Contrasting the Chinese *bei*-passive and the English *be*-passive through dependency distance

Yonghui Xie
Faculty of Linguistic Sciences
Beijing Language and Culture University
1098751719@qq.com

Ruochen Niu✉
Faculty of Linguistic Sciences
Beijing Language and Culture University
niuruochen@126.com

Haitao Liu
Department of Linguistics
Zhejiang University
htliu@163.com

Abstract

Many previous works have employed dependency distance to measure the syntactic difficulty of various languages. However, little is known concerning the properties of dependency distance within certain constructions across languages, such as the passive. The current study intends to contrast the Chinese *bei*-passive and the English *be*-passive using different measures of dependency distance against self-built treebanks. Results show that: (1) In terms of the dependency distance pointing from the predicate verb, the *bei*-passive exhibits a longer distance compared with the *be*-passive, consistent with general properties. (2) Regarding the dependency distance pointing from the head noun, both passives show a similar distance, following general properties. (3) With respect to the dependency distance between a head noun and the predicate verb, the *bei*-passive displays a shorter distance than the *be*-passive, contrary to general properties. Moreover, these findings suggest that the general properties of dependency distance in languages like Chinese and English may fail to fully explain those in certain language constructions, such as the passive.

1 Introduction

Languages exhibit considerable variation with regard to the productivity[1] of their passives. For instance, certain languages, such as Chadic languages, have no passives at all. In contrast, some languages present passives on a limited class of transitive and ditransitive verbs, and not with intransitive verbs at all. Conversely, some languages, like Bantu languages, essentially allow all verbs to passivize (Keenan, 1985). Chinese and English are two genetically different languages, with Chinese belonging to the Sino-Tibetan family and English belonging to the Indo-European family. Despite their differences, both languages have the passive. Typologically, Mandarin Chinese is regarded to be low on the scale of productivity of the passive voice compared to English, which is known for its extensive use of the passive.

Chinese employs some devices to express the passive meaning. The most important passive marker in Chinese is 被 *bèi*, while there are also alternatively colloquial forms such as 让 *ràng*, 叫 *jiào*, 给 *gěi*, and 为…所 *wéi*…*suǒ*. However, the frequencies of the four forms are relatively low (Xiao, 2015), and the first three have not been fully grammaticalized as the passive markers due to their primary usage as lexical verbs, meaning 'allow', 'call', and 'give'. Additionally, there are certain lexical verbs that inherently convey a passive meaning, such as 挨 *ái* 'endure', 受 *shòu* 'be subjected to', and 遭 *zāo* 'meet with', as well as some notional passives with unmarked forms. Nevertheless, these two usages are not the grammatical category of the passive in a strict sense (Zhang, 1953; Tang, 2006). All in all, the

[1] Productivity is a general term in linguistics referring to the limitless ability to use language—any natural language—to say new things. (Nordquist, 2020).

Chinese passive is represented by the predominant *bei*-passive.

The categorial status of the passive mark *bei* in the *bei*-passive is controversial, and it can be grouped into three categories: the first is a preposition like the preposition 'by' in English; the second is a verb similar to the verb 'get' in English; and the third is a dual-status marker—a passivization morpheme like the English counterpart '-en', or a preposition like the 'by'. The present study argues, in line with traditional analyses (Chao, 1968; Zhu, 1982; McCawley, 1992; Tsao, 1996), that the Chinese *bei*-passives can be analogous to the *be*-passive in English, with *bei* being treated on a par with the preposition 'by' in English (i.e., the first category). Therefore, we choose *bei*-passive and *be*-passive as the primary focus in our research.

In addition to the above contrast of the syntactic status of the passive marker in Chinese and English, there are various perspectives of the syntactic comparison between the Chinese and English passive, including their syntactic structures (Chu, 1973), syntactic functions (Xiao et al., 2006; McEnery and Xiao, 2010; Xiao, 2015), and syntactic nature of passivization (Huang, 1999; Pan and Hu, 2021). However, few studies have considered the syntactic difficulty of the passive across languages.

Dependency distance (Heringer et al., 1980; Hudson, 1995; Liu, 2007) is defined as the linear distance between two syntactically related words, and it is considered as a predictor of syntactic difficulty (Liu et al., 2017). In view of parsing models of dependency grammar, dependency distance presents a means of measuring the working memory burden imposed on language processing (Hudson, 1995; Liu et al., 2017; Niu and Liu, 2022).

Previous research has utilized dependency distance to annotate and examine various languages in the world (Liu, 2008; Johannsen et al., 2015; De Marneffe et al., 2021; Yan and Liu, 2021), and provided evidence that the mean dependency distance of Chinese is longer compared to that of English (Liu, 2008). However, limited knowledge exists regarding the dependency distance property specific to certain constructions across languages, such as the passive. It is worth noting that the general properties of dependency distance in languages may not fully capture those observed in specific language constructions. In addition, previous studies have focused more on the distance of specific dependency relations (Xu and Liu, 2015; Ouying et al., 2022), and less on the in-depth investigation of all the dependencies pointing from a specific word (e.g., the dependencies pointing from the predicate verb). Therefore, a more fine-grained investigation is needed on the cross-language constructions with different measures related to dependency distance.

To address the above issues, this study aims to contrast the Chinese *bei*-passive and the English *be*-passive from a fine-grained perspective of dependency distance. Section 2 describes the data and the methodology. The statistical results are presented in Section 3 and subsequently analyzed in Section 4. In Section 5, this paper is ended with conclusions.

2 Methodology

This section introduces the data collection, dependency distance measures and methods used in the contrastive analysis between the two passives. The *bei*-passive takes the form of NP1+*bei*+(NP2)+VP, and the *be*-passive follows the structure of NP1+*be*+VP+(by+NP2).

2.1 Data

To collect the Chinese and English data for our study, ToRCH2009, 2014, 2019 Corpus[2] (2014) and Yiyan English-Chinese Parallel Corpus (Xu and Xu, 2021) were used as the main data sources, respectively. These corpora are balanced million-word corpora constructed following the model of the Brown Corpus, covering various genres. More importantly, the texts in these corpora have the relatively recent publication dates, all from 2009 onwards. To avoid the possible disturbing effects of different genres, data samples of equal sizes were randomly selected from news, fiction, and general genres for our study.

The following scripts were written to retrieve the available passive sentences from the above two open corpora. To begin with, these corpora were processed including sentence splitting, word tokenization, part-of-speech tagging, and dependency parsing using Stanford CoreNLP (Manning et al., 2014).

[2] The download website is http://corpus.bfsu.edu.cn/info/1072/1015.htm

Next, our scripts identified the candidate passive sentences by selecting those that contain a specific dependency relation tagged as "nsubjpass"[3] and governed by "ROOT"[4]. These candidate sentences were then subjected to one or multiple rounds of filtering for extracting the required sentences. Notably, the Chinese and English sentences were processed differently during this stage.

For Chinese candidate sentences, non-passive sentences were manually filtered out. Then, the annotations related to predicate verbs were also manually revised to ensure their accuracy and consistency. Regarding English candidate sentences, it was observed that some sentences use the structure "*be* (as a linking verb) + past particle" instead of "*be* (as an auxiliary verb) + past particle", as in the sentence "Her voicemail was gone". To filter these sentences, the scripts further utilized regular expression to identify specific patterns suggesting that the *be* in "*be* + past particle" is functioning as an auxiliary verb rather than a linking verb. These patterns included the presence of progressive tenses (e.g., "is being done"), future tenses (e.g., "will be done"), predicate verbs followed by "by" (e.g., "done by"), or the inclusion of temporal or other adverbials (e.g., "done yesterday"). In addition, a manual check of the part "be + past particle" was performed to ensures that true passive sentences were retained for further analysis.

In the end, 600 *bei*-passive sentences and 600 *be*-passive sentences with dependency annotations were selected. Table 1 provides the description of the two newly constructed corpora.

	Genre	Number	In total
Chinese *bei*-passive corpus	news	200	600
	general text	200	
	fiction	200	
English *be*-passive corpus	news	200	600
	general text	200	
	fiction	200	

Table 1: Description of the passive corpora.

[3] The "nsubjpass" dependency relation, also known as the "passive nominal subject", refers to a noun phrase which is the syntactic subject of a passive clause.

2.2 Dependency distance factors

Five dependency distance (DD) factors used in our study were defined and measured, focusing on the head noun in NP1 and the predicate verb in VP along with their respective dependents. These factors include: the DD of the subject relation (SUBDD), the mean DD pointing from the head noun (NMDD), the furthest DD pointing from the head noun (NFDD), the mean DD pointing from the predicate verb (VMDD), and the furthest DD pointing from the predicate verb (VFDD) (Table 2).

Factor	Code	Description
DD of subject relation	SUBDD	the dependency distance between a head noun and the predicate verb
mean DD pointing from head noun	NMDD	the mean dependency distance between a head noun and its dependents
furthest DD pointing from head noun	NFDD	the furthest dependency distance between a head noun and its dependents
mean DD pointing from predicate verb	VMDD	the mean dependency distance between a predicate verb and its dependents
furthest DD pointing from predicate verb	VFDD	the furthest dependency distance between a predicate verb and its dependents

Table 2: Five dependency distance factors.

Dependency distance was measured as the linear absolute difference between a governor and its dependent (Liu et al. 2009). For example, in the sentence "The findings were then validated in a separate group of patients" (Figure 1), the SUBDD (i.e., the DD between "findings" and "validated") is |5-2|=3. Later, the NMDD/VMDD is the mean of all dependency distances governed by a head noun/predicate verb, while the NFDD/VFDD is the maximum among all dependency distances dominated by a head noun/predicate verb. In Figure 1, there is only one dependency governed

[4] The "root" dependency relation points to the root of the sentence, and is typically represented by a finite verb.

by the head noun “findings”, thus the NMDD and NFDD are both 1. The dependency distances governed by the predicate verb “validated” can be obtained as follows: 3 2 1 4, so the VMDD is 2.5 and the VFDD is 4.

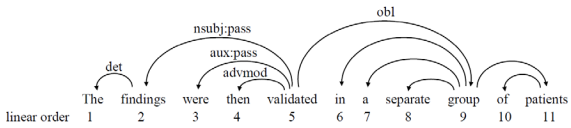

Figure 1: Dependency structure of a sample sentence.

2.3 Method

Each dependency distance factor was computed for *bei*-passives and *be*-passives in order to determine their respective patterns in use. Subsequently, to assess the importance of these five factors in distinguishing the passives of the two languages, a random forest model was employed to rank factors. The random forest model is known for its highly accuracy among current classification algorithms. It is capable of generating an out-of-bag score during the run, which serves as an evaluate metric for the model, without the need for cross-validation. Finally, significance tests were performed to evaluate the statistical significance of the above computed patterns in use and rankings of the factors.

3 Results

This section presents the statistical results for dependency distance factors.

3.1 Patterns in use of the five factors

The eigenvalues of the five factors were calculated for both 600 *bei*-passives and 600 *be*-passives, and their corresponding averages were then computed. These statistical results are presented in Figure 2.

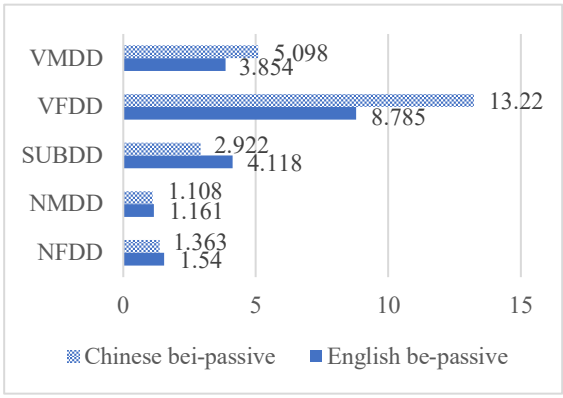

Figure 2: The averages of the five factors counted in two passives.

From Figure 2, it can be observed that the VMDD and VFDD in the *bei*-passive are longer compared with that in the *be*-passive (5.098 > 3.854, 13.22 > 8.785), indicating a pattern in use for long dependency distances between a predicate verb and its dependents in the *bei*-passive. On the other hand, the SUBDD in the *bei*-passive is shorter than that in the *be*-passive (2.922 < 4.118), so the *be*-passive tends to have a long dependency distance for the subject relation. Additionally, the NMDD and NFDD in the *bei*-passive are slightly shorter compared with the *be*-passive (1.108 < 1.161, 1.363 < 1.540). This suggests that the pattern in use for dependency distances between a head noun and its dependents are similar in both passives.

3.2 Importance ranking of the five factors

The input dataset for training the random forest classifier on the two passives consisted of the eigenvalues of the five factors calculated from each of the 1200 passive sentences, along with their corresponding language tags. The dataset was divided into a training set and a test set, with a ratio of 9:1. The results indicated an out-of-bag score of 0.6667 and a total of 21 tree nodes. As the least important feature was removed, the out-of-bag score gradually decreased. By retaining all five factors, the classification importance ranking shown in Figure 3 was obtained.

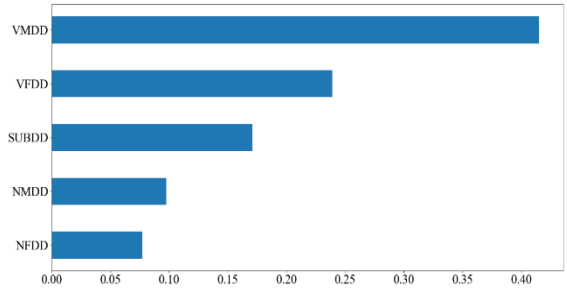

Figure 3: The ranking of classification importance.

According to Figure 3, when distinguishing the *bei*-passives and the *be*-passives, the two factors related to predicate verbs (VMDD and VFDD) have the highest contribution, followed by SUDD. On the other hand, the two factors related to head nouns (NMDD and NFDD) have a relatively smaller impact on the classification.

3.3 Statistical significance of five factors

The Kolmogorov-Smirnov test showed that the five factors were non-normally distributed in the 1200 passives. Consequently, the Mann-Whitney U test was employed to assess the factor significance. Table 3 presents the results of this test, which reveal the highly significant effects of the VMDD, VFDD, and SUBDD in differentiating *bei*-passives and *be*-passives (all $p < 0.001$). Furthermore, the NFDD also shows a little significance between the two types of passives ($p < 0.01$), while the NMDD does not have a significant effect ($p > 0.05$).

Factor	Significance
VMDD	0.000***
VFDD	0.000***
SUBDD	0.000***
NMDD	0.58
NFDD	0.006**

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 3: The results of the significance tests.

In summary, our results imply the three characteristics of syntactic difficulty in the two kin ds of passives:

- The Chinese *bei*-passive significantly shows long dependency distances pointing from predicate verbs, which plays a crucial role in distinguishing the two passives;
- The English *be*-passive shows long dependency distances of subject relations;
- Besides, the dependent distance pointing from head nouns exhibits similarities between *bei*-passives and *be*-passives.

4 Discussion

In this section, we firstly attempt to provide some explanations and interpretations for the three findings summarized at the end of Section 3.2. Then, we compare these findings with previous studies conducted on Chinses and English. The aim is to gain a deeper understanding of the common patterns and unique characteristics of the syntactic difficulty of the passives.

4.1 Long dependency distances pointing from predicate verbs in *bei*-passives

According to Ferrer-i-Cancho (2004) and Jiang and Liu (2015), mean dependency distances tend to increase as sentence length increases. Based on this, we hypothesized that sentence length is a factor that contributes to the longer VFDD and VMDD in *bei*-passives other than *be*-passives.

To verify this, the linear relationships between VFDD (or VMDD) and sentence length were plotted in both passive and be-passive sentences, as shown in Figure 4 (or Figure 5). Evidently, both the VFDD and VMDD exhibit a trend of gradual increase with the growth of the sentence length. Furthermore, we calculated separately the mean sentence length of the two passives, and found that the *bei*-passive has an average length of 22.05, which is significantly longer than that counted in the *be*-passive with 19.66 ($p < 0.01$). Therefore, we can conclude that the inherently longer sentence in *bei*-passives renders the longer dependency distance pointing from predicate verbs.

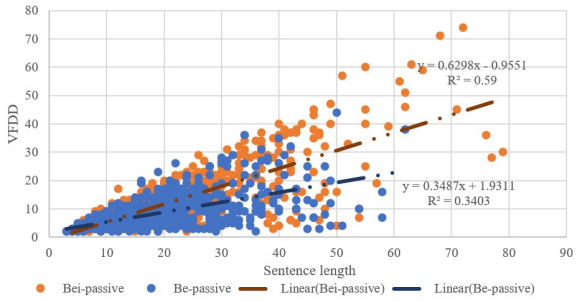

Figure 4: The relationship between VFDD and sentence length.

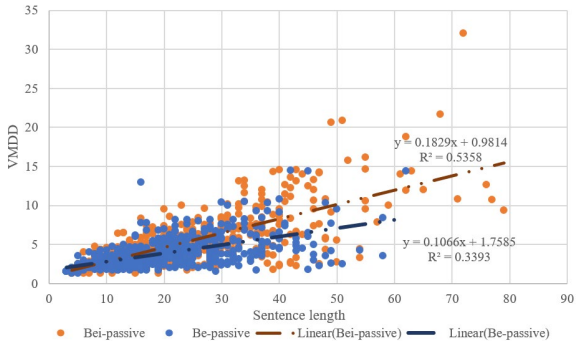

Figure 5: The relationship between VMDD and sentence length.

Next, we explored the reasons from the perspective of specific dependency relations. From the distributions of VFDD (Figure 6) and VMDD (Figure 7), it was observed that when VFDD > 25 or VMDD > 8, their percentage in *bei*-passives starts to exceed that in *be*-passives. This suggests that the longer VFDD and VMDD in *bei*-passives are mainly caused by long dependency relations. We then investigated the long dependencies in our corpus, and identified that they frequently involve conjunct relations between predicate verbs and other verbs in subordinate clauses.

The long dependency distance of this conjunct relation can be attributed to grammar regularities in languages. In English, the syntactic structure is relatively complete and it follows standardized forms without much dispersion. This is achieved through the use of formal means, such as morphological markers and connectives. On the other hand, Chinese lacks extensive morphological means, and it relies more on function words and word order to express various grammatical relations. Thus, Chinese exhibits a more flexible syntactic structure, and its form tends to be dispersed, giving rise to its unique flowing sentences (Lian, 1992; Lv, 1979). Figuratively, Chinese sentences are often likened to the bamboo, where each clause is a small bamboo section composed of a semantic-phonetic chunk, and connected in chronological order (Pan, 1997). Within dependency grammar, these bamboo sections are typically connected by their verbs. In other words, the predicate verb in the main clause serves as the root of the Chinese sentence, linking other verbs in subordinate clauses. In this sense, when there are multiple clauses, the conjunct relation becomes longer.

In summary, the second reason for the longer VFDD and VMDD in *bei*-passives is related with the longer dependency distance of conjunct relations between the predicate verb and other verbs in subordinate clauses. This characteristic is a result of the unique bamboo-like syntactic structure of Chinese.

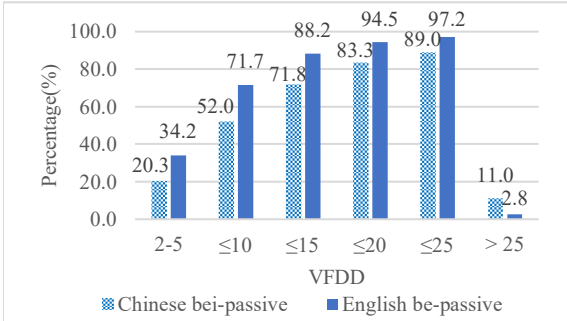

Figure 6: The distribution of VFDD.

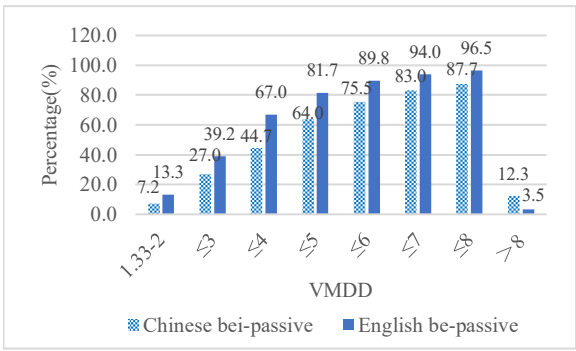

Figure 7: The distribution of VMDD.

In addition to the two explanations mentioned above, our calculations did not reveal any significant differences in the genre effect (news, general and fiction genes) or the semantic impact (adversity of predicate verb) between *bei*-passives and *be*-passives.

Previous studies (Jiang and Liu, 2015; Li, 2020) have shown that the mean dependency distance is generally longer in Chinese than in English based on the overall dependency analysis. This contrast result is consistent with our findings regarding the differences between Chinese and English passives in term of the dependency distances pointing from predicate verbs. Thus, the longer VFDD and VMDD are the common regularities observed in Chinese and Chinese special constructions.

4.2 Long dependency distances of subject relations in *be*-passives

To explain the longer SUBDD in *be*-passives rather than *bei*-passives, it is necessary to first examine the characteristics of frequently intervening words

between a head noun and the predicate verb in the two types of passives, respectively.

Figure 8 displays the distribution of SUBDD for both passives. From this figure, it is evident that in *bei*-passives, the percentage of one intervening word (SUBDD = 2) accounts for half of the distribution (55.3%), and two intervening words (SUBDD = 3) are also quite common (24.8%). Furthermore, the analysis of our corpora reveals that the one word is typically *bei*, while the two words fall into the three cases: "*bei* + the agent of a predicate verb" (e.g., *bèi tā-men* 'BEI they'), "*bei* + the premodifier of a predicate verb" (e.g., *bèi tí-qián* 'BEI in advance'), or "the premodifier of *bei* + *bei*" (e.g., *yě bèi* 'also BEI').

As for *be*-passives, there is a notable prevalence of one or two intervening words (SUBDD = 2, 3) of subject relations, accounting for 32.7% and 32.3% respectively (Figure 8). Also, the percentage of three intervening words (SUBDD = 4) also exceeds 10%. Our corpora suggest that when one intervening word is present, it typically corresponds to the simple present tense or simple past tense forms of *be* (e.g., "is, are, was, or were"). Meanwhile, two intervening words often involve the progressive, perfect or future tense forms of *be* (e.g., "be being, have been, or will be"), or the combination of *be* and the adverbs that modify the predicate verb (e.g., "be suddenly"). In addition, the instances with three intervening words feature the combination of non-simple tenses forms of *be* and verb modifiers (e.g., "will be suddenly").

By comparing the 1-2 frequently intervening words in the *bei*-passive with the 1-3 frequently intervening words in the *be*-passive, it becomes evident that the *be*-passive requires the use of different tenses as intervening words. This requirement is a result of English rich morphology and cannot be ignored (Greenbaum, 1996). On the other hand, Chinese does not have this form requirement. Thus, the presence of these abundant tense forms in the *be*-passive may be one of the reasons for its long SUBDD.

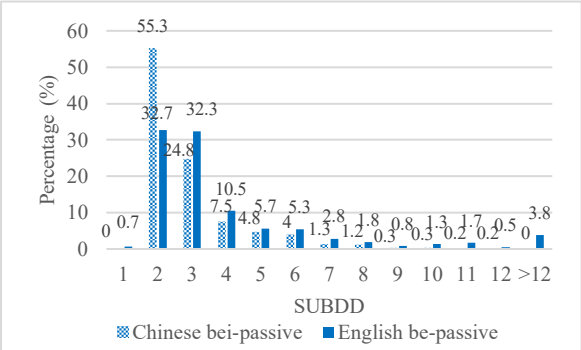

Figure 8: SUBDD distribution of *bei*-passives and be-passives.

Moreover, Figure 8 provides some additional and interesting information for contrasting *bei*-passives and *be*-passives. Firstly, both passives exhibit a long-tail distribution in terms of the SUBDD. Secondly, the percentage of SUBDD > 12 in the *be*-passives is 3.8%, while the *bei*-passives have none. This suggests that the long-tail effect is more pronounced in the *be*-passives. Thirdly, we observed an interesting occurrence of SUBDD = 1 in *be*-passives (0.7%), where there are no intervening words between head nouns and predicate verbs. This finding was unexpected. Upon analyzing our corpora, we discovered that these passives occurred in interrogative sentences, such as "How was Cookie infected?". Conversely, this phenomenon failed to be observed in the subject relation of the *bei*-passive.

Lastly, when the SUBDD is expanded to 3, the *bei*-passives and the *be*-passives cover almost 80% and 60% of the dependencies, respectively; Also, when the SUBDD is expanded to 6, both the *bei*-passives and the *be*-passives cover nearly 90% of the dependencies. This implies that there is a tendency to minimize the dependency distance within specific dependency relations of the certain construction. Notably, this tendency is not limited to different languages (Liu, 2008; Temperley, 2007; Buck-Kromann, 2006), but extends to specific dependency relations, as proved by Li (2020). Thus, dependency distance minimization holds a universal linguistic regularity.

	Chinese	**English**	***Bei*-passive**	***Be*-passive**
SUBDD	2.69	2.58	2.922	4.118

Table 4: SUBDD in different language environment.

In previous study (Li, 2020), it is Chinese that tends to have a longer SUBDD (2.69) compared with English (2.58). Whereas, our investigation reveals that the SUBDD difference is expressed in the opposite way in *bei*-passives and *be*-passives (Figure 2 and Table 3). In this sense, our finding provides additional insights to prior research.

Additionally, Table 4 shows that there is no considerable difference in SUBDD between Chinese and Chinese *bei*-passives, but the SUBDD in English *be*-passives is significantly larger than that in English. Based on our previous analysis of the intervening components between a head noun and the predicate verb within *be*-passives, it can be inferred that progressive, perfect and future tenses, as well as preverbal modifiers, occur more frequently in *be*-passives than in non-passive constructions. The increased frequency of such components may lead to longer distances of subject relations in *be*-passives. Other reasons remain to be analyzed further.

4.3 Similar dependency distances pointing from head nouns for two passives

From Figure 2, it can also be observed that the NMDD and NFDD of both *bei*-passives and *be*-passives are about 1-2, indicating that the average length of NP1s is around 2-3. Wang (2018) performed a study on the average length of noun phrases in both Chinese and English, revealing that their average length is approximately 4. This suggests that NP1s in *bei*-passives and *be*-passives tend to be short and concise. In the following analysis, we further explore the reasons behind their consistent pattern of shorter NP1s.

When it comes to the formation of the passive, there are various methods in the world's languages. However, both Chinese *bei*-passives and English *be*-passives share a similar operation: they foreground and topicalize the patient of the action (i.e., NP1 in both passives), and background or eliminate the agent of the action (i.e., NP2 in both passives) (Keenan, 1985). According to the principle of short-before-long preference (Hawkins, 1994) or the heavy noun phrase shift (Arnold, 2000), the foregrounding NP1 in both passives tends to be shorter as it takes on a more prominent position in the sentence.

5 Conclusion

By employing the fine-grained features of dependency distance, we have contrasted the syntactic difficulty of the Chinese *bei*-passives and the English *be*-passives. Our findings are as follows:

- In comparison with the *be*-passive, the *bei*-passive demonstrates a longer dependency distance pointing from the predicate distance, primarily due to the long sentences and the long conjunct relations. This finding is in line with previous research that has identified differences between Chinese and English in terms of the overall dependency distances.

- Both passives show a shorter dependency distance pointing from the head noun compared with regular noun phrases in both Chinese and English. This suggests that passives tend to use more concise subjects, aligning with the short-before-long preference.

- Regarding the dependency distance between a head noun and the predicate verb, the *be*-passive exhibits a longer distance than the *bei*-passive owing to the abundant tense forms in English. This finding is contrary to the difference observed between Chinese and English.

To conclude, our findings suggest that the general properties of dependency distance in languages like Chinese and English do not fully account for the specific characteristics observed in certain language constructions, such as the passive. These certain constructions may have their own distinct dependency features within the complex language system, characterized by emergence and uncertainty.

Moreover, dependency distance is also considered an indicator of cognitive load. Based on our study, we can speculate that the predicate verb as a governor poses a heavier cognitive load in the *bei*-passive, whereas the head noun as a governor bears a similar cognitive load in both passives. Additionally, the subject relation in the *be*-passive requires comparatively a fewer cognitive burden.

Limitations and implications

At the same time, it is important to acknowledge the limitations of this study, some of which may be addressed through future research.

First, the size of our corpora is relatively small, and subsequent studies can increase the corpus size

to further validate the reliability of the conclusions drawn from our analysis of 1200 sentences.

Besides, although we have manually proofread the important parts of the automatically annotated sentences, a comprehensive review of all the annotations in the future would enhance the rigor of research.

Also, the explanations for the three syntactic difficulty characteristics in Section 4 may not be exhaustive or in-depth enough. Further reasons could be analyzed, such as cognitive loads and lexical semantics.

Last, considering the limits of dependency distance, future studies could explore possible alternative accounts of syntactic difficulty involved in the passives.

Ethics Statement

The Chinese *bei*-passive and English *be*-passive sentences used in our study were collected from real and publicly available corpora. Our data is available to the public solely for research purposes. The annotation tool Stanford CoreNLP, which we utilized for analysis, is both open-source and reputable in the natural language processing field. Besides, the authors declare that there is no conflict of interest.

Acknowledgments

This research project is supported by Science Foundation of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (23YJ170002).

References

Jennifer E. Arnold, Losongco Anthony, Thomas Wasow and Ginstrom Ryan. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1):28–55.

Matthias Buch-Kromann. 2006. *Discontinuous grammar: A dependency-based model of human parsing and language learning*. Diss. Copenhagen Business School.

Yuen Ren Chao. 1968. *A grammar of spoken Chinese*. University of California Press, Berkeley and Los Angeles, USA.

Chauncey C. Chu. 1973. The passive construction: Chinese and English. *Journal of Chinese Linguistics*, 1:437–470.

Ramon Ferrer-i-Cancho. 2004. Euclidean distance between syntactically linked words. *Physical Review E*, 70(5):056135.

Sidney Greenbaum. 1996. *The Oxford English Grammar*. Oxford University Press, Oxford, UK.

John A Hawkins. 1994. *A performance theory of order and constituency*. Cambridge University Press, Cambridge, UK.

Hans Jürgen Heringer, Bruno Strecker, and Rainer Wimmer. 1980. *Syntax: Fragen-Lösungen-Alternativen*. Wilhelm Fink Verlag, München.

C.-T. James Huang. 1999. Chinese passives in comparative perspective. *Tsing Hua Journal of Chinese Studies*, 29:423–509.

Richard A. Hudson. 1995. Measuring syntactic difficulty. Available at http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf.

Jingyang Jiang, Haitao Liu. 2015. The effects of sentence length on dependency distance, dependency direction and the implications–Based on a parallel English–Chinese dependency treebank. *Language Sciences*, 50:93-104.

Anders Johannsen, Héctor Martínez Alonso, and Barbara Plank. 2015. Universal dependencies for danish. In *International Workshop on Treebanks and Linguistic Theories (TLT14)*, pages 157, Warsaw, Poland. Institute of Computer Science Polish Academy of Sciences ul.

Edward Keenan. 1985. Passive in the world's languages. In Tim Shopen (ed.), *Language typology and syntactic description*, Volume 1: Clause Structure, 243–281. Cambridge University Press, Cambridge, UK.

Wenwen Li. 2020. 基于依存树库的英汉语主语句法计量特征的对比研究 [A contrastive study of quantitative syntactic properties of Chinese and English subjects based on Dependency Treebank]. *Bulletin of Linguistic Studies*, 25:173-191.

Shuneng Lian. 1992. 论英汉句法的基本特征 [On the basic characteristics of English-Chinese syntax]. *Journal of Xiamen University (Arts & Social Sciences)*, (3):122-126.

Haitao Liu. 2007. Probability distribution of dependency distance. *Glottometrics*, 15:1–12.

Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159-191.

Haitao Liu, Richard Hudson, and Zhiwei Feng. 2009. Using a Chinese treebank to measure dependency distance. *Corpus Linguistics and Linguistic Theory*, 5(2):161-174.

Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of life reviews*, 21:171-193.

Shuxiang LV. 1979. *汉语语法分析问题 [Chinese grammar analysis issues]*. The Commercial Press, Beijing, China.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55-60, Baltimore, Maryland USA. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal dependencies. *Computational linguistics*, 47(2):255-308.

Noriko Akatsuka McCawley. 1992. On the treatment of Japanese passives. In *Proceedings of the Eighth Regional Meeting*, pages 256–270. Chicago Linguistic Society.

Tony McEnery and Richard Xiao. 2010. *Corpus-based contrastive studies of English and Chinese*. Routledge, London, UK.

Ruochen Niu and Haitao Liu. 2022. Effects of Syntactic Distance and Word Order on Language Processing: An Investigation Based on a Psycholinguistic Treebank of English. *Journal of Psycholinguistic Research*, 51(5):1043-1062.

Richard Nordquist. 2020. Definition and Examples of Productivity in Language. Retrieved from https://www.thoughtco.com/productivity-language-1691541.

Jinghui Ouyang, Jingyang Jiang, and Haitao Liu. Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51 (2022):100603.

Haihua Pan and Xiaoshi Hu. 2021. The passive construction in Chinese. In Mark Aronoff (ed.), *Oxford research encyclopedias: Linguistics*, 1–29. Oxford University Press, Oxford, UK.

Wenguo Pan. 1997. *汉英语对比纲要 [Outline of Chinese-English comparison]*. Beijing Language and Culture University Press, Beijing, China.

Siying Tang. 2006. 汉语被动句的三个句法问题 [Three syntactic issues of Chinese passives]. In Fuyi Xing (ed.), *汉语被动表述问题研究新拓展 [New developments in the issues on passive expressions in Chinese]*, pages 92-99. Central China Normal University Press, Wuhan, China.

David Temperley. 2007 Minimization of dependency length in written English. *Cognition*, 105:300–333

Feng-fu Tsao. 1996. On verb classification in Chinese. *Journal of Chinese Linguistics*, 24:138–191.

Hua Wang. 2018. 汉英名词短语长度的计量研究 [A quantitative study on the length of Chinese and English noun phrases]. In Haitao Liu (ed.), *计量语言学研究进展 [Advances in quantitative linguistics]*, 268-280. Zhejiang University Press, Hangzhou, China.

Richard Xiao. 2015. Contrastive corpus linguistic studies of English and Chinese: the cases of passive constructions and classifiers. *Journal of Chinese Linguistics Monograph Series*, (25):82-114.

Richard Xiao, Tony McEnery, and Yufang Qian. 2006. Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in contrast*, 6(1):109-149.

Chunshan Xu and Haitao Liu. 2015. Can familiarity lessen the effect of locality? A case study of Mandarin Chinese subjects and the following adverbials. *Poznan Studies in Contemporary Linguistics*, 51(3):463-485.

Xiuling Xu and Jiajin Xu. 2021. 燚炎英汉平行语料库的创建 [Yiyan English-Chinese Parallel Corpus]. *Corpus Linguistics*, (01):3.

Jianwei Yan and Haitao Liu. 2021. Morphology and word order in Slavic languages: Insights from annotated corpora. *Вопросы языкознания [Linguistic Questions]*, 4:131-159.

Zhigong Zhang. 1953. *汉语语法常识 [Elementary knowledge of Chinese grammar]*. China Youth Press, Beijing, China.

Dexi Zhu. 1982. *语法讲义 [Lectures on grammar]*. The Commercial Press, Beijing, China.

. 2014. “ToRCH2009 现代汉语语料库”建成 [ToRCH2009: Texts of Recent CHinese]. *Corpus Linguistics*, (01):103-104.