



Research Paper

# Lexical Analysis between E-C Translated News Articles and Chinese News Articles: A Lexical Grading Perspective

Yanfei Fang

(School of English Studies, Shanghai International Studies University)

Corresponding Author: Yanfei Fang

**ABSTRACT:** The graded lists of Chinese characters and words of English-Chinese Translated News Articles parallel corpus of Yiyao and the self-built Chinese news of People's Daily are formed by AntWordProfile 1.5.1. First, both lists have 10 levels. The higher the frequency level is, the larger the number of characters and word is. Second, the graded lists of Chinese characters and words of Chinese translation of Yiyao and People's Daily (PD) corpora are compared. As for the graded list of Chinese characters, the frequency level and the number of characters used for the same Chinese-featured words in Yiyao are lower and less than those in the PD. The frequency level of English-featured words in both is low, and the number of words in Yiyao is more than that of PD. Unique characters of Yiyao and PD have typical features of English and Chinese respectively. With the comparison of the graded word lists, especially the Chinese and English featured words, the English-to-Chinese translation of Yiyao can be improved, such as the reduction of *de* (auxiliary, is used after the modifier) in the noun modifiers, conjunctions, and *bei* (passivization) sentences. Third, the character and word lists of both translated Chinese and original Chinese texts are enlightening for translation researchers and translators. The former could further analyze the characteristics of translation variants and the latter could select the appropriate translation strategies to improve the translation quality.

**KEYWORDS:** E-C translated news, native Chinese news, word grading, comparison

Received 26 Oct., 2023; Revised 06 Nov., 2023; Accepted 08 Nov., 2023 © The author(s) 2023. Published with open access at [www.questjournals.org](http://www.questjournals.org)

## I. INTRODUCTION

News translation highlights objectivity compared with literary translation (Fang, 2011). Translators need to pay particular attention to cultural differences and the expression habits of the translated language to facilitate the translation of a more readable and higher quality text.

The graded word list and graded word list are lists of individual Chinese characters and Chinese Words respectively and can reflect the linguistic characteristics. Comparing the graded list of Chinese characters (GLCC) in translated texts (TT) and that of source texts (ST) in original language, and graded word list (GWL) in TT and ST, the linguistic difference between TT and ST can be derived, such as Chinese or English featured words, unique words or characters used in the translation. Translators then can further improve the quality of translation according to those differences.

This paper takes English-to-Chinese translation news and original Chinese news as the example to explore their differences by two graded lexical lists, namely, graded character and word lists.

## II. GRADED CHARACTER AND WORD LISTS IN TRANSLATION STUDIES

*The List of Frequently Used Modern Chinese Characters* contains 3,500 commonly used Chinese characters. It was released by the State Language and Character Commission (SLCC) and the State Education Commission of China in 1988. *The General Standardized Chinese Character List*, contains 8,105 characters. It was released by the State Ministry of Education and the State Language Commission of China in 2013. The characters are divided into three levels, with 3,500, 3,000 and 1,605 characters in Level 1, Level 2 and Level 3 respectively. The list meets the needs of Chinese character use in various fields, and has a strong applicability in basic education, information, dictionary compilation, and public services (X. Wang, 2016).

English vocabulary grading is more detailed compared with Chinese word grading. Based on the word frequencies of the British national corpus (BNC) and the Corpus of Contemporary American English (COCA),

Paul Nation (2016) presents the vocabulary lists which include 25,000 words in ten levels of vocabulary. The first 2,000 to 3,000 words overlap with the General Service List (GSL). There are two levels of the General Service List, each containing 1,000 words. These words are commonly used words in English and have a high frequency in BNC and COCA. In the AntWordProfile software (Anthony, 2021), the default graded word lists are two general service word lists and the Academic Word List (AWL).

Word frequency count and grading are important for word information processing and applied research. In translation research, lexical characteristics are adopted to examine translation quality for lexical measurement. Word frequency analysis softwares (e.g., AntWordProfile and Range) are used to derive word lists in translations as a quantitative indicator of word frequency, e.g., J. Wang et al., (2021). In foreign language teaching, the vocabulary grading list can be used in daily teaching and test questions and writing. Foreign language learners can learn vocabulary according to the degree of frequency level to improve learning efficiency (T. Wang, 1987; Huang, 1988).

Comparing the lexical features of TT and ST helps improve the quality of translation. For instance, semantic prosody (Zhu & Hu, 2014), word class (Ding, 2020), syntactic structure (Jiang et al., 2016) and cohesion and coherence (J. Xu and X. Xu, 2016) are studied. In addition, the lexical, syntactic and discursal features of the translated and original texts can be analyzed comprehensively. For example, Shu and Li (2021) compare the English translation of the prose *Sheba Songs* and *Ashima* with the original English prose *Beowulf*, and analyze the lexical diversity and syntactic complexity, high-frequency words and low-frequency words of the three texts, and further explore the reasons.

### III. RESEARCH DESIGN

The corpora consist of the main corpus of the news of Yiyen English-Chinese Parallel Corpus (*Yiyen* hereinafter) (X. Xu and J. Xu, 2021) and two self-built People’s Daily corpora (*PD* hereinafter) as the reference corpora. *Yiyen* was created by Xu Xiuling and Xu Jiajin, from Beijing Foreign Studies University, based on the model of the Brown corpus. The selected news corpus mainly consists of three parts: news reports, editorials and newspaper commentaries. The English-to-Chinese (E-C) translation and the original English texts are kept aligned. *PD* corpora are original Chinese texts and also composed of the above three parts. The time of news release in the two *PD* corpora is close to that of the *Yiyen* corpus (from 2008 to 2015). The number of tokens and file size of main corpus and reference ones are close (see Table 1).

**Table 1:** The composition of *Yiyen* and *PD* corpora

	<i>Yiyen</i>		<i>PD-1</i>		<i>PD-2</i>			
	Original English text	E-C translation	Original Chinese text	Original Chinese text				
<b>Sub-section</b>	types	tokens	types	tokens	types	tokens	types	tokens
news reports	11,431	89,621	2755	181,510	2,711	180,797	2,855	187,018
editorials	8,475	54,957	2497	108,645	2,513	91,961	2,604	91,441
newspaper commentaries	7,448	36,035	2548	78,410	2,218	80,495	2,460	79,171
Total tokens	180,613		368,565		353,253		357,630	
File size	1.4M		1.1M		1.2M		1.2M	

The following three questions are answered: (1) What are the similarities and differences between the Chinese translation of news from *Yiyen* and the original texts of Chinese news from *PD* in terms of featured characters and unique characters? (2) What are the differences of the featured words in different frequency levels among the corpora? (3) What are the implications of lexical grading for E-C translation of news?

#### IV. THE GRADED LIST OF CHINESE CHARACTERS (GLCC) AND GRADED WORD LIST (GWL) IN *YIYAN* AND *PD* CORPORA

##### 4.1 GLCC: English-featured and Chinese-featured characters, and unique characters

GLCC is formed by the frequency and cumulative percentages of the single words in the E-C translation of *Yiyan*, *PD-1* and *PD-2*. The frequency levels and the span of word frequencies are as follows: 0.01% to 10.00% for Level 1 characters, 10.01% to 20.00% for Level 2 characters, 20.01% to 30.00% for Level 3 characters, etc. Characters in Level 1 are the most frequently used ones while characters in Level 10 are the least frequently used ones. GLCCs of *Yiyan* and *PD* are formed partly shown in Table 2.

Table 2: GLCCs of *Yiyan* and *PD* (part)

Level	count	<i>Yiyan</i> Chinese	count	<i>PD-1</i> Chinese	count	<i>PD-2</i> Chinese
Level 1	5	的 一 在 是 人	9	的 国 中 和 一 发 在 人 是	9	的 一 国 在 人 中 是 了 有
Level 2	14	这 了 他 国 有 不 们 个 中 和 为 我 会 大	19	会 大 展 有 不 了 要 年 为 业 作 新 进 全 经 建 实 生 对	23	不 和 大 年 为 会 发 上 业 行 日 要 对 出 生 作 法 地 个 以 到 民 这
Level 3	26	年 到 上 来 时 以 对 能 家 说 美 斯 地 作 要 而 经 发 出 可 生 多 于 自 者 过	28	民 出 化 地 政 力 学 成 上 能 主 工 社 家 部 行 重 个 动 方 理 以 体 来 加 济 高 这 产 日 合 到 开 党 区	32	工 能 家 时 市 来 公 成 部 政 新 进 用 方 全 动 实 经 现 多 力 化 机 于 月 关 等 务 学 理 过 体 本 长 区 高 展 前 网 建
Level 4	41	学 行 就 些 也 但 成 得 现 最 后 之 其 如	39	设 长 关 强 时 资 现 基 公 问 治 平 事 领 推 心 各 前 共 系 水 员 创 度 增 电 们	44	与 者 面 也 电 就 主 加 设 海 报 目 事 心 得 其 问 量 城 管 题 记 据 而 从 系 改 入 外 还 资 农 更 向 性 道 所 着 万 医 北 联 持 如 并 解 门 水 种 府 已 意 统 代 活 但
Level 5	57	她 被 奥 当 着 比 子 动 很 克 利 力 主 与 拉 面 新 法 德 小 影 外 无 亚 看 起 长 体 果 尼 想 文 入 名 好 正 总 科 布 或 格	56	元 项 交 品 里 其 价 省 才 施 流 万 想 传 如 府 所 门 示 投 海 论 第 信 取 天 域 此 今 计 运 办 名 先 由 极 权 融 件 接 南 处 李 格 纪 律 走 团 果 去 克 划 手 据 少 物 林 又 康 营 率 么 超 充 兴 王 直 住 较 培 遇 县 备 已 须 模 判 青 清 双 王 盟 乡 则	62	风 然 放 认 被 决 今 接 使 州 院 老 没 处 结 传 由 少 斯 际 李 二 江 形 想 整 再 况 县 军 条 张 校 乡 环 俄 投 别 举 界 给 空 知 类 县 观 究 势 型 协 史 便 察 税 站 航 围 销 则 居 望 响 负 普 余 巴 官 米 货 铁 行 远 财 复 配 王 容 健 那 污
Level 6	80	两 展 目 什 少 持 基 究 且 至 导 欧 片 量 伊 向 济 平 区 何 金 应 管 件 色 式 空 保 求 语 元 二 儿 确 图 响 官 仅 立 取 设 支 未 再 流 项 瑞 包 孩 组 打 显 改 权 失 曾 施 识 境 居 岁 住 贝 功 脑 预 院 博 欢 络 往 否 姆 试 亿 装 况 群 严 音 张 降 走 萨 穿 广 菲	75	元 项 交 品 里 其 价 省 才 施 流 万 想 传 如 府 所 门 示 投 海 论 第 信 取 天 域 此 今 计 运 办 名 先 由 极 权 融 件 接 南 处 李 格 纪 律 走 团 果 去 克 划 手 据 少 物 林 又 康 营 率 么 超 充 兴 王 直 住 较 培 遇 县 备 已 须 模 判 青 清 双 王 盟 乡 则	80	危 宝 咱 丝 鲜 害 假 竞 姓 哈 讯 杨 俩 剧 拿 握 啥 彻 吃 附 挺 露 星 帮 订 滑 荣 透 驻 背 储 累
Level 7	118	潘 射 退 雪 跟 判 莎 校 养 艺 宇 章 爆 措 宫 娜 束 徒 赢 遇 赞 初 房 估 河	102	扶 七 请 贡 哈 挺 扬 扎 哈 丰 跨 涉 顺 俩 吸 咱 厂 腐 套 挑 写 艳 纳 派 绣	111	
Level 8	186		151		171	
Level 9	335		287		315	
Level 10	2,406		2,439		2,571	

The number of types of *Yiyan*, *PD-1*, and *PD-2* are 3268, 3205, and 3419, respectively. In GLCCs, the numbers of characters of three corpora tend to increase as the frequency levels increase, and the numbers of characters in each frequency level among three corpora are relatively close.

##### English-featured characters and Chinese-featured characters:

The high-frequency featured characters reflect the language features in the original and translated news. In Table 2, the words underlined with a horizontal line are the characters that keep the feature of English in the E-C translation that in *Yiyan*, and the words underlined with a wavy line are the Chinese-featured words in the original news in *PD*. Part of English-featured characters and Chinese-featured characters as well as their counts are shown indicated in Table 3 with the Level in the bracket.

**Table 3: GLCCs of *Yiyan* and *PD***

English- featured characters	<i>E-C translation of Yiyan</i>	<i>PD-1</i>	<i>PD-2</i>	Chinese- featured characters	<i>E-C translation of Yiyan</i>	<i>PD-1</i>	<i>PD-2</i>
奥	498 ( 5 )	102 ( 9 )	127 ( 9 )	咱	2	4	7
克	590 ( 5 )	279 ( 8 )	226 ( 8 )	俩	4	6	5
拉	464 ( 5 )	186 ( 8 )	186 ( 8 )	啥	2	7	8
姆	124 ( 9 )	21 ( 10 )	57 ( 10 )	咋	3	6	9
萨	87 ( 9 )	43 ( 10 )	32 ( 10 )	挺	2	13	16

The English-featured characters, such as “奥” (*ao*), “克” (*ke*), “拉” (*la*) and “姆” (*mu*), are mostly the transliteration of names of people or places in the news. They are mainly distributed in the fifth to tenth frequency levels. Although they exist in three corpora, the frequency levels in *Yiyan* are higher than those in the two *PD* reference corpora, and the number of English-featured characters in *Yiyan* is also larger than that of two *PD* corpora. This indicates that E-C translation of *Yiyan* retains the characteristics of the original English news and uses the English-featured words more frequently.

The Chinese-featured characters, such as “咱” (*zan*, we), “俩” (*lia*, two), “啥” (*sha*, what) and “挺” (*ting*, much) are mainly from Level 6 to Level 10. The numbers of them in two *PD* reference corpora are larger than that in *Yiyan*. The frequency of Chinese-featured characters used in E-C translations is lower than that of the original Chinese news of *PD*. For instance, there are two “咱” (*zan*, we) in *Yiyan*. “We” in (1a) is translated as “咱们” (*zanmen*, we) in (1b), a more informal way, to be closer to readers. “咱们” (*zanmen*, we) in (2a), is an addition to the translation in (2b).

(1a) The students are framed as coddled infants, as if perhaps we should educate college students in a more spartan manner -- placing classrooms in lions' dens. (From editorials)

(1b) E-C translation in *Yiyan*: 抗议学生被刻画成熊孩子，简直好像咱们教育大学生是想让他们当斯巴达勇士一样——索性把教室搁在狮子洞里好了。

(2a) “Even the gender conditioning was better back then!” (From news report)

(2b) E-C translation in *Yiyan*: “咱们那时候的东西连性别针对性都更强！”

### Unique characters:

A unique character is a word that is used only in that corpus with the ten levels of another corpus as a reference. It is shown in AntWordProfile 1.5.1 as level 0. There are 538 unique characters (total character frequency reaches 1,602) of *Yiyan* when *PD* corpora are references, including “猎 (*lie*), 卦 (*gua*), 詹 (*zhan*), 谍 (*die*), 猜 (*cai*), 绑 (*bang*), 窃 (*qie*), 狄 (*di*), 芭 (*ba*), 渲 (*xuan*), 珀 (*bo*), 琰 (*yan*), 侣 (*lv*), 烷 (*wan*), 狙 (*ju*), 浏 (*liu*), 庇 (*bi*), 掠 (*lue*) and 狩 (*shou*)”. Some are English-featured characters. Similarly, unique characters of *PD-1* involve “龚 (*gong*), 龄 (*ling*), 鼎 (*ding*), 黔 (*qian*), 鹭 (*lu*), 麒 (*qi*) and 龔 (*yan*)” while unique characters of *PD-2* include “魏 (*wei*), 魂 (*hun*), 魄 (*po*), 鹤 (*he*), 龚 (*gong*), 鼎 (*ding*), 鸪 (*zhen*), 麾 (*hui*) and 黝 (*you*)”. The unique characters of both *PD* corpora have vivid Chinese features. The original news and the E-C translation of the news can be clearly distinguished by the unique characters.

4.2 GWL: English-featured and Chinese-featured words, and unique characters

In addition to GLCC, word frequency grading also plays an important role in the comparison of translated texts and original texts shown in GWWL in Table 4. *Yiyan* and *PD* are imported into AntWordProfile 1.5.1 for lexical grading. The words are graded into ten levels. The cumulative percentage of Level 1 is from 0.01% to 10.00%, Level 2 is from 10.01% to 20.00%, Level 3 is from 20.01% to 30.00%, etc. Some words in each level are shown in Table 4. Those with a higher frequency in E-C translation of *Yiyan* than those in *PD* are marked as horizontal underlines, and those with a lower frequency than in *PD* are not marked. The number of words in each frequency level is: Level 1: 5; Level 2: 43; Level 3: 120; Level 4: 223; Level 5: 388; Level 6: 704; Level 7: 1304; Level 8: 2,726; Level 9: 6,948; Level 10: 31,212.

Table 4: English-featured and Chinese-featured words in *Yiyan* and *PD*

Word	Level	Token	<i>Yiyan</i>	<i>PD-1</i>	<i>PD-2</i>	Word	Level	Token	<i>Yiyan</i>	<i>PD-1</i>	<i>PD-2</i>
的	1	31,547	<u>12,335</u>	9,290	9,922	对于	4	327	<u>130</u>	78	119
在	1	6,656	<u>2,704</u>	1,815	2,137	实施	4	326	14	185	127
是	1	4,321	<u>1,679</u>	1,287	1,355	美元	4	313	<u>146</u>	63	104
我们	2	1,320	<u>730</u>	303	287	中央	4	309	16	191	102
他们	2	967	<u>685</u>	120	162	能力	4	300	31	155	114
被	2	766	<u>400</u>	138	228	而且	4	218	<u>113</u>	38	47
她	3	671	<u>457</u>	123	91	尽管	5	186	<u>122</u>	23	41
服务	3	619	<u>48</u>	265	306	比较	6	104	18	42	44
人民	3	574	<u>43</u>	199	332	即使	6	67	<u>39</u>	13	15
但是	3	527	<u>409</u>	41	77	物流	7	58	1	20	37
群众	3	439	<u>1</u>	245	193	领土	8	31	<u>16</u>	5	10
如果	3	432	<u>222</u>	85	125	军	8	30	2	14	14
因为	3	363	<u>232</u>	41	90	视角	9	14	9	3	2

In Level 1, the numbers of “的” (auxiliary, is used after the modifier), “在” (in) and “是” (be) in *Yiyan* is more than those in two PD corpora, with the gap of tokens of 3000, 1000 and 300, respectively. This shows that translators should reduce the use of “的” (auxiliary, is used after the modifier), “在” (in) and “是” (be) in E-C translation.

We import parallel texts from *Yiyan* and retrieve the high-frequency words in BFSU Paraconc 1.2.1 to compare the high-frequency word characteristics of translation with those of the original Chinese news. The search for “的” (auxiliary, is used after the modifier) yielded 6,262 sentences. In E-C translation of *Yiyan*, there were 12,335 “的” (auxiliary, is used after the modifier) in 6,262 sentences, with an average of 2 “的” (auxiliary, is used after the modifier) per sentence. The average number of “的” (auxiliary, is used after the modifier) per sentence is 2.

In (4), there are four “的” in the translation. They can be reduced. For example, “的” in “日本和韩国的领导人” (*riben he hanguo de lingdaoren*, the leaders of Japan and South Korea) and “两国领导的能量” (*liangguo lingdao de nengliang*, the energy of the leaders of the two countries) can be omitted. “极低的支持率” (*jidi de zhichi lv*, extremely low approval rate) and “拖住后腿的状态” (*tuozhu houtui de zhuangtai*, the state of holding back) can also be omitted by restructuring. “领导人” (leaders) appeared 128 times in two PD corpora. Among them, 122 are not preceded by “的”. As shown in Figure 1 of Antconc 3.5.9, there are “中韩领导人” (*zhong han lingdaoren*, leaders of China and South Korea), “中国与东盟领导人” (*zhongguo yu dongmeng lingdaoren*, leaders of China and ASEAN), etc. “的” is often in noun modifiers which can be adjectives, nouns, prepositional phrases, definite clauses, multilayer prepositions, V-ed participles, V-ing participles, complements and apposition (Zhu, 2021; Biber and Gray, 2010). In Chinese translation, words or phrases with “的” are often translated. Based on the comparison, the use of “的” to in the translation of noun modifiers should be reduced.



(4a) Clinton’s trip occurs when the Japanese and South Korean leaders are weakened by dismally low approval ratings and hamstrung by legislative inertia.

(4b) E-C translation in *Yiyan*:克林顿此次出访正值日本和韩国的领导人处于极低的支持率和被立法程序拖住后腿的状态中，因而两国领导的能量都被削弱。

Concordance Hits 128			File
Hit	KWIC		
1	... 潘实习生 焦惠芳 ) 中共 新一 届	领导人 上任 以来 对 公职人员 进行 史	r mrb_2 【分词】.txt
2	领导人 对话 的 大门, 对于 这样 的	领导人, 中国 人民 是 不 欢迎 的。	r mrb_2 【分词】.txt
3	优秀 儿子, 我国 民族 工作 的 杰出	领导人, 中国 共产 党 的 亲 密 朋 友,	r mrb_1 【分词】.txt
4	拜 靖 国 神 社 后 表 示 希 望 会 晤 中 韩	领导人, 中 方 很 快 做 出 可 以 预 见 的	r mrb_2 【分词】.txt
5	原则 共 识 创 造 气 氛 和 条 件。 中 芬	领导人 互 致 建 交 60 周 年 贺 电 新 华 社	r mrb_1 【分词】.txt
6	健康 稳 定 发 展。 进 入 新 世 纪, 两 国	领导人 互 访 不 断, 政 治 关 系 更 加 成 熟	r mrb_1 【分词】.txt
7	会, 同 尼 政 府、 政 党 和 制 宪 会 议	领导人 亲 切 交 谈, 就 中 尼 关 系 广 泛	r mrb_1 【分词】.txt
2	其 他 20 个 经 济 体 领 导 人 ( 或 经 济	领导人 代 表 ) 出 席 峰 会。 截 至 目 前	r mrb_2 【分词】.txt
3	参 拜。 过 去 被 看 做 有 问 题 言 行 的	领导人 会 因 引 发 外 交 争 端 引 咎 辞 职	r mrb_1 【分词】.txt
4	议 题。 此 次 会 议 在 第 十 二 次 中 欧	领导人 会 晤 前 夕 召 开, 是 中 欧 就 对	r mrb_1 【分词】.txt
5	议 题。 此 次 会 议 在 第 十 二 次 中 欧	领导人 会 晤 前 夕 召 开, 是 中 欧 就 对	r mrb_1 【分词】.txt
6	至 13 日 在 美 国 夏 威 夷 举 行。 此 次	领导人 会 议 主 题 是 “ 紧 密 联 系 的 区 域	r mrb_1 【分词】.txt
7	河 内 出 席 了 第 十 三 次 中 国 与 东 盟	领导人 会 议 并 发 表 讲 话。 温 家 宝 说	r mrb_1 【分词】.txt
8	内 举 行 的 第 十 三 次 东 盟 与 中 日 韩	领导人 会 议 并 讲 话。 温 家 宝 说,	r mrb_2 【分词】.txt
9	发 展 进 步 温 家 宝 出 席 中 国 与 东 盟	领导人 会 议 本 报 河 内 10 月 29 日	r mrb_1 【分词】.txt
8	议 式。 温 家 宝 出 席 东 盟 与 中 日 韩	领导人 会 议 本 报 河 内 10 月 29 日	r mrb_1 【分词】.txt
10	将 呼 吁 各 方 按 照 亚 太 经 合 组 织 历 次	领导人 会 议 达 成 的 共 识, 坚 定 地 维	r mrb_1 【分词】.txt
18	将 呼 吁 各 方 按 照 亚 太 经 合 组 织 历 次	领导人 会 议 达 成 的 共 识, 坚 定 地 维	r mrb_1 【分词】.txt
19	得 积 极 务 实 成 果: 一 是 落 实 去 年	领导人 会 议 通 过 的 《 APEC 领 导 人 会	r mrb_1 【分词】.txt
20	参 拜 靖 国 神 社, 并 称 想 与 中 韩	领导人 会 谈。 但 是 整 个 讲 话 中, 与	r mrb_2 【分词】.txt
21	与 尼 泊 尔 政 府、 政 党 和 制 宪 会 议	领导人 会 面 1 月 14 日, 国 务 院 总 理	r mrb_1 【分词】.txt
22	表 示 要 ‘ 不 附 带 任 何 条 件 与 中 韩	领导人 会 面 ’, 这 是 一 种 要 挟 行 为	r mrb_2 【分词】.txt
23	以 来 第 八 次 访 华, 指 出 中 德 两 国	领导人 保 持 密 切 交 往 体 现 了 两 国 关 系	r mrb_1 【分词】.txt
24	纪 的 挑 战。 “ 世 界 最 强 大 国 家 的	领导人 做 出 反 思。 变 革 举 步 维 艰。	r mrb_1 【分词】.txt
25	合 志 朋 友 会 后 温 家 宝 与 东 盟	领导人 共 同 出 席 了 中 国 一 东 盟 虚 拟	r mrb_1 【分词】.txt

Figure 1: Search results for “领导人” (leaders) in PD corpora in Antconc

In Level 2, there are “我们” (*wo'men*, we), “问题” (*wenti*, problem), “社会” (*shehui*, society), “他们” (*tamen*, they), “被” (*bei*, passivization), “国家” (*guojia*, cpuntry), “工作” (*gongzuo*, work) and “经济” (*jingji*, economy). The numbers of personal pronouns “我们” (*wo'men*,we) and “他们” (*tamen*, they) in *Yiyan* are more than those of *PD* corpora, with the gap of 400 and 500 respectively. In *Yiyan*, there are 528 sentences containing “我们” (we) as the word frequency reaches 730 times. Among them, 445 (84.28%) are subject, 75 (14.20%) are modifiers and 8 (1.52%) are object. For instance, “我们” (*wo'men*,we) appears twice in both (5b) and (6b). However, the first “我们” (*wo'men*, we) in both cases can be omitted in the translation. On the contrary, the number of “我们” (*wo'men*, we) in the two *PD* reference corpora is 500, which is still lower than that in *Yiyan*. The distribution is as follows: 308 (61.60%) as subject; 182 (36.40%) as modifiers; 10 (2.00%) as object. It can be seen that the subject “我们” (we) should be appropriately reduced or even be omitted in the translation.

(5a) If we could turn the clock back, we would not have let that occur. (From news report)

(5b) E-C translation in *Yiyan*:如果**我们**能让时钟倒转，**我们**不会让它发生。

(6a) “In some cases, we would say we have a confident result,” Dr. Stott said. (From news report)

(6b) E-C translation in *Yiyan*:斯托特博士说：“某些情况下，**我们**可以说**我们**有信心下定论。

In addition, “被” (*bei*, passivization) in Level 2 should be noted. The number of it is 400 in *Yiyan* while reaching only 138 and 228 respectively in *PD-1* and *PD-2*. Merely 10% keep the passivization in E-C passive translation and they are turned into active, subject-object structure, verb-object structure and omission, etc. (Wang and Liu, 2018). The translation of *Yiyan* uses passive structure excessively, which makes the

number of the word “被” (*bei*, passivization) exceeds that of *PD*. In (7), both “be viewed as” and “be seen as” are translated as “被视为” (*bei shi wei*, be viewed as). The deletion of the word “被” (*bei*, passivization) does not affect the meaning of the sentence and even makes the sentence clearer.

(7a) Such polling could be viewed as a preemptive defense against a charge of exclusiveness, but it can also be seen as an attempt to purvey common sense, rather than snobbery or defensiveness, and, in the end, just to tell the truth. (From newspaper commentaries)

(7b) E-C translation in *Yiyan*: 如此投票可被视为对于“排外性”指控的一种先发制人式防御；但也可以被视为提供常识的尝试，而非对上谄媚或防御。说到底，它就是说出了真相而已。

In Level 3, the words “她” (*ta*, she), “但是” (*danshi*, but), “如果” (*ruguo*, if) and “因为” (*yinwei*, because) in C-E translation of *Yiyan* are much more than those of *PD*, with the number of them reaching about 350, 350, 100, and 170 respectively. “她” (*ta*, she) is similar to the pronouns “我们” (*wo’men*, we), and “他们” (*tamen*, they) in Level 2. It is mainly in subject and is not omitted in the translation. The words “但是” (but), “如果” (*ruguo*, if) and “因为” (*yinwei*, because) are conjunctions. They “links clauses, sentences or groups of sentences” (Guo, 2002, p. 235). Linking devices and forms in English are varied since it highlights explicit coherence with a strong focus on form, while Chinese emphasizes meaning and the linking component is not necessary (Lian, 2010).

(8a) Kan, who favors joining the TPP talks, faces resistance from his own party amid a backlash from Japan’s farmers who benefit from tariff protection. (From news report)

(8b) E-C translation in *Yiyan*: 菅直人希望加入泛太平洋战略伙伴合作谈判，但是面对本党派的反对，因为日本农民会从关税保护中受益。

Adopting BFSU Paraconc 1.2.1, we retrieved the parallel corpus between English and Chinese in *Yiyan*. 406 sentences contain “但是” (*danshi*, but), 222 sentences contain “如果” (*ruguo*, if), and 216 sentences contain “因为” (*yinwei*, because). For example, there is no obvious conjunction in the original text in (8a), while translation adopts two consecutive conjunctions of “但是” (*danshi*, but) and “因为” (*yinwei*, because) to reflect the logic of the sentences in (8b).

Similarly, the numbers of “而且” (*erqie*, and) and “因此” (*yinci*, therefore) in Level 4, “即使” (*jishi*, even if) in Level 6, and “即使是” (*jishi shi*, even if) in Level 9 in E-C translation of *Yiyan* are more than those in *PD*. Explicit words of coherence and cohesion in translation can reduce the reader’s thinking and inferring efforts, while it can reduce the fluency and authenticity of the translation (J. Xu and X. Xu, 2016, p. 101).

## V. CONCLUSIONS

By comparing GLCC and GWL of *Yiyan* and *PD* corpora obtained by AntWordProfile 1.5.1, and analyzing the featured characters, unique characters and featured words, the following conclusions can be drawn: (1) GLCC of *Yiyan* and *PD* show that the higher the frequency level is, the larger the number of characters. The frequency level and the number of words used in the Chinese translation of the same English words are more than those in *PD* corpora; the number of unique characters in the E-C translation (e.g., “咱” (*zan*, we), “啥” (*sha*, what), “挺” (*ting*, much) and “咋” (*za*, what) is less than those in the original Chinese news of *PD* and they are all in the lower frequency level; 538 characters in *Yiyan* are unique words with English features while unique words of *PD* have obvious Chinese features. (2) With the compassion of GWL of *Yiyan* and *PD*, the number of “的” (*de*, auxiliary, is used after the modifier), pronouns in subject, explicit conjunctions and “被” (*bei*, passivization) in translation are much more than those in the original text. (3) Lexical grading is enlightening for both researchers and translators of news translation. It can not only help translation researchers provide a new research perspective to understand the actual situation of translation variants, but also provide translators with convenience in word selection and word usage in translation practice.

## REFERENCES

- [1] Anthony, L. AntWordProfiler (Version 1.5.1) [Computer Software]. Tokyo, Japan: Waseda University, 2021. Available from <https://www.laurenceanthony.net/software>
- [2] Biber, D. & Gray, B. Challenging stereotypes about academic writing: complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 2010. (1): p. 2-20.
- [3] Ding, G. Q. A corpus-based study on word class relations in English-Chinese translation. *Foreign Language Teaching and Research*, 2020. **52**(5): p. 773-785+801.
- [4] Fang, M. Z. *A Dictionary of Translation Studies in China*. Shanghai: Shanghai Foreign Language Education Press. 2011.
- [5] Guo, R. *A study of word class in modern Chinese*. Beijing: The Commercial Press. 2002.
- [6] Huang, B. G. *University English syllabus (science and engineering) --An attempt to grade vocabulary from Levels 1 to 4 of the syllabus*. *Foreign LanguageWorld*, 1988. (2): p. 13-17.
- [7] Huang, Y. Y. Adhering to the principle of “Three Close proximity to foreign propaganda” and dealing with the difficult problems in foreign propaganda translation. *Chinese Translators Journal*, 2004. **25**(6): p. 29-30.
- [8] Jiang, Y., Zhang, Y. X. & Han, H. J. A vector distance algorithm for similarity between human translation and online translation of English passives. *Journal of Xi’an Jiaotong University(Social Sciences)*, 2016. **36**(3): p. 110-116.
- [9] Lian, S. N. *A comparative study of English and Chinese*. Beijing: Higher Education Press. 2010.
- [10] Nation, P. *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins Publishing Company, 2016.
- [11] Shu, J. & Li, W. Linguistic features of English translation of Sheba Songs based on corpus. *Journal of South-Central University for Nationalities (Humanities and Social Sciences)*, 2021. **41**(4): p. 155-160.
- [12] Wang, J. Q., Yu, X. & Wu, W. N. A study on translation quality evaluation based on lexicometric features. *Chinese Translators Journal*, 2021. (5): p. 113-120.
- [13] Wang, K. F. & Liu, D. J. A study on the Chinese translations of English passives: A super-sized English-Chinese parallel corpus-based approach. *Journal of Foreign Languages*, 2018. **41**(6): p. 79-90.
- [14] Wang, T. F. Principles and practices of compiling graded basic vocabulary lists. *Foreign LanguageWorld*, 1987. (2): p. 9-12.
- [15] Wang, X. M. On the Social Dimension of The General Purpose Normalized List of Chinese Characters. *Applied Linguistics*, 2016. (3): p. 12-19.
- [16] Xu, J. J. & Xu, X. L. A comparable corpus based investigation into the cohesive explicitation of translational English. *Foreign Languages and Their Teaching*, 2016. (6): p. 94-102+122+150.
- [17] Xu, X. L. & Xu, J. J. The building of Yiyen English-Chinese Parallel Corpus. *Corpus Linguistics*, 2021. **8**(1): p. 149-151.
- [18] Zhu, Y. F. & Hu, K. B. The semantic preferences and semantic prosody of bei passives — A corpus-based contrastive study. *Journal of Foreign Language*, 2014. **37**(1): p. 53-64.
- [19] Zhu, Y. F. & Li, X. A quantitative study on lexical features of translated Chinese: Based on the corpus of E-C translated news articles and the corpus of Chinese news articles. *Foreign Languages in China*, 2019. **16**(2): p. 81-90.
- [20] Zhu, Z. Y. The relationship between noun phrase complexity and L2 Learners’ Chinese-to-English translation proficiency and quality. *Journal of PLA University of Foreign Languages*, 2021. **44**(5): p. 119-127.