

基于语料库的国防部例行记者会汉英口译词汇特征研究

杜海琪

上海海事大学外国语学院, 上海

收稿日期: 2023年8月18日; 录用日期: 2023年9月28日; 发布日期: 2023年10月9日

摘要

目前针对国防部例行记者会的研究还比较少, 且研究大都集中于部分词汇和句子, 对口语词汇整体的研究则少之又少。本文采用语料库研究法, 自建语料库, 以近五年来的国防部例行记者会汉英口译词汇为研究对象, 对标准类符/形符比、高频词、动词名词化和平均词长四个词汇特征进行定量分析, 并与同体裁的美国国防部新闻发布会发言语料进行对比, 分析两者的不同及差异形成的原因并给出建议。本研究的发现主要如下: 相较于新闻发布会, 例行记者会口译语料库具有更高的词汇复杂度, 词汇密度也比较高, 口译笔译化的倾向也很明显, 这些也与简化、范化的翻译共性相佐。造成这一现象的原因主要是因为政府例行记者会是公共政治传播的重要形式, 发言人既要传递重要的政策信息也要阐述政府的立场, 因此用词较为正式准确, 这也在一定程度上解释了口译词汇特征要比原创英语要复杂。

关键词

语料库, 国防部例行记者会, 军事汉英口译, 词汇特征

A Corpus-Based Study of the Lexical Features in Chinese-English Interpreting of the Regular Press Conference of the Ministry of National Defense

Haiqi Du

School of Foreign Languages of Shanghai Maritime University, Shanghai

Received: Aug. 18th, 2023; accepted: Sep. 28th, 2023; published: Oct. 9th, 2023

Abstract

At present, there are few researches on the regular press conference of the Ministry of National Defense, and most of them focus on some words and sentences, while the whole oral vocabulary is rarely studied. Using the corpus research method and with self-built corpus, this paper takes the Chinese-English interpreting vocabulary of regular press conference of the Ministry of National Defense in the past five years as the research object, and makes a quantitative analysis of four lexical features, namely the standard token/type ratio, high-frequency words, verb nominalization and average word length, and compares them with the same type of US department of defense press conference speech corpus. Then it will analyze the differences between the two and the reasons for the differences and give suggestions. The findings of this study are as follows: Compared with the US press conference, the regular press conference interpreting corpus has higher lexical complexity and lexical density, and the tendency of interpreting and translating is also obvious, which is not consistent with the generality of simplified and standardized translation. The main reason for this phenomenon is that the regular press conference of the government is an important form of public political communication. The spokesman not only needs to convey important policy information but also needs to elaborate the position of the government. Therefore, the words used are more formal and accurate, which also explains that the lexical features of interpretation are more complicated than those of original English to a certain extent.

Keywords

Corpus, The Regular Press Conference of the Ministry of National Defense, Military Chinese-English Interpretation, Lexical Features

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

我们把译语语言和目的语语言规范之间的差异称为“翻译共性”，具体表现为译语语言的显化(explicitation)、简化(simplification)和范化(normalization)倾向[1]。显化指对源语中暗含的、需从上下文推导出的信息在译入语中给与明示[2]。显化最主要的表现是扩充句子和语篇，增添修饰词、额外信息等。简化指“译者无意识地简化语言或信息或两者兼而有之”[3]，换言之就是翻译文本比原创文本的语言更简单，比如用词范围缩小，常用词增加，句型简化等。范化指译者消除原文本中独特的文本特征，用译入语中规范典型的方式来替代，例如，减少特色词，增加常用词，规范标点符号等。

2. 文献综述

2.1. 兴起和发展

Holmes [4]在“The Name and Nature of Translation Studies”一文中提出了诸多翻译研究成为一门真正独立学科的理由。自此以后，对翻译的本质及规律的探索方兴未艾，其中就包括对翻译共性的研究。Baker [5]在“Corpus Linguistics and Translation Studies: Implications and Application”一文中首次提出“翻译共性”的假设，将其定义为翻译文本而非原创文本的典型特征。学术界一般把对“翻译共性”的研究分为前语

料库时期和基于语料库的时期，而此文也正是这二者的分界线[6]。Baker 提出了六大“翻译共性”子假设：1) 显化；2) 消歧及简化；3) 合乎语法规则；4) 避免重复；5) 目标语特征夸大表征；6) 某些语言特征的特定分布。上述子假设中最受学术界关注的当属简化、显化和范化。

就研究视角而言，除关注翻译的过程和结果外，黄立波[7]从翻译文体学视角考察译出、译入两个方向以及合译文本的语言特征，庞双子，王克非[8]利用语料库语言学和描写语言学的结合，考察透过翻译的书面语言接触。就研究路径而言，庞双子，胡开宝[9]认为只有显隐并重，将传统的语言学研究路径和翻译学研究路径二者结合，才能真正探析出翻译的语言显化现象，推动翻译共性研究进步。就研究切入点而言，除了传统的词汇和句法层面的研究，Kruger [10]选择语篇和跨文化交际层面出发，对翻译共性进行综合系统性研究。

2.2. 基于语料库的翻译共性研究

Monker Baker 是语料库翻译研究的创始人之一，她最早提出了运用语料库语言学工具对翻译过程进行描述性研究，住持建立了世界上第一个翻译英语语料库(TEC) [11]。而今天距 Baker 教授于 1993 年提出语料库翻译学已经三十年，这三十年来，Sara Laviosa [12]提出，语料库语言学和翻译研究的结合形成了“一种连贯的、全面而丰富的研究范式。该范式涉及翻译理论、描写和实践等方面的问题，也已成为当代描述性译学研究的新范式”。吴昂，黄立波[13]就语料库方法应用于翻译研究取得的翻译共性研究成果提出问题，认为针对翻译共性的研究不应只局限于翻译本身，而应从具体语言对、译者因素、文体类型等多个角度进行考察。胡显耀，曾佳[14]认为翻译共性研究出现了新的趋势，即译入语独特性降低、源语干扰和异常搭配。董敏，冯德正[15]对汉语科技译文逻辑关系的显化策略实证研究，研究结果证实了汉语翻译文本的显化假设并阐明了语言学动因。

近年来，学者们越来越多地采用语料库的方法进行翻译共性的研究，内容不仅包括笔译范围，也有口译的内容。但是，目前的大部分研究还是聚焦于笔译语料库，且口译语料库的研究以部分词汇篇章的研究为主，因此本文基于口语语料库进行总体的词汇特征研究还是很有必要的。

3. 语料与研究方法

3.1. 研究问题

- 1) 国防部例行记者会口译和美国国防部新闻发布会的词汇特征有哪些差异？
- 2) 造成国防部例行记者会口译和美国国防部新闻发布会词汇特征差异的原因是什么？

3.2. 研究对象

本文以中国国防部例行记者会为核心内容，设计建立了一个军事汉英口译语料库(口译语料)。该语料库分为两个子库。第一子库为军事汉英口译平行语料库，主要收录国防部发言人 2018 年 5 月至 2023 年 5 月共计 56 次例行记者会上的发言与现场口译内容，该内容直接从中国国防部网站下载，共计 248,628 词。第二子库为军事英语原创语料库(原创语料)，收录了美国国防部 2023 年 1 月至 5 月新闻发布会的现场发言和回答。第二子库可直接从美国国防部网站上找到原文，字数共计 258,463 词。第二语料库与第一语料库虽然时间上跨度不同，但是字数差别不大，因此具有一定的可比性。

3.3. 研究工具

本文主要采用两种语料库研究的常用工具。进行语料库数据统计时，主要使用的是 WordSmith，由英国利物浦大学 Mike Scott 设计。利用 AntConc 来对语料库文本进行分析，检索文本高频词汇等。

3.4. 研究步骤

- 1) 首先，自建语料库，网站下载后转化为纯文本；
- 2) 其次，利用 WordSmith 和 AntConc 对语料库进行定量分析；
- 3) 最后，依据数据进行分析并得出结论。

4. 研究结果与讨论

4.1. 词汇丰富度——类符/形符比

类符/形符比(TTR)是用来衡量文本词汇丰富度最重要的标准之一。形符指单词，在一个文本中的形符可以重复计算，类符指的是不重复计算的形符数。单纯的类符和形符数不能反映出文本的本质特征，但是两者的比率可以在一定程度上反映出文本用词的变化性[16]。但是这一比率会受到文本长度的影响，所以本文用标准类符/形符比(STTR)来衡量词汇的丰富度。

Table 1. STTR of interpreting and original corpus
表 1. 口译语料、原创语料的标准类符/形符比

	口译语料	原创语料
形符数	247,580	256,575
类符数	9222	8787
类符/形符比	3.77%	3.45%
标准类符/形符比	41.92%	38.26%

从表 1 可以看出，口译语料库的标准化类符/形符比为 41.92%，大于原创语料的 38.26%。但是，这一差距并不十分明显，也就是说在口译语料和原创语料之间，词汇使用的丰富度上并无明显差异。并且这一结果与简化的翻译共性相反。出现这一现象，一个原因是口译的笔译化倾向，本身就在脑中的表达方式已经具有很高的词汇丰富度，导致即使是口译表达也会有更丰富的词汇。原因二是例行记者会的形式，在中国国防部例行记者会上，记者们会就各个方面进行问答，涵盖范围很广，所以涉及到的词汇自然特别多。因此，词汇复杂度就需要维持在一个水平上。所以，口译语料库的标准类符/形符比高于原创语料库。

4.2. 高频词

Table 2. A comparison of words of high frequency in the interpreting and original corpus
表 2. 口译语料、原创语料的高频词汇比较

次序	口译语料	频数	频率(%)	原创语料	频数	频率(%)
1	THE	19,383	7.80	THE	12,668	4.90
2	AND	12,951	5.21	TO	9811	3.80
3	OF	8725	3.51	AND	6882	2.66
4	TO	6437	2.59	THAT	6416	2.48
5	IN	4561	1.83	OF	6178	2.39
6	CHINA	3419	1.38	IN	4096	1.58
7	A	3204	1.29	YOU	4090	1.58
8	MILITARY	3197	1.28	WE	3921	1.52
9	IS	2387	0.96	I	3913	1.51
10	ON	2380	0.96	A	3571	1.38

词频统计是语料库研究最基本的手段。Sinclair [17]指出, 在语料库中出现的那些高频词的分布都是固定的, 所以即使排列顺序有微小的改变也有可能是有意义的。

首先, 如表 2 所示, 在原创语料库中我们可以看出, 前十位除代词以外没有出现实词, 而在口译语料库中, 第 6 位和第 8 位就分别出现了实词 China 和 Military。这一结果也验证了口译词汇密度较高这一点。China 出现频率如此高, 一是因为中国国防部例行记者会本身传达的是中国政府和军队的观点, 许多表达必是以 China 做主语。二是因为汉语和英语不同, 汉语中存在大量的无主句, 翻译过来的主语往往都是 China、Chinese, 所以 China 的出现频率才会如此高。

其次, 在原创语料库中, You、We、I 出现在了高频词前十位中, 而这三个代词并未出现在口译语料库前十位高频词中, 这一结果也显示了中国国防部例行记者会和美国国防部新闻发布会对话方式的不同。因为在发布会上, 会有大量的发言人与记者的对话, 在对话中以 I、You 互相指称, 但是这样的对话方式在记者会上并不明显。

再次, and 和 the 这两个词在口译语料中出现的频率明显高于在原创语料中出现的频率。这一特点也符合显化和范化的翻译特征。汉语是意合的语言, 英语是形合的语言。在汉语中, 句子与句子之间通常不会有明显的连接词, 一般通过句意连接, 汉语翻译过来往往会需要加一些连接词表示关系, 因此符合显化这一特征。定冠词 the 的使用频率高是因为, 译员会在口译中一些定冠词可有可无的情况下也会加上定冠词, 这比原创语料更加遵守语法规则, 反映了译员的口语规范化。

4.3. 动词的名词化

名词化是语法中产生词汇高密度集中的主要来源, 它具有信息浓缩、表达简洁、结构紧凑和逻辑性强的特征。因此, 名词化大多使用在科技、法律、政论等正式文体中[18]。

Table 3. A comparison of verb nominalization phenomenon in interpreting corpus and original corpus
表 3. 口译语料、原创语料动词名词化现象比较

	口译语料频数	口译语料频率	原创语料频数	原创语料频率
-tion	6069	2.44%	2985	1.15%
-sion	677	0.27%	548	0.21%
-ment	2405	0.97%	1792	0.69%
-ence	754	0.30%	475	0.18%
总计	9905	3.98%	5800	2.23%

表 3 为口译语料与原创语料中“-tion”“-sion”“-ment”和“-ence”这四类最常用的动词名词化后缀的使用情况。由表 3 可知, 口译语料动词名词化使用频率为 3.98%, 明显要高于原创语料 2.23%的使用频率。这四类后缀口译语料使用频率均高于原创语料, 且“-tion”的使用频率高了一倍, 说明在口译中用词往往带有名词化的倾向。Halliday [19]认为, 书面语中名词现象的出现是因为作者可以通过名词化的使用来减少对语境的依赖。并且在展现现实世界时, 可以把它当作客观物体来看待, 文章也会显得更加客观正式。动词名词化中失去了原有的动态形象, 转变成一种抽象的、不涉及任何动作的静态概念。所以, 口译大量使用名词化词语有利于形成客观文体风格, 同时这也反应了口译语料中出现了较为明显的书面语特征, 口译笔译化的趋势也在这里得到了证实。

4.4. 平均词长

平均词长是特定文本中所有符号与字母之间的比率, 换句话说就是文本中以字母为单位测量出的单

词的平均长度。通常来说，一个单词由 2 至 4 个字母组成，平均词长约为 4 个字母。平均词长越短就说明单词越简单，文本也就越容易理解[20]。通过 WordSmith 统计所得数据如表 4 所示。

Table 4. A comparison of average word length in interpreting and original corpus
表 4. 口译语料、原创语料的平均词长比较

研究参数	口译语料	原创语料
平均词长	5.28	4.62
平均词长标准差	2.88	2.59

由表 4 可知，口译语料的平均词长为 5.28 高于原创语料的 4.62，同时口译词长的平均词长标准差为 2.88 也略高于原创语料的 2.59，说明原创语料的词长分布略微稳定一些。另外，口语语料的平均词长高于原创语料，说明其用词整体难度略高于英语的用词习惯，文本对于读者来说存在一定的阅读难度，而母语使用者在口语中更多使用的是较为简单的词汇，避免复杂词汇。

5. 结论

通过以上的语料库对比分析，本文认为造成上述现象的原因有两个因素值得考虑。一是修辞偏好差异。政府记者会是政治传播的重要场所，需要准确表达政府的相关立场，因此用词需要准确正式，语言越正式，文本复杂性也会越高。而西方注重劝说，注重听者，与一定的可协商性因此用词会偏简单一些[21]。二是语境差异，中美记者会的交际特点不同。比如，中国的记者会基本上采取一问一答的模式，讲话风格偏严肃，关系比较疏远。而美国比较随意，发言人与记者互动较频繁，双方关系也较为亲近。因此，中方的口译语言复杂度要高于美方。对此给出的建议是可以提高国防部记者会汉英口译接受度，可以适当降低词汇难度密度，提高口译产品的接受度。还可以引进新的口译质量评估标准，重视听众的接受习惯，更加注重目的语语言词汇特征。从而提高听众接受度，易于军事口译质量的进步，进一步提升军事外宣效果。

参考文献

- [1] 冯建中, 殷东豪. 基于语料库的国防部例行记者会汉英口译词汇特征研究[J]. 外语研究, 2016, 33(3): 69-73. <https://doi.org/10.13978/j.cnki.wywj.2016.03.012>
- [2] Vinay, J.P. and Darbelnet, J. (1958/1977) *Sylistique comparée du français et de l'anglais. Méthode de traduction*, Transl. Juan Sager and M. J. Hamel as *Comparative Stylistics of French and English: A Methodology for Translation*. John Benjamins, Amsterdam and Philadelphia.
- [3] 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>
- [4] Holmes, J.S. (1972/1988) *The Name and Nature of Translation Studies*. In: Holmes, J.S., Ed., *Translated Papers on Literary Translation and Translation Studies*, Rodopi, Amsterdam, 67-80. https://doi.org/10.1163/9789004486669_008
- [5] Baker, M. (1993) *Corpus Linguistics and Translation Studies: Implications and Applications*. In: Baker, M., Francis, G. and Tognini-Bonelli, E., Eds., *Text and Technology: In Honour of John Sinclair*, John Benjamins, Amsterdam, 233-252. <https://doi.org/10.1075/z.64>
- [6] 计霄雯, 刘芹. 国内外翻译共性研究述评与反思[J]. 语言与翻译, 2022(2): 43-46.
- [7] 黄立波. 译出还是译入: 翻译方向探究——基于语料库的翻译文体考察[J]. 外语教学, 2011, 32(2): 96-101. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2011.02.012>
- [8] 庞双子, 王克非. 透过翻译的语言接触研究: 翻译文本的“显化”特征对原生文本的历时影响[J]. 外语教学与研究, 2018, 50(2): 253-267+320-321.
- [9] 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.

-
- [10] Kruger, H. and Van Rooy, B. (2016) Constrained Language: A Multidimensional Analysis of Translated English and a Non-Native Indigenised Variety of English. *English World-Wide*, **37**, 26-57. <https://doi.org/10.1075/eww.37.1.02kru>
 - [11] 胡显耀. 用语料库研究翻译普遍性[J]. 解放军外国语学院学报, 2005(3): 45-48+81.
 - [12] Laviosa, S. (1998) The Corpus-Based Approach: A New Paradigm in Translation Studies. *Meta*, **43**, 474-479. <https://doi.org/10.7202/003424ar>
 - [13] 吴昂, 黄立波. 关于翻译共性的研究[J]. 外语教学与研究, 2006(5): 296-302+320.
 - [14] 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.
 - [15] 董敏, 冯德正. 英汉科技翻译逻辑关系显化策略的语料库研究[J]. 外语教学, 2015, 36(2): 93-96. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2015.02.020>
 - [16] 杨惠中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2002.
 - [17] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press, New York.
 - [18] 杨廷君, 张建理. 中国优秀博士论文英文摘要词汇特征研究[J]. 北京第二外国语学院学报, 2013, 35(4): 1-8.
 - [19] Halliday, M.A.K. (1985) *Spoken and Written Language*. Deakin University Press, Victoria.
 - [20] 祝朝伟, 李润丰. 基于语料库的庞德中国典籍英译译者风格探析[J]. 外语教学, 2023, 44(4): 75-82. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2023.04.018>
 - [21] 符荣波, 王克非. 基于跨模式类比语料库的汉英口译词汇特征研究[J]. 外语教学与研究, 2021, 53(6): 912-923+961. <https://doi.org/10.19923/j.cnki.fltr.2021.06.010>