

A Corpus-based Study on Translation Universals: Taking *Call to Arms* as an Example

Hongjuan Yang, Linfang Zhang*

College of Foreign Languages, Shandong University of Science and Technology, Qingdao 266590, Shandong, China

*Correspondence Author

Abstract: *With the rapid development of corpus linguistics, corpus-based research on translation universals has become a hot spot in the field of translation studies. The studies of translation universals are conducive to better understanding the translation process and provide guidance for translation practice. This study takes Call to Arms as the research corpus and original English novels as the reference corpus. Based on comparable corpora, this research compares Call to Arms with original English novels at both lexical and syntactic levels to analyze the translation universals embodied in Call to Arms. It is found that in terms of lexical richness, lexical density and average word length, Call to Arms reflects the tendency towards simplification. From the perspective of high-frequency words, the explicitation is validated in Call to Arms. With respect to average sentence length, the simplification of Call to Arms is not confirmed. From the aspect of connective structures, Call to Arms shows the tendency towards explicitation and normalization. This study attempts to provide a useful reference for corpus-based studies on translation universals and benefit relevant translation practice.*

Keywords: Translation universals; *Call to Arms*; Corpus.

1. INTRODUCTION

With the development of corpus linguistics, a wide range of translation studies has been carried out. Among them, corpus-based research on language comparison and translation universals has gradually become a hot spot. In studies on translation universals in English translations of Chinese novels, the research objects are mostly contemporary Chinese novels from the 21st century, while there is less research on modern vernacular Chinese novels. To date, there have been few studies focusing on the exploration of translation universals with *Call to Arms* as the research object.

Call to Arms is a collection of Chinese modern vernacular novels written by Mr. Lu Xun. Regarded as a milestone in the history of Chinese literature, *Call to Arms* has a profound impact on other writers who write in vernacular. It is also selected as the top Chinese novel of the 20th century by Asiaweek. *Call to Arms* occupies an important position in the literary arena at home and abroad. Moreover, at present, the studies on *Call to Arms* mainly center on qualitative translation analysis, and thus the research on translation universals combined with quantitative analysis is still a new attempt. Therefore, this study takes *Call to Arms* as the research object and conducts a comparative analysis between it and original English novels in terms of vocabulary and syntax to explore translation universals reflected in *Call to Arms*, aiming to provide useful references and insights for corpus-based research on translation universals.

2. PREVIOUS STUDIES ON TRANSLATION UNIVERSALS

Translation universals, also known as universal features of translation, are first proposed by Baker as “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (243). Translation universals refer to the fact that the translated language, as an objective variety of language, exhibits some regular features on the whole with respect to the original target language (Wang 44). Chesterman divides translation universals into two categories: S-universals and T-universals (7). The former is based on the interlanguage contrastive relationship between the source text and the translated text, which focuses on the way the translator handles the source text. The latter is based on the intralingual analogy between the translated text and the original text in the target language. It focuses on the translator’s treatment of the target language (Wang and Huang 102).

After many empirical studies, translation universals are mainly summarized as simplification, explicitation, normalization and so on, which may be reflected in the lexical and syntactic aspects of translated texts. Explicitation refers to the tendency to express information that may be implicit in the source text. In addition, explicitation is not only reflected in form but also in meaning. Explicitation in meaning refers to the fact that the

implicit meaning of the source text is made explicit in the translation, making the meaning preciser and the logic clearer (Ke 306). Simplification is the translator's subconscious simplification of the lexical, syntactic and stylistic aspects of the source text in translation and the translated text shows an overall tendency to be less difficult than the source text or the original text in the target language (Wang and Huang 102). Normalization refers to the tendency of translated language to follow or even exaggerate typical features and expressions in the target language (Baker 183). The unique textual features of the source text are consciously or unconsciously transformed by translators to be consistent with the traditional textual features of the target language, which means eliminating the uniqueness of the source text to make the translated text normative.

The results of empirical studies on translation universals are inconsistent, ranging from proof to denial of the assumptions of translation universals. And the research on translation universals covers various types of translated texts. For example, Feng et al. investigate collocations in the translations of business-related texts through comparable corpora to examine the translation universals of simplification and explicitation embodied in these business-related translations from Chinese to English (677). The findings of this study confirm the simplification and explicitation. Moreover, collocations seem to be simpler in form and clearer in meaning in translated texts than in original English texts. As for machine-translated texts, a Chinese-English translation corpus generated by WeChat Translator, an online machine translation programme for the Chinese social media WeChat, is studied to explore whether and to what extent these Chinese-English translations exhibit the tendency towards simplification and explicitation. And the results show that although simplification can not be confirmed, the trend of normalization in WeChat translations can be verified (Luo and Li 31). There is not yet a consensus on the specific contents of translation universals, but the relevant theoretical elaboration and empirical studies mainly focus on the three aspects of explicitation, simplification and normalization.

In the studies on translation universals in English translations of Chinese novels, the focus is mostly on contemporary Chinese novels from the 21st century, whereas the research on modern vernacular Chinese novels is relatively scarce. Moreover, there has been a lack of research on translation universals with *Call to Arms* as the research object. In view of this, based on comparable corpora, this study compares *Call to Arms* with original English novels at the lexical and syntactic levels and attempts to explore the universal features of translation embodied in *Call to Arms*.

3. RESEARCH DESIGN

The research corpus of this study is a self-built English translation corpus (named Corpus of *Call to Arms*, abbreviated as CCA). The English version selected for this study is the well-known one translated by Yang Xianyi and Gladys Yang, which was published in 1981. According to statistics, the number of tokens in CCA is 50,517.

The reference corpus used in this study (named Corpus of Original English Novels, abbreviated as COEN), is a collection that includes the novel subcorpus of the Crown Corpus and the novel subcorpus of the CLOB Corpus. The CLOB Corpus is a one-million-word Brown family corpus of British English, while the Crown Corpus is a one-million-word Brown family corpus of American English. The reference corpus, i.e., COEN, contains various types of novels, with a token count of 507,067, making it representative to a certain extent.

Firstly, the two corpora are proofread to remove spelling errors, garbled words, redundant information, extra spaces, etc, and then are tagged for part-of-speech by using CLAWS4, a part-of-speech tagger developed by Lancaster University. The two corpora are retrieved and statistically analyzed by using the corpus tool AntConc 3.5.9. The lexical-level comparison includes four aspects: lexical richness, lexical density, average word length and high-frequency words. The syntactic-level comparison includes two aspects: average sentence length and connective structures.

4. ANALYSIS AND DISCUSSION

4.1 Lexical Richness

Lexical richness refers to the number of different words in a corpus of the same length, which can be measured by the type-token ratio (TTR) of the corpus (Hu 215). The type-token ratio is the ratio of types to tokens in a certain corpus and is directly proportional to the richness and diversity of the writers' vocabulary used in their writings. The standardized type-token ratio (STTR) is a standardized value of TTR, which can be a more reliable and accurate measure of comparing lexical richness in corpora of different sizes. STTR is positively correlated with

lexical richness in given corpora, i.e. the bigger the STTR, the greater the lexical richness.

Table 1: Tokens, types, TTR, STTR of CCA and COEN

	CCA	COEN
Tokens	50517	507067
Types	5673	29158
TTR (%)	11.23	5.75
STTR (%)	42.85	47.14

Table 1 shows that the capacity of COEN is much larger than that of CCA, thus it is more wise and scientific to compare the values of STTR at this time. The STTR of CCA (42.85) is smaller than that of COEN (47.14), which indicates that the lexical richness of *Call to Arms* is lower compared to original English novels. And the vocabulary used in *Call to Arms* is not as varied and abundant as that of original English novels. This result is consistent with the findings of Laviosa who points out that the lexical richness of English translated texts is lower than that of original English texts (60). Therefore, in terms of lexical richness, CCA confirms the hypothesis of translation universals and reflects the tendency towards simplification.

4.2 Lexical Density

Lexical density is the ratio of content words to the total number of words, which is a measure of the size of the amount of information in the text. And lexical density can measure the differences in the amount of information and the difficulty of different texts (Hu 216). The higher the lexical density, the higher the proportion of content words used in the text, the more information it conveys, and accordingly, the more difficult the text is to process and comprehend. According to English content words are mainly classified into four categories: verbs, nouns, adjectives and adverbs, which can express stable lexical meanings. In contrast, English function words (mainly including pronouns, conjunctions, prepositions and articles) do not signal stable lexical meanings.

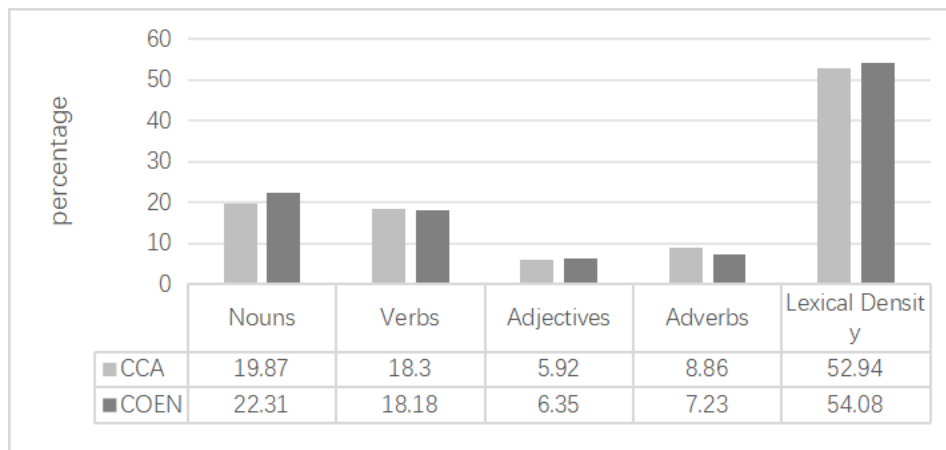


Figure 1: Percentage of content words and lexical density of CCA and COEN

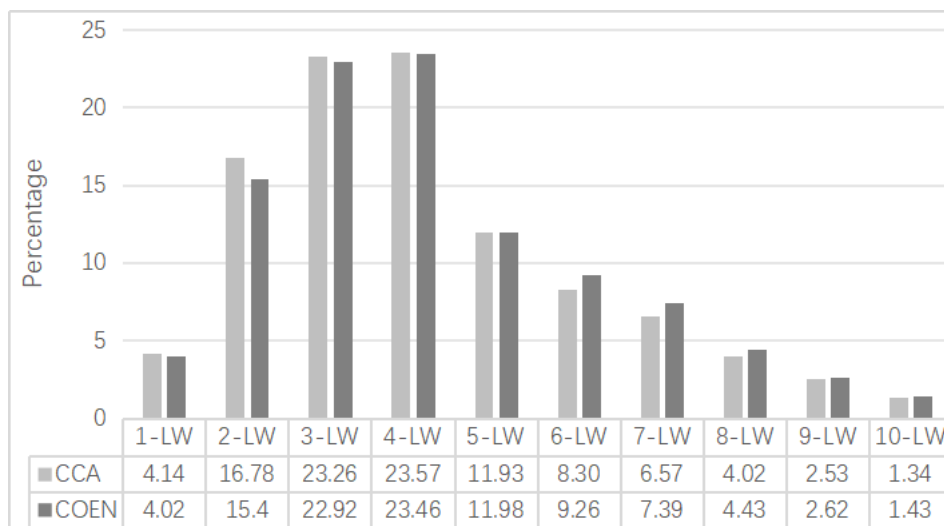
According to Figure 1, it can be concluded that the lexical density of CCA is lower than that of COEN (52.94% < 54.08%). And there is a significant difference in lexical density between them ($X^2=7.28$, $p<0.05$). This indicates that the amount of information of CCA is smaller than that of COEN and CCA is less difficult to read than COEN. Laviosa suggests that the lexical density of English translations is lower than that of original English texts (61). This result is consistent with Laviosa's finding and confirms the simplification of CCA.

4.3 Average Word Length

The average word length is considered to be one of the quantitative indicators for validating simplification. The ratio of the total number of letters to the total number of tokens in the corpus is the average word length, which reflects the complexity of vocabulary used in writing to some extent. Generally speaking, the longer the word length, the more complex the word is. The average word length of translated texts is often smaller than that of the original texts in the target language, which reflects simplification in translation. The average word length of the two corpora is obtained and the frequency distribution of the number of words with 1-10 letters are also calculated. The results are shown in Table 2 and Figure 2.

Table 2: Average word length of CCA and COEN

	CCA	COEN
Average word length	4.25	4.35
Number of words with fewer than 5 letters	32382	313130
Percentage of words with fewer than 5 letters	64.10%	61.75%

**Figure 2:** Percentage of words with 1-10 letters in CCA and COEN

As can be seen from Table 2, the average word length of CCA is smaller than that of COEN (4.25<4.35). And the proportion of words with fewer than 5 letters in CCA (64.10%) is significantly higher than that of COEN (61.75%), which suggests that words with fewer than 5 letters are used more in *Call to Arms*. Therefore it can be inferred that CCA tends to use simpler words. Moreover, according to Figure 2, words consisting of 2-5 letters are in a dominant position in both corpora. The proportions of words with 1, 2, 3, and 4 letters in CCA are all greater than in COEN. However, as for longer words of 5 letters and above, the percentages of them in CCA are all smaller than those of COEN, which means longer words are used less in CCA. Therefore, Figure 2 also indicates a simpler vocabulary of CCA. This may be due to the fact that the translators use relatively simple words to improve the readability of their translations and make their translations easier to understand for readers in the target language. These results presented in Table 2 and Figure 2 validate the tendency toward simplification in CCA.

4.4 High-frequency Words

Table 3 shows the ranking of the top ten high-frequency words in CCA and COEN. The top five most frequently used words in both CCA and COEN are “The”, “And”, “To”, “A”, and “Of”, except that “To” and “And” are in different positions. Among the last five high-frequency words, “He” ranks sixth in CCA, higher than “He” in COEN which ranks ninth. Moreover, more “His” are used in CCA because “His” is in the ninth position on the list of top ten high-frequency words in CCA, while in COEN “His” is not in the top ten high-frequency words list.

Table 3: Top ten high-frequency words in CCA and COEN

	CCA		COEN	
1	The	2532	The	28707
2	To	1366	And	13517
3	And	1220	To	12772
4	A	1184	A	12659
5	Of	997	Of	11030
6	He	953	In	7615
7	Was	850	Was	7353
8	In	772	I	7137
9	His	660	He	6797
10	Had	623	It	5754

In order to further examine the reasons for the high-frequency use of “He” and “His”, this study searches the sentences containing these two words in the CCA by using “He” and “His” as search terms respectively. Then

some example sentences are selected and the corresponding Chinese sentences in the source text are also extracted. Examples are as follows.

Example 1:

ST: 那时偶或来谈的是一个老朋友金心异，将手提的大皮夹放在破桌上，脱下长衫，对面坐下了，因为怕狗，似乎心房还在怦怦的跳动。

TT: The only visitor to drop in occasionally for a talk was my old friend Jin Xinyi. Having put his big portfolio on the rickety table he would take off his long gown and sit down opposite me, looking as if his heart was still beating fast because he was afraid of dogs.

Example 2:

ST: 早上小心出门，赵贵翁的眼色便怪：似乎怕我，似乎想害我。

TT: This morning when I went out cautiously, Mr. Zhao had a strange look in his eyes, as if he were afraid of me, as if he wanted to murder me.

The words “他” and “他的” can not be found in these two examples. Instead, the translation reveals more third person pronouns “He” and “his”. In Example 1, a series of descriptions are directed at an old friend of the main character, so “he” is used to specify the agent of the action and “his” is used to specify the belongings of things. In the second example, translators use the collocation “in his eyes” and add two subjects “He” to make the meaning of the translations clearer. The high frequency of the use of “he” and “his” confirms the explicitation of CCA. The aim is to better convey the meaning of the source text and make the logic clearer, thus improving the coherence of the discourse and making it more smoother for the target language readers to read and understand.

4.5 Average Sentence Length

Laviosa proposes a hypothesis regarding simplification that in the English comparable corpus the average sentence length would be lower in translated texts than in non-translated ones (62). The average sentence length correlates with the simplification of translated texts. Therefore this part examines the simplification of CCA according to the comparison of average sentence length between CCA and COEN. The average sentence length of CCA and COEN are presented in Table 4.

Table 4: Average sentence length of CCA and COEN

	CCA	COEN
Number of sentences	3073	37840
Average sentence length	15.92	13.13
Sentence length standard deviation	11.49	10.42

Table 4 shows that the average sentence length of CCA is longer than COEN. This implies that CCA are more inclined to use longer sentences, which is inconsistent with Laviosa’s findings. By comparing the source text and the translated text, this study holds that the possible reason for the longer average sentence length of CCA is the influence of the source text. This study finds out that the source text makes extensive use of semicolons to connect two parts, which inevitably increases the sentence length of the source text. When translators translate such sentences, they also use a semicolon in their translations. In this case, the translators focus on being faithful to the content and form of the source text. They try to maximize the correspondence between the translated text and the source text so as to reconstruct the original context in their translations. Alternatively, translators translate a sentence with two parts joined by a semicolon into a complete and coherent sentence. This is more conducive to the reading and understanding of the target language readers.

4.6 Connective Structures

Translated texts use connective structures more frequently than non-translated texts, which is related to the explicitation of translated texts. Blum-Kulka proposes the explicitation hypothesis that the translation process makes the translated text more lengthy than the source text, and the explicitation may be characterized by the increased use of cohesive devices in the translated text (19). Therefore, This section explores the extent of syntactic explicitness of CCA by focusing on the use of connective structures in both corpora. Since there are numerous and complex connective structures in English, this study refers to Huang’s approach to research on

connective structures and classifies them into the following three types: causal, adversative and conditional (74-77). The results are shown in Figure 3.

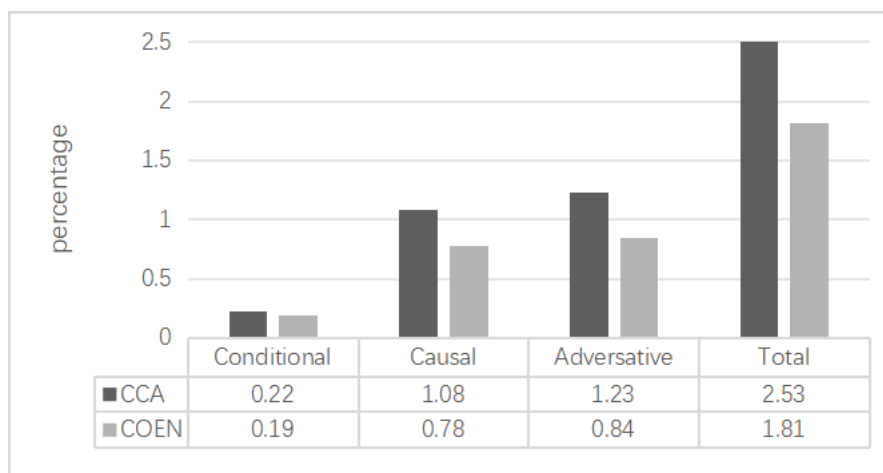


Figure 3: Percentage of connective structures

The results show that the use of all three types of connective structures in CCA is greater than that of COEN. The total use of the connective structures in CCA is significantly higher than COEN ($X^2=126.93$, $p<0.05$), which is consistent with the explicitation hypothesis. This is due to the linguistic differences between English and Chinese. English is a hypotactic language, while Chinese is a paratactic language. The hypotaxis refers to the fact that words or clauses in a sentence are connected with each other by means of linguistic forms to express grammatical meanings and logical relations. The parataxis means that words or clauses are not connected with each other by means of linguistic forms, and the grammatical meanings and logical relations in a sentence are expressed through the meanings of the words or clauses (Lian 48). Therefore, English focuses on explicit connections while Chinese emphasizes implicit coherence. The connective structures and forms in English are not only large in number and variety, but also used very frequently. In Chinese, they are used much less frequently. Therefore, on one hand, the higher proportion of connective structures in CCA is used to convey the meanings more explicitly and better clarify the logical relations of the source text, which reveals the tendency towards explicitation in CCA. On the other hand, the translators take the initiative to add to the use of connective structures in order to fit the linguistic characteristics of the target language. In this regard, the overuse of connective structures also validates the normalization of CCA.

5. CONCLUSION

This study compares *Call to Arms* and the original English novels at the lexical and syntactic levels so as to explore the translation universals embodied in *Call to Arms*. In terms of lexical richness, lexical density, and average word length, *Call to Arms* exhibits the tendency towards simplification. Regarding high-frequency words, the explicitation of *Call to Arms* is confirmed. However, as for average sentence length, *Call to Arms* does not exhibit the simplification feature. Concerning connective structures, the explicitation and normalization of *Call to Arms* are validated. This study also examines the impact of the original text and the differences between English and Chinese on the translation universals in *Call to Arms*. This study aims to provide empirical evidence and valuable reference for corpus-based research on translation universals. This study compares the linguistic characteristics of translated English and non-translated English at both lexical and syntactic levels by constructing monolingual comparable corpora. Future research can be further explored by combining bilingual parallel corpora with comparable corpora, thereby enabling a more comprehensive investigation of translation universals.

REFERENCES

- [1] Baker, Mona. "Corpus Linguistics and Translation Studies: Implications and Applications." *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker et al., John Benjamins Publishing Company, 1993, pp. 233-250.
- [2] Baker, Mona. "Corpus-based Translation Studies: The Challenges That Lie Ahead." *Terminology, LSP and Translation*, edited by Harold Somers, John Benjamins Publishing Company, 1996, pp. 175-186.

- [3] Blum-Kulka, Shoshana. "Shifts of Cohesion and Coherence in Translation." *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, edited by Juliane House and Shoshana Blum-Kulka, Gunter Narr, 1986, pp. 17-35.
- [4] Chesterman, Andrew. "Hypothesis about Translation Universals." *Claims, Changes and Challenges in Translation Studies*, edited by Gyde Hansen et al., John Benjamins, 2004, pp. 1-13.
- [5] Feng, Haoda, et al. "Form and Meaning in Collocations: A Corpus-Driven Study on Translation Universals in Chinese-to-English Business Translation." *Perspectives*, vol. 26, no. 5, 2018, pp. 677-690.
- [6] Hu, Xianyao. "A Corpus-Based Study on the Lexical Features of Chinese Translated Fiction". *Foreign Language Teaching and Research*, no. 3, 2007, pp. 214-220+241.
- [7] Huang, Libo. *A Study of Translation Universals Based on Chinese-English/English-Chinese Parallel Corpora*. Fudan University Press, 2007.
- [8] Ke, Fei. "Explicitation and Implicitness in Translation". *Foreign Language Teaching and Research*, no. 4, 2005, pp. 303-307.
- [9] Laviosa, Sara. *Corpus-based Translation Studies: Theory, Findings, Applications*. Rodopi, 2002.
- [10] Lian, Shuneng. *Contrastive Studies of English and Chinese*. Higher Education Press, 1993.
- [11] Luo, Jinru, and Dechao Li. "Universals in Machine Translation?: A Corpus-Based Study of Chinese-English Translations by WeChat Translate." *International Journal of Corpus Linguistics*, vol. 27, no. 1, 2022, pp. 31-58.
- [12] Wang, Kefei, and Libo Huang. "Terms in Corpus-based Translation Studies". *Journal of Sichuan International Studies University*, no. 6, 2007, pp. 101-105.
- [13] Wang, Kefei. *Exploring Corpus-based Translation Studies*. Shanghai Jiao Tong University Press, 2012.