

Quantitative Analysis on verb valency evolution of Chinese¹

Bingli Liu

Institute of Chinese and Culture
Education Studies, Huaqiao University,
Xiamen, 361021, China

bingli_liu@yahoo.com

Chunshan Xu

School of Foreign Studies, Anhui
Jianzhu University, Anhui, 230601,
China

adinxu@126.com

Abstract

The paper aims at studying the evolution of syntactic valency of Chinese verbs. We construct three corpora of ancient classical Chinese, ancient vernacular Chinese and modern vernacular Chinese. From these corpora, ten main verbs are selected to probe into the evolution of their valency, namely, their complements and adjuncts. The paper reveals that the syntactic structures has a trend toward complex. The ancient classical Chinese and the ancient vernacular Chinese are similar in sentence structure. With the transformation from the ancient vernacular to the modern vernacular, syntactic complexity increases dramatically, indicating drastic changes in sentence structure.

1. Introduction

Valency is a property of words (Tesnière, 1959). It refers to the ability of words to syntactically or semantically to combine with other words (Liu, 2007). It is determined by the meaning of the word itself. The valency relations realized in sentences are dependence relationships between words (Liu, 2009). Quantitative investigations into valency may reveal some syntactic and semantic features of human language. Based on the German valency dictionary, Köhler studies some quantitative characteristics of the German verbs valency (Köhler, 2000; 2005; 2007). Čech et al. (2010) studied Czech valency framework distribution and verified the hypothesis about the relationship between the number of valency frames and the word length. And they proposed the concept of "full valency" without distinguishing between complements and declaratives. Liu(2011) conducted quantitative studies into English verb valency and concludes that the number of meanings of English verbs obeys the positive negative binomial distributions -- the more meanings a word has, the bigger the valency is.

However, most of these studies focus on the synchronic description of the word valency. but ignore the diachronic changes of word valency. Meanwhile, the description and study of the valencys in a corpora are not balanced in every historical period and there is a lack of

¹ Huaqiao University's Academic Project Support by the Fundamental Research Funds for the Central Universities. No.19SKGC-QG04

uniform descriptive principles and methods. Our studies try to answer the question that how the main verbs valency evolve during the long period from ancient classical Chinese to modern vernacular Chinese.

2. Methods and Materials

This study statistically analyzes 2,817 examples from ancient classical Chinese, ancient vernacular Chinese and modern vernacular Chinese. Being different from the existing researches on valency, this paper is concerning with the macroscopic and the diachronic picture of Chinese verb valency. The study explores the evolution of the "syntactic value" of verbs in real corpus. Not only will the present study help us to learn more about the context of language development, but also indicate the possible role of the valency in natural language processing.

The diachronic researches involve comparison and contrast of different diachronic stages of a language. Historically, Chinese can be roughly divided into ancient classical Chinese, ancient vernacular Chinese and modern vernacular Chinese. The ancient classical Chinese dates back from 1600 BC to 618 AD; the ancient vernacular Chinese dates back from 618 to 1911; 1912 is often taken as the year dividing the modern vernacular Chinese and the ancient vernacular Chinese.

In order to reflect the overall linguistic properties of each period, we try to cover as many genres as possible when constructing the corpora. The corpus of ancient classical Chinese language include *Zuozhuan* (narrative chronicle), *Lv Shi Chun Qiu* (a book on political theory), *Liu Tao* (a book on military strategies), *Shangshu* (government archives), *Mencius* (quotations from a sage), *Xunzi* (a book on philosophical treatise), *Zhanguoce*, *Shiji*, *Han Shu*, *Sanguozhi*, *Houhanshu* (5 books on history), *Guoxiaoshuogoucheng*, *Shishuoxinyu* (novels); the corpus of ancient vernacular Chinese include samples from *Dunhuangbianwenji*, *Qingpingshantanghuaben*, *Xixiangji*, *Sanguoyanyi*, *Chukepaianjinqi*, *Erkepaianjinqi*, *Shuihuzhuan* and *Xiyouji* (novels or playbooks); the corpus of modern vernacular Chinese mainly include samples from novels.

Ten verbs are selected because of their diachronic lexeme stability: 走(walk), 听(listen), 到(arrive), 爱(love),有(have), 为(be), 能(can), 来(come), 使(let), 愿(wish).

In total, we annotated 3,128 sentences, of which 1,383 are from ancient classical Chinese, 813 from ancient vernacular Chinese, and 932 from modern vernacular Chinese. Table 1 shows the frequencies of sentence containing the above 10 verbs.

	ancient classical Chinese	ancient vernacular Chinese	modern vernacular Chinese
到(arrive)	166	52	102
来(come)	91	97	78
爱(love)	98	51	70
能(can)	211	100	100
使(let)	210	62	97

听(listen)	71	76	72
为(be)	201	100	100
有(have)	207	116	104
愿(wish)	53	108	151
走(walk)	75	51	58
Total	1383	813	932

Table 1. Number of sentence containing the 10 verbs

To study the valency of verbs, we need sentences where the verbs appear. We take the following criteria to select sentences:

- (1) The verb is used in the active voice
- (2) The verb has similar semantic meaning across different periods

Then we begin to annotate the sentences which include the verbs chosen according to the dependency grammar. Basically, syntactic dependency can be roughly divided into two types: complement and adjunct(Liu, 2011). Complement relationships involve arguments like subjects, objects or complements. The adjunct relationships often involve adverbials and attributives.

3. Results and Discussion

3.1 Increasing complexity of subjects and objects

According to their syntactic complexity, subjects and objects of these verbs can be divided into two categories: simple subject and complex subject. Simple subjects are single words, such as nouns and pronouns, while complex subjects include phrases, such as numerical-classifier phrase, noun phrases and verb phrases. Figure 1 and Figure 2 show the ratio of complex subjects and complex objects in three forms of Chinese.

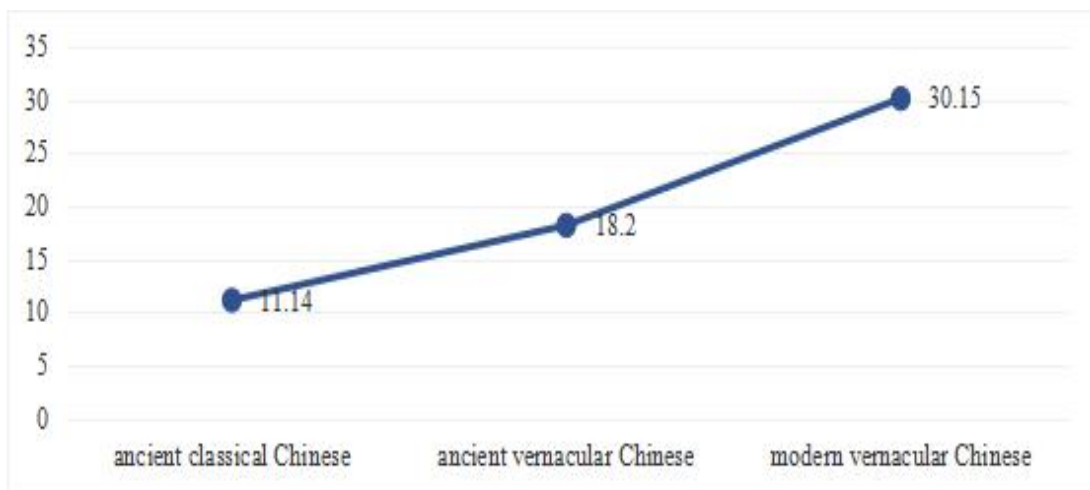


Figure 1. Average ratio of the complex constituent in the subject argument (%)

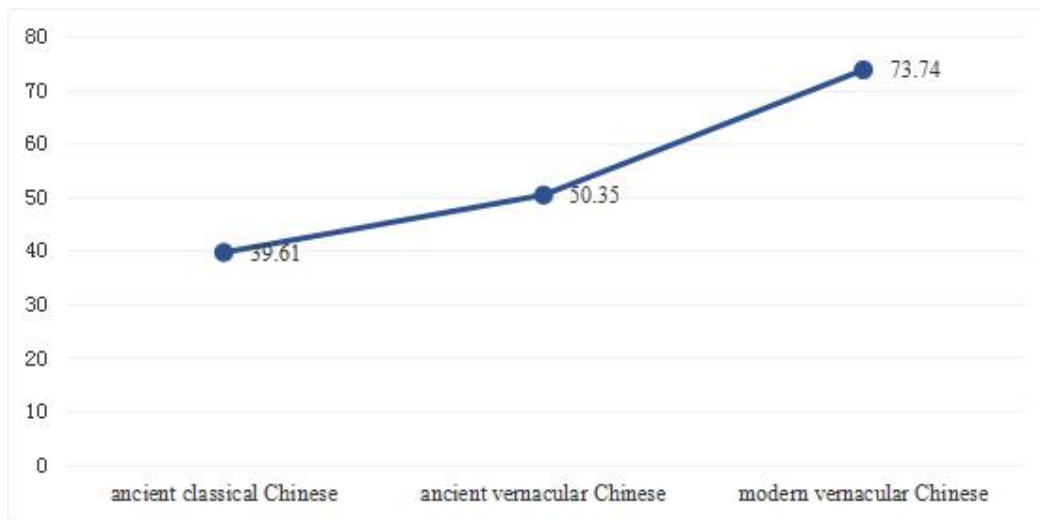


Figure 2. Average ratio of the complex constituent in the object argument (%)

Generally, the complexity of the subjects increases in ancient classical language, ancient vernacular Chinese and modern vernacular Chinese, as indicated by the increasing ratio of complex subjects. The complexity of the objects also increases diachronically, as indicated by the increasing ratio of complex objects in Figure 2. This suggests a tendency in Chinese to evolve into more complexity.

H0 is that all the complex sign has nothing to do with the time. We use SPSS to do the chi-square test. Result for Figure 1 and Figure 2 are $p < 0.001$. It means the difference in Figure 1 and Figure 2 are both highly significant. The reason for this significant difference is complex since language is a complex adaptive system. Maybe because writing is more and more convenient with the society development or the human thinking becomes more complex.

3.2 Increasing use of Complements and Adjuncts

Among the ten verbs, seven verbs may take complements: 到 (arrive), 听 (listen), 走 (walk), 来 (come), 为 (be), 有 (have), 爱 (love). Figure 3 shows the average ratio of other constituent in the complement.

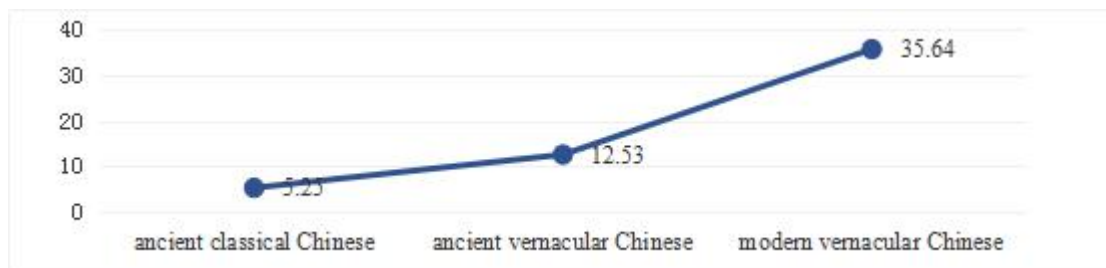


Figure 3. Average Ratio of other constituent in the complement (%)

Diachronically, these verbs show a growing tendency to take complements. From ancient classical Chinese to ancient vernacular Chinese, the proportion of complements increases by 7.28% ; from ancient vernacular Chinese to modern vernacular Chinese, the proportion

increases by 23.11%. At the same time, the types of complements increase from 5 to 9. In short, the diversity and the frequency of complements both increase.

Chi-square test result for Figure 3 is that $p < 0.001$. It means the difference is also highly significant. Besides the reasons mentioned above, maybe it is also relevant to the factors inside the language such as parts of speech function.

Not only have the complements been used more frequently, but also the adjuncts. In the present study, we are mainly concerning with two types of adjunct valency: the adverbial and the topic

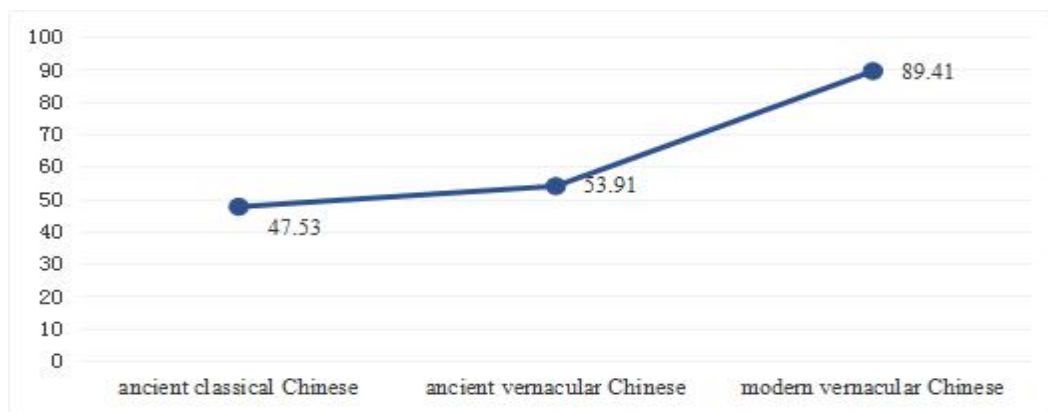


Figure 4. Average Ratio of the adjunct in the sentence (%)

The statistical data from figure 4 shows that the frequencies of these two types of adjunct valency increases diachronically. From the ancient classical Chinese to the ancient vernacular Chinese, the proportion increases by 6.38%, while from the ancient vernacular to the modern vernacular, the percentage increases, drastically, by 35.5%. These results strongly suggest that the tendency toward more complexity is not merely found in nominal constructions, or, in the valency patterns of nouns, but also in verbal constructions, or, in the valency patterns of verbs.

Chi-square test result for Figure 4 is $p < 0.001$. It means the difference is also highly significant ($p < 0.001$). And it means syntactic structures has a trend toward complex.

The findings presented in the above tables are diagrammed in Figure 5.

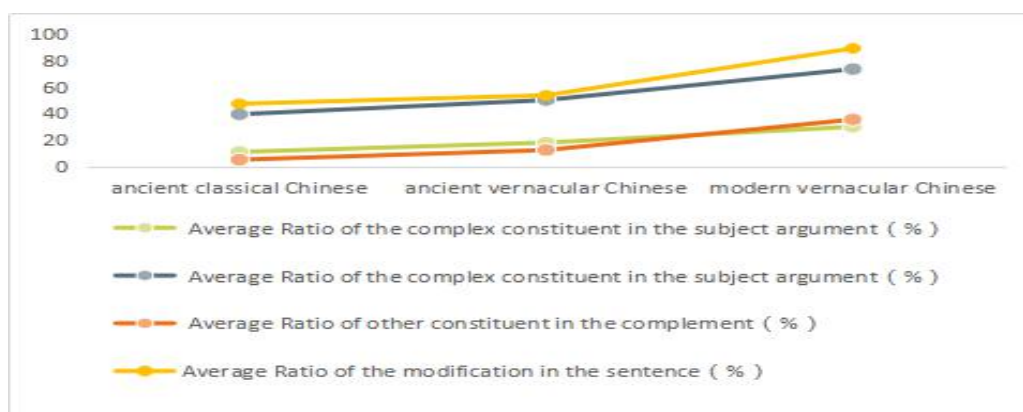


Figure 5. The valency evolution of the main verbs in the three forms of Chinese (%)

4. Conclusion

This quantitative study suggests that Chinese syntax changes gradually with time. As the Chinese language passed through the three stages, that is, the ancient classical Chinese, the ancient vernacular Chinese and the modern vernacular Chinese. The syntactic structure indicates a tendency toward increasing complexity. In other words, the valency patterns of both nouns and verbs have evolved into growing complexity. Moreover, the ancient classical Chinese and the ancient vernacular Chinese are more similar in valency patterns. The transition from the ancient vernacular Chinese to the modern vernacular Chinese seems to be drastically increased in the syntactic complexity.

And the main corpus of this paper is written language, which does not reflect the whole picture of Chinese. The next step will be to expand the scope of the study and adopt more representative oral corpus to carry out statistical analysis so as to explore the evolvement of Chinese valency laydown as far as possible.

References

- Liu Haitao. 2007. Building and using a Chinese dependency treebank. *GrKG/Humankybernetik*, 48(1):3 -14.
- Liu Haitao. 2009. Probability Distribution of Dependencies Based on a Chinese Dependency Treebank, *Journal of Quantitative Linguistics*, volume. 16:256-273.
- Liu Haitao. 2011. Quantitative Properties of English Verb Valency. *Journal of Quantitative Linguistics*, volume 18:207-233.
- Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Klincksieck, Paris.
- Radek Čech, Peter Pajas and Ján Mačutek. 2010. Full Valency. Verb Valency without Distinguishing Complements and Adjuncts. *Journal of Quantitative Linguistics*, Volume 17: 291-302.
- Reinhard Köhler. 2005. Quantitative Studies of Valency of German Verbs. *Glottmetrics*, Volume 9:13-20.
- Reinhard Köhler and Gabriel Altmann. 2000. Probability Distributions of Syntactic Units and Properties. *Journal of Quantitative Linguistics*, Volume 7:189-200.
- Reinhard Köhler. 2007. Quantitative Analysis of Syntactic Structures in the Framework of Synergetic Linguistics. *Studies in Fuzziness and Soft Computing*, Volume 209:191-209.