 翻译共性理论

“翻译共性”这一概念最早是由Mona Baker (1993 [13])提出的。她指出: “翻译共性即通常出现在翻译文本而不是源语文本中的特征, 并且不是特定语言系统干扰的结果”(Baker, 1993: 243 [13])。这意味着她将翻译语言视为一种独立的语言, 并强调其表现出的不同于非翻译语言的普遍特征。胡开宝(2011:86 [14])解释道, Mona Baker关于翻译共性的定义具有双重意义, 首先, 翻译共性概率分布在特定的语言(目标语言)之中, 依靠统计归纳而体现; 其次, 翻译共性是翻译过程本身的结果, 具有独立于源语言和目标

语言系统的差异。Mona Baker (1996 [15])将翻译共性分为四种类型,即简化、显化、范化和整齐化。本文主要从简化和显化两个维度出发,探讨中国 2020 年政府工作报告英译文的语言特征。

1983 年,Blum-Kulka 和 Levenston 指出,词汇简化是指用更少的词来表达思想的过程及产物(1983:119 [16])。Mona Baker (1996: 181 [15])将简化定义为“在翻译中简化语言的倾向”,“包括让读者更容易理解事物”。有一些语言表征可以用来描述简化趋势,Mona Baker (1996 [15])认为,句子的平均长度和标点改变都可以作为简化的体现。Baker (1996: 176 [15])认为简化是“译者下意识地简化语言或信息或两者都有”。因此,这一趋势在目的语文本中表现为相对于源语和原语,用词减少,常用词语重复,实词减少,虚词增多,语言复杂性普遍降低(胡开宝, 2011:101 [14])胡显耀(2007 [17])在针对汉语翻译小说词语特征进行研究时,通过词语的变化性以及词表常用词描述了翻译简化趋势。

庞双子和胡开宝(2019: 63 [18])指出,“显化”概念最早是由 Vinay 和 Darbelnet 在 1958 年提出的,表示“对原语中暗含的,但可以从上下文中推导出的信息在译语中加以明示”(Vinay & Darbelnet, 1958: 8 [19]; 转引自胡显耀、曾佳, 2011: 57 [20])。其后,Blum-Kulka (1986 [21])在研究翻译衔接措施显化时提出了显化假说,指出,“显化是一种整体趋势,即在翻译中把事情解释清楚,而不是把它们隐舍起来。”Mona Baker (1996:180 [15])认为,显化趋势的表现包括文本长度,“译文通常比原文要长”。此外,“在词汇上,翻译中明确事物的倾向可以通过解释性词汇和连词的使用或过度使用来表达”(Baker, 1996: 181 [15])。

本研究将结合语料库检索工具,立足于翻译共性理论中的简化趋势与显化趋势,具体从高频词比例、词语变化性、平均句长以及词类分布四个角度,通过单语类比和双语平行比较两个途径来描写政府工作报告译文的语言特征,并未后续政府工作报告翻译提供一定的参考,以期进一步提高政府文书外译质量,服务于向世界传播好中国声音,塑造好中国形象。

### 3. 研究设计

政府工作报告是中国的一种重要公文形式,其内容主要分为两部分:一是回顾并总结前一年或者前五年的政府工作情况,二是汇报新一年中国政府的工作计划、目标及举措。历年的国务院政府工作报告是中国社会以及政府政策发展变化的缩影,同时也是国外了解中国的重要参考,以及中国向世界传播中国声音,让世界人民进一步了解中国的重要渠道。

本研究所采用的 2020 年中国国务院政府工作报告中文语料来自于中华人民共和国中央人民政府官方网站([www.gov.cn](http://www.gov.cn)),以下简称为 RWGC,共 11331 字,5667 词。英语译文来自于中国日报网站对新华网的转载([www.language.chinadaily.com.cn](http://www.language.chinadaily.com.cn)),其标题为 Report on the Work of the Government,以下简称为 RWGE,共 9660 词。英语原创文本选择了体裁比较接近的美国 2020 年国情咨文 Remarks by President Trump in State of the Union Address,来自于美国白宫官方网站(<https://www.whitehouse.gov/>),以下简称 SUA,共 6314 词。

本研究对三个文本分别进行清理建库,使用中国科学院计算所软件室研发的 ICTCLAS Version 1.0 对中文文本进行词汇切分及标注,德国斯图加特大学 Helmut Schmid 开发的 TreeTagger3 对英语文本进行词类标注,英国语料库专家 Mike Scott 教授设计的 Wordsmith6 进行单语检索以及 EditPad Pro 进行文本编辑文档内容筛选。借助上述语料库工具,本文通过单语类比和双语平行比较两个途径,即政府报告英译文和美国国情咨文进行比较,政府报告英译文和政府报告中文文本进行比较,从高频词比例、词语变化性、平均句长以及词类分布四个角度来考察政府报告英译文的语言特征。

## 4. 研究结果与分析

### 4.1. 高频词比例

Olohan (2004: 77 [22])指出,“词表是语料库中出现的所有单词的列表,附带单词出现的频数”。高

频词表是指词表中出现频率较高的单词的列表, 根据研究文本的大小和高频词的一般选择标准, 本研究通过 Wordsmith6 生成了三个子语料库的词表, 将词表中单词按频次高低排序, 选取了前 20 的单词制成了高频词表, 如表 1 所示。

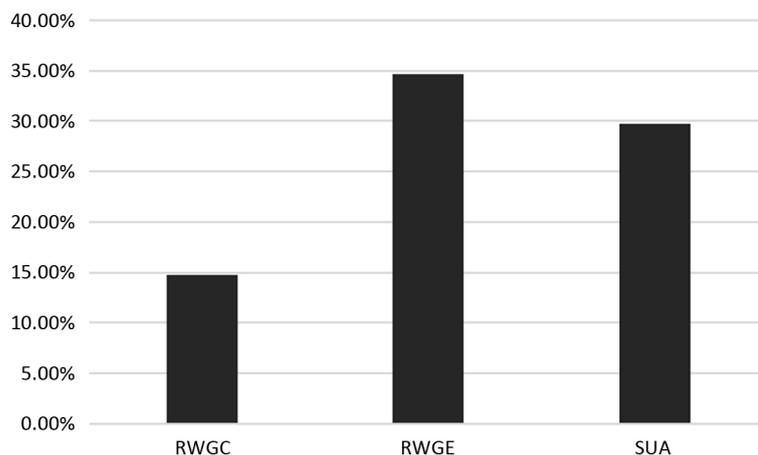
**Table 1.** The high-frequency word list of three corpora

**表 1.** 子语料库的高频词表

排序	RWGC	频次	RWGE	频次	SUA	频次
1	和	100	AND	634	THE	287
2	的	97	THE	508	AND	221
3	发展	84	OF	335	TO	166
4	要	61	WE	294	OF	149
5	就业	39	WILL	287	IN	114
6	企业	38	TO	254	IS	100
7	经济	36	IN	198	A	97
8	加强	34	FOR	110	OUR	95
9	建设	34	A	106	WE	95
10	支持	32	DEVELOPMENT	74	YOU	71
11	保	31	ON	69	THAT	70
12	改革	31	BE	68	ARE	64
13	社会	31	OUR	64	FOR	63
14	疫情	31	WITH	59	HAVE	51
15	等	29	ALL	54	I	50
16	政策	28	PEOPLE	53	ON	40
17	新	27	HAVE	52	WITH	40
18	推进	26	THAT	50	AMERICAN	36
19	为	25	WORK	42	THIS	35
20	促进	24	SUPPORT	39	VERY	34
	合计	838		3350		1878

观察表 1 可以发现, 相比政府工作报告中文语料库(RWGC), 政府工作报告英译文语料库(RWGE)虚词较多, 而相比美国国情咨文语料库(SUA), RWGE 受中文原文影响而带有更多的实词。借鉴胡显耀(2007 [17])通过计算高频词使用频率占各语料库总词频的比例来争鸣翻译小说词语操作的简化趋势, 本文也以此方法计算了各子语料库的高频词比例, 即高频词次数之和与总词数之比。

高频词比例高说明文本中较少常用词语的重复使用次数较多, 以此可以体现文本的简化趋势, 即高频词比例越高, 简化趋势越明显。通过比较三个子语料库的高频词比例, 发现英译文本的高频词比例(34.68%)高于英语原创文本(29.74%), 也高于中文原创文本(14.79%), 具有明显的简化趋势, 如上述图 1 所示。



**Figure 1.** The proportion of high-frequency words in three corpora  
**图 1.** 子语料库的高频词比例

## 4.2. 词语变化性

“词语的变化性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。”(胡显耀, 2007: 215 [17])类符形符比即语料库中类符(type)与形符(token)的比率,类符是指不同的词语,形符是指所有的词形。于红(2016: 80 [23])指出,“TTR 的数值越高,表明文本的词汇丰富度较高,文本的阅读难度较大;反之,词汇丰富度越低,文本的阅读难度越小。”本研究的三个子语料库存在语料大小差异,因而采用标准类符形符比(Standardised TTR, 简称 STTR)进行分析,即应用 Wordsmith6 进行标准化处理,计算各个 1000 词的平均 TTR。

**Table 2.** The STTR of three corpora  
**表 2.** 子语料库的标准类符形符比

文本	RWGC	RWGE	SUA
Tokens	5667	9660	6314
Types	1876	2015	1762
STTR	56.66	43.71	45.83

根据表 2, 英译文本(RWGE)的 STTR(43.71)略低于英语原创文本(SUA)的 STTR(45.83), 也低于中文原创文本(RWGC)的 STTR(56.66), 词汇丰富度较低, 文本的阅读难度较小, 具有明显的简化趋势。

## 4.3. 平均句长

Mona Baker (1996:180 [15])指出:“译文通常比原文长。”句子越长,其中包含的信息就越多,因而可以以此描述显化趋势。平均句长是指平均每个句子中包含词数, Wordsmith6 的 Wordlist 功能可以直接计算出语料库的平均句长,结果如表 3 所示。

**Table 3.** The average sentence length of three corpora  
**表 3.** 子语料库的平均句长

文本	RWGC	RWGE	SUA
句数	371	438	402
平均句长	15.13	21.87	15.5
标准值	15.08	19.12	10.05

观察表 3, 我们可以发现, 英译文本(RWGE)的平均句长(19.12)大幅高于英语原创文本(SUA)的平均句长(10.05), 也高于中文原创文本(RWGC)的平均句长(15.08), 具有明显的显化趋势。译者会将原文中隐含的意思明确地表达出来, 从而产生一定程度的冗余。结合政府工作报告的平行语料库观察发现, 译文中确实存在冗余现象, 例如, 将原文中“我们一定要努力改进工作, 切实履行职责, 尽心竭力不辜负人民的期待”译为“*We must strive to improve our work and fulfill our responsibilities and make every effort to live up to the people's expectations*”实际上, 原文所讲的只是政府会尽职负责这一层意思, 而英译文对原文进行了亦步亦趋的翻译, 造成了武光军(2010 [8])提到的迁移性冗余问题, 他还进一步指出, “翻译中的迁移性冗余指的是, 并不按照译文语言的内在组合规律, 而是将原文中的语言组合形式直接迁移到译文中, 从而造成译文的冗余性表达, 并影响译文的交流效果的现象”(武光军, 2010: 64 [8])。因此平均句长所体现的翻译显化趋势告诫我们在进行外宣翻译时, 不能逐字逐句地机械转换, 而是要考虑到汉英语言思维模式的差异, 根据国外受众的思维习惯对原文进行适当加工, 从而提高翻译质量。

#### 4.4. 词类分布

Maevae Olohan (2004: 81 [22])指出, 词汇密度“是指语料库中内容词与功能词的比例”。还指出, “词汇密度是通过计算内容词占总词数的百分比得到的”。词汇密度越低, 语料库中内容词的数量越少, 所占比例越低, 信息量和难度也就越低。本研究认为, 译文中词汇密度越高, 说明倾向越明显。

Table 4. The POS of three corpora

表 4. 子语料库的词类分布

词类	RWGC	RWGE	SUA
名词	1760	2964	1835
动词	2181	1480	1082
副词	187	235	396
形容词	506	1188	589
词汇密度	81.77%	60.73%	61.80%

表 4 显示, 英译文本的词汇密度(60.73%)稍低于英语原创文本(61.80%), 也低于中文原创文本(81.77%), 显化趋势并不明显。然而英译文本的名词使用情况(12.30%)高于英语原创文本(9.33%)也高于中文原创文本(8.93%)存在超用现象。显化趋势体现在解释性词汇和连词的使用或过度使用层面(Baker, 1996: 181)。名词存在超用情况之外, 观察表 1 高频词表可以发现, “和”与“and”都出现在了高频词表的前列, 进一步计算分析发现, 英译文本中的“and”的使用频率(6.56%)高于英语原创文本(3.50%), 也大幅高于中文原创文本(1.76%), 同样存在超用现象。

## 5. 结语

本研究以中国 2020 年政府工作报告英译文为研究对象, 基于翻译共性理论, 主要从简化趋势和显化趋势两个层面出发, 结合与政府工作报告中文原文以及美国 2020 年国情咨文进行对比, 对政府工作报告的英译文进行了描写与分析。结合语料库检索工具, 本文从高频词比例、词语变化性、平均句长以及词类分布四个角度细致分析了政府工作报告英译文的语言特征。

本文研究发现, 中国 2020 年政府工作报告英译文和美国国情咨文原文以及中文源语文本相比, 英译文高频词比例较高, 词汇变化性较低, 词汇丰富度较低, 具有明显的简化趋势。此外, 政府工作报告英译文平均句长较长, 虽然词汇密度较低, 但是存在名词和连词“and”的超用现象, 有所体现显化趋势。

进一步深入文本分析发现,2020年政府工作报告英译文本存在迁移性冗余问题,连词“and”的大幅超用也体现出了翻译文本存在衔接僵硬问题。望以此为戒,政府文书翻译是我国外宣工作的重要组成部分之一,其翻译质量不仅关乎中国声音的对外传播,也关乎中国国际形象的塑造,只有逐步打破刻板僵化的文书翻译模式,深入考虑英汉语言模式的差异,考虑到国外受众的思维模式以及目标语的相应文本表达习惯,从而可以进一步提高我国政府文书的翻译质量。本文主要结合语料库检索工具对翻译文本进行了多角度描写分析,有所发现存在的翻译问题,关于报告中具体的翻译问题有待于进一步深入文本分析并提出更具针对性的改进建议,相关研究有待于进一步深入。

## 参考文献

- [1] 童孝华. 翻译的主体意识——2014年政府工作报告翻译心得[J]. 中国翻译, 2014, 35(4): 92-97.
- [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, 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] 赵琦. 权力话语理论视域下政论文英译策略——以2014年政府工作报告中双引号词语为例[J]. 广西师范大学学报(哲学社会科学版), 2016, 52(4): 128-133.
- [5] 司炳月, 高松. 外宣文本中英级差资源分布与翻译——以2019年政府工作报告双语文本为例[J]. 上海翻译, 2019(5): 14-20.
- [6] 王瑞昀. 概念隐喻的认知及其跨文化英译研究——以《政府工作报告》译文为例[J]. 上海对外经贸大学学报, 2016, 23(2): 87-96.
- [7] 田苗. 基于概念隐喻的2015年政府工作报告英译研究[J]. 黑龙江高教研究, 2016(6): 157-160.
- [8] 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.
- [9] 吴文子. 2005年《政府工作报告》英语译文中一些值得商榷的译例[J]. 上海翻译, 2006(2): 76-78.
- [10] 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.
- [11] 朱晓敏. 批评话语分析视角下的《政府工作报告》英译研究(一)——基于语料库的第一人称代词复数考察[J]. 外语研究, 2011(2): 73-78+112.
- [12] 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.
- [13] Baker, M. (1993) Corpus Linguistics and Translation Studies: Implications and Applications. *American Journal of Physiology*, 274, 321-327. <https://doi.org/10.1113/jphysiol.1993.sp019637>
- [14] 胡开宝. 语料库翻译学概论[M]. 上海: 上海交通大学出版社, 2011.
- [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/Philadelphia, 175-186. <https://doi.org/10.1075/btl.18.17bak>
- [16] Blum-Kulka, S. and Levenston, E. (1983) Universals of Lexical Simplification. In: Færch, C. and Kasper, G., Eds., *Strategies in Inter-Language Communication*, Longman, London, 119-139.
- [17] 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(3): 214-220 + 241.
- [18] 庞双子, 胡开宝. 翻译共性中的显化问题研究[J]. 现代外语, 2019, 42(1): 61-71.
- [19] Vinay, J. and Darbelnet, J. (1958/1995) *Comparative Stylistics of French and English: A Methodology for Translation*. John Benjamins Publishing Company, Amsterdam. <https://doi.org/10.1075/btl.11>
- [20] 胡显耀, 曾佳. 基于语料库的翻译共性研究新趋势[J]. 解放军外国语学院学报, 2011, 34(1): 56-62 + 127-128.
- [21] Blum-Kulka, S. (1986) Shifts of Cohesion and Coherence in Translation. In: House, J. and Blum-Kulka, S., Eds., *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, Gunter Narr, Tübingen, 17-35.
- [22] Olohan, M. (2004) *Introducing Corpora in Translation Studies*. Routledge, London. <https://doi.org/10.4324/9780203640005>
- [23] 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.