

This is the accepted version of the publication Wu, B., Cheung, A. K., & Xing, J. (2021). Learning Chinese political formulaic phraseology from a self-built bilingual United Nations Security Council corpus: A pilot study. *Babel*, 67(4), 500-521. The Version of Record is available online at: <https://doi.org/10.1075/babel.00233.wu>

## **Learning Chinese Political Formulaic Phraseology from a Self-Built Bilingual**

### **United Nations Security Council Corpus: a Pilot Study**

Wu Baimei Hong Kong Polytechnic University

Andrew K.F. Cheung Hong Kong Polytechnic University

Xing Jie Guangdong University of Foreign Studies

This is a pilot study investigating formulaic phraseology most frequently used in highly formulaic political documents by examining a self-built bilingual parallel corpus of 43 speeches delivered in United Nations Security Council (UNSC) meetings by Chinese representatives. This study also probes into corpus-based approaches to explore formulaic phraseology and demonstrates a method of retrieving Chinese formulaic phraseology from the UNSC corpus. Formulaic phraseology is often seen in political discourse and can be defined as a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated—that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. This study begins with a literature review of formulaic phraseology, including features of formulaic phraseology and significance of formulaic phraseology to interpreting, and then exhibits a four-step retrieval process with the Sketch Engine software program to acquire Chinese formulaic phraseology from the corpus to fill in the gap of previous studies. Key functional units of Sketch Engine, including Wordlist, N-grams, and Concordance, are used to extract formulaic

phraseology from the UNSC corpus. Methodological issues involved in identifying formulaic phraseology, such as the length of phraseology, quantitative criteria (frequency and dispersion thresholds) are also discussed in the study. Three types of formulaic phraseology were identified: 1. to greet representatives and other members and express appreciation; 2. to express concerns about the topic of the meeting; and 3. to express China's viewpoints on the topic of the meeting. It may be helpful for translators and/or interpreters to study this categorization of formulaic phraseology and apply it in practice. Overall, this research may inform the training of Chinese into English simultaneous interpreting and empirical study is expected to verify the correlation between learning formulaic phraseology and interpreting performance.

Keywords: formulaic phraseology, corpus, United Nations Security Council, simultaneous interpreting

## **1. Introduction**

Political discourse is understood to be formulaic, institutionalized, and authoritative. Formulaic language is widely used in natural language, and whole chunks can be retrieved from memory at the time of use (Altenberg, 1998; Pawley & Syder, 1983; Sinclair, 1991; Wray, 2002). Studies of language patterns in the corpora of international organizations have suggested that the texts and speeches produced by these organizations show intrinsic formulaic patterns (Aston, 2018; Biel, Koźbiał, & Wasilewska, 2019; Jablonkai, 2010).

Formulaic expressions identified in parliamentary corpora have been incorporated into English language teaching materials, suggesting that the use of formulaic phraseology may support English language learning. Jablonkai (2010) analyzed the structures and functions of recurrent English word sequences in a European Union (EU) corpus and explored their pedagogical implications for language courses on English for EU purposes. Aston (2018) demonstrated a methodology for acquiring English formulae from a corpus of European Parliament proceedings, seeking to provide second language learners with a learning model based on formulaic phraseology.

Being familiar with formulaic patterns may facilitate the simultaneous interpretation (SI) of political speeches. This is particularly important for multilingual international organizations such as the EU and the United Nations (UN), whose daily work is largely dependent on the accurate and timely SI of speeches. Learning formulaic phraseology from existing corpora may ease the burden on professional interpreters whose organizations require them to demonstrate language mastery and avoid linguistic interference (de Laet, 2012).

The use of formulaic phraseology may be correlated with better interpreting performance. Eyckmans (2007) found that fluency in sight translation by trainee interpreters was related to the use of formulaic phraseology in the target language. Plevoets and Defrancq (2018) used a European Parliament corpus to investigate the relationship between cognitive load in SI and the linguistic features of the source language. They reported that formulaic phraseology could ease the cognitive load associated with comprehension of the source language and production in the target

language. However, there have been few studies of the use of formulaic phraseology by Chinese UN speakers, although formulaic phraseology is one of the linguistic features of the UN and other international organizations, and Chinese is one of the UN's six official languages.

To fill this gap in the literature, this study investigated the Chinese formulaic phraseology most frequently used in a self-built UN Security Council (UNSC) corpus and demonstrated a method of retrieving this phraseology from the corpus. Section 2, reviews previous studies of formulaic phraseology, including the methodologies used to acquire formulaic phraseology from corpora. Section 3 demonstrates the procedures used to identify Chinese formulaic phraseology in the abovementioned UNSC corpus. Finally, Section 4 discusses the findings and concludes the study.

## **2. Chinese language and the UN**

Few studies have explored either the linguistic features of the Chinese used at the UN in general or the formulae used by Chinese UN speakers in particular. Chinese researchers affiliated with institutes in China may shy away from studying the use of Chinese at the UN because representing China is a politically sensitive process. Interpreting scholars may also be reluctant to study the interpretation of Chinese UN speeches, because both the statements and the corresponding SI renditions are likely to be based on prepared written texts. In addition, research on the use of Chinese at the UN may have been limited by the small number of statements delivered in Chinese at UN meetings.

The use of the Chinese language at the UN is inevitably a sensitive topic, for historical reasons. The Nationalist government of the Republic of China (ROC), which fled from mainland China to Taiwan in 1949, represented China at the UN from 1945 to 1971. During this period, Chinese diplomats at the UN used mostly English and/or French (Tao 2001, Yao & Deng 2019). Therefore, translation and interpreting into Chinese at the UN occurred only infrequently before 1971, when Resolution 2758 was passed to restore all the lawful rights of the Communist government of the People's Republic of China as the legitimate Chinese representative at the UN and its affiliated agencies. Although China's opposition prevented Taiwan from re-entering the UN, the Taiwanese government was allowed to send its health minister to attend the World Health Assembly, a specialized UN agency, as an observer in 2016. However, the Taiwanese speaker used English, instead of Chinese, to address the Assembly. This was probably a counter-measure intended to differentiate the Taiwan region from China, as the UN acknowledges it an integral part of China's territory.

The Chinese language is underutilized at the UN. The UN provides translation and interpretation from and to Chinese, as one of its six official languages. However, China is the only member state to use Chinese as an official language (Kawashima, 2019). Most Chinese diplomats at international organizations speak English well and may not require the use of SI when the source language is English (Cheung, 2019). As the Chinese language has only limited circulation among international elites, such as senior diplomats, researching and bringing a spotlight on the use of Chinese language at the UN may lead to its status as a UN official language being questioned (Tonkin, 2011).

However, although the use of Chinese at the UN may be a sensitive and controversial issue, investigating this issue may contribute to the debate of language policy in general and language diversity at international organization in particular.

Many oral statements in Chinese and their SI renditions are based on written scripts. Chinese diplomats who are fluent in English may be sympathetic to the challenges of interpreting from Chinese to English due to the differences between the two languages. Therefore, most Chinese speakers at the UN give the Chinese interpreters' booth written statements in advance, allowing them to prepare (Liu, 1988). If they have received written statements well in advance, most Chinese interpreters will translate them into English to be read out as the SI renditions. As a result, the English SI rendition of a Chinese statement may be more akin to reading aloud a written translation of a prepared statement than to the SI of an impromptu speech. Nevertheless, investigating SI renditions based on prepared translations can broaden our understanding of a practice that is common at international organizations.

As China is the only UN member state to use Chinese, the Chinese booth usually interprets into, rather than from, Chinese. The Chinese booth only interprets into English, or more rarely into French, when a speaker uses Chinese. As a result, although there may be ample data on Chinese as a target language, data for Chinese as a source language may be lacking. However, using data from the Security Council (SC) may help overcome this problem. The SC meets almost daily at the UN headquarters. Delegates representing China, which is one of the five permanent members of the SC, also attend and make statements at these regular meetings. SI is normally provided in

all six official languages at the daily meetings of the SC at the UN headquarters. Video recordings of these meetings in the floor language and English are available on the UN website.

### **3. Previous Related Studies**

#### **3.1 Formulaic phraseology in previous studies**

Formulaic phraseology is a key element of language use and learning, mainly due to its pragmatic value. Formulaic language is prevalent in both written discourse and spoken discourse. However, the diversity of formulaic language makes it very tricky to define and categorize (Schmitt, 2010).

Wray (2002) listed more than 50 terms used in the literature to describe formulae, such as “lexical(ized) bundles,” “multiword items,” “formulae,” “formulaic language,” and “formulaic phraseology.” The first two terms emphasize the micro level, the lexicalized features of formulaic language, and the last three terms focus on the macro level, the fixed and recurrent characteristics of formulaic language in the context of a passage as a whole. This study adopts the term “formulaic phraseology” because the word “formulaic” conveys the sense of recurrence and fixedness, and the word “phraseology” describes the way in which language is used or expressed.

Although researchers have generally agreed that formulaic phraseology is widely used in natural language and can be holistically recalled or produced from memory, they have not reached a consensus on a specific and definite characterization of formulaic phraseology. Pawley and Syder (1983) defined formulaic language based on its linguistic pattern, namely “lexicalized sentence stems,” which are “units of clause

length or longer whose grammatical form and lexical content is wholly or largely fixed” (p. 191). Similarly, Nattinger and DeCarrico (1992) focused on the lexicalized features of a formulaic sequence, which they referred to as a “lexical phrase,” defined as something between “lexicon” and “syntax.” Both of the latter studies defined formulaic expressions based on their forms, but neither distinguished formulaic expressions from other word strings.

Moon (1997) then proposed three criteria, institutionalization, fixedness, and non-compositionality, to distinguish formulae from other multiword phrases. Institutionalization was defined as the conventionalized degree of a multiword unit in the language. Fixedness was used to examine whether a multiword unit was frozen as a sequence of words. Non-compositionality indicated that a multiword unit should be treated as having a fixed and unitary meaning instead of being interpreted word by word.

However, many studies have failed to observe that not all formulaic expressions are continuous; many are structurally incomplete. Biber et al. (1999) claimed that only 5% of the lexical bundles in academic prose and 15% of the lexical bundles in conversation can be regarded as complete structural units. Formulaic phraseology can be defined as “a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated—that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray & Perkins, 2000: 1). Therefore, many formulaic phrases are discontinuous, consisting of fixed elements and semantically constrained slots; for



example, "... make it clear that ..." can be modified adjectivally into "... make it entirely clear that ..." or "... make it abundantly clear that..."

This study adopts Wray and Perkins's (2000) definition of formulaic phraseology because it goes beyond the lexical level to provide a comprehensive and detailed description of formulaic phraseology that highlights the two most prominent features of formulaic phraseology, namely its prefabricated nature and ease of retrieval.

### **3.2 Corpus-based research on formulaic phraseology**

This section reviews previous studies that have used corpus-based approaches to explore formulaic phraseology, focusing on the methodological issues involved in identifying formulaic phraseology, such as the length of lexical bundles, quantitative criteria (frequency and dispersion thresholds), and retrieval procedures.

Biel et al. (2019) argued that formulaic phraseology is "much more pervasive than we were able to see with the naked eye" (p. 67). The most common way of identifying formulaic language is to use corpus linguistics software to extract formulaic phraseology from a corpus based on predetermined frequency and range criteria. The software program most commonly used to extract formulaic phraseology is WordSmith Tools, which has three main functions: Concord, KeyWords, and WordList.

Biber et al. (1999) proposed a pioneering approach to identifying recurrent sequences in a corpus. They investigated lexical bundles in the Longman Spoken and Written English Corpus based mainly on the bundles' frequency and range; that is, a bundle was formulaic if it occurred at least 10 times per million words (pmw) and in at

least five texts. The Biberian method was then modified by a number of studies to fit different research purposes.

Regarding the length of formulaic phraseology, most studies have examined three-word and four-word sequences, because they are much more frequent than lengths. However, some studies have focused on four-word sequences, because four-word expressions carry more distinct discourse functions than three-word expressions and contain many three-word expressions (Cortes, 2004; Hyland, 2008).

In terms of quantitative criteria, the frequency and dispersion thresholds used in previous studies have varied. The frequency cut-offs have ranged from 10 occurrences pmw (e.g., Biber et al., 1999) to 50 occurrences pmw (e.g., Breeze, 2013), with 40 occurrences pmw as the conventional threshold. This variation can be attributed to the use of differently sized corpora. When a smaller corpus is used, a higher frequency cut-off is desirable. Dispersion or range criteria may also be set to ensure the representativeness of the list; that is, to exclude the possibility that a certain formula is used in only one or a few of the speeches in a corpus. It has generally been agreed that a sequence should appear in at least three to five texts in a corpus (e.g., Biber & Barbieri, 2007) to exclude the idiosyncrasies of individual writers/speakers. However, Gries (2008) argued that this text count threshold may not be accurate enough, and put forward a more sensitive dispersion measure, known as deviation of proportions (DP). DP calculates the absolute difference between the expected percentage and the observed percentage of certain formulaic phraseology in each portion of the corpus, and then sums the absolute differences and divides the sum by two. The lower the result,

the more evenly distributed the formulaic phraseology is in the corpus. However, this study used the abovementioned text count threshold instead of DP because all of the texts in the corpus derived from Chinese delegates at the UNSC and were therefore homogenous.

In addition to the dispersion threshold, some studies have discussed the association threshold, such as mutual information (MI). Biber (2009) compared two metrics used to identify extracted sequences, namely the MI score and simple frequency. Biber (2009) observed that although the MI score evaluates the strength of collocations between words, it might not be an appropriate way to identify formulaic sequences, because it is only suitable for two-word collocations. Multiword sequences with high MI scores tend to be technical expressions (which Biber termed “collocations”) rather than formulaic sequences (“lexical bundles”). In fact, formulaic sequences have relatively low MI scores but high frequency scores. This suggests that the frequency-based method is the primary method for the identification of formulaic sequences.

Previous studies have used similar retrieval procedures. Wood (2015) summarized the retrieval procedures used in formulaic phraseology research as follows. First, the researcher sets frequency and range cut-offs to be used in the scan of the corpus by analysis software. This creates a list of high-frequency sequences. Once the software has yielded a list of formulaic phraseology that satisfies the predetermined frequency and range criteria, the researchers refine the list. Certain types of formulaic sequences, such as overlapping items and topic-specific terms, are excluded. Wood (2015) added that some formulaic phraseology is non-contiguous with internal fillable slots and thus

cannot be retrieved by software programs. Therefore, further exploration of the concordances of sequences on the list is necessary.

Notably, most studies have explored formulaic phraseology in languages such as English and Polish; they have rarely considered Chinese. Unlike English, Chinese is an isolating and non-inflectional language with a low morpheme to word ratio, and Chinese formulaic phraseology is generally non-contiguous with internal fillable slots. As a result, a frequency-based list can only serve as the preliminary result; further examination of concordances is necessary to find other discontinuous formulaic phraseology. In addition, as an ideogrammatic language, Chinese has no spaces as boundaries between characters and is written as an unseparated string of characters. Therefore, an additional process of segmentation is needed before Chinese formulaic phraseology can be identified from corpora by corpus software programs such as AntConc and WordSmith Tools. Segmentation is defined as “the process of segmenting text strings into word tokens, i.e. defining words (as opposed to characters) in a running text” (Xiao & Hu, 2015, p. 47). The segmentation tools used in previous studies were the Chinese Lexical Analysis System, developed by the Institute of Computing Technology, Chinese Academy of Sciences (Xiao & Hu, 2015), and SegmentAnt 1.1.0, developed as a tool to accompany AntConc (Gu, 2019). Both Xiao and Hu (2015) and Gu (2019) found that these two segmenters were able to achieve a fairly high precision rate; only minor manual correction was needed. However, another time-saving online corpus linguistic software program, Sketch Engine (Kilgarriff et al., 2004), can analyze Chinese corpora without the need for segmentation. In other words, Sketch Engine is

capable of processing original untokenized texts in Chinese. Details of Sketch Engine are presented in Section 4.2.

#### **4. Learning formulae from a UNSC corpus**

The UNSC corpus used in this study was a bilingual parallel corpus comprised of 43 selected speeches delivered by Chinese representatives at 60 UNSC meetings (from the 8,440<sup>th</sup> to the 8,499<sup>th</sup>) and their English interpretations by UN interpreters. All of the transcripts of the Chinese speeches and their English interpretations were downloaded from the UN's official document system (<https://documents.un.org>). Generally, the interpreting modality of these speeches was SI with text. To create the UNSC corpus, the researcher first collected 60 UNSC meeting records in Chinese and English from January 3, 2019 to April 1, 2019, and then deleted 17 that did not include speeches delivered by Chinese representatives. The resulting 43 Chinese speeches were matched with their English interpretations. The word counts of the Chinese and the English texts were 35,230 and 25,548, respectively.

The rest of this section discusses the features of the selected UNSC speeches, the significance of learning formulaic phraseology from the UNSC corpus, and the process of retrieving formulaic phraseology from the corpus.

##### **4.1 Features of the UNSC speeches**

The speeches in this corpus had three distinctive features. First, they were homogeneous in terms of topic and structure, as they all centered on security and political conditions in Middle Eastern and African countries such as Syria and Mali. The speeches consisted of three parts: greeting the presented members and expressing appreciation for the work

done by relevant parties; briefly analyzing the current situation in the region under discussion; and offering recommendations for follow-up work or stating China's position on a certain issue. Accordingly, the speeches in the UNSC corpus were highly formulaic, institutionalized, and authoritative, and were thus considered ideal materials for the study of formulaic phraseology.

Second, the speeches by Chinese representatives had been written to be read out rather than delivered extemporaneously. Therefore, they tended to be examples of written rather than oral discourse. Written discourse may impose an additional processing load on interpreters. More specifically, written discourse delivered in a formal context or on a formal occasion is likely to be carefully crafted, with greater information density, more low-frequency words, more subordinate clauses, and less linguistic redundancy than daily speech (Russo, Bendazzoli, & Sandrelli, 2006; Seeber, 2017). Thus, more of interpreters' cognitive capacity may be taken up by the need to interpret a low-frequency lexicon, adjust word order, and render intensive information (i.e., function words) simultaneously. To minimize the additional process load, interpreters should be familiar with the style of the text as well as formulaic expressions in the text. Moreover, as UNSC speeches focus on key international issues and represent the stances of countries and regions, representatives take great care to select each word carefully and deliver each piece of information accurately. Similarly, interpreters are required to render all of the information with precision.

Last, as all 43 speeches had been prepared to be read aloud, the average speech rate was about 190 characters per minute (cpm), faster than the ideal speech rate for

interpreting, i.e., 150-180 cpm (Li, 2010). The delivery was quite fluent, with few pauses or hesitations. As a result, the interpreters had to be responsive to keep up with the speakers. Additionally, when the interpreters interpreted from Chinese into English, the interpretations were longer than the source speeches, which can be attributed to the linguistic differences between Chinese and English. As Wang (2012) noted, Chinese is monosyllabic, and English is multisyllabic. When using a multisyllabic language to simultaneously interpret a monosyllabic language, the interpreter may be left behind by the speaker, leading to errors and omissions. Furthermore, Chinese is considered a high-context language, whereas English is a low-context language; information implied and unsaid in Chinese usually needs to be spoken aloud in an English rendition. In other words, there is “an overall tendency to spell out things” (Baker, 1996, p. 180), which requires additional time and effort. Given these features of speeches in the UNSC corpus, it is fair to say that familiarity with formulaic phraseology may help simultaneous interpreters to cope with challenges such as a high information density and rapid speech.

#### **4.2 Retrieval of formulaic phraseology from the UNSC corpus**

This study used Sketch Engine as the corpus analysis software program. Sketch Engine, launched in 2004, is a well-rounded corpus tool that enables users to create, upload, and manage their own corpora, and contains many ready-to-use corpora in different languages. According to Kilgarriff et al. (2014), Sketch Engine covers almost all languages with more than 50 million speakers and offers “high-level resources” for numerous languages, including Chinese (p. 18). As mentioned in Section 2.3, Sketch

Engine can identify Chinese characters and automatically segment uploaded Chinese texts using a language-specific tool known as a “segmenter.” This facilitates corpus analysis. Kilgarriff et al. (2014) pointed out that Sketch Engine has been adopted by various users, including lexicographers (e.g., Oxford University Press) and members of language teaching communities. It also helps translators/interpreters to identify terminology and phraseology in a certain domain.

The key functional units of Sketch Engine include Word Sketch, Thesaurus, Wordlist, N-grams, and Concordance. Word Sketch summarizes and categorizes the recurring grammatical and collocational patterns of a certain word, along with the frequency of these patterns, in several columns. The Thesaurus tool creates a thesaurus based on the collocations of words. It is presented in the form of either a list or a word cloud. Wordlist works on the token level and is used to identify frequently occurring single tokens in the corpus. The N-grams tool is designed to mine formulaic phraseology on the basis of a set n-grams length (two, three, four, five tokens, etc.) and frequency cut-off from the corpus. The frequency list generated by the N-grams tool serves as a useful starting point for further analysis of statistically significant formulaic phraseology. Concordance is the fundamental tool for corpus analysis in Sketch Engine, as it shows how a word is used in the texts. The Concordance tool displays results in two formats, namely the Key Word In Context (KWIC) format and sentence format. Concordance lines in the KWIC format display the search characters in the middle, with a number of tokens on each side, whereas the concordance lines in the sentence format exhibit a set of complete sentences in which the search item occurs. Notably, Sketch





Word	Count	Word	Count	Word	Count
1 国际社会	114	18 应继续	23	35 人道局势	19
2 各方	102	19 领土完整	22	36 解决方案	18
3 社会应	57	20 和领土完整	22	37 人道主义	18
4 国际社会应	57	21 和领土	22	38 通过对话	17
5 和平与	45	22 和人民	22	39 若开	17
6 的通报	38	23 稳定与	21	40 继续为	17
7 政治解决	34	24 特别代表	21	41 气候变化	17
8 能力建设	32	25 有关各方	21	42 恐怖融资	17
9 与发展	32	26 有关各	21	43 将继续	17
10 和平稳定	31	27 政治进程	21	44 各国	17
11 地区国家	29	28 地区和平	21	45 发展中国家	17
12 问题的	28	29 中方支持	21	46 为维护	17
13 的努力	26	30 与安全	21	47 中方愿	17
14 恐怖主义	26	31 所作	20	48 若开邦	16
15 南苏丹	26	32 我感谢	20	49 苏丹政府	16
16 非洲国家	24	33 建设性作用	20	50 联合国宪章	16
17 和平进程	24	34 和平与安全	20		

**Figure 2.** List of n-grams of two to four consecutive tokens occurring seven or more times in the UNSC corpus (sorted by frequency)

Word	Count	Word	Count	Word	Count
1 一个	8	18 中方将	15	35 了重要	8
2 一带	14	19 中方将继续	11	36 人员和	8
3 一带一	14	20 中方希望	7	37 人道主义	18
4 一带一路	14	21 中方感谢	9	38 人道主义援助	8
5 一路	14	22 中方愿	17	39 人道局势	19
6 一贯支持	8	23 中方支持	21	40 人道援助	15
7 上的	7	24 中方欢迎	8	41 人道救援	12
8 与发展	32	25 中方注意	8	42 今天的	7
9 与合作	7	26 中方注意到	7	43 仍面临	7
10 与安全	21	27 中方赞赏	12	44 代表的	9
11 与稳定	7	28 中方高度	10	45 作的	14
12 两国	11	29 为推动	8	46 作的努力	8
13 个国家	7	30 为维护	17	47 保持团结	7
14 中方一贯	9	31 主义援助	8	48 共同努力	10
15 中方对	9	32 主权和	7	49 共同应对	8
16 中方对此	13	33 也门各	16	50 出兵国	8
17 中方对此表示	9	34 也门各方	13		

**Figure 3.** List of n-grams of two to four consecutive tokens occurring seven or more times in the UNSC corpus (sorted by word)

Some of the n-grams in the list were clearly topic-related terms (e.g., “南苏丹”; “特别代表”; “若开邦问题”; “人道主义”), which do not conform to the definition of formulaic phraseology. It was also clear from the list sorted by word that some of the n-grams overlapped (e.g., “中方注意” and “中方注意到”). Hence, the next step was to eliminate topic-specific terms and overlapping n-grams. After the preliminary screening of the n-grams, some formulaic phraseology could be directly identified from the list, such as “发挥 建设性 作用,” which occurred 14 times in the corpus. Closer examination of the filtered n-grams list revealed that “重要 作用,” an expression similar to “发挥 建设性 作用,” occurred seven times in the corpus. Given this similarity, the researcher checked the concordance lines for “重要 作用” and found that “重要 作用” was always preceded by the verb “发挥.” Therefore, this study used the filter context function to explore the patterning of “发挥...作用” by adding “发挥” as the context filter for “作用” within the seven tokens to the left. There were 46

occurrences of “发挥...作用,” as shown in Figure 4. Most of the fillable slots between “发挥” and “作用” were adjectives, echoing Schmitt’s (2005) finding, discussed above.

维护叙以双方停火和地区安全稳定发挥的建设性作用，支持联脱部队在条件允许情况下全面重返任务区  
在人道主义援助领域的合作，愿继续发挥建设性作用。</s><s>  
家经济共同体及区域和次区域组织继续发挥斡旋作用，帮助各方通过和平方式化解纠纷。</s><s>联合国  
宗旨和原则，充分发挥联合国及其安理会的中心作用。</s><s>国际反恐斗争包括预防与打击恐怖融资，  
里安全能力建设、协助马里和平进程发挥了重要作用，做了大量工作。</s><s>希望联马团继续同马里政  
治对话、民族和解与国家重建等方面发挥了重要作用，积极推动伊拉克同科威特等地区国家加强双边关  
层面综合稳定团为促进中非和平稳定发挥的重要作用，支持特别代表和联中稳定团积极履职。</s><s>中  
>中方赞赏东盟在解决若开邦问题上发挥的重要作用 和所作的积极努力。</s><s>国际社会应该鼓励缅甸  
问题的解决，为维护有关地区和平稳定发挥了重要作用。</s><s>中方对此表示欢迎。</s><s>当今世界各  
自部署以来，为维护南苏丹和平稳定发挥了重要作用。</s><s>国际社会应继续支持联合国驻南苏丹特派  
里安全能力建设、协助马里和平进程发挥了重要作用。</s><s>联马团继续履职，对马里及萨赫勒地区  
南部非洲发展共同体等区域组织继续发挥斡旋主导作用。</s><s>联刚稳定团继续根据刚果(金)政府的需  
国际反恐主体责任，充分发挥联合国及安理会的主导作用，全面落实安理会相关决议及《联合国全球反恐战  
方案。</s><s>三是继续发挥联合国的斡旋主渠道作用。</s><s>希望也门各方加强与格里菲斯特使的合作  
当前的机会窗口，坚定支持联合国发挥斡旋主渠道作用，支持秘书长特使的工作，尊重叙利亚的主权、独  
治对话和谈判。</s><s>要坚持发挥联合国主渠道作用，也门各方应加强与格里菲斯特使、罗勒斯格德主  
><s>国际社会应坚定支持联合国发挥斡旋主渠道作用，支持秘书长特使工作，推动叙利亚各方根据“叙人  
</s><s>联合国应继续加大努力，发挥斡旋主渠道作用。</s><s>安理会成员应继续保持团结，发出一致声  
的压力。</s><s>中方支持联合国发挥斡旋主渠道作用，推动叙利亚各方根据“叙人主导、叙人所有”原则  
在联合国机构主导的国际人道援助活动中发挥积极作用，在力所能及范围内逐年扩大援助规模。</s><s>  
共同努力，继续为推动也门问题的解决发挥建设性作用。</s><s>8442 吴海涛先生(中国):我感谢钱伯斯  
协调配合，共同为地区和平稳定和发展发挥建设性作用。</s><s>中方将继续支持钱伯斯特别代表和西  
实现达尔富尔地区的和平、稳定与发展发挥建设性作用。</s><s>8448 姚绍俊(中国):中方感谢萨拉梅特  
亚早日结束过渡阶段、恢复安全与稳定发挥建设性作用。</s><s>8449 马朝旭先生(中国):我感谢姆拉德  
><s>对中东有重要影响的各方也应该发挥建设性作用。</s><s>中国始终致力于推动中东和平进程，支  
发展的推动者，为实现中东和平发挥积极和建设性作用。</s><s>8450 吴海涛先生(中国):主席先生，欢  
国际社会应为推动妥善解决科索沃问题发挥建设性作用。</s><s>中方支持联科团履职。</s><s>希望联  
><s>中方将继续为政治解决乌克兰危机发挥建设性作用。</s><s>8462 祝贺亨尼斯-普拉斯哈特女士担任  
进也门政治进程、缓解人道局势发挥积极的建设性作用。</s><s>8465 吴海涛先生(中国):中方感谢秘书  
, 为维护布隆迪和地区长期稳定与发展发挥建设性作用。</s><s>中方欢迎非盟、东非共同体(EAC)等区  
续为布隆迪实现和平稳定和可持续发展发挥建设性作用。</s><s>8467 马朝旭先生(中国):首先我欢迎中  
实现达尔富尔地区的和平、稳定与发展发挥建设性作用。</s><s>8471 中方感谢联合国人道主义事务协  
国愿为促进非洲和平、稳定与发展继续发挥建设性作用。</s><s>8475 我感谢裴凯儒特使所作通报，赞  
调，为推进叙利亚问题政治解决发挥积极和建设性作用。</s><s>8476 中方高度关注委内瑞拉局势，支  
度施压，为推动缅孟双方双边解决问题发挥建设性作用。</s><s>联合国难民署和开发计划署应与缅甸政  
责为维护地区和平、安全与稳定发挥积极和建设性作用。</s><s>关于乌克兰问题，包括欧安组织在内的  
愿继续为南苏丹实现和平、稳定与发展发挥建设性作用。</s><s>8485 当前，阿富汗局势处于关键阶段，  
调，为推进叙利亚问题政治解决发挥积极和建设性作用。</s><s>8495 我感谢迪卡洛副秘书长和拉克鲁瓦

**Figure 4.** Results for “发挥” with context filter “作用” within the seven tokens to the left

However, Chinese formulaic phraseology tends to have non-contiguous patterns with fillable slots, demanding further exploration of the concordance lines. Therefore, the third step was a targeted search of the concordances of the remaining n-grams, coupled with the Word Sketch tool, to identify any formulaic phraseology patterns. There were 38 occurrences of the two-gram “的 通报” (“*de ... briefing(s)*”). As shown in Figure 5, combining the collocations and Word Sketch of “的 通报” yielded 30 out of 38 instances co-occurring with the verb “感谢” (“to thank”). The whole pattern can be summarized as follows: “A + 感谢 + B + (所作)的通报” (A often refers to “我” or “中方,” whereas B often refers to a person’s name and title), which was interpreted as *A thanks the briefing made by B*. Figure 5 also indicates that this formulaic phraseology frequently occurred at the beginning of the speeches. The meanings of some of the formulaic phraseology vary between contexts. For instance, the meaning of “中方 愿” depends on the subsequent context. In Figure 6, the concordance lines from 5 to 8 mean that China wishes to emphasize the following points. The other lines refer to China’s commitment to working with other parties, and “中方 愿” in this context should be interpreted as *China is willing/ready to*.

<S> 中方感谢制裁委主席弗罗内茨卡大使的通报。 </S>

---

<S> 我也想感谢沃伦科夫副秘书长、比林斯利主席刚才所作的通报,以及布库女士的发言。 </S>

---

<S> 我感谢洛科克副秘书长的通报。 </S>

---

<S> 中国感谢凯塔助理秘书长的通报。 </S>

---

<S> 中方感谢委员会主席加尼大使的通报。 </S>

---

<S> 中方感谢联合国人道主义事务协调厅盖拉尼主任的通报。 </S>

---

<S> 首先,我感谢迪卡洛副秘书长的通报。 </S>

---

<S> 我欢迎德里昂外长阁下主持今天的会议,也欢迎马伊加总理出席今天的会议,感谢古特雷斯秘书长所作的通报。 </S>

---

<S> 主席先生,中方感谢吉尼特特使的通报。 </S>

---

<S> 中方高度评价多米尼加倡议召开本次公开辩论会,感谢迪卡洛副秘书长、施泰纳署长、卡巴教授和盖特谢尔女士刚才所作的通报。 </S>

---

<S> 感谢凯塔助理秘书长所作的通报。 </S>

---

<S> 首先,我感谢欧盟外交与安全政策高级代表莫盖里尼女士的通报。 </S>

---

<S> 中方欢迎欧洲安全与合作组织轮值主席、斯洛伐克外长莱恰克阁下出席会议,感谢他刚才所作的全面的通报。 </S>

---

<S> 首先,我想感谢泽鲁居伊女士的通报,感谢特别代表女士为促进刚果(金)和平稳定所做的努力。 </S>

---

<S> 首先,我感谢格里菲斯特使和洛科克副秘书长所作的通报。 </S>

---

<S> 中方感谢联合国秘书长特别代表泽鲁居伊女士的通报,也听取了莫迪女士的通报。 </S>

---

<S> 我感谢姆拉德诺夫特别协调员的通报。 </S>

---

<S> 我感谢联合国秘书长缅甸问题特使比尔格纳女士所作的通报,对你所作的努力表示赞赏。 </S>

---

<S> 首先,我感谢秘书长科索沃问题特别代表塔宁先生的通报,欢迎塞尔维亚第一副总理兼外长达契奇先生出席会议并发言,我也听取了齐塔库女士的发言。 </S>

---

<S> 感谢马谢乌特别代表的通报。 </S>

---

<S> 中方感谢萨拉梅特别代表的通报,也感谢舒尔茨大使刚才作为制裁委主席所作的通报。 </S>

---

<S> 我感谢山本忠通特别代表的通报,对你所作努力表示赞赏。 </S>

---

<S> 马朝旭先生(中国):我感谢姆拉德诺夫特别协调员的通报。 </S>

---

<S> 我感谢费多托夫执行主任和阿德里克·乌孔加执秘的通报。 </S>

---

<S> 中方感谢沃伦科夫副秘书长和科宁克斯主任所作的通报,欢迎古特雷斯秘书长提交应对“伊斯兰国”威胁国际和平与安全的最新报告。 </S>

---

<S> 中方感谢联合国秘书长南苏丹问题特别代表希勒先生的通报,也听取了Jial女士的通报。 </S>

---

<S> 我感谢迪卡洛副秘书长和拉克鲁瓦副秘书长的通报。 </S>

---

<S> 中方感谢萨拉梅特别代表的通报,也感谢舒尔茨大使刚才作为制裁委主席所作的通报。 </S>

---

<S> 首先,我感谢迪卡洛副秘书长的通报。 </S>

---

<S> 我感谢钱伯斯特别代表的通报。 </S>

Figure 5. Occurrences of “感谢” with “的 通报” within 15 tokens in the UNSC corpus



作基金,支持非方加强海上执法能力建设。</s><s>中方愿 继续为支持非洲国家实现持久和平与共同繁荣而做出自己的  
 进程,乐见一个团结、稳定、繁荣的欧洲。</s><s>中方愿 继续与欧盟共同努力,通过"一带一路"倡议等合作平台,推动和  
 国充分参与,以利于特派团更好履行授权。</s><s>中方愿 继续为南苏丹实现和平、稳定和发展发挥建设性作用。</s><s>  
 点问题的政治解决,并为此做出积极贡献。</s><s>中方愿 进一步深化同委员会的合作,将继续与各方一道,为完善国际  
 隆迪政府为维护自身和平稳定所作努力。</s><s>中方愿 强调以下两点:一是应尊重布隆迪主权、独立、统一和领土  
 削弱非洲枪声"仍面临挑战。</s><s>针对当前形势,中方愿 强调以下三点。</s><s>首先,进一步加强联合国与非盟合作。  
 良好发展势头,推动各方进一步落实协议。</s><s>中方愿 强调以下三点:第一,南苏丹各方担负落实重振协议的主要责任  
 平稳,改善人道状况具备更加有利的环境。</s><s>中方愿 谈以下看法。</s><s>第一,叙利亚各方应从国家前途和人民  
 谈判达成具有广泛包容性的政治解决方案。</s><s>中方愿 同各方共同努力,继续为推动也门问题的解决发挥建设性作用  
 固了中非友好关系,深化了中非务实合作。</s><s>中方愿 同国际社会一道,继续支持非洲加强和平安全能力建设,促进  
 伦比亚各方增进互信,积极落实和平协议。</s><s>中方愿 同安理会其他成员一道,支持马谢乌特别代表的工作和哥伦比亚  
 政治解决创造良好的环境。</s><s>最后,我想重申,中方愿 同各方共同努力,继续为推进也门政治进程、缓解人道局势  
 责任逐步、有序移交给刚果(金)安全力量。</s><s>中方愿 同国际社会一道,继续为刚果(金)早日实现和平、稳定与发展  
 来为维护达区和平稳定作出了积极贡献。</s><s>中方愿 与国际社会共同努力,继续为实现达尔富尔地区的和平、稳定  
 来为维护达区和平稳定作出了积极贡献。</s><s>中方愿 与国际社会共同努力,继续为实现达尔富尔地区的和平、稳定  
 管、知识产权、智库等多领域务实合作。</s><s>中方愿 与国际社会一道,为阿富汗的安全、稳定、发展与繁荣做出贡  
 的核心,关系到该地区的长治久安与发展。</s><s>中方愿 再次强调,"两国方案"是化解巴以冲突的根本出路。</s><s>

**Figure 6.** Occurrences of “中方 愿” in the UNSC corpus (sorted by first word to the right)

While all of the instances discussed above derived from the n-grams list, certain other formulaic sequences starting with one-gram words could not be detected using the N-grams tool. Therefore, the researcher used the Wordlist tool to generate one-gram words that occurred more than 10 times in the corpus. For instance, “欢迎” appeared 30 times in the corpus and the concordance lines for “欢迎” showed that it co-occurred with “主持” or “出席” 13 times (Figure 7). Thus “欢迎 + NAME + 出席/主持” was also an example of formulaic phraseology. Interestingly, it recurrently appeared close to the formulaic phrase “A + 感谢 + B + (所作)的通报” mentioned above. The discovery of this pattern may help interpreters to make predictions about subsequent sentences.

---

<S> 中方 欢迎 卡马拉外长出席本次会议。 </S>

---

<S> 热烈 欢迎 印尼外长雷特诺女士阁下出席今天的公开辩论会,并作发言。 </S>

---

<S> 马朝旭先生(中国):主席先生,首先我 欢迎 你来纽约主持本次会议。 </S>

---

<S> 欢迎 赤几外长奥约诺阁下来纽约主持本次会议。 </S>

---

<S> 首先,我感谢秘书长科索沃问题特别代表塔宁先生的通报,欢迎 塞尔维亚第一副总理兼外长达契奇先生出席会议并发言,我也听取了齐塔库女士的发言。 </S>

---

<S> 首先我 欢迎 中非共和国外交和侨民事务部长拜波-泰蒙女士阁下出席此次安理会会议。 </S>

---

<S> 我 欢迎 委内瑞拉外交部长阿雷亚萨阁下出席今天的公开会。 </S>

---

<S> 中方 欢迎 赤道几内亚外长奥约诺主持今天的会议,感谢迪卡洛副秘书长、非盟消弭非洲枪声高级代表拉姆拉先生,以及古恩登先生的通报。 </S>

---

<S> 中方 欢迎 欧洲安全与合作组织轮值主席、斯洛伐克外长莱恰克阁下出席会议,感谢他刚才所作的全面的通报。 </S>

---

<S> 中国支持召集关于预防与打击恐怖筹资问题的公开会, 欢迎 勒德里昂外长阁下来纽约主持此次会议。 </S>

---

<S> 我 欢迎 勒德里昂外长阁下主持今天的会议,也欢迎马伊加总理出席今天的会议,感谢古特雷斯秘书长所作的通报。 </S>

---

<S> 我欢迎勒德里昂外长阁下主持今天的会议,也 欢迎 马伊加总理出席今天的会议,感谢古特雷斯秘书长所作的通报。 </S>

---

<S> 欢迎 外长阁下来纽约主持此次公开会。 </S>

---

**Figure 7.** Occurrences of “欢迎” with “出席” or “主持” within 15 tokens in the UNSC corpus

In sum, the four steps involved in the proposed method of retrieving formulaic phraseology from a corpus are as follows. 1) Use the N-grams tool to create an eligible n-gram list with set frequency and range cut-offs; 2) refine the list by filtering out overlapping items and topic-specific terms; 3) contextualize the remaining n-grams by exploring the concordances to identify formulaic phraseology that fits the definition; and 4) and go through the one-gram list retrieved by the Wordlist tool and identify formulaic phraseology using concordances. The concordance tool played an important role in the retrieval process, as it aided not only in sorting formulae and their possible variants but also in identifying the potential positions and contexts of the formulae in the speech.

## 5. Discussion and Conclusion

Using corpora to study language patterns provides systematic and nuanced insights into formulaic phraseology. This study mined formulaic phraseology from a UNSC corpus



using the Sketch Engine software program. The formulaic phraseology retrieved is listed in Appendix 1. As explained in Section 3.1, all of the speeches in the corpus were made up of three parts: greeting the present members and expressing appreciation for the work done by relevant parties; briefly analyzing the current situation in the region under discussion; and recommending further endeavors or stating China's position on a certain issue.

The formulaic phraseology extracted can be classified accordingly. The first 7 formulae were used by the representatives to greet other members and express appreciation; the next 7 expressions concerned situations in certain regions; and the last 10 formulaic expressions were associated with China's viewpoints on certain issues. It may be helpful for translators and/or interpreters to study this categorization of formulaic phraseology and apply it in practice.

To conclude, this study investigated formulaic phraseology in highly formulaic political documents by examining a self-built corpus of speeches delivered at UNSC meetings by Chinese representatives. The study proposed a way to systematically retrieve Chinese formulaic phraseology from corpora. The proposed retrieval procedures can be summarized into four steps, as follows.

- 1) Generation of an n-grams list—use the N-grams tool to create an n-grams list with set frequency and range cut-offs.
- 2) Preliminary refinement—refine the list by filtering out overlapping items and topic-specific terms.

- 3) Contextualization of the list—explore the concordances of the remaining n-grams to identify formulaic phraseology that fits the definition.
- 4) Examination of one-gram list—look through the one-gram list and identify formulaic phraseology by concordances.

One of the limitations of this study is the narrow scope of the data in the corpus. The study used speeches delivered by Chinese representatives at UNSC meetings, chiefly because the UNSC holds frequent meetings, providing a sufficient volume of data for analysis. However, as the data reflect the work of UN interpreters only, they are homogenous in this respect. Future research should examine other types of political texts using a similar approach. Experimental studies should be conducted to investigate whether learning formulaic phraseology can reduce interpreters' processing burden.

## References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: Oxford University Press.
- Aston, G. (2018). Acquiring the language of interpreters: A corpus-based approach. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making way in corpus-based interpreting studies* (pp. 83-96). Singapore: Springer.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language*

- engineering in honour of Juan C. Sager (pp. 175-186)*. Amsterdam; Philadelphia: J. Benjamins Pub.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.
- Biber, D., & Barbieri. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.
- Biel, Ł., Koźbiał, D., & Wasilewska, K. (2019). The formulaicity of translations across EU institutional genres: A corpus-driven analysis of lexical bundles in translated and non-translated language. *Translation Spaces*, 8(1), 67-92.
- Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, 18(2), 229-253.
- Cheung, A. K. (2019). The hidden curriculum revealed in study trip reflective essays. In Sawyer, D., Austermühl, F., & Raído, V. (eds.) *The Evolving Curriculum in Interpreter and Translator Education: Stakeholder perspectives and voices*, Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Cortes, Viviana. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.

- De Laet, F. (2012). Teaching and training sight translation – a multitasking activity. In (Ren, W. ed.) *Interpreting in the age of globalization*. Beijing: Foreign Language Teaching and Research Press.
- Eyckmans, J. (2007). Taking SLA research to interpreter-training: Does knowledge of phrases foster fluency? In F. Boers, J. Darquennes & R. Temmerman (Eds.), *Multilingualism and applied comparative linguistics, Volume 1: Pedagogical perspectives*. Cambridge: Cambridge Scholar Publishing, 89–105.
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International journal of corpus linguistics*, 13(4), 403-437.
- Gu, C. (2019). *Interpreters' Institutional Alignment and (Re) construction of China's Political Discourse and Image: A Corpus-Based CDA of the Premier-Meets-the-Press Conferences* (Doctoral dissertation, The University of Manchester).
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29(4), 253-267.
- Kawashima, T. (2019). English use by the heads of state at the United Nations General Assembly: Biennial survey of 1,540 speeches between 2004 and 2018. *English Today*, 1-23. doi:10.1017/S0266078419000464

- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress* (pp. 105-116). Lorient: EURALEX.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36.
- Li, C. (2010). Coping strategies for fast delivery in simultaneous interpretation. *The Journal of Specialized Translation*, 13, 19-25.
- Liu, L. Y. (1988). Conference interpretation at the United Nations. *Foreign Language and Research*. 74(2), 42-46.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 40-63). New York: Cambridge University Press.
- Nattinger, J., & De Carrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.
- Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh (m). *Interpreting*, 20(1), 1-28.

- Russo, M., Bendazzoli, C., & Sandrelli, A. (2006). Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: Extended analysis of EPIC (European Parliament Interpreting Corpus). *Forum*, 4(1), 221-255.
- Schmitt, N. (2005). Formulaic language: Fixed and varied. *Elia: Estudios De Lingüística Inglesa Aplicada*, 6, 13-39.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan
- Seeber, K. (2017). Interpreting at the European Institutions: Higher, faster, stronger= La interpretación en las instituciones europeas: más rápido, más alto, más fuerte. *CLINA*, 3(2), 73-90.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tao, D. G. (2001). Simultaneous interpreting of Chinese at the United Nations. *Zongheng*, 8, 62-64.
- Tonkin, H. (2011). Language and the United Nations: A preliminary review. *Léger*, 1(1), 2-5.
- Wang, J. H. (2012). Effects of chunk teaching strategy on accuracy of students' conference interpreting: An empirical study. *Chinese Translators Journal*, 33(2), 47-51.
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London ; New York: Bloomsbury Academic, An imprint of Bloomsbury Publishing Plc.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge [England]: Cambridge University Press.

Xiao, R., & Hu, X. (2015). *Corpus-based studies of translational Chinese in English-Chinese translation*. Berlin: Springer.

Yao, B., & Deng, X. L. (2019). Interviews with former faculty members and students of the UN Training Program for Interpreters and Translators at Beijing Foreign Studies University. *Translation Horizon*, 2019 (1), 149-164.

#### **Appendix 1. Extracted formulaic phraseology from UNSC corpus**

1	祝贺.....担任.....	congratulate sb. on its assumption of...
2	欢迎.....主持.....	welcome sb. to preside over ...
3	欢迎.....出席	welcome the presence of sb.
4	感谢.....的通报	thank sb. for his/her briefing
5	向.....致以.....	convey...to sb.
6	.....取得（积极）进展	... register (positive) progress
7	对.....所作的努力	efforts in ...
8	中方注意到.....	China noticed that...
9	对.....造成（严重）影响	seriously affect...
10	为.....提供支持	provide support for...

11	以.....为重	put...first
12	致力于.....	make commitment to...
13	为.....创造条件	create conditions for...
14	在.....基础上	on the basis of...
15	加强.....能力建设	enhance the capacity-building in...
16	中方呼吁.....	China calls on...
17	根据.....的要求	as requested by...
18	根据授权.....	...based on its mandate
19	发挥.....(xx) 作用	play a (xx) role in ...
20	A 是解决 B 的唯一途径	A is the only way to resolve B
21	根据.....原则	uphold the principle of...
22	通过 A 达成.....的解决方案	reach a...solution through A
23	恪守.....宗旨和原则	abide by the purposes and principles of...
24	中方愿同 A 一道, .....	China is ready to work with A to...

**Address for correspondence**

Ms Baimei Wu

Department of Chinese and Bilingual Studies

Hong Kong Polytechnic University

Kowloon, Hong Kong

rosebm.wu@hotmail.com

Co-author information

Andrew K.F. Cheung



Hong Kong Polytechnic University  
[andrewkfcheung@gmail.com](mailto:andrewkfcheung@gmail.com)

Xing Jie  
Guangdong University of Foreign Studies  
[xingjie.xj@gmail.com](mailto:xingjie.xj@gmail.com)