A Corpus-Based Study of N_1-N_2 Words in Archaic Chinese

By

Jane Chanell

17970091

Supervisor: Associate Professor
Robert Mailhammer

Western Sydney University
School of Humanities and Communication Arts

June 2019

A thesis submitted in partial fulfilments of the requirements for the degree of
Master of Research

Acknowledgements

I would like to express my sincere gratitude to my supervisor Associate Professor Robert Mailhammer, for his whole-hearted support of my Master's study and research, for his patience, immense knowledge, invaluable assistance and for providing valuable guidance along every step of the way. Without his unwavering support and constant encouragement, I would never have been able to complete this thesis.

I would also like to thank Dr Chong Han for being kind enough to provide details of the Chinese online corpus used for this linguistic research. A special thank you to Weicong Li, for providing all the essential technical support for working with the online corpus.

I am very grateful to Dr Geoff Hyde for his invaluable advice on my writing.

STATEMENT OFAUTHENTICATION

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Jane Chanell

(Name)

(Signature)

20/06/2019

(Date)

Table of Contents

List of Tables

List of Figures

Abstract

In Modern Chinese, the majority of words with a combination of two characters have been identified as compounds. However, the general consensus is that compounds or disyllabic words did not exist in early Archaic Chinese (before 220 BC). While some previous research has reported the occurrence of disyllabic words or compounds, the only compounds identified in Archaic Chinese were proper nouns and reduplicative words. The aim of this study is to investigate more thoroughly the origin of disyllabic words in the history of Mandarin. It focuses, in particular, on the nominal combination of two nouns (N_1-N_2), the most frequently occurring and highly productive combination in later periods. The research adopts a corpus-based approach to analyse a sample of texts spanning a period of over 3000 years. The findings show that nominal words with N_1-N_2 sequences originated in early Archaic Chinese, and these N_1-N_2 words were structurally formed using a range of linguistic rules. The occurrence of words in a set of nominal N_1-N_2 words (identified in the early Archaic period) decreased during later periods, probably a result of the uneven distribution of genres in the corpus, and changes in language use due to transformations of cultural and political systems. The main conclusions drawn from this research are that compounding was not only the consequence of the phonological simplification of the Medieval period, and that other types of compounds, in addition to proper nouns and reduplicates, occurred in Archaic Chinese. Further research is recommended to assess the constituents of compounds phonologically, morphologically, and semantically to better understand the order of sequences of the constituents of compounds and the historical disappearance of some compounds.

1. Introduction

Words can be categorised into simple and complex words; generally, a complex word may consist of a root (base or stem) and one or more affixes (e.g., *believable* and *unbelievable*), or more than one root in a compound (e.g., *policeman* and *son-in-law*). Complex words with two syllables, commonly termed disyllabic words, are quite common in Mandarin, the main language of China (Chao 1968, Duanmu 2007, Li & Thompson 1981, Packard 2000). While complex words can be of many types, such as noun-noun (N_1-N_2), adjective-adjective (A_1-A_2), verb-verb (V_1-V_2), adjective-noun (A-N), the N_1-N_2 form is the most frequently occurring in Mandarin and, linguistically, is considered to be highly productive (Chao1968, Huang 1998, Li & Thompson 1981:53, Qin & Duanmu 2019). As such, compounds of the N_1-N_2 form would be the most likely form of disyllabic words to be discoverable in the earliest records of the language.

N_1-N_2 words in Mandarin are made of two monosyllabic nouns, such as *yi-fu* 'clothes' and *cha-bei* 'teacup', and are derived through morphological processes, including compounding and affixation. In general, compounding and affixation are seen as distinct processes. A number of complex words with two syllables are borrowed from other languages, e.g., *ka-fei* 'coffee' and *sha-la* 'salad'. Words formed by compounding and derivation are the main focus of this study due to the purpose of the study, which will be discussed below. Compounds and derivatives are sub-categories of disyllabic words and these three terms will be used throughout the study.

Compounding in Mandarin has been considered as a core attribute of the language, and the use of compounding has increased during the language's long history (Duanmu 2007, Li 1993, Li 1981, Li 2013, Karlgren 1926, Pulleyblank 2000, Ting 2002, Wang 1957, Wu 2001). As such, compounding has consistently been a major focus in studies of the evolution of the language.

Research on morphological processing in Archaic Chinese, and in particular, on the issue of how compounds were represented in the lexicon, has led to significant differences among scholars. Some research has asserted that compounding came into existence in Medieval Chinese due to the pressure of phonological changes in the Medieval Chinese period that

resulted in a simplification of the syllable structures used in Archaic Chinese (Chao 1968, Karlgren 1926, Li 1993, Li & Thompson 1981, Pulleyblank 2000, Tai & Chan 1999, Ting 2002, Wang 1957). The strongest position is that held by Karlgren (1926) and Wang (1957) who argued that the phonological simplification was the only driver of the emergence and rise of compounding, which they proposed only began after the Archaic Chinese period. In this model, the phonological changes resulted in words that were distinguishable in Archaic Chinese being converted to homophones in Medieval Chinese. The presence of an enormous number of homophones, and the resulting hampered communication, acted as forces that promoted the use of compounds. In Karlgren (1926) and Wang's (1957)view, if a language has a rich sound system, as did Archaic Chinese, then compounding will not occur; the presence of many homophones is a prerequisite for compounding. However, from other work, it is clear that homophones did not suddenly emerge, for the first time, in Medieval Chinese, as they also occurred in Archaic Chinese (Schuessler 2006, Feng 1998). For example, 鹽 **rjams* 'to salt' and 豔 **rjams* 'beautiful' were phonologically the same, yet they represented two different morphemes or characters in early Archaic Chinese. The existence of homophones in Archaic Chinese is problematic to Karlgren (1926) and Wang's (1957) proposals, because if homophones were only the result of an historically-documented simplification event, then homophones should not exist in Archaic Chinese. Likewise, if compounding always occurs when a language has many homophonous words (Karlgren 1926, Li 1993, Wang 1957), then the existence of homophonous words in Archaic Chinese suggests the possibility of compound words being used at this time.

Li (1993) proposed a cause-effect relationship between phonological simplification and the increasing use of compounds. Li's assertion was based on matching his own findings, regarding the increased use of compounds across Chinese history with Wang's work that showed a reduction in the sound inventory across similar periods. Nevertheless, other work shows that phonological simplification affected not only simple words in Archaic Chinese, but also complex words formed through complete reduplication of syllables. For instance, the initial consonant cluster **hr*- in the Archaic word 赫赫 ***hrak-hrak*** 'illustrious' was reduced to a single consonant *h*- in *hak-hak* in the transition to Medieval Chinese (Schuessler 2006:253, Wu 2001). The proposed cause-effect relationship cannot be assumed to be the only valid explanation of the origins of compounding even if sound simplification, and increased homophony, does accelerate the process.

Pulleyblank (2000) suggested that, because of its rich sound system, in Archaic Chinese, new words were derived only by changing consonants or vowels within monosyllables, rather than by compounding. It is true that some words were derived through these processes, referred to, linguistically, as non-concatenative operations (Haspelmath & Sims 2010:34). For example, 傳 **dron* 'to transmit' was derived from 轉 **tron* 'to turn around' by transforming the initial voice from a voiceless **-t* to a voiced **-d* (Pulleyblank 2000:36). However, it does not follow that non-concatenative operations were only the means by which new words are formed in Archaic Chinese. Words can also be derived through derivational processes.

Indeed, while it may be true that a rich sound inventory could mean that simple words were sufficient for a language, with new words derived through processes within monosyllables. Nevertheless, these possibilities do not rule out the existence and contribution of compounding processes.

Indeed, some researchers, who have begun looking into the origin of disyllabic words, believe that compounding existed long before Medieval Chinese (Feng 1998, Jin 2017, Kennedy 1951, Tao 1996, Wu 2001, Yu 1990). Studying the Shang period (1675-1029 BC), Tang (2007) argued that some proper nouns were formed by the combination of characters, and can be considered disyllabic or polysyllabic words. Even though they were written together, and occupied a single character's space, such words semantically referred to a completed concept and phonologically maintained the original sounds of their component characters; each character in the combination needed to be pronounced. For example, 𓏲𓇋 *wu-yue* 'May' consists of two characters 𓏲 *wu-* 'five' and 𓇋 *-yue* 'moon'; it needs to be pronounced as two syllables, that is, as *wu-yue,* and its meaning refers to a particular month of a year. These proper nouns are therefore thought of disyllabic or polysyllabic words, by virtue of their meaning and the phonological rules used in their formation (Lei 2009:9, Li 2008:196, Pan, as cited in Wu 2001:78, Yu 1990:89).

Based on Tang's (2007) observations, Wu (2001) examined ancient books and records that were written in the Archaic Chinese period. He showed that the majority of disyllabic or compounding words were formed through reduplication of syllables, e.g., 格格 *ge-ge* 'princess', 夜夜 *ye-ye*, nightly', and 人人 *ren-ren* 'people'. He thus proposed that reduplication of syllables was the main process to form disyllables in Archaic Chinese.

Even though Wu (2001) had reported more categories of disyllabic words in Archaic Chinese (in addition to proper nouns), reduplication is considered a word formation process of limited productivity. His inability to detect processes of greater productivity could, however, be a consequence of his approach, in that his sources were limited, his data was collected manually, and there are questions about the criteria he used to distinguish words from phrases. Both Tang (2007) and Wu (2001) suggested the need for further investigation into disyllabic words, to clarify whether other types of processes, apart from proper noun formation and syllabic reduplication, could be involved.

Given that the origins of disyllabic words in Mandarin remain essentially under-explored**,** the purpose of this study is to provide the first digital corpus-based investigation of the origins of compound words, excluding complete reduplicative words, in early Archaic Chinese (before 220 BC). Its particular focus is the origin and diachronic development of disyllabic N_1-N_2 words and patterns.

The research questions are:

Q1 Did disyllabic nominal words with N_1-N_2 sequences originate in Archaic Chinese? If so:

Q2 What patterns of N_1-N_2 words existed in this period?

Q3 If a set of nominal N_1-N_2 words can be identified in early Archaic Chinese, what were their frequencies and patterns when they were traced diachronically throughout Chinese history?

With respect to the origins of disyllabic words, this study predicted that disyllabic N_1-N_2 words originated in Archaic Chinese. If this prediction were true, the study inferred that disyllabic N_1-N_2 words were structurally formed using a range of linguistic rules, and that this set of nominal N_1-N_2 words and their patterns increased in frequency from Archaic Chinese to Modern Chinese.

The Sheffield Corpus of Chinese (SCC) is the main source for this study. The large variety of its historical textual records and its wide coverage of the time periods of Chinese history (Hu & McLaughlin 2007), mean that data from it can be used as quantifiable evidence for drawing generalisations about language use (Kennedy 2014:1-7). Its digital annotations, and integrated search system, facilitate studies of this kind, even if some steps (e.g., the classification of N_1-N_2 compounds as words or phrases) are best done manually.

This thesis is structured as follows:

Chapter 2, *Review of Related Literature*, discusses Mandarin being used throughout the history and the division of Chinese history into periods. It then examines the discrepancy between characters and words in Mandarin. Next it provides a comprehensive review of the origins of disyllabic words, covering such topics as the sound simplification being the sole driver of compounding; research into the simple forms of compounds in ancient Chinese (proper nouns and reduplicated words). The criteria for distinguishing words and phrases are discussed next. The deficiencies of our current knowledge, as revealed in these discussions, is shown to justify the current project, and demonstrates that an examination of N_1-N_2 nominal compounds is a promising approach to help expand our knowledge.

Chapter 3, *Research Methods*, discusses and justifies the research strategy. It introduces the corpora adopted in this study, and the criteria for identifying compounds in this study. It then discusses potential problems related to data collection and explains how the data were processed before being analysed to respond to the research questions.

Chapter 4, *Results*, reports on the results from the data analysis in relation to the research questions. It presents the results of the frequencies of N_1-N_2 phrases and N_1-N_2 disyllabic words, and the frequencies of N_1-N_2 derivatives and N_1-N_2 compounds in early Archaic Chinese. It then reports the findings on the diachronic development of a set of N_1-N_2 disyllabic words identified in early Archaic Chinese, and their patterns.

Chapter 5, *Discussion*, provides a detailed discussion of the results and explanations for the various phenomena related to the origins of N_1-N_2 disyllabic words and the word formation processes involved. The issues or observations that were unanticipated or fall beyond the consideration of the current study are also discussed and interpreted.

Chapter 6, *Conclusion*, revisits the overall aim and research questions of this study. It contains a summary of the findings of the data analysis, followed by a discussion of the implications of the findings and the limitations of this study. Finally, it provides suggestions for further research on the origin and development of N_1-N_2 words.

2. Review of Related Literature

This chapter aims to provide a comprehensive review of the origin of disyllabic words in the history of Mandarin Chinese. It first discusses the use of the Mandarin language, past and present, and the discrepancy between character and word in Mandarin Chinese. Next, it reviews the history of disyllabic words in Mandarin, focusing largely on research into the increasing use of compounds, and the hypotheses associated with this phenomenon. In particular, it examines various proposals that posit the phonological simplification of the Medieval period as the main driver of compounding. Lastly, research regarding the criteria for identifying Mandarin compounds is examined.

2.1 The use of Mandarin throughout Chinese history

This study focuses on the origin and historical development of disyllabic words in Mandarin. Before approaching the core issues, it is useful to examine the pivotal role of Mandarin as the standard spoken language used throughout China's history.

Mandarin is one of the major languages of the world and is the most important of the seven main varieties of Chinese, in terms of the population of its speakers and the geographical spread of its usage. The other varieties of Chinese (Wu, Gan, Xiang, Hakka, Yue and Min) are shown in Figure 1 (Ramsey 1987:87). Mandarin speakers currently cover by far the greatest geographical area, including much of northern, southern and western China, including several provinces and cities, as depicted in Figure 2. This area also includes the Central Plain, where the majority of the Chinese population lived before the 19th century (Zhang 2010, Zhang 1994:51). The Mandarin used in the Central Plain is comprised of Northern and Southern Mandarin (Luo & Mei 2004:295); and there are further subcategories for each of these two varieties, based on the pronunciation used in a particular city or province, for example, Beijing (Peking) Mandarin and Nanjing Mandarin (Dai 2017:40, Zhang 2010, Geng1992).

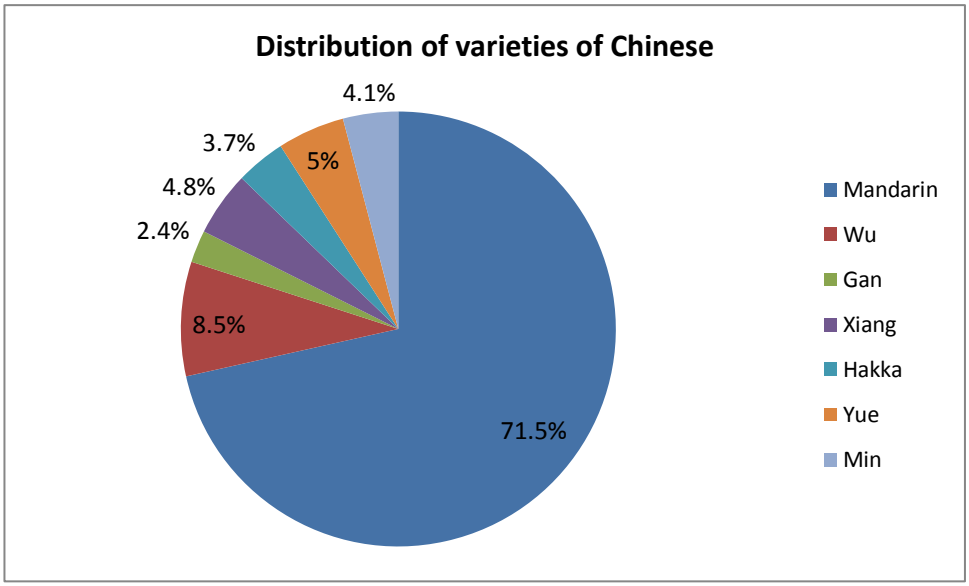

Figure 1: Usage of varieties of Chinese languages (data from Ramsey 1987:87)

Throughout China's history, Mandarin has always been perceived and used as a national lingua franca or a koine language (Mair 1994, Norman 1988). In contemporary China, standard Mandarin, which is one of its varieties, is used as the official language by all speakers. Earlier, however, during the Ming and Qing dynasties (1368 AD – 1911 AD), knowledge of Mandarin was only required if one wished to have a successful career within the empire (Mair 1994:728-729). While the Ming and Qing dynasties saw Mandarin used for the first time as the national spoken form for governance (Coblin 2000:537, Yang 1986, Mair 1994:728), in earlier times ,Mandarin to was referred by terms (*Putonghua*, *Guan-hua*, *Yayin)* that indicate it was already a koine language (Dai 2017: 29 & 34, Mair 1994:728-730).

Mandarin's usefulness as a koine language was underpinned by several features. Firstly, as mentioned above, it had a wide geographical spread (Figure 2), particularly on the Central Plain, and even though the regional dialects had slightly different phonologies, they were similar enough for people from different areas to converse in a mutually intelligible manner. Unlike languages that use a phonographic writing system (such as English), Mandarin is logographically based, and this means the written language is not directly linked to pronunciation. A separation between the written and spoken forms of a language provides great advantages for communication amongst people who use different dialects: They can use the same written form. Indeed, most of the characters used in Archaic Chinese are still readable by speakers of Modern Chinese. In China, northern and southern Mandarin

speakers can communicate to a significant extent in writing, using either *Wenyanwen*[1] or *Baihuawen*[2].

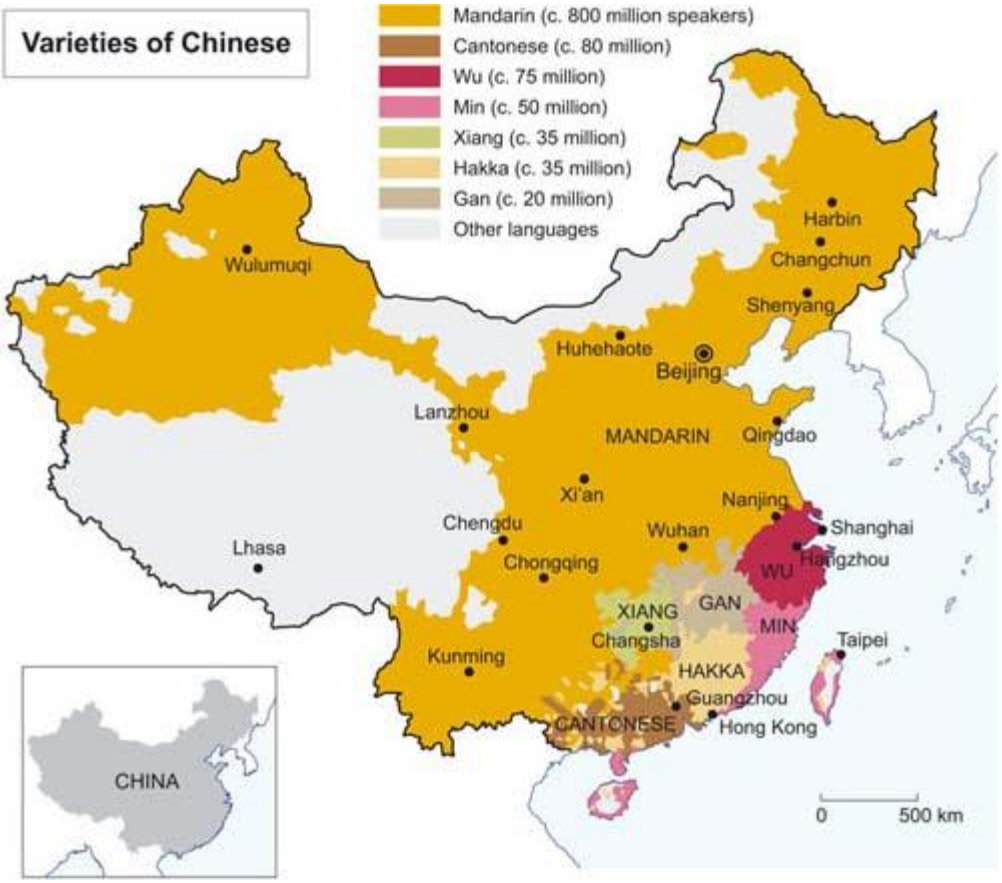

Figure 2: Distribution map of the seven major of varieties of Chinese
(The Open University, as cited in Chinese dialects in China, 2009)

Secondly, as early as the Xia dynasty (2146 BC - 1675 BC), the capital cities for most dynasties have always been established within the Central Plain, where the varieties of Mandarin or *Guan-hua* have served as a lingua franca (Twitchett, Fairbank & Franke 1994). There is a traditional and collective Chinese phrase called *Zhong-guo-si-da-gu-du* 中国四大古都 'Four Great Ancient Capitals of China': they are Beijing, Nanjing, Luoyang and Xi'an (Chang'an) (Mookherjee 2013). Accordingly, Nanjing Mandarin was used as the national lingua franca when the first capital of the Ming dynasty (AD 1368 – AD 1644) was set up within Nanjing city (Luo & Mei 2004, Yang 1986), while, at other times, Peking Mandarin was the spoken form used across the empire when Beijing was the capital, for example, in both the Yuan (1206 AD – 1368 AD) and Qing (1644 AD – 1911 AD) dynasties (Luo & Mei 2004, Dai 2017).

The discussion has already broached the practice of describing Chinese history in terms of 'periods', for example, with respect to the ruling dynasties of the time. Linguistic studies of the history of Mandarin have adopted a range of approaches in setting up the subdivisions within what is, by some estimations, almost a 4000-year timeline. Given the focus of the current project, it is important to evaluate these different approaches, particularly with regard to the intentions of the particular studies.

The time-points were chosen as divisional markers for Mandarin's timeline, and the terms used for each sub-division, have depended, at least in part, on the purpose of the study and/or the major linguistic changes under consideration (e.g., phonological, grammatical and/or lexical). But, even similar studies of the Chinese lexicon have sometimes adopted different approaches. For example, the studies conducted by Pan (1989) and Hu and McLaughlin (2007) used different periodisations. While Pan's (1989) approach focused on important periods in history (Table 1), Hu and McLaughlin's (2007) study of the Sheffield Corpus of Chinese (SCC) was based on lexical changes and posited three major periods and seven further subdivisions (Table 2).

Table 1: Pan's (1989) division of Mandarin's timeline, based on historical periods

Period	**Time frame**
Old Chinese	1675 BC – 207 BC
Middle Chinese	206 BC – AD 580
Pre-Modern Chinese	AD 581 – AD 1919
Modern Chinese	AD 1919 – present

Table 2: Hu and McLaughlin's (2007) division, based on lexical changes

Period	**Name of dynasty**	**Time frame**
Archaic Chinese 1200 BC – AD 220	Pre-Qin	1200 BC – 206 BC
	Western-Han and Eastern-Han	206 BC – AD 220
Middle Chinese AD 220 – AD 1368	Wei,Jin and Sothern-Northern Dynasties	AD 220 – AD 581
	Sui,Tang and Five Dynasties	AD 581 – AD 979
	Song and Yuan	AD 860 – AD 1368
Modern Chinese AD 1368 – AD 1911	Ming	AD 1368 – AD 1644
	Qing	AD 1364 – AD 1911

In consideration of the historical chronology of the Chinese dynasties and the SCC corpus as the primary sources of evidence, this study adopts the periodisation that mainly maintains the same periods after Pre-Qin (206 BC) in accordance with the SCC corpus. Additionally, the Shang dynasty has been added to the SCC's Archaic Chinese period, thus the Archaic Chinese period for this study would start from 1676 BC instead of 1200 BC and has been subdivided into four periods as represented in Table3.

Table 3: N1-N2 combination for lexical periodisation in the history of Chinese

Period		Name of dynasty	Time frame
Proto Chinese		Xia	2146 BC–1675 BC
Archaic Chinese (AC)	Early Archaic 1676 BC-206 BC	Shang	1676 BC–1029 BC
		West-Chou	1030 BC – 771 BC
		East-Chou	770 BC – 256 BC
		Qin	221 BC– 206 BC
	Later Archaic	Western-Han and Eastern-Han	206 BC– AD 220
Medieval Chinese (MedC)	Early Medieval	Wei,Jin & Sothern-Northern Dynasties	AD 220 – AD 581
	Middle Medieval	Sui,Tang and Five Dynasties	AD 581 – AD 979
	Later Medieval	Song and Yuan	AD 860– AD 1368
Modern Chinese (ModC)	Early Modern	Ming	AD 1368–1644
	Later Modern	Qing	AD 1644–1911

This periodisation of Archaic Chinese is more meaningful for the language changes investigated in this study. To more accurately reflect the language changes in the Archaic Chinese period, it is especially important to include the earliest possible period, that is, the Shang dynasty. The four periods within the early Archaic Chinese period align with the historical chronology of the Chinese dynasties that are noted in the Table.

In this section, we have discussed Mandarin as a standard spoken form that has been consistently used throughout Chinese history, and established the periodisation to be applied throughout this study.

2.2 Discrepancy between character and word in Mandarin

It is essential to discuss the discrepancy between characters and words in Mandarin Chinese before further considering compounds or disyllabic words. Without understanding what a word is in Mandarin, one cannot discuss word formation. Equally, morphological facts cannot be properly understood unless a word or character is known.

In the literature of general linguistics, the term *word* refers to a complete meaningful unit of language which is sometimes placed, in a hierarchy of grammatical constituents, above the morpheme and below the phrase; such constituents appear in many of the world's languages (Packard 2000:9). The pivotal point is that a word can stand alone. However, the definition of the term *word* is a much debated matter in the study of Mandarin.

Duanmu (2007:114), in discussing Chinese words, made the statement that "the notion 'word' was absent in Chinese linguistics until it was introduced in the twentieth century". The *zi* 'character' used in Chinese corresponded to the term *word* used in English before the twentieth century and it was only after the 20th century that the *zi* 'word' differs from the *zi* 'character'. In Duanmu's view, there were no morphological processes in Mandarin language before the 20th century, and a word was always represented by a morpheme or syllable. This view fails to differentiate between characters and words. For example,

(1)

a. 天地孰得
   tian *di* *shu* *de*
   heaven earth which suitable
   'the natural order and geography are most favourable to whom'
   (*The Art of War. Before 600 BC*)

b. 天地之大也
   tian *di* *zi* *da* *ye*
   heaven earth of big modal-PARTICLE
   'this is such a big world'
   (*The Doctrine of mean 551 BC – 479 BC*)

(1a) and (1b) are extracted from two different texts, each by a different author, but written in the same time period (before 500 BC). Although (1a) and (1b) include the same sequence of two characters *tian* 天 'heaven' and *di* 地 'earth', they are syntactically and semantically different from one another. The sequence *tian-di* in the expression (1a) is a phrase because the meaning of *tian-di* is derived from its components; also, a conjunction, such as *yu* 'and', can be inserted between the two nouns. This suggests that these two characters are the individual words. In contrast, *tian-di* is a compound in the expression (1b). This is because the meaning of the combination is not simply and structurally interpreted from its constituents (in which it has created a specific or opaque meaning); also, no word can

coherently be inserted between the two elements. This indicates that these two characters in (1b) are bound morphemes.

Giving another example:

(2)

他 們 是 諸 侯

ta men shi zhu hou

he plural-mark were every marquis

'they were governors'

It is clear that in English the sentence is comprised of three words, and in Mandarin there are five characters in the sentence. Both languages have the same sentence structure and word order, that is, subject, verb and object (S-V-O). As can be seen, the subject *they* in English has to correspond to the *ta men* in Mandarin before the verb were in English and *shi* in Mandarin. In the same vein, the word *governors* must correspond to *zhu hou* 諸侯, as the object. This example would suggest that *ta-men* and *zhu-hou* are words, even if each of them contains two characters. In particular, the word *zhu-hou* 'governor' was repeatedly used as early as 220 BC and the two characters/sounds always occur together, with two syllables or segments, in writing or speech, respectively (Wu 2001:137).

A number of issues can make it difficult to distinguish between words and characters in any given language. One of these is the type of writing system used. English, for example, which employs a running text and an alphabetic writing system, differs from a character-based system in a number of important ways. To begin with, words are the elemental writing unit in the contemporary alphabetic system, even though there were no spaces between words in English medieval manuscripts (Saenger 1997:4). The length of each word is flexible and depends on how many letters are included in the word. Words are generally delimited clearly by a preceding white space and a following white space or punctuation (although some compounds can be written with a space between two components if they are free morphemes).

However, in Mandarin, characters are the writing unit, and each character corresponds to one morpheme and one syllable, and the structure of each character is fixed. Characters may be a single word when they function as a free morpheme, or may be the parts of a word when they act as a bound morpheme, or act as an affix that cannot stand alone, requiring another free or bound morpheme to be attached to form a word. Unlike the contemporary

alphabetical system of English, the boundary between words in Mandarin is not clear. Mandarin running text is written as an unseparated string of words, even if the semantic boundary between morphemes is distinct. Even though there is a fixed-space between characters, this space is only used to separate characters or morphemes, regardless of how closely adjacent characters may be tied together as a word.

The second issue that causes confusion between words and characters is the form of Mandarin dictionary entries. In a Mandarin dictionary, each headword is an individual character or morpheme, regardless of whether the morpheme is free or bound. The pronunciation provided for each character always consists of a single syllable. That is, each character is represented as an independent unit and is defined as having at least one meaning. The entries of such a character-based dictionary could lead to the conclusion that each character represents a monosyllabic word (DeFrancis, 1984:190-9).

Furthermore, although morphemes and syllables are components of words and represent the smallest units in a word, they are different linguistic concepts. A morpheme is a meaningful morphological unit of a language and is a form-meaning combination, whereas a syllable is a phonological unit and does not create meaning. A word in English can consist of either only one morpheme or syllable, such as the word *book*, or more than one morpheme or syllable, such as the word *worked*. Nevertheless, morphemes and syllables do not always overlap so neatly. For example, the word *books* is made up of two morphemes, *book-* and *-s*, but it has only one syllable. By contrast, in Mandarin a morpheme is always represented by a character, and the character is always its single-syllable unit of pronunciation. This property of the Mandarin language may mislead other language users into thinking that a character always corresponds to a word.

In this section, we have examined the discrepancy between characters and words in Mandarin Chinese, and these two concepts will be used throughout this project.

2.3 History of disyllabic words, and compounding, in Mandarin

For much of Mandarin's later history, the use of disyllabic words has been considered a core attribute of the language, and thus an important topic in the study of the language's evolution. There has been much debate about morphological processing in Archaic Chinese, and as to how compounds were represented in the lexicon. This has led to much uncertainty about when compounds were first used in Mandarin.

2.3.1 The increase in using compounds

Much literature has been published on compounding in Mandarin, with the most common position being that the use of compounds increased during history (Duanmu 2007, Li 1993, Li 1981, Li 2013, Karlgren 1926, Pulleyblank 2000, Ting 2002, Wang 1957, Wu 2001). Duanmu (2007) and Li (2013) proposed that the use of disyllabic compounds were basically absent in Archaic Chinese but then underwent a dramatic increase, eventually reaching the situation where over 70 % of all words in Modern Chinese are compounds. In an in-depth diachronic investigation of verbal (V_1-V_2) Chinese compounds, Li (1993) showed that the frequency of compounding words gradually increased from 0.32% to 7.77% between the 5th Century BC and the 17th Century AD, as represented in Table 4 (Li 1993:130).

Table 4: Occurrence of compounds (data from Li 1993:130)

Source	Date	Total number of compounds	Percentage
Spring and Autumn Annals	5th Century BC	four nominal, one resultative	0.32
Records of the Grand Historian	2nd Century AD	68 (34 verb and 34 others)	0.80
The Platform Scripture	8th Century AD	173 (64 verb and 109 others)	2.04
Yuan Drama	13th Century AD	500 (98 verbs and 402 others)	5.88
S & W World	17TH Century AD	661 (112 verbs and 549 others)	7.77
Family	1931	1203 (273 verbs and 930 others)	14.15

While Li's investigation was confined to the sequences of V_1-V_2, the results demonstrated the broader pattern of the history of compound use, as viewed by many scholars. As the language evolved, these researchers proposed that compounding was increasingly used to form complex words.

2.3.2 Sound simplification accounts

The research on compounding discussed above is in broad agreement with work that has asserted that the lexicon of Archaic Chinese was mainly monosyllabic and that the appearance of compounding in Medieval Chinese was driven by the phonological simplification that occurred at this time (Karlgren 1926, Li 1993, Pulleyblank 2000, Wang 1957). For these scholars, the simplification of the sound system -largely a reduction in syllable structure- was proposed as the only driver of the appearance, and a gradual rise to primacy, of compound words. Below, these ideas are discussed in more detail.

The accounts of Karlgren (1926) and Wang (1957)

Karlgren (1926:42) proposed that Archaic Chinese had a rich sound inventory. The language was made "with simple words" that were sufficient to meet communication needs. This view was supported by Wang (1957:342-4) who concluded that, in Medieval Chinese the "polysyllabification of Chinese words is a logical consequence of the sound simplification" because there was "a very complex sound system" in Archaic Chinese that precluded the need for compounds.

In Karlgren (1926:42) and Wang's (1957:342) accounts, the use of compounds only became necessary after phonological simplification: words that had been distinguishable in Archaic Chinese became identical in Medieval Chinese. Thus, the large number of resulting homophones would have reduced the communicative effectiveness of the language, which was compensated for by the rise of compounding (Karlgren 1926:42, Wang 1957:342). For example, *yi* has more than one meaning, even when it has the same tone. yi(1), where (1) indicates the first tone in Mandarin, has many possible meanings, including: 'a', 'one', 'garment', and 'to treat'.

There is no doubt that the phonological syllable structures used in Archaic Chinese changed in some aspects in the transition to Medieval Chinese, as is evident in the reduction of consonant clusters and the loss of final consonants (Baxter 1992, Haudricourt 1954 & 1972, Pulleyblank 2000, Sagart 1999, Schuessler 2006 & 2009). For instance, consonant clusters were reduced to a single consonant in (3a), (3b), and (3c), and the final consonant was dropped from their syllables in (3d), (3e), and (3f) (Karlgren 1926:43, Schuessler 2006:249-565).

(3)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	酣	*han(1)*	*< gam*	*< *glam*	'be tipsy, drunk'
b	甘	*gan(1)*	*< kam*	*< *klam*	'be sweet'
c	藍	*lan(2)*	*< lam*	*< *gram*	'indigo'
d	梅	*mei(2)*	*< mə*	*< *məʔ*	'plum'
e	屎	*shi(3)*	*< si*	*< *lhiʔ*	'dung'
f	夷	*yi(2)*	*< ji*	*< *ljəj*	'be level'

The *ABC Etymological Dictionary of Old Chinese* (Schuessler, 2006:553-557) also includes individual morphemes or characters that were phonologically and morphologically distinct in Archaic Chinese but which became identical in Medieval Chinese. The initial consonant clusters **rj-* in (4c), (4d), and (4f), were reduced into a single initial consonant *j-* in the transition to Medieval Chinese. Similarly, the final consonant clusters **-m̚* in (4b) and (4e), and **-ms* in (4d) and (4f) were also simplified into a single final consonant *-m*. As a result, these six syllables were merged into a single, phonologically indistinct syllable *jam* in Medieval Chinese, but they maintained their original meaning from Archaic Chinese. In other words, these six semantically different words shared a common structure and are still used in the present day.

(4)

		Modern Chinese	Medival Chinese	Archaic Chinese	
a	炎	*yan(2)*	*< jam*	*< *liam* ，	‘to burn’
b	燄	*yan(4)*	*< jam*	*< *lam̚*	‘blazing up fire’
c	鹽	*yan(2)*	*< jam*	*< *rjam*	‘salt’
d	鹽	*yan(4)*	*< jam*	*< *rjams*	‘to salt’
e	剡	*yan(4)*	*< jam*	*< *jam̚*	‘pointed, sharp’
f	豔	*yan(4)*	*< jam*	*< *jams* or *<*rjams*	‘beautiful’

With respect to the consonant **-s* at the final position of the syllables in (5), (5a) and (5b) these were two different morphemes with different syllable structures in Archaic Chinese, even though they are represented by the same character (Schuessler 2006:554). The final consonant **-s* in the verb **rjams* was lost in the transition to Medieval Chinese (and the **r-* sound was also lost for both the verb and the noun). While both the verb and the noun shared a phonologically identical syllable structure in Medieval Chinese, they were morphologically different.

(5)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	鹽	*yan(2)*	*< jam*	*< *rjam*	‘salt’
b	鹽	*yan(4)*	*< jam*	*< *rjams*	‘to salt’

There is no question, according to the examples demonstrated above, that homophonous words emerged in Medieval Chinese, and this was correlated with a reduction in the complexity of the consonant clusters and the dropping of final consonants.

However, example (4) also indicates something that counters the views of Karlgren (1926) and Wang (1957). It makes clear that in Archaic Chinese, the syllable structure **rjams* (in (4d) and (4f)) were phonologically the same, yet they represented two different morphemes or characters. Further examples of the same are listed below:

(6)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	淋	*lin(2)*	*<lim*	*< *rəm*	'to pour, drench'
b	臨	*lin(2)*	*<lim*	*< *rəm*	'to approach'',

(Schuessler 2006:359)

(7)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	識	*shi(4)*	*<tsə*	*< *təkh*	'to remember, record'
b	織	*zhi(4)*	*<tsə*	*< *təkh*	'woven cloth'

(Schuessler 2006:469 & 615)

(8)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	信	*xin(4)*	*< sin*	*< *sin*	'to believe'
b	薪	*xin(1)*	*< sin*	*< *sin*	"firewood'
c	新	*xin(1)*	*< sin*	*< *sin*	'new, renew'
d	辛	*xin(1)*	*< sin*	*< *sin*	'bitter'

(Schuessler 2006:538-539)

As can be seen from (6), (7), and (8), in Archaic Chinese, morphemes from the same group shared the same phonological syllable structures, e.g., **rəm* in group (6), **təkh* in group (7), and **sin* in group (8). However, they did not share meanings, that is, they were morphologically different. These examples demonstrate that homophonous words did not have their origins in Medieval Chinese, but also occurred in Archaic Chinese. Thus the

conditions that were thought to be necessary for compounding were already present in Archaic Chinese.

The existence of homophones in the sound inventory of Archaic Chinese is problematic to Karlgren (1926) and Wang's (1957) proposed relationship between phonological simplification and the appearance and rise of compounding. It argues against the sound inventory of Archaic Chinese being sufficiently rich to preclude the need for compounding. Karlgren (1926) and Wang's (1957) argument could be an example of a teleological linguistic argument (Vincent 1978:409), that is, in the absence of any evidence that compounding did not exist in Archaic Chinese, this nevertheless seemed an attractive idea to them, because the phonological simplification of the Medieval period suggested there would have been less need for compounds in the earlier times.

Another problem to Karlgren (1926) and Wang's (1957) studies is that they did not consider other processes that might lead to multi-syllable words being used, even if compounding itself was absent in Archaic Chinese. Many disyllabic words, for example, are formed by the process of affixation. These affixes are normally bound morphemes and are represented by single syllables or characters. The number of syllables, morphemes and characters are interdependent. That is, if there is one character, then there is one morpheme and one syllable. For example, the derived words *jun-zi* 君子 gentlemen-affix 'gentlemen' and *zhan-zhe* 戰者 to fight-affix 'fighter' are each comprised of two morphemes and two syllables.

In summary, the claim that the phonological simplification was only the possible cause of compounding in the development of Mandarin is problematic.

The account of Li (1993)

Li (1993:127) proposed that the simplification of sounds that occurred in the Medieval period was the direct cause of the appearance and rise of compounding in the period thus further promoting the ideas of Karlgren (1926) and Wang (1957), discussed above. Li's argument is supported by his results that report that the occurrence of compounds gradually increased from 0.32% to 7.77% between the 5th Century BC and 17th Century AD (see Table 4). In Li's view, his results were in line with Wang's observation of the sound inventory being reduced by 50% between the 6th and 8th century, and then being reduced by a further 50% by the 14th (Wang, 1957:342).

While there is an obvious correlation between the two phenomena, this does not mean that the reduced sound inventory was the only cause of the increased use of compounds. Although sound simplification results in homophones that may accelerate the use of compounding, as argued in the previous section, we cannot extrapolate backwards to deduce that there were no homophones and no compounds in Archaic Chinese. In addition to the teleological nature of that line of thinking, the claim that compounds only appeared in Medieval Chinese is complicated by the sorts of Archaic Chinese words that were phonologically simplified. There is linguistic evidence that, in the transition to Medieval Chinese, the sounds of both simple and complex words were reduced. For example, it has been suggested that the final consonant *-*k* in (9a) **hrak-hrak* and (9b) **krak-krak* was also reduced, thus forming *hak-hak* and *kak-kak*, in the Medieval Chinese period (Schuessler 2006:253 & 276). Likewise, the consonant *-*h* in (9c) was also dropped (Schuessler 2006:561). This indicates that words of variable complexity were open to phonological simplification, including those formed through complete reduplication. According to Wu's research (2001:161-2), complete reduplication was the main compounding process occurring in early Archaic Chinese, and this issue will be discussed further in this study.

(9)

		Modern Chinese	Medieval Chinese	Archaic Chinese	
a	赫赫	*he(4)-he*	*<hak-hak*	*<*hrak-hrak*	' illustrious'
b	格格	*ge(2)-ge*	*<kak-kak*	*<*krak-krak*	'princess'
c	夜夜	*ye(4)-ye*	*<ja-ja*	*<*jah-jah*	'nightly'

Secondly, Li (1993:130) found only four examples of nominal compounds in the earliest period: *da-fu* big-husband 'senior official' (a rank in Archaic China), *fu-ren* husband-people 'madam or wife', *gong-zi* duke-son 'prince' and *tian-wang* heaven-king 'emperor or king'. These four compounds are proper nouns, and were not structured in the same way as the V_1-V_2 sequences that he studied more intensively (in terms of how their meaning related to their constituents). Given that Li (1993:130) did not discuss the criteria used to identify nominal compounds (as he did for V_1-V_2 compounds), his characterisation of the historical evolution of compound use, as shown in Table 4, is therefore open to question.

It is, perhaps, also surprising that the investigated text, *Chun Qiu* 'Spring and Autumn', contains between 8000 – 8500 characters (Li 1993:96), but "hardly any compounds were

found" (Li 1993:130). If what is said about the *Chun Qiu* is true, that is, it is "a common practice among the court historians of those days to record events according to the day, the month, the season and the year during which the events took place", as Li (1993:88) suggested, it would be expected to include a much larger number of nominal proper nouns. For instance, various proper nouns not noted by Li (1993:130), such as the place name *chao-ting* 朝廷 'royal or imperial court', the dynasty name *chu-qiu* 春秋 'spring and autumn', the book title *shi-shu* 詩書 'the book of poetry', and *zhu-hou* 諸侯 every-marquis 'marquis', all occurred in almost every text in the literature of the Pre-Qin period (before 220 BC) (Wu 2001:137).

Thirdly, the relationship between sound simplification and the appearance and rise of compounding is not a strict one. In many correlations, it is difficult to determine which is the cause, and which is the effect. In fact, according to Wu (2001), if there is any cause-effect relationship as suggested by Karlgren (1926), Wang (1957) and Li (1993), it might not be that sound simplification drives compounding, but that the linguistic availability of compounding (driven other forces) enables sound simplification. According to this view, the rich sound system of early Archaic Chinese meant that the language had a sufficient store of monosyllables; and thus, some monosyllables (or more precisely, morphemes) began to be used to form compounds (Huang, as cited in Wu, 2001:378). In turn, the increasing inventory of compounds would reduce the need for any richer variety of monosyllabic words: there would already be enough variety to form millions of compounds, and thus more than enough to meet communication needs (Xu, as cited in Wu, 2001:379). While this would not necessarily lead to a simplification of sounds, just a deceleration of phonological diversification, and at the least, the original sound system should be maintained (Wu 2001:379).

In summary, Li's argument, that compounding came into existence in Medieval Chinese solely because of the accompanying simplification of sound, is problematic.

The account of Pulleyblank (2000)

Pulleyblank (2000) has argued that because of rich sound inventory in Archaic Chinese, words were formed or created only by changing (replacing or dropping) consonants or vowels within monosyllables, rather than compounding. The examples provided by Pulleyblank are as follows: the transitive verb **dron* 'to transmit' in (10b), was derived from

an intransitive verb **tron* 'to turn around' in (10a) by replacing the initial consonant from the voiceless **-t* to the voiced **-d* (Pulleyblank 2000:36). The adjective **huʔ* 'good' in (11a) and the verb **huh* 'to love' in (11b) were formed by alternating the articulation between the plosive **-ʔ* and the fricative **-h* (Pulleyblank 2000:30). The accusative noun **sləkh* in (12a) was derived from the verb **mlək* in (12b) by replacing the voiceless alveolar fricative**s-* to the bilabial nasal **m-*, and the glottal fricative **-h* to the velar plosive **-k* simultaneously (Pulleyblank 2000:37).

(10)

		Archaic Chinese	
a	轉	**tron*	'to turn around'
b	傳	**dron* < **tron*	'to transmit'

(11)

		Archaic Cheese	
a	好	**huʔ*	'good'
b	好	**huh*	'to love'

(12)

		Archaic Cheese	
a	食	**sləkh*	'to give food to'
b	食	**mlək*	'eat'

These three examples are highly credible cases of the word formation processes invoked: no segmentable morphemes or syllables were added, and they all remain monosyllabic. In linguistic terms, these are known as non-concatenative operations or processes, because, in morphological terms, the process does not rely on the stringing together of multiple morphemes (Haspelmath & Sims 2010:34). Nevertheless, the existence of such examples does not mean that non-concatenative operations were, in Archaic Chinese, the only the means by which new words could be formed, as Pulleyblank (2000) suggested.

As mentioned previously, words can also be formed by attaching derivational affixes that add an extra syllable to a stem. For example, the suffix *-zhe* ('-er'), is commonly combined with verbs to form nouns, such as *xue-zhe* study-suffix 'scholar' and *zhan-zhe* fight-suffix 'fighter'. The meaningless suffix *-zi* can be attached to a noun morpheme to produce a

disyllabic noun, e.g., *li-zi* 李子 plum-suffix 'plum' and *wa-zi* 襪子 sock-suffix 'sock'. Like derivational affixes, inflectional affixes can also be used to form disyllabic words. For example, the human noun plural suffix *-men*, can be added to singular pronouns, e.g., *wo* 我 'I' and *ta* 他 'he', to change their grammatical meaning, e.g., *wo-men* 我們 'we' and *ta-men* 他們 'they. Overall, however, very few inflectional affixes occur in Mandarin Chinese.

This indicates that Pulleyblank's interpretation overlooks some morphological processes, such as derivational and inflectional affixation, that are fundamental to forming words of more than one character. In Mandarin Chinese, a word that consists of one character or more should be a unit that is characterised by syntactic and semantic independence and integrity (Li & Thompson 1981:13). Also, compounding could still occur regardless of, or in addition to, these processes.

Further, Pulleyblank (2000:26) stated that "the traditional characterisation of Chinese as monosyllabic is not far off the mark", from the perspective of morpheme structure. However, Pulleyblank (2000:44) also suggested that there was "no lack of compounds" in Archaic Chinese, an idea based on his hypothesis that the meaning of a string of characters cannot be inferred transparently from their parts. Examples of this, given by Pulleyblank are: *bai-xing* hundred-surname 'common people' and *gua-ren* solitary-man 'I, self-reference'. Clearly, these two positions are contradictory, because there cannot be monosyllabic compounds.

In summary, there are problems with Pulleyblank's ideas, that is, that Archaic Chinese was monosyllabic, because its rich sound inventory meant that a sufficient variety of words could be formed by non-concatenative operations, without the need for compounding.

Drawing together this evaluation of the accounts discussed in section 2.3.2, there are numerous issues with the claim that the sound simplification was the only driver of the appearance and rise of disyllabic words in Medieval Chinese. For example, homophones existed prior to this period, in the Archaic Chinese period. Some research has analysed data without specifying the criteria applied in choosing the data. In other work, disyllabic words have been discussed without accounting for the fact that words with derivational affixes are an important subcategory of the group. These limitations suggest the possibility that disyllabic or compounding words existed in Archaic Chinese and indicate the need to further investigate their origin.

2.3.3 The appearance of Compounding

While some scholars claim that lexical entries of Archaic Chinese were almost exclusively monosyllabic words (Karlgren 1926, Li 1993, Pulleyblank 2000, Wang 1957), some researchers, exploring the origins of disyllabic words in Mandarin, argue that it is not true; they propose that compounding existed long before Medieval Chinese (Feng 1998, Jin 2017, Kennedy 1951, Tao 1996, Wu 2001, Yu 1990). For example, some compounding proper nouns have been reported as occurring in the oracle bone inscriptions (Jin 2017:89, Tao 1996:68, Wang 2017:39, Yu 1990:87), and reduplicative disyllabic words have also been found in Archaic Chinese (Wu 2001:161-2). Both types existed before sound simplification took place, and are discussed in detail below.

Compound proper nouns in the oracle bone inscriptions

Compounding words are first attested in the oracle bone inscriptions, the earliest known form of Chinese writing, from the Sang period (1675-1029 BC). They are traditionally known *Hewen* (合文) 'combined characters'. *Hewen* refers to the process of combining characters, which were largely ideographic, and usually written together, occupying a single character's space (Wang 2017:39, Yu 1990:89). Phonologically, each *Hewen* maintains the original shapes of its combined characters, and each character must be pronounced, even though a *Hewen* was written as an individual character (Lee 2012:93-95, Li 2008:196, Wang 2017:39). For example, the *Hewen* in (13c) is pronounced as three syllables i.e., *shi-er-yue* and other examples in (13) are read as two syllables, respectively, that is, *wu-yue*, *liu-yue*, and *zu-jia.*

(13)

	Hewen	components			Pin-yin	English
a.		five	moon		*wu-yue*	'May'
b.		six	moon		*liu-yue*	'June'
c.		ten	two	moon	*shi-er-yue*	'Decembers'
d.		first	ancestor		*zu-jia*	'King'

(Yu 1990:85)

A majority of the studies of the *Hewen* in the oracle bone inscriptions have reached a consensus that the *Hewen* are dominated by proper nouns (e.g., months in (13a) and (13c), special names in (13d); and places), and that the relation between components is mainly hierarchical, modifier-to-modified (Jin 2017:89, Tao 1996:68, Wang 2017:39, Yu 1990:87). Given that the *Hewen* were not formed by mere random concatenation, it is not surprising that their structures were relatively stable (Jin 2017:89, Wang 2017:89, Yu 1990:87).

Tang (2007) reviewed the literature from the period of the Shang dynasty to West Chou dynasty (1675 – 771 BC) and argued that proper nouns formed by combining characters in the Shang period can be thought of disyllabic or polysyllabic words, as shown in (13) (Tang 2007:251). The basis of Tang's argument is that these proper nouns have specific meanings and represent a particular person, place or object, similarly to the English words, *White House* and *greenhouse*. For instance, apart from (13c) consisting of three characters, (13a), (13b), and (13d) were made of two characters correspondingly, and, semantically they refer to a completed concept. These proper nouns, thus, are disyllabic or polysyllabic words according to their meaning and the phonological rules of the *Hewen* mentioned above (Lei 2009:9, Li 2008:196, Pan, as cited in Wu 2001:78, Yu 1990:89).

In contrast, in 1994 the research of Guo (as cited in Wu, 2001:308) proposed that *Hewen* should not be thought of as disyllabic or polysyllabic compounds based on the results of *Hewen* occurred in the Oracle Dictionary that were mainly proper nouns in subordinative relations. In Guo's view, the *Hewen shang-di* 上帝 superior-emperor 'god', for example, should be considered as a phrase, because the modifier *shang* 'superior' in this subordinative *Hewen* merely functioned as an adjective to narrow or describe the noun *di* 'emperor'. Within this combination, the original components still maintained the functions and meanings that they had in the Shang period, even if the phrase later acted as a compounding word (after the West Chou period, post 770 BC). Indeed, it is true that the combination *Shang-di* would be a phrase if its parts are free morphemes and its meaning is interpreted as 'superior-emperor' rather than 'god'. According to historical records, *shang-di* occurred in oracle bones inscription refers to 'god' (Guo 2010:64).

While there are clear differences in the two interpretations of the *Hewen* described above, overall, it seems that Tang's account is the more reasonable. First of all, theoretically, the constituents of a compound can be free or bound morphemes, and these constituents may partially or completely lose their original meanings, and create a new meaning for the

combination. In practice, the name of a place, official or tribe occurred frequently in every text of the Shang period (Wu 2001:308). Many of these proper nouns are the names of a particular person, place, organisation, or object, and have special meanings, thus distinguishing that person, place, organization or object from others (Wu 2001:309). In addition, these proper nouns are relatively fixed in the language rather than being just a random combination (Lei 2009:11, Li 2008, Jin 2017:89, Wang 2017:40). They longer maintain the meanings of their constituents, which themselves do not refer to any particular person, place, organization, or thing. Perhaps, those *Hewen* that semantically refer to a completed concept can, at the least, be conceived of as precursors of compound words.

While compound proper nouns are first attested in oracle bone inscriptions, the question arises if any other types of compounds existed in Archaic Chinese, a point which is discussed below.

Compounds formed by the reduplication of syllables

Looking for evidence of compounds in addition to those that are proper nouns, Wu (2001) examined eleven ancient books and records from before the Pre-Qin period (before 220 BC), and manually gathered and analysed data, having specified the criteria used in his classification. From his analysis, Wu identified compounding as the main contributing factor to di-syllabication or poly-syllabication in Archaic Chinese. His results showed that about 20% to 40% of all the words in each book or record were compounds, excluding personal names, names of places , and names of dynasties (Wu 2001:363). He reported that the majority of compounds found in early Archaic Chinese were reduplicative compounds, such as the complete reduplication, *ren-ren* 'people' or the partial reduplication *zhi-zhu* 'spider'; over 60 % of reduplicated compounds involved complete reduplication (Wu 2001:161-2). The details of his analysis are shown in Table 5. These results might not match the observations of other scholars, due to the use of different criteria, but his work represents a valid argument that compound words were used much more frequently in Archaic Chinese than others have proposed.

Table 5: Appearance of compounds before Pre-Qin (Wu 2001:363)

Name of text		Time frame	Compound	Words in texts	%
Shangshu	The classic of history	before 400 BC	374	1924	19.4
Shijing	Book of odes	1100 BC - 600 BC	974	3450	28.2
Lunyu	The Analects	551 BC - 479 BC	329	1479	22.2
Zuozhuan	The Commentary of Zuo	770 BC- 476 BC	1185	4177	28.3
Mozi	The Book of Master Mo	475 BC -221 BC	1336	3977	33.6
Mengzi	Mencius	372 BC -289 BC	651	2240	29
Zhuangzi	Zhuangzi	369 BC – 286 BC	1965	5170	38
Shangjunshu	The Book of Lord Shang	390 BC -338 BC	447	1353	33
Xunzi	Xunzi	316 BC – 237 BC	1356	3753	36
Hanfeizi	Hanfeizi	280 BC – 233BC	1484	3762	39.7
Lüshi Chunqiu	Lüshi Chunqiu	before 238 BC	1148	3992	28.7
			11249	35277	31.89

In addition, Wu (2001:349) provided the results from Xiang's investigation into the relation of the constituents of compounds in the text, *Shijin* 'Book of Odes'; this was done in order to discover the rules that applied to disyllabic word formation in early Archaic Chinese. Even though much of the *Shijin* was published in the Spring Autumn and Warring States period (770 BC – 221 BC), some sections were written earlier, in the West Chou period (1029 BC – 771 BC). Also, it can be assumed that the language used in the text was indicative of the language used before this period, given that language change is a gradual process. Xiang's investigation shows that over 50 % of the 974 compounds that occurred in the *Shijin* are subordinative (about 30%) and coordinative compounds (over 20 %) (Xiang, as cited in Wu, 2001:349). Xiang's observations show that Archaic Chinese compounds were structurally formed using various rules of syntax, and that subordinative and coordinative relations were the most favoured types in compound formation (Feng, 1998:209, Xiang, as cited in Wu, 2001:349).

Even though Wu (2001) shows that more compounds in very early Chinese are not just proper nouns (as indicated by the oracle bone inscriptions), the compounds identified by Wu' study were, nevertheless, produced by a limited linguistic word formation process, that is, reduplication. However, this finding might be an outcome of the methods used, such as the choice of source texts and the criteria used to identify compounds.

For example, Wu's study relied heavily on semantic properties and did not take into account that Chinese characters are not always free morphemes. Identifying whether morphemes are free or bound should be done prior to analysing the meanings of individual components and their interrelationships. If one morpheme of a given combination of two is

a bound morpheme, then this combination has to be, definitionally, a compound or disyllabic word. Thus, Wu's classification of compounds, without taking into consideration the function of the morphemes, is inappropriate. For example, the word *fu-mu* 父母 father-mother 'parents' would, in theory, classified as a phrase according to Wu's semantic criteria, because the meaning of the given combination is interpretable from the components (Wu 2001:134). However, in practice, *fu-mu* was actually classified as a compound in Wu's study, based on his criterion of "more frequently occurring" (Wu 2001:132). Although the meaning of *fu-mu* is indeed interpretable from its constituents, both *fu-* and *-mu* are bound morphemes and need to be attached with another free or bound morpheme to form a compound, e.g., *fu-qin* 'father' and *mu-qin* 'mother'.

It is evident that if the features of morphemes were considered first, the semantic criteria would not clash with that of "more frequently occurring". Wu included "more frequently occurring" as one of the criteria to classify compounds based on the approach of Zhang & Yan (as cited in Wu, 2001:137), who had proposed that the majority of Chinese compounds originated from phrases that employed the same sequence of characters. They claimed that the process of conversion in Archaic Chinese involved three stages: at the start, a combination of monosyllabic words was used infrequently, then it was regularly used, and finally, it was fossilised as a compound.

However, a more reliable approach to classifying a given combination of two characters as a compound or a phrase, depends, not on frequency of use, but on the function of the morphemes (whether they are free or bound forms), the semantic relationship between a given combination and its morphemes, and stress patterns (Chao 1968:361-2). Although a "more frequently occurring"' combination of two free morphemes can be a compound, this depends on whether its meaning derives from its constituents or not. For example:

(14)

a. 天 子
   tian-zi
   heaven-son
   'emperor'

b. 百姓
   bai-xing
   hundred-surname
   'common people'

In these examples, while both morphemes in each of (14a) and (14b) are free forms, they cannot be separated when they are interpreted as 'emperor' and 'common people'. The meaning of the whole combination is not a composition of its morphemes, and a new meaning has been created. This indicates that the classification of *ti-zi* and *bai-xing* is more accurately a reflection of the fact that they are not semantically transparent, than of the fact that they are "more frequently occurring".

Wu's findings that complete reduplication was the main process to derive compounds or disyllables, and the criteria he applied to identify compounds from phrases with the same combination of characters prompted a further investigation into the origin of compounds that excludes complete reduplications.

2.3.4 Review of the criteria for identification of compunds

While compounding is one of the more unanimously agreed upon processes, at least in prototypical cases, different views exist as to how to identify Chinese compounds. This debate centres around whether the term *compound* should only refer to two or more free morphemes that are bound together, or should also include bound morphemes other than affixes, or even cover derivational morphemes as well. Under almost no framework does compounding also include affixation though there is a slightly grey area. The key issues, as discussed below, are delineating compounds from phrases, and delineating compounds from derivatives.

In *The Morphology of Chinese*, Packard (2000) argues that polysyllabic words with at least one bound morpheme should be excluded from the category of compounds (Packard 2000:78). A similar view is proposed by Zhou and his colleagues (1999), in the study of Morphology, Orthography, and Phonology in Reading Chinese Compound Words. Packard explained that some components of disyllabic words no longer functioned as a free morpheme (although they had in Archaic Chinese) and thus these words had been lexicalised as simple words, not compounds during the period of their development (Duanmu 1998:135, Packard 2000:281). For example, taken from Packard (2000), if we consider *di-tie* 地鐵 earth-railroad 'subway', Packard (2000) suggests that the morpheme *-tie* is a bound root in *di-tie* when used with its new meaning 'railroad', but it is free morpheme when used with its original meaning 'iron'. Under this interpretation, *di-tie* has been lexicalised as a simple word, and it is a single morpheme with two syllables. Nevertheless, the word *di-tie* can also be

analysed as having a bound morpheme *-tie*, even if *-tie* lost its original meaning as Packard suggested. This interpretation also applies to the English word *cranberry* in which at first sight seems to a simple word. In fact, it can be analysed into *cran-* and *-berry* regardless of whether *cran-* is defined as an empty or unique morpheme.

Another problem with excluding words with a bound morpheme from the category of compounds is that does not take into account that some constituents of disyllabic words never acted as free morphemes. For example, the word *zhu-hou* 諸侯 every-marquis 'marquis' occurred in almost every text in the literature before 220 BC (Wu 2001:137). The first component *zhu-* in the word is a bound morpheme that cannot stand alone as a free morpheme and has to be attached to another bound morpheme *-jiu* or free morpheme *-hou* to form the word *zhu-jiu* 諸舅 every-mother's brother 'uncle' or *zhu-hou* 'marquis'. *zhu-, -hou,* and *-jiu* are the components of the word *zhu-hou* and *zhu-jiu*, which remain semantically and grammatically transparent. This suggests that not all compounds are composed of free morphemes, and not all words consisting of one or more bound morphemes are lexicalised.

Li & Thompson (1981:46) proposed a much wider definition than Packard (2000). They stated that compounds are "all polysyllabic units that have certain properties of single words and that can be analysed into two or more meaningful elements, or morphemes, even if these morphemes cannot occur independently". This, however, fails to clarify whether some derived words should be included as compound words. For example, the derived word *cha-tou* to plug-affix 'plug' can be analysed into two meaningful morphemes *cha-* and *-tou*, and would thus satisfy their criteria. In addition, the bound morpheme *-tou* occurs in other Chinese words as an affix, such as *shi-tou* 石頭 stone-affix 'rock' and *wai-tou* 外頭 outside-affix 'outside'. Li & Thompson's definition of the compound generates confusion, by implying that derived words should be considered in the same category as compounds.

In contrast, Chao (1968:359) held the view that a compound is a combination of two or more free morphemes, or bound morphemes other than affixes, brought together to form a single word. In Chao's view, compounding forms a word out of two or more root morphemes. Although, in English, native compounding roots are typically free morphemes. such as *fire* and *place* in the word *fireplace*, some compounds are borrowed from Latin and Greek and the components of these borrowed compounds retain their original status as bound morphemes (Kemmer 2015), e.g., *photograph* and *iatrogenic*. While Mandarin does

not have a large amount of borrowed words from Latin and Greek, it has undergone changes during a longer period of language development. For various reasons, some morphemes are of indeterminate status. First, a morpheme might be free in at one point in history but bound in others. For example *yan* 言, is a bound morpheme in modern Chinese but it was a free word in Archaic Chinese. Second, the status might depend on the usage of the morpheme. A useful example of a morpheme being bound in one usage but free in another is the morpheme *ge* 'song' that functions as a free morpheme in the expression *Ta chang ge* 她唱歌 'she sing a song'. In many another usages, however, it functions as a bound morpheme, such as in *ge-hou* 'voice', *ge-shou*, 'singer' and *ge-song*, 'extol'.

In practice, however, discussions about whether a sequence of characters should be perceived as a compound or a phrase has been among linguists. The same combination of two syllables can be either a word or a phrase at a different point in processing time. For example, in:

(15) 一筆好買賣
yi bi hao mai mai
one-strock good buy sell
'a good deal in trade'

(16) 買賣公平
mai mai gong ping
buy sell fair flat
'fair in buying and selling'

Mai mai 買賣 in (15) is a compound word, whereas in (16) it functions as a phrase. The meaning in (15) cannot be explained from consideration of its components and a new meaning is created; in (16), however, the meaning is derived from from its components.

Chao's view (1968) appears to sit between those of Packard (2000) and Li & Thompson (1981). Chao's view would seem the most appropriate and objective of the three, better reflecting the development of the Chinese lexicon, and thus can be applied in determining criteria for the classification of compounds in the current study. These criteria will be discussed further in the Research Methods section.

2.3.5 Motivation

In previous research on the topic of Chinese compounds, diverse interpretations regarding the origin and development of Chinese compounds have been proposed. These include: the simplification of the sound system in Medieval Chinese acting as the sole driver of compounding (Karlgren 1926, Li 1993, Pulleyblank 2000, Wang 1957); early words forming within monosyllables as a reason for the lack of compounding in early Archaic Chinese (Pulleyblank 2000); proper nouns being the only form of compounds in the Shang period (1675 – 1029 BC) (Jin 2017, Tao 1996, Wang 2017, Yu 1990); and reduplication of syllables being the main process to form compounds in early Chinese (Wu 2001). The discussion of these proposals in the previous sections has identified that several points require further examination.

Firstly, while research on some syntactical combinations (e.g., V_1-V_2) in the Archaic Chinese period has been undertaken, as discussed in the Literature Review, studies focused on the N_1-N_2 combination are underrepresented. Investigating only the N_1-N_2 combination would naturally tend to limit analysis to compounds that are themselves nouns. Although the N_1-N_2 combination can also be used to form verbs and adjectives, the number of such compounds is quite small. Huang (1998) examined the range of possible Chinese complex word structures using the dictionary of *Guoyu Ribao Cidian* (*GRC*) 'Mandarin Daily Dictionary'. The author found that over 98% of all N_1-N_2 combinations (6910 out of 7021) are nominal compounds and 55% of all nominal compounds (6910 out of 12481) consist of two nouns N_1-N_2, as shown in Table 6 (Huang 1998:264). This data suggests that the N_1-N_2 combination is highly productive and creative (Chao1968, Li & Thompson 1981:53, Packard 2000:126-7). This may facilitate the discovery of the origin of N_1-N_2 words in early Archaic Chinese and make it possible to track the historical development of their use within the SCC corpus.

Table 6: Number of disyllabic compounds in GRC (Huang 1997: 264)

Structure	Noun	Verb	Adjective	Total
noun-noun N_1-N_2	6910	21	90	7021
adjective-noun A-N	2961	-	198	3159
verb-noun V-N	1581	2940	378	4899
verb-verb V_1-V_2	276	3730	103	4109
Others	753	1587	2063	4403
Total	12481	8278	2832	23591

In addition, the prominence of nominal compounds in Mandarin can probably be traced back to Chinese antiquity, that is to the earliest language used in the Shang period (1675 - 1029 BC) as discussed in section 2.3 (Bao 2009, Lei 2009, Li 2008, Tang 1997, Wu 2001, Yu 1990). Since some of nominal compounds (proper nouns) used in the Shang period were continuously used until the West Chou period (1029–771 BC) (Tang, as cited in Wu, 2002:342), this would provide a consistent thread in the use of nominal compounds. This would facilitate a diachronic analysis of syntactic patterns and semantic relations during the development of Mandarin compound use. Further, concentrating on the N_1-N_2 combination will allow this study to focus on what are the majority of two-syllable combinations (some nominal compounds are comprised of more than two syllables, such as *Xue-li-hong*, red-in-snow, 'mustard green' and *li-bai-er*, 'Tuesday'). This concentration would also delimit this study's database, avoiding other two-syllable combinations, such as A-N, V-N and V_1-V_2, and put the study in a better position to discover principles and generalisations that may also apply to future research.

Secondly, there are conflicting reports on the origin of compounds as discussed previously. Some studies claim that Archaic Chinese was almost solely monosyllabic, while other literature argues that compounding existed, but it was of a very limited nature and range. Regarding the former view, Chao (1968:139) used his observations about "a language in which every syllable has a meaning" to reach the conclusion that for Archaic Chinese "the so-called 'monosyllabic myth' is in fact one of the truest myths in Chinese mythology". This conclusion has influenced some sinologists for decades, such as Li and Thompson (1981:14). Having knowledge of the difference between characters and words would facilitate a better understanding of the morphological processes of complex words in Mandarin.

Thirdly, previous studies have for the most part gathered data manually, which may affect the accuracy of the data. For example, the number of compound words collected from the *Jinwen* were reported differently by scholars, such as 165 (Xu 1992), 237 (Tang 1986), and 242 (Wu 2001). To improve the accuracy of data collection and analysis, using an online corpus is highly promising. Automated processing is faster, and more data can be collected in a short-term project. Thus, more accurate conclusions can be drawn because of the larger data set. Further, the data is likely to more consistent, in that the potential for human error, when using a well-established corpus, is reduced.

The overall aim of this study is to investigate the origin and development of compounds, in particular, of disyllabic words, and the relationships between their two components. This study first critically evaluates models and frameworks relevant to choosing the syntactic and semantic criteria with which compounds and phrases of the same two characters can be distinguished (focusing on the N_1-N_2 combination in particular). This study then uses the texts of the SCC corpus to gather solid empirical evidence related to the origin and development of compounds or disyllabic words throughout Chinese history.

This chapter has provided a review of previous studies' accounts of the origin of disyllabic words in Mandarin. It has made clear that these explanations were problematic or incomplete, and it has justified the current project, a corpus-based examination of N_1-N_2 words, as an ideal case study to help resolve these issues.

3. Research Methods

This chapter discusses and justifies the research strategy, a digital corpus-based study. Section 3.1 provides an introduction to the corpora used in this study. Section 3.2 establishes criteria for identifying compounds. Section 3.3 then discusses potential problems related to data collection, and the final section presents data collection techniques (focused on digital data) and analysis procedures.

3.1 Sources of data

This project, a study of the historical changes of linguistic phenomena, required a body of written texts as analysed sources. These could be used to identify particular syntactic structures and elements, and thus infer changes within the language. For such purposes, online or computerised linguistic corpora have been established, drawing upon a large variety of historical textual records. These provide quantifiable evidence for making generalisations about language use (Kennedy 1998:1-7). Computerised corpora are more effective than physical texts, because the digitisation results in the researcher making fewer errors when analysing large volumes of material (Kennedy 1998:5), and assist in mapping the data and distribution of the linguistic emergence in traceable texts in history to exhibit the frequency of the elements being analysed.

The main digital resource used in this study is the Sheffield Corpus of Chinese (SCC), one of the newest and most up-to-date online corpora. It provides coverage of the full historical record; its wide range of fully marked-up Chinese texts is enhanced with an integrated search and analysis tool, which is not a feature of most other computerised Chinese corpora. For example, the collection of the text samples from the Lancaster Corpus of Mandarin Chinese (LCMC) and the Academia Sinica Balanced Corpus of Modern Chinese (Sinica Corpus) are confined to contemporary Chinese texts (Xiao & Hu 2015:39-41). While the Centre of Chinese Linguistics at Peking University (CCL) and the Academia Sinica Ancient Chinese Corpus at the Institute of Linguistics in Taiwan (ASACC) both cover the same time periods as the SCC (from 1000 BC to Modern Chinese) (Hu & McLaughlin 2007:421), and the CCL has the largest sample collection of archaic Chinese texts, neither of these corpora provide marked-up systems for linguistic analysis; they are also deficient in organised text categories (Xiao & Hu 2015:39).

The texts chosen and organised in the SCC corpus are well-suited to the objective determination of the frequencies of the various morphological and syntactical phenomena that are the focus of this project. Instead of prioritising the volume of texts in the corpus, the SCC corpus includes a wide range of 17 text types (e.g., drama, fiction, including general and historical, legal documents, and philosophy). This means the data are more representative. Although the SCC corpus only contains 40 texts in total, which can be converted into about 432,670 Chinese characters, this is more than sufficient in terms of the linguistic features and analyses that can be associated with the corpus. For example, 21 word classes, 49 associated special categories, and 112 distinct tag labels have been included in the SCC corpus (Appendix 1).

Most importantly, all the sample texts in the SCC corpus are annotated with a mark-up system, accessible by an integrated search and analysis tool that displays: (i) the original emergence of any specified sequence of characters and any particular word class; and (ii) the source texts for any identified pattern. The search system can also calculate and display the frequencies of specific forms for a certain time period or genre. This corpus thereby facilitates the identification of disyllabic words throughout Chinese history, particularly by examining the sample texts from the early Archaic Chinese period, the main focus of this study. The marked-up texts and their distribution by genre in the early Archaic Chinese period are shown in Table 7.

The eight samples of early Archaic Chinese in the SCC corpus are well-accepted as the classic works of literature of the period. They are distributed across four different genres, (history, legal works, philosophy and warfare); these disparate fields provide different patterns of occurrence for the particular linguistic pattern that is the focus of this project (the combination of two nouns). Thus, the SCC provides a reliable representation of this feature. While the SCC corpus overall covers 17 text types, as mentioned previously, only four text types are included in the early Archaic Chinese portion. *Shang Shu, '*The Classic of History', a historical text, is a compilation of documentary records related to events in China's ancient history. The documents and speeches collected in *Shang Shu* were written by rulers and officials from before and during the early Zhou period (1029 – 771 BC) and are the best examples of very early Chinese language use (Hu & McLaughlin 2007).

Table 7: Texts and text types in early Archaic Chinese (from Hu & McLaughlin 2007)

Title of the book/text	Author	Time	Genre
The Classic of History 《尚書》		4th century or earlier	history
The Book of Lord Shang 《商君書》	Shang-Yang 商鞅	390BC - 338BC	legal works
The Doctrine of the Mean 《中庸》	Kong-Zi (Confucius) 孔子	551BC - 479BC	philosophy
The Great Learning 《大學》	Kong-Zi (Confucius) 孔子	551BC - 479BC	philosophy
The Analects 《論語》	Kong-Zi (Confucius) 孔子	551BC - 479BC	philosophy
Mencius 《孟子》	Meng-Zi (Mencius) 孟子	372BC - 289BC	philosophy
The Classic of the Tao and Its Virtue 《道德經》	Lao-Zi 老子	770BC - 476BC	philosophy
The Arts of War 《孫子兵法》	Sun-Wu 孫武	6th century BC	warfare

Shang Jun Shu, 'The Book of Lord Shang', a legal work written around the 3rd century BC*,* is regarded as a foundational work of Chinese legalism, and records the theories and specific measures of the Shang Yang Reform. The book includes a large number of ordinances, essays, and courtly petitions attributed to various reformers, as well as discourses delivered at the court. Its purpose was to maintain societal order through written laws that meted out strict punishments, or rewards, for the actions of citizens.

In the philosophy category, five works from three different authors (*Kong-Zi, Meng-Zi* and *Lao-Zi*) have been included; these works provide rich sources of language use in the early Archaic Chinese period. *Dao De Jing,* 'The Classic of the Way and Its Virtue', by *Lao-Zi* (770 BC- 476 BC) is viewed as one of the core texts for both philosophical and religious Daoism in the Chinese way of thinking. *Lun Yu,* 'The Analects', recorded Confucius' teachings and conversations with his disciples in the 5th century BC. Both *Zhong Young* 'The Doctrine of the Mean' and *Da Xue* 'The Great Learning' were written by *Kong-Zi,* and discuss the Confucian philosophy of life. 'The Great Learning' enlightens people how to achieve a gateway to perfect virtue, while 'The Doctrine of the Mean' teaches readers how to conduct themselves in the world. *Meng Zi* (Mencius) by *Meng-Zi* (372 BC - 289 BC) is a record of anecdotes and conversations on topics in moral and political philosophy, between Mencius and various rulers.

The text of the *Sun Zi Bing Fa* 'The Art of War', written around 771 – 476 BC, is the earliest Chinese text on military matters, and has been an influential work on military

strategy. Having a different focus to any of the other texts, it offers a useful additional window to observe the appearance of N_1-N_2 sequences.

According to the SCC corpus, all of the sample texts discussed above are non-literary writing (Hu, McLaughlin & Williamson 2007:420). They are valuable resources for any examination of the language used in routine situations in ancient times. They comprise sufficient evidence for addressing the questions about language change posed in this study.

3.2 Criteria for identifying compounds

In Chapter 2, different criteria for distinguishing words and phrases in Mandarin Chinese were discussed. Using Chao's (1968) formulation of the criteria, this section establishes a general working definition of compounds. This allows us to delimit the current data and to differentiate compounds and phrases that consist of the same two forms.

Chao (1968:361) provided the most fundamental criterion, the "Freedom of Parts", that is generally applied to distinguish words from phrases. According to the "Freedom of Parts", if one component of an expression is a bound form, then, regardless of whether the construction is subordinative or coordinative, the expression should be considered a word. For example, both *fei-ji* fly-machine 'airplane' and *fang-ke* house-customer 'tenant' are compound words because one component of each expression (e.g., *-ji* in *fei-ji* and *fang-* in *fang-ke*) is a bound morpheme. This criterion is generally valid: if the combination is to be a compound, then either both, or one, of its constituents have to be of a bound form (Lu 1975).

Where an expression consists of two free forms, such as *shou-jiao* hand-feet 'movement', additional criteria need to be applied, according to Chao (1968:360). He proposed that for such an expression to be a compound, it should meet one of the following conditions:

(17)

A. The combination of the two morphemes is voiced in a neutral tone.
B. The constituents of the combination are inseparable from each other
C. The internal structure is exocentric
D. The meaning of the combination is not a simple synthesis of the meanings of its two constituents ('semantic non-compositionality'),

(Chao 1968: 360)

Consider criterion (17A) first. While this is valid in Modern Chinese, it may not reliably identify compounds in Archaic Chinese, especially for this study. Archaic Chinese was not a

tonal language; tones were only introduced after the simplification of the phonological system in the Medieval Chinese period. It is thus impossible to specify whether any part of the two combined forms has a neutral tone or not, without knowledge of how the sounds were produced in Archaic Chinese.

Secondly, let us consider Chao's criterion (17B) that the constituents of the combination cannot be separated by the insertion of a third morpheme. This is considered a useful test to differentiate compounds and phrases. The *de*-insertion, whereby the particle *de,* 'of', is inserted between a modified-modifier structure (such as A-N and N_1-N_2), works particularly well for nominal combinations (Lu 1964:21).

(18)

a. 手腳
shou *jiao*
hand feet
'movement'

b. **shou* *de* *jiao*
hand of feet

(19)

a. 大腳
da *jiao*
big feet
'big feet'

b. *da* *de* *jiao*
big of feet
'big feet'

The insertion test shows us that *de* cannot be inserted between *shou* and *jiao* in (18b) and thus this expression can be considered as a compound; conversely *da-jiao* in (19), which does allow the insertion, is a phrase. An additional condition for the *de*-insertion test is that, according to Zhang (1992: 52), if an expression consists of an A-N or N_1-N_2 nominal, the structure cannot be changed into (A de N) or (N_1 de N_2), respectively, when it occurs in the accusative position.

(20)

a. 牛 肉
niu rou
ox meat
'beef'

b. *niu de rou*
ox of meat
'ox's meat (beef)'

(21)

a. 買 一 斤 牛 肉
mai yi jin niu rou
buy one jin niu rou
'to buy 500 grams of beef'

b. **mai yi jin niu de rou*
buy one jin niu of rou
'to buy 500 grams of ox's meat (beef)'

The additional condition of the *de*-insertion test explains the seeming contradiction between (20a) and (21a), in that the expression *niu de rou* is acceptable in (20b), but is odd in (21b), where it occurs in the accusative position. Therefore, the additional condition needs to be met when the *de*-insertion test is applied.

Considering the internal structure (17C) criterion proposed by Chao, an exocentric structure is one in which "the syntactic form class of the head of the compound is not the same as that of a phrase in which the compound occurs" (Feng 1998:201). The class of the expression below (which is a noun), from Chao (1968:362), is not the same as the class of its head (a verb) and thus the expression *tian fang* is a compound.

(22) *tian fang* (VO→N)
fill room
'second wife (to a widower)

However, *tian-fang* can also be perceived as an idiom, which is not a word, and idioms are also opaque. Thus Chao's (17C) is an inappropriate criterion.

Taking the combinations of two N_1-N_2 as an example, although this combination of two nouns can be formed nominal compounds, verb compounds or adjective compounds, over 90 % of these combinations would be nominal compounds as mentioned in section 2.3.

(23) $N_1 N_2 \rightarrow$ Noun
(24) $N_1 N_2 \rightarrow$ Verb
(25) $N_1 N_2 \rightarrow$ Adjective

Since the focus of the current study is the combinations of N_1-N_2, the additional condition (17C) from Chao (1968:362), may not be related to the current study.

Finally, let us consider semantic non-compositionality (17D) suggested by Chao (1968: 363). In Chao's criteria, if the meaning of the whole combination is simply synthesised from the meaning of its constituents, then the expression is a phrase; otherwise, the combination is a compound as in (26).

(26)

水 土
shui *tu*
water soil
'environment'

汝 平 水 土
ru *pin* *shui* *tu*
you improve water soil
'improve the environment'

(*The Classic of History*)

(27)

車 馬
ju *ma*
carriage horse
'carriage'

百 姓 聞 王 車 馬 之 音
bai *xing* *wen* *wang* *ju* *ma* *zhi* *yin*
hurdred surname hear king carriage horse of sound
'people hear the sound of the king's carriage'

(*Mencius*)

In (27), it is unclear whether the combination *ju-ma* is a phrase or a compound, given that the meaning of the whole combination retains that of one of its constituents. To solve the

problem, Feng (1998:204) suggested the following additional criteria for semantic non-compositionality.

(28) In any context, if X and Y are free forms of a combination X Y, and if the semantic interpretation of X is "x" and Y is "y", and if the meaning of the X Y combination (<XY>) is one of the following forms:

a. <X Y> = x (first constituent of X Y),
b. <X Y> = y (second constituent of X Y), or
c. <X Y> = z (neither X or Y)

then the combination of X Y is a compound. (Feng 1998:204)

According to (28), therefore, *ju-ma* in (27) is a compound.

Based on the discussion above, a simple set of criteria for identifying compounds in N_1-N_2 combinations in Archaic Chinese can now be formulated.

For any given N_1-N_2 combination in Archaic Chinese to be a compound word,

(29) at least one of N_1-N_2 is a bound form

(30) if both N_1-N_2 are free forms, then,

a. in the structure of the coordinative combination, the semantic non-compositionality criterion (28) can be applied
b. in the structure of the subordinative combination, *de* cannot be inserted between N_1-N_2, when located in the accusative position, and the semantic non-compositionality criterion (28) also applies.

The application of these criteria in sequential order will facilitate the delimitation of the data to be examined in the current study. It will enable the differentiation of compounds and phrases that contain the same combination of two syllables.

3.3 Problems in collecting data

Although a computerised linguistic database is often a very useful resource, it can sometimes be suboptimal when Chinese texts are studied, particularly when *Wenyanwen* is involved. As mentioned in Chapter 2, the Chinese writing system is characterised by a lack of any boundary between words; this is evident in most Chinese digital corpora. Although the SCC corpus annotates various recurrent linguistic patterns or categories (as shown in Appendix 1) in the chronological texts, the lack of word boundaries means that some

patterns are not clear. For example, a disyllabic word in Mandarin is usually derived from two monosyllabic characters or morphemes as its core semantic components. But disyllabic words are not directly identified by the SCC tagging system because it does not show whether a particular two-character sequence is a word or a phrase. In addition, the fact that a morpheme may have multiple roles in Mandarin Chinese morphemes (discussed in Chapter 2) can lead to further challenges when trying to categorise a particular two-character sequence as a compound word, a phrase or a derivative. It should be noted, however, that differentiating between derivatives and compounds may be not necessary to answer the first research question of this study.

A notable example of a morpheme with multiple roles in Mandarin Chinese is *zi* 子, which has many forms and meanings. It can be a measure word in the expression (31) and a pronoun in (32):

(31) 擔 子
dan *zi*
load-MEASURE zi
'load'

一 擔 子 米
yi *dan* *zi* *mi*
one load zi rise
'a pole-load of any weight or size of rise' (Chao 1968:239)

(32) 你
ni
you
'you'

子何恃而往?
zi *he* *ci* *er* *wan?*
you how depend and go
'what will you go with?' (*Weixue*)

In most cases, *zi* 子 functions either as a bound morpheme or a derivational affix (although all affixes are also bound morphemes), depending on the context it occurs in. This can be illustrated briefly by the following expressions:

(33) 父子
fu *zi*
father children
'father and children'

父子不相見
fu *zi* *bu* *xiang* *jian*
father children not mutually see
'father and children do not see each other' (*The Mencius*)

(34) 妻子
qi *zi*
wife children
'wife'

妻子好合
qi *zi* *hao* *he*
wife child good marry
'good marriage with a nice wife' (*Shijing.Tangdi)*

While the *-zi* in (33) performs as a bound morpheme in the sequence *fu-zi* 父子'father and children', the *-zi* in (34) acts as a derivational affix, such that the nominal noun *qi-zi* 妻子 'wife' is derived from the bound morpheme *qi-* 妻. Even though both (33) and (34) have the same morpheme *-zi*, they are semantically different. In addition, as also the case in English, nouns in Chinese can be derived from verbs by attaching a suffix *-zi*, such as *pian*-zi 騙子 'swindler'.

Furthermore, an N_1-N_2 combination can sometimes function as a compound word in one context, but acts as a phrase in another. Such usages can coexist in the same time period or even in the same text. For example:

(35)

a. 天地孰得?
tian *di* *shu* *de*
heaven earth which suitable
'the natural order and geography are most favourable to whom?'
(*The Art of War*)

b. 天地之大也
tian *di* *zi* *da* *ye*
heaven earth of big modal-PARTICLE
'this is such a big word' (*The Doctrine of Mean*)

As discussed in (15), *tian-di* in (35a) and (35b) both occurred in early Archaic Chinese. *tian-di* in (35a) is a phrase since (i) its meaning is a simple synthesis of the meanings of its two constituents; and (ii) a conjunction could be inserted between its two parts. On the other hand, *tian-di* in (35b) is a compound, because its meaning cannot be inferred from its constituents and a word cannot be inserted between them.

As is clear from the above examples, classifying data without reference to the context could lead to inaccuracies. Therefore, in this study, it is important to manually classify data by applying the criteria established in section 3.2.

3.4 Data collection and analysis procedures

This study retrieved all N_1-N_2 sequences (tokens) in the SCC corpus by using its search and analysis tool's parts of speech (POS) tag (for nouns), and then further categorised the tokens into seven periods, from 1200 BC to 1911 AD (as shown in Chapter 2.1). The frequency of these nominal N_1-N_2 combinations occurring across the seven periods is given in Table 8.

Table 8: Distribution of the nominal N_1-N_2 combinations occurring in seven periods in the SCC

Time frame	Early Archaic	Later Archaic	Early Medieval	Middle Medieval	Later Medieval	Early Modern	Later Modern	total
Token	4804	3565	4699	3662	5945	12931	4335	39941

In order to accurately trace the origins of the use of N_1-N_2 disyllabic words in Chinese history, the analysis focused on the 4804 N_1-N_2 tokens that occurred in early Archaic Chinese (see Appendix 2). This will help address a major aim of this study, that is, the discovery of the origin of disyllabic words in the earliest period of Chinese history covered by the SCC corpus.

The first step was to filter the 4,804 N_1-N_2 tokens for unique values that display the various types of the nominal N_1-N_2 forms. The frequencies of all the N_1-N_2 types were then entered in an Excel spreadsheet. This filtering procedure produced 1,807 N_1-N_2 types from 4804 N_1-N_2 tokens and these 1,807 N_1-N_2 types were organised from the most frequent to the least frequent (see Appendix 2). This process identified the most frequent N_1-N_2 forms in the early Archaic Chinese period of the SCC corpus, and the top 299 were chosen for further linguistic analysis (see Appendix 3). Taken together, these 299 N_1-N_2-sequence types occurred a total of 2,643 times, thus making up 54.97% of all occurrences; also, each type of N_1-N_2 sequence occurred at least four times. There is a connection between lexicalisation

and frequency: the higher the frequency of the N_1-N_2 sequence, the more likely it is to be lexicalised (Fernández-Domínguez 2010:193 & 202).

The SCC tagging system is effective for identifying honorifics and proper nouns (including dynasties, places, personal names and titles of books) as these are singled out as individual categories in the SCC corpus (Appendix 1 from N-NI to N-NO). Therefore, this study first identified, and removed, these honorific and proper nouns so as to reduce the amount of manual analysis required. Each of the remaining N_1-N_2 sequences was then manually classified by applying the criteria for identifying compounds established in section 3.2.

These first steps provide a general picture of the frequency and nature of different linguistic types of N_1-N_2 sequences in early Archaic Chinese. These data will now serve as the basis of the main aim of this study, to identify the origin, and track the development, of disyllabic words in the history of Mandarin Chinese.

The next step is to categorise each of the derivatives and compounds obtained from the previous steps into particular (i) derived words, according to their derivational affixes attached, such as *-zi* and *-zhe*; and (ii) compound words, according to the syntactic relations between constituents, such as subordinative or coordinative. The purpose of this more detailed classification is to discover derivational and compounding patterns that emerged in early Archaic Chinese. This step will help to identify the linguistic rules or patterns that disyllabic words followed and which were the most favoured patterns. Once these patterns have been identified, words that follow them can be singled out for more in-depth analyses.

The final step is to focus on the diachronic development of the N_1-N_2 disyllabic words, including derivational N_1-N_2 words and compound N_1-N_2 words (as identified in the previous steps), across the seven periods of Chinese history. This will help to determine the developmental patterns of disyllabic words, including derivatives and compounds, and to identify the favoured word formation processes involved in disyllabic word formation in the history of Mandarin Chinese. This is central to addressing the research questions of this project.

In summary, this chapter introduced the SCC online corpus and discussed its benefits and limitations. It has proposed criteria for identifying compounds. This was followed by an explanation of how the data would be processed before being analysed with respect to the research questions. The results of the analysed data will be presented in the next section.

4. Results

This chapter reports the results from the data analysis mentioned earlier. Sections 4.1 and 4.2 present the results of studies of the distribution of, respectively, N_1-N_2 phrases and N_1-N_2 disyllabic words, and word formation patterns, in early Archaic Chinese. Section 4.3 reports the findings from an examination of the diachronic development of the use of the same N_1-N_2 disyllabic words and their patterns.

4. 1 Distribution of N_1-N_2 combinations in early Archaic Chinese

This section reports the distribution of N_1-N_2 phrases and N_1-N_2 disyllabic words that together made up the 299 N_1-N_2-sequence types (out of 1807 N_1-N_2-sequence types) that were obtained from the most frequently occurring types of N_1-N_2-sequence in the earliest period in the SCC corpus (Appendix 2 & 3).

Table 9 shows the number, and percentage, of types and tokens for each N_1-N_2 sequence in the early Archaic Chinese texts of the SCC corpus. Table 9 classifies the personal nouns as an individual N_1-N_2 form. Personal nouns are excluded from the N_1-N_2 disyllable category because Chinese personal names usually consist of a surname (which is often a single word, although there are a small number of disyllabic surnames, such as *Ouyang* 歐陽 and *Situ* 司徒) and a given name, which can occur as one or two characters. If a given name is constructed of two characters, these two characters can be either the components of a disyllabic word (which is a compound) or two single words. Although Chao (1968:514-516) held the view that personal names in Chinese should be considered as compounds, in most cases, the given names in Chinese are semantically transparent (Kałużyńska 2015:107).

Table 9: Distribution of 299 N_1-N_2 forms in early Archaic Chinese

N_1-N_2-sequence		**types**	**ratio %**	**tokens**	**ratio %**
Phrasal N_1-N_2		60	20.07	479	18.12
Disyllabic N_1-N_2		200	66.89	1702	64.40
	derivational N_1-N_2	42	14.05	463	17.52
	compounding N_1-N_2	158	52.84	1239	46.88
Personal name		39	13.04	462	17.48
Total		299	100	2643	100

For the sake of comparing words formed from different processes, such as derivation and compounding, the N_1-N_2 disyllabic words (excluding personal nouns) were subdivided into two groups, N_1-N_2 derivatives and N_1-N_2 compounds. Figure 3 uses the data of Table 9 to present the frequency of tokens and types in a more graphical manner.

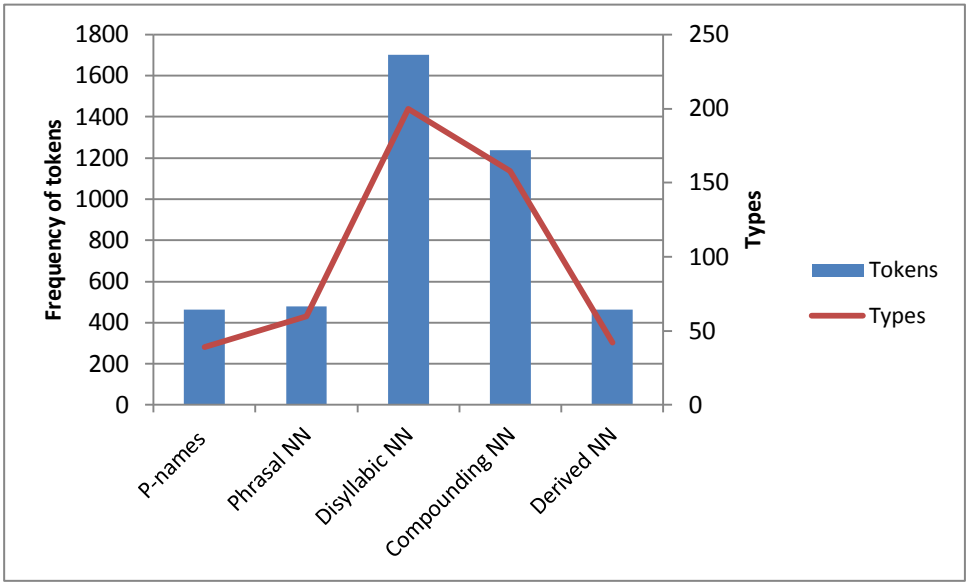

Figure 3: Distribution of 299 N_1-N_2 forms in early Archaic Chinese

As indicated in Table 9, there exists a variety in the distribution of the 299 N_1-N_2 sequence samples between their different forms in the earliest period of the SCC corpus. These forms include phrases, derivatives, compounds and personal names. In general, with regard to the sequence of N_1-N_2 whether they are individual words or phrases, N_1-N_2 disyllabic words emerge much more frequently than the N_1-N_2 phrases in both their token (1702 versus 479) and type frequencies (200 versus 60), with 64.4% versus 18.12% of the total 2643 N_1-N_2-sequence tokes and 66.89% versus 20.07% of the total 299 N_1-N_2-sequence types.

With respect to word formation processes, it was found that on the whole, both tokens and types of N_1-N_2 compounds were much more common than those of N_1-N_2 derivatives. The frequency of compounds is approximately 2.7 times that of derivatives, with regard to tokens, and three times, with regard to types. This outcome indicates that compounding processes were the most favoured ways of forming disyllabic words in the early Archaic Chinese period.

Due to the high frequencies of the N_1-N_2 disyllabic words in the Early Ancient period, this study now focuses on a more detailed investigation, that is the patterns of the N_1-N_2 disyllabic words.

4. 2 Distribution of word formation patterns in early Archaic Chinese

This section presents a more detailed investigation of the distribution of word formation patterns. These patterns comprise the derivational structures with suffixes *-zi* and *-zhe*, and the compounding structures with subordinative, coordinative, and [NX] and [XN] patterns in early Archaic Chinese.

The type, token and percentage of each N_1-N_2 disyllabic pattern are shown in Table 10. The percentages were calculated by dividing the number of types and tokens by the total number of N_1-N_2 derivatives and the total number of N_1-N_2 compounds, respectively, in order to compare their patterns under the same word formation process (e.g., derivation and compounding). The proper nouns shown in Table 10 include titles of books, and names of dynasties and counties.

Table 10: Distribution of different patterns of N_1-N_2 disyllabic words in early Archaic Chinese

disyllabic N_1-N_2	**forms**	**types**	**ratio %**	**tokens**	**ratio %**
derivation	*-zi* 子	6	14.29	181	39.09
	-zhe 者	36	85.71	282	60.91
	total derivation	**42**		**463**	
compounding	proper nouns	15	9.49	104	8.39
	coordinative	43	27.22	298	24.05
	subordinate	100	63.29	837	67.56
	total compounding	**158**		**1239**	
	[NX] [XN]	**59**	**37.34**	**543**	**43.83**
	[-ren] [-人]	15	9.49	163	13.16
	[ren-] [人-]	2	1.27	9	0.73
	[min-] [民-]	3	1.90	18	1.45
	[-min] [-民]	9	5.70	55	4.44
	[guo-] [国-]	2	1.27	22	1.78
	[-guo] [-国]	5	3.16	38	3.07
	[fu-] [夫-]	2	1.27	14	1.13
	[-fu] [-夫]	4	2.53	36	2.91
	[dao-] [道-]	1	0.63	5	0.40
	[-dao] [-道]	3	1.90	16	1.29
	[bai-] [百-]	6	3.80	91	7.34
	[qian-] [千-]	4	2.53	27	2.18
	[wan-] [万-]	3	1.90	49	3.95
Total		**200**		**1702**	

As seen in Table 10, 17 different patterns of N_1-N_2 disyllable formation were found in the 200 N_1-N_2 disyllables; these patterns include the two derivational suffixes *-zi* and *-zhe;* and compounding with subordinative, coordinative and [NX] and [XN] patterns, the latter being of 13 different types. It is noteworthy that some of the N_1-N_2 compounds used in compiling Table 10 may have been classified in more than one way. For example, the word *nong-min* 農民 farming-people 'peasant' is classified as a subordinative compound in terms of syntactic relation between two elements of the compound, while it is also classified as the pattern [N-*min*, N-民] pattern in terms of it having frequently occurred. It may be argued the [NX] and [XN] patterns classified as compounding here could have been classified as derivational. While this point has some merit, it not relevant to the current study, because all affixes are bound morphemes.

In terms of derivational processes, *-zi* and *-zhe* are the main derivational suffixes to form N_1-N_2 derivatives in early Archaic Chinese. In particular, *-zhe* is extremely common, found in 85.71% of types and 60.91% of tokens, in comparison to the derivational suffix *-zi* (14.29 % and 39.09 %, respectively).

With regards to the syntactic relationship between the constituents of N_1-N_2 compounds, including coordinative and subordinative compounds, subordinative compounds were the more common form (63.29% of types, and 67.56% of tokens) compared to the coordinative (27.22% and 24.05%, respectively).

With respect to [NX] and [XN] patterns, the pattern [N-*ren*], [*bai*-N], and [N-*min*] were used more frequently than [*dao*-N], [*fu*-N], [*ren*-N], and [*guo*-N] with respect to both their types and tokens. The pattern [*bai*-N] showed the second highest frequency in tokens, although its types do not match the same level of occurrence. Looking at [NX] and [XN] patterns as a group, they comprised 37.34% of all types and 43.83% of all tokens.

In order to build a picture of the development of disyllabic words in the history of Mandarin, the same N_1-N_2 disyllabic samples are now closely examined, according to the frequency of N_1-N_2 disyllabic words.

4. 3 The diachronic development of N_1-N_2 disyllables

This section focuses on the historical patterns of the use of certain N_1-N_2 disyllabic words already identified in the early Archaic Chinese texts of the SCC corpus, and the use of various patterns by which disyllabic words were formed.

Table 11 shows the tokens and frequencies, across all time periods of the SCC, of a set of N_1-N_2 disyllabic words that originated in early Archaic Chinese (as described in section 4.2). As also shown in the table, these consist of N_1-N_2 derivatives and N_1-N_2 compounds. The frequencies represent the average number of N_1-N_2 occurrence per ten thousand Chinese characters in the SCC corpus, a measure that was used because of the different sizes of the sub-corpora (as shown in the *Character* column).

Table 11: Distribution of N_1-N_2 disyllables and patterns in all time periods in the SCC

Period	**Time frame**	**Character**	**N_1-N_2 derivative (42 types)**			**N_1-N_2 compounds (158 types)**			**N_1-N_2 disyllables**	
			token	**frequency**	**%**	**token**	**frequency**	**%**	**token**	**frequency**
Archaic Chinese	1200 BC – 206 BC	72130	463	64.19	27.2	1239	171.77	72.8	1702	235.96
	206 BC – AD 220	37540	105	27.97	21.04	394	104.95	78.96	499	132.92
Medieval Chinese	AD 220 – AD 581	42450	68	16.02	20.54	263	61.96	79.46	331	77.98
	AD 581 – AD 979	40740	77	18.90	22.38	267	65.54	77.62	344	84.44
	AD 860 – AD 1368	64310	65	10.11	16.84	321	49.91	83.16	386	60.02
Modern Chinese	AD 1368 – AD 1644	130240	103	7.91	14.51	607	46.61	85.49	710	54.52
	AD 1644 – AD 1911	45280	14	3.09	6.92	188	41.52	93.08	202	44.61
Total		432690	895			3279			4174	

* Frequencies have been converted into frequencies per 10K words. Percentages represent the proportion of disyllabic words that were either derivatives or compounds.

Figure 4 uses the data of Table 11 to present the frequency of tokens in a more graphical manner. The time periods appear on the x-axis and the count per ten thousand characters on the y-axis:

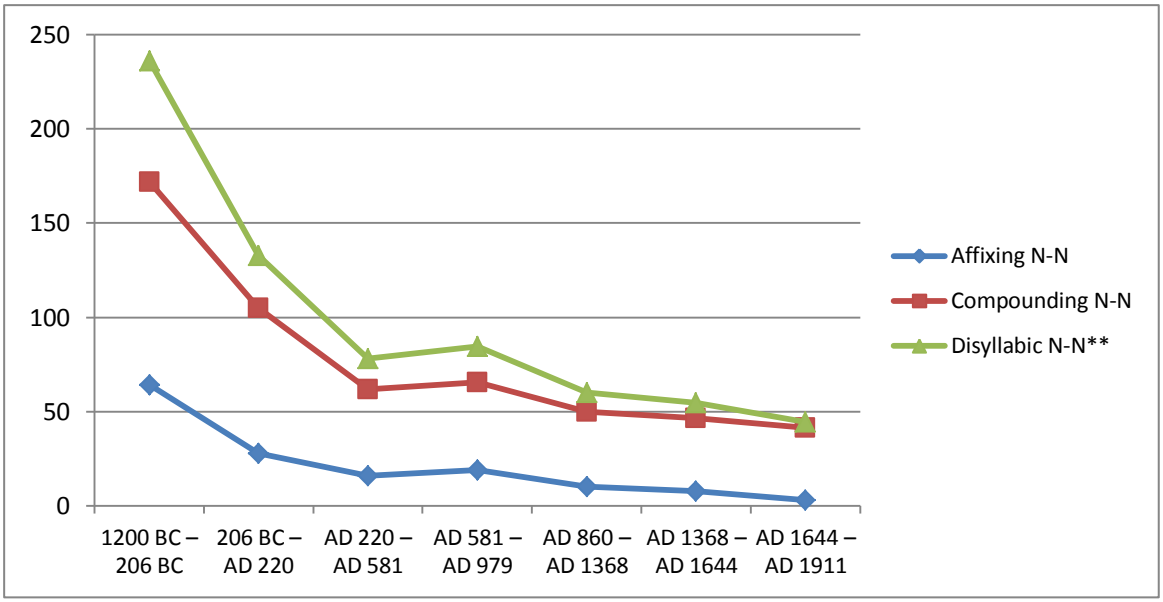

Figure 4: Distribution of N_1-N_2 disyllables and patterns in all time periods in the SCC

* Frequencies have been converted into frequencies per 10K words. **N_1-N_2 disyllables are made up by N_1-N_2 derivatives and N_1-N_2 compounds

4.3.1 Distribution of N_1-N_2 disyllabic words in the SCC

Table 11 and Figure 4 show that frequency of the set of N_1-N_2 disyllabic words identified in early Archaic Chinese texts decreased sharply from 235.96 (frequency per 10,000 characters) in the early Archaic Chinese period to 77.98 at the beginning of the Medieval Chinese period. After this sharp decrease, while the use of N_1-N_2 disyllabic words increased slightly in a short period from AD 581 – AD 979, it declined gradually from a frequency of 88.44 in Middle Medieval Chinese to 44.61 in later Modern Chinese.

4.3.2 Distribution of N_1-N_2 derivatives and compounds in the SCC

As well as showing the historical pattern of the entire set of N_1-N_2 disyllabic words, Figure 4 also shows the pattern of the larger set's component sub-sets, i.e., the N_1-N_2 derivatives and N_1-N_2 compounds. As was the case for N_1-N_2 disyllables overall, the frequencies of both N_1-N_2 derivatives and N_1-N_2 compounds also decreased, from 64.19 and 171.77 (frequency per 10,000 characters) in early Archaic Chinese to 3.09 and 41.52 in later Modern Chinese, respectively. It can be seen from Figure 4, while N_1-N_2 derivatives and N_1-N_2 compounds have been in co-occurrence in all time periods in the history of Mandarin Chinese, N_1-N_2 compounds have consistently been more common than N_1-N_2 derivatives. Figure 4 shows a downward trend for the occurrence of N_1-N_2 derivatives and N_1-N_2 compounds from the first

time period (1200 BC – 206 BC) to the third period (220 – 581). During these time periods, the occurrences of N_1-N_2 derivatives and N_1-N_2 compounds are distributed 27.2% versus 72.8% in early AC and about 20% versus 80% for both later Archaic Chinese and early Medieval Chinese. This downward trend (from early Archaic Chinese to early Medieval Chinese, 1200 BC – 581 AD) was followed by an upward trend in the next time period (middle Medieval Chinese, 581 – 979). After that, the downward trend was re-established (from later Medieval Chinese to later Modern Chinese, 860 – 1911). By the end of Modern Chinese, the ratio of distribution between the occurrences of N_1-N_2 derivatives and N_1-N_2 compounds is 6.92% versus 93.08%.

The overall trends of the occurrences of derivatives and compounds in Figure 4 are themselves comprised of several disparate trends. The trends of the N_1-N_2 derivatives that are formed by individual affixes, and of the N_1-N_2 compounds that are constructed by various syntactic patterns, are now analysed more deeply.

4.3.3 Distribution of suffixes *–zi* and *–zhe* in the SCC

Table 12: Distribution of N_1-N_2 derivatives by word formation structures across all time periods in the SCC

	Time frame	Characters	derivational *-zi*			derivational *-zhe*		
			token	frequency	%	token	frequency	%
Archaic Chinese	1200 – 206 BC	72130	181	25.09	39.09	282	39.10	60.91
	206 BC – AD 220	37540	61	16.25	58.10	44	11.72	41.90
Medieval Chinese	220 – 581	42450	48	11.31	70.60	20	4.71	29.40
	581 – 979	40740	55	13.50	71.43	22	5.40	28.57
	860 – 1368	64310	27	4.20	41.54	38	5.91	58.46
Modern Chinese	1368 – 1644	130240	86	6.60	83.44	17	1.31	16.56
	1644 – 1911	45280	9	1.99	64.40	5	1.10	35.60
	Total	432690						

* Frequencies have been converted into frequencies per 10K words. Percentages represent the proportion of all derivatives that were either derivational *-zi* or derivational *-zhe*.

Figure 5 is produced using the data of Table 12; the results are shown with the time periods on the x-axis and the count per ten thousand characters on the y-axis:

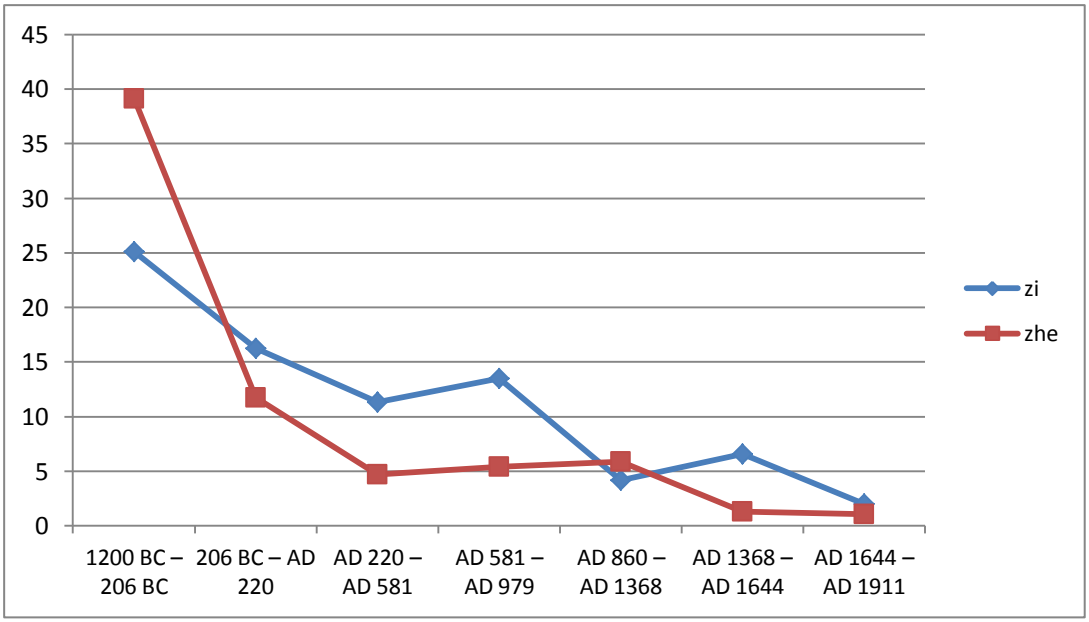

Figure 5: Distribution of N_1-N_2 derivatives across all time periods in the SCC

While Figure 5 shows that the derivational suffixes *-zi* (N-*zi*) and *-zhe* (N-*zhe*) have coexisted from early Archaic Chinese to later Modern Chinese, a large proportion of their occurrences in the texts are found in early Archaic Chinese period: 60.91% of N-*zhe*, in early and middle Medieval Chinese 70.60% and 71.43% of N-*zi* and in early Modern Chinese 83.44 of N-*zi*. N-*zi* and N-*zhe* words are almost evenly distributed in later Archaic Chinese (N-*zi*, 58.1%; N-*zhe*, 41.9%) and later Medieval Chinese (N-*zi*, 41.54%; N-*zhe*, 58.46%).

The finding suggests that the general occurrence of both N-*zi* and N-*zhe* words declined during the history of Mandarin Chinese, as documented in the SCC corpus. Although N_1-N_2 derivatives predominated over N-*zhe* words in early Archaic Chinese, the opposite pattern was observed in all subsequent periods. A detailed discussion and explanation of these phenomena will be provided in the next chapter.

4.3.4 Distribution of N_1-N_2 compounds in the SCC

Figure 6 is also produced using the data of Table 13. The results are shown with the time periods on the x-axis and the count per ten thousand characters on the y-axis:

As can be seen in Figure 4, the use of N_1-N_2 compounds largely stabilised from the beginning of the Medieval Chinese period after having a decline in the Archaic Chinese period. The overall trend of the N_1-N_2 compounds in Figure 4 is comprised of the trends of subordinative and coordinative compounds. Figure 6 shows a more detailed examination of these trends.

Table 13: Distribution of N_1-N_2 compounds by word formation structures across all time periods in the SCC

	Time frame	Characters	subordinative			coordinative			[NX] [XN]	
			token	frequency	%	token	frequency	%	token	frequency
Archaic Chinese	1000 – 206 BC	72130	754	104.53	71.67	298	41.31	28.33	543	75.28
	206 BC – AD 220	37540	200	53.28	58.65	141	37.56	41.35	116	30.90
Medieval Chinese	220 – 581	42450	136	32.04	67.33	66	15.55	32.67	60	14.13
	581 – 979	40740	126	30.93	56.76	96	23.56	43.24	104	25.53
	860 – 1368	64310	101	15.71	48.10	109	16.95	51.90	88	13.68
Modern Chinese	1368 – 1644	130240	221	16.97	54.17	187	14.36	45.83	215	16.51
	1644 – 1911	45280	44	9.72	49.44	45	9.94	50.56	30	6.63
	Total	432690								

* Frequencies have been converted into frequencies per 10K words. Percentages represent the proportion of all compounds that were either subordinative compounds or coordinative compounds.

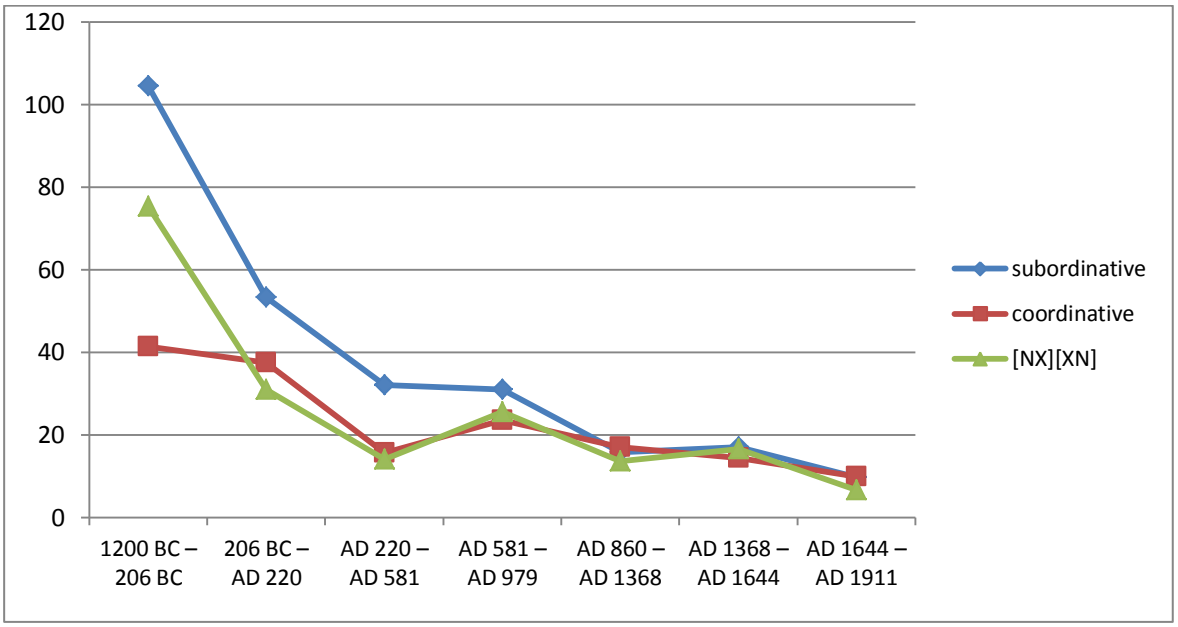

Figure 6: Distribution of N_1-N_2 compound structures across all time periods in the SCC

Figure 6 shows that subordinative and coordinative compounds also always coexisted across the seven periods of the SCC corpus. However, a large proportion of them occurred in early Archaic Chinese texts (71.67% of subordinative N_1-N_2 words in the SCC) and in early Medieval Chinese (67.33 % of subordinative N_1-N_2 words). While subordinative compounds

predominated over coordinative compounds in early Archaic Chinese, the two types are almost evenly distributed from the middle of Medieval Chinese to the end of Modern Chinese (56.76% versus 43.24% in middle Medieval Chinese, 48.10% versus 51.90% in later Medieval Chinese, 54.17% versus 45.83% in early Modern Chinese, and 49.44% versus 50.56% in later Modern Chinese). These results will be discussed further in the next chapter.

Figure 6 also shows that using [NX] and [XN] patterns were used to form disyllabic words across all time periods in the SCC. As was the trend for subordinative N_1-N_2 words, the [NX] and [XN] patterns were commonly used to form N_1-N_2 compounds in early Archaic Chinese (a frequency of 75.28 per 10,000 characters), but there was a rapid decline in early Medieval Chinese (frequency of 14.13). There was a much more gradual decline from the middle of the Medieval Chinese period to the end of the Modern Chinese period.

4. 4 Summary of the results

This chapter has, by analysing the SCC corpus, demonstrated that disyllabic nominal words with an N_1-N_2 sequence originated in early Archaic Chinese. It has been shown that, in this period, these N_1-N_2 disyllabic words were formed structurally, not from random combinations, but from derivational, syntactic and [NX] and [XN] patterns. The study has revealed that the N_1-N_2 disyllabic words were much more frequently formed by compounding than by derivation. Also, the derived N-*zhe* words and subordinative N_1-N_2 words occurred more frequently than the derived N-*zi* words and coordinative N_1-N_2 words correspondingly. The study has also found that, across all the time periods covered by the SCC, the frequency of occurrence of a set of N_1-N_2 disyllabic words identified in early Archaic Chinese, declined. The rates of decline of the overall set, and of its component subsets relative to each other, showed some variation from one period to another.

A general discussion and explanation of the results of the occurrences of N_1-N_2 disyllabic words in early Archaic Chinese, of the various forms of these N_1-N_2 disyllabic words, and their corresponding long-term historical patterns, will be provided in Chapter 5.

5. Discussion

This chapter provides a detailed discussion of the results and an explanation for the various phenomena related to the origin of N_1-N_2 disyllabic words in early Archaic Chinese, the processes by which they were formed and the historical development of these words from early Archaic Chinese to later Modern Chinese. The discussion is organised according to the study's research questions. Section 5.1 presents a discussion of the origin of N_1-N_2 disyllabic words in early Archaic Chinese. Section 5.2 discusses the various patterns by which N_1-N_2 disyllabic words in this same period. Section 5.3 discusses diachronic investigations of N_1-N_2 disyllabic words.

5.1 The existence of N_1-N_2 disyllables in early Archaic Chinese

The current study hypothesised that a wide range of disyllabic words existed in early Archaic Chinese (before 206 BC), and this claim was substantiated by the results of the analysis of N_1-N_2 expressions in the SCC corpus. The findings demonstrate that N_1-N_2 disyllabic words occurred much more frequently than N_1-N_2 phrases in early Archaic Chinese. Of 2,643 N_1-N_2 expressions, 64.4% were N_1-N_2 disyllabic words and only 18.12% were N_1-N_2 phrases. In this analysis, personal names had been excluded from both of these expression types, as explained in section 4.1. This study also found that many types of N_1-N_2 disyllabic words were used in the period: 299 types of N_1-N_2 expressions were identified, of which more than 66% were N_1-N_2 disyllabic words. This indicates that N_1-N_2 disyllabic words are a highly productive linguistic form, given that it is generally held that the higher the number of types, the more word formation processes are involved (Fernández-Domínguez 2010:198).

These results (as shown in Table 9) are significant as they reveal the existence of N_1-N_2 disyllabic words in the early archaic period regardless of their types or frequencies, or whether they were formed through derivation or compounding. As mentioned in the literature review, there are various forms of disyllabic words, such as derivational and compounding, depending on which affixes are attached to form derivatives, and which classes of constituents are involved to form different classes of compounds, such as N_1-N_2, A-N, N-V and V_1-V_2. The existence of one category can be logically inferred from its sub-categories. That is, if any forms of derivatives (i.e.,N-affix) or any types of these compounds

(i.e., N_1-N_2) existed in Archaic Chinese, disyllabic words are therefore generally perceived to have originated in the early archaic period.

This finding differs from previous scholars' contentions that monosyllabic words predominated in Archaic Chinese, and that compounding processes were, accordingly, also extremely rare (Karlgren 1926, Li 1993, Li and Thompson 1981, Pulleyblank 2000, Wang 1957). While Li (1993:130), for example, reported that the frequency of compounds in Archaic Chinese was only 0.32%, in this study, which employed a more reliable methodology, the frequency was 17.18 %, as shown in Table 11. This study has therefore been unable to support the proposals of Karlgren (1926), Pulleyblank (2000), Wang (1957) and Li (1993), that the use of compounding was a direct result of the simplification of the sound system in the Medieval Chinese period, leading to great number of potentially confusing homophones. The results of this study indicate, however, that many compound words coexisted together with the rich sound system of the early Archaic Chinese period. This suggests that compounding in early Archaic Chinese was not the result of the inadequate sound inventory, given that it is well established that the simplification of sounds only began in the Medieval Chinese period.

In addition, these results argue against the proposal of Pulleyblank (2000), that the morphological processes of word formation in Archaic Chinese were restricted to the changing of consonants or vowels within a monosyllable. It is clear that other processes, e.g., derivation or compounding, were also available to reduce the possible confusion due to homophonous syllables. While we cannot say that the frequency of compounding is completely unrelated to the richness of any given period's sound systems, neither does it appear that the sound simplification is the main driver of compounding, as proposed by Li (1993) and Wang (1957).

Some earlier studies have also reported that compounding existed in Archaic Chinese (Cheng 1981, Li 2009, Guo 1994, Jin 2017, Kennedy 1951, Ma 1981, Sagart 1999, Tang 2007, Tao 1996, Wu 2001, Wang 2017, Yu 1990). However, our findings do not support those studies' findings that reduplication was the main compounding process in Archaic Chinese as proposed by Wu (2001), or that proper nouns were the only word types compounded, as suggested by Jin (2017), Tao (1996), Wang (2017), and Yu (1990).

In short, our findings contribute to the literature by enriching our understanding of the origins of disyllabic words in early Archaic Chinese, demonstrating that they were much more common than previously thought, possibly because they were previously misclassified as phrases.

5.2 The existence of various compounding patterns

As to whether N_1-N_2 disyllabic words were formed structurally or randomly in early Archaic Chinese, the findings indicate the former: that a variety of word formation rules or patterns, including derivation, compounding, and [NX] and [XN] patterns, were involved. Certain processes and patterns in early Archaic Chinese were found to be more commonly used. First, the findings display distinctive differences in the occurrence of certain suffixes. As shown in Table 10, the use of the suffix *-zhe* is six times higher than that of *-zi*, as indicated by types, and one and a half times as frequent, in terms of tokens. What is interesting in this data is that the percentage of its types, 85.71%, does not keep the same pace with the percentage of its frequencies, 60.91%. The high type frequency of the suffix *-zhe* indicates that this process had high productivity in early Archaic Chinese. This result also confirms that there was a strong positive correlation between the richness of types and productivity; as proposed earlier, productivity is associated with a high ratio of types : tokens (Fernández-Domínguez 2010:210 & 215). While this is the first direct comparison between *-zi* and *-zhe*, *-zhe* itself is considered by many authors (Chao 1968:221, Packard 2000:73, Wu 2001:280) to be one of the most fertile affixes or bound morphemes, even if there is disagreement as to which of these categories *-zhe* belongs to.

Second, the results also demonstrated that subordinative and coordinative N_1-N_2 compounds occurred very commonly in early Archaic Chinese, comprising about 90% of all N_1-N_2 compounds. These findings generally corroborate the results of Cheng (1981), Feng (1989), Liu (2003) and Wu (2001), who suggested that subordinative and coordinative relations were the most favoured types of compound formation in early Archaic Chinese. In this study, subordinative structures were the much more common type, accounting for approximately 65% of all N_1-N_2 compounds (compared to 25% for coordinative compounds), as indicated by the numbers of both types and tokens. As such, the results from the SCC suggest that subordinative and coordinative compounds, at least of the N_1-N_2 type, are more frequent than reported previously. Wu (2001:349) observed that 50% of the compounds in

Shijin are subordinative and coordinative compounds, with about 30% being subordinative and 20% coordinative (as mentioned in the literature review, section 2.3).

While both the current study and that of Wu (2001:349) focused on the same period, i.e. early Archaic Chinese, the syntactic structures examined are not the same. Wu's analysis included all types of compounds, including: verb + verb (V_1-V_2), adjective + adjective (A_1-A_2), adjective + noun (A-N), and even subject + predicate (S-P), verb + resultative complement (V-R) and verb + object (V-O). In the current study only N_1-N_2 disyllabic words were included. Another possible explanation for the different findings is that the criteria applied for identifying compounds used in the two studies may differ. For example, in the current study, duplicated disyllabic words, such as *ren-ren* 人人 people-people 'everyone', were excluded. In Wu's study, not only were they included, but they represented the most common form of disyllabic words (Wu 2001).

Finally, the current study found that a great number of compounding words were formed structurally from [N-X] and [X-N] patterns in early Archaic Chinese. The results suggest that words with [N-X] or [X-N] structures were used more regularly than coordinative compounds in both types (37.34% versus 27.22%) and tokens (43.83% versus 24.05%) as shown in Table 10. The most common individual pattern (of the [N-X] or [X-N] form) in early Archaic Chinese was [N-*ren*] (N-人): it represented 30.02% (163/543 as shown in Table 10) of all the occurrences of [N-X] words and [X-N] words. This finding corroborates the ideas of Wu (2001:284), who suggested that the [N-*ren*] construction was the most productive way to form words in Archaic Chinese. The findings also suggest that while the patterns [N-X] and [X-N] coexisted in the Early Archaic Chinese period, interestingly, for any given morpheme that occurred in both patterns (e.g. [N-*ren*] and [*ren*-N]), the former pattern (e.g. [N-*ren*]) was always the more common.

It is difficult to explain this phenomenon, but it might be related to the syntactic relations between the two components. There are two ways in which the components of [N-X] and [X-N] words can relate to each other: (i) modifier-modified (subordinative structure); (ii) parallel (coordinative structure). The pattern [N-X] can be applied to form subordinative compounds (sub-[N-X]), such as *di-ren* 敵人 enemy-ren 'enemy', *zhong-ren* 衆人 all-ren 'everybody', or coordinative compounds (coor-[N-X]), such as *ren-min* 人民 people-people 'people'. As already noted, in early Archaic Chinese, N_1-N_2 disyllabic words were much more

likely to be formed by subordinative, rather than coordinative, structures. By extension, disyllabic words of the subordinative [N-X] (sub-[N-X]) or the subordinative [X-N] (sub-[X-N]) structure are much more likely than those of the coordinative [N-X] (coor-[N-X]) or the coordinative [X-N] (coor-[X-N]) form. For the sake of clarity, this is illustrated in the following formulas:

(36)

a. sub-[N-X] > coor-[N-X]
b. sub-[X-N] > coor-[X-N]

(Note: ">" means much frequent than)

The formulas demonstrate a key relation between sub-[N-X] and sub-[X-N], and make it easier to address the difficulty of the explanation. It is likely that the pattern sub-[N-X] occurred more frequently than the pattern sub-[X-N] because words that are most frequently used as content words have a strong tendency to be a head of that word which is modified members (Wu 2000:284). The component X is, therefore, the head of sub-[N-X]; likewise, the head of sub-[X-N] is the component N. If the latter one exists, the [X-N] should be coor-[X-N] because there are no two heads in subordinative compounds. Therefore, (36a) and (36 b) cannot co-exist. Thus,

(37) sub-[N-X] > sub-[X-N]

In summary, the evidence indicates which were the preferred methods or patterns by which N_1-N_2 disyllabic words were formed in early Archaic Chinese. The evidence also confirms that N_1-N_2 disyllabic words were not formed simply by the random concatenation of two morphemes or nouns, and were by no means serendipitous. The results provide further support for the hypothesis posed at the beginning of this study, that N_1-N_2 disyllabic words in early Archaic Chinese were formed by structurally diverse processes. Thus, they corroborate the results of some previous work on a similar topic, which suggested that archaic Chinese compounds were formed in accordance with the rules of syntax (Feng 1989, Wu 2001:284).

5.3 Diachronic investigations of N_1-N_2 disyllabic words

Regarding the diachronic development of N_1-N_2 disyllabic words in Mandarin Chinese, this was studied by analysing the occurrence, in later period of history, of a set of N_1-N_2 disyllabic words that had been identified in the early Archaic Chinese texts. Interestingly, and somewhat unexpectedly, instead of these words being used more frequently with time, there was a marked decline. At first sight, these results might seem to be at odds with previously published studies (Liu 2003:107, Feng 1998:208), that have suggested compounding increased markedly in the medieval Chinese period; in the following sections some reasons for this apparent disagreement are suggested.

5.3.1 Explaining the declining tendency in N_1-N_2 words

The diachronic patterns of N_1-N_2 disyllabic word occurrence, across all time periods of the SCC, are shown in Figure 4. It is clear that that was a steep decline immediately after the early Archaic period, and a gradual decline from the middle Medieval period onwards. The occurrences of both N_1-N_2 derivatives and N_1-N_2 compounds, as reflected in the percentage of words found (per 10,000 characters) was high in the early Archaic period but fell off sharply during the later Archaic period. This is probably due to these N_1-N_2 disyllabic words recurrently being used or the small number of other N_1-N_2 disyllabic words was created. There was no apparent decline for all types of compounds, in the following period, and then the frequencies of N_1-N_2 derivatives and N_1-N_2 compounds dropped again from later Medieval Chinese to later Modern Chinese. Thus, instead of indicating anything about the frequency of compounding generally, the results suggest that some of the specific N_1-N_2 disyllabic words that originated in early Archaic Chinese gradually declined.

Regarding this decline, there are several possible reasons. Firstly, it might be, in part, a methodological artefact. While, as discussed in the methodology section, all sample texts in the SCC are examples of non-literary writing (Hu, McLaughlin & Williamson 2007:420), this is not sufficient to ensure uniformity of language use across the corpus. Across the various time periods covered by the SCC, there is an uneven representation of genres amongst the sample texts of each period. For example, all of the texts that belong to the Legal Works and Warfare genres are located in the early Archaic Chinese period. Likewise, of the six samples of the philosophy genre, five are from the early Archaic Chinese period. The language used in each of these diverse fields has its own distinct terminologies, and thus the uneven

representation of genres could lead to differing patterns of word occurrence. This can be illustrated by consideration of the sample text, *The Art of War*《孫子兵法》. *wei-di* 圍地, *qu-di* 衢地, *jiao-di* 交地, *zheng-di* 爭地 and *zhong-di* 重地 are words that describe locations in military contexts, and it is possible that they might have been used in Medieval or Modern Chinese. But, the SCC is unlikely to reveal this, because none of the SCC texts from these periods belongs to the military genre. The uneven distribution of the sample genres of the SCC means we must be careful when interpreting the apparent historical decline of the N_1-N_2 disyllabic words examined in this part of the study.

A second reason for the decreased occurrence is that the N_1-N_2 disyllabic words identified in early Archaic Chinese may in fact have either disappeared or declined in use in the Modern Chinese or even the Medieval Chinese periods. They may have been replaced by new words with similar meanings due to social and cultural changes. The language used at any point in history has a close connection with the developmental stage of the language speakers' culture and political systems. This is particularly true for a language with a 4000-year history, such as Mandarin Chinese. For instance, *xian-wang* 先王 before-king 'former king', *wang-zhe* 王者 king-suffix-*zhe* 'king', and *huang-di* 皇帝 emperor-emperor 'emperor' all refer to emperors, kings or rulers in the history of China. *xian-wang* and *wang-zhe* generally refer to several famous emperors in the Ancient times (before West Chou 771 BC), such as the *xian-wang* only refers to Yao 堯, Shun 舜, and Yu 禹 before Xia dynasty (before 2000 BC), and the *wang-zhe* generally refers to Tang 湯 in Shang dynasty (1675 BC – 1029 BC) and Wen 文 and Wu 武 in West-Chou (1029 BC – 771 BC) (Cheng et al. 1998:1693). The word *huang-di*, in contrast, did not occur until the imperial systems had been established and the first emperor *Qin Shi Huang* took the title *huang-di* 皇帝 'emperor' in 221BC. Apart from *xian-wang*, *wang-zhe* and *huang-di*, there are other words that also refer to emperors in Chinese history, such as *sheng-ren* 聖人 sage-people 'sage' and *tian-zi* 天子 heaven-son 'the emperor'; as to which was used, this was influenced by context and the position of the speaker or writer.

Similarly to *xian-wang*, the word *zhu-hou* 諸侯 every-marquis 'marquis' specifically referred to a person who was the governor of a vassal state in the Ancient times, and as such it is also the product of early Archaic Chinese. It was commonly used before *Qi Shi Huang*

united all of China under one rule, but not after this, since there were no longer any vassal states.

The examples above illustrate how social changes could result in the decreasing occurrence of certain N_1-N_2 disyllabic words. A detailed list of occurrences of the words associated with emperors and the like, as detected in the SCC, is provided below.

Table 14: Historical decline of certain N_1-N_2 words that occurred less frequently after Archaic Chinese

N_1-N_2 word		Early Archaic Chinese	Later Archaic Chinese	Early Medieval Chinese	Middle Medieval Chinese	Later Medieval Chinese	Early Modern Chinese	Later Modern Chinese	Total
shen-ren 'sage'	聖人	71	14	4	3	9	2		103
zhu-hou 'marquis'	諸侯	39	37	3	2	1			82
xian-wang 'king'	先王	34	18	2		4			58
jun-zi 'monarch'	君子	107	28	2	23	2	7	1	170
fu-zi 'teacher'	夫子	34	1		26	5		1	67
wang-zhe 'king'	王者	21	4	1		2		1	29
tian-zi 'emperor'	天子	16	31	31	4	4	6	2	94
total		322	133	43	58	27	15	5	603
10K		44.64	35.43	10.13	14.24	4,20	1.15	1.1	

In summary, the uneven distribution of the sample genres in the SCC corpus, and changing word usage associated with cultural and political systems are possible reasons for the observed decrease in occurrence of certain disyllabic words that originated in early Archaic Chinese.

5.3.2 More frequent use of N_1-N_2 compounds and N-*zhe* derivatives

Although the occurrence of all types of N_1-N_2 disyllabic words decreased across the periods studied here, N_1-N_2 compounds were less affected than N_1-N_2 derivatives (Table 11). In fact, while the percentage representation of the two types was 72.8 : 27.2 (compounds : derivatives) in early Archaic Chinese, in later Modern Chinese it was 93.08 : 6.92. N_1-N_2 derivatives might be more likely to be lost, because, after the early Archaic period, derivational affixes were sometimes substituted, by bound morphemes, to form compounds with similar meanings.

For example, from this study's data, the bound morphemes *-fu* 夫 'man' and *-min* 民 'people' can be substituted for the suffix *-zhe* in the early Archaic derived word *nong-zhe* 農

者 farming-person ‘peasant’ to create the new compounding words, *nong-min* 農民 farming-people ‘peasant’ or *nong-fu* 農夫 farming-husband ‘peasant’. These new compounds maintain a similar meaning as the derived *nong-zhe*. While the suffix *-zhe* in early Archaic Chinese appears to be a highly versatile morpheme, and thus can be replaced by many other morphemes, as suggested by Chao (1968:221), the affix *-zi* is less versatile, and is mainly used to nominalise a bound morpheme, thus forming a nominal compound (Chao 1968:237), such as *qi-zi* 妻子 wife-suffix-*zi* ‘wife’ and *jun-zi* 君子 monarch-suffix-*zi* ‘gentleman’. This phenomenon could also account for the uneven distribution of the derived N-*zhe* words and N-zi words in all time periods of the SCC corpus, as indicated in Table 10 and Figure 3.

5.3.3 Changes in the occurrence patterns of sub-N_1-N_2 and coor-N_1-N_2 compounds

In this sub-section, the focus will be on a significant change in the occurrences of suborinative N_1-N_2 words and coordinative N_1-N_2 words between early Archaic and Modern Chinese (indicated in Table 13 and Figure 6). First, however, it will be helpful to discuss a set of the most frequent subordinative and coordinative compounds in the SCC corpus (as shown in Table 15). In particular, the compounds of most interest are those that occurred more frequently after the middle Medieval Chinese period.

It is important to understand the internal relations between two elements of a compound semantically and grammatically in order to explain this distributional pattern of change. Apart from the factors explained in section 5.3.1and 5.3.2, another reason for this change may be the degree of lexicalisation, that is, to what extent the constituents of a compound word remain semantically transparent. In a semantically transparent compound word, the meaning may be inferred from the meanings of its constituents. In the examples below however, in words such as *zuo-you* 左右 left-right ‘influence’, *zi-ran* 自然 oneself-way ‘nature’, and *bai-xing* 百姓 hundred-surname ‘common people’ in (15b), (15e) and (15o), each compound is semantically opaque. The meaning of each compound cannot be understood by consideration of constituents, regardless of whether they are exhibit a hierarchical modifier-modified relation (e.g., *bai-xing* 百姓 ‘common people’) or a parallel relation, with synonyms (e.g., *zi-ran* 自然 ‘nature’), or antonyms (e.g., *zuo-you* 左右 ‘influence’). In each case, the meaning of a morpheme within the compound is not associated with the independent meaning of that same morpheme. Such words show the

highest degree of lexicalisation; conversely, it can be said that they exhibit the weakest connection between the word and its constituents (Packard 2000:217).

Table 15: N_1-N_2 compounds that occurred more frequently after the middle Medieval Chinese period

	N_1-N_2	literal & meaning	Early Archaic Chinese	Later Archaic Chinese	Early Medieval Chinese	Middle Medieval Chinese	Later Medieval Chinese	Early Modern Chinese	Later Modern Chinese	total
a	*tian-di* 天地	heaven-earth 'world'	24	24	4	9	13	10	5	89
b	*zuo-you* 左右	left-right 'influence'	10	8	9	7	17	16	2	69
c	*fu-mu* 父母	father-mother 'parents'	21	6		7	6	20	4	64
d	*zhang-fu* 丈夫	a unit of length-husband 'husband'	7	2		11	8	35		63
e	*zi-ran* 自然	oneself-way 'nature'	4	3	1	3	18	6	8	43
f	*shang-xia* 上下	top-bottom 'world'	16	8	3	3	4	6	2	42
g	*ri-yue* 日月	sun-moon 'time'	9	10		9	7	4		39
h	*xiong-di* 兄弟	young brother-old brother 'fraternity'	6	3	4	3	1	15	1	33
i	*yi-f* 衣服	clothier-garment 'clothing'	4	1	1	2		19	5	32
j	*fu-fu* 夫婦	husband-wife 'couple'	5		1	4		12	2	24
k	*cao-mu* 草木	grass-trees 'plants'	6	5		1	2	4	4	22
l	*peng-you* 朋友	friend-friend 'friend'	14					5	2	21
m	*dao-ru* 道路	way-road 'path'	5		3	5	2	2	2	19
n	*zi-sun* 子孫	son-grandson 'descendant'	5	1		3	1	1	2	13
o	*bai-xing* 百姓	hundred-surname 'people'	38	14	11	23	3	5	4	98
p	*xiao-ren* 小人	somellness-person 'people'	25	5	1	5	6	40	4	86
q	*zhong-ren* 衆人	many-people 'people'	6			1	2	61	5	75
i	*wei-ren* 為人	act as-person 'behaviour'	9	12	8	5	1	8	7	50
s	*si-hai* 四海	four-direction 'world'	17	8	6	8	1	3	2	45
t	*wan-wu* 萬物	ten thousand-thing 'every things'	27	7		1	3	6		44
u	*bai-guan* 百官	hundred-official 'official'	11	8	3	4	7	3		36
v	*si-fang* 四方	four-direction 'world'	11	4	1	5	3	10	1	35
w	*ren-xing* 人心	people-heart 'feelings'	5	5	1	2	16	3		32
x	*qian-li* 千里	thousand-a unit of distance 'far away'	11	1		6	8	4	1	31

While highly lexicalised words are morphologically constructed as componential compounds, from a semantic perspective, they are more likely to be non-componential compounds (Libben 1998:38). A completely lexicalised word tends to be perceived as an individual unit rather than a combination of semantically transparent words (Libben 1998:36-39). If a compound is perceived as non-componential, its component generally cannot be easily replaced by another morpheme or constituent. That is, the internal structure of such a word is relatively stable, compared to that of a componential compound, and they are thus more likely to be used, across different time periods, with a consistent form and meaning. For example:

(38)

a. 聖 人 無 常 心, 以 百 姓心 為 心
sheng ren wu chang xin, yi bai xing xin wei xin
sage people no often heart, because hundred surname heart act heart
'the sage's impermanence is based on the hearts of the people'
(*The Classic of the Tao and Its Virtue. Before 500 BC*)

b. 百 姓 開了城 門
bai xing kai le cheng men
hundred surname open LE city door
'the people have opened the gates of the city'
(*The Scholars. 1701 -1754*)

c. 增加百姓 收 入
zeng jia bai xing shou ru
add plus hundred surname receive enter
'increasing people's income' (*Xinhuanet 2017*)

(39)

a. 丈夫生而 願 為之有室
zhang fu sheng er yuan wei zhi you shi
MEASURE husband born and willing act of have room
'a husband is willing to marry a wife after he was born.'
(*Meccius. Before 300 BC*)

b. 嫁 得此丈 夫

Jia de ci zhang fu

marry TENSE this MEASURE husband

'(she) married a husband.'

(*Two collections of striking the table in amazement. 1580-1644*)

c. 丈夫和妻子在家庭中的角色是什么？

zhang fu he qi zi zai j ia ting zhong de jia se

MEASURE husband and wife children at home yard of corner of colour is what such

'what is the role of husband and wife at home?'

(*Gotquestions 2019)*

(40)

a. 日 月忽 其 不淹 兮

ri yue hu qi bu yan xi

sun moon suddenly its no flood PARTICLE

time passes quickly and cannot stay long.

(*Qu-yuan. Before 278 BC*)

b. 日月 如梭

ri yue ru suo

sun moon like shuttle

'time elapse quickly'

(*Zaojunet 2016*)

The compound *bai-xing* 'common people' in (38a) occurred in *The Classic of the Tao and Its Virtue,* written by *Lao-Zi* around 500 BC. The same word with the same meaning in (38b) was used in the book of *The Scholars* by *Wu jing-zi*, 2,200 years later, and is still in use today (38c). Both (38a) and (38b) are found in the SCC corpus. Likewise, the word *zhang-fu* 丈夫 a unit of length-man 'husband' in (39) and *ri-yue* 日月 sun-moon 'time' in (40) also originated in the early Archaic period (39a) and (40a), were used in Modern Chinese (39b) and continue to be used in contemporary Chinese (39c) and (40b).

In addition, the type of grammatical relationship between the two components of a compound can also affect the distributional pattern change between subordinative N_1-N_2 words and coordinative N_1-N_2 words. When the two components are semantically synonymous (e.g., *xiong-di* 兄弟 'brother' in (15h), *fu-fu* 夫婦 'couple' in (15j), etc.) or antonymous (e.g., *tian-di* 天地 'world' in (15a), *zuo-you* 左右'influence' in (15b), etc.) the compounds have a strong tendency to be increasingly used, after they have originated (e.g.

Table 15, a-n). Such compounds have reduced semantical combinatorial compatibility, i.e. the individual constituents, when they in a synonymous or antonymous relationship, are less suitable to be matched with one another. In contrast, when the morphemes of a compound form a subordinative compound, with a hierarchical relationship between the two components, there are more ways in which matching can occur (e.g. Table 15, o-x).

A useful example of this is the entry *zuo* 左 'left', found in *Gu-dai han-yu ci-dian* 'Dictionary of Ancient Chinese' (Cheng et al. 1998:2084). In this text, twelve items with the pattern of [*zuo*-X] [左-X] are listed; in eleven of them, there is a hierarchical relationship between the two elements, whereas only one has a parallel relationship, the word *zuo-you* 左右 left-you 'influence', an antonymous combination. Similarly, the dictionary lists fifteen compounds under the entry *you* 右 'right', but none of these exhibits a parallel relationship between the components, and even the antonym *zuo* 左 'left' is excluded (1998:1907).

This example exemplifies another point: that if both constituents of a compound are in a parallel relationship, whether that be synonymous or antonymous, the difficulty of creating new matches is exacerbated by the impossibility of reversing the sequence of the two components (Chao 1968:372). This difficulty arises from the phonological tonal system: the sequence of two constituents in a parallel compound always follows the following order of tones: the level tone, the rising tone, the departing tone, and the entering tone (Ding 1975, Liu 2003:128, Liu 2002:48, Wu 2001:260). For example, the compound *zuo-you* 左右 has to be constructed such that the left side is the morpheme *zuo-* 左, it being a level tone, while the morpheme *-you* 右 must be on the right side, being a departing tone. In addition, social norms can influence the sequence of two constituents of a parallel compound (Wu 2001:260). For example, the sequence of components in the words *fu-qi* 夫妻 husband-wife 'husband and wife' and *jie-mei* 姐妹 old sister-young sister 'sisters' are restricted by gender issues (the subordination of women in Confucianism) and the family pecking order, respectively.

In the early Archaic Chinese period, there were only a few exceptions where we see the reversed and unreversed sequences, e.g. (41) and (42)

(41)

a. 室 家

Shi *jia*

bedroom family

'family'

(*Confucius*)

b. 家室

Jia *shi*

Family bedroom

'family' (*Shijing*)

(42)

a. 圖 書

tu *shu*

picture book

'publications'

(*Hanfeizi*)

b. 書圖

shu *tu*

book picture

'publications'

(*Hanfeizi*)

Such cases of variability in disyllabic compounds are generally considered to be restricted to compounds used in early historical periods (Feng 1997:223). However, a small number of compounds with interchangeable components co-exist today, as shown in (43), albeit this is not an important issue for the current study.

(43)

a. 兄 弟

xiong *di*

old brother young brother

'fraternity'

b. 弟兄

di *xiong*

young brother old brother

'fraternity'

In summary, it appears likely that semantical and grammatical relations between the two components of a subordinative N_1-N_2 or coordinative N_1-N_2 compound, and the degree of lexicalisation of the compound as a whole, have an impact on the changing distributional pattern of the occurrence of such compounds across Chinese history.

5.4 Summary of discussions

This chapter has provided a detailed discussion of the existence of N_1-N_2 disyllabic words in early Archaic Chinese and the linguistic patterns by which various types of such words are formed. Specifically, this chapter has described and explained the declining tendency of a set of N_1-N_2 disyllabic words during the historical periods covered in the SCC corpus. It has also explained why both compounding and use of the derived N-*zhe* form have been being frequently applied to form N_1-N_2 disyllabic words. Finally, this chapter has presented an account of how the occurrences of subordinative N_1-N_2 words and coordinative N_1-N_2 words have changed across Chinese history.

6. Conclusion

This research has presented a corpus-based study of the origin and development of N_1-N_2 words in the history of Mandarin Chinese. This focus emerged from hypotheses made in previous studies on Chinese morphological processes: that compounding occurred in the Medieval Chinese period due to the simplification of sound systems (Karlgren 1926, Li 1993, Pulleyblank 2000, Wang 1958); that non-concatenative operations or processes were only the means to form new words (Pulleyblank 2000); that reduplication was the main processes to derive compounds in Archaic Chinese (Wu 2001); and that proper nouns were the only type of compound to appear in the oracle bone inscriptions (Jin 2017, Tao 1996, Wang 2017, Yu 1990).

Returning to the research questions posed at the beginning of this study, they were:

Q1 Did disyllabic nominal words with N_1-N_2 sequences originate in Archaic Chinese? If so:

Q2 What patterns of N_1-N_2 words existed in this period?

Q3 If a set of nominal N_1-N_2 words can be identified in early Archaic Chinese, what were their frequencies and patterns when they were traced diachronically throughout Chinese history?

It is now possible to state that nominal words with N_1-N_2 sequences originated in early Archaic Chinese, and these N_1-N_2 words were structurally formed according to a wide range of linguistic rules. Nominal N_1-N_2 words identified in the early archaic period were found less frequently in later periods.

This chapter contains three sections. Section 6.1 summarises the findings of this study and offers conclusions based on the findings. Section 6.2 addresses the implications of this study, and section 6.3 notes the limitations of this work and discusses recommendations for future research.

6. 1 Summary of findings and conclusions

This study identified and examined 299 N_1-N_2 sequences in early Archaic Chinese texts from the SCC corpus. This corpus-based examination provides a representative view of the occurrence of N_1-N_2 compounds or phrases, and of the structural patterns of N_1-N_2 words in

early Archaic Chinese. By then examining the occurrence of this set of N_1-N_2 words across all the time periods covered by the SCC corpus, the study also provides insights into the changing usage patterns of N_1-N_2 words, and their linguistic patterns, from Archaic Chinese to Modern Chinese.

6.1.1 On the occurrence of the N_1-N_2 combination

The literature review of Mandarin compound word formation reviewed the arguments used by scholars who have claimed that Archaic Chinese was primarily monosyllabic. However, the results of this investigation show that, in early Archaic Chinese, N_1-N_2 disyllabic words occurred much more frequently than N_1-N_2 phrases, as indicated by the numbers of both tokens and types. In terms of word formation, N_1-N_2 compounds were much more frequent, in tokens and types, than N_1-N_2 derivatives. The conclusion that can be drawn from this research on the issue of the origin of compounding is that compounds originated in early Archaic Chinese rather than in Medieval Chinese, the common position of previous research (Karlgren 1926, Li 1993, Pulleyblank 2000, Wang 1958). Further, it indicates that compounding was not simply the consequence of the phonological simplification of the Medieval Chinese period, and, thus, that other processes must also have contributed to the use of compounding.

6.1.2 On the distribution of N_1-N_2 word patterns

The second major finding was that the N_1-N_2 words identified in the early Archaic period were formed through different processes. This finding differs from previous research that proposed that early Archaic words were formed solely by non-concatenative processes (Pulleyblank 2000), or, if concatenative processes were involved, they were of a basic nature (e.g. reduplication). Regarding derivational processes, this study has shown that *-zi* and *-zhe* were the main suffixes used to form N_1-N_2 derivatives in early Archaic Chinese, and that, in compounds, the two components of a word were most commonly related, syntactically, in a subordinative or coordinative relationship. This study found that the most common types of N_1-N_2 disyllabic words were those formed with the derivational suffix *-zhe*, and subordinative compounds. In addition, the investigation of N_1-N_2 compounds has shown that words formed by [N-X] and [X-N] patterns occurred frequently. Based on these results, the conclusion can be drawn that N_1-N_2 words in early Archaic Chinese were formed structurally,

by application of a range of syntactical rules; they were neither formed solely by non-concatenative processes, nor by random, or simple, reduplicative, concatenation of two morphemes or nouns.

6.1.3 On the diachronic development of N_1-N_2 words

By focusing on the later occurence of set of N_1-N_2 words identified in the SCC's early Archaic texts, this study has shown that the usage of these N_1-N_2 words decreased sharply before the early Medieval Chinese period, and continually declined from middle Medieval Chinese to later Modern Chinese, after increasing slightly in a short period from AD 581 – AD 979. While this does not support the prediction that the use of N_1-N_2 words increased across time , this study provided a range of possible explanations for the observed decline. These include artefactual concerns, such as the uneven representation of genres in the corpus, and linguistic drivers, such as historical changes in the cultural and political systems that differentially favour the persistence of certain words.

The data show that N_1-N_2 derivatives and N_1-N_2 compounds occurred in all time periods in the history of Mandarin Chinese. The proportional use of the derivational suffixes *-zi* and *-zhe* changed over time: the use of *-zhe* decreased after the early Archaic Chinese period, which itself might have led to the increased use of *-zi* increase. This suggestion is motivated by the fact that *-zhe* and *-zi* were, according to our data, the only derivational suffixes available for use. There was a similar change in the proportional use of subordinative compounds and coordinative compounds over time. Before the middle Medieval Chinese period, subordinative compounds were the more frequently used, while after the middle Medieval Chinese period, the use of subordinative compounds and coordinative compounds was almost evenly distributed. The conclusion that can be drawn from this research on diachronic development is that the occurrence of words of the N_1-N_2 words identified in early Archaic Chinese had significantly declined by the Modern Chinese period. Further, the proportional use of derivational N-*zi* forms and coordinative compounds increased during the same time-frame, while there was the corresponding decline in the use of derivational N-*zhe* forms and subordinative compounds.

6. 2 Implications of this study

The arguments of this study on the origin of Chinese compounds differ from previous examinations. First, in previous examinations, scholars argued that the origin of compounding was only the result of homophones, and homophones were only caused by the simplification of sounds in Medieval Chinese. In previous accounts, compounding and the rich sound inventory did not co-exist in the same period. This study has presented a new approach on N_1-N_2 disyllabic words: that compounding and the rich sound inventory co-existed in early Archaic Chinese. This new approach suggests that compounding coming into existence was therefore not caused by the simplification of sounds, though the simplification of the sound system may link to or accelerate the development of compounds. This approach therefore reached a new understanding that there were other processes that caused the development of compounding, not only phonological change.

Secondly, this study differs from previous research in its methodological approach to investigating the origin of compounding in Archaic Chinese. Previous studies either did not involve empirical research or collected data manually, which collects less data in comparison to automated processing. Conclusions or generalisations that were drawn from the lack of empirical data or limited data were less accurate. This study presented a unique empirical research work by investigating the combination of N_1-N_2 and linguistic patterns of N_1-N_2 words in early Archaic Chinese in the SCC corpus. Through the investigation, this study has demonstrated that digital corpus can contribute significantly to research on the origin and the development of the Chinese language. This study was able to examine the earliest texts in the history of the Chinese language for the discovery of the origin of N_1-N_2 words, and was able to obtain different types of N_1-N_2 words to demonstrate the origin of N_1-N_2 words being formed structurally through different word formation processes. Further, this study was able to collect a large number of data from different time periods for the purpose of understanding developmental patterns of the use of N_1-N_2 words in the history of Chinese. The conclusions that were drawn from this study were, therefore, more accurate in comparison to research with limited or no data.

6. 3 Limitations and future research

Although this study constitutes a representative investigation of N_1-N_2 disyllabic words in early Archaic Chinese, their historical usage, and their linguistic patterns, there are some limitations with the corpus used for this study.

As discussed in section 5.3 the sub-collections of the SCC corpus do not all contain similar samplings of the various genres found in the collection overall. Nevertheless, there is no other corpus available that provides a fully marked-up searchable system covering a similarly wide time frame. If a corpus were established with a much more even distribution of genres for each period, this would enable a more thorough investigation of N_1-N_2 disyllabic words in early Mandarin, and permit more robust generalisations about their diachronic development.

This study has provided evidence relevant to the three research questions in Chapter 1. However, the findings of this study also point to the need for a more in-depth investigation into the constituents of compounds, phonologically, morphologically, and semantically. Future research could usefully explore the following questions further.

Previous studies have claimed that the sequence of two constituents in a disyllabic word is subject only to: (i) the need to align with the permissible order of different tones; and (ii) various social norms. However, this claim cannot account for the compound *di(4)-xiong(1)* (where (4) and (1) indicate the entering tone and the level tone) 弟兄 'fraternity' from neither the perspective of the tones used nor the requirements of any social norms. Likewise, the proposal cannot help in compounds where the two constituents have the same tones, such as *dao(4)-lu(4)* 道路 'path'. Future research could pay attention on this issue.

Regarding the decreased occurrence of the identified N_1-N_2 compounds in later periods, future research could investigate if these compounds truly disappeared or were lexicalised into individual, simple words, such as happened with the English word *window*[3]. Further investigations could also look at whether the constituents of compounds were partially or completely lost, or if they changed their original state or function due to morphological processes.

Notes

1. *Wenyanwen* is also known as Classical Chinese or Literary Chinese. The former refers to the written form of the Archaic Chinese period of Chinese literature, from the end of the Spring and Autumn period (early 5th century BC) to the end of the Han dynasty (AD 220). The latter refers to the form of written Chinese used from the end of the Han Dynasty to the early 20th century. The difference between Classical Chinese and Literary Chinese is that Literary Chinese gradually absorbed elements of the vernacular languages. Both are referred to as *Wenyanwen* in the history of Mandarin Chinese.
2. *Baihuawen,* also known as Standard Written Chinese or Modern Written Chinese, refers to the vernacular written form of Chinese as opposed to the *Wenyanwen*, and it formally replaced *Wenyanwen* in the early 1920s.
3. The word *window* (which refers to a pane of glass or plastic in a house, car, or something you look through) was originally a compound word borrowed in Middle English *windoge* from Old Norse *vindauga (*literally: wind eye). Over time, the compound became dissociated from its components and ceased to be a compound. Thus, *window* was converted into a simple word in Modern English.

7. References

Arcodia, Giorgio Francesco. 2007. Chinese: A language of compound words. *Selected proceedings of the 5th décembrettes: Morphology in Toulouse* 79–90.

Bao, Huifang (暴慧芳). 2009. *汉语古文字合文研究* [The study on words-character in Ancient Chinese]. Southwest University 硕士. http://oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=2009104654.nh&DbName=CMFD2009&DbCode=CMFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnR3h5Mjd3WG1WVVF3czVVY244ZHJiND0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (17 September, 2018).

Baxter, William H. 2010. *A handbook of Old Chinese phonology*. Vol. 64. Walter de Gruyter.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Cheng, Fuhua, Yongan Chu, chengtao Guo, Tianxiang Jin, Lin Li & Xiaoxiang Wand. 1998. *Gu dai han yu ci* dian [Dictionary of Ancient Chinese]. 第 1 版.; Di 1 ban. Beijing: Shang Wu Yin Shu Guan.

Coblin, W. South. 2000. A Brief History of Mandarin. *Journal of the American Oriental Society* 120(4). 537–552. doi:10.2307/606615. http://www.jstor.org/stable/606615 (28 June, 2018).

Dai, Ligang (戴黎刚). 2017. 晚清的官话音系及其性质 [The Phonology of Mandarin in Late Qing Dynasty And Its Basic Dialect]. *Research in Ancient Chinese Language* (04). 28-41+103. http://oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=GHYY201704004&DbName=CJFDLAST2018&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1FhcEE0RVVzYTFBY3B4ZnV5enlQMTJDOWx5bXFYMD0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (19 October, 2018).

DeFrancis, John. 1984. *The Chinese language: fact and fantasy*. Honolulu: University of Hawaii Press.

Ding, Bang-xin. 1975. Tonal Relationship between the Two Constituents of the Coordinate Constructions in the Analects, the Meng-tze and the Book of Odes. *Bulletin of the Institute of History and Philology Academia Sinica* 17–52. doi:10.6355/BIHPAS.197512.0017. http://www.airitilibrary.com/Publication/alDetailedMesh?docid=10124195-197512-47-1-17-52-a.

Dong, Xiufang. 2012. Lexicalization in the history of the Chinese language. *Newest trends in the study of grammaticalization and lexicalization in Chinese* 235–274.

Duanmu, San. 1998. Wordhood in Chinese. *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* 135–196.

Duanmu, San. 2007. *Phonology of Standard Chinese*. Oxford, United Kingdom: Oxford University Press. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=415943 (29 September, 2018).

Feng, Shengli. 1998. Prosodic structure and compound words in Classical Chinese. In Packard Jerome L (ed.), *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese*. Walter de Gruyter.

Fernández-Domínguez, Jesús. 2010. Productivity vs. Lexicalization: Frequency-Based Hypotheses on Word-Formation. *Poznań Studies in Contemporary Linguistics* 46(2). 193–219. doi:10.2478/v10010-010-0010-x. https://www.degruyter.com/view/j/psicl.2010.46.issue-2/v10010-010-0010-x/v10010-010-0010-x.xml (19 February, 2019).

Geng, Zhensheng (耿振生). 1992. *Mingqing dengyunxue tonglun (明清等韵学通论)* [A study on Ming and Qing rhyme tables]. Beijing: Langue and Culture Press.

Gotquestions. 2019. 丈夫和妻子在家庭中的角色是什么? [What is the role of husband and wife in the family?]. *Got Questions Ministries*. https://www.gotquestions.org/Chinese/Chinese-roles-husband-wife.html (11 April, 2019).

Guo, Jingyun (郭静云). 2010. The God's Belief in Yin Shang and the New Explanation of the "Emperor". *Cultural Relics in Southern China* 2010(2). 63–67.

Haspelmath, Martin & Andrea Sims. 2010. *Understanding Morphology*. London, United Kingdom: Routledge. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=615894 (16 September, 2018).

Haudricourt, André G. 1972. *Problèmes de phonologie diachronique*. Vol. 1. Société pour l'étude des langues africaines.

Haudricourt, André-Georges. 1954. *De l'origine des tons en vietnamien*. Imprimerie Nationale.

Hu, X., J. McLaughlin & N. Williamson. 2007. Syntactic Positions of Prepositional Phrases in the History of Chinese: Using the Developing Sheffield Corpus of Chinese for Diachronic Linguistic Studies. *Literary and Linguistic Computing* 22(4). 419–434. doi:10.1093/llc/fqm017. https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqm017 (18 July, 2018).

Hu, Xiaoling & Jamie McLaughlin. 2007. Sheffield Corpus of Chinese. *University of Sheffield*. https://www.dhi.ac.uk/scc/db/scc/manual.html (14 July, 2018).

Huang, Shuanfan. 1998. Chinese as a headless language in compounding morphology. *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* 261–284.

Huang, Zhiqiang (黄志强). 1986. 西周, 春秋时代汉语构词法概论 [An introduction to the formation of Chinese in the Western Zhou dynasty and the Spring and Autumn period]. *Seeking Truth* 3.

Jin, Xinxin. 2017. The characteristics showed in compound graph marks of the inscription on bones or tortoise shells. *Journal of Huainan Normal University*.

Kałużyńska, Irena. 2015. Traditional Chinese Generation Names. *Onomastica* 59. 107–122. doi:10.17651/ONOMAST.59.6. http://rcin.org.pl/dlibra/publication/78088 (19 February, 2019).

Karlgren, Bernhard. 1926. *Philology and ancient China*. Vol. 8. H. Aschehoug & Company.

Kemmer, Suzanne. 2015. Words in English: Types of Word Formation. https://www.ruf.rice.edu/~kemmer/Words/wordtypes.html (5 August, 2018).

Kennedy, George A. 1951. The Monosyllabic Myth. *Journal of the American Oriental Society* 71(3). 161–166. doi:10.2307/595185. http://www.jstor.org/stable/595185 (14 August, 2018).

Kennedy, Graeme. 2014. *An introduction to corpus linguistics*. Routledge.

Lee, Kyookap (李圭甲). 2012. 甲骨文与金文及战国文字合文字形比较分析 [A Comparative Study of the Hewen (合文) on Structures of Oracle-bone Script and Bronze Inscriptions and the Warring States Writing]. *The Study of Chinese Characters* 45–53.

Lei, Liming (雷黎明). 2009. 论楚简合文 [Chujian Hewen]. *Journal of Ningxia University(Humanities & Social Sciences Edition)* 31(06). 7–12. http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=NXDX200906003&DbName=CJFD2009&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnR3h5Mjd3WG1WVVF3czVVY244ZHJiND0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (17 September, 2018).

Li, Charles N & Sandra A Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Li, Fengxiang. 1993. *A diachronic study of v-v compounds in Chinese*. United States -- New York: State University of New York at Buffalo Ph.D. http://search.proquest.com/docview/304112043/abstract/204A9C00368B41A3PQ/1 (19 August, 2018).

Li, Jian. 2013. *The rise of disyllables in old Chinese: The role of Lianmian words*. United States -- New York: City University of New York Ph.D. http://search.proquest.com/docview/1314417250/abstract/454A826CBA4F486BPQ/1 (21 August, 2018).

Li, Jing (李静). 2008. 东巴文合文研究 [Study on Combined Graphs of Dongba Characters]. *Lanzhou Academic Journal* (12). 196–199. http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=LZXK200812062&DbName=CJFD2008&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnR3g5Tjl3M2tGaURKQVZabndUdTNQYz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (18 September, 2018).

Li, Rulong (李如龙). 2009. 论汉语的单音词 [Monosyllabic Words in Chinese]. *Linguistic Research* (02). 1–7. http://oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=YWYJ200902002&DbName=CJFD2009&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnUzRJNlBSNTdjekVyd1BVUzgwR092ND0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (28 August, 2018).

Liao, Wei-wen Roger. 2014. Morphology. In Huang, Cheng-Teh James, Yen-hui Audrey Li, Andrew Simpson & John Wiley (eds.), *The handbook of Chinese linguistics*. Wiley Online Library.

Libben, Gary. 1998. Semantic Transparency in the Processing of Compounds: Consequences for Representation, Processing, and Impairment. *Brain and Language* 61(1). 30–44. doi:10.1006/brln.1997.1876. http://www.sciencedirect.com/science/article/pii/S0093934X97918769 (22 March, 2019).

Liu, Chenghui (劉承慧). 2003. 古漢語實詞的複合化 [On Compounding in Classical Chinese]. 古今通塞: 漢語的歷史與發展 (第三屆國漢學會議論文集・ 語言組)》(南港: 中央研究院語言學研究所籌備處), 107–140.

Liu, Chenghui. 2002. On the Origin and Early Development of Chinese Compounds. *National Tsing Hua University* 32(2). 469. http://nthur.lib.nthu.edu.tw/dspace/handle/987654321/75616.

Lü, Shuxiang (吕淑湘）. 1963. *現代漢語單雙音節問題初探* [A preliminary study of the problem of monosyllabism and disyllabism in modern Chinese]. Chinese Language.

Lu, Zhiwei (陸志韋). 1964. *漢語的構詞法 [Chinese morphology] Revised rdition*. Peijing: Kexue Chubanshe.

Lu, Zhiwei (陸志韋). 1975. *漢語的構詞法* [Chinese morphology]. Zhonghua Book Company.

Luo, Jierui & Zhulin Mei. 2004. Some Thoughts on the Early Development of Mandarin. *Dialect* (04). 295–300. http://eng.oversea.cnki.net.ezproxy.uws.edu.au/kcms/detail/detail.aspx?dbCode=cjfd&QueryID=52&CurRec=118&filename=FYZA200404002&dbname=CJFD2004&uid=WEEvREcwSlJHSldRa1FhdkJkVWI3b1ZxVFBnSFVwQjkxSkk4VDdiVG9MVT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!! (29 June, 2018).

Ma, Zhen (马真). 1981. 先秦复音词初探 (续完) [Polysyllabic Words in Pre-Qin (continued)]. *China Academic Journals (CD Edition) Electronic Publishing House*. http://oversea.cnki.net.ezproxy.uws.edu.au/kcms/detail/detail.aspx?recid=&FileName=BDZK198101008&DbName=CJFD1981&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnUzRJNlBSNTdjekVyd1BVUzgwR092ND0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (28 August, 2018).

Mair, Victor H. 1994. Buddhism and the Rise of the Written Vernacular in East Asia: The Making of National Languages. *The Journal of Asian Studies* 53(3). 707–751. doi:10.2307/2059728. http://www.jstor.org/stable/2059728 (8 August, 2018).

Mookherjee, Sheema. 2013. China's ancient capital rises again. *BBC*. http://www.bbc.com/travel/story/20131025-chinas-ancient-capital-rises-again (5 June, 2019).

Norman, Jerry. 1988. *Chinese*. Cambridge University Press.

Packard, Jerome L. 2000. *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge, United Kingdom: Cambridge University Press. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=201428 (30 July, 2018).

Pan, Yunzhong（潘允中). 1989. *汉语词汇史概要* [Chinese vocabulary history]. Shanghai Ancient Books Publishing House.

Pulleyblank, Édwin G. 2000. Morphology in Old Chinese. *Journal of Chinese Linguistics* 28(1). 26–51. http://www.jstor.org/stable/23754003 (18 August, 2018).

Qin, Zuxuan & San Duanmu. 2019. An acceptability judgment study of 1+1 NN compounds in Chinese. *Lingua*. doi:10.1016/j.lingua.2019.02.004. http://www.sciencedirect.com/science/article/pii/S0024384118305564 (3 April, 2019).

Ramsey, S. Robert. 1987. *The languages of China*. Princeton, N.J.: Princeton University Press.

Saenger, Paul. 1997. *Space Between Words: The Origins of Silent Reading*. Stanford University Press.

Sagart, Laurent. 1999. *Roots of Old Chinese*. Amserdam, Netherlands: John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=710266 (26 July, 2018).

Schuessler, Axel. 2006. *ABC Etymological Dictionary of Old Chinese*. Honolulu, United States: University of Hawaii Press. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=3413378 (23 August, 2018).

Schuessler, Axel. 2009. *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensa*. Honolulu, United States: University of Hawaii Press. http://ebookcentral.proquest.com/lib/uwsau/detail.action?docID=3413409 (5 April, 2019).

Tai, James HY & Marjorie KM Chan. 1999. Some reflections on the periodization of the Chinese language. *Studies in Chinese historical syntax and morphology: Linguistic essays in honor of Mei Tsu-lin* 223–239.

Tang, Yuanfa (唐元发). 2007. 先秦复音词的产生与发展 [The Formation And Development of Disyllabic and Polysyllabic Words in The Pre-Qin period]. *Journal of Zhejiang University of Technology (Social Science)* (03). 251–254. http://oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=SZGD200703002&DbName=CJFD2007&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnUzRJNlBYZ3B0ZUMrWWpuM1JYbTJOcz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (28 August, 2018).

Tang, Yuming. 1986. Jinwen fuyinci jianlun—jianlun hanyu fuyinhua de qiyuan 金文複音詞簡論—兼論漢語複音化的起源 [A brief discussion on polysyllabic words in the inscription on Ancient Bronze: with a discussion on the origin of polysillabization]. Renleixue lunwen xuanji 人類學論文選集 [A paper collection of anthropology]

Tao, Wang. 1996. Colour terms in Shang oracle bone inscriptions. *Bulletin of the School of Oriental and African Studies* 59(1). 63–101.

The Open University. 2009. Beginners' Chinese. *OpenLearn*. https://www.open.edu/openlearn/languages/chinese/beginners-chinese/content-section-0 (5 June, 2019).

The University of Sheffield. 2007. Sheffield Corpus of Chinese. *The Sheffield Corpus of Chinese*. https://www.dhi.ac.uk/scc/ (14 July, 2018).

Ting, Pang-Hsin. 2002. Morphology in Archaic Chinese. (Ed.) Laurent Sagart & Pang-Hsin Ting. *Journal of Chinese Linguistics* 30(1). 194–210. http://www.jstor.org/stable/23754869 (21 August, 2018).

Twitchett, Denis C., John King Fairbank & Herbert Franke. 1994. *The Cambridge History of China: Volume 6, Alien Regimes and Border States, 907-1368*. Vol. 6. Cambridge University Press.

Vincent, Nigel. 1978. Is sound change teleological. *Recent developments in historical phonology* 4. 409.

Wang, Li. 1958. Hanyu shigao [Lectures on Chinese]. *Beijing: Zhonghua Shuju*.

Wang, Jing (王景). 2017. 甲骨文"合文"现象初探 [Hewen in Oracle]. *Heihe Journal* 231(3). 39–40. doi:10.14054/j.cnki.cn23-1120/c.2017.03.018.

Wu, Zongwen (伍宗文). 2001. *先秦漢語複音詞研究* [Study on the Chinese polysyllabic words in Pre-Qin]. Bashu Publishing House.

Xiao, Richard & Xianyao Hu. 2015. Corpora and Corpus Tools in Use. *Corpus-Based Studies of Translational Chinese in English-Chinese Translation*, 37–66. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-41363-6_4. http://link.springer.com/10.1007/978-3-642-41363-6_4 (14 October, 2018).

Xinhuanet. 2017. 2017 年百姓"钱袋子"将有哪些新变化？ [What new changes will be made to the "money bag" of the people in 2017?]. *Xinhuanet*. http://www.xinhuanet.com//politics/2017lh/2017-03/06/c_1120578507.htm (10 April, 2019).

Xu, Weihan (许威汉). 1992. *汉语词汇学引论* [Introduction to Chinese Vocabulary]. The Commercial Press.

Yang, Paul Fu-mien. 1986. "The Portuguese-Chinese dictionary of Matteo Ricci: a historical and linguistic introduction." Proceedings of the Second International Conference on Sinology. Taipei: Academia Sinica. Vol. 1 of the Section on Language and Script, pp. 191-242

Yu, Suisheng. 1990. The origins of combining characters and phonograms of Oracle and Naxi Dongba. *Journal of the Central University for Nationalities* (01). 85–89. doi:10.15970/j.cnki.1005-8575.1990.01.020. http://gb.oversea.cnki.net/kcms/detail/detail.aspx?recid=&FileName=ZYMD199001019&DbName=CJFD1990&DbCode=CJFD&uid=WEEvREcwSlJHSldRa1Fhb09jSnZnR3g5Tm96bTQrZlp1SEhxejluZ2Y1Zz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!! (18 September, 2018).

Zaojunet. 2016. 造句 [Making sentences]. http://zaojv.com/2413647.html (6 June, 2019).

Zhang Jing. 1994. *Yu yan yu yong yu fa*. Zhengzhou shi: Wen xin chu ban she.

Zhang, Hongming. 1992. *Topics in Chinese phrasal tonology*. United States -- California: University of California, San Diego Ph.D. http://search.proquest.com/docview/303984748/abstract/47250B779B7649CBPQ/1 (6 June, 2019).

Zhang, Yulai (张玉来). 2010. 《中原音韵》时代汉语声调的调类与调值 [Tone Categories and Tone Pitches of Chinese in the Time of Zhongyuan Yinyun]. *Research in Ancient Chinese Language* (Research in Ancient Chinese Language) 87(2). 11–25. http://oversea.cnki.net.ezproxy.uws.edu.au/kns55/detail.aspx?QueryID=0&CurRec=12&dbcode=cjfd&dbname=CJFD2010&filename=GHYY201002003.

Zhou, Xiaolin, William Marslen-Wilson, Marcus Taft & Hua Shu. 1999. Morphology, Orthography, and Phonology Reading Chinese Compound Words. *Language and Cognitive Processes* 14(5–6). 525–565. doi:10.1080/016909699386185. http://www.tandfonline.com/doi/abs/10.1080/016909699386185 (14 October, 2018).

Appendix 1

Tag labels, word classes, and special categories

Tag label	Word class	Category
AJA	Adjective	Non_predicate (e.g. 溜清, 噴香)
AJB	Adjective	Non_predicate_AA (e.g. 薄薄, 蕩蕩)
AJC	Adjective	Non_predicate_AAB (e.g. 黯黯然)
AJD	Adjective	Non_predicate_AABB (e.g. 熟熟馴馴)
AJE	Adjective	Non_predicate_ABAB (e.g. 筍條筍條)
AJF	Adjective	Non_predicate_ABB (e.g. 酸蔭蔭)
AVA	Adverb	general (e.g. 約莫, 直截)
AVB	Adverb	AA (e.g. 常常, 往往)
AVC	Adverb	negative (e.g. 未, 休)
CJA	Conjunction	coordinating (e.g. 和, 但是)
CJB	Conjunction	subordinating (e.g. 假若, 因為)
CLA	Classifier	(e.g. 粒, 幅)
EPA	Expression	direction (e.g. 庵北, 其西)
EPB	Expression	formulaic (e.g. 端的, 不期)
EPC	Expression	genitive_zhi_N (e.g. 之理, 之屬)
EPD	Expression	genitive_zhi_suo_V (e.g. 之所恃)
EPE	Expression	location (e.g. 廳外, 崖下)
EPF	Expression	nominal (e.g. 吊孝的)
EPG	Expression	order (e.g. 吸前)
EPH	Expression	time (e.g. 嘉祐中, 慶曆中)
FMA	Functional_morpheme	adverbial (e.g. 地)
FMB	Functional_morpheme	aspect_durative (e.g. 着)
FMC	Functional_morpheme	aspect_experiential (e.g. 過)
FMD	Functional_morpheme	aspect_perfective (e.g. 了)
FME	Functional_morpheme	causative (e.g. 使)

FMF	Functional_morpheme	complement (e.g. 得)
FMG	Functional_morpheme	emphatic (e.g. 所)
FMH	Functional_morpheme	general (e.g. 聿)
FMI	Functional_morpheme	genitive (e.g. 之)
FMJ	Functional_morpheme	objective (e.g. 把)
FMK	Functional_morpheme	passive (e.g. 見, 被)
FML	Functional_morpheme	plural (e.g. 們)
FMM	Functional_morpheme	relative (e.g. 的)
IDA	Idiom	(e.g. 斐然成章)
ITA	Interjection	(e.g. 嗚呼, 哎)
LCA	Localizer	(e.g. 上, 后)
NNA	Noun	common (e.g. 劍客, 糧食)
NNB	Noun	AA (e.g. 根根, 人人)
NNC	Noun	AAB (e.g. 三三行, 萬萬慈)
NND	Noun	AABB (e.g. 般般件件)
NNE	Noun	ABAB (e.g. 一對一對)
NNF	Noun	ABAC (e.g. 僮男僮女)
NNG	Noun	ABB (e.g. 一層層, 汗珠珠)
NNH	Noun	ABCB (e.g. 千世萬世)
NNI	Noun	honorific (e.g. 貴庚, 仙鄉)
NNJ	Noun	proper (e.g. 黃巾)
NNK	Noun	proper_dynasty_name (e.g. 春秋戰國)
NNL	Noun	proper_person_name (e.g. 蘧伯玉)
NNM	Noun	proper_place_name (e.g. 黃山)
NNN	Noun	proper_title (e.g. 孫子兵法)
NNO	Noun	proper_year_name (e.g. 天章)
NMA	Numeral	cardinal (e.g. 十八, 千)
NMB	Numeral	indefinite (e.g. 數十, 幾百)

NMC	Numeral	ordinal (e.g. 第一, 第八)
ONA	Onomatopoeia	AA (e.g. 哇哇, 嘻嘻)
ONB	Onomatopoeia	AAA (e.g. 騰騰騰, 撒撒撒)
ONC	Onomatopoeia	AABB (e.g. 隱隱轟轟)
OND	Onomatopoeia	ABBC (e.g. 撲通通冬, 吉丁丁璫)
ONE	Onomatopoeia	general (e.g. 耶櫓咿啞)
PNA	Pronoun	demonstrative (e.g. 這, 其)
PNB	Pronoun	honorific (e.g. 寡人, 在下)
PNC	Pronoun	personal (e.g. 我們, 俺)
PND	Pronoun	possessive (e.g. 我的, 厥)
PNE	Pronoun	reciprocal (e.g. 彼此)
PNF	Pronoun	reflexive (e.g. 自己)
PPA	Preposition	(e.g. 據, 至於)
PRA	Particle	tag (e.g. 吧, 乎)
PTA	Punctuation	general_separating_mark (‘。’,‘！’,’？’,‘，’,)
PTB	Punctuation	left bracket (e.g.『, 《, or「)
PTC	Punctuation	right_bracket (e.g.』, 》, or 」)
PTD	Punctuation	secondary_separating_mark (e.g.‘・’,‘、’)
QWA	Question_word	general (e.g. 為何, 甚麼)
QWB	Question_word	tag (e.g. 麼)
UND	Unidentified	(e.g. □)
VBA	Verb	general (e.g. 剛, 頂)
VBB	Verb	AA (e.g. 演演, 走走)
VBC	Verb	AAB (e.g. 散散心)
VBD	Verb	AABB (e.g. 哭哭啼啼)
VBE	Verb	ABAB (e.g. 接待接待)
VBF	Verb	ABAC (e.g. 包長包短)
VBG	Verb	ABB (e.g. 哭啼啼)

VBH	Verb	ABCB (e.g. 手之舞之)
VBI	Verb	bei_V (e.g. 被戮)
VBJ	Verb	bu_V (e.g. 不宜)
VBK	Verb	copular_shi (e.g. 是)
VBL	Verb	copular_shi _negative (e.g. 不是)
VBM	Verb	existential_you (e.g. 有)
VBN	Verb	existential_you_negative (e.g. 未有)
VBO	Verb	jian_V (e.g. 見信, 見教)
VBP	Verb	modal_auxiliary (e.g. 必, 該)
VBQ	Verb	modal_auxiliary_negative (e.g. 不必, 不該)
VBR	Verb	reciprocal_xiang_V (e.g. 相會, 相辭)
VBS	Verb	reflexive_zi_V (e.g. 自寬, 自縊)
VBT	Verb	stative (e.g. 惆悵, 廣厚)
VBU	Verb	stative_comparative (e.g. 更深)
VBV	Verb	stative_superlative (e.g. 最早)
VBW	Verb	suo_V (e.g. 所積, 所吟)
VBX	Verb	V_bu_V (e.g. 念不念, 定不定)
VBY	Verb	V_hua (e.g. 羽化)
VBZ	Verb	V_lai (e.g. 宣來, 討來)
VBAA	Verb	V_N (e.g. 守法, 聽話)
VBBB	Verb	V_potential_bu_RVC* (e.g. 睡不穩)
VBCC	Verb	V_potential_de_RVC (e.g. 躲得過)
VBDD	Verb	V_qu (e.g. 消去, 拿去)
VBEE	Verb	V_RVC (e.g. 學成, 生出)
VBFF	Verb	V_V (e.g. 敘說, 思慮)
VBGG	Verb	V_yi_V (e.g. 畫一畫, 嘗一嘗)
VBHH	Verb	V_yu (e.g. 起於)
VBII	Verb	V_zhi (e.g. 刑之)

VBJJ	Verb	yi_V (e.g. 一訪, 一望)

Appendix 2

Tokens and types occurred in the early Archaic Chinese period in the SCC

1	君子	107
2	孟子	91
3	聖人	71
4	其所	56
5	諸侯	39
6	孔子	38
7	百姓	38
8	先王	34
9	夫子	34
10	萬物	27
11	一人	25
12	小人	25
13	天地	24
14	仁者	23
15	子貢	23
16	文王	23
17	其民	22
18	百里	22
19	曾子	21
20	父母	21
21	王者	21
22	大夫	20
23	其君	19
24	昔者	19
25	周公	18
26	子路	18
27	商書	17
28	四海	17
29	上下	16
30	天子	16
31	農戰	16
32	三軍	15
33	國家	15
34	皐陶	15
35	管仲	15
36	伊尹	14
37	朋友	14
38	武王	14
39	其國	13
40	其心	13
41	大國	13
42	孫子	13
43	道者	13
44	其道	12
45	賢者	12
46	其政	11
47	千里	11
48	四方	11
49	子思	11
50	戰者	11
51	百官	11
52	知者	11
53	禽獸	11
54	萬乘	11
55	萬邦	11
56	齊人	11
57	他日	10
58	伯夷	10
59	其事	10
60	其位	10
61	其子	10
62	其身	10
63	十者	10
64	富者	10
65	左右	10
66	死地	10
67	許子	10
68	三年	9
69	三者	9
70	上帝	9
71	世主	9
72	仁義	9
73	其親	9
74	天命	9
75	夫民	9
76	夷子	9
77	妻子	9
78	官爵	9
79	小國	9
80	弟子	9
81	技藝	9
82	日月	9
83	明君	9
84	為人	9
85	百工	9
86	鬼神	9
87	鳥獸	9
88	中庸	8
89	仁政	8
90	其人	8
91	其德	8
92	其志	8
93	其知	8
94	千乘	8
95	右傳	8
96	善者	8
97	天道	8
98	子之	8
99	我者	8
100	民力	8
101	農民	8
102	顏淵	8
103	黎民	8
104	丈夫	7
105	三月	7
106	中國	7
107	什一	7
108	今日	7
109	仲子	7
110	其言	7
111	勝者	7
112	國人	7
113	姦民	7
114	子夏	7
115	宗廟	7
116	湯武	7
117	禮樂	7
118	詩書	7
119	貧者	7
120	三載	6
121	乃祖	6
122	二者	6
123	兄弟	6
124	先知	6
125	兩者	6
126	兵者	6
127	其一	6
128	其家	6
129	其故	6
130	其時	6
131	其母	6
132	其然	6
133	其父	6
134	凶年	6
135	前日	6
136	四岳	6
137	圍地	6
138	學者	6
139	宰我	6
140	小子	6
141	山林	6
142	庶人	6
143	忠信	6
144	故知	6
145	春秋	6
146	晏子	6
147	智者	6
148	死者	6
149	然友	6
150	父子	6
151	百畝	6
152	義者	6
153	草木	6
154	衆人	6
155	衢地	6
156	輕者	6
157	五者	5
158	人心	5
159	仲尼	5
160	侯王	5
161	兆民	5
162	八者	5
163	其仁	5
164	其力	5
165	其本	5
166	其極	5
167	其樂	5
168	匹夫	5
169	反間	5
170	古者	5
171	商賈	5
172	善人	5
173	土地	5
174	圯地	5
175	地形	5
176	堯舜	5
177	大臣	5
178	天時	5
179	天者	5
180	夫婦	5
181	子孫	5
182	子張	5
183	尊賢	5
184	庶民	5
185	敵人	5
186	明德	5
187	此時	5
188	民利	5
189	民壹	5
190	洪水	5
191	爵祿	5
192	玄德	5
193	王命	5
194	生民	5
195	百世	5
196	皇天	5
197	神農	5
198	管叔	5
199	股肱	5
200	至德	5

201	葛伯	5
202	虞書	5
203	衆者	5
204	誠者	5
205	說者	5
206	農者	5
207	道路	5
208	遠人	5
209	鄰國	5
210	重地	5
211	重者	5
212	長者	5
213	險阻	5
214	一日	4
215	三代	4
216	九族	4
217	九江	4
218	五刑	4
219	五穀	4
220	五色	4
221	交地	4
222	人君	4
223	仁人	4
224	六淫	4
225	六者	4
226	其二	4
227	其利	4
228	其居	4
229	其意	4
230	其正	4
231	其法	4
232	其用	4
233	其私	4
234	其行	4
235	其門	4
236	其餘	4
237	制度	4
238	勇民	4
239	十歲	4
240	千人	4
241	千歲	4
242	君臣	4
243	哀公	4
244	四時	4
245	四難	4
246	國用	4
247	地者	4
248	士卒	4
249	夏書	4
250	夙夜	4
251	大事	4
252	大學	4
253	大德	4
254	大王	4
255	大道	4
256	夷狄	4
257	姦宄	4
258	嬰兒	4
259	學問	4
260	將軍	4
261	山川	4
262	巧言	4
263	弱者	4
264	彊國	4
265	後人	4
266	怯民	4
267	愚者	4
268	散地	4
269	數年	4
270	新邑	4
271	旌旗	4
272	明主	4
273	時子	4
274	曾西	4
275	有若	4
276	朝廷	4
277	楚人	4
278	此處	4
279	死生	4
280	民務	4
281	法者	4
282	溝壑	4
283	爭地	4
284	王子	4
285	王道	4
286	管氏	4
287	終始	4
288	自然	4
289	蝨官	4
290	衣服	4
291	車馬	4
292	軍市	4
293	軍爭	4
294	輕重	4
295	農夫	4
296	野人	4
297	門人	4
298	飲食	4
299	高后	4
300	三危	3
301	三官	3
302	三日	3
303	三苗	3
304	上者	3
305	上賢	3
306	下者	3
307	世俗	3
308	九河	3
309	九變	3
310	事者	3
311	二月	3
312	五人	3
313	五典	3
314	人倫	3
315	人言	3
316	今世	3
317	令色	3
318	仲弓	3
319	來年	3
320	來者	3
321	倉廩	3
322	元首	3
323	內間	3
324	八音	3
325	共工	3
326	其亂	3
327	其光	3
328	其兵	3
329	其名	3
330	其命	3
331	其善	3
332	其性	3
333	其治	3
334	其禮	3
335	其罪	3
336	其能	3
337	其途	3
338	其過	3
339	刑人	3
340	前後	3
341	勇士	3
342	北方	3
343	北狄	3
344	十乘	3
345	厥命	3
346	名利	3
347	君命	3
348	四端	3
349	四門	3
350	國務	3
351	墨者	3
352	夏禮	3
353	大人	3
354	大小	3
355	大旱	3
356	天吏	3
357	太山	3
358	夫人	3
359	奇正	3
360	子遊	3
361	孝公	3
362	孝弟	3
363	孝慈	3
364	季氏	3
365	小大	3
366	小民	3
367	崔子	3
368	帝位	3
369	常官	3
370	平陸	3
371	後世	3
372	微子	3
373	心者	3
374	戰地	3
375	故國	3
376	敵間	3
377	文武	3
378	斥澤	3
379	於陵	3
380	星辰	3
381	是非	3
382	朝服	3
383	東北	3
384	梓匠	3
385	樊遲	3
386	此道	3
387	武丁	3
388	死間	3
389	殷禮	3
390	每事	3
391	民心	3
392	水火	3
393	江海	3
394	沮澤	3
395	滔天	3
396	父兄	3
397	牛羊	3
398	狄人	3
399	率然	3
400	王季	3
401	王良	3
402	生者	3
403	生間	3
404	疾農	3
405	百人	3
406	百步	3
407	百縣	3
408	盜賊	3

409	短長	3
410	社稷	3
411	神祇	3
412	祭祀	3
413	禮者	3
414	私利	3
415	私門	3
416	程子	3
417	童子	3
418	罪人	3
419	羣臣	3
420	老弱	3
421	聖者	3
422	臣主	3
423	致知	3
424	萬世	3
425	藪澤	3
426	處士	3
427	蚳黿	3
428	西夷	3
429	親戚	3
430	許行	3
431	諸身	3
432	豪傑	3
433	貧賤	3
434	貴賤	3
435	輪輿	3
436	辯慧	3
437	邦君	3
438	鄉人	3
439	重輕	3
440	長幼	3
441	長短	3
442	陳相	3
443	陽貨	3
444	霸王	3
445	顏色	3
446	饑歲	3
447	驩兜	3
448	魚鱉	3
449	黑水	3
450	齊國	3
451	一歲	2
452	一理	2
453	一端	2
454	一篇	2
455	一簞	2
456	一簣	2
457	一道	2
458	三人	2
459	三嗅	2
460	三家	2

461	上世	2
462	世事	2
463	世祿	2
464	丘陵	2
465	中人	2
466	中心	2
467	中道	2
468	九地	2
469	九川	2
470	九德	2
471	九經	2
472	亂臣	2
473	二女	2
474	二生	2
475	五危	2
476	五味	2
477	五教	2
478	五月	2
479	五服	2
480	五民	2
481	五火	2
482	五畝	2
483	五禮	2
484	五聲	2
485	五行	2
486	五里	2
487	五間	2
488	人主	2
489	人力	2
490	人食	2
491	仕者	2
492	他邦	2
493	俗人	2
494	倉府	2
495	先君	2
496	內外	2
497	兩死	2
498	兩端	2
499	八月	2
500	公事	2
501	公家	2
502	兵力	2
503	兵革	2
504	其任	2
505	其使	2
506	其先	2
507	其反	2
508	其味	2
509	其地	2
510	其妙	2
511	其守	2
512	其官	2

513	其急	2
514	其情	2
515	其教	2
516	其斯	2
517	其氣	2
518	其生	2
519	其神	2
520	其祿	2
521	其禁	2
522	其聲	2
523	其舍	2
524	其苗	2
525	其衆	2
526	其財	2
527	其賢	2
528	其辭	2
529	其類	2
530	其食	2
531	其首	2
532	冉子	2
533	冉有	2
534	冉牛	2
535	刑罰	2
536	刑者	2
537	力行	2
538	功名	2
539	勇者	2
540	動者	2
541	務學	2
542	勝兵	2
543	勝敗	2
544	勝負	2
545	勢者	2
546	北海	2
547	匹婦	2
548	十人	2
549	十步	2
550	十里	2
551	十鎰	2
552	千金	2
553	南方	2
554	叔齊	2
555	古人	2
556	司徒	2
557	后稷	2
558	君斷	2
559	吾國	2
560	吾心	2
561	吾計	2
562	周人	2
563	商酤	2
564	善修	2

565	善政	2
566	善民	2
567	喪禮	2
568	喬木	2
569	器用	2
570	四境	2
571	四夷	2
572	四者	2
573	四鄰	2
574	四體	2
575	國力	2
576	國法	2
577	園囿	2
578	地力	2
579	地方	2
580	塗炭	2
581	壹空	2
582	壺口	2
583	外交	2
584	多少	2
585	多歲	2
586	大功	2
587	大者	2
588	大邪	2
589	天將	2
590	太王	2
591	夫地	2
592	夫天	2
593	夫物	2
594	女子	2
595	威儀	2
596	嬖人	2
597	子噲	2
598	子游	2
599	安危	2
600	安民	2
601	宋人	2
602	宣王	2
603	室家	2
604	宮室	2
605	家斷	2
606	寡者	2
607	實者	2
608	將者	2
609	導者	2
610	小事	2
611	少師	2
612	少者	2
613	山澤	2
614	岐山	2
615	岷山	2
616	工師	2

617	干戈	2
618	幽谷	2
619	府庫	2
620	庶官	2
621	康誥	2
622	庸德	2
623	弱水	2
624	彭蠡	2
625	從周	2
626	從者	2
627	御者	2
628	德行	2
629	心力	2
630	忠臣	2
631	患難	2
632	懷山	2
633	戎狄	2
634	戮人	2
635	戰戢	2
636	戰道	2
637	放勳	2
638	政事	2
639	政舉	2
640	故事	2
641	故道	2
642	敗兵	2
643	教者	2
644	文公	2
645	文祖	2
646	斯民	2
647	新民	2
648	日費	2
649	昆弟	2
650	明道	2
651	時日	2
652	時雨	2
653	晉國	2
654	晝夜	2
655	暴君	2
656	朝夕	2
657	木石	2
658	本末	2
659	杜墊	2
660	林放	2
661	林木	2
662	桐宮	2
663	桐柏	2
664	桓公	2
665	棺槨	2
666	楊氏	2
667	樂者	2
668	正者	2

669	此章	2
670	此篇	2
671	死制	2
672	民生	2
673	水土	2
674	汙池	2
675	沈同	2
676	河內	2
677	河東	2
678	河海	2
679	泄柳	2
680	法令	2
681	法度	2
682	泰山	2
683	流沙	2
684	淫者	2
685	深淵	2
686	淵泉	2
687	湯誓	2
688	火攻	2
689	烝民	2
690	犧牲	2
691	犬馬	2
692	狂夫	2
693	狐狸	2
694	狐貉	2
695	猛獸	2
696	猶人	2
697	玉人	2
698	王公	2
699	王驩	2
700	甘龍	2
701	甲冑	2
702	疾戰	2
703	百乘	2
704	百年	2
705	百歲	2
706	百獸	2
707	百穀	2
708	皋陶	2
709	盛服	2
710	盜跖	2
711	瞽者	2
712	知慧	2
713	禍福	2
714	禮義	2
715	秋毫	2
716	窮寇	2
717	糧食	2
718	細過	2
719	絕地	2
720	罪誅	2

721	羔裘	2
722	羽旄	2
723	老者	2
724	耒耜	2
725	耳目	2
726	聖王	2
727	聲色	2
728	臣下	2
729	臧倉	2
730	自邇	2
731	舊服	2
732	舊邦	2
733	良將	2
734	良田	2
735	荒草	2
736	萬事	2
737	萬民	2
738	衆寡	2
739	衆樹	2
740	衣裳	2
741	衰者	2
742	裕利	2
743	襄陵	2
744	西河	2
745	言談	2
746	誑事	2
747	諸人	2
748	谿谷	2
749	貨財	2
750	賓客	2
751	賞者	2
752	賢君	2
753	賢將	2
754	越人	2
755	蹙頞	2
756	軍士	2
757	輜重	2
758	逆旅	2
759	遠方	2
760	遠者	2
761	遠近	2
762	邪官	2
763	郊社	2
764	都邑	2
765	鄉党	2
766	鄉黨	2
767	酒食	2
768	酣奭	2
769	里仁	2
770	金粟	2
771	金鼓	2
772	鐘鼓	2

773	長上	2
774	長生	2
775	間然	2
776	閔子	2
777	阿衡	2
778	陳良	2
779	雅言	2
780	雍也	2
781	靈沼	2
782	靈臺	2
783	靜者	2
784	食口	2
785	餘力	2
786	高陽	2
787	麀鹿	2
788	麋鹿	2
789	齊明	2
790	龍門	2
791	一兩	1
792	一勺	1
793	一卷	1
794	一員	1
795	一國	1
796	一宅	1
797	一役	1
798	一心	1
799	一月	1
800	一束	1
801	一死	1
802	一牛	1
803	一生	1
804	一石	1
805	一等	1
806	一者	1
807	一言	1
808	一賞	1
809	一身	1
810	一轂	1
811	一鍾	1
812	一門	1
813	一隅	1
814	七世	1
815	七寸	1
816	三仕	1
817	三公	1
818	三分	1
819	三寶	1
820	三屬	1
821	三帛	1
822	三思	1
823	三旬	1
824	三歸	1

825	三遾	1
826	三王	1
827	三生	1
828	三省	1
829	三禮	1
830	三過	1
831	三里	1
832	三重	1
833	三鼎	1
834	上兵	1
835	上士	1
836	上官	1
837	上將	1
838	上章	1
839	上風	1
840	下士	1
841	下官	1
842	下流	1
843	下風	1
844	不佞	1
845	世禁	1
846	丘垤	1
847	丘役	1
848	丘者	1
849	丘隅	1
850	中原	1
851	中古	1
852	中士	1
853	中者	1
854	中門	1
855	丹朱	1
856	主人	1
857	主心	1
858	主用	1
859	乃粒	1
860	乙丑	1
861	九人	1
862	九天	1
863	九夷	1
864	九層	1
865	九山	1
866	九州	1
867	九澤	1
868	九載	1
869	亂軍	1
870	亂邦	1
871	予手	1
872	予足	1
873	事事	1
874	二人	1
875	二代	1
876	二國	1

877	二子	1
878	二日	1
879	二本	1
880	互鄉	1
881	五事	1
882	五品	1
883	五器	1
884	五子	1
885	五尺	1
886	五庸	1
887	五惇	1
888	五旬	1
889	五玉	1
890	五瑞	1
891	五用	1
892	五章	1
893	五臟	1
894	五言	1
895	五載	1
896	五邦	1
897	五采	1
898	五霸	1
899	五音	1
900	五鼎	1
901	人事	1
902	人口	1
903	人情	1
904	人欲	1
905	人臣	1
906	人行	1
907	人道	1
908	人願	1
909	什二	1
910	什四	1
911	仁廉	1
912	今人	1
913	他人	1
914	令尹	1
915	仲冬	1
916	仲夏	1
917	仲由	1
918	仲秋	1
919	任譽	1
920	任賢	1
921	伊摯	1
922	伊訓	1
923	伏妤	1
924	伏羲	1
925	伯牛	1
926	佯北	1
927	來世	1
928	來朝	1

929	倉口	1
930	倍欲	1
931	偏家	1
932	偽者	1
933	傅巖	1
934	傾者	1
935	儒者	1
936	元日	1
937	先人	1
938	先公	1
939	先天	1
940	先後	1
941	先時	1
942	先生	1
943	先祖	1
944	兗州	1
945	內作	1
946	內省	1
947	全伍	1
948	全卒	1
949	全國	1
950	全旅	1
951	全軍	1
952	兩實	1
953	兩生	1
954	兩虛	1
955	八佾	1
956	八口	1
957	八年	1
958	公倉	1
959	公利	1
960	公卿	1
961	公門	1
962	六合	1
963	六宗	1
964	六官	1
965	六尺	1
966	六師	1
967	六律	1
968	六旬	1
969	六親	1
970	六馬	1
971	兵士	1
972	兵家	1
973	兵法	1
974	兵甲	1
975	兵衆	1
976	其交	1
977	其令	1
978	其制	1
979	其功	1
980	其勢	1

981	其嚴	1
982	其城	1
983	其壽	1
984	其宜	1
985	其害	1
986	其富	1
987	其尾	1
988	其己	1
989	其弊	1
990	其弱	1
991	其徒	1
992	其惡	1
993	其文	1
994	其書	1
995	其材	1
996	其根	1
997	其梯	1
998	其業	1
999	其榮	1
1000	其樸	1
1001	其機	1
1002	其欲	1
1003	其此	1
1004	其理	1
1005	其節	1
1006	其結	1
1007	其美	1
1008	其腹	1
1009	其色	1
1010	其萃	1
1011	其虛	1
1012	其衰	1
1013	其說	1
1014	其謀	1
1015	其象	1
1016	其貴	1
1017	其辯	1
1018	其農	1
1019	其鄉	1
1020	其間	1
1021	其闕	1
1022	其陽	1
1023	其險	1
1024	其順	1
1025	其養	1
1026	其體	1
1027	其默	1
1028	其齊	1
1029	典刑	1
1030	冀州	1
1031	冉求	1
1032	冢宰	1

1033	凶事	1
1034	凶服	1
1035	分數	1
1036	刑戮	1
1037	刑賞	1
1038	利劍	1
1039	利害	1
1040	則天	1
1041	前世	1
1042	前功	1
1043	前章	1
1044	剛柔	1
1045	功臣	1
1046	勁者	1
1047	勇功	1
1048	動靜	1
1049	勾踐	1
1050	北者	1
1051	北辰	1
1052	匠人	1
1053	匡人	1
1054	匡章	1
1055	十世	1
1056	十倍	1
1057	十室	1
1058	十手	1
1059	十畝	1
1060	十目	1
1061	十裏	1
1062	千室	1
1063	千物	1
1064	千駟	1
1065	半古	1
1066	南子	1
1067	南容	1
1068	南海	1
1069	南蠻	1
1070	卜稽	1
1071	危邦	1
1072	卿相	1
1073	厚薄	1
1074	原思	1
1075	去彊	1
1076	右袂	1
1077	司空	1
1078	吉事	1
1079	吉凶	1
1080	吉月	1
1081	同姓	1
1082	同寅	1
1083	后羿	1
1084	吏卒	1

1085	君上	1
1086	君道	1
1087	君長	1
1088	吳人	1
1089	吾友	1
1090	吾才	1
1091	吾義	1
1092	吾間	1
1093	呂牙	1
1094	周監	1
1095	周禮	1
1096	周霄	1
1097	和叔	1
1098	唐棣	1
1099	商民	1
1100	善賈	1
1101	善道	1
1102	喪祭	1
1103	嘉樂	1
1104	器備	1
1105	器械	1
1106	器皿	1
1107	四宿	1
1108	四帝	1
1109	四朝	1
1110	四目	1
1111	四聰	1
1112	四肢	1
1113	四表	1
1114	四軍	1
1115	四通	1
1116	四達	1
1117	因間	1
1118	囿方	1
1119	國位	1
1120	國俗	1
1121	國功	1
1122	國政	1
1123	國本	1
1124	國相	1
1125	國粟	1
1126	圍師	1
1127	坐者	1
1128	城邑	1
1129	域中	1
1130	執事	1
1131	堤防	1
1132	堯典	1
1133	塚宰	1
1134	墨氏	1
1135	墨翟	1
1136	墾田	1

1137	士人	1
1138	壬癸	1
1139	壯農	1
1140	夏后	1
1141	外事	1
1142	外人	1
1143	外內	1
1144	外誘	1
1145	多寡	1
1146	夜戰	1
1147	夜風	1
1148	大吏	1
1149	大孝	1
1150	大官	1
1151	大川	1
1152	大敵	1
1153	大甲	1
1154	大知	1
1155	大舜	1
1156	大葬	1
1157	大車	1
1158	大軍	1
1159	大麓	1
1160	天井	1
1161	天德	1
1162	天心	1
1163	天性	1
1164	天成	1
1165	天明	1
1166	天災	1
1167	天牢	1
1168	天物	1
1169	天羅	1
1170	天門	1
1171	天陷	1
1172	天隙	1
1173	太原	1
1174	太師	1
1175	太廟	1
1176	太牢	1
1177	太甲	1
1178	太行	1
1179	太誓	1
1180	夫山	1
1181	夫微	1
1182	夫政	1
1183	夫水	1
1184	失飪	1
1185	夷道	1
1186	夾右	1
1187	奇物	1
1188	奮武	1

1189	好人	1
1190	妖孽	1
1191	妻帑	1
1192	妾婦	1
1193	委積	1
1194	姦蝨	1
1195	威者	1
1196	婦人	1
1197	媧汭	1
1198	子產	1
1199	子禽	1
1200	子罕	1
1201	子者	1
1202	子華	1
1203	子襄	1
1204	子道	1
1205	孔彰	1
1206	孔氏	1
1207	孔門	1
1208	存亡	1
1209	孟孫	1
1210	孟津	1
1211	孟賁	1
1212	孟軻	1
1213	季秋	1
1214	季路	1
1215	學民	1
1216	安定	1
1217	安平	1
1218	宋朝	1
1219	宋王	1
1220	宗器	1
1221	官人	1
1222	官任	1
1223	官制	1
1224	官士	1
1225	官斷	1
1226	官道	1
1227	定公	1
1228	宰予	1
1229	容色	1
1230	容貌	1
1231	寇賊	1
1232	富家	1
1233	寒暑	1
1234	寒疾	1
1235	寡欲	1
1236	寢衣	1
1237	實學	1
1238	實用	1
1239	寶藏	1
1240	將相	1

1241	將過	1
1242	尊官	1
1243	小官	1
1244	小德	1
1245	小敵	1
1246	小畝	1
1247	小者	1
1248	小鮮	1
1249	屈伸	1
1250	山梁	1
1251	山谿	1
1252	岱宗	1
1253	岱禮	1
1254	岳陽	1
1255	島夷	1
1256	峻宇	1
1257	崑岡	1
1258	嶓冢	1
1259	川流	1
1260	川谷	1
1261	巨擘	1
1262	巫匠	1
1263	市脯	1
1264	市賈	1
1265	布帛	1
1266	帝典	1
1267	帝堯	1
1268	帝王	1
1269	帷裳	1
1270	常人	1
1271	常勢	1
1272	常山	1
1273	常形	1
1274	干祿	1
1275	平章	1
1276	幽州	1
1277	幽明	1
1278	幽都	1
1279	庶事	1
1280	康子	1
1281	庸者	1
1282	庸言	1
1283	廉潔	1
1284	廊廟	1
1285	廢國	1
1286	廣狹	1
1287	建德	1
1288	弘毅	1
1289	張儀	1
1290	強者	1
1291	强梁	1
1292	形名	1

1293	彭更	1
1294	影響	1
1295	征利	1
1296	後日	1
1297	徐州	1
1298	得失	1
1299	得者	1
1300	從人	1
1301	徵招	1
1302	心法	1
1303	心腹	1
1304	忌憚	1
1305	忌諱	1
1306	志意	1
1307	忠直	1
1308	忸怩	1
1309	忿速	1
1310	性善	1
1311	怨民	1
1312	怪石	1
1313	怯者	1
1314	怵惕	1
1315	恒位	1
1316	恥辱	1
1317	惡壽	1
1318	惡田	1
1319	惡臭	1
1320	惡貨	1
1321	惰歸	1
1322	愛子	1
1323	慈仁	1
1324	慍色	1
1325	憂患	1
1326	憲章	1
1327	戎衣	1
1328	成人	1
1329	成湯	1
1330	成覛	1
1331	我家	1
1332	戰事	1
1333	戰勢	1
1334	戰卒	1
1335	戰日	1
1336	戰色	1
1337	戶牖	1
1338	拳拳	1
1339	持載	1
1340	掛形	1
1341	揚州	1
1342	支形	1
1343	故習	1
1344	教道	1

1345	敬忠	1
1346	敵國	1
1347	數人	1
1348	數月	1
1349	數者	1
1350	數里	1
1351	文德	1
1352	文明	1
1353	文獻	1
1354	文章	1
1355	文綵	1
1356	斯夫	1
1357	新子	1
1358	方土	1
1359	方策	1
1360	旄倪	1
1361	日亡	1
1362	日日	1
1363	日章	1
1364	旻天	1
1365	昆侖	1
1366	昆夷	1
1367	昊天	1
1368	明命	1
1369	明日	1
1370	明王	1
1371	明聖	1
1372	明衣	1
1373	昧谷	1
1374	昭公	1
1375	時使	1
1376	時俗	1
1377	時制	1
1378	時措	1
1379	晉文	1
1380	晝戰	1
1381	晝氣	1
1382	晝風	1
1383	景公	1
1384	景春	1
1385	智名	1
1386	智慧	1
1387	暘谷	1
1388	暮氣	1
1389	曠夫	1
1390	曲制	1
1391	月守	1
1392	有顏	1
1393	朔方	1
1394	朝儛	1
1395	朝暮	1
1396	朝氣	1

1397	木者	1
1398	本物	1
1399	朽木	1
1400	材能	1
1401	杖者	1
1402	松柏	1
1403	枉木	1
1404	栗栗	1
1405	栽者	1
1406	桓文	1
1407	桓魋	1
1408	楊朱	1
1409	槁壤	1
1410	樵采	1
1411	橐籥	1
1412	次第	1
1413	欽明	1
1414	歡樂	1
1415	正人	1
1416	正席	1
1417	正心	1
1418	正月	1
1419	此世	1
1420	此地	1
1421	此心	1
1422	此者	1
1423	此言	1
1424	武城	1
1425	歷山	1
1426	歸師	1
1427	比干	1
1428	毛毯	1
1429	氄毛	1
1430	民資	1
1431	水流	1
1432	水草	1
1433	江漢	1
1434	沛澤	1
1435	沽酒	1
1436	法術	1
1437	法語	1
1438	泰伯	1
1439	洿池	1
1440	流水	1
1441	浮雲	1
1442	海濱	1
1443	淫民	1
1444	深谿	1
1445	深間	1
1446	湯誥	1
1447	溝洫	1
1448	溫恭	1

1449	滄浪	1
1450	漆雕	1
1451	漢朝	1
1452	潢井	1
1453	澤梁	1
1454	激水	1
1455	濟河	1
1456	火人	1
1457	火力	1
1458	火庫	1
1459	火積	1
1460	火輜	1
1461	火隊	1
1462	無辜	1
1463	煙火	1
1464	熊羆	1
1465	燕居	1
1466	燕毛	1
1467	營窟	1
1468	營軍	1
1469	物者	1
1470	牲殺	1
1471	狂簡	1
1472	狐裘	1
1473	狼戾	1
1474	獯鬻	1
1475	玄冠	1
1476	玄牡	1
1477	率性	1
1478	玉衡	1
1479	王室	1
1480	王庭	1
1481	珠玉	1
1482	班師	1
1483	琅玕	1
1484	琅邪	1
1485	瑕謫	1
1486	瑚璉	1
1487	瑟琴	1
1488	瑤琨	1
1489	璞玉	1
1490	璿璣	1
1491	瓜祭	1
1492	甘誓	1
1493	甘酒	1
1494	甘露	1
1495	生物	1
1496	田數	1
1497	田畝	1
1498	申棖	1
1499	申詳	1
1500	男女	1

1501	異名	1
1502	異端	1
1503	疏食	1
1504	疲者	1
1505	白刃	1
1506	百殃	1
1507	百疾	1
1508	百辟	1
1509	百金	1
1510	皇上	1
1511	益稷	1
1512	盛饌	1
1513	盜夸	1
1514	瞽瞍	1
1515	矢弓	1
1516	祖廟	1
1517	祝鮀	1
1518	神器	1
1519	神宗	1
1520	神祇	1
1521	神紀	1
1522	禎祥	1
1523	禘嘗	1
1524	禮儀	1
1525	禮法	1
1526	禹貢	1
1527	私交	1
1528	私勞	1
1529	私覿	1
1530	私道	1
1531	秦國	1
1532	稅斂	1
1533	穆公	1
1534	積水	1
1535	立人	1
1536	箕子	1
1537	篤信	1
1538	篤敬	1
1539	簡子	1
1540	籩豆	1
1541	糞土	1
1542	糧道	1
1543	紀綱	1
1544	素具	1
1545	素衣	1
1546	累土	1
1547	終南	1
1548	絕世	1
1549	絕山	1
1550	絕澗	1
1551	絲絮	1
1552	經綸	1

1553	緇衣	1
1554	緝熙	1
1555	緡蠻	1
1556	縕袍	1
1557	縹緲	1
1558	縻軍	1
1559	繩約	1
1560	美玉	1
1561	群羊	1
1562	羽山	1
1563	老彭	1
1564	聖賢	1
1565	聲名	1
1566	聲音	1
1567	肉味	1
1568	肉敗	1
1569	肺肝	1
1570	背丘	1
1571	胡齕	1
1572	胤征	1
1573	能臣	1
1574	脫兔	1
1575	膠漆	1
1576	膠鬲	1
1577	臣子	1
1578	自專	1
1579	自身	1
1580	至文	1
1581	至聖	1
1582	至道	1
1583	臺池	1
1584	舊人	1
1585	舊惡	1
1586	舍人	1
1587	舜典	1
1588	舟楫	1
1589	舟車	1
1590	良工	1
1591	良民	1
1592	色惡	1
1593	芻蕘	1
1594	芻藁	1
1595	若水	1
1596	草萊	1
1597	荊州	1
1598	荊棘	1
1599	荼毒	1
1600	莊子	1
1601	莊嶽	1
1602	莊暴	1
1603	菉竹	1
1604	菏澤	1

1605	菜羹	1
1606	華嶽	1
1607	萬人	1
1608	萬夫	1
1609	萬章	1
1610	葉公	1
1611	蒲盧	1
1612	蒹葭	1
1613	蒼生	1
1614	薈龜	1
1615	薄冰	1
1616	藪薈	1
1617	藁椁	1
1618	處女	1
1619	虛實	1
1620	蜂蠆	1
1621	蠅蚋	1
1622	蠶繅	1
1623	蠻夷	1
1624	蠻貊	1
1625	血氣	1
1626	衆力	1
1627	衆星	1
1628	衆草	1
1629	行潦	1
1630	衡漳	1
1631	衡陽	1
1632	衣衾	1
1633	衣錦	1
1634	衣食	1
1635	裳衣	1
1636	裹糧	1
1637	褐夫	1
1638	褻服	1
1639	襲明	1
1640	西伯	1
1641	西戎	1
1642	西禮	1
1643	要妙	1
1644	角招	1
1645	言說	1
1646	計謀	1
1647	誠意	1
1648	諸夏	1
1649	諸掌	1
1650	諸斯	1
1651	諸歲	1
1652	諸父	1
1653	諸罟	1
1654	諸臣	1
1655	諸道	1
1656	謁者	1

1657	識者	1
1658	變化	1
1659	變者	1
1660	豫州	1
1661	財貨	1
1662	貧富	1
1663	貴人	1
1664	貴爵	1
1665	買者	1
1666	費宰	1
1667	費留	1
1668	賊子	1
1669	賢人	1
1670	賢聖	1
1671	賤工	1
1672	走獸	1
1673	足恭	1
1674	跛行	1
1675	蹊道	1
1676	躁心	1
1677	身者	1
1678	躬行	1
1679	車戰	1
1680	車甲	1
1681	軍政	1
1682	軍食	1
1683	轉附	1
1684	轍迹	1
1685	辭氣	1
1686	農時	1
1687	迅雷	1
1688	近者	1
1689	述而	1
1690	追者	1
1691	通形	1
1692	逝者	1
1693	進者	1
1694	道學	1
1695	道心	1

1696	道法	1
1697	達人	1
1698	達巷	1
1699	遠形	1
1700	遠水	1
1701	邇言	1
1702	邦畿	1
1703	鄙事	1
1704	鄙夫	1
1705	鄰国	1
1706	鄰里	1
1707	酒肉	1
1708	釋明	1
1709	重華	1
1710	重資	1
1711	金玉	1
1712	金革	1
1713	鈇鉞	1
1714	銳卒	1
1715	銳氣	1
1716	錯法	1
1717	鎡基	1
1718	長子	1
1719	門者	1
1720	閏月	1
1721	闕殆	1
1722	闕疑	1
1723	關楗	1
1724	關睢	1
1725	關雎	1
1726	阼階	1
1727	陟方	1
1728	陰陽	1
1729	陰雨	1
1730	陳代	1
1731	陳臻	1
1732	陳賈	1
1733	陶唐	1
1734	陷阱	1

1735	隄防	1
1736	隘形	1
1737	險形	1
1738	險易	1
1739	險阨	1
1740	雅頌	1
1741	雉兔	1
1742	雌雉	1
1743	雍州	1
1744	雕牆	1
1745	雞犬	1
1746	雨水	1
1747	雪宮	1
1748	雷雨	1
1749	雷霆	1
1750	雷震	1
1751	雷首	1
1752	霖雨	1
1753	霜露	1
1754	靈圄	1
1755	面目	1
1756	革車	1
1757	音聲	1
1758	頑童	1
1759	飛廉	1
1760	飛鳥	1
1761	食飲	1
1762	飣粥	1
1763	餌兵	1
1764	餘命	1
1765	餘歲	1
1766	餘者	1
1767	餘財	1
1768	餱糧	1
1769	餽鼈	1
1770	饋糧	1
1771	饒野	1
1772	首尾	1
1773	馬牛	1

1774	馳車	1
1775	駟馬	1
1776	驕子	1
1777	驟雨	1
1778	高下	1
1779	高人	1
1780	高子	1
1781	高山	1
1782	高祖	1
1783	高者	1
1784	高陵	1
1785	魚餧	1
1786	魚鱷	1
1787	魯頌	1
1788	鳥夷	1
1789	鳳凰	1
1790	鳳鳥	1
1791	鴃舌	1
1792	鷙鳥	1
1793	麑裘	1
1794	麒麟	1
1795	麻冕	1
1796	麻縷	1
1797	黃帝	1
1798	黃泉	1
1799	黃鳥	1
1800	黍稷	1
1801	黍稻	1
1802	黎獻	1
1803	黻冕	1
1804	黼黻	1
1805	齊桓	1
1806	齊王	1
1807	龍子	1
		4804

Appendix 3

299 N_1-N_2 more frequently occurred in earli Archaic Chinese of the SCC

	character	Pin-yin	literal translation	meaning translation	token
1	君子	Jūn-zǐ	gentlemen-son	monarch, gentleman	107
2	孟子	mèng-zǐ	personal name	personal name	91
3	聖人	shèng-rén	saint-person	sage	71
4	其所	qí-suǒ	its-place	one's place	56
5	諸侯	zhū-hóu	every-marquis	dukes or princes, marquis	39
6	孔子	kǒng-zǐ	personal name	personal name	38
7	百姓	bǎ-xìng	hundred-surname	common people	38
8	先王	xiān-wáng	ancestor-king	emperors of Xia Shang and Chou	34
9	夫子	fū-zǐ	husband-son	man (teacher),master	34
10	萬物	wàn-wù	ten thousand-thing	everything	27
11	一人	yī-rén	one-person	one person	25
12	小人	xiǎo-rén	smallness-person	common people	25
13	天地	tiān-dì	heaven and earth	world	24
14	仁者	rén-zhě	benevolence-people	benevolence	23
15	子貢	zi-gòng	personal name	personal name	23
16	文王	wén-wáng	personal name	personal name	23
17	其民	qí-mín	its-people	one's people	22
18	百里	bǎi-lǐ	hundred-a unit of distance	the size of land	22
19	曾子	céng-zi	personal name	personal name	21
20	父母	fù-mǔ	father-mother	parents	21
21	王者	wáng-zhě	king-person	emperors	21
22	大夫	dà-fū	bigness-man	one kind of official name,	20
23	其君	qí-jūn	its- minister	one's minister	19
24	昔者	xī-zhě	former times-person	former	19
25	周公	zhōu-gōng	personal name	personal name	18
26	子路	zi-lù	personal name	personal name	18
27	商書	shāng-shū	trade-book	the name of a book	17
28	四海	sìh-ǎi	four-sea	whole world or country	17
29	上下	shàng-xià	top-bottom	world	16
30	天子	tiān-zǐ	heaven-son	emperor	16
31	農戰	nóng-zhàn	farming-war	developing strategy	16
32	三軍	sān-jūn	three- armed force	three kind of armed force	15
33	國家	guó-jiā	country-home	state	15
34	皋陶	gāo-táo	personal name	personal name	15
35	管仲	guǎn-zhòng	personal name	personal name	15
36	伊尹	yī-yǐn	personal name	personal name	14
37	朋友	péng-yǒu	friend-friend	friend	14
38	武王	wǔ-wáng	personal name	personal name	14
39	其國	qí-guó	its-country	one's country	13
40	其心	qí-xīn	its-heart	one's heart	13

41	大國	dà-guó	bigness-country	sovereign state	13
42	孫子	sūnzi	personal name	personal name	13
43	道者	dào zhě	way-person	Taoist	13
44	其道	qí dào	its-way	one's way	12
45	賢者	xiánzhě	virtuousness-person	sage	12
46	其政	qí zhèng	its- politics	one's politics	12
47	千里	qiānlǐ	thousand-a unit of distance	far away	11
48	四方	sìfāng	four- direction	every where	11
49	子思	zi sī	personal name	personal name	11
50	戰者	Zhàn zhě	fight-person	fighter	11
51	百官	bǎi guān	hundred-official	officials of ranks and descriptions	11
52	知者	zhì zhě	be aware of-person	who's wise	11
53	禽獸	qínshòu	birds-beasts	beast in AC	11
54	萬乘	wàn chéng	ten thousand-chariot	many chariots , sovereign state	11
55	萬邦	wàn bāng	ten thousand-state	all states,	11
56	齊人	qí rén	Qi-people	people from Qi	11
57	他日	tā rì	his-day	his day	10
58	伯夷	bóyí	personal name	personal name	10
59	其事	qí shì	one's-business	one's business	10
60	其位	qí wèi	one's-position	one's position	10
61	其子	qí zi	one'- son	one's son	10
62	其身	qí shēn	one's-body	one's body	10
63	十者	shí zhě	ten-people	the ten	10
64	富者	fù zhě	richness-people	who's rich	10
65	左右	zuǒyòu	left-right	influence	10
66	死地	sǐdì	deadness-position	location in military	10
67	許子	xǔzi	personal name	personal name	10
68	三年	sān nián	three-year	three years	9
69	三者	sān zhě	three-people	the three	9
70	上帝	shàngdì	top-king	god	9
71	世主	shì zhǔ	lifetime-host	lord	9
72	仁義	rényì	benevolence- righteousness	righteousness	9
73	其親	qí qīn	one’s-parents or relative	one’s parents or relative	9
74	天命	tiānmìng	heaven-life	destiny	9
75	夫民	fū mín	husband-people	common people	9
76	夷子	yí zi	personal name	personal name	9
77	妻子	qīzi	wife-son	wife	9
78	官爵	guānjué	official-peerage	official post	9
79	小國	xiǎo guó	smallness-country	developing nation	9
80	弟子	dìzǐ	young brother-son	follower	9
81	技藝	jìyì	skill-skill	skill	9
82	日月	rì yuè	sun-moon	time	9
83	明君	míngjūn	wisdom-gentlemen	wise monarch	9
84	為人	wéi rén	act as-person	behavior	9
85	百工	bǎi gōng	hundred-labor	varied officers	9

86	鬼神	guǐshén	ghost-nerve	spirits	6
87	鳥獸	niǎo shòu	bird and beast	bird and beast	2
88	中庸	zhōngyōng	middle-to use	moderation	8
89	仁政	rénzhèng	benevolence-politics	policy of benevolence	8
90	其人	qí rén	its-people	others	8
91	其德	qí dé	one's-virtue	one's virtue	8
92	其志	qí zhì	one's-ambition	one's ambition	8
93	其知	qí zhī	one's-knowledge	one's knowledge	8
94	千乘	qiān chéng	thousand-chariot	chariot	8
95	右傳	yòu chuán	right-pass on	proper noun	8
96	善者	shàn zhě	kindness-person	good or kind person	8
97	天道	tiān dào	heaven-way	natural law	8
98	子之	zǐ zhī	personal name	personal name	8
99	我者	wǒ zhě	I or me-person	I or me	8
100	民力	Mínlì	people-strength	human capacity	8
101	農民	nóngmín	farming-people	peasant	8
102	顏淵	yányuān	personal name	personal name	8
103	黎民	límín	numerousness-people	common people	8
104	丈夫	zhàngfū	a unit of length-husband	husband	7
105	三月	sān yuè	three-moon	March	5
106	中國	zhōngguó	middle-country	China	7
107	什一	shí yī	ten-one	tenth	7
108	今日	jīnrì	today-today	today	7
109	仲子	zhòngzi	personal name	personal name	7
110	其言	qí yán	one's-words	one's words	7
111	勝者	shèngzhě	winning-person	winner	7
112	國人	guórén	country-people	people who were living DA-YIN	7
113	姦民	jiān mín	evil-people	evildoer	7
114	子夏	zi xià	personal name	personal name	7
115	宗廟	zōngmiào	ancestor-temple	ancestral temple	7
116	湯武	tāng wǔ	Shang-tang Chou-wu-wang	Shang-tang Chou-wu-wang	7
117	禮樂	lǐ yuè	rites and music	rites and music	7
118	詩書	shīshū	poem-book	title of book	7
119	貧者	pínzhě	poorness-person	pauper	7
120	三載	sān zài	three-load	three years	6
121	乃祖	nǎi zǔ	honor	honor	6
122	二者	èr zhě	two-person	the two	6
123	兄弟	xiōngdì	young brother-old brother	brothers	6
124	先知	xiānzhī	before-be aware of	farmer	6
125	兩者	liǎng zhě	two-person	both	6
126	兵者	bīng zhě	soldier-person	soldier	6
127	其一	qí yī	one of them	one of them	6
128	其家	qí jiā	one's-home	one's home	6
129	其故	qí gù	its-reason	reason	6
130	其時	qí shí	its-time	actually	6

131	其母	qí mǔ	one's-mother	one's mother	6
132	其然	qí rán	one’s-right	actually	6
133	其父	qí fù	one's-father	one's father	6
134	凶年	xiōng nián	terribleness-year	famine	6
135	前日	qiánrì	front-day	before	6
136	四岳	sì yuè	name of place	name of place	6
137	圍地	wéi de	be surround- position	location in military	6
138	學者	xuézhě	learn-person	scholar	6
139	宰我	zǎi wǒ	personal name	personal name	6
140	小子	xiǎozi	smallness-son	common people	6
141	山林	shānlín	mountain and forest	mountain and forest	6
142	庶人	shù rén	numerousness-person	peasant	6
143	忠信	zhōngxìn	honest-faith	loyalty	6
144	故知	gùzhī	accident-be aware of	old friend	6
145	春秋	chūnqiū	spring and autumn	spring and autumn	6
146	晏子	yànzi	personal name	personal name	6
147	智者	zhìzhě	wisdom-person	wise man	6
148	死者	sǐzhě	deadness-person	deceased	6
149	然友	rán yǒu	personal name	personal name	6
150	父子	fùzǐ	father and son	father and son	6
151	百畝	bǎi mǔ	hundred- a unit of area	vast land	6
152	義者	yì zhě	righteous-person	the righteousness	6
153	草木	cǎomù	grass-trees	plants	6
154	衆人	zhòngrén	many-people	common people	6
155	衢地	qú de	many-people	terrain in military	6
156	輕者	qīng zhě	light-person	the light	6
157	五者	wǔ zhě	five-person	the five	5
158	人心	rénxīn	people-heart	public feelings	5
159	仲尼	zhòng ní	personal name	personal name	5
160	侯王	hóu wáng	marquis-king	officer	5
161	兆民	zhàomín	omen-people	common people	5
162	八者	bā zhě	eight-person	the eight	5
163	其仁	qí rén	one's-benevolence	one's benevolence	5
164	其力	qí lì	its-force	one's force	5
165	其本	qí běn	its-origin	one's origin	5
166	其極	qí jí	its-top	its pole	5
167	其樂	qí lè	one’s-cheerfulness	one’s cheerfulness	5
168	匹夫	pǐfū	lone-husband	ordinary man	5
169	反間	fǎn jiàn	reversed-space	spy	5
170	古者	gǔ zhě	old-person	the before	5
171	商賈	shānggǔ	trade-business	merchant	5
172	善人	shànrén	kindness-people	the kind	5
173	土地	tǔdì	soil-land	territory	5
174	圮地	yí de	bridge-position	location in military	5
175	地形	dìxíng	land-shape	terrain	5

176	堯舜	yáoshùn	personal name	personal name	5
177	大臣	dàchén	bigness-official	minister	5
178	天時	tiān shí	heaven-time	laws of nature	5
179	天者	tiān zhě	heaven-person	heaven	5
180	夫婦	fūfù	husband-wife	couple	5
181	子孫	zǐsūn	son-grandson	descendant	5
182	子張	zi zhāng	personal name	personal name	5
183	尊賢	zūn xián	honor	honor	5
184	庶民	shùmín	numerousness-people	common people	5
185	敵人	dírén	enemy-people	enemy	5
186	明德	míng dé	wisdom-moral	commend bright morality	5
187	此時	cǐ shí	this-time	at moment	5
188	民利	mín lì	people-benefit	interests of the people	5
189	民壹	mín yī	people-one	concentration	5
190	洪水	hóngshuǐ	flood-water	flood	5
191	爵祿	juélù	official ranking and earning	official ranking and earning	5
192	玄德	xuán dé	personal name	personal name	5
193	王命	wángmìng	king-order	king's commandment	5
194	生民	shēng mín	growth-people	people	5
195	百世	bǎishì	hundred-lifetime	long period of time	5
196	皇天	huángtiān	king-heaven	heaven	5
197	神農	shénnóng	personal name	personal name	5
198	管叔	guǎn shū	personal name	personal name	5
199	股肱	gǔgōng	thigh-arm	right-hand man	5
200	至德	Zhì dé	the highest ethic	the highest ethic	5
201	葛伯	gé bó	personal name	personal name	5
202	虞書	yú shū	title of book	title of book	5
203	衆者	zhòng zhě	crowd-person	people	5
204	誠者	chéng zhě	honesty-person	honest	5
205	說者	shuō zhě	speaking-person	speaker	5
206	農者	nóng zhě	farming-person	farmer	5
207	道路	dàolù	way-road	roadway	5
208	遠人	yuǎn rén	distance-people	foreigner	5
209	鄰國	lín guó	neighbor-country	neighboring states or countries	5
210	重地	zhòngdì	heavy-position	location in military	5
211	重者	zhòng zhě	heavy-person	heavy	5
212	長者	zhǎngzhě	long-person	the old	5
213	險阻	xiǎnzǔ	danger-obstructive	danger	5
214	一日	yī rì	one-day	one day	4
215	三代	sāndài	three-generation	Xia, Shang and Chou	4
216	九族	jiǔzú	nine-race	nine generations	4
217	九江	jiǔjiāng	nine-river	name of place	4
218	五刑	wǔxíng	five- punishment	five forms of punishment	4
219	五穀	wǔgǔ	five-grain	grains	4
220	五色	wǔsè	five-color	colors	4

221	交地	jiāo de	joint-position	location in military	4
222	人君	rén jūn	people-gentleman	monarch, emperor	4
223	仁人	rén rén	benevolence-people	benevolent man	4
224	六淫	liù yín	six-obscenity	six excessing bad behaviors	4
225	六者	liù zhě	six-person	the six	4
226	其二	qí èr	its-two	Second	4
227	其利	qí lì	its-benefit	one's Profit	4
228	其居	qí jū	one's-home	one's home	4
229	其意	qí yì	its-meaning	It's meaning	4
230	其正	qí zhèng	its-position	the positive	4
231	其法	qí fǎ	its-law	the law	4
232	其用	qí yòng	its-use	It's use	4
233	其私	qí sī	its-private	the private	4
234	其行	qí xíng	one's -behavior	one's behavior	4
235	其門	qí mén	its-door	It's door	4
236	其餘	qíyú	its-remaining	the remaining	4
237	制度	zhìdù	system-degree	regulation, system	4
238	勇民	yǒng mín	bravery-people	valiant	4
239	十歲	shí suì	ten-annum	ten years	4
240	千人	qiān rén	the thousand-people	many people	4
241	千歲	qiānsuì	the thousand-annum	many years	4
242	君臣	jūnchén	gentleman-minister	dukes or minister	4
243	哀公	āigōng	personal name	personal name	4
244	四時	sì shí	four-time	four seasons	4
245	四難	sì nán	four-difficulty	punishment, law, arming and joint a war	4
246	國用	guó yòng	country-use	policy of use	4
247	地者	de zhě	land-person	the land	4
248	士卒	shìzú	solider-hurry	soldiers	4
249	夏書	xià shū	personal name	personal name	4
250	夙夜	sùyè	morning-night	often	4
251	大事	dàshì	bigness-thing	important thing	4
252	大學	dàxué	bigness-study	education	4
253	大德	dà dé	bigness-virtue	virtues	4
254	大王	dàwáng	bigness-king	monarch, emperor, king	4
255	大道	dàdào	bigness-way	principle, theory	4
256	夷狄	yídí	personal name	personal name	4
257	姦宄	jiānguǐ	evil-bad	malefactor	4
258	嬰兒	yīng'ér	baby-son	baby	4
259	學問	xuéwèn	study-knowledge	knowledge	4
260	將軍	jiāngjūn	general-army	general	4
261	山川	shānchuān	mountain-river	mountain	4
262	巧言	qiǎo yán	artfulness-word	blarney	4
263	弱者	ruòzhě	weakness-person	the weak	4
264	彊國	jiàng guó	boundary-country	powerful nation	4
265	後人	hòu rén	back-people	descendant	4

266	怯民	qiè mín	coward-people	untouchable	4
267	愚者	yúzhě	fool-person	the foolish	4
268	散地	sàn dì	looseness-position	location in military	4
269	數年	shù nián	number-year	years	4
270	新邑	xīn yì	new-city	new city	4
271	旌旗	jīngqí	flag-banner	flag	4
272	明主	míng zhǔ	wisdom-host	sagacious monarch or emperor	4
273	時子	shí zi	personal name	personal name	4
274	曾西	céng xī	personal name	personal name	4
275	有若	yǒu ruò	personal name	personal name	4
276	朝廷	cháotíng	court-courtyard	royal or imperial court	4
277	楚人	chǔ rén	suffering-people	people	4
278	此處	cǐ chù	this-place	here	4
279	死生	sǐ shēng	dead-life	importance	4
280	民務	mín wù	civil affairs	civil affairs	4
281	法者	fǎ zhě	law-person	the law, legal person	4
282	溝壑	gōuhè	gutter-gully	ravine	4
283	爭地	zhēng dì	fight-position	location in military	4
284	王子	wángzǐ	personal name	personal name	4
285	王道	wángdào	king-way	idea of governing	4
286	管氏	guǎn shì	personal name	personal name	4
287	終始	zhōngshǐ	end-beginning	ending and beginning	4
288	自然	zìrán	oneself-way	natural	4
289	蝨官	shī guān	louse-official	Eunuch	4
290	衣服	yīfú	clothier-garment	clothes	4
291	車馬	chē mǎ	vehicle-horse	carriage	4
292	軍市	jūn shì	arm-market	military market	4
293	軍爭	jūn zhēng	arm-fight	war strategy	4
294	輕重	qīngzhòng	lightness-heavy	weight, important	4
295	農夫	nóngfū	farming-husband	farmer, peasant	4
296	野人	yěrén	field-people	peasant	4
297	門人	ménrén	door-people	students	4
298	飲食	yǐnshí	drink and eating	having dinner	4
299	高后	gāo hòu	personal name	personal name	4
300					2643