From Modern to Ancient Chinese: A Corpus Approach to Beneficiary Structure

Yu-Yin Hsu[1] and Tao Wang[2]

[1] Hong Kong Polytechnic University, [2] Sichuan University
yyhsu@polyu.edu.hk; caichcd@Outlook.com

Abstract. This study reports the results of two sets of corpus studies on the use of beneficiary structures (*wèi-dòng shì*), one in modern and the other in ancient Chinese. First, we analyzed the semantic associations of the word *wèile* 'do something for something/someone' in modern Chinese, using two corpora and the word-embedding model. The results were in line with semantic analyses proposed in the Semantic-Map Model. Second, based on an examination of all the sentences expressing beneficiary meanings in *Zuo's Commentary* and *Mencius*, we established that the beneficiary structure in those works involves a light-verb structure that should be syntactically distinguished from other such structures that introduce causative and intentional events. As well as providing some new evidence regarding the semantic content of the *wèi-dòng shì* in modern Chinese, we present structural evidence of its source, which can be dated to the pre-Qin period, as shown by the examples in the two target ancient-Chinese texts.

Keywords: *Chinese GigaWord Corpus 2.0*, beneficiary structure, word-embedding, modern Chinese, *Zuo's Commentary*, *Mencius*

1 Introduction

In linguistic typology, Modern Chinese is generally regarded as an isolated language. Recently, some Chinese sentences have appeared with uncommon forms that are unlike the typical analytical structures of modern Chinese, and more similar to structures in synthetic languages. These include examples like *bèi xiāngqīn* 'to be match-made', in which the verb *xiāngqīn* 'blind-date' usually does not allow passivization, and *děng-huǐ'er duǎn wǒ* 'Later, P(rivate)-M(ail) me', the meaning of which was formerly expressed using the ditransitive verb *fā* 'send' (as in *děng-huǐ'er fā duǎnxìn gěi wǒ* 'Later, send a P-M to me'). In the recent linguistic literature, light-verb theory is often used to account for these special sentences in modern Chinese [1, 2]. However, in light of another special type of sentence in modern Chinese that will be discussed in Section 2, below, we will argue that in addition to the light-verb structures proposed in the existing literature, there is another type, called *wèi-dòng* beneficiary structure, whose existence is related to linguistic typology's semantic-map theory [3] (see Section 3), and its syntactic source can be found based on data from the pre-Qin dynasty era around the time in 221 BC. Our results for modern Chinese are obtained via inspecting the dynamic meanings of the *wèi-dòng* structure as used in the *Chinese GigaWord Corpus 2.0* [4], and via applying the word-embedding model [5] to more than 240,000 sentences containing *wèi-le* 'do something for something/someone' in modern Chinese (see Section 4). With the semantic results obtained from Sections 3 and 4, we then analyze the sentences conveying *wèi-dòng* meanings in the ancient historical texts, *Zuo's Commentary* and *Mencius*, and use the results to provide a syntactic analysis of the *wèi-dòng* structure in Section 5. Lastly, we briefly sum up our findings and their implications in Section 6.

2 Meanings of *wèi-dòng* and the light-verb theory

Recently, researchers have called attention to some special Chinese sentences that are atypical of an analytic language ([1, 2]). For example, the meaning of (1a) has traditionally been said in the form (1b), and the sentence in (2a) has until recently been more commonly said as (2b).

(1) a. Qǐng diàn fúwù.chù.
please call service.center
'Please call the service center.'
b. Qǐng dǎ.diànhuà gěi fúwù.chù.
please make.phone.call to service.center
'Please make a phone call to the service center.'

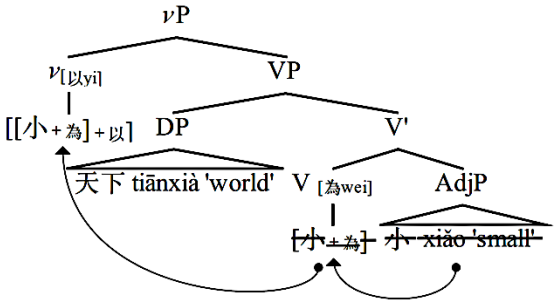
(2) a. Lǎo Wáng bèi tuìxiū le.
Lao Wang passive retire perfective
'Lao Wang got retired.'
b. Lǎo Wáng bèi qiǎngpò tuìxiū le.
Lao Wang passive force retire perfective
'Lao Wang has been forced to retire.'

Tsai [2] has pointed out how light-verb theory can explain such special patterns in modern Chinese. According to that theory, the three most common light verbs are *shǐ-dòng* (causative), *chéng-shì* (become), and *yì-dòng* (intentional) ([6, 7, 8]). A classic analysis of light verbs is that they are denominal verbs [7]. As shown in (3), in the lexical-syntax derivation (L-Syntax), the word *shelf* is moved up from its original position inside a post-verbal prepositional phrase, and merges first with the light verb BECOME, and then moves and merges with the light verb CAUSE, so that *shelf* obtains verbal functions and can serve as a verb taking a direct object, *the book*. Another similar example is 'bench the player' meaning 'order the player to sit on the bench rather than continue playing baseball', where the noun 'bench' serves as a verb taking a direct object.

(3) 'shelf the book' ([5], citing [7])
…… [$_{vP}$ CAUSE [$_{VP}$ the book [$_{V'}$ BECOME [$_{PP}$ ON the shelf]]]
→ [[[shelf + ON] + BECOME] + CAUSE] t_V t_P t_N

Light-verb theory has also been used to analyze *shǐ-dòng* and *yì-dòng* expressions found in ancient Chinese ([9, 10]): e.g., in *xiǎo tiānxià* in (4), *xiǎo* 'small' is not an adjective, but expresses a verbal meaning 'to regard something as small'. This structure is illustrated in (4b) ([9, 10]).

(4) a. Dēng Tài.shān ér xiǎo tiānxià.
Climb Mount-Tai and small world
'Climb Mount Tai and deem the world to be small.'
b.



However, we would like to point out that some expressions in modern Chinese cannot be explained semantically by the three common types of light verbs mentioned above. For example, the (a) examples in (5) and (6) would typically be uttered as the (b) examples, and their meanings are not related to causation, becoming, and/or the intentional 'taking/being' as specified in the theoretical literature on light verbs. However, all of them contain a sense of 'benefiting someone or something' or 'benefiting the public good' (i.e., 'beneficiary'). Other examples involved with similar structures include *bèi zhàn* 'prepare for war', *jiù huǒ* 'fight fires', *yǎng bìng* 'recuperate from illness', and *máng kǎoshì* 'be busy preparing for exams'.

(5) a. dǎsǎo wèishēng
clean hygiene
'to clean up for the sake of hygiene/health'
b. wèile wèishēng ér dǎsǎo
for/consider hygiene then clean
'to clean up for the sake of hygiene/health'

(6) a. xiàn shēn zǔguó
sacrifice body nation
'to sacrifice one's life for one's country'

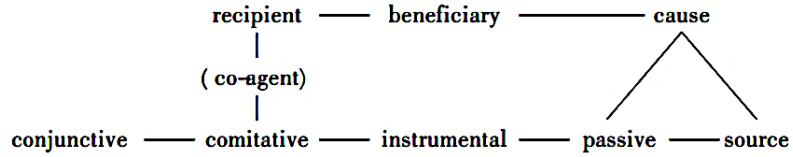
b. wèile zǔguó ér fèngxiàn zìjǐ.de shēngmìng
for/consider nation then sacrifice one's life
'to sacrifice one's life for one's country'

We also note that the meanings of *wèile* in (5) and (6) show some differences; that is, in (5) it is related to the aim or goal that "causes" the events to happen, whereas in (6) *wèile* shows a beneficiary relationship. In semantic theory, these different meanings are identified as cause and beneficiary meanings. Below, we look at how these two concepts are accounted for in semantic-map theory.

3 The meanings of *wèile* and semantic-map theory

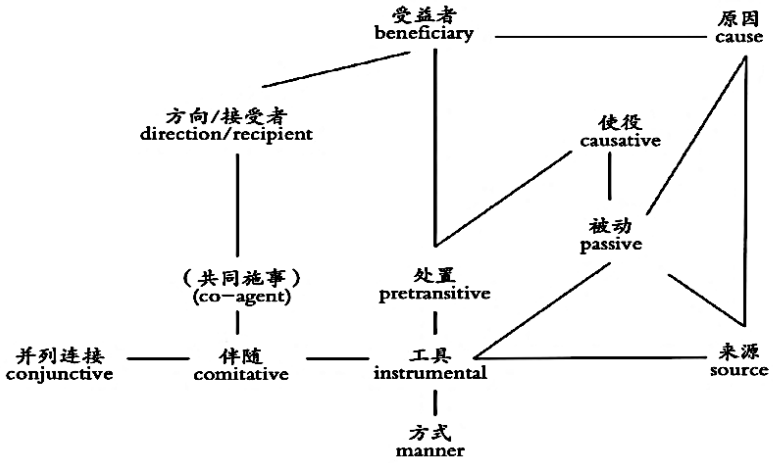
The semantic-map theory presents semantic concepts with similar functions in close proximity to one another in a universal semantic space, and holds that if a linguistic form assumes two semantic functions, then those functions must be adjacent in that space ([3]). Given that there are two semantic functions associated with the expression *wèile* in modern Chinese, cause and beneficiary, it is expected under the semantic-map model that these two functions should be adjacent in the universal semantic map, and this is indeed the case: in the base graph of the "tools" class (7), the beneficiary function ('for someone') is adjacent to the cause function ('for this reason') [3].

(7) The semantic base of the "tools" class ([3])



Based on an examination of the semantic base of the "tools" class of the universal semantic map [3] in light of materials from modern Chinese dialects, Zhang [11] constructed a semantic map (8) corresponding to the semantic relations of *passivity*, *causation*, *disposal*, *tool*, and *beneficiary* in Chinese.

(8) Base map of "disposal" and its related functions ([11])



The adjacency of the beneficiary function and the cause function in (8) provides important theoretical support for the dual meanings associated with the form *wèile*. Based on this result, in Section 4, we use a corpus to investigate the correlation between *wèile*'s meaning and the meanings of some other terms related to the cause and beneficiary functions in modern Chinese. Then, in Section 5, we go on to investigate how such meanings are expressed as well as the related argument types in the corpora of *Zuo's Commentary* and *Mencius*.

4 *Wèile* in modern Chinese and word embedding

4.1 *Wèile* in the Center for Chinese Linguistics corpus

From the Center for Chinese Linguistics (CCL) corpus of modern Chinese [12], we extracted the entries containing *wèile*. After eliminating redundant components, this yielded more than 130,000 effective linguistic data points. Since the CCL corpus has not been annotated with parts of speech, we performed our semantic analysis based on a 1:100 scale distributed sampling of 1,300 records.

Overall, purpose and beneficiary meanings were *wèile*'s major senses in the dataset. However, many more *wèile* sentences expressed purpose than beneficiary readings (939 vs. 361). These two types of meanings appeared in various linguistic units such as words, phrases, and clauses. It was because of these observations that we decided to perform a more robust semantic analysis of the use of *wèile* in modern Chinese.

4.2 A word-embedding analysis of *wèile*

Given that modern Chinese allows some synthetic structures like (5a) and (6a), and the fact that the use of the term *wèile* in modern Chinese can introduce either a purpose or beneficiary meaning, we used the word-embedding model [5] to examine whether such semantic associations of *wèile* can be observed in a larger-scale corpus. For this, we used the *Chinese GigaWord Corpus 2.0* [4], which uses the Chinese Word Sketch (CWS) system built by Academia Sinica and SkE Team.

The beneficiary structure (*wèi-dòng shì*) in modern Chinese is primarily expressed by the verb *wèile.* However, the CWS system only allows a monosyllabic search of 为 *wèi*/*wéi*, not *wèile.* According to the *Xinhua Dictionary* online [13], the character 为 has two primary meanings in modern Chinese. When it is used to mean "to do, execute, and become", it is not relevant to the beneficiary structure at issue. Its meaning becomes relevant only when it is "purpose, substitute/give, or help". To avoid uses irrelevant to the beneficiary structure, we first turned to the *Xinhua News Agency in Mainland corpus* (Gigaword2xin) available in the CWS system (containing 311,660,000 tokens across 992,261 documents) and extracted 101,868 sentences containing *wèile*. Next, we used the Doc2Vec model [14] to process all the semantic relations between *wèile* and the other words in the extracted sentences.

Doc2Vec calculates the cosine vectors of words' distribution in a given set of documents through an unsupervised learning algorithm, and returns the probability of the occurrence of each word while retaining information on word order within sentences and the semantic impacts of words in their respective contexts. Possibly because the limit of textual content, our results showed that among the 50 words most closely related to *wèile*, only five were adjectives or verbs, with all others being nouns. Moreover, the word with the highest degree of similarity to *wèile* was only 0.5254 similar. Next, we compared four terms that are often discussed as semantic features of the beneficiary structure, i.e., *yuányīn* 'reason/cause', *mùdì* 'purpose', *hǎochù* 'advantage', and *shòuhuì* 'benefit'. In (9), we present the degree of similarity of these words to *wèile* as calculated by the Doc2Vec model. The fact that the results are influenced by the contexts of words in a given text is evidenced by the differences among the results for each word pair.

(9) Degrees of similarity of *wèile* to words associated with the beneficiary structure (Gigaword2xin)

a. *wèile* and *yuányīn* 'reason/cause': 0.2450
b. *wèile* and *mùdì* 'purpose': 0.1443
c. *wèile* and *hǎochù* 'advantage': 0.2401
d. *wèile* and *shòuhuì* 'benefit': 0.2197

The Doc2Vec model also allows us to calculate the semantic associations formed between more than one word and their relations to other words in a given text. Our comparison of the semantic intersection of *wèile* with reason (*wèile+reason/cause*) and *wèile* with benefit (*wèile+benefit*) yielded a very high similarity, 0.895. This suggests that *wèile* expressing 'cause for reasons' and 'for benefit' are closely correlated.

The Doc2Vec model was also used to identify the 50 words in the *wèile* dataset we extracted from Gigaword2xin that had the highest degree of semantic association with *wèile+reason/cause* (i.e., meaning 'for the sake of the reason') and with *wèile+benefit* (i.e., meaning 'for the sake of the good'). Interestingly, we observed a difference in how these two semantic combinations of *wèile* affected other words. As shown in (10), both these subsets had the same number of adjectives and of prepositions/conjunctions. However, the 'for the sake of the reason' subset included many more nouns than verbs, while the 'for the sake of the good' had a similar number of nouns and verbs. Although further research will be needed to explain the correlation between word-class selection and the associated semantic set, differences related to word-class distributions may indicate some qualitative differences between semantic and structural interactions.

(10) Counts (percentages) of the parts of speech represented by the 50 words most similar to *wèile*

	Nouns	Verbs	Adjectives	Prepositions/ Conjunctions
wèile + reason/cause	30 (60%)	5 (10%)	10 (20%)	5 (10%)
wèile + benefit	17 (34%)	18 (36%)	10 (20%)	5 (10%)

In addition, we wanted to establish if there were differences between the use of causative structure (*shǐ-dòng*) and intentional structure (*yì-dòng*) in our chosen corpora. Again using the Doc2Vec model, we took the words *zàochéng* 'cause' and *dǎozhì* 'result in' as representative of the causative structure and compared them with *wèile.* Likewise, we compared *wèile* against the words *rènwéi* 'consider' and *yǐwéi* 'think', taken as representative of intentional structure. The results show that the semantic similarity between these concepts and *wèile* in the corpus texts tended to be low, as shown in (11a-b) for *shǐ-dòng* structure, and in (11c-d) for *yì-dòng* structure.

(11) Degrees of similarity of *wèile* to *shǐ-dòng* and *yì-dòng* in Gigaword2xin

a. *wèile* and *zàochéng* 'to cause': 0.1048
b. *wèile* and *dǎozhì* 'to result in': 0.1507
c. *wèile* and *rènwéi* 'to consider': 0.1285
d. *wèile* and *yǐwéi* 'to think': 0.1506

We then extracted further 147,613 sentences containing *wèile* from another, larger corpus in the CWS system, the *Central News Agency in Taiwan corpus* (Gigaword2cna), which contains 501,456,000 tokens across 1,769,953 documents. We used the same Doc2Vec model to identify the semantic relationships within this dataset, and as shown in (12), the results were basically similar to those we derived from the other corpus, but with lower levels of correlation between *wèile* and the class of reason/purpose. Nonetheless, when we compared the semantic intersections *wèile+reason/cause* and *wèile+benefit*, we again found very similar results (0.9017). In other words, in Gigaword2cna as in Gigaword2xin, *wèile*'s semantic association with related words are similar.

(12) Degrees of similarity of *wèile* to words associated with the beneficiary structure (Gigaword2cna)

a. *wèile* and *yuányīn* 'reason/cause': 0.1449
b. *wèile* and *mùdì* 'purpose': 0.1355
c. *wèile* and *hǎochù* 'advantage': 0.2297
d. *wèile* and *shòuhuì* 'benefit': 0.219

In summary, supplementing our findings arrived at via semantic map theory, our corpus studies indicate that the beneficiary structure as expressed by *wèile* in modern Chinese differs from both *shǐ-dòng* (causative) and *yì-dòng* (intentional). In the next section, we further test this analysis through a syntactic examination of the beneficiary structure in two ancient Chinese documents, *Zuo's Commentary* and *Mencius*.

5 The beneficiary (*wèi-dòng*) structure in ancient Chinese

The foregoing study of modern Chinese has shown that the similarity between "cause for reasons" and "cause for benefit" is high: i.e., 0.895 and 0.902 cosine similarities. Our preliminary examination of ancient Chinese texts suggested that the meaning of phrases in the class 'benefit' was associated not only with people but also with events. Likewise, it indicated that the associations of phrases in the class of 'purpose' may not be limited to a single event, and can also be persons. These findings imply that there could be two sub-classes of the beneficiary structure, since the object selections of these two meanings are not complementary, and either can take either a human or an event as its argument.

Zuo's Commentary (375-351 BC) is an ancient Chinese narrative history covering a period from 722 to 468 BC. *Mencius* (ca. late 4th century BC) is a collection of anecdotes about and conversations of the philosopher Mengzi and other Confucian thinkers, on topics in moral and political philosophy. In *Zuo's Commentary* [15], we identified 110 occurrences of the *wèi-dòng* structure, of which 22 were "benefited" and 88 showed a purpose relationship. We found that the purpose structure could be used to describe a cause-purpose event related to a person or to an event or process. For example, *qǐ zhī* in (13a) means 'to open the door for him,' while *qín wáng* in (13b) means 'to help and support actions related to the best interests of the emperor/the royal family.'

(13) a. 大叔 完 聚…… 將 襲 鄭， 夫人 將 啓 之。
Tàishū wán jù…… jiāng xí Zhèng, fūrén jiāng qǐ zhī.
Taishu finish gathering about.to attack state.Zheng lady plan open.for him
'Taishu finished gathering troops and was about to attack the state of Zheng. The lady was planning to open the gate for him to enter.'

b. 求 諸侯, 莫 如 勤 王。
Qiú zhūhóu, mò rú qín wáng
request duke not like support.for emperor
'It is better to facilitate actions related to the best interests of the Emperor than to make a request to the Dukes.'

Similarly, the "benefit" function in *Zuo's Commentary*, while most commonly associated with a person, is not limited to persons as objects. For example, *bǔ qī Jìngzhòng* in (14) means 'to perform divination about marrying his daughter to Jingzhong'. Thus, in this beneficiary structure, the verb phrase *qī Jìngzhòng* 'to marry [his daughter to] Jingzhong' functions as the object.

(14) 懿氏 卜 妻 敬仲。
Yìshì bǔ qī Jìngzhòng
Yishi tell.fortune marry Jingzhong
'Yishi performed divination about marrying his daughter to Jingzhong.'

In *Mencius*, we identified 26 occurrences of the *wèi-dòng* structure, of which only two were of the 'for benefit' class. One of these two examples, *sǐ qí-zhǎng* in (15a), means 'to sacrifice their lives for the leader', and *mò zhī sǐ* in (15b) interpreted as *mò sǐ zhī* meaning 'not to sacrifice their lives for the officers.'

(15) a. 斯民 親 其 上, 死 其 長 矣。
Sīmín qīn qí shàng, sǐ qí zhǎng yǐ
people close their leader die their leader then
'People worship their leader and are willing to die for him.'

b. 吾 有 司 死 者 三十三.人, 而 民 莫 之 死 也。
Wú yǒu sī sǐ zhě sānshísān.rén, ér mín mò zhī sǐ yě.
I have officer die one 33.person and people not them die.for particle
'I have 33 officers died, but no one is willing to die for them.'

The other 24 examples all had cause-purpose meanings. For example, *qǐng shìshī* in (16) means 'to request for the position of being a judge.'

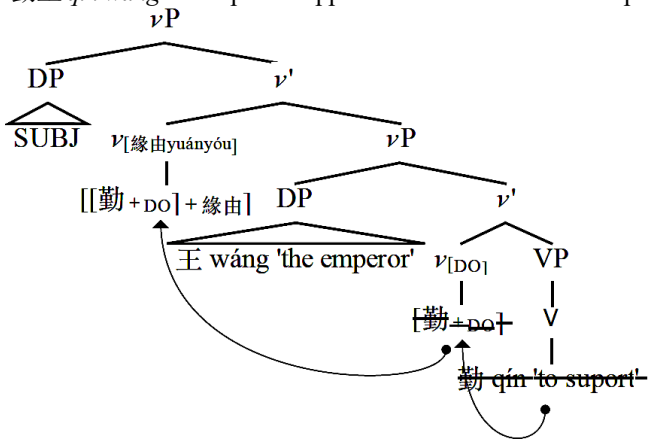
(16) 子 之 辭 靈丘 而 請 士師
Zǐ zhī cí Língqiū ér qǐng shìshī
you possessive leave Lingqiu then ask.for judge
'Your leaving Lingqiu to request for the position of being a judge.'

We also found one example of a prepositional phrase being used to modify the beneficiary structure. In (17), *wèi wǒ* 'for me' modifies the purpose structure (i.e., *yán zhī* 'to say good things for it') to introduce the person (*wǒ* 'I') who receives the benefit.

(17) 子 盍 為 我 言 之?
Zi hé wèi wǒ yán zhī
you why for me say it
'Why don't you put in a good word about it for me?'

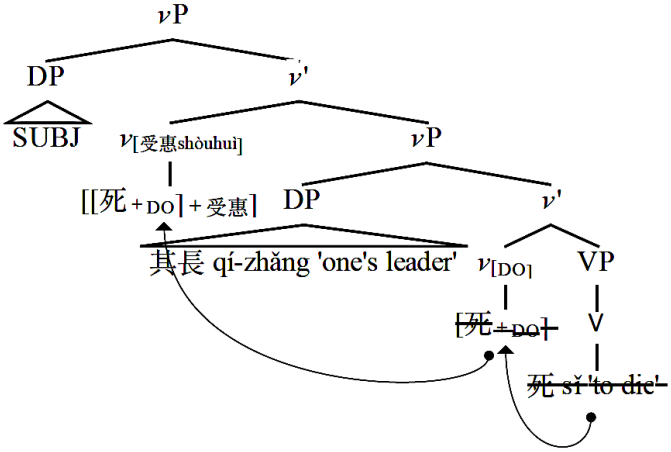
Based on the semantic distinction between the beneficiary structure under discussion and other typical light-verb structures, along with prior light-verb analyses of ancient Chinese [9, 10], we propose that two light verbs that are unique to the beneficiary sense should be identified in syntax. They are 緣由 *yuányóu* (for cause-purpose) and 受惠 *shòuhuì* 'benefit'. For example, the phrase *qín wáng* 'to help and support actions related to the emperor/the royal family' in (13b) can be derived through a structure like (18), in which the verb *qín* 'to support' moves to pick up the active feature (for direct object) and then moves to the light verb *yuányóu* to express cause-purpose regarding the object noun phrase, *wáng*. The verb *qín* then moves to *v* to license the subject of the whole event.

(18) 勤王 *qín wáng* 'to help and support actions related to the emperor/the royal family'



When the sentence is about doing something for benefit, another light verb *shòuhuì* 'benefit' is involved. For example, *sǐ qí-zhǎng* 'to die for one's leader' in (15) can be derived through a structure like (19), in which the verb *sǐ* 'to die for' moves to pick up the active feature (for direct object) and then raises to the light verb *shòuhuì* to express beneficial effects on the object noun phrase *qí-zhǎng*. The verb *sǐ* then moves to *v* to license the subject of the whole event.

(19) 死其長 *sǐ qí-zhǎng* 'to die for one's leader'



6 Conclusion

Important theoretical support for the semantics of different uses of *wèile*, in keeping with semantic-map theory, has been provided by this paper's investigation of the semantic correlations between *wèile* in modern Chinese, on the one hand, and on the other, some key words related to typical light-verb structures in two corpora of modern Chinese and those sentences expressing the beneficiary structure in two ancient Chinese texts.

For our corpus study of modern Chinese, we used a basic word-embedding model to analyze all those sentences containing *wèile* in two large datasets within the *Chinese GigaWord Corpus 2.0*. Our results, though preliminary, show that analyses of word embedding can indeed be influenced by which texts are used, and yet, that the beneficiary structure at issue is semantically different from the three typical light-verb structures that we also investigated. Then, our corpus study of ancient Chinese usage of the beneficiary structure, based on *Zuo's Commentary* and *Mencius*, found that such structure differed semantically from other light-verb structures, and that the subtypes of its beneficiary associations could probably be expressed by two light verbs, one being *yuányóu* (used to express cause-purpose meanings), and the other, *shòuhuì* 'benefit'.

The existence of the beneficiary data in *Zuo's Commentary* and *Mencius* also provides some evidence regarding the structural source of the emergence of the new beneficiary structure in modern Chinese. Given that typologically, ancient Chinese is generally considered a synthetic language, the sharp difference in the number of beneficiary-structure occurrences between *Zuo's Commentary* (110 cases), which dates from the early 4th century BC, and *Mencius* (26 cases), from perhaps three quarters of a century later, may imply a transformation of the Chinese language from a synthetic to an analytic one, partially within that period.

References

1. Huang, C. T. J., Liu, N.: The Syntax and Semantics of the New Non-canonical Bei(被) XX Construction. Linguistic Sciences. 13(03), 225-241 (2014) (in Chinese)
2. Tsai, W.-T.: On the Distribution and Interpretation of Inner and Outer Light Verbs in Chinese. Linguistic Sciences. 15(04), 362-376 (2016) (in Chinese)
3. Haspelmath, M.: The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Tomasello, M. (ed.): The New Psychology of Language, vol. 2. Erlbaum, New York 211–243 (2003)
4. Huang, C. R.: Tagged Chinese Gigaword version 2.0 (2009) https://catalog.ldc.upenn.edu/LDC2009T14
5. Mikolov, T., Yih, W.-T. Zweig, G.: Linguistic regularities in continuous pace word representations. 1, 746–751 (2013)
6. Lin, T.-H. Jonah: Light Verb Syntax and the Theory of Phrase Structure. Ph.D. dissertation, University of California, Irvine (2001)
7. Hale, K., Keyser, S. J.: Prolegomenon to a Theory of Argument Structure. MA, The MIT Press, Cambridge (2002).
8. Tang, S.-W.: Formal Chinese Syntax. Shanghai Educational Publishing House, Shanghai (2010) (in Chinese)
9. Feng, S.-l.: Light Verb Movement in Modern and Classical Chinese. Linguistic Sciences. 1, 3-16 (2005) (in Chinese)
10. Feng, S.-l.: Light verb syntax between English and classical Chinese. In: Li, Y.-H. A., Simpson, A., Tsai, W.-T. (ed.): Chinese Syntax in a Cross-linguistic Perspective. Oxford University Press, New York 229-250 (2015)
11. Zhang, M.: On the 'Constant' and 'variables' Towards Spatial Maps and Semantic Maps: Take the Relation of Passive, Causative, Disposal, Tools, Beneficiaries in Chinese. Lecture at the Institute of languages of the Chinese Academy of Social Science (2008) (in Chinese)
12. Zhan, W.-D., Guo R., Chen, Y.-R.: The CCL Corpus of Chinese Texts: 700 million Chinese Characters, the 11th Century B.C. - present. Peking University (2003) http://ccl.pku.edu.cn:8080/ccl_corpus
13. XinHua Dictionary Online (*Xianshang XingHua Zidian*) https://zidian.911cha.com/zi4e3a.html
14. Le Q., Mikolov,T.: Distributed representations of sentences and documents. In: The 31th International Conference on Machine Learning, Vol. 32. Beijing 1188-1196 (2014)
15. Yang, B.-J.: Annotation of Zuo's Commentary (Revised Edition). Zhonghua Book Company. Beijing (1990) (in Chinese)