International Journal of Knowledge and Language Processing
Volume 12, Number 1, **2021**

www.ijklp.org
KLP International ©2021 ISSN 2191-2734
pp.48–61

The Construction and Statistical Analysis of Pre-Qin Ancient Chinese WordNet

Huidan Xu, Siyu Chen, Jingjing Cai, Lin Cao, Chen Wan, and Bin Li
School of Chinese Language and Literature, Nanjing Normal University,
Nanjing 210097, China
libin.njnu@gmail.com

Selected paper from CLSW2019

ABSTRACT. *Pre-Qin ancient Chinese (PQAC) plays an important role in the history of Chinese development. In previous works, most research is focused on sense explanation, while few can show the general vocabulary and the conceptual system characteristic of Pre-Qin Dynasty. In this paper, we construct a preliminary wordnet for Pre-Qin ancient Chinese (named PQAC WordNet(PQAC-WN)) that contains 45,498 Pre-Qin basic words and 63,230 semantic classes. PQAC-WN organizes information based on semantic relationships and establishes lexical semantic mappings among Pre-Qin ancient Chinese, modern Chinese and English. In PQAC-WN, there are 45,168 semantic classes that can be mapped with the CCD table, accounting for about 71.4%; and there are 18,062 semantic classes that cannot find corresponding concepts in the CCD table, accounting for about 28.6%. On this basis, the semantic richness comparison is used to mine the cultural factors behind the vocabulary system.*

Keywords: Pre-Qin ancient Chinese, WordNet, semantic class analysis, lexical semantic

1. **Introduction.** Pre-Qin ancient Chinese plays a significant role in the history of Chinese development. Research on the topic requires knowledge of the earliest use of Pre-Qin Chinese. The Pre-Qin vocabulary research falls into the following categories [1]: research on the development of language via vocabulary, such as Tan Shuwang et al. [2]; research on the vocabulary itself, such as Wu Baoan et al. [3]; research on a certain vocabulary of Pre-Qin Chinese, such as Teng Huaying et al. [2]; research on the vocabulary system of Pre-Qin Chinese, such as Yu Wei et al. [12]; and research on certain books, such as Han Feizi (Che Shuya et al. [4]) and Lu Shi Chun Qiu (Zhang Shuangdi et al. [5]). However, these studies are currently limited to the field of ancient Chinese or lexicology, and lack research on the prepositional vocabulary.

Based on the "Great Chinese Dictionary", this paper forms a Pre-Qin word list. Using computer and database technology, we manually mark the existing English-Chinese bilingual resources CCD and build the Pre-Qin ancient Chinese WordNet. This WordNet contains 45,498 Pre-Qin basic words and 63,230 semantic classes. A statistical analysis of the concept semantics of the Pre-Qin vocabulary is then conducted on the basis of the Pre-Qin ancient Chinese WordNet.

2. **Related Work.** Since 1980s, the theory and technology of natural language semantic analysis has been a research hotspot in computational linguistics, and the most important basis of the representation of linguistic knowledge used for semantic analysis is semantic dictionary. WordNet is an English semantic dictionary, developed by Princeton University in the United States [6-7]. It organizes English words into meaning classes and describes more than ten semantic relationships between different semantic classes. It is one of the most important dictionaries in the world.

WordNet is an online dictionary reference system that is driven by current psycholinguistic studies of human vocabulary memory. The traditional dictionary system always organizes entry information by alphabetical order. Some meaning-related words are randomly distributed throughout the system. They ignore the organization of synonymous information in the dictionary system [8]. Therefore, it is very difficult to find similar or related words in such a dictionary system. So we need a convenient and intelligent semantic dictionary, and WordNet can meet our needs. WordNet organizes vocabulary into a synonym set (Synset), where each set indicates a vocabulary concept (Concept), and each vocabulary concept forms semantic relations such as synonym, antisense, and hyponymy to form a relatively complete lexical semantics.

In recent years, research on WordNet abroad has developed rapidly, and there have been many other word networks, such as Hebrew, Latin, Sanskrit and Ancient Greek. Among them, CoreNet is a Chinese-Japanese-Korean word network based on a shared semantic hierarchy [9]. EuroWordNet, as a multi-language database, contains four languages: Dutch, Italian, Spanish and English, and have also been extended to cross-language information retrieval or to compare different word networks in order to study lexical semantic resources and their specificity in German, French, Estonian and Czech [10].

In addition, there are currently a few studies of Pre-Qin vocabulary, but most of them focus on vocabulary research of a specific semantic class. For example, Teng Huaying [11] takes the Pre-Qin Chinese time words as the research topic and intends to create a performance system that contains the Pre-Qin Chinese time-domain category. Yu Wei [12] takes the Pre-Qin Chinese architectural vocabulary as the research object, and it is a generational study of a specific semantic category. Most of these studies focus on the specific semantics of Pre-Qin vocabulary and lack an overview of the Pre-Qin vocabulary system. Moreover, the ancient Chinese dictionary currently constructed is often reinterpreted and is seldom applied, and it is difficult to form a systematic lexical semantic network for people's learning and use.

Nanjing Normal University selects 25 Pre-Qin biographies in accordance with

self-designed word-sharing and labeling specifications (in order of length: Zuo Chuan, Guan, Han Feizi, Lu Shi Chunqiu, Li Ji, Mozi, Xunzi, Mandarin, Yili, Zhuangzi Zhou Li, Gong Yang Chuan, Xunzi Chunqiu, Gu Liangchuan, Mencius, Shijing, Shangshu, Chu Ci, Zhou Yi, Shang Junshu, Analects, Laozi, Sun Tzu, Wu Zi, Xiao Jing, totaling 1.35 million Chinese characters.) to carry out a comprehensive work of word-sharing and word-based calibration, and establishes a Pre-Qin Chinese finishing corpus (Shi Min et al. [13]).

In summary, since there is no vocabulary network mapping Pre-Qin vocabulary to modern English, the construction of such a multi-lingual vocabulary network is of great significance to the study of Pre-Qin vocabulary, the development of Chinese vocabulary concepts, and the comparison of English-Chinese concepts. Therefore, this paper envisages the construction of the Pre-Qin ancient Chinese WordNet and realizes the mutual mapping of Pre-Qin Chinese, modern Chinese and modern English through semantic relations. This will promote the study of Pre-Qin vocabulary in the field of linguistic information processing, and has positive significance for exploring the characteristics of vocabulary, studying the development of vocabulary in ancient and modern times, and language and culture.

3. The Construction of Pre-Qin Ancient Chinese WordNet.

3.1. **Basic Resources.** In previous studies, experts manually create a WordNet to achieve the best results in accuracy, but it is costly. Therefore, there has also been a fully automated or semi-automated method by extracting WordNet between English and target languages. For example, Jordi Atserias, Salvador Climent and others started from the existing vocabulary resources to explore the automatic construction of multilingual vocabulary knowledge base, and finally formed the Spanish WordNet [14~16]. However, due to the scarcity of language resources, these methods cannot be applied to the construction of the Pre-Qin ancient WordNet. Specifically, there is no bilingual dictionary or corpus between Pre-Qin Chinese and English, and there is also no corpus resource between ancient Chinese and modern Chinese that can be directly used to extract words from translation pairs [17]. Therefore, in order to build the trilingual mapping of Pre-Qin Chinese, modern Chinese and English, it is necessary to build a complete database of vocabulary resources. The Pre-Qin ancient Chinese WordNet database constructed in this paper contains two data tables, Chinese Concept Dictionary (CCD) and Pre-Qin word list.

The Chinese Concept Dictionary (CCD) is a Chinese-language semantic dictionary compatible with WordNet developed by the Institute of Computational Linguistics of Peking University [18]. It is based on the WordNet (version 1.6) released in 1997. On the one hand, it inherits the concept of international standard WordNet and its semantic relationship and lexical relationship. On the other hand, CCD is not just a simple Chinese version of WordNet [19]. It has been adjusted according to Chinese characteristics and cultural habits. From the perspective of relational semantics, CCD uses synset to define concepts and relationships between concepts to describe relationships between semantics [20]. This feature distinguishes CCD from other semantic dictionaries and enables it to better target various applications in the field of Chinese information processing.

We have set up 20 fields in the CCD data table, mainly with serial number (CCDID), word of speech (POS), English of target words, synonym collection of modern Chinese (Synset, CSynset), and definition of Chinese and English (Definition/ CDefinition) and so on. In WordNet and CCD, hyponymy between concepts is the main relationship in the structure, and it is attached with other relationships (such as: opposite relationship, partial overall relationship, synonymous antisense relationship, etc.), constituting the whole vocabulary concept network. In the construction of the Pre-Qin ancient Chinese WordNet database, CCD is used as a modern Chinese translation of WordNet and Pre-Qin word list, playing a mediating role in linking Pre-Qin Chinese to English.

TABLE 1. CCD EXAMPLE

Nominal semantic class corresponding to CCD	
OFFSET	*03798428*
POS	*N*
CATEGORY	*07*
SYNSET	*Morality*
CSYNSET	*伦理品德品格品行道德*
DEFINITION	*concern with the distinction between good and evil or right and wrong; right or good conduct*
CDEFINITION	*涉及到好与坏对与错的行为*
HYPONYM	*01484340；01485475；01486456；01835142；01835619*
ANTONYM	*038014990101*
SIMILARTO	*03714294*

TABLE 2. PRE-QIN WORD REPRESENTATION EXAMPLE

Some words that express "morality" in the Pre-Qin lexicon			
ID	*WORDS*	*SENSE*	*EXP*
20222	*道*	*道德，道义。*	*《孟子公孙丑下》：“得道者多助，失道者寡助。”*
20260	*德*	*道德，品德。*	*《易乾》：“君子进德修业。”*
43883	*行*	*品行，德行。*	*《书酒诰》：“天降威，我民用大乱丧德，亦罔非酒惟行。”*
97043	*倫質*	*伦理，人伦道德之理。*	*《逸周书武纪》：“土地未削，人民未散，國權未傾，倫質未移，雖有昏亂之君，國未亡也。”*

The Pre-Qin word list is filtered from the Chinese vocabulary diachronic database. Li Bin et al. [21] manually tagged more than 300,000 entries in the Great Chinese Dictionary and more than 800,000 documentary evidences, finally constructing a Chinese diachronic vocabulary database. We filtered the documentary dynasty in this database and obtained

45,498 entries from the Pre-Qin (~-221), and 63,230 from the Pre-Qin period. On this basis, according to the WordNet framework, vocabulary information is organized to form the Pre-Qin word list. In the Pre-Qin word list, vocabulary is grouped into synonym sets (Synset). Each synset represents a different concept, and is associated with other words through lexical relationships. The Pre-Qin word list contains a total of 12 fields, providing the corresponding IDs of the Pre-Qin vocabulary in the CCD table, modern Chinese interpretation, example sentences and the earliest use cases, and constructs a relatively complete Pre-Qin Chinese vocabulary network, laying the foundation for the later establishment of WordNet.

3.2. **Mapping Methods.** In the procedure of constructing the WordNet, we manually map Pre-Qin words to modern English. In the CCD and Pre-Qin word list, the field CCDID is both added. According to the principle that the same index items are shared by the symmetry classes in different languages, the association between the CCD and the Pre-Qin word list is established, thereby implementing the tri-lingual mapping. The following is an example of the Pre-Qin word "備禦":

TABLE 3. THE PRE-QIN WORD “備禦”

WORDS	*SENSE*	*EXP*	*CCDID*
備禦	*防备*	*《国语周语中》：“將民之與處而離之，將災是備禦而召之，則何以經國？”*	*01483858v*

TABLE 4. CORRESPONDING CHINESE-ENGLISH BILINGUAL RECORDS IN CCD TABLE

CCDID	*SYNSET*	*CSYNSET*	*DEFINITION*	*CDEFINITION*
01483858v	*keep_one's_eyes_skinned keep_one's_eyes_open*	*小心戒备提防留心*	*pay attention; be watchful*	*注意的警惕的*

In the process, some Pre-Qin ancient words cannot be mapped to English. The Pre-Qin word list contains a large number of with distinct characteristics of the times and cultural characteristics. It is difficult to accurately and properly map into the conceptual system of English vocabulary. However, due to "specificity", such vocabulary is of great significance for studying the special lexical semantics of Pre-Qin Dynasty.

4. **Statistical Analyses.** Based on the Pre-Qin ancient Chinese WordNet, we provide a semantic overview of the Pre-Qin vocabulary. In addition, in comparing the concepts in Pre-Qin word list with CCD, we discover some special words and semantic categories in Pre-Qin ancient Chinese that can be used for further exploration of cultural connotation encoded in the language.

4.1. **Basic Data.** According to the previous text, the Pre-Qin word list is based on the selection of the Great Chinese Dictionary. The word-list organizes vocabulary information according to semantic relations, each of which corresponds to a new semantic class that appeared in Pre-Qin Dynasty. Multiple new semantic classes can appear in the same

vocabulary, and a new semantic class can be shared by multiple words. The word-list has 63,230 semantic classes, and we get the result using database tools showing that the list contains a total of 45,498 basic words of Pre-Qin ancient Chinese.

TABLE 5. STATISTICS OF MONOSEMY AND POLYSEMY

NUMBER OF ITEMS	*NUMBER OF TYPES*	*PROPORTION*
1	*38640*	*84.9%*
2	*3936*	*8.7%*
3	*1083*	*2.4%*
4	*519*	*1.1%*
5	*342*	*0.8%*
6	*209*	*0.5%*
>=7	*769*	*1.7%*

TABLE 6. PRE-QIN WORDS WITH THE MOST SENSES(TOP TEN)

TYPE	*NUMBER OF ITEMS*
爲	*46*
發	*39*
辟	*35*
與	*34*
食	*32*
方	*31*
將	*31*
齊	*31*
至	*30*
服	*29*

We then carried out word length statistics on these 45,498 Pre-Qin words. Disyllabic words have a significant advantage over single words because we ignore the frequency information of the use of entries. The Pre-Qin word list contains a total of 35,847 disyllabic words, accounting for about 78.8%, with an average word length of 2.01 characters. The words with a length of 4 or more are idioms. The specifics of the statistics for the length of words are shown in the table below:

TABLE 7. LENGTH STATISTICS OF THE PRE-QIN WORD LIST

WORD LENGTH	*NUMBER OF TYPES*	*PROPORTION*
1	*6209*	*13.6%*
2	*35847*	*78.8%*
3	*672*	*1.5%*
4	*2629*	*5.8%*
>4	*141*	*0.3%*

4.2. **Annotation Data.** The labeling process is the mapping process between Pre-Qin ancient Chinese semantic classes and English semantic classes to achieve two main purposes: (1) to obtain the characteristic words in Pre-Qin ancient Chinese, which can be shown by the unmapped semantic classes (2) to obtain the richness of Pre-Qin semantic classes, which is achieved by linking words representing the same semantic class with the same CCDID in the successfully mapped semantic classes.

4.2.1. **Semantic Class Richness.** We first analyze the common semantic classes that Pre-Qin word list and CCD map to each other. According to the results of the first annotation, the new semantic classes produced by the Pre-Qin vocabulary can be mapped to 45,168 classes of CCD, accounting for about 71.4%, and 18,062 corresponding concepts cannot be found in CCD, accounting for about 28.6%.

This part analyses 45,168 semantic classes that can be successfully mapped, and reflects the richness of the semantic classes of this dynasty by counting the number of words that represent the same semantic class in the Pre-Qin period.Nouns and verbs constitute the most important part of the vocabulary system. This paper classifies nouns and verbs according to the WordNet system, organizes them according to their semantic relations, carries out semantic category richness statistics respectively, and compares them with modern Chinese. On this basis, we get the most different semantic categories between Pre-Qin vocabulary and modern Chinese vocabulary. This will help us to dig into the characteristic Pre-Qin culture reflected by the semantics of vocabulary.

TABLE 8. STATISTICS OF NOMINAL SEMANTIC CLASSES

The top ten richest nominal semantic classes				
ITEM	*ID OF CCD*	*EXAMPLE OF CCD*	*TYPES*	*EXAMPLE OF PQWN*
1	*06188759n*	*官僚官吏官员*	*196*	*六卿 老 公士 外官*
2	*00014887n*	*位置地方场所*	*123*	*州 鄗 次*
3	*04783039n*	*地名*	*96*	*郊 棘 翼*
4	*06299747n*	*国土国家*	*84*	*山川 土疆*
5	*06711088n*	*山*	*77*	*密山 鎮 柞*
6	*06789983n*	*江河河川*	*61*	*河 江 沂*
7	*10536941n*	*宝玉宝石石头*	*58*	*璧 琥 琨*
8	*05452645n*	*灾劫难危难*	*50*	*孽 長蛇*
9	*00795487n*	*善举善行德行*	*48*	*賢能 中德*
10	*03601056n*	*兵器战具武器*	*45*	*器用 備 兵*

Table 8 lists the top ten richest nominal semantic categories of the Pre-Qin ancient Chinese WordNet, and Table 9 is the richness comparison of the corresponding semantic classes between the Pre-Qin ancient Chinese WordNet and modern Chinese. We also select the top ten most different. It should be noted that the most varied in the richness of ancient and modern nouns of the category is "place names", covering all the place names in

Pre-Qin era, that is, in the mapping process, specific place names are dealt with as the upper category. Similarly, the semantic classes ranked 2 to 9 in Table 9 are the upper category of basic-level terms and belong to category words, which are not analyzed in detail in this paper. Considering the richness of the semantic classes in the ancient Chinese WordNet, combined with the significance of the difference between ancient and modern Chinese, we chose to take the semantic class of "兵器" as an example to reveal the Pre-Qin characteristic culture reflected by the vocabulary of the weapon.

TABLE 9. COMPARISON OF NOMINAL SEMANTIC CLASSES

The top ten most different Nominal semantic classes						
ITEM	*ID OF CCD*	*ANCIENT CHINESE*		*MODERN CHINESE*		*ANCIENT/MODERN RATIO*
		TYPES	*EXAMPLE*	*TYPES*	*EXAMPLE*	
1	*04783039n*	*96*	*滑、皇*	*1*	*地名*	*96*
2	*06711088n*	*77*	*羽、陵*	*1*	*山*	*77*
3	*04767588n*	*41*	*安、端*	*1*	*副词*	*41*
4	*06188759n*	*196*	*工、僚*	*6*	*官僚*	*32.7*
5	*02490563n*	*24*	*豆、罍*	*1*	*容器*	*24*
6	*03025339n*	*22*	*鼓、八音*	*1*	*乐器*	*22*
7	*02851548n*	*20*	*辟、成器*	*1*	*工具*	*20*
8	*00014887n*	*123*	*次、墟*	*7*	*位置*	*17.6*
9	*03338120n*	*31*	*絲、良功*	*2*	*丝织品*	*15.5*
10	*03601056n*	*45*	*兵、戈兵*	*3*	*兵器*	*15*

In the Pre-Qin era, the meaning of "兵器" was mapped to 45 words, which far exceeds the number of modern Chinese words (" 兵器 战具 武器"), and the specific words are listed as follows:

TABLE 10. THE PRE-QIN WORDS EXPRESSING 'WEAPON'

備、兵、祋、度、斧、惠、矛、椎、鈹、鉅、鉞、鐫、鑺、兵刃、兵甲、兵弩、兵革、兵械、兵器、凶器、天兵、器用、器用、器械、彫戈、守具、官兵、寢戈、戈甲、戈矛、戈兵、戈盾、戈戟、戈劍、戎事、戎昭、戎器、武車、武翼、砥刃、矢石、短兵、精鋭、良兵、鋒刃

"The great events of the nation is military." The Pre-Qin period is an important stage in the history of China with great social changes and frequent war activities [22]. In the Pre-Qin period of about three thousand years, there were more than 800 wars recorded in the historical books. Frequent and fierce wars became the fundamental driving force to promote the development of weapons, and weapons in national strength also occupied a more and more prominent position. Weapons separated from tools of production and developed from stone weapons to bronze weapons and iron weapons, producing " 鐫 " which can be used for wood-breaking, "鈹" shaped like swords, cane weapon "祋" and so on [23]. It can be seen that the increase in the status and variety of weapon are reflected in

the Pre-Qin vocabulary system, which makes words expressing "weapons" richer and richer. This will support the study of the Pre-Qin military.

Previous studies on weapons vocabulary are mostly research on specialized books, included in the classification study of "military vocabulary". They often take the way of Exegetics, paying attention to the investigation of military vocabulary's classification and significance. For example, Luo Weilei [24] classified the military words in "Zuo Zhuang", and mainly examined the etymological meanings of some ancient weapons words; Zhang Xiangyou [25] expounded the categories and characteristics of military words in "Shuo Wen Jie Zi", not only analyzing the meanings of words, but also exploring the military culture, the level of the development of ancient productivity, and the ancient etiquette culture behind words. By constructing the Pre-Qin ancient Chinese WordNet , we can carry out systematic research on the Pre-Qin weapon vocabulary and reveal the Pre-Qin military culture more comprehensively.

TABLE 11. STATISTICS OF VERBAL SEMANTIC CLASSES

The top ten richest nominal semantic classes				
ITEM	*ID OF CCD*	*EXAMPLE OF CCD*	*TYPES*	*EXAMPLE OF PQWN*
1	*00109250v*	*摞积丛集*	*100*	*積 併 柴 崇*
2	*01643863v*	*协助帮助援助*	*76*	*比 庳 弼 承*
3	*01656373v*	*管辖治理*	*56*	*徹 爲 艾 比*
4	*01703421v*	*治罚处治*	*49*	*兩 創 濟 服*
5	*01547388v*	*赏奖励奖赏*	*45*	*慶 敘 封 疏*
6	*00250254v*	*死去世殉难*	*40*	*吪 喪 殭 憚*
7	*00244173v*	*遵守遵循*	*39*	*發 律 守 修*
8	*01231606v*	*念思念想念*	*39*	*憶 存 服 維*
9	*01215448v*	*怕害怕恐惧*	*37*	*惡 憂 恐 聳*
10	*00176836v*	*伤害伤耗损伤*	*35*	*墜 敗 椓 斃*

Table 11 lists the top ten richest verbal semantic categories, from which we can catch a glimpse of the political culture, reward and punishment system, the view of death, social stability and many other aspects of the Pre-Qin era. Combined with the differences in Table 13 in the richness of semantic classes, we chose to take the semantic class "遵守、遵循" as an example, and analyzed the political thought environment of the Pre-Qin Dynasty reflected by the corresponding Pre-Qin words. The specific terms are listed below:

TABLE 12. THE PRE-QIN WORDS EXPRESSING 'OBSERVE'

發、律、守、修、循、軌、遹、鉛、類、修道、允迪、率由、率法、率常、率道、勖帥、帥意、循牆、守禮、安制、安節、宗原、敵國、恪守、惇帥、慎法、懷刑、由義、由豫、由禮、畜道、踐期、踐繩、道法、遵道、遵繩、誠士、順心、順理

As can be seen from the examples provided by the Pre-Qin word list, this semantic class is often used to express "observance of etiquette, compliance with national law, and adherence to morality", such as “發” in the sentence “君臣上下貴賤皆發焉”, “軌” in “心不畏時之禁，行不軌時之法，此大亂之道也”, “由義” in “吾身不能居仁由義，謂之自棄也”, etc. To some extent, this reflects the political thought and different doctrines at that time upheld different concepts of governance: Confucianism advocates "etiquette and music" and adheres to etiquette education; the legalist advocates the rule of law as the core thought to govern the country and requires people to follow the law. The unique political and cultural factors of the Pre-Qin are reflected in the vocabulary, which makes the Pre-Qin semantic category of "遵循" richer than modern Chinese. Consistent with the previous text, this example also starts with semantic richness, providing evidence for the study of the culture behind vocabulary.

"A social language can reflect its corresponding culture, and one of its ways is in terms of vocabulary. [26]" There are many previous studies on lexical culture, such as "Chinese Vocabulary and Culture", written by Mr. Chang Jingyu [27]. On the basis of analyzing a large number of corpora, the cultural classification of Chinese vocabulary is carried out based on the cultural factors of Chinese, and the cultural meaning of the words is discussed from many angles in this book. However, the traditional study of vocabulary culture is mostly limited to example description and analysis, which cannot be quantitative. From the perspective of semantic classes, this paper explores the hidden culture of the Pre-Qin vocabulary according to semantic richness, and makes a more systematic and comprehensive study of lexical culture in a quantitative way.

TABLE 13. COMPARISON OF VERBAL SEMANTIC CLASSES

The top ten most different Verbal semantic classes						
ITEM	*ID OF CCD*	*ANCIENT CHINESE*		*MODERN CHINESE*		*ANCIENT/MODERN RATIO*
		TYPES	*EXAMPLE*	*TYPES*	*EXAMPLE*	
1	*00244173v*	*39*	*律, 軌*	*2*	*遵守*	*19.5*
2	*00240992v*	*15*	*摯, 極*	*1*	*到达*	*15*
3	*00929407v*	*13*	*傅, 附*	*1*	*依附*	*13*
4	*01643863v*	*76*	*俌, 夾*	*6*	*协助*	*12.7*
5	*00073088v*	*12*	*慬, 時*	*1*	*停止*	*12*
6	*00323209v*	*20*	*斂, 裁*	*2*	*抑制*	*10*
7	*01617687v*	*20*	*矢, 式*	*2*	*履行*	*10*
8	*00069684v*	*19*	*災, 痛*	*2*	*危害*	*9.5*
9	*01656373v*	*56*	*徵, 艾*	*6*	*治理*	*9.3*
10	*00415041v*	*28*	*廢, 出*	*3*	*丢弃*	*9.3*

4.2.2. **Unmapped Words.** Then, we discuss the Pre-Qin vocabulary that is difficult to map accurately or cannot map to CCD. According to the results of the first annotation, there are 18,062 Pre-Qin words that cannot find corresponding concepts in CCD, accounting for about 28.6%. Most of these words are historical vocabulary and classical Chinese vocabulary such as some examples provided in Table 14. They indicate the unique semantic

concepts in Pre-Qin Dynasty and we cannot find corresponding words in modern Chinese. Besides, there are some special words, with the unique color of Pre-Qin ancient Chinese. In the following table they are classified as small categories for analysis.

TABLE 14. EXAMPLES OF UNMAPPED WORDS

Pre-Qin WORD	*CHINESE DEFINITION*
來	*特指已嫁女子回娘家省亲。*
則	*古指二百平方里以下的采邑。*
匱	*古代无盖斜口盛粮食的竹器。*
倮	*单衣无甲者。*
兩	*古代军队编制单位，二十五人为一两。*
丱	*古时儿童束发成两角的样子。*
儺	*古代的一种风俗，迎神以驱逐疫鬼。*

There are many special lexical semantic phenomena in the Pre-Qin word list, and the following three categories are illustrated by examples.

(1) Antonomous Compound Word

Contradiction is universal, and language is no exception. In the Pre-Qin vocabulary, there are many “positive and negative integration” phenomena at the level of combination, such as the word "信誕", whose morphemes meanings are opposite, but can coexist in the same language unit. Traditional scholars call it anantony compound words. The “positive and negative integration” in this paper refers to the linguistic phenomenon that two morphemes with opposite meanings are combined in the same Chinese language unit (the unit in this paper is limited to words) [28]. This constitutes a language unit, but there are obvious contradictions outside. Table 15 lists some examples of “antonomous compound word” in the Pre-Qin vocabulary.

TABLE 15. EXAMPLES OF ANTONOMOUS COMPOUND WORD

INTEGRAL WORDS OF POSITIVE AND NEGATIVE	*DEFINITION*
信誕	*诚实和欺诈。*
謗譽	*毁谤和称誉。*
背向	*背对和面向。*
表裏	*表面和内部。*
冰炭	*冰块和炭火。*
步驟	*缓行和疾走。*
傳習	*传授和学习。*
辭受	*推辞和接受。*
敵與	*敌国和盟国。*
俯仰	*低头和抬头。*
寒暍	*寒冻和暴热。*
鴻殺	*粗大和细小。*
雨暘	*雨天和晴天。*
與奪	*奖励和惩罚。*

This special lexical phenomenon is rare and mostly exists in Chinese or some Indo-European languages. The phenomenon of positive and negative integration in modern Chinese has been studied before, but the study is shallow and does not involve the vocabulary field of ancient Chinese. In the construction of the Pre-Qin ancient Chinese WordNet, we tried to map these words to English, but found it very difficult. The antonym compound words reflect people's dialectical understanding of things, and there are differences in the understanding of things among different language users, because of which, the " Integral Word of Positive and Negative " that can correspond to each other in the two languages are extremely rare.

(2) Integral Word of Whole And Part

In Pre-Qin ancient Chinese, especially in single words, a word can represent the whole of an object and also the different parts of the object. For example, one of the righteous items of the Pre-Qin word " 犀 " is " animal name, commonly known as rhinoceros", but it can also refer to rhino skin, rhino horn; the word "屋" can refer to "house" and "roof"; "畝" is not only a plot unit that generally refers to farmland or fields, but can also refer to ridges.This phenomena may have something to do with metonymy rhetoric in ancient Chinese. In the process of mapping to English, the difficulties are mainly due to the fact that some parts are divided too carefully. There are no independent words in English to express them. People often use phrases to express them. For example, the word "犀牛角" is often translated into "rhinoceros horn" in English.

(3) Monosyllabic Words with Implicit Semantic Degree

Among the monosyllabic words of Pre-Qin Chinese, there are some words with very complex meanings. They not only have core concepts, but also imply different degrees of information. We take the word "沃" appearing in the sentence "啓乃心，沃朕心。若藥弗瞑眩，厥疾弗瘳。" as an example. The explanation for it in the Great Chinese Dictionary is "sincere advice", but if mapped to English vocabulary, this word can only correspond to "advise, counsel". It can be seen that the semantic degree of the two do not match. This is also seen in the Pre-Qin word "效", which means " serve with all one's heart and soul ", but the mapping of the English word is "serve"; the Pre-Qin word "褋" means "overlapping and densely stacked together", however, its corresponding English word is "pile-up". All of these Pre-Qin words map to the English concepts in an inappropriate degree.

As mentioned earlier, polysyllabic words, especially disyllabic words in Pre-Qin Chinese, far outnumber single words in types but are far less than single words in tokens. Therefore, ancient Chinese, especially Pre-Qin ancient Chinese, is still monosyllabic words-based, occupying an absolute advantage. Ancient Chinese is characterized by "simple refinement, semantic accuracy", and monosyllabic words often need to contain rich information in order to express meanings accurately in communication.

5. **Conclusion.** Based on the "Great Chinese Dictionary", we build the Pre-Qin word list. Using the computer and database technology, we manually mark the existing English-Chinese bilingual resources CCD and construct the Pre-Qin ancient Chinese WordNet. Main tasks are as follows:

(1) We explain the construction process of the Pre-Qin ancient Chinese WordNet. Based on the Great Chinese Dictionary, we have established a Pre-Qin word list containing 63,230 meanings, and mapped them to the CCD table by manual labeling. Among them, there are 45168 Pre-Qin semantic classes mapped into English, accounting for 71.4%; but there are 18062 that do not have corresponding semantic classes in CCD, accounting for 28.6%.

(2) Based on the established Pre-Qin word list, we elaborate the outline of the Pre-Qin vocabulary. With the help of database tools, there are 38,488 lexical lexicons in the vocabulary, accounting for 84.9%, and 6858 polysemous words, accounting for 15.1%. In regard to word length, the vocabulary contains 35,847 double words, accounting for 78.8%, and the average word length is 2.01.

(3) Based on the manual mapping, we analyze the Pre-Qin ancient Chinese WordNet in terms of semantic richness. Firstly, the richness statistics of the semantic classes successfully mapped with CCD are analyzed. On the basis of this, the semantic richness of Chinese is compared chronologically, and the Pre-Qin culture reflected by the difference is mined. Secondly, we also analyze the unclassified semantic class.

In the future, we will continue to proofread the existing corpus to improve the accuracy of the annotation. Second, we will conduct a detailed analysis of the Pre-Qin semantic classes which has failed to be mapped to the CCD and research the characteristic of the Pre-Qin vocabulary. Third, in comparison with WordNet in other languages in the world, we will do researches from the perspective of national culture and cognitive psychology. Fourth, we will use database technology to visualize the Pre-Qin ancient Chinese WordNet and establish an online retrieval platform for the public.

Acknowledgment. This work is partially supported by National Social Science Foundation (18BYY127), Construction Project of Superiority Disciplines in Jiangsu Universities, and Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences project (2017STD006). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Liu Liu. The Automatic Acquisition and Application of the Characteristics of the Chinese Word[D]. Nanjing Normal University, 2014.

[2] Tan Shuwang. The Development of Chinese from the Warring States to the Eastern Han Dynasty According to Mencins Zhangju[J]. Research In Ancient Chinese Language,2001(02):62-66.

[3] Wu Baoan. The study of core words in Han Dynasty[D]. The Huazhong University of Science and Technology, 2007.

[4] Che Shuya. Han Feizi Vocabulary Research[D]. Barwon Book Club, 2004.

[5] Zhang Shuangdi. The Brief Comment of Vocabulary in Lu Buwei[J]. Journal of Peking University, 1989(05):58-68.

[6] Miller.G, Beckwith.R, Fellbaum.C, Gross.D. Introduction to Wordnet: An Online Lexical Database in Fie Paperson Wordnet[J]. CLS report, Cognitive Science Laboratory, Princeton University, 1993.

[7] Miller G. Psychology and communication: In Communication,Language,and Meaning. Basic Books, NewYork, 1973.
[8] Yao Tianshun, Zhang Li, Gao Zhu. Overview about WordNet[J]. Language and Character Application, 2001(01):27-32.
[9] Choi K S, Bae H. A Korean-Japanese-Chinese Aligned Wordnet with Share Semantic Hierarchy[C]. International Conference on Digital Libraries: Technology & Management of Indigenous Knowledge for Global Access. DBLP, 2003.
[10] Vossen P. Introduction to EuroWordNet[J]. Computers and the Humanities, 1998, 32(2-3):73-89.
[11] Teng Huaying. Pre-Qin Chinese Time Vocabulary System Research[J]. Language Construction, 2015(35):81-82.
[12] Yu Wei. The Study of architectural vocabulary in Pre-Qin ancient Chinese[D]. Northeast Normal University, 2018.
[13] Shi Min, Li Bin, Chen Xiaohe. CRF-based Pre-Qin Chinese Word-marking Integration Study[J]. Chinese Informatics Journal, 2010, 2(24):39-45.
[14] X Farreres, G Rigau, H Rodriguez. Using WordNet for Building WordNets, 1998.
[15] Barbu E, Mititelu V. Automatic Building of Wordnets. Proceedings of recent advances in natural language processing IV (pp. 217–226), 2007.
[16] Atserias J, Climent S, Farreres X, et al. Combining Multiple Methods for the Automatic Construction of Multilingual WordNets[J], 1997.
[17] Zhang Y, Li B, Dai X. PQAC-WN: constructing a wordnet for Pre-Qin ancient Chinese[J]. Language Resources & Evaluation, 2017, 51(2):1-21.
[18] Yu Jiangsheng, Yu Shiwen. The Structure of Chinese Concept Dictionary[J]. Journal of Chinese Information Science, 2002(04):12-20+44.
[19] Liu Yang, Yu Jiangsheng, Yu Shiwen. CCD Construction Model and VACOL Auxiliary Software Design and Implementation[C]. Applied Linguistics, 2003.
[20] Jiu Hongying, Yu Shiwen. CCD and its Application[J]. Journal of Guangxi Normal University, 2003(01):98-103.
[21] Li Bin, Liu Xueyang. Study on the Time-evolving Measurement of Chinese Vocabulary Based on the Chinese Dictionary[J]. Journal of Nanjing Normal University (Social Science edition), 2018(05):152-160.
[22] Huang Pumin. The general Situation of the Development of Military Thought in the Pre-Qin Dynasty and its Characteristics[J]. Journal of Jinan University (Social Science edition), 2000(04):1-8.
[23] Zheng Wenchao. The Development and Contemporary Value of War Views in Pre-Qin[J]. Journal of Xi'an Political College, 2013, 26(01):119-123.
[24] Luo Beilei. A Study on Military Words in ZuoZhuan[D]. Guangxi Normal University,2004.
[25] Zhang Xiangyou. The Study of Military Vocabulary in Shuo Wen Jie Zi[D]. Xinan University,2007.
[26] Carol R.Ember, Melvin R.Ember. Cultural Anthropology[M]. Liaoning Publishers, 1988.
[27] Chang Jingyu. Chinese Vocabulary and Culture[M]. Peking University Press, 1995.
[28] Yuan Guang. The Study of Positive and Negative Integration in Chinese[D]. Guangxi University, 2011.