A Cognitive Approach to Investigating Two-Plus-Two Constructions in Chinese Four-Character Idioms

Haoran Yang

This thesis is submitted in fulfilment of the requirements for the

degree of Doctor of Philosophy

Linguistics and English Language

Lancaster University

2022

Abstract

Chinese idioms comprise word strings of various lengths, ranging from three to eight characters (Luo, 2015). Four-character idioms (FCIs) constitute the largest group among all Chinese idioms. Different syntactic patterns have been identified among FCIs, namely, 1+1+1+1, 1+3, and 2+2, whereby each digit stands for the number of characters that constitute a syntactic unit. Among these, the 2+2 construction (henceforth, AABB) is found to be most widely distributed (Wang et al. 2013). Two types of 2+2 FCIs have been identified in the present study: (a) interchangeable 2+2 FCIs whose two units can replace each other (i.e., AABB or BBAA) and (b) non-interchangeable 2+2 FCIs whose two units cannot substitute each other (i.e., only AABB but not BBAA). For instance, 黑白混淆 (*hēi-bái-hùn-xiáo*: black-white-mix-confuse, "to garble things up like mixing black and white colours together") can be re-constructed as 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white), but 做贼心虚 (*zuò-zéi-xīn-xū*, become-thief-heart-empty, "to feel guilty like a thief having stolen something") cannot be re-constructed as 心虚做贼 (*xīn-xū-zuò-zéi*, heart-empty-become-thief).

Prior studies (Chen 2001; Su 2002; Tao 2002; Zuo 2006; Nall 2008) have identified combinatory relationships in FCIs from a Construction Grammar perspective (Goldberg 1995, 2006). However, none has provided an in-depth diachronic account of the differences between interchangeable and non-interchangeable 2+2 FCIs in terms of internal constituency and propositional act functions (Croft, 2001). Similarly, structural mismatches between AABB and BBAA constructions in interchangeable FCIs have also not been adequately addressed in the literature. Finally, not much attention has been given to the partly schematic negative 2+2 construction [不 (*bù*, not) A 不 (*bù*, not) B] in terms of its functions.

This thesis contains 8 chapters. Chapter 1 is the introduction which explains the aims and scope of this study. Chapter 2 is the literature review providing a description of idioms and idiomaticity. In particular, it deals with the basic concepts of Chinese

idioms' classification and the research motivation for the Chinese FCIs. Chapter 3 is the literature review about the Construction Grammar and explains how construction grammar can be applied to Chinese FCI research. Chapter 4 is devoted to data collection and methodology. Chapter 5 makes a comparison between interchangeable and non-interchangeable 2+2 FCIs, while Chapter 6 is centred on AABB and BBAA patterns of interchangeable FCIs. Chapter 7 gives an account of the 2+2 [*bù* A *bù* B] construction in terms of internal constituency, propositional act function, and semantic prosody. Finally, chapter 8 is for the findings and conclusion.

The present thesis argues that the internal constituency of Chinese 2+2 FCIs may affect their propositional act functions (cf. Croft 2001) in context and further lead a diachronic differentiation of interchangeable idioms vs non-interchangeable idioms. The former will appear to follow a directional path of constructional change, while the latter a non-directional one. This research also shows that three different mechanisms (attraction, differentiation, and substitution) may dictate the diachronic change between AABB and BBAA. This work aims to make a valuable contribution to the study of FCI constructions as it sets to explain (a) how interchangeable and non-interchangeable idioms evolve over time and (b) how the 2+2 [*bù* A *bù* B] construction shows a different behaviour than the general 2+2 constructions in terms of internal constituency, propositional act functions, and semantic prosody.

Finally, the present analysis sheds new theoretical light not only on the linguistic representation of Chinese FCIs based on constructional schematicity, but also on the diachronic relationship between idiomaticity and creativity. Corpus data were obtained from Xinhua Dictionary of Idioms (Xu, 2002), the BLCU Corpus Center (BCC), zhTenTen Corpus and the Centre for Chinese Linguistics, PKU (CCL) and data manipulation and analysis of FCIs was implemented with Rstudio.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to my supervisor Vittorio Tantucci, who took me as his Ph.D. student in the first place, and who had always offered me constructive and invaluable advice in the completion and improvement of my thesis. I was most grateful that he, as a true mentor, not only gave me the freedom to enquire into the topic of my research, but also showed me the kind of care and empathetic understanding that I most needed through the Covid-19 pandemic as well as all the ups and downs during my thesis writing. I was particularly indebted to him for helping me obtain the student funding from both the China Scholarship Council and Lancaster University, and also for kindly writing job reference letters for me, without which this doctoral thesis would not have been possible.

I would also like to extend my sincere thanks to the following scholars who have played a crucial role in my academic career. To Associate Professor Yanhua Wang (Hubei University of Education), who turned the dry and dreary linguistic theories (general linguistics) into an interesting subject matter during my undergraduate years. To Professor Wenli Ma (Wuhan University of Technology), built in me the solid and fundamental skills for linguistic research, later making it possible for me to be an exchange student at National Taiwan University (NTU). To Professor Karen Steffen Chung (NTU), who guided me into the field of phonetics and sociolinguistics. Lastly to Professors I-wen Su and Chia-rung Lu (NTU), who opened the door for me to the world of Cognitive Linguistics.

Indeed, I could not have undertaken this journey without the financial and spiritual support of my family, notably, my grandparents and parents. My grandparents are the kind of people who taught me important life lessons. Firstly, *learning is a life-long process*. At 85 years old, they still managed to learn to use the smartphones and necessary Internet skills to acquire new knowledge and connect with other family members. Secondly, *stay calm, sensible, and positive in the face of all challenges*. Life had not been easy for my grandparents, who survived not only World War II but also decades of poverty in China since 1949. However, they stood their ground and braved

all the difficult times with resolve and composure as they had faith that all suffering would eventually end and be met with happy tears. The year of 2022 is especially worth celebrating for two things: their Diamond Marriage Anniversary (60th) and my graduation.

My parents named me 浩然 (Hàorán), which means *honesty* and *integrity* in Chinese. These two qualities represent their expectations of me and have shaped what I am today: being kind, honest, and brave. My parents never force me to do anything that I am reluctant to do; instead, they teach me to negotiate and communicate with them whenever we have disagreements. Although they only received high school education, they set an example for me in every aspects of live. Especially, they are able to accept who I truly am and who I want to be. I also would like to thank all of my family members even they do not talk to me very often, but I know from the bottom of my heart that they are ready to support me whenever I need.

Lastly, I would also like to thank my colleagues and beloved friends who have helped, supported, and accompanied me through my four-years study period in the UK. These include, but are not restricted to, Lie Sun, Chuanxi Zou, Man Tian, Bo Liu, Hong Pan, Jiahui Sun, Beibei He, Yang Hong, TRAN Dieu Linh, Jiahao Li, Yulian Liang, Jia Zhang, Jiacheng Xie, Zhenhao Cao, Shengnan Liu, Yingnian Tao, Shijie Zhang, Liwen Bing, Haoyu Chen, Lucyann Featherstone, Jane Hulland, Joy Payne, Sayjda Talib, Ian Newman, John Bandman, and Andrew H. C. Chuang.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research background and significance of this study

Chinese idioms not only epitomize 5,000 years of the Chinese culture, but are considered one of the most significant heritages across China's history. Chinese idioms comprise word strings of various lengths, ranging from three to twelve characters (Liu, 2007). Confined to a limited number of characters, the Chinese idioms are nevertheless capable of coding a speaker's thought in a concise fashion. Actually, they remain widely used in today's Chinese speech communities. Different syntactic patterns have been identified among Four Character Idioms (FCIs), namely, 1+1+1+1, 1+3, 3+1 and 2+2. 喜怒哀乐 (*xǐ-nù-āi-lè*, happy-angry-sad-joyous, "ups and downs") is an example of 1+1+1+1 construction, which refers to four different human emotions encoded in four distinct characters. 打退堂鼓 (*dǎ-tuì-táng-gǔ*, beat-return-hall-drum, "to give up") is a 1+3 construction idiom, while 当局者迷 (*dāng-jú-zhě-mí*, when-game-person-baffled, "the person on the spot is baffled") is a 3+1 construction. An example of a 2+2 construction is 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river, "a long journey during which numerous obstacles are encountered") (see section 2.3). To specify the scope and nature of this proposed study, only the 2+2 relationship of the four-character idioms will fall into the concern of future analysis because 2+2 is the most widely distributed forms among all (Wang et al. 2013). Besides, the language that this study aims to investigate is that of Mandarin Chinese, in that speakers of other forms or dialects of the Chinese language (e.g., Taiwanese Chinese or Cantonese) may approach the language in a subtly different way in terms of speaker cognition.

The objective of this thesis is to discuss the different internal constituencies of the 2+2 constructions in Chinese FCIs based on Construction Grammar, with the aim to inquire into not only their formal but also cognitive and functional dimensions. This thesis also aims to investigate the various combinatory relationships between the two units. This thesis contributes to the study of Chinese idioms in a crosslinguistic

perspective, with a particular focus on the four-character idioms. Also, it offers a construction-based approach to these idioms and brings in recent work on creativity and diachronic change, that is, attraction, substitution and differentiation. Besides, it explores how the internal constituency in Chinese idioms affects their constructional meanings in a particular 2+2 relationship.

This thesis employs Goldberg's (1995) CxG approach, believing that the understanding of a Chinese four-character idiom must go beyond the sheer formal tradition, that is, cognitive and functional perspectives must be brought in to examine the construction of these idioms. In other words, the understating of a Chinese FCI should be based on the construction as a whole instead of viewing the four characters as distinct and unrelated lexemes. Construction Grammar, as argued by Su (2002), can be a practical approach to investigating the four-character Chinese idioms, in that it takes into account not only formal considerations but also cognitive and functional aspects of a construction, providing a more thorough explanation for their actual language use. Moreover, it brings to the fore innovative perspectives into cross-linguistic constructions.

By inspecting the Chinese four-character idioms from a usage-based angle, this study argues that internal constituency plays a key role in the development of Chinese FCIs regarding interchangeability, combinatory patterns, collocational preferences, propositional act functions, productivity, schematicity, constructional meaning, change and variation, as well as semantic prosody. The research will endeavor not only to account for cultural literacy of the Chinese language, but to shed new light on the typologically features of Chinese idioms from a usage-based perspective.

1.2 Key Notions related to this research

There are some important notions needed to be clarified before I proceed to elaborate my research methods and discussions. Firstly, it is important to make the distinction between type and token frequency as they are fundamental notions when

discussing the lexicon and language use (Berg, 2014). Baayen (2001) and Bybee (1985) have importantly distinguished type frequency from token frequency. According to them, type frequency refers to the number of different expressions a particular pattern has, while token frequency the number of times the same construction occurs in text. In other words, type frequency reflects the schematic representation of form and meaning, but token frequency language use. For example, the English definite article "the" is a construction with a single type frequency, but it is the most token-frequent instantiation in the contemporary language (Traugott & Trousdale, 2013).

Secondly, it is also important to make the distinction between synthetic languages and analytic languages because this research investigates Mandarin Chinese, which is an analytic one and which does not feature lexical inflections. Synthetic language is a language whose syntactic relations within sentences are expressed by inflection (e.g., tense, person, gender, number, mood, voice, and case) or by agglutination (word formation by means of morpheme, or word unit, clustering). Analytic (isolating) language refers to any language that uses specific grammatical words or particles rather than inflection to express syntactic relations within sentences (Tauli,1945, p.80-85). An analytic language is commonly identified as an isolating language.

Thirdly, the notions about productivity and schematicity need to be spelt out because they are two important elements when discussing the language change phenomena from a usage-based perspective. Productivity refers to knowledge about different lexical items which may be used in the slots of a construction (Perek, 2018). For example, affixation is productively used to mark the past tense in English (e.g. help — helped), but the past tense is sometimes marked by change in the stem vowel (e.g. stick — stuck). Therefore, "when new verbs are introduced into the language, their past tense is usually formed by the more productive method of affixation rather than by the vowel change — the past tense of *skype* ('to make a video call via the internet') is *skyped*, not, for example **skope* (based on write-wrote)" (Traugott & Trousdale, 2013, pp. 17-18). Schematicity refers to "the level of detail that is stored in the representation of a construction" (Perek, 2020, p. 144). In English, productivity tends to increase over

time because different lexicons can be fitted in the slots of a construction. The changes in productivity and schematicity of Chinese FCIs will be discussed in Chapter 5.

Fourthly, the concept of iconicity will be used to account for the interchangeability of Chinese FCIs. In cognitive linguistics, iconicity (Simone, 1995) refers to the similarity between a sign and its meaning. This underpins three iconicity principles – quantity principle, proximity principle, and sequential order principle (Croft, 1999). The quantity principle states that formal complexity corresponds to conceptual complexity, the proximity principle that conceptual distance tends to match with linguistic distance, and the sequential order principle that the sequential order of events described is mirrored in the speech chain.

Finally, the concept of grammaticalisation is also important and relevant to this research as it explains how Chinese FCIs evolve over time to become grammatical elements or markers. Grammaticalisation refers to the process by which words or constructions that originally had a non-grammatical or lexical meaning become grammatical markers that express grammatical relationships or categories. This process usually involves the gradual erosion of the original lexical or semantic content of the word or construction, and its evolution into a more abstract and grammatical form (Hopper & Traugott, 2003).

1.3 Outline of the Chapters

This thesis is structured as follows: Chapter 1 is the introduction which articulates the research background and significance of the study. Also, it introduces some key notions relevant to my research and brings in the research questions for this study.

Chapter 2 is the literature review providing a description of FCIs. In particular, it deals with (a) a definition of both idioms in general and Chinese idioms in terms of their similarities and differences, and (b) the combinatory patterns and classifications of Chinese idioms and the research motivation.

Chapter 3 is the literature review on Construction Grammar. More specifically, it discusses (a) the development of Construction Grammars and a general idea of

construction grammars as well as their central tenets, (b) how different construction grammars can be applied to Chinese FCI research.

Chapter 4 is devoted to data collection and methodology. Specifically, I introduce the corpora consulted in the thesis and explain how I select, normalize, and annotate Chinese idioms in a systematic way.

Chapter 5 makes a comparison between interchangeable and non-interchangeable 2+2 FCIs in terms of type frequency and token frequency. Also, this chapter illustrates the changes in token frequency, schematicity, and productivity of the 2+2 Chinese FICs over time.

Chapter 6 is centred on the AABB and BBAA constructions of interchangeable FCIs. In particular, it deals with (a) the methods that I use to distinguish AABB from BBAA, (b) the mechanisms (attraction, differentiation, substitution) that influence the diachronic change of the two forms, and (c) the types of mechanisms that most frequently apply to the interchangeables.

Chapter 7 gives an account of the 2+2 [*bù* A *bù* B] construction in terms of internal constituency, propositional act function, and semantic prosody. Specifically, it discusses how the 2+2 [*bù* A *bù* B] construction is different from regular 2+2 FCIs. Moreover, it investigates how internal constituency influences its constructional meaning and semantic prosody.

Finally, Chapter 8 is the conclusion in which I discuss the findings and limitations of my study as well as the suggestions for future research.

CHAPTER 2 Literature review on idioms and idiomaticity

2.1 Outline

This chapter reviews idioms in general and then with specific attention paid to Chinese idioms from a constructionist perspective. The literature review on idioms and idiomaticity can be divided into four parts. The first part gives a definition of both idioms in general and Chinese idioms in terms of their similarities and differences. The second part discusses the combinatory patterns of Chinese idioms regarding their distribution patterns. The third part provides two different ways to classify Chinese FCIs. Lastly, the fourth part reviews how idiomaticity is studied in earlier research and accounts for the research motivation for the Chinese FCIs.

2.2 Definitions of idioms in general and Chinese Idioms

The term idiom has basically two meanings; one meaning refers to "the ability to speak a fluent and appropriate version of a language" (Grant & Bauer 2004, p. 39), which is also referred to as "native-like selection" (Pawley & Syder 1983, p. 191). The second meaning is a widely quoted definition, which can be found in the Oxford English Dictionary (OED): A form of expression, grammatical construction, phrase, etc., peculiar to a language; a peculiarity of phraseology approved by the usage of a language, and often having a significance other than its grammatical or logical one. This definition obviously does not measure up for a precise and watertight definition of idioms since it is only vaguely paraphrased. Idioms are fixed expressions that reflect conventionalized pairings of form and meaning. The transparency of an idiomatic expression is gradient as idioms can be more or less compositional (Clausner & Croft, 1997; Cruse, 2000). For instance, a drop in the bucket (American English) or a drop in the ocean (British English) can refer to a very small quantity (OED). The meaning of said expressions could be obtained through a speaker's embodied experience with liquid and size, yet their holistic meanings go beyond mere combinations of semantic constituents. There are a number of important works on idioms and their relation to constructions. Fillmore

(1988) argues that idioms are a type of construction instead of simply fixed phrases. The idioms are combined with a more flexible frame that specifies the syntactic and semantic properties of the idiom in different contexts, while Nunberg (1994) argues that idioms should be regarded as patterns of lexical co-occurrence instead of syntactic constructions. This means idioms are created when certain words are associated with each other through repeated usage in particular contexts, which leads to the result that a fixed expression is not easily decomposable into its component parts. Besides, McGlone (1996) argues that idioms involve a degree of conventionalization and fixedness, and Coulson and Oakley (2005) argues that idioms are best understood as mappings between different domains of knowledge, rather than simply as fixed expressions. Overall, these studies have provided valuable insights into the nature of the idiom and the ways in which it is processed by the human mind.

Kracht (2011: 57) highlights the importance of compositionality in idiomatic language by arguing that "the meaning of a complex expression is a function of the meanings of its parts and the mode of composition by which it has been obtained from these parts". Swinney and Cutler (1979) argue that idioms can be treated as complex words, being represented and understood as a whole unit. For instance, *kick the bucket*, which has to be learned as a complete unit rather than treated as a semantic combination of all its constituent items. It is also posited that an idiom can receive different weights of contributions in representation from its constituent words (Gibbs & Nayak, 1989), and that an idiom cannot be understood without the semantic relations between its constituent words being processed. For instance, the idiom *pop the question* is itself about a specific question (composition-based), which is conventionally interpreted as proposing marriage. In other words, idiomaticity (Gibbs, 2002) is a matter of compositional gradience, that is, some idioms are rather transparent in meaning while others must be approached as a learned unit due to high semantic opacity. I agree with this idea that idioms show a degree of gradience. In Chinese, the idiom 不慌不忙 (*bù-huāng-bù-máng*, not-hustle-not-bustle, "calm and unhurried") is transparent in meaning, while 不破不立 (*bù-pò-bù-lì*, not-eradicate-not-build, "if there is no destruction,

there can be no construction") should be understood as a learned construction because it conveys a hypothetical condition, that is, 破 (*pò*, eradicate) is the condition for 立 (*lì*, build).

The frameworks of Makkai (1972) and Sinclair (1987) are the most representative of classical idiom studies, and they are profusely cited across different cognitive-linguistic fields. Makkai conducts a dichotomous study of English idioms where gradience is not taken into account. He divides idioms into two types: (a) idioms of encoding, and (b) idioms of decoding. According to his definition, an *idiom of encoding* is an expression which a speaker would not realize is a conventional way of saying what it means without knowing that fact. An *idiom of decoding*, on the other hand, is an expression whose interpretation could not be figured out by someone using only independently learned linguistic conventions. In other words, decoding idioms are those whose meanings could not be understood by simply knowing the words and grammar of a linguistic chunk (e.g., *kick the bucket, pull someone's leg, let the cat out of the bag*). Encoding idioms are expressions which make sense to the hearer/reader based on the meanings of the independent words of an idiom, but which he/she would not *a priori* know are ways of conveying those meanings in the language (*answer the door, perform surgery, wide awake, heavy smoker, deep voice, bright red, mindless chatter, distinct possibility, healthy attitude, blow one's nose*). Based on these two types, 不慌不忙 (*bù-huāng-bù-máng*, not-hustle-not-bustle, "calm and unhurried") is an encoding idiom, while 不破不立 (*bù-pò-bù-lì*, not-eradicate-not-build, "if there is no destruction, there can be no construction") is a decoding idiom. However, such decoding vs. encoding types are not widely used in studies of Chinese FCIs because this dichotomous approach rejects other possibilities between the two ends of its classification. That is, firstly, speakers of the same speech community may display different levels of literacy towards the same expression, and secondly, human language is oftentimes situated in a dynamic context rather than in a perfectly homogeneous state. Therefore, there may possibly exist a dynamic interpretation of the same word, much less an idiomatic phrase or expression. Indeed, little research can be carried out on Chinese FCIs on the basis of

this classification. Therefore, I propose three further diagnostics (i.e., interchangeability, internal constituency and symmetricity) in section 2.3.3 in order to address not only the interchangeability of Chinese FICs but also their propositional act functions and their internal constituency.

It is important to remark that there are different outlooks on the so-called Chinese idioms. It is generally considered problematic to simplistically categorize as "idioms" the Chinese characters (words) that are conventionally grouped into a specific pattern (based on alignments of different word counts). These fixed strings of characters may behave in ways similar to idioms in English or other Indo-European languages; however, they are different in some ways — in particular, in their scope and origin (of literary sources). Without a consistent view on the definition of the subject matter, most previous studies refer to said linguistic phenomenon (strings of characters) as "Chinese idioms" due to a common feature shared between regular idioms (idioms in general) and such fixed strings of words — that is, both are highly conventionalized chunk-like expressions.

For a working definition, I have turned to the Chinese dictionary 辞海 (*cí-hǎi*, *Cihai Dictionary of the Chinese Language*) and 现代汉语词典 (*xiàn-dài-hàn-yǔ-cí-diǎn*, *Modern Chinese Dictionary*). According to the Chinese dictionary 辞海 (*cí-hǎi*, *Cihai Dictionary of the Chinese Language*), a Chinese idiom is defined as a form of expression that could be used to convey one's thoughts, and that has become prevalent in a given speech community. Also, the sources of these idioms must derive mostly from the Chinese classics, poems, literary works, ballads or common sayings that are widely known to and used in a society. Based on this definition, firstly, people may challenge whether all prevalent expressions that can be used to express thoughts are idioms. For example, "No" is certainly prevalent in the speech community, and can be used to express thoughts (e.g., disagreement or disapproval), but it should not be recognized as an idiom. Secondly, people may challenge whether idioms must derive from literary works. This implies that only people with decent education get to use idioms; however, Chinese idioms are indeed used across different walks of life. In fact,

there are a large number of idioms that originate from the classics, such as *Water Margin, The Journey to the West, A Dream of Red Mansions*, and *Romance of the Three Kingdoms*. That is, those who have never received proper education may still be able to make use of the idioms in Chinese society as they are an entrenched part of the Chinese language.

Another dictionary, 现代汉语词典 (*xiàn-dài-hàn-yǔ-cí-diǎn, Modern Chinese Dictionary*), defines Chinese idioms as set phrases and short sentences usually appearing in pithy forms with concise meanings, which must be generally acknowledged in a given society, and which must have been in constant use for generations by the common folk of the language. This second definition is even broader than the first one as short sentences with concise meanings are also defined as idioms.

Based on 辞海 (cí-hǎi, *Cihai Dictionary of the Chinese Language*) and 现代汉语词典 (xiàn-dài-hàn-yǔ-cí-diǎn, *Modern Chinese Dictionary*), typical Chinese idioms are composed of four characters, although there remain a small number of Chinese idioms comprising three, five, six, seven, or even eight characters. For example, 莫须有 (*mò-xū-yǒu*, not-must-have, "baseless") is a three-character idiom, 隔行如隔山 (*gé-háng-rú-gé-shān*, differ by-field-just like-differ by-mountain, "difference in profession makes one feel worlds apart") a five-character idiom, 有眼不识泰山 (*yǒu-yǎn-bù-shí-tài-shān*, have-eyes-not-know-mountain Tai, "fail to recognise someone's great talent") a six-character idiom, 近水楼台先得月 (*jìn-shuǐ-lóu tái-xiān-dé-yuè*, close to-terrace-first-get-moon, "the advantage of being in a favored position") a seven-character idiom, and 己所不欲勿施于人 (*jǐ-suǒ-bù-yù-wù-shī-yú-rén*, yourself-where-not-want-not-impose-on-others, "do not do others what you would not have them do to you") an eight-character idiom.

Narrower uses restrict idiom to a particular kind of unit: one that is fixed and semantically opaque or metaphorical, or, traditionally, "not the sum of its parts", for example, *kick the bucket* or *spill the beans*. Such units are sometimes called pure idioms (Fernando and Flavell 1981: passim; Cowie 1988: 133). In broader uses, idiom is a general term for many kinds of multiword item, whether semantically opaque or not. In

fact, Chinese FCIs also exhibit narrower and broader uses. For example, 不破不立 (*bù-pò-bù-lì*, not-eradicate-not-build, "if there is no destruction, there can be no construction") is a narrower use because it is fixed and semantically opaque, while 不慌不忙 (*bù-huāng-bù-máng*, not-hustle-not-bustle, "calm and unhurried") is a broader use due to the opacity of the idiom. Such a distinction between narrow and broad uses implies that some expressions may or may not be regarded as idioms. Thus, it is essential to specify the rules for my data collection for this research (see section 4.3).

In conclusion, idioms, Chinese and English alike, are not mere fixed strings of words. The definition of Chinese idioms is broader than that of idioms in a general sense. Both Chinese and English idioms show some degree of gradience. The distinction between decoding and encoding idioms can be applied to Chinese FCIs, but this distinction is not the focus of my research. Instead, three other classifications based on interchangeability, constituency, and symmetricity are proposed in order to discuss the differences between interchangeables and non-interchangeables. Phrases may or may not be strictly regarded as idioms; therefore, it is important to define and identify the idioms to be analysed for this research (see Section 4.3).

2.3 Combinatory patterns of Chinese Idioms

Chinese idioms comprise three, four, five, six, seven, or even eight characters. Chinese three-character idioms take up 0.18% in all Chinese idioms, four-character idioms 95.11%, five-character idioms 1.16%, six-character idioms 1.01%, seven-character idioms 0.85%, and eight-character idioms 1.45% (Luo, 2015).

It is argued that FCIs have chalked up the largest share in all Chinese idioms，and all Chinese FCIs share one surface structure — that is, they are made up of four characters. However, four different combinatory patterns can be identified in FCIs, namely, 1+1+1+1, 1+3, 3+1, and 2+2 (each digit standing for the number of characters). Firstly, an example of 1+1+1+1 construction is 喜怒哀乐 (*xǐ-nù-āi-lè*, happy-angry-sad-joyous)，which refers to four different human emotions encoded in four distinct

characters (Chinese emotion words). In other words, the four characters of this construction are separate words where no syntactic correlation holds between any two individual characters that modulates the idioms' ultimate interpretation. Taking all four characters together, the idiom refers to the ups and downs in one's life.

Secondly, 打退堂鼓 (*dǎ-tuì-táng-gǔ*, beat-return-hall-drum, "to give up") is a 1+3 construction idiom. The idiom literally means "*beat the return drum*", but it should be metaphorically interpreted as "*to give up*" in actual language use. Semantically, this idiom functions on the 1+3 construction whose last three words must be taken as a holistic element because it is a noun phrase (退堂鼓, *tuì-táng-gǔ*, return-hall-drum, "backing out"). Without viewing the 打 (*dǎ*, beat) and 退堂鼓（*tuì-táng-gǔ*) as two elements, it would be difficult to arrive at a sound understanding of this idiom.

Thirdly, there is also a 3+1 construction in Chinese FCIs. For instance, 当局者迷 (*dāng-jú-zhě-mí*, when-game-person-baffled, "the person on the spot is baffled") must be understood as the combination of the first three characters and the very last character (迷, *mí*, baffled) because 当局者 (*dāng-jú-zhě*, when-game-person, "the person on the spot") is an NP. It would be impossible to understand the idiom if 当局者迷 (*dāng-jú-zhě-mí*) were realized as 当 (*dāng*, when) plus 局者迷 (*jú-zhě-mí*, game-person-baffled) or 当 (*dāng*, when) + 局 (*jú*, game) + 者 (*zhě*, person) + 迷 (*mí*, baffled); the former is a 1+3 construction, while the latter a 1+1+1+1 construction.

The 2+2 construction is found most widely distributed (93.2%) in all Chinese FCIs, while the other three constructions (1+1+1+1, 1+3, 3+1) altogether comprise only 6.8% (Luo, 2015). An example of a 2+2 construction is 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river, "a long journey during which numerous obstacles are encountered") which is formed of 千山 (*qiān-shān*, a thousand mountains) plus 万水 (*wàn-shuǐ*, ten thousand rivers). In a 2+2 construction, each unit comprises two characters: the first two characters form the first unit, and the other two characters the second. The interpretation of this idiom has to be predicated on the 2+2 construction. That is, without processing 千山 (*qiān-shān)* and 万水 (*wàn-shuǐ*) as two

units, it would be difficult to arrive at a holistic understanding of 千山万水 (*qiān-shān-wàn-shuǐ*), where both units jointly construct the meaning of this idiom.

To sum up, regular Chinese idioms comprise word strings of various lengths, ranging from three to eight characters. Four-character idioms (FCIs) constitute the largest group among all Chinese idioms. Different syntactic patterns have been identified among FCIs, namely, 1+1+1+1, 1+3, and 2+2. Among these, the 2+2 construction is found to be most widely distributed (Wang et al. 2013). The processing of the 2+2 construction is different from that of other constructions because the interpretation of an FCI lies in its patterns of internal constituency, that is, the patterns of 1+1+1+1, 1+3, 3+1, and 2+2. To specify the scope and nature of this study, only the 2+2 relationships of the FCIs will fall into the concern of future analysis (see Section 2.3.4 for the motivation).

2.4 Classifications of Chinese Four-character Idioms

As discussed in section 2.3.1, decoding and encoding types are not widely used in studies of Chinese FCIs. Therefore, I propose three classifications: interchangeability, internal constituency, and symmetricity. I then proceed to address the differences between interchangeables and non-interchangeables in terms of interchangeability and symmetricity (see Section 5.4) and the mismatch between propositional act functions and internal constituency (Section 5.5 and Section 7.3).

Firstly, in terms of interchangeablity, Chinese FCIs can be categorized into two types — interchangeable FCIs and non-interchangeable FCIs. The interchangeable 2+2 FCIs (interchangeables) are those whose two units can replace each other (i.e., AABB or BBAA), while the non-interchangeable 2+2 FCIs are those whose two units cannot be substituted for each other (i.e., only AABB but not BBAA). For instance, 黑白混淆 (*hēi-bái-hùn-xiáo*: black-white-mix-confuse, "to garble things up like mixing black and white colours together") can be re-constructed as 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white), but 做贼心虚 (*zuò-zéi-xīn-xū*, become-thief-heart-empty, "to feel

guilty like a thief having stolen something") cannot be re-constructed as 心虚做贼 (*xīn-xū-zuò-zéi*, heart-empty-become-thief). This classification helps to address the differences between interchangeables and non-interchangeables regarding interchangeability (see Section 5.4).

Secondly, Chinese 2+2 FCIs can be categorized into different types based on their internal constituency (Wang, 2019). Seven different types can be found in Chinese 2+2 FCIs (see below).

i. Type 1: [NP NP],

ii. Type 2: [AP AP],

iii. Type 3: [VP VP],

iv. Type 4: [[N V] [N V]],

v. Type 5: [[V N] [V N]],

vi. Type 6: [NP VP],

vii. Type 7: [VP NP].

All seven construction types can be found across both interchangeables (n = 428) and non- interchangeables (n = 428) (see 3.3 for idioms selection criterion). Table 2 gives exemplars of 428 interchangeables based on the seven FCI types, and Table 3 those across the 428 non-interchangeables.

I provide a sample of interchangeables based on the seven different FCI types in terms of internal constituency. As shown in Table 1, the seven types figure in all 428 interchangeables.

Interchangeable idiom exemplars	Types

才子佳人 (*cái-zǐ-jiā-rén*, an adorable couple of lovers)	Type 1[NP NP]
佳人才子 (*jiā-rén-cái-zǐ*, an adorable couple of lovers)	
光明正大 (*guāng-míng-zhèng-dà*, just and honorable)	Type 2 [AP AP]
正大光明 (*zhèng-dà-guāng-míng*, just and honorable)	
不闻不问 (*bù-wén-bù-wèn*, show no interest in something)	Type 3 [VP VP]
不问不闻 (*bù-wèn-bù-wén*, show no interest in something)	
胆战心惊 (*dǎn-zhàn-xīn-jīng*, tremble with fear)	Type 4 [[N V] [N V]]
心惊胆战 (*xīn-jīng-dǎn-zhàn*, tremble with fear)	
摆尾摇头 (*bǎi-wěi-yáo-tóu*, be well pleased with oneself)	Type 5 [[V N] [V N]]
摇头摆尾 (*yáo-tóu-bǎi-wěi*, be well pleased with oneself)	
黑白混淆 (*hēi-bái-hùn-xiáo*, to garble things up)	Type 6: [NP VP]
混淆黑白 (*hùn-xiáo-hēi-bái*, to garble things up)	Type 7 [VP NP]

Table 1 Exemplars of the seven types in interchangeables

Type 1 FCIs ([NP NP] construction) can be realized in the forms of 才子佳人 (*cái-zǐ-jiā-rén*, gifted-man-gifted-woman, "an adorable couple of lovers") and 佳人才子 (*jiā-rén-cái-zǐ*, gifted-woman-gifted-man, "an adorable couple of lovers"), whereby 才子 (*cái-zǐ*, gifted man) and 佳人 (*jiā-rén*, beautiful woman) are both noun phrases. Examples for Type 2 FCIs ([AP AP] construction) are 光明正大 (*guāng-míng-zhèng-dà*, light-bright-upright-big, "just and honorable") and 正大光明 (*zhèng-dà-guāng-míng*, upright-big-light-bright, "just and honorable"), whereby both 光明 (*guāng-míng*, honorable) and 正大 (*zhèng-dà*, just) serve as adjectival phrases. Type 3 FCIs ([VP VP] construction) can be realized in 不闻不问 (*bù-wén-bù-wèn*, not-listen-not-ask, "show no interest in something") and 不问不闻 (*bù-wèn-bù-wén*, not-ask-not-listen, "show no interest in something"), where 不闻 (*bù-wén*, not care about) and 不问 (*bù-wèn*, not ask about) are both verb phrases.

Type 4 FCIs ([[N V] [N V]] construction) feature such examples as 胆战心惊 (*dǎn-*

zhàn-xīn-jīng, gallbladder-shiver-heart-beat, "tremble with fear") and 心惊胆战 (*xīn-jīng-dǎn-zhàn*, heart-beat-gallbladder-shiver, "tremble with fear"), where both of the two units — 胆战 (*dǎn-zhàn*, gallbladder-shiver) and 心惊 (*xīn-jīng*, heart-beat) — can be analysed as the construction of [Noun plus Verb] in terms of internal constituency. Type 5 idioms ([[V N] [V N]] construction) can be illustrated by 摆尾摇头 (*bǎi-wěi-yáo-tóu*, wag-tail-shake-head, "be well pleased with oneself) and 摇头摆尾 (*yáo-tóu-bǎi-wěi*, shake-head-wag-tail, "be well pleased with oneself"), where both 摇头 (*yáo-tóu*, shake head) and 摆尾 (*bǎi-wěi*, wag tail) are examples of construction [Verb plus Noun].

Type 6 FCIs ([NP VP] construction) can be illustrated as 黑白混淆 (*hēi-bái-hùn-xiáo*, black-white-mix-confuse, "to garble things up"), whose first two characters together (黑白, *hēi-bái*, black and white) acts as a noun phrase, while whose last two characters together (混淆, *hùn-xiáo*, mistake with) a verb phrase. An example for type 7 FCIs (VP+NP construction) is 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white, "to garble things up"), where the first two characters together (混淆, *hùn-xiáo*, mistake with) constitute a verb phrase, and where the last two characters together (黑白, *hēi-bái*, black and white) a noun phrase.

What needs to be noted is that in types 1, 2, 3, 4, and 5, both units in a construction belong in the same word class (phrasal category), that is, [NP NP], [AP AP], [VP VP], [[N V] [N V]], and [[V N] [V N]]. In other words, inter-switching the positions of the two units of an idiom does not necessarily lead to any change in the word-class of the internal constituents (phrasal category). For example, both 才子佳人 (*cái-zǐ-jiā-rén*, gifted-man-beautiful-woman) and 佳人才子 (*jiā-rén-cái-zǐ*, beautiful-woman-gifted-man) still operate as the [NP NP] construction. They are two functionally similar forms of an interchangeable idiom (more in-depth discussions of the differences between the two functionally similar forms will be found in Chapter 6).

However, the case with types 6 and 7 is different due to the fact that inter-switching the two units of an idiom in types 6 and 7 shall lead to the change of the constituents' word class. Take for example, 黑白混淆 (*hēi-bái-hùn-xiáo*, black-white-mix-confuse, "to garble things up") and 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white, "to garble

things up"), where the former is the combination of NP (黑白, *hēi-bái*) and VP (混淆, *hùn-xiáo*) while the latter is one of VP (混淆, *hùn-xiáo*) and NP (黑白, *hēi-bái*). Therefore, when the constructional form changes (Types 6 and 7), the grammatical features should change accordingly.

I also provide a sample of non-interchangeables based on the seven different FCI types. As shown in Table 2, the seven types figure in all 428 non-interchangeables.

Idiom	Types
能工巧匠 (*néng-gōng-qiǎo-jiàng*, a skilled craftsman)	Type 1[NP NP]
富丽堂皇 (*fù-lì-táng-huáng*, splendorous and majestic)	Type 2 [AP AP]
吹吹打打 (*chuī-chuī-dǎ-dǎ*, all the instruments of music were played)	Type 3 [VP VP]
兔死狐悲 (*tù-sǐ-hú-bēi*, to have sympathy with a like-minded person in distress)	Type 4 [[N V] [N V]]
量体裁衣 (*liàng-tǐ-cái-yī*, act according to actual circumstances)	Type 5 [[V N] [V N]]
昙花一现 (*tán-huā-yī-xiàn*, a flash in the pan)	Type 6 [NP VP]
不修边幅 (*bù-xiū-biān-fú*, not caring about one's appearance)	Type 7 [VP NP]

Table 2 Exemplars of the seven types in non-interchangeables

Consider the case of 能工巧匠 (*néng-gōng-qiǎo-jiàng*, capable-worker-skilled-craftsman, "a skilled craftsman"), which corresponds to a Type 1 ([NP NP] construction). The two NPs of the idiom — 能工 (*néng-gōng*, capable worker) and 巧匠 (*qiǎo-jiàng*, skilled craftsman) — cannot be inter-switched. Type 2 ([AP AP] construction) can be illustrated by 富丽堂皇 (*fù-lì-táng-huáng*, rich-splendid-magnificent-grand, "splendorous and majestic), where both 富丽 (*fù-lì*, "splendorous") and 堂皇 (*táng-huáng*, majestic) function as adjective phrases, taken together a two-fold AP modifier. An example of Type 3 ([VP VP] construction) is 吹吹打打 (*chuī-chuī-dǎ-*

dǎ, blow-blow-beat-beat, "all the instruments of music were played"), where 吹吹 (*chuī-chuī*, blow-blow) and 打打 (*dǎ-dǎ*, beat-beat) are both verb phrases. These two verb phrases cannot be inter-positioned.

Type 4 ([[N V] [N V]] construction) can be illustrated by the idiom 兔死狐悲 (*tù-sǐ-hú-bēi*, rabbit-die-fox-grieves, "to have sympathy with a like-minded person in distress"), where both 兔死 (*tù-sǐ*, rabbit-die) and 狐悲 (*hú-bēi*, fox-grieve) are the [Noun plus Verb] construction. An example of Type 5 ([[V N] [V N]] construction) is 量体裁衣 (*liàng-tǐ-cái-yī*, measure-body-tailor-suit, "act according to actual circumstances"), where 量体 (*liàng-tǐ*, measure-body) and 裁衣 (*cái-yī*, tailor the suit) both fall into the [Verb + Noun] construction. Type 6 ([NP VP]) can be exemplified by the idiom 昙花一现 (*tán-huā-yī-xiàn*, epiphyllum-flower-once-appear, "a flash in the pan"), where the first two characters together form a noun phrase 昙花 (*tán-huā*, "epiphyllum"), and the last two characters a verb phrase 一现 (*yī-xiàn*, "blossom once"). The idiom 不修边幅 (*bù-xiū-biān-fú*, not-prune-margin-cloth, "not caring about one's appearance") falls within Type 7 ([VP NP]) because the first unit 不修 (*bù-xiū*, "not care") is a VP, while the second 边幅 (*biān-fú*, "dress and manner") is an NP.

All seven types can be found across both interchangeables (n = 428) and non-interchangeables (n = 428). However, the numbers for each FCI type are different between the two. Section 4.2.3 will discuss the distribution differences in the seven types. This classification also helps to investigate the mismatch between propositional act functions and internal constituency (Section 5.5 and Section 7.3)

Thirdly, regarding symmetricity, Chinese FCIs can be categorized into two groups — symmetrical FCIs and asymmetrical FCIs. More specifically, symmetrical FCIs are idioms whose two units share the same word class (Types 1, 2, 3, 4, and 5), while asymmetrical FCIs are those whose two units do not belong in the same word class (Types 6 and 7) (Wang, 2011).

The classifications based on interchangeability and symmetricity may be conducive to investigating the differences between interchangeables and non-

interchangeables. This is because some symmetrical idioms are non-interchangeables, while some asymmetrical idioms are interchangeables (see Section 5.2.3).

One the one hand, if a 2+2 FCI is a symmetrical structure (Types 1, 2, 3, 4 and 5), its two units can be inter-switched in word order. For instance, the Type 1 FCI (NP + NP) 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river) can be re-constructed as 万水千山 (*wàn-shuǐ-qiān-shān*, ten thousand-river-a thousand-mountain). However, the syntactic relations between the two units of the 2+2 FCIs are more complicated than they appear to be. There are certain constraints that strongly regulate their word order. Although some 2+2 FCIs operate on symmetrical structures, they are in reality non-interchangeable FCIs in nature (e.g., Type 4 FCIs: [[N V] [N V]]). For instance, 水涨船高 (*shuǐ-zhǎng-chuán-gāo*, water-rise-boat-lift) cannot be re-aligned as 船高水涨 (*chuán-gāo-shuǐ-zhǎng*, boat-lift-water-rise) due to cause-and- effect iconicity, which implies that 水涨 (*shuǐ-zhǎng*, water-rise) is a necessary and sufficient condition for 船高 (*chuán-gāo*, boat-lift). Therefore, 水涨 (*shuǐ-zhǎng*, water-rise) must happen before 船高 (*chuán-gāo*, boat-lift), which makes this symmetrical idiom non-interchangeable.

On the other hand, there are some idioms whose two units are asymmetrical (Types 6 and 7), but the two units can still be inter-switched. For example, Type 6 FCI ([NP VP]) 黑白混淆 (*hēi-bái-hùn-xiáo*: black-white-mix-confuse, "to garble things up like mixing black and white colours together") can be re-constructed as Type 7 FCI ([VP NP]) 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white).

In conclusion, there are four methods for classifying Chinese FCIs (i.e., manner of coding, interchangeability, internal constituency, and symmetricity). The very first (encoding vs. decoding) is not considered because it is not widely used in the studies of Chinese FCIs. This research has adopted the remaining three methods, which have provided a solid theoretical foundation (explanatory adequacy) for answering my research questions.

2.5 Research motivation for the Chinese Four-character Idioms

Despite some prior studies on Chinese idioms (Wang & Yu, 2010; Wu, 2016; Qi, 2016; Lu, Tsai, Su & Liu, 2018), there has not been much literature on cognitive approaches to analyzing or re-analyzing Chinese FCIs, let alone the 2+2 constructions. More specifically, Wang & Yu (2010) argued that it is important to know the emotion prediction of an idiom (appreciative, neutral, and derogatory), but their research only discussed how internal constituency may directly affect emotion orientation from a machine learning approach by influencing the types of features that are available and effective for predicting emotional content. By understanding the unique features of different languages, machine learning models can be tailored to better capture the emotional nuances of language use. However, their research did not take pragmatics into consideration, while pragmatics arguably is key to interpreting Chinese idioms (Qi, 2016). Wu (2016) focused on the word order of antonyms in Chinese FCIs, arguing that iconicity plays a key role to determine which word comes before the other. For example, in the top-down relationship, "up" always comes before "down" because the word "up" is often associated with concepts such as height, ascent, and positivity, while the word "down" is associated with concepts such as depth, descent, and negativity. While this is a reasonable observation, Wu's data came entirely from Chinese dictionaries. To extend Wu's research, I have collected data not only from dictionaries but also from the BLCU (Beijing Language and Culture University) Corpus Centre (BCC), Centre for Chinese Linguistics, Peking University (CCL), and zhTenTen (see Section 4.2). Furthermore, I shall also distinguish complimentary antonyms from relational antonyms and discuss the word order of synonyms in FCIs (Section 7.4). Jiang (2020) found that Chinese speakers process idioms differently depending on whether they are presented in the traditional or simplified character form. This suggests that the visual complexity of the characters plays a role in how idioms are processed. Although Qi's research (2016) relied on a corpus approach and emphasised the importance of pragmatics in understanding Chinese FCIs, her research is restricted to four specific 2+2 idioms (千山万水, *qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river vs.

万水千山 *wàn-shuǐ-qiān-shān* and 千言万语, *qiān-yán-wàn-yǔ*, a thousand-word-ten thousand-speech vs. 万语千言, *wàn-yǔ-qiān-yán*, ten thousand-speech-a thousand-word). Instead, my research is not restricted to a particular 2+2 idioms, but focuses on 428 2+2 construction idioms.

Chen (2001) posited that Chinese idiomatic expressions should be treated as constructional idioms, arguing that one way to interpret the FCIs is by characterizing their specific patterns, and then treating the remaining characters as supplementary information that contributes to this construction. For example, 千 X 万 Y (*qiān-X-wàn-Y*, a thousand-X-ten thousand-Y) is a construction, where X and Y can be replaced by different paradigmatic items. In this manner, the construction can then generate idioms like 千秋万代 (*qiān-qiū-wàn-dài*, a thousand-generation-ten thousand-generation, "from generation to generation"), 千刀万剐 (*qiān-dāo-wàn-guǎ*, a thousand-cut-ten thousand-cut, "punishment by hacking process"), 千推万阻 (*qiān-tuī-wàn-zǔ*, a thousand-shirk-ten thousand-impede, "unwilling to do"), 千姿万态 (*qiān-zī-wàn-tai*, a thousand-gesture-ten thousand-posture, "in thousands of postures"), and 千军万马 (*qiān-jūn-wàn-mǎ*, a thousand-army-ten thousand-horse, "magnificent army with thousands of men and horses"). The FCIs, therefore, can be seen as a rather productive construction. In other words, such idiom-like constructions can operate in a partly schematic fashion (Fillmore et al., 1988; Taylor, 2002) in that they are composed partly of constructional patterns and partly of distinct semantic constituents, which sheds light on the study of the partly schematic [不 (*bù*, not) A 不 (*bù*, not) B] construction. Similarly, Lu et al. (2018) studied the partly schematic construction 大 X 大 Y (da-*X*-da-*Y*), claiming that the understanding of different constructional meanings of idioms is not only difficult but also strenuous and complicated because constructions are innovative and dynamic in nature. In the process of understanding an idiom, people can often arrive at many novel extended meanings based on this idiom's conventional interpretation. In other words, language is never a static phenomenon, but rather an epiphenomenon that emerges out of a speaker's mind (Hopper, 1988); hence, language meaning is a dynamic

on-going process. Of course, there are connections between the original meaning and the extended meaning based on the same composite patterning of an idiom.

There are some other studies on Chinese FCIs from a different perspective. Kong (2014) only focused on the idioms containing numbers, while the rhetorical features of Chinese FCIs were investigated by Tsou (2012). Some scholars are interested in the study of Chinese idioms containing colors (Meng, 2022), while Lai (2015) studied the 1+3 construction. Moreover, Zuo (2006) investigated the transition from a 1+3 construction to a 2+2 construction. There are gaps in the existing literature on the cognitive processes involved in the comprehension and production of Chinese idioms. One such gap is the lack of empirical investigation into the cognitive processes involved in both comprehending and producing Chinese idioms. Additionally, more research is needed to explore the relationship between Chinese idioms and the cultural values and beliefs that they reflect. Furthermore, there is a need for more cross-linguistic studies of idioms to compare how different cultures conceptualize the world through their use of language. Although some prior studies may have noted the existence of combinatory/compositional relationships in FCIs based on Construction Grammar (Goldberg 1995, 2006), none has provided an in-depth diachronic account of the differences between interchangeable and non-interchangeable 2+2 FCIs in terms of internal constituency or propositional act functions (Croft, 2001). Similarly, structural mismatches between AABB and BBAA constructions in interchangeable FCIs have not been adequately studied. Moreover, not much attention has been given to the partly schematic 2+2 construction [*bù* A *bù* B] in terms of its functions. By investigating the Chinese FCIs in a 2+2 construction, my research topics give new insights into the nature of constructions and diachronic change in terms of token frequency and functions that idioms serve in different contexts. Studying Chinese idioms in such a cognitive manner can help us to understand the cognitive processes that underlie language use and explains the ways in which Chinese speakers conceptualize and think about the world.

Hence, this thesis seeks to tackle native Chinese speakers' sense-making of the FCIs with reference to their structural, cognitive, and functional aspects, taking into

account components of culture, history, and properties of language.

2.6 Summary

Chinese idioms and English idioms share similarities and differences. There are different combinatory patterns and classifications of Chinese FCIs. Four different syntactic patterns have been found in FCIs, namely, 1+1+1+1, 1+3, and 2+2. Among them, the 2+2 construction is the most widely distributed construction. When it comes to the classifications of Chinese FCIs, Chinese FCIs can be categorized into interchangeables and non-interchangeables in terms of interchangeablity. They can also be categorized into symmetrical FCIs and asymmetrical FCIs regarding symmetricity. Besides, Chinese 2+2 FCIs can be categorized into different types based on their internal constituency, that is, i. Type 1 [NP NP], ii. Type 2 [AP AP], iii. Type 3 [VP VP], iv. Type 4 [[N V] [N V]], v. Type 5 [[V N] [V N]], vi. Type 6 [NP VP], and vii. Type 7 [VP NP]. The motivations for this research are based on the literature gaps in the field, and this thesis contributes to the typologically features of Chinese idioms from a usage-based perspective.

Chapter 3 Literature review on the Construction Grammar

3.1 Outline

This chapter is about the theoretical frameworks, which can be divided into two parts (3.2-3.3). The first part (3.2) introduces the development of Construction Grammars proposed by different scholars and gives a general idea of construction grammars as well as their central tenets. The second part (3.3) accounts for Construction Grammars' general applications and explains how they are applied to Chinese FCIs. In other words, this section specifies the approaches I adopt for this research because these different constructionist frameworks each have their own strengths that may contribute.

3.2 Different constructionist frameworks

Construction grammars are a group of theories within the field of Cognitive Linguistics. There are some representative frameworks. I will present them in chronological order.

Berkeley Construction Grammar (BCG) was first developed in the late 1980s to early to mid-1990s (Fillmore, 1985, 1988; Fillmore, Kay & O'Connor, 1988; Michaelis, 1994; Michaelis & Lambrecht, 1996; Lakoff, 1987). An early example of "There-Constructions" was given by George Lakoff (1987, p.465). He argued that "a cognitive grammar can provide both an adequate description and explanation of the complexities of there-constructions, while other theories cannot". In his book *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (1987), he claimed that there is a major difference between the deictic (e.g., There's Harry with his red hat on) and existential (e.g., There was a man shot last night) there-constructions. The former is a locative adverb that picks out a location and is understandable only relatively to the context when the sentence is uttered, while the latter is not location but the existence of an event. Besides, other differences between the two types of there-constructions

involve subjecthood, negatability, embeddability, and alternation (Lakoff, 1987, p. 468). In the end, he argued that the meaning of an expression derived not only from the meaning of its parts, but also from the pragmatic meaning of the construction. The main idea put forth by Construction Grammar is that our understanding of language can be completely explained by our knowledge of constructions. Constructions are symbolic units that establish a connection between linguistic forms and their meanings. (Hilpert, 2014). That is, form-meaning parings are the fundamental units of the human language.

Secondly, another representative approach is called Sign-Based Construction Grammar (SBCG), which was developed by Paul Kay, Ivan Sag, and Laura A. Michaelis. The innovation of SBCG is the introduction of the distinction between signs and constructs (Sag, 2007). The most important claim of SBCG is that the lexicon provides a model for the syntax-semantics interface instead of viewing syntax, semantics, and lexicon as independent modules. As argued by Sag (2007, p. 70), SBCG can be recognised as a formalized version of BCG in that SBCG has "clearer empirical prediction and falsifiability, enhanced comparability of analyses across languages, and a general theoretical clarity". Recent SBCG research has extended to the lexicalist model of idiomatically combining expressions (Sag, 2012).

The next important framework is Goldberg's Construction Grammar (CxG). Goldberg (1995) highlighted four fundamental principles that distinguish CxG from Generative Grammar: (a) examining phrasal and clausal phenomena as grammatical constructions, (b) prioritizing the surface form of expressions and avoiding transformations or derivations, (c) asserting that constructions form an interconnected network, and (d) recognizing and carefully analysing crosslinguistic variations and the influence of general cognitive processes in developing crosslinguistic generalizations. (Hoffmann & Trousdale, 2013). In her view, constructions are connected to each other in networks through four inheritance relations — polysemy link, subpart link, metaphorical extension, and instance link (Goldberg, 1995). The notion of polysemy link focuses on the sematic relations between a particular construction and its extensions, while the subpart link is a construction that is the subpart of another

construction, but that exists independently. The instance link refers to a "special case" of a construction, and the metaphorical extension link is found when two constructions are related by a metaphorical mapping (Goldberg, 1995, p. 75-81). These four inheritance relations can be used to explain how constructions are related with one another.

Radical Construction Grammar (RCG) was proposed by William Croft (2001). His RCG is based on the idea that form is semantically motivated. It deals with the internal structure of constructions and takes cross-linguistic factors into account. It argues that syntactic categories are not only language-specific but also construction-specific (Croft, 2001). RCG offers innovative approaches to grammatical categories, generalizations and universals, also it takes both language-internal and cross-linguistic variation into consideration. That is, word classes are gradient and highly dependent on propositional act functions (see Section 3.4; see also Hollmann 2012, 2013 on the key role of psychological sub-schemas for word class differentiation).

The next approach is Cognitive Grammar (CG) proposed by Ronald Wayne Langacker in 2008. CG mainly focuses on the semantic meanings of constructions. Langacker argued that even the parts of speech (abstract grammatical units) are semantically motivated and involve certain conceptualizations.

Two other frameworks that are also worth mentioning are Fluid Construction Grammar (FCG) and Embodied Construction Grammar (ECG). Luc Steels (2011, p. 3) developed FCG as a framework enables researchers to show their discoveries in a precise manner and to evaluate the consequences of their theories in terms of language parsing, production, and acquisition. (Hoffmann & Trousdale, 2013). ECG was developed by Benjamin Bergen and Nancy Chang. The key point of ECG is that mental simulation plays a crucial role in language processing.

In summary, there are seven representative frameworks, that is, Berkeley Construction Grammar, Sign-Based Construction Grammar, Goldbergian Construction Grammar (CxG), Radical Construction Grammar (RCG), Cognitive Grammar (CG), Fluid Construction Grammar (FCG), and Embodied Construction Grammar (ECG). Of

all the inventory-based approaches to grammar, this study will base its theoretical analytics and metrics primarily on Goldberg's Construction Grammar (CxG), William Croft's Radical Construction Grammar (RCG), and Langacker's Cognitive Grammar (CG), in hopes of using such cognitive approaches to account for the different patterns found in Chinese four-character idioms with reference to form and meaning. Section 3.3 will give a detailed discussion about how said three Construction Grammars can be applied to the study of Chinese idioms.

3.3 Construction Grammars and Chinese idioms

This section aims to explain 1) the general applications of Construction Grammars, 2) how the three primary approaches that are adopted in this study — Construction Grammar (CxG), Radical Construction Grammar (RCG), and Cognitive Grammar (CG), and 3) why this research have taken a construction-based approach.

Goldberg (1995) proposed the following definition for identifying a grammatical construction:

> C is a CONSTRUCTION iffdef C is a form-meaning pair <Fi, Si> such that some aspect of Fi (Form) or some aspect of Si (Semantics) is not strictly predictable from C's component parts or from some other previously established constructions. (p. 4)

Goldberg (2006) later essentially modified her definition in that certain expressions can be realized without being idiosyncratic. Therefore, she broadened the definition of constructions as follows:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions

even if they are fully predictable as long as they occur with sufficient frequency. (p. 5)

Chinese idioms can be undertaken in this modified Goldbergian framework. For instance, 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river) literally means "a thousand mountains and ten thousand rivers". In actual language use, however, this idiom should be interpreted as "a long journey during which numerous obstacles are encountered". The construal of the idiom should not be realized merely based on the four characters viewed as distinct lexemes, but on the construction itself which embraces aspects of syntax, semantics, and pragmatics (Goldberg, 1995, 2006). The understanding of this idiom depends on two cognitive processes. Firstly, the four characters in the idiom are grouped into two units — 千 (*qiān*, a thousand) and 山 (*shān*, mountain) as a unit and 万 (*wàn*, ten thousand) and 水 (*shuǐ*, river) as another. Secondly, the two units are integrated as a whole (construction). This construction's meaning (a long journey during which numerous obstacles are encountered) is not largely predictable from its four distinct components (千 *qiān*, 山 *shān*, 万 *wàn*, 水 *shuǐ*), but depends on how these components are processed as a whole and how the construction is used in context. Therefore, Chinese idioms can be regarded as constructions as they completely meet the Goldbergian definition.

Next, I will illustrate the general applications of CxG and how it is used to investigate Chinese idioms. CxG is widely used to investigate human language activities in the enterprise of cognitive linguistics. It helps explain such linguistic phenomena as sentences, idioms, morphological items, information packaging, language processing, language acquisition, and language change, depending on which constructional approach is being employed (Hilpert, 2014). For example, CxG helps explain how a native speaker may exploit the "way-construction" even if a verb is conventionally viewed as an intransitive one (Goldberg 1995), as seen below:

(1) The boy cried his way into his bedroom.

(2) Sally smiled her way into the debate.

Cry and smile are both found as intransitive verbs in the dictionary, that is, neither of the two is syntactically allowed to take a direct object. However, the way-construction provides a "solution" to such syntactic constrains. *The boy cried his way into the bedroom* literally means the boy walked into his bedroom while he was crying, and *Sally smiled her way into the debate* is understood as Sally walked (or joined) the debate with a smile on her face — metaphorically referring to a specific personality trait of hers, that is, possibly her confidence. The two sentences can be schematised as [Subj V-ed PPron *way into* NP]. A schema is a linguistic pattern which specifies a potentially infinite multitude of phrases, sentences, or arguments (Corcoran, 2006). [Subj V-ed PPron *way into* NP] is regarded as a partly schematic pattern consisting of a fixed part (*way into*) and a schematic/open part (subject, personal pronoun, NP). The partly schematic "way-construction" allows intransitive verbs to follow noun phrases. Similarly, this approach can be used to investigate partly schematic Mandarin idiomatic constructions such as [不 (*bù, not*) A 不 (*bù, not*) B], which comprises a fixed part ([不 (*bù, not*)]) and a schematic part (A and B can be replaced by different nouns, adjectives and verbs) (see detail in Chapter 7).

As noted by Hilpert (2014), idioms are productive expressions that permeate ordinary language, not mere fixed strings of words. CxG then seeks to address idiomatic expressions by considering (1) whether idioms deviate from canonical patterns, (2) whether they carry non-compositional meaning, (3) whether they have idiosyncratic constraints, and (4) whether they have collocational preferences. These can also be used to examine Chinese FCIs. For example, Section 2.3 discusses the four different combinatory patterns of Chinese idioms as opposed to canonical patterns, and Section 7.3 identifies certain idioms are frequently used to modify certain nouns or verbs (i.e., collocational preferences).

CxG can also be used to research language variation and change (constructional variation) in investigation of (1) whether there is more than one way to represent the same construction, (2) variation in syntactic constructions (e.g., relative clauses), and (3) the analysis of variation between distinct constructions. Moreover, CxG can be also

used to analyze constructional variation across groups of speakers as well as constructional changes across time (Hilpert, 2014). Chinese FCIs feature different types of constructional variation, which is discussed in Chapter 6. For example, two patterns (AABB and BBAA) have been identified in the interchangeable idioms. These two patterns may or may not share the same syntactic features, and each pattern exhibits a different tendency in terms of token frequency through time (see detailed discussion in Chapter 6).

CxG is mostly applied to syntax-related studies. It is even used to analyze morphological patterns, particularly non-conventional morphemes, such "gl" in "glitter" (Croft, 2001). CxG is used to study morphological productivity, paradigmatic organization, non-compositional meanings, and simultaneous affixation; moreover, it provides solutions to such morphological puzzles as affix ordering and compounding. In fact, Chinese FCIs whose units and components are not grouped randomly, but are organized in a certain logical manner (see Section 7.2) or some syntactic ordering (see Section 2.3). Section 5.4 also discusses the productivity of interchangeable and non-interchangeable FCIs.

Taking a further step, CxG seeks to also explore pragmatic dimensions of language such as information packaging, presupposition and assertion, activation, and topic and focus. Other related issues may include cleft constructions, dislocation and related constructions, and island constraints (Hilpert, 2014). Moreover, Construction Grammars posit that "the interpretation of linguistic utterances can involve an interaction of grammar and context which vastly exceeds in complexity, formal structure and wealth of interpretive content" (Kay, 2001, p.1). In fact, this application can also be used to examine Chinese FCIs because one form of an interchangeable idiom (AABB) is mainly used to serve as predicate, while the other form (BBAA) is used to serve as subject or object. This means that Chinese FCIs package information based on different contexts.

Langacker (2008) posited that most of the expressions we use are symbolically complex, and to some extent, analysable into smaller symbolic elements. "Grammar

consists of the patterns for constructing such expressions. Accordingly, the expressions and the patterns are referred to as constructions" (Langacker, 2008, p. 161-167). Constructions can be regarded as symbolic assemblies. A symbolic structure (Σ) consists of the pairing of a semantic structure and a phonological structure. Symbolic structures can combine with one another to form a more elaborate symbolic structure: [Σ1] + [Σ2] = [Σ3]. The three symbolic structures can constitute a symbolic assembly. Note that [Σ3] itself can also be combined with another symbolic structure to form a more elaborate symbolic structure: [Σ3] + [Σ4] = [Σ5]. In this manner, words, phrases, clauses, and sentences can be progressively formed. At a particular level of organization, the component structures [Σ1] and [Σ2] are grouped together to form the composite structure [Σ3]. For example, components [Σ1] (jar) and [Σ2] (lid) can be combined to form the expression [Σ3] (jar lid). At a "higher" level, [Σ3] (jar lid) can act as a component structure to integrate with another component structure [Σ4] (factory) to form the expression [Σ5] (jar lid factory).

Along the same line of thought, an FCI can be viewed as a composite structure which can be further divided into smaller symbolic elements. For example, the four components [Σ1] 千 (*qiān*, a thousand), [Σ2] 山 (*shān*, mountain) [Σ3] 万 (*wàn*, ten thousand), and [Σ4] 水 (*shuǐ*, river) can be grouped into an FCI. Firstly, [Σ1] and [Σ2] are grouped into [Σ5] 千山 (*qiān-shān*, a thousand mountain), while [Σ3] and [Σ4] are integrated into [Σ6] 万水 (*wàn-shuǐ*, ten thousand river). Secondly, [Σ5] and [Σ6] make up the composite structure [Σ7] 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand mountain ten thousand river). Based on Langacker's idea, I define a character (component structure) in an FCI as a component, while two characters (composite structure) in an FCI as a unit, and four characters as a construction.

In fact, there are different combinatory patterns for the seven types of 2+2 FCIs (see Section 2.3). Therefore, this research aims to look at how each 2+2 FCI is organized in particular terms of the ways the characters of an FCI are grammatically and semantically arranged. The different combinatory patterns and their subtypes can be summarized by table 3.

31

Types	Subtypes	Examples
Type 1 [NP NP]	[[ADJ N] [ADJ N]]	轻嘴薄舌 (*qīng-zuǐ-bó-shé*, "hasty and rude")
	[[N N] [N N]]	驴年马月 (*lǘ-nián-mǎ-yuè*, "impossible date")
Type 2 [AP AP]	[[N ADJ] [N ADJ]]	冰清玉洁 (*bīng-qīng-yù-jié*, "spotless")
	[[ADJ ADJ] [ADJ ADJ]]	孤苦伶仃 (*gū-kǔ-líng-dīng*, "solitary and impoverished")
Type 3 [VP VP]	[[ADV V] [ADV V]]	未艾方兴 (*wèi-ài-fāng-xīng*, "rapidly expanding")
	[[V V] [V V]]	吹吹打打 (*chuī-chuī-dǎ-dǎ*, "to advertise something or to arouse people's attention")
Type 4 FCI [[N V] [N V]]	[[N V] [N V]]	燕舞莺歌 (*yīng-gē-yàn-wǔ*, "prosperity abounds")
Type 5 FCI [[V N] [V N]]	[[V N] [V N]]	捕风捉影 (*bǔ-fēng-zhuō-yǐng*, "groundless accusations")
Type 6 FCI [VP NP]	[VP NP]	不咎既往 (*bù-jiù-jì-wǎng*, "overlook someone's past mistakes")
Type 7 FCI [NP + VP]	[NP + VP]	昙花一现 (*tán-huā-yī-xiàn*, "flash in the pan")

Table 3 Seven different combinatory patterns and their subtypes

Firstly, two combinatory subtypes have been identified as the Type 1 FCI ([NP NP]): (i) [ADJ N] [ADJ N], and (ii) [N N] [N N]. An example of the first subtype is [∑7] (轻嘴薄舌 *qīng-zuǐ-bó-shé*, "hasty and rude"), where [∑5] (轻嘴, *qīng-zuǐ*, light- mouth) and [∑6] (薄舌, *bó-shé*, thin-tongue) can be respectively analyzed into: [[∑1] [ADJ (轻, *qīng*, light)] + [∑2] [N (嘴, *zuǐ*, mouth)]] and [[∑3] [ADJ (薄, *bó*, thin)] + [∑4] [N (舌, *shé*,

tongue)]]. The second subtype is the [N N] [N N] construction. This can be shown in the idiom 驴年马月 (*lǘ-nián-mǎ-yuè*, "impossible date"), where both units can also be further analyzed into — [N (驴, *lǘ*, mule)] + [N (年, *nián*, year)] and [N (马, *mǎ*, horse)] + [N (月, *yuè*, month)].

Secondly, two subtypes can also be identified in the Type 2 FCI ([AP AP]): (i) [N ADJ] [N ADJ], and (ii) [ADJ ADJ] [ADJ ADJ]. An instance of the first sub-type is [∑7] 冰清玉洁 (*bīng-qīng-yù-jié*, "spotless"), where [∑5] (冰清, *bīng-qīng*, clear as ice) and [∑6] (玉洁, *yù-jié*, clean as jade) can be further analyzed into: [[∑1] [N (冰, *bīng*, ice)] + [∑2] ADJ (清, *qīng*, clear)] + [[∑3] [N (玉, *yù*, jade)] + [∑4] ADJ (洁, *jié*, clean)]. The second subtype can be illustrated as 孤苦伶仃 (*gū-kǔ-líng-dīng*, "solitary and impoverished"), where (孤苦, *gū-kǔ*, lonely and painful) and (伶仃, *líng-dīng*, "alone and helpless") can be further analyzed into: [[ADJ (孤, *gū*, lonely)] + [ADJ (苦, *kǔ*, painful)]] + [[ADJ (伶, *líng*, alone)] + [ADJ (仃, *dīng*, helpless)]]. It can be found that the basic components of the Type 2 FCI are also nouns and adjectives, and they can be grouped into a unit which can be either [N + ADJ] or [ADJ + ADJ].

Thirdly, The Type 3 FCI ([VP VP]) features two different combinatory patterns: (i) [ADV + V] + [ADV + V], and (ii) [V + V] + [V + V]. The instance of the first subtype is the [ADV + V] + [ADV + V] construction. This can be shown in the idiom 未艾方兴 (*wèi-ài-fāng-xīng*, "rapidly expanding"), where both units can be further analyzed into — [[未 (*wèi*, not yet) + [艾 (*ài*, flourish)]] + [[方 (*fāng*, currently) + [兴 (*xīng*, expand)]]. The second sub-type can be realized as 吹吹打打 (*chuī-chuī-dǎ-dǎ*, "to advertise something or to arouse people's attention"), where [∑5] (吹吹, *chuī-chuī*, blow and blow) and [∑6] (打打, *dǎ-dǎ*, beat and beat) can be further analyzed into: [[∑1] V (吹, *chuī*, blow) + [∑2] V (吹, *chuī*, blow)]] + [[∑3] V (打, *dǎ*, beat) + [∑4] V (打, *dǎ*, beat)]].

The Type 4 FCI ([[N V] [N V]]) has a different combinatory pattern which can be illustrated by the idiom 燕舞莺歌 (*yīng-gē-yàn-wǔ*, "prosperity abounds"), where [∑1] (燕舞, *yàn-wǔ*, swallow dances) and [∑2] (莺歌, *yīng-gē*, warbler sings) can be respectively analyzed into: [[N (燕, *yàn*, swallow)] + [V (舞, *wǔ*, dance)]] + [[N (莺, *yīng*, warbler)] + [V (歌, *gē*, sing)]].

The combinatory pattern in the Type 5 FCI [[V N] [V N]] can be illustrated by the idiom 捕风捉影 (*bǔ-fēng-zhuō-yǐng*, "groundless accusations"), where (捕风, *bǔ-fēng*, chasing the wind) and (捉影, *zhuō-yǐng*, clutching at shadows) can be respectively broken down into: [[V (捕, *bǔ*, chase)] + [N (风, *fēng*, wind)]] and [[V (捉, *zhuō*, clutch) + [N (影, *yǐng*, shadow)]].

The combinatory pattern in the Type 6 FCI ([VP NP]) can be showed in the idiom 不咎既往 (*bù-jiù-jì-wǎng*, "overlook someone's past mistakes"), where 不咎 (*bù-jiù*, "not censure") is a VP and 既往 (*jì-wǎng*, "past misdeeds") is an NP. Lastly, the combinatory pattern in the Type 7 FCI ([NP + VP]) can be shown in the idiom 昙花一现 (*tán-huā-yī-xiàn*, "flash in the pan"), where 昙花 (*tán-huā*, "the night blooming cactus") is an NP and [∑2] (一现, *yī-xiàn*, shows once) is a VP.

Note that in Types 1-5, the different orders (AABB and BBAA) are sub-constructional variants because the two units in a construction belong to the same word class (phrasal category), that is, [NP NP], [AP AP], [VP VP], [[N V] [N V]], and [[V N] [V N]]. The different orders does not lead to the change in the word-class of the internal constituents (phrasal category). They can also be called as allostructions (Cappelle, 2006; Nesset & Janda, 2023). Allostructions are patterns of form and meaning in language that arise from contextual usage rather than being predetermined or fixed. These patterns can be observed across various linguistic levels, such as phonology, morphology, syntax, and discourse. They usually represent a shift in thinking about the nature of language and cognition, emphasizing the importance of usage-based approaches to language learning and processing, as well as the role of context and interaction in shaping the structure and meaning of language. However, in Types 6-7, the different orders (AABB and BBAA) yield separate constructions. For example. 黑白混淆 (*hēi-bái-hùn-xiáo*, black-white-mix-confuse, "to garble things up") is the combination of NP (黑白, *hēi-bái*, black-white) and VP (混淆, *hùn-xiáo*, mix-confuse), while 混淆黑白 (*hùn-xiáo-hēi-bái*, mix-confuse-black-white, "to garble things up") is one of VP (混淆, *hùn-xiáo*, mix-confuse) and NP (黑白, *hēi-bái*, black-white).

The concepts of attraction, substitution, and differentiation (De Smet et al., 2018)

are fundamental in the field of diachronic construction grammar, which aims to elucidate how constructions change over time and contribute to language evolution. These concepts can be applied to the study of Chinese FCIs to explain how different orders (AABB and BBAA) evolve and interact with each other. Attraction refers to how competing forms become more similar in function, while substitution involves a single form occupying the functional domain at the expense of all others. Differentiation, on the other hand, refers to a reduction in the functional overlap between two forms. These concepts can also be operationalized within RCG, which emphasizes the relationship between constructions and their pragmatic and discourse functions. RCG treats these concepts as dynamic processes that emerge from usage in context rather than fixed or predetermined phenomena.

There are several reasons why I examine Chinese idioms in a construction-based approach. The first reason is that this approach helps to understand how individual characters are grouped into an idiom by breaking it down into its component and understands the relationships between those parts. As discussed earlier, the understanding of 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river) depends on two cognitive processes. The four characters in the idiom are grouped into two units (千 (*qiān*, a thousand) and 山 (*shān*, mountain) vs. 万 (*wàn*, ten thousand) and 水 (*shuǐ*, river)). Then, the two units are integrated as a whole (construction). The character needs to follow certain semantic or syntax rules to be grouped into a unit. In Chinese, it can be found that there are combinations of 千 (*qiān*, a thousand) + 山 (*shān*, mountain) and 万 (*wàn*, ten thousand) + 水 (*shuǐ*, river), but the units 千 (*qiān*, a thousand) + 水 (*shuǐ*, river) and 万 (*wàn*, ten thousand) + 山 (*shān*, mountain) are not acceptable to Chinese speakers as no one use such units. Such units can also not be found in any Corpus or Dictionary.

Secondly, this approach helps readers gain a deeper understanding of the idiom's overall meaning. That is to say, it can help you to better understand the literal and figurative meanings of the idioms and how they can be used in different contexts. For example, the literal meaning of 打退堂鼓 (*dǎ-tuì-táng-gǔ*, beat-return-hall-drum, "to give

up") is "*beat the return drum*". The understanding of the literal meaning must be based on the 1+3 construction whose last three words should be regarded as a holistic unit because it is a noun phrase (退堂鼓, *tuì-táng-gǔ*, return-hall-drum). In actual language use, this idiom should be metaphorically interpreted as "*to give up*". Without viewing the 打 (*dǎ*, beat) and 退堂鼓（*tuì-táng-gǔ*) as two constructions, it would be difficult to understand the literal and figurative meanings of the idiom.

Thirdly, investigating the idiom in a construction-based approach can help to understand its origin and historical significance in Chinese language. For example, you are able to identity which form comes before the other in the interchangeables. Besides, it helps to examine the changes of token frequency in the interchangeables.

In conclusion, Chinese FCIs fit in the definition of Construction Grammar, which can be applied to investigate Chinese FCIs regarding combinatory patterns, collocational preferences, partly schematic construction, productivity, change and variation, and pragmatics.

3.4 A constructional approach to 'propositional act functions'

Just as I discussed previously, CxG can be used to explore pragmatic considerations of Chinese idioms. The notions that I mainly borrow are from Croft's (2001) Radical Construction Grammar (RCG) which proposes a constructional approach to three basic propositional act functions that lay the foundation for the three-way distinction of parts of speech. This approach is able to explain the mental representations and processing of language through the analysis of form-meaning pair constructions, which capture regularities in language use. That is to say, investigating Chinese idioms through RCG can gain insights into how Chinese speakers understand and use them in discourse by analyzing the constructional properties of the idioms. Also, it can gain a more comprehensive understanding of how these idioms are constructed and interpreted by Chinese speakers, which can have implications for language teaching, cross-cultural communication, and cognitive linguistics research.

His hypothesis (1991) is that the semantic classes of OBJECTS, PROPERTIES, and ACTIONS are the TYPOLOGICAL PROTOTYPES of speech acts of referring, attributive, and predicating, respectively. Predication, reference, and modification are pragmatic (communicative) functions, or as Searle describes them, PROPOSITIONAL ACTS (see Searle 1969: 23–4; Croft 1990b; Croft 1991: 109–11). I adopt this framework for two reasons. Firstly, this is a rigorous and consistent framework which leads to universal prototypes of the correspondence between word classes and use, i.e., in terms of reference to an object, predication of an action, and modification by a property. This is especially useful for Chinese, as in this language there are no clear-cut boundaries for word classes. Secondly, both Chinese adjectival modifiers and adverbial modifiers can be classed as serving the act of modification based on this framework (see examples (5) and (6)). Simply put, such classification is useful in better investigating the relationships between functions of the idioms and their internal constituencies in different contexts (see Section 4.5).

The act of REFERENCE identifies a referent and establishes a cognitive file for that referent, thereby allowing for future referring expressions to be coreferential with the first referring expression. For example,

(3) 你 拣　　 个 吉日良辰

　　nǐ jiǎn　　 gè jí-rì-liáng-chén

　 you pick up one lucky-day-fine-time

　　 "You pick up a good day."

The Story of the West Chamber (Yuan Dynasty, 1,271-1,368 A.D.)

In (3), the idiom 吉日良辰 (*jí-rì-liáng-chén*, lucky-day-fine-time, "a lucky day") is used as an object which is subcategorised by the verb 拣 (jiǎn, pick up) and which serves the function of REFERENCE.

The act of PREDICATION ascribes something to the referent. Predication does not establish a cognitive file for the state of affairs that is predicated, but instead

prototypically reports relatively transitory states of affairs, often in a narrative sequence. For example,

(4) 蔡廷栋　　　力竭声嘶,　　　　　　　知　斗　不　过。

　　càitíngdòng lì-jié-shēng-sī,　　　　zhī　dòu bù guò

　　cai tingdong energy-out-voice-hoarse, know fight not win

　　"Mr. Cai runs out of energy, knowing that he is not able to win the fight."

<div align="right">The Romance of the Western Empress Dowager YanShi (Minguo, 1,912-1,349 A.D.)</div>

The idiom 力竭声嘶 (*lì-jié-shēng-sī*, energy-out-voice-hoarse, "running out of energy") is used as the predicate in (4) as it makes clear what the subject 蔡廷栋 (*cài-tíng-dòng*, a person's name) does. When idioms act as the predicate, they fulfil the function of PREDICATION.

The act of MODIFICATION (of referents) functions to enrich a referent's identity by an additional feature of the referent, denoted by the modifier. For example,

(5) 侯、童　二　位　　　是 光明磊落　　　　　　　　　的 侠士。

　　hóu, tóng èr　wèi　　shì guāng-míng-lěi-luò　　　　　de xiá-shì

　　hou, tong two person   be upright/open righteous and open-hearted   PART Knight.

　　"Mr. Hou and Mr. Tong are two frank and straightforward heroes."

<div align="right">Yongzheng swordsman diagram (Minguo, 1,912-1,349 A.D.)</div>

In (5), the idiom 光明磊落 (*guāng-míng-lěi-luò*, "frank and straightforward") is used as the attributive adjective function to modify the noun 侠士 (*xiáshì*, "knight"). This idiom 光明磊落 (*guāng-míng-lěi-luò*, "frank and straightforward") adds an additional feature to the referent 侠士 (*xiáshì*, "knight"). Thus, it can be viewed as the act of MODIFICATION. Similarly,

(6) 小心谨慎　　　　　　　　地　　侍奉　皇帝，

xiǎo-xīn-jǐn-shèn　　　　　　de　　shìfèng huáng-dì,

small-heart-careful-watchful PART serves emperor,

来　讨他 的　　欢心 。

lái　　tǎo tā de　　huān-xīn

come get his PART pleasant heart

<Someone> serves the emperor carefully in order to win his heart.

Ancient and Modern Love Sea (Minguo, 1,912-1,349 A.D.)

In (6), the idiom 小心谨慎 (*xiǎo-xīn-jǐn-shèn*, small-heart-careful-watchful, "cautious and timid") is used as the adverbial that modifies the verb 侍奉 (*shì-fèng*, "serve"). In this case, the idiom can also be seen as the act of MODIFICATION. Note that although the detailed usages of the idioms in (5) and (6) are slightly different (the former is an adjectival modifier and the latter an adverbial modifier), but they can be categorized into the act of MODIFICATION based on Croft's model. Also, when idioms are used as complements, they can be regarded as serving the modification function.

Note that the relationships between different types of FCIs and propositional act functions can be diverse. For instance, the Type 1 idiom ([NP NP]) can function as a reference in Example (3), but as a modification in Example (16). Similarly, the Type 3 idiom ([VP VP]) can function as a predication in Example (26), but as a modification in Example (27). While Types 1-7 may not be actual cognitive constructs, they play an important role in language acquisition. As Hollmann (2013) argues, super-schemas and sub-schemas enable individuals to rapidly classify and recognize objects and events in the world. However, these schemas may not be consciously recognized by ordinary speakers, as noted by Croft (1998).

To conclude, when Chinese FCIs are used as subjects and objects, they can be categorised as having a referring function, and Chinese FCIs serve as the modification function if they are used as attributives, adverbials, or complements. Similarly, Chinese

FCIs are categorised as predication when they are used as a predicate. This classification is useful for the present study because it facilitates the annotation of idioms in context and provides a method for classifying the pragmatic function of Chinese idioms in different contexts and at different stages of change.

3.5 Research Questions

1. What are the functional and formal differences between interchangeable and non-interchangeable four-character idioms (FCIs) and how do they change over time from Tang Dynasty (618-907 A.D.) to the Minguo Period (1,912-1,349 A.D.)?

2. What are the functional differences between the two forms of interchangeables (AABB and BBAA) and how do they change over time in terms of token frequency?

3. How is the [*bù* A *bù* B] construction different from regular 2+2 constructions?

4. How does internal constituency of four-character idioms (FCIs) affect constructional meaning and semantic prosody in Chinese?

Hypothesis 1: Symmetricity and iconicity limits the interchangeablity of Chinese FCIs.
Hypothesis 2: Constructions with reference or predication functions may over time develop more modification uses.
Hypothesis 3: The constructional meaning of [bù A bù B] is based on the relationship between lexical items A and B.

3.6 Summary

Construction Grammar (CxG), Radical Construction Grammar (RCG), and Cognitive Grammar (CG) are the three primary usage-based approaches that are adopted in this study to investigate FCIs in terms of their combinatory patterns, collocational

preferences, schematicity, productivity, change, and usage. The concepts (attraction, substitution and differentiation) can be operationalised within RCG, and RCG facilitates the annotation of idioms in context.

Chapter 4 Data and methodology

4.1 Outline

In this chapter I will focus on the corpora consulted in the thesis, the idioms selection criterion, normalization criterion and also the rules about how I annotate the idioms. That is, I will explain in more detail how the data were extracted and coded, and how decisions were taken to classify AABB and BBAA as sub-constructional variants of a single more abstract construction or not.

4.2 Corpora consulted in the thesis

The corpus data are retrieved from three sources: (1) the BLCU (Beijing Language and Culture University) Corpus Centre (BCC)[1], (2) Centre for Chinese Linguistics, Peking University (CCL)[2], (3) zhTenTen[3] and (4) Xinhua Dictionary of Idioms.

With a total of 15 billion characters, BCC comprises various categories such as (a) newspapers and periodicals (2 billion characters), (b) literary classics (3 billion characters), (c) microblogs (3 billion characters), (d) scientific/technological publications (3 billion characters), (e) ancient Chinese text (2 billion characters), and (f) uncategorized sources (1 billion characters). BCC contains a comparatively more comprehensive data categorization of Modern Mandarin, for example, newspapers/periodicals, literary classics, microblogs, and scientific/technological publications (Yu et al., 2002).

The Centre for Chinese Linguistics, Peking University (CCL) is a large corpus that reflects actual language uses in different historical periods of Mandarin Chinese, with a total of 1,569,802,369 characters that features 1,170,199,473 Modern Mandarin characters. CCL also includes different categories listed as follows: (a) the oral Chinese

---

[1] http://bcc.blcu.edu.cn/ Last accessed: 21/7/2022
[2] http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai Last accessed: 21/7/2022
[3] https://www.sketchengine.eu/zhtenten-chinese-corpus Last accessed: 21/7/2022

counts 3,081,723 characters, (b) history and biography novel 8,799,888, (c) Chinese composition 48,286,885, (d) newspapers 839,973,730, (e) literary classics 85,241,162, (f) TV and films 21,359,547, (g) crosstalk[4] 3,480,086, (h) Internet Chinese 54,690,142 and (i) Translated literature 90,046,147.

Both BCC and CCL thus make possible a comparison between different genres in terms of register and language use. One advantage of CCL is that it comes with a spoken section that is absent in the BCC (Zhan et al., 2019). This allows for a comparison between written and spoken forms of modern Mandarin Chinese. Another advantage of CCL is that it provides the word count (characters) in each Chinese historical period (Zhan, Guo, & Chen, 2003) — which eases the normalization of the data (see section 3.5 for more details about the normalization criteria).

The zhTenTen has also been used in the thesis. It is a Chinese corpus made up of texts collected from the Internet. This corpus comprises more than 15.9 billion words. I use this corpus to ensure whether there are more [不 (*bù, not*) A 不 (*bù, not*) B] idioms that should be regarded as interchangeables (see detail in 7.2)

Finally, the Xinhua Dictionary of Idioms (新华成语字典: *xīn-huá-chéng-yǔ-zì-diǎn*; henceforth, XDI), an authoritative and well-established source of Chinese idioms, is also consulted due to its wealth of more than 8,000 idiom entries, each supplemented with references of and examples from Chinese literary classics. Besides, XDI also provides examples based on pragmatic considerations offering Chinese native speakers a comprehensive reading of an idiom in proper contexts. Note that the target language of the research is Mandarin Chinese, as speakers of other forms or dialects of the Chinese language may approach the language in a subtly different way (e.g., Taiwanese Southern Min, Taiwanese Chinese, Cantonese, etc.).

---

[4] Crosstalk is typically performed as a dialogue between two performers.

4.3 Idioms selection criteria

The Xinhua not only has the idioms arranged in a concise way, but it also indicates whether the 2+2 sequence can be syntactically reversed. Therefore, I have resorted to XDI as the main benchmarking source for interchangeable idioms rather than establishing a new criterion to accommodate such idioms. For instance, the idiom 扶危济困 (fú-wēi-jì-kùn, helping those in distress) is the combination of the first unit 扶危 (fú-wēi, help those who were in difficulty) and the second unit 济困 (jì-kùn, rescue the desperately poor). The dictionary specifically points out that the idiom 扶危济困 (fú-wēi-jì-kùn, helping those in distress) can also be annotated as 济困扶危 (jì-kùn-fú-wēi, helping those in distress), where 济困 (jì-kùn, rescue the desperately poor) is the first unit and 扶危 (fú-wēi, help those who were in difficulty) the second. These two idioms are referred to as an "interchangeable type". On the other hand, the idioms which only operate on one form are labelled as "non-interchangeable". For example, 量体裁衣 (*liàng-tǐ-cái-yī*, act according to actual circumstances), where the first unit 量体 (*liàng-tǐ*, measure the body) and the second unit 裁衣 (*cái-yī*, tailor the suit) both fall into the [Verb + Noun] construction, but the two units cannot be inter-switched.

In Chinese FCIs, the 2+2 construction is a type. There are two sub-types (interchangeables and non-interchangeables) that fall under the 2+2 construction. In order to make a comparison between types of interchangeables and non-interchangeables, I sorted out two groups of idioms — 428 interchangeables and 428 non-interchangeables, respectively. To be specific, each group comprises 428 subtypes (see below for the steps for idioms selection). There are three steps to retrieve interchangeables and non-interchangeables (856 types) and their occurrences (17, 966 tokens) for this thesis.

The first step for data retrieval was sifting through the XDI manually for all the candidates of interchangeable idioms. This step yielded a total of 428 interchangeables (types).

The second step was to select 428 non-interchangeables in order to make a comparison with the 428 interchangeables that have been found from the dictionary.

Then I retrieved 4,474 types of both interchangeable and non-interchangeable idioms from the BCC Online Dictionary of Chinese Idioms as the BBC Online dictionary does not distinguish interchangeables from non-interchangeables. Note that BCC is a tagged corpus, which allows me to retrieve all the idioms. I then looked through all the 4,474 idioms (types), and removed the interchangeables and those that do not enter the 2+2 construction, yielding a total of 3,834 non-interchangeables (types), of which I randomly selected 428 types. Finally, a list of 428 non-interchangeable types has been randomly retrieved from the BCC. These two steps led to two groups of idioms for a balanced comparison: 428 interchangeables (types) and 428 non-interchangeables (types).

The third step was the search for the occurrences of the 428 interchangeables and 428 non-interchangeables in the CCL for the ensuing usage-based analysis. Between the Tang Dynasty and Minguo period, the 428 interchangeables (types) yielded 8,381 occurrences (raw tokens), and the 428 non-interchangeables (types) counted 9,585 occurrences (raw tokens). The annotation of those occurrences is discussed in section 3.4. These raw data laid the foundation for discussing the differences between the interchangeables and non-interchangeables in Chapter 4.

To summarise, Chinese FCIs can be analysed with reference to the type frequency vs token frequency distinction. Two balanced groups of idioms were sorted out (i.e., 428 interchangeables and 428 non-interchangeables), and each group generated specific occurrences. The type and token frequencies of interchangeables and non-interchangeables can be summarised in Table 4 below.

Type	Chinese 2+2 FCIs	
Subtypes	428 interchangeables	428 non-interchangeables
Occurrences from diachronic section in CCL (Token frequency)	8,381 occurrences	9,585 occurrences

Table 4 The type and token frequencies of interchangeables and non-interchangeables

Note that in all 856 interchangeables and non-interchangeables, if the two units in a construction belong to the same word class (phrasal category), the different orders (AABB and BBAA) are sub-constructional variants. If inter-switching the two units of an idiom leads to the change of the constituents' word class, the different orders (AABB and BBAA) are different constructions. Therefore, Types 1-5 are sub-constructional variants ([NP NP], [AP AP], [VP VP], [[N V] [N V]], and [[V N] [V N]]), while Types 6-7 are new constructions ([NP VP] and [VP NP]). There are two factors constraining the usage of AABB and BBAA patterns. The first factor is related to the rhyme and rhythm (Chang & Owen, 2010). Classical Chinese poetry typically adheres to specific tonal patterns and rhythmic structures. For example, in order to fit the tonal pattern and maintain the desired rhyme scheme, the order of 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river, "a long journey during which numerous obstacles are encountered") may be adjusted to 万水千山 (*wàn-shuǐ-qiān-shān*, ten thousand-river- a thousand-mountain, "a long journey during which numerous obstacles are encountered") as the former ends in the third tone, while the latter ends in the first tone. The second factor is related to iconicity (see details in section 6.4.2). Also, a more detailed discussion on the three methods on distinguishing AABB and BBAA in interchangeables can be found in Chapter 6.

Section 4.4 will discuss how these occurrences are annotated, and Section 4.5 will elaborate on how the data were normalized in terms of the data size of each dynasty.

4.4 Inclusion criteria for the idioms

A construction may specify not only syntactic, but also lexical, semantic, and pragmatic information (Fillmore, Kay, and O'Connor, 1998: 501). Therefore, I took into account the internal constituency of a Chinese FCI, but also other factors concerning their usage. In order to look at how different types of FCIs are used in different contextual environments, I manually annotated the occurrences one by one (8,381 for interchangeables and 9,585 for non-interchangeables). It was therefore

necessary to establish the inclusion criteria for my data. Here are five examples to illustrate the criteria for the analysis of the idioms that I have annotated. (7) is an occurrence that should be included in my analysis; (8) and (9) are two tricky examples whose rules need to be specified as to how I annotate them; examples (10) and (11) are cases that should be excluded.

(7)　两　只　黑白混淆　　　　　的　　眼, 只管　溜来　　溜去
　　　liǎng zhī hēi-bái-hùn-xiáo　　de　　yǎn, zhǐ-guǎn liū lái　　liū qù
　　　two　CL black-white-mix-mistake PART eyes, just　　look here look there
　　　"The two confusing eyes look just here and there"

The Sequel of Liudong's story (Minguo, 1,912-1,349 A.D.)

In (7), 黑白混淆 (*hēi-bái-hùn-xiáo*: black-white-mix-confuse, "to garble things up like mixing black and white colours together") serves as an attributive adjective to modify the NP 眼 (*yǎn*, eyes).

(8) 陈妇　虽　　是 扬花水性　　　　　却　知道　玉成　的　死。
　　chénfù　suī　shì yang-huā-shuǐ-xìng, què zhī-dào yùchéng de　sǐ
　　chen.Mrs. although be flirtatious　　　　but know　yùchéng PART death
　　"Although Mrs. Chen is fickle (cares nothing about others), she knows something of YuChen's death."

A Romance of the Three Hundred Years of Yan History (Minguo, 1,912-1,349 A.D.)

In English, the adjective after a copula is usually called a predicative (Burton-Roberts, 2016), but there is no such function in Chinese syntax. However, this can still be labelled as predicate (谓语, *wèi-yǔ*) as the English predicative functions in similar way as a predicate in Chinese (Ren, 2016). In (8), 扬花水性 (*yáng-huā-shuǐ-xìng*, raise-flower-water-character, "fickle") is placed after the copula 是 (shì) which also functions

as an emphatic marker in Chinese (Wiedenhhof, 2015). It places a distinctive focus on Mrs. Chen's fickle personality. Thus, I categorized the idiom 扬花水性 (*yáng-huā-shuǐ-xìng*) as a predicate based on its syntactic position, as the copula 是 (shì) can be omitted without affecting the meaning of the sentence. In Chinese, if the copula 是 (*shì)* cannot be omitted, the idiom following the copula should be regarded as an object. Cases as such were only 24 out of 8,381 retrieved from CCL.

(9) 你　本　　　是 冰清玉洁　　　的　女子

　　nǐ　běn　　　shì bīng-qīng-yù-jié　de　nǚzǐ,

　　you originally be spotless　　　PART woman

　　"You were a spotless woman."

<div align="right">Ancient and Modern Love Sea (Minguo, 1,912-1,349 A.D.)</div>

In (9), the phrase 冰清玉洁的女子 (*bīng-qīng-yù-jié-de-nǔ-zǐ*, ice-clear-jade-clean-de-woman-person, "a spotless woman") occurs as an NP after the copula (是, *shì*), and the NP is treated as a predicate (as in the second example). However, 冰清玉洁 (*bīng-qīng-yù-jié*, ice-clear-jade-clean, "spotless") should arguably not be characterized as a predicate. In fact, it is the NP here that functions as the predicate, rather than the idiom 冰清玉洁 (*bīng-qīng-yù-jié*) itself. The NP is composed of the idiom 冰清玉洁 (*bīng-qīng-yù-jié*) and the head noun 女子 (*nǚzǐ*, woman), with the former adjectivally modifies the latter. Therefore, I categorized 冰清玉洁 (*bīng-qīng-yù-jié*) as an attributive in this case.

The (10) is one that I did not include in the analysis of FCIs because it comes in the form of a dictionary entry (direct quotation), and it does not occur in the form of a naturalistic utterance.

(10) 耳提面命：　　形容　　当面　　倾听　殷切　恳诚　　的 教诲　和 希望。

ěr-tí-miàn-mìng: xíngróng dāngmiàn qīngtīng yīnqiè kěnchéng de jiàohuì hé xīwàng

"The idiom 耳提面命 (ěr-tí-miàn-mìng) is used to describe a situation where people hope they are being listened to patiently."

This is a sentence explaining what the idiom means.

<div align="right">Guofan Zeng's Family Letter (Qing, 1,644-1,912 A.D.)</div>

(11) 此　去　好　凭　三寸　　舌,　　再来　　不　值　　一文 钱。

　cǐ　qù hǎo　píng sān-cùn　shé,　　zài-lái　bù　zhí　　yī-wén-qiān

This go good use three-inch tongue, next time not worth a penny.

"This time it is worth talking with them, but next time it isn't worth a penny."

<div align="right">Clear Words to Illustrate the World (Ming, 1,368-1,644 A.D.)</div>

In example (11), 不值一文 (*bù-zhí-yī-wén*, not-worth-a-penny) is not treated as an idiom in use, but one that is 'rebracketed'. The first two characters 不值 (*bù-zhí*, not worth) serve as a verb phrase, and the last two characters 一文 (*yī-wén*, a small amount of) is used as a measure word to modify the noun 钱 (*qián*, money). The sentence is therefore analysed as "不值 (*bù-zhí*, not-worth)" + "一文钱 (*yī-wén-qián*, a-penny)", instead of a whole chunk "不值一文 (*bù-zhí-yī-wén*, 'not worth a penny')" + "钱 (*qián*, money)". The reason for which I excluded this example is that it was so retrieved as the corpus was not capable of differentiating between "mention" and "use" because instances of the two cannot technically be clearly distinguished from each other in the corpus (Sperber & Wilson, 1981, 1986; Wilson, 1999).

Based on the notion of propositional act functions proposed by Croft (2001), Chinese 2+2 FCIs serve the reference function when they are used as subjects and objects, and if they are used as attributives, adverbials, or complements they serve as the modification function. Lastly, when FCIs are used as a predicate, they are categorised as predication.

4.5 Normalization criteria

One of the differences between CCL and BCC is that the former provides the word count (characters) in each Chinese historical period (Zhan, Guo, & Chen, 2003) — which eases the normalization of the data, while BCC shows only the total number of characters from all the periods as a whole. For example, in CCL the Tang dynasty (618-907 A.D.) features 9,002,907 characters, the Song dynasty (960-1,279 A.D.) 34,816,689, the Yuan dynasty (1,271-1,368 A.D.) 961,884, the Ming dynasty (1,368-1,644 A.D.) 21,038,301, the Qing (1,644-1,912 A.D.) dynasty 48,109,077, and the Minguo period (1,912-1,349 A.D.) 35,371,339. The normalisation of my dataset has been achieved after searching for the 428 interchangeables in CCL, which yielded 8,381 context-based instances (tokens), (Section 3.4 discussed the annotation rules for these examples). The number of instances in the Tang dynasty (618-907 A.D.) is 120, Song 553, Yuan 80, Ming 1,356, Qing 3,346, and Minguo 2,946. As these are row frequencies, we cannot simply conclude that the frequency of interchangeables in the Tang dynasty (120) is higher than that of the Yuan dynasty (80), due to a huge gap in data size between these two periods (Tang: 9,002,907; Yuan: 961,884). Therefore, I normalized the instances in each dynasty in order to achieve a fair balance in word count based on the total of characters in each dynasty. As shown in Table 5, the number of characters in the Qing (1,644-1,912 A.D.) is 50 times greater than that in the Yuan dynasty (1,271-1,368 A.D.). Thus, I used the median corpus size across the dynasties to normalize the data and prevent any distortion due to outliers or skewed distribution (Leclerc, 1993; McMorris, 2000), as the median is usually the preferred measure of central tendency when the distribution is not symmetrical (Mulder, 1978; Schrijver, 1979). Note that the normalisation (see below) was a post-hoc normalisation process, which took place after the manual annotation of the occurrences (Tantucci 2020; Culpeper & Tantucci 2021).

Given the six dynasties, the median should be the average of the middle two dynasties Ming (21,038,301 characters) and Song (34,816,689 characters), which is 27,927,495 (characters). Based on the median corpus size, I then normalised the number

of instances that I retrieved from each dynasty. For example, the median (27,927,495 characters) is 1.33 times greater than the word count of the Ming dynasty (1,368-1,644 A.D.), so the normalized instances of the Ming dynasty (1,368-1,644 A.D.) should accordingly increase to 1,800 (1.33 times of the original). In other words, I randomly selected 444 out of the original 1,356 instances of the Ming dynasty (1,368-1,644 A.D.) to suffice for a balanced normalization. This principle was then applied to data totals of other dynasties. Note that the Yuan dynasty (1,271-1,368 A.D.) may be problematic with a relatively small word count (961,881; 1/29 of the median), which entailed a high number of "virtual instances" for data normalization. In order to balance the CCL, such a methodological decision unavoidably presents some caveats. However, it is arguably a better solution rather than 'out casting' a whole Dynasty from the diachronic analysis. The contrast between original and normalized data for interchangeables can be found in Table 5 below.

Dynasty	Original Data		Normalized Data	
	Word count	Instances	Word count	Instances
Tang (618-907 A.D.)	9,002,907	120	27,927,495	372
Song (960-1,279 A.D.)	34,816,689	533	27,927,495	428
Yuan (1,271-1,368 A.D.)	961,884	80	27,927,495	2,323
Ming (1,368-1,644 A.D.)	21,038,301	1,356	27,927,495	1,800
Qing (1,644-1,912 A.D.)	48,109,077	3,346	27,927,495	1,942
Minguo (1,912-1,349 A.D.)	35,371,339	2,946	27,927,495	2,326
In total	130,400,197	8,381	167,564,970	9,192

Table 5 Original and normalized data for interchangeables (from Tang to Minguo) in CCL

The table shows the occurrences of interchangeable idioms in different Chinese historical periods from the Tang dynasty (618-907 A.D.) to the Minguo period (1,912-

1,349 A.D.). Before data normalization, token frequency of interchangeables rises from 120 (Tang) to 533 (Song), and rapidly drops to 80 (Yuan), later rapidly surging to 1,356 (Ming) and 3,346 (Qing). It might not be easy to explain from a diachronic perspective why the Yuan dynasty (1,271-1,368 A.D.) undergoes a sharp decrease, or why the Ming dynasty (1,368-1,644 A.D.) undergoes an abrupt upswing. However, with normalized data, a more reasonable tendency is made possible. Namely, that the Tang dynasty (618-907 A.D.) and the Song dynasty (960-1,279 A.D.) have a small number of context-based instances that grows from 372 to 428. This may imply that the token frequency of interchangeable idioms has seen a considerable growth ever since the Song dynasty (960-1,279 A.D.), reaching its peak in the Yuan dynasty (1,271-1,368 A.D.). The number of the interchangeables has escalated from 1,800 (Ming), to 1,942 (Qing), and to 2,236 (Minguo). Compared with their earlier dynasties, the Yuan dynasty (1,271-1,368 A.D.), the Ming dynasty (1,368-1,644 A.D.), the Qing dynasty (1,644-1,912 A.D.), and the Minguo period (1,912-1,349 A.D.) are found with significantly higher frequencies of such idioms.

This same method also applies to non-interchangeables. The contrast between original and normalized data for non-interchangeables can be found in Table 6 below, which shows the occurrences of non-interchangeable idioms in different Chinese historical periods from the Tang dynasty (618-907 A.D.) to the Minguo period (1,912-1,349 A.D.).

	Original Data		Normalized Data	
Dynasty	Word count	Instances	Word count	Instances
Tang (618-907 A.D.)	9,002,907	80	27,927,495	239
Song (960-1,279 A.D.)	34,816,689	611	27,927,495	488
Yuan (1,271-1,368 A.D.)	961,884	43	27,927,495	1,248
Ming (1,368-1,644 A.D.)	21,038,301	1,361	27,927,495	1,807
Qing (1,644-1,912 A.D.)	48,109,077	3,381	27,927,495	1,958
Minguo (1,912-1,349 A.D.)	35,371,339	4,108	27,927,495	3,241
In total	130,400,197	9,584	167,564,970	8,982

Table 6 Original and normalized data for non-interchangeables (from Tang to Minguo) in CCL

The post-hoc normalisation had two advantages. Firstly, it achieved a fair balance in word count based on the total of characters in each dynasty (for both interchangeables and non-interchangeables). Secondly, it made the two groups of occurrences more balanced as those for interchangeables increased from 8,831 (raw token frequency) to 9,192 (normalized token frequency), while those for non-interchangeables decreased from 9,584 (raw token frequency) to 8,982 (normalized token frequency). The normalized occurrences (9,192 for interchangeables; 8,982 for non-interchangeables) laid the foundation for addressing the token differences between the two (see Chapter 5).

According to Traugott and Dasher (2001), approximate stages of change in the history of Chinese can be classified as follows: Late Middle Chinese (LMC, 600 — 1250), Early Mandarin (EMand, 1250 — 1800), and Modern Mandarin (MdMand, 1800 — present). Please note that these stages are conventions of periodization for the Chinese language. Thus, Late Middle Chinese belongs with the Tang (618-907 A.D.) and the Song (960-1,279 A.D.) dynasties, Early Mandarin the Yuan (1,271-1,368 A.D.), Ming (1,368-1,644 A.D.), Qing (1,644-1,912 A.D.) dynasties, and Modern Mandarin the Minguo period (1,912-1,349 A.D.) as Table 7 shows.

Dynasty	Approximate stages
Tang (618-907 A.D.)	Late Middle Chinese (LMC, 600 — 1250)
Song (960-1,279 A.D.)	
Yuan (1,271-1,368 A.D.)	Early Mandarin (EMand, 1250—1800)
Ming (1,368-1,644 A.D.)	
Qing (1,644-1,912 A.D.)	
Minguo (1,912-1,949 A.D.)	Modern Mandarin (MdMand, 1800—present)

Table 7 Time frames for Chinese idioms

The rationale for categorizing the six dynasties into three primary historical stages was that language change does not necessarily come in discernible features between two different consecutive dynasties. This method helped to combine several independent dynasties as a group (in a wider time frame) in order to discover any significant patterns based on which FCIs changed over time.

Chapter 5 The differences between interchangeable idioms and non-interchangeable idioms

5.1 Outline

In this chapter I will focus on the internal constituency of interchangeable and non-interchangeable idioms, their type frequency, token frequency, and functions. More specifically, section 5.2 discusses the different distributions of interchangeables and non-interchangeables idioms in seven types and argues that symmetricity and iconicity are the two factors that limit the (non-)interchangeability of Chinese FCIs. Section 5.3 makes a comparison between interchangeables and non-interchangeables in terms of token frequency, and section 5.4 discusses the productivity and schematicity of Chinese 2+2 FCIs. At last, section 5.5 investigates how the interchangeables and non-interchangeables change over time in terms of function.

5.2 Distributions of interchangeables and non-interchangeables idioms in seven types

This section discusses the differences between interchangeables and non-interchangeables in terms of type frequency. I investigated all the interchangeables (n = 428) and non-interchangeables (n = 428) that feed into seven different types of internal constituency and counted the numbers for each type. I also accounted for whether they were symmetrical or not, that is, whether the first internal constituent (say a NP) would be of the same class as the second (i.e., one more NP). See Table 8 for the type frequencies of interchangeables and non-interchangeables in each type.

Types	Interchangeables	Non-interchangeables	Symmetricity
Type 1 [NP NP]	138	71	
Type 2 [AP AP]	62	38	
Type 3 [VP VP]	34	75	Symmetrical
Type 4 [[N V] [N V]]	66	15	
Type 5 [[V N] [V N]]	106	91	
Type 6 [NP VP]	11	97	asymmetrical
Type 7 [VP NP]	11	41	
Total	428	428	

Table 8 Type frequencies of interchangeables and non-interchangeables idioms in terms of internal constituency

There are two interesting observations from Table 8. Firstly, interchangeables display a higher type frequency in symmetrical structures (Types 1, 2, 3, 4, 5) when compared with non-interchangeables. On the other hand, a higher type frequency is found with non-interchangeables in asymmetrical structures (Types 6 and 7) when compared with interchangeables.

With this in mind, I hypothesized that structural symmetricity would play a crucial role in determining whether a four-character idiom is interchangeable or not. I treated all the symmetrical idioms as one category (Types 1, 2, 3, 4, 5), and all the asymmetrical idioms as another (Types 6 and 7). Then, I performed a Chi-square (χ^2) test of independence to verify my hypothesis, as χ^2 tests can be used to determine whether there is a significant association between two categorical variables from a single population (Desagulier, 2017). This allows us to test is there some correlation is present between structural symmetricity and (non-)interchangeability. In the model, Pearson residuals indicate the difference between the observed and estimated probabilities and whether such a difference is significant (Menard, 2002). Figure 1 is an association plot (Levshina, 2015), determining which Pearson residuals represent significant deviations from the expected values at a given of statistical significance. The model computes the

observed frequency and the expected frequency with regard to the total number of observations. Each cell of the data is represented by a tile whose height is proportional to the corresponding Pearson residual and whose width is proportional to the square root of the expected counts (Desagulier, 2017).

The plot displays bars that either "grow" or "fall". If a bar "grows" above the baseline (dotted line), the residual is considered positive, i.e., the observed frequency is greater than expected. If it "falls" below the baseline, the residual is then considered negative, i.e., the observed frequency is smaller than expected. If a standardized residual value is greater than 1.96 or less than –1.96, the cell makes a statistically significant contribution to the obtained χ^2-statistic value at the significance level of 0.05 (Desagulier, 2017). The width of the bars indicates frequency; namely, the wider the bar, the higher the frequency.
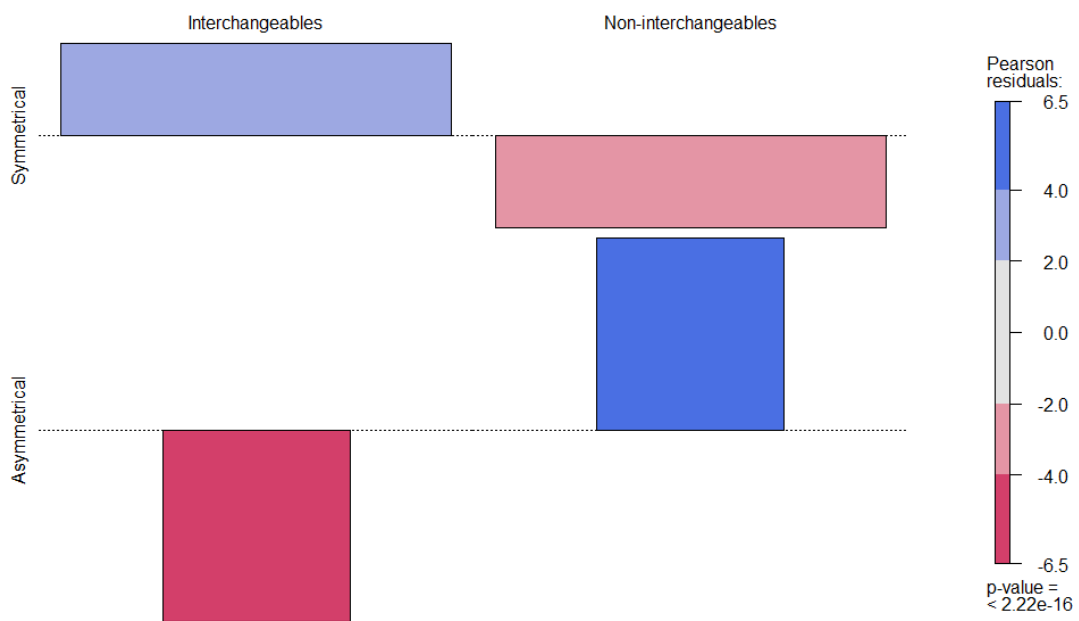

Figure 1 Association plot of residuals: interchangeables and non-interchangeables in symmetrical and asymmetrical

Figure 1 shows the Pearson residuals based on the chi-square differences between observed and predicted frequencies, with a highly significant mismatch (<2.22e–16) between interchangeables (AABB and BBAA) and non-interchangeables (X-squared =

101.66, df = 1, p-value < 2.2e-16). Note that interchangeables show significantly "positive" presence of symmetrical structures, while non-interchangeables indicated relative absence of symmetrical structures. This entails that interchangeables have a significantly stronger correlation with symmetrical structures over asymmetrical ones, compared with non-interchangeables. On the other hand, non-interchangeables have attained 'positive' presence of asymmetrical structures over symmetrical ones, compared with interchangeables. Hence, a conclusion could be drawn that internal constituency (Symmetricity VS Asymmetricity) plays a crucial role in determining the (non-)interchangeability of an idiom.

Secondly, it can be found that the numbers of interchangeables in symmetrical structures (Types 1, 2, 3, 4, 5) are greater than those of non-interchangeables except for Type 3 ([VP VP] construction). More specifically, there are 34 interchangeables and 75 non-interchangeables in Type 3. That is to say, symmetrical structures (Types 1, 2, 3, 4 and 5) have a higher type frequency in interchangeables than non-interchangeables, but Type 3 is an exception. One of the reasons for this lies in iconicity (Simone, 1995). In discussion of the symmetrical structures, the sequential order principle limits the interchangeability of Type 3 FCIs. For example, the Type 3 idiom 过目成诵 (*guò-mù-chéng-sòng*, "can recite after taking a cursory glance") operates on the sequential order principle; it is, therefore, non-interchangeable. This is to say that the first unit 过目 (*guò-mù*, "taking a cursory glance") must come before the second unit 成诵 (*chéng-sòng*, "can recite") due to their sequential order in the real world. Similarly, another example of Type 3 is 闻风而动 (*wén-fēng-ér-dòng*, "act at once on hearing the news"), where the first unit 闻风 (*wén-fēng*, "hearing the news") must precede the second unit 而动 (*ér-dòng*, "act according"). Iconicity applies to most Type 3 non-interchangeables, but such sequential order is not found between the two units in Type 3 interchangeables. For example, 不闻不问 (*bù-wén-bù-wèn*, not to pay much attention to) and 不问不闻 (*bù-wèn-bù-wén*, not to pay much attention to), where either unit — 不闻 (*bù-wén*, not to hear) or 不问 (*bù-wèn*, not to ask about) — can come first as a distinct action. Thus, more non-interchangeables than interchangeables are found for Type 3 (symmetrical

structure), precisely due to the iconic order of the internal constituents.

In conclusion, symmetricity and iconicity are the two factors that limit the (non-)interchangeability of Chinese FCIs (see hypothesis 1).

5.3 A comparison between interchangeables and non-interchangeables in terms of token frequency

After a detailed search for the 428 interchangeables and 428 non-interchangeables in CCL (diachronic section), I found 8,381 occurrences for interchangeables and 9,585 occurrences for non-interchangeables (see Chapter 4). The raw frequencies of interchangeables and non-interchangeables over three different periods of the CCL are shown in Table 9 (raw frequencies are provided for the Chi-square test in section 5.4).

	LMC (600 — 1250)	EMand (1250—1800)	MdMand (1800-present)
AABB	425	2486	1590
BBAA	238	2316	1368
Interchangeables	663	4802	2958
Non-interchangeables	691	4785	4108

Table 9 The raw frequencies of interchangeable and non-interchangeable expressions in three different periods in CCL

Note that we cannot simply make a token frequency comparison between interchangeables and non-interchangeables based on the raw data due to the fact that there is a remarkable gap in data size among Late Middle Chinese, Early Mandarin and Modern Mandarin. For this reason, I normalised the raw data based on each sub-corpus to improve the descriptive statistic interpretation of the dataset (see Data and Methodology in Chapter 4). At this point, it is finally possible to look at how the token frequencies of interchangeables and non-interchangeables change over time. The normalized frequencies of interchangeables and non-interchangeables are illustrated in

Table 10.

Normalized	LMC (600 — 1250)	EMand (1250—1800)	MdMand (1800-present)
AABB	511	3307	1250
BBAA	289	2758	1076
Inters	800	6065	2326
Non-inters	727	5013	3241

Table 10 The normalized frequencies of interchangeables and non-interchangeables expressions in three different periods in CCL

The normalized frequencies of the development of interchangeables and non-interchangeables over three periods is plotted in Figure 3 below.
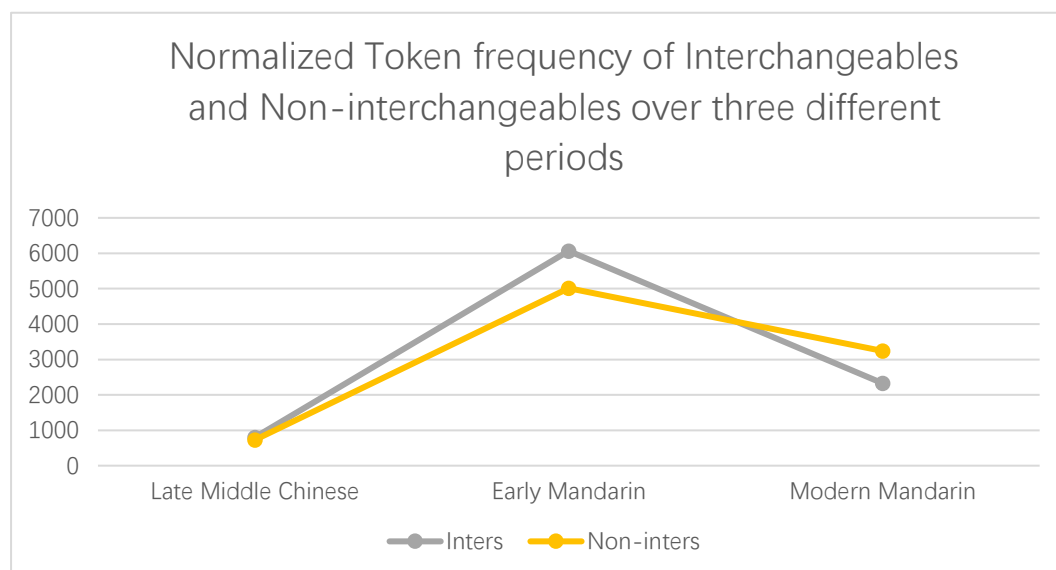
Figure 2 Normalized Token frequency of Interchangeables and Non-interchangeables over three different periods

The first interesting observation that can be made Figure 2 is that the frequency of Chinese FCIs is clearly rising between Late Middle Chinese (600 — 1250) and Early Mandarin (1250 — 1800) and then it drops in the transition from Early Mandarin Chinese to Modern Mandarin (1800 — present). In other words, the frequencies of interchangeables and non-interchangeables have reached the highest point in Early

Mandarin, before appreciably falling in Modern Mandarin. This trend may be explained by two factors.

Firstly, Early Mandarin is a prosperous period for Chinese fiction, which positively influences the use of Chinese FCIs. The rise of a "money economy" and urbanization in the Late Middle Chinese era led to the professionalization of entertainment which was further encouraged by the spread of printing the rise of literacy, and education (Lu, 1976). This form of literary fiction shows its full role in society and literary value, breaking the monopoly of orthodox poetry in literary history (Hsia, 2016). The themes of fiction in Early Mandarin cover wars and violence, chivalry and martial arts, adventures and fantasies, religions and moral values, deities and demons, human relationships and conflicting desires, courtesans and prostitutes, and loyalism and nationalism (Chang, 2010). Such significant development of literary production is likely to have decisively contributed to the rising usage of Chinese FCIs throughout this time span.

Having said that, it is also important to speculate as to why there was then a drop of FCIs during the following period. This may be partly due to the May Fourth Movement, which advocates the use of 白话文 (*bái-huà-wén*, "written vernacular Chinese") instead of 文言文 (*wén-yán-wé*n, "literary Chinese"). In fact, before the year 1911, the official style of the written language in China is 文言文 (*wén-yán-wé*n). It corresponds to the written Chinese language in use from the end of the Han Dynasty (220 CE) to the early 20th century. The written style of 文言文 (*wén-yán-wé*n) remained prescriptively stable over the years, thus increasingly diverging from the natural development of the spoken language (Pulleyblank, 1995). The 文言文 (*wén-yán-wé*n, "literary Chinese") was classically regarded as the socially more 'important' language and by the late imperial period it 'was considerably different from the colloquial (e.g. it applied a monosyllabic lexicon in contrast with the colloquial polysyllabic one), while the vernacular imitated spoken style and thus was closer to colloquial' (cf. Pan and Kádár, 2011: 26).

The May Fourth Movement was a political, cultural and anti-feudal movement that

advocated scientific and democratic social values, affecting the language use. It started out of student protests in Beijing on the fourth of May, 1919. Students and scholars demanded political and cultural reforms inspired by Western-style democracy though such movements in China has often placed a heightened focus on nationalism (Liu, 2020). This led to the birth of a new anti-traditionalist intellectual class that criticised core elements of traditional Chinese culture and the Confucian ideology. Intellectuals from the May Fourth Movement agreed that 文言文 (*wén-yán-wé*n, "literary Chinese") style was a 'dead language' and claimed that literature should now be written in vernacular Chinese.

The famous writer Shi Hu (1891–1962) was the ideological father of this literary revolution and introduced the terminology 白话文 (*bái-huà-wén*, "written vernacular Chinese") to address the new writing style born from The May Fourth Movement (Tantucci & Wang, 2020). Shi Hu gave the guidelines to create a new form of literature aiming to avoid classical allusions, discard stale and outworn literary phrases, and adopt vernacular words and expressions (Wang, 2010). The May Fourth Movement is regarded as the starting point where vernacular Chinese gained currency over and eventually replaced the use of literary Chinese (文言文, *wén-yán-wén*) (Liang & Yang, 2020). Literary Inquisition (文字狱, *wén-zì-yù*, "imprisonment due to writings") and speech crimes took place in the Qing dynasty (1,644-1,912 A.D.). This unavoidably negatively affected the usage of classical figures of speech, of which FCIs constitute a prototypical example. The Qing dynasty (1,644-1,912 A.D.) was particularly notorious for its crackdowns on dissenting voices, that is, oppression and unease brought about by the Manchurian occupation. The Literary Inquisition occurred towards the end of Early Mandarin, may have played a key part in the sharp decrease of Chinese FCIs.

5.4 A comparison between interchangeables and non-interchangeables in terms of comparative frequency

This section discusses the differences between interchangeables and non-interchangeables in terms of comparative frequency. I thus fitted a Chi-square test to look at the differences in distribution between interchangeables ("AABB" and "BBAA") and non-interchangeables over the three periods of Late Middle (LMC, 600 — 1250), Early Mandarin (EMand, 1250—1800) and Modern Mandarin (MdMand, 1800—present).
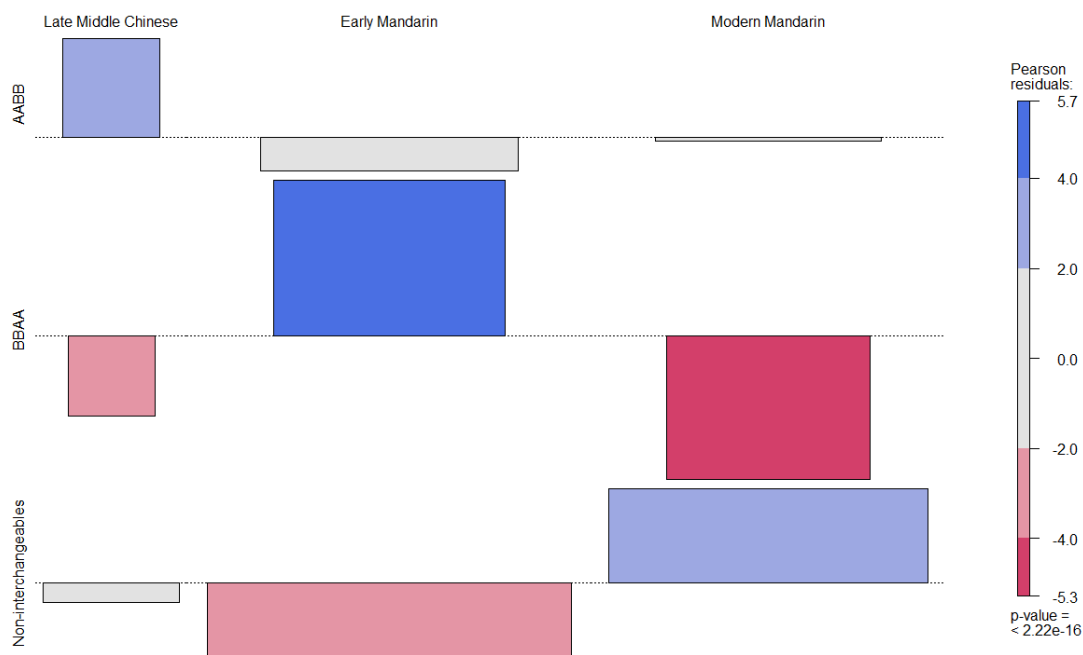
Figure 3 Association plot of residuals: interchangeable idioms (AABB and BBAA) and non-interchangeable idioms in three periods

Figure 3 shows the Pearson residuals based on the chi-square differences between observed and predicted frequencies, with a highly significant mismatch ($<2.22e–16$) between interchangeables (AABB and BBAA) and non-interchangeables (X-squared = 104.11, df = 4, p-value < 2.2e-16).

The first interesting observation about Figure 4 is that AABB shows significantly "positive" residuals in Late Middle Chinese in contrast with BBAA. In other words, AABB features a significantly higher distribution than BBAA in Late Middle Chinese (AABB appeared ahead of BBAA). For example, the AABB idiom 佳人才子 (*cái-zǐ-jiā-rén*, gifted scholar) is first found in Late Middle Chinese, but there is no record of the

BBAA form 才子佳人 (*jiā-rén-cái-zǐ*, gifted scholar) until Early Mandarin. In general, AABB has a higher token frequency than BBAA in the first stage, implying its higher popularity in Late Middle Chinese.

A second important finding is that BBAA in Early Mandarin shows a "positive" trend which is in sharp contrast with its usage in Late Middle Chinese. This means that the token frequency of BBAA increases sharply from Late Middle Chinese to Early Mandarin. There are two reasons for this trend. Firstly, new artistic and literary forms, e.g. 小令散曲 (*xiǎo-lìng-sǎn-qǔ*, verse), 戏剧 (*xì-jù*, drama), and 话本 (*huà-běn*, script), have appeared in Early Mandarin. Note that verse is colloquial and flexible in style (Zhang, 2012), drama is versatile in forms (Yu, 2019), and script for story-telling is loose in format (Li, 2010). The stylistic features in the three forms are relatively more flexible than those in Late Middle Chinese because the new literary forms are based on oral language. This may have sped up the spread of the new BBAA schema. Secondly, when new constructions emerge, they often "spread by gradually increasing their frequency of use over time" (Bybee & McClelland, 2005, p. 387). This means that speakers use the new constructional schema (BBAA) more frequently, accounting for the increase of comparative frequency in BBAA (the second form of the interchangeable idioms). For example, there are 28 instances of BBAA 才子佳人 (*jiā-rén-cái-zǐ*, gifted scholar) in Early Mandarin, while the instances of AABB 佳人才子 (*cái-zǐ-jiā-rén*, gifted scholar) count only 22 (These are normalized data). To sum up, new artistic and literary forms have led to the first use of BBAA, and such new constructions are assumed to have increased token frequency.

The third finding from Figure 4 is that the bar widths of interchangeables (AABB and BBAA) and non-interchangeables in Early Mandarin are wider than those in Late Middle Chinese and Modern Mandarin. This means that both AABB and BBAA are widely used in Early Mandarin and that the frequencies of both interchangeables and non-interchangeables have reached their highest point in Early Mandarin. The reasons for this are arguably fiction development, May Fourth Movement, and the Literary Inquisition, which have already been discussed in Section 5.3. Thus, my argument is that Early Mandarin is a period where AABB and BBAA are competing with one another (see Chapter 6 for an in-depth discussion about constructional competition).

Two very important elements of the development of FCIs during the three periods

are productivity and schematicity. In view of Chinese FCIs, interchangeables do become more productive in Early Mandarin because of the starting usage of the new BBAA schema. In other words, interchangeables can fit either in AABB or BBAA, which means interchangeable idioms increase in productivity and schematicity.

Although it is important to recognize that there is no predictable timeframe for the interaction of productivity and nonproductivity (Traugott and Trousdale, 2013), I nonetheless argue that Early Mandarin (second stage) is a period where Chinese FCIs are the most productive, and where the two functionally similar forms (AABB and BBAA) enter a phase of competition. Productivity may be short-lived while non-productive patterns may persist for a long time (Nørgård-Sørensen, Heltoft, and Schøsler 2011: 38). In other words, interchangeables show a decreased tendency in productivity in Modern Mandarin (third stage). This phenomenon will be further illustrated in Chapter 6, which discusses how the two functionally similar forms are competing over time.

Finally, it is important to note that interchangeables (both AABB and BBAA) show a significantly "negative" trend in Modern Chinese, which is in sharp contrast with non-interchangeables during the same period. This entails that over time non-interchangeables have somehow become more preponderant compared with interchangeables. The reason why non-interchangeables have become more prominent compared with interchangeables is based on the fact that when interchangeables (AABB and BBAA) compete with each other, the token frequency of interchangeables actually decreases over time. More specifically, Type 1 ([NP NP]), Type 2 ([AP AP]), Type 4 ([N V] [N V]), and Type 5 ([[V N] [V N]]) interchangeables undergo attraction, which means either AABB or BBAA in each type mainly serves a similar function. This entails that one of the two forms may suffice for serving the function in different contexts, leading to the decreased token frequency of Type 1, Type 2, Type 4, and Type 5 in interchangeables when compared with non-interchangeables. Type 6 ([NP VP]) and Type 7 ([VP NP]) undergo differentiation whereby each construction finds its own function. This implies that AABB and BBAA in Type 6 and 7 serve a different function

in different contexts, which means that they are not as interchangeable as they used to be. On the contrary, Type 3 ([VP VP]) interchangeables undergo substitution, which means only AABB or BBAA prevails. This also makes the token frequency of interchangeables decrease (see detail in Chapter 6).

5.5 The differences between interchangeable idioms and non-interchangeable idioms in terms of function

Comparisons in internal constituency, token frequency, and comparative frequency have so far been made between interchangeables and non-interchangeables. It is now worth looking at the differences between interchangeables and non-interchangeables in terms of their function.

Using the normalized data from Table 10 in Section 5.3, I fitted a conditional inference tree (Hothorn et al., 2006) to further investigate the mismatch between interchangeables and non-interchangeables. Figure 5 shows the output of a conditional inference tree model (Tagliamonte & Baayen, 2012) which gathers unbiased corpus-driven convergences of FCIs' interchangeability, as well as its functions, and dynasties, as different dimensions concurring (Tantucci and Wang, 2018) to the spontaneous encoding of Chinese FCIs. I look at FCI's propositional act functions (Croft, 2001) as the outcome variable of my analysis. One of the advantages of this method is that it plots statistically significant patterns of formal and functional distribution that intersect hierarchically with one another. It is then possible to locate the most important predictor when we classify propositional functions and to investigate how FCIs change over time in terms of their functions. More specifically, the conditional inference tree allows us to predict the distribution of propositional act functions between interchangeable and non-interchangeable idioms.
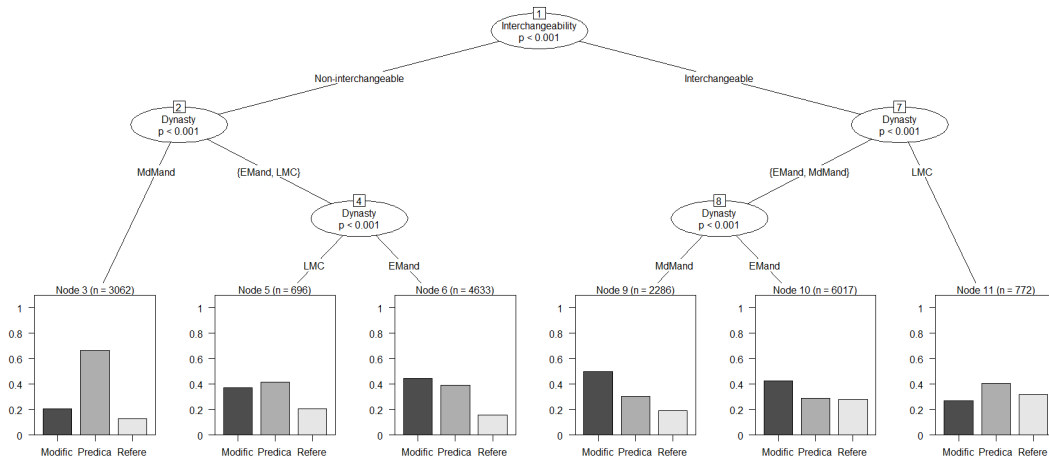
Figure 4 Conditional inference tree of FCIs' function and dynasty

The plot is obtained with the "ctree" function of R package "party" (Levshina, 2015:291). Note that the tree above has no connection with a generative one. It is completely usage-based, computing holistically probabilities among dimensions like form, meaning, context, and pragmatic factors. In Figure 4, conditional dependencies among variables exclusively depend on statistical significance. In other words, the higher the node, the more significant the "conditional decision" (Tantucci & Wang, 2018). It is clear that for the computational classification of propositional functions of FCIs, the most important predictor appears to be "interchangeability" as it is at the top of the conditional inference tree. That is, in terms of FCI's functions in a given context, interchangeability is the most decisive independent variable.

Illocutional concurrences (IC) refer to significant intersections of the variables subsumed by the dimensions (Tantucci & Wang 2018, p. 69). Namely, ICs encompass converging factors at different levels of verbal experience that contribute, both locally (i.e. at the morphosyntactic level) and peripherally (i.e. at the illocutionary level), to the encoding of contextually and culturally situated speech acts or pragmemes (Mey, 2001; Capone, 2005; Tantucci, 2016a). The first illocutional concurrence (IC) that is worth noting runs from nodes 1 to 6, showing the functional usage of non-interchangeables over three historical periods. This IC is extracted from the original plot in Figure 4 above:
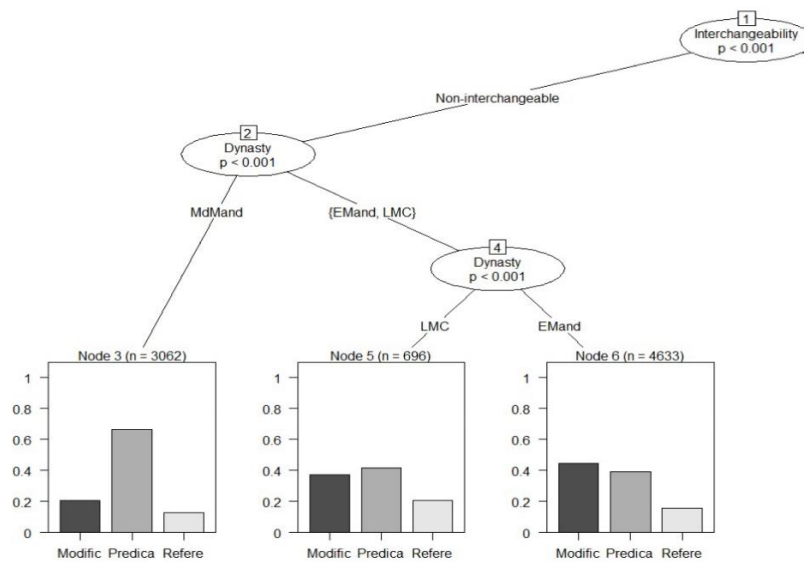
Figure 5 One illocutional concurrence (IC) of non-interchangeables in terms of propositional functions

In Figure 5, it is worth noting that the diachronic functional change in non-interchangeable idioms has a different first split in node 2, leading to two sub-groups: a) Modern Mandarin and b) Late Middle Chinese together with Early Mandarin (Nodes 1-2).

In Late Middle Chinese (Nodes 1-2-4-5), the percentages that non-interchangeables serve as modification, predication, and referencing are 38%, 42%, and 20%, respectively (Nodes 1-2-4-5), and such distributions are similar in Early Mandarin with 44% in Modification, 41% in predication, and 18% in referencing (Nodes 1-2-4-6). However, in Modern Mandarin, non-interchangeables serve such roles as modification, predication, and referencing (20%, 70%, and 10% respectively). The constructional change of non-interchangeables is therefore not linear.

More specifically, the percentage of the modification function that non-interchangeable idioms serve increases from 38% in Late Middle Chinese (Nodes 1-2-4-5) to 44% in Early Mandarin (Nodes 1-2-4-6), and then decreases to 20% in Modern Mandarin (Nodes 1-2-3). The percentage of the predication function that non-interchangeable idioms serve decreases from 42% in Late Middle Chinese (Nodes 1-2-4-5) to 38% in Early Mandarin (Nodes 1-2-4-6) and then increases to 70% in Modern

Mandarin (Nodes 1-2-3). The percentage of the referencing function that non-interchangeable idioms serve shows a decreased tendency from 20% in Late Middle Chinese (Nodes 1-2-4-5) to 18% in Early Mandarin (Nodes 1-2-4-6) to 10% in Modern Mandarin (Nodes 1-2-3).

This means that the constructional change of non-interchangeables is not linear, that is, the tendency of modification increased in Early Mandarin, and then decreased in Modern Mandarin. In other words, the diachronic function change in non-interchangeables is non-linear, and is determined by their internal constituency. As discussed in Section 5.2, internal constituency of non-interchangeables has fewer symmetrical structures (Types 1, 2, 3, 4, and 5). This means that the internal constituency of non-interchangeables is less predictable, which affects the usage of Chinese FCIs, further interfering with the directionality of their constructional change. This also explains why the percentage of modification in non-interchangeables increases from 38% in Late Middle Chinese (Nodes 1-2-4-5) to 44% in Early Mandarin (Nodes 1-2-4-6), and then decreases to 20% in Modern Mandarin (Nodes 1-2-3), which is contrary to the hypothesis 2.

Quite different is the diachronic functional change in interchangeable idioms, with its first split in node 7 which leads to two sub-groups: a) Late Middle Chinese and b) Early Mandarin and Modern Mandarin as shown in Figure 6.
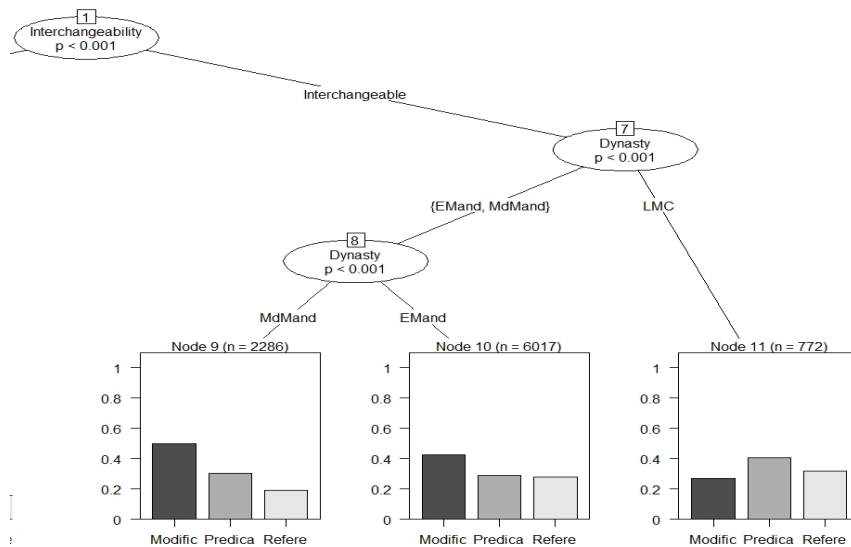
Figure 6 One illocutional concurrence (IC) of interchangeables in terms of propositional functions

The second illocutional concurrence (IC) that is worth noting runs from node 7 to 11, showing the functional usage of interchangeables over three historical periods. This IC is extracted from the original plot in Figure 4 above. In Late Middle Chinese, the percentages of interchangeables are shown as 28% (modification), 40% (predication), and (32%) referencing. In Early Mandarin non-interchangeables serve such roles as modification, predication, and referencing (40%, 30%, and 30% respectively), and such distributions in Modern Mandarin are 50% (modification), 30% (predication) and 20% (referencing). The constructional change of interchangeables is linear as Figure 5 shows.

More specifically, the percentage of the modification function that interchangeable idioms serve increases from 28% in Late Middle Chinese (Nodes 1-7-11) to 40% in Early Mandarin (Nodes 1-7-8-10), and then it increases to 50% in Modern Mandarin (Nodes 1-7-8-9). However, the percentage of the predication function that interchangeable idioms serve decreases from 40% in Late Middle Chinese (Nodes 1-7-11) to 30% in Early Mandarin and Modern Mandarin. The referencing function also shows a decreased tendency from 32% (LMC) to 30% (EMand) to 20% (MdMand).

This means that, different from what we can see for non-interchangeables, the

constructional change of interchangeables is indeed linear because the increasing tendency of modification keeps rising (see hypothesis 2), while the percentages of referencing and predication functions tend to drop. Simply put, the diachronic function change in interchangeables appears to be linear. This is due to the fact that the internal constituency of interchangeables is comparatively more predictable (with more symmetrical structures), which entails a more matching correspondence between their internal constituency and their propositional act functions. This makes the development of constructional change in interchangeables more linear.

Another difference between interchangeables and non-interchangeables in terms of function is that non-interchangeables in Modern Mandarin more often serve the predication function which takes up more than 60% in Nodes 1-2-3, while interchangeables are more frequently used for modification which takes up more than 50% in Nodes 1-7-8. This means that interchangeables are somehow more often associated with the modification function while non-interchangeables are frequently linked with the predication function in Modern Mandarin. Consider example (12) below:

(12) 蓦然间　　　瞧见　　一 个 明眸皓齿　　　　　　的　　美人儿

　　　 mò-rán-jiān qiáo-jiàn yī gè míng-móu-hào-chǐ　　　 de　　mĕi-rén-er

　　　 suddenly　　 see　　 a CL bright-eyes-white-teeth PART beautiful people

　　　 "Suddenly, <someone> has seen a beautiful lady who has bright eyes and white teeth"

　　　　　　　　　 History of Miyagi in the Ming Dynasty (Modern Mandarin, 1,912-1,949 A.D.)

In (12), the interchangeable idiom 明眸皓齿 (*míng-móu-hào-chǐ*, bright-eyes-white-teeth) is used as an attributive adjective (the modification) to modify its following noun 美人儿 (*mĕi-rén-er*, beautiful lady) in a novel from the Minguo period (Modern Mandarin).

(13) 你 助纣为虐,　　　　　　　　　　使　韩宝、吴志广　　对抗　　朝廷?

Nǐ zhù-zhòu-wéi-nüè,　　　　　　　　shǐ　hánbǎo, wúzhìguǎng duì-kàng cháotíng?

you help Emperor Zhou do bad things, make hanbao, wuzhiguang　fight　　the royal court?

"You took the side of the evil-doer, making Bao Han and Zhiguang Wu counterwork the royal court?"

<div align="right">Yongzheng swordsman diagram Modern Mandarin (1,912-1,949 A.D.)</div>

In (13), the non-interchangeable idiom 助纣为虐 (*zhù-zhòu-wéi-nüè*, "take the side of the evil-doer") serves the predication function in Modern Mandarin as the idiom is a verb which follows the subject 你 (*nǐ*, you).

An important assumption of this thesis is that that internal constituency affects the usage of Chinese FCIs. In other words, the differences between interchangeables and non-interchangeables in terms of function partly lie in their differences of internal constituency. In order to shed new light on how different idiom types are used in actual language use, I fitted a Multiple Correspondence Analysis (MCA; Nenadic and Greenacre, 2007). This model allows one to model associations between different variables by calculating their chi-square distance (cf. Tantucci and Wang, 2019 on MCA for diachronic analysis). These associations are then represented graphically as a map, which eases the interpretation of the data structures — the closer the distance between variables, the stronger the statistical correspondence. The interaction between idiom types and their functions can be captured on a two-dimensional space with a multiple correspondence analysis (see Figure 8).
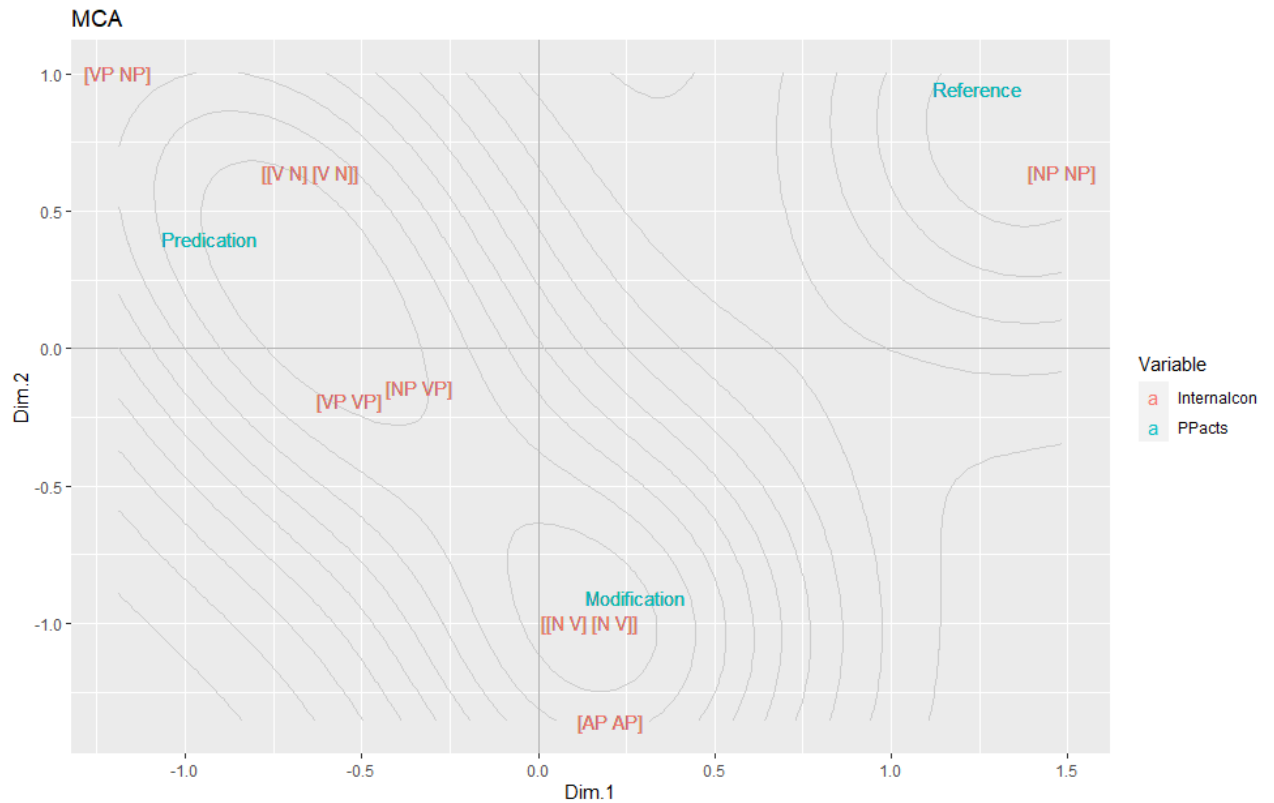
Figure 7 Two-dimensional correspondence of idiom types and functions

In Figure 7, the two dimensions represent 100% of variation, which is considered a perfect approximation for MCA visualization. The triangular distribution of density areas reflects the disposition of the seven types of internal constituency in red (Types 1 to 7) and three propositional acts in green (modification, predication, and referencing). The figure illustrates the variable that significantly affects the usage of the interchangeable FICs — internal constituency (in red).

It can be observed that there is an association between the predication function (Dim. 1: -1, Dim. 2: 0.5) and Types 3, 5, 6, 7 ([VP VP], [[V N] [V N]], [NP VP], [VP NP]), suggesting that these types are more often used as predication. The reason for this is that verbs are more often used as the predicate in actual language use, and these types can be realized as a verb based on their constructions. It can also be found that the referencing function is located near Type 1 [NP NP] FCIs (Dim. 1: 1.3, Dim. 2: 0.7), meaning that Type 1 FCIs more often serve the referencing function. Type 1 is nonetheless more often used as referencing compared with other Types. The main

reason for this is that a noun phrase is prototypically associated with the propositional act function of reference. Therefore, the propositional act function of Type 1 FCIs (two NP phrases) is also predicted to be reference due to a match between their internal constituency and their usages as a whole. Moreover, a strong association is also found between the modification function (Dim. 1: 0.3, Dim. 2: -1) and Types 2, 4 ([[N V] [N V]], AP+AP). This means both [[N V] [N V]] and AP+AP constructions are more often used to modify an entity.

In Section 5.2, I have found that non-interchangeables display a higher type frequency in Types 3 ([VP VP] construction), Type 6 ([NP VP]) and 7 ([VP NP] construction) when compared with interchangeables. From the dataset, types 3, 6 and 7 in non-interchangeables have yielded 5,580 occurrences, while those types in interchangeables have only yielded 1,188 occurrences. This may lead to the result that non-interchangeables in Modern Mandarin more often serve the predication function as these types are more often used for predication. Similarly, interchangeables display a higher type frequency in Types 2 and 4 ([AP AP], [[N V] [N V]]) when compared with non-interchangeables. There are 2883 occurrences that are retrieved from Types 2 and 4 in interchangeables, while only 1111 occurrences that are found from Type 2 and 4 in non-interchangeables. This may also lead to the result that interchangeables in Modern Mandarin more often serve the modification function as these types are more often used for modification.

In conclusion, Type 1 FCIs tend to combine with the referencing function because they can be regarded as a 'noun chunk' that is composed of two NP units. The reason why Types 3, 5, 6, and 7 FCIs correlate with the predicate function can be ascribed to its VP units being holistically treated as a verb. Types 2 and 4 FCIs tend to combine with the modification function because they are a combination of AP units. I, therefore, argue that the differences in actual usage of these seven types are largely determined by the internal constituency of the FCIs, and the distribution differences further cause the differences of interchangeables and non-interchangeables in terms of function.

5.6 Summary

Symmetricity plays a crucial role in determining the (non-)interchangeability of an idiom. However, iconicity strongly constrains the interchangeability of Chinese FCIs. The token frequency of the FCIs had reached the highest point in Early Mandarin (1250 — 1800), and then it decreased in Modern Mandarin (1800 — present), presumably because of the development of literary works in Early Mandarin and the May Fourth Movement in Modern Mandarin. Productivity and schematicity of interchangeables increase in Early Mandarin (1250 — 1800) because of the new BBAA schema, while the token frequency of the interchangeables decreases in Modern Mandarin (1800 — present) due to the fact that the token frequencies of the interchangeables (AABB and BBAA) decreased as a result of competition (see Chapter 5). The internal constituency of the FCIs affects not only the idioms' propositional functions, but also their constructional changes. That is to say, Type 1 FCIs tend to combine with the referencing function, Types 3, 5, 6, and 7 correlate with the predicate function, and Types 2 and 4 perform the modification function. The constructional change of the interchangeables is linear, while that of non-interchangeables is not. This is plausibly due to the fact that the change of interchangeables is comparatively more predictable (with more symmetrical structures) than the one of non-interchangeables.

Chapter 6 Differences between AABB and BBAA structures of interchangeable idioms

6.1 Outline

This chapter will focus only on interchangeables. More specifically, it will look at the differences between AABB and BBAA [5] in terms of type frequency, token frequency, and function. Section 6.2 illustrates the methods that I use to distinguish AABB from BBAA, and it also provides some examples of AABB and BBAA. Section 6.3 introduces the mechanisms that govern the diachronic change of the two forms. Section 6.4 discusses the differences between AABB and BBAA in terms of type frequency, token frequency, and their respective functions from a diachronic perspective. Section 6.5 further illustrates the types of mechanisms that are most frequently applied to the interchangeable idioms.

6.2 Distinguishing AABB from BBAA in interchangeables

There are 428 interchangeables (AABB and BBAA) that were retrieved in Chapter 4 (see section 4.3: idioms selection criterion), out of which there are 138 Type 1 interchangeables ([NP NP]), 62 Type 2 ([AP AP]), 34 Type 3 ([VP VP]), 66 Type 4 ([[N V] [N V]]), 106 Type 5 ([[V N] [V N]]), 11 Type 6 ([NP VP]), and 11 Type 7 ([VP NP]) (see section 5.2 for distributions of interchangeables and non-interchangeables idioms in seven types). This section discusses the three methods for categorising them functionally and formally.

6.2.1 A time-based method for distinguishing AABB from BBAA

The first method is to check the time of appearance of the idioms in CCL. The form which appears earlier than the other form is labelled as AABB, whereas the newer one

---

[5] AABB refers to those interchangeable idioms which appeared first in our dataset, while BBAA appeared later.

is labelled as BBAA. For example, 皓齿明眸 (*hào-chǐ-míng-móu*, white-teeth-bright-eyes) and 明眸皓齿 (*míng-móu-hào-chǐ*, bright-eyes-white-teeth) are a pair of Type 1 interchangeables ([NP NP]). The idiom 明眸皓齿 (*míng-móu-hào-chǐ*) is first recorded in Tang dynasty, while 皓齿明眸 (*hào-chǐ-míng-móu*) first appears in Ming Dynasty (1,368-1,644 A.D.) as shown below:

(14) 明眸皓齿　　　　　今　何在?

　　　míng-móu-hào-chǐ,　　jīn　hé-zài?

　　　bright eyes white teeth now where?

　　　Where has the young beautiful lady been?

<div style="text-align: right">Du Fu's poems (Tang Dynasty, 618-907 A.D.)</div>

(15) 朱颜绿发,　　　　　　皓齿明眸

　　　zhū-yán-lǜ-fā,　　　　hào-chǐ-míng-móu

　　　dark red face green hair, white teeth and bright eyes

　　　<Someone> is young and beautiful with white teeth and bright eyes.

<div style="text-align: right">Water Margin (Ming Dynasty, 1,368-1,644 A.D.)</div>

Based on the CCL, 明眸皓齿 (*míng-móu-hào-chǐ*, 618-907 A.D.) appears earlier than 皓齿明眸 (*hào-chǐ-míng-móu*, 1,368-1,644 A.D.). Thus, 明眸皓齿 (*míng-móu-hào-chǐ*) is categorized as AABB, and is labelled as [NP NP]1, whereas 皓齿明眸 (*hào-chǐ-míng-móu*) is categorized as BBAA, and is labelled as [NP NP]2.

6.2.2 A token frequency-based method for distinguishing AABB from BBAA

In those scenarios where AABB and BBAA may appear in the same period, it is essential to posit a second criterion to distinguish AABB from BBAA, which was looking at the token frequencies of the two forms. My methodological assumption is

that, if two interchangeable types both first appear during the same period in the corpus, the form which has a higher token frequency is likely to have arisen before form with a lower token frequency. Note that frequency does not necessarily reflect earlier stages of development, but it can be a feasible criterion to suggest that one form is more widespread than the other in history. In those cases when two idioms first appear in the same period, such frequency-based method is perhaps the only inferential way to distinguish AABB from BBAA. For example, 千山万水 (*qiān-shān-wàn-shuǐ*, a thousand-mountain-ten thousand-river) and 万水千山 (*wàn-shuǐ-qiān-shān*, ten thousand-river-a thousand- mountain) both are first recorded in Tang Dynasty (618-907 A.D.) as examples shown below:

(16) 而 过　万水千山　　　　　　　　　　之　　险,　　贵　有　馀　香

　　　ér guò wàn-shuǐ-qiān-shān　　　　　　zhī　xiǎn,　　guì　yǒu yú　xiāng

但 but pass ten thousand rivers and a thousand mountains PART obstacles, worth has left fragrance

"The meaning of overcoming all kinds of difficulties is that you will get a sense of achievement."

<div align="right">Tang Wen Supplements (Tang dynasty, 618-907 A.D.)</div>

(17) 今　则　千山万水,　　　　　　　　　杳 隔　二　途

　　 jīn zé qiān-shān-wàn-shuǐ,　　　　　yǎo gé　èr　tú

　　 now be a thousand mountains and ten thousand rivers, far gap two roads

　　 "Currently there is a huge distance between us, and we are far away from each other like living in two different places."

<div align="right">Quotations from Master Wu Ben (Tang dynasty, 618-907 A.D.)</div>

These two idioms both first appear in Tang Dynasty (618-907 A.D.), so 千山万水 (*qiān-shān-wàn-shuǐ*) cannot be distinguished from 万水千山 (*wàn-shuǐ-qiān-shān*) merely based on their first appearance in history. Although these two forms appear at during the same period, the token frequency of 万水千山 (*wàn-shuǐ-qiān-shān*, 8 occurrences)

is higher than that of 千山万水 (*qiān-shān-wàn-shuǐ*, 1 occurrence). Thus, 万水千山 (*wàn-shuǐ-qiān-shān*) is categorized as AABB ([NP NP]1), and 千山万水 (*qiān-shān-wàn-shuǐ*) BBAA ([NP NP]2).

Note that there are cases where the two forms have appeared in the same period with the exact token frequency. This is where I checked the token frequencies in their following dynasty until finding the one with a higher token frequency. For example, 安贫乐道 (*ān-pín-lè-dào*, accept-poverty-happy-spiritual, "be content with poverty and devoted to spiritual things") and 乐道安贫 (*lè-dào-ān-pín*, happy-spiritual-accept-poverty, "devoted to spiritual things and be content with poverty") are a pair of Type 5 interchangeables ([[V N] [V N]]). These two idioms first appear in the Tang Dynasty (618-907 A.D.) as examples shown below:

(18) 安贫乐道, 　　　　　　　不 以 宠 辱 担惊

　　　ān-pín-lè-dào, 　　　　　bù yǐ chǒng rǔ dān-jīng

　　　accept poverty happy spiritual, not base doting abusing be afraid of

　　　"Being happy to lead a simple and virtuous life, not worrying about getting or losing something."

　　　　　　　　　　　Epitaph Compilation Sequel (Tang dynasty, 618-907 A.D.)

(19) 乐道安贫 　　　　　　　者

　　　lè-dào-ān-pín 　　　　　　zhě

　　　happy spirituals accept poverty person

　　　"One who is devoted to a spiritual life despite poverty."

　　　　　　　　　　　Wang Wei's Poems (Tang dynasty, 618-907 A.D.)

Based on the CCL, the token frequencies of the two idioms are both 1 in Tang Dynasty (618-907 A.D.). A hapax is a word or an expression that appears only once in a single text or corpus (Lardilleux & Lepage, 2007). As these two expressions only occur once in Tang, they can be referred to as hapax legomena, which is an important indicator in

corpus-based approaches to language change (Hilpert 2015; Tantucci & Di Cristofaro 2020).

It was first thought that a hapax legomenon presents a problem for corpus-based research because its low frequency fails to provide sufficient statistical data for analyzing fields like word alignment or statistical machine translation (Schrader, 2006). Neologisms (newly coined words) and misspelt words are the two main cases that contribute to hapax legomena (Lardilleux & Lepage, 2009). For neologisms, the number of hapax legomena depends on language change on its own, but for misspelt words, the number of hapax legomena depends largely on the quality of the corpus. In other words, misspelled words are usually unique occurrences (hapaxes), but their overall frequency remains relatively low. (Lardilleux & Lepage, 2009).

However, there are several crucial aspects about hapax legomena which make them an important factor in language change. Firstly, hapaxes take up a large proportion of word tokens (Cartoni 2006) in that they generally represent around 40% of words in a corpus, but this percentage may vary based on a) the richness of the vocabulary and b) the degree of synthesis of the language (Lardilleux & Lepage, 2009). The richer the vocabulary in a corpus or text, the higher proportion of hapaxes. If a language displays a higher level of synthesis, it will generate a larger vocabulary — therefore, more examples of hapaxes. For example, in Inuktitut, a highly synthetic language of Eastern Canada, more than 80% of hapaxes are found, as reported by Langlais et al. (2005). Secondly, hapaxes actually can contribute to the majority of the word alignments (Lardilleux & Lepage, 2009). Thirdly, hapax legomena are a positive indicator for the productivity of a morpheme (Pierrehumbert & Granell, 2018). However, Chinese is an isolating language and does not include many inflectional changes. Therefore, in this study, while hapax legomena are not an important indicator for morphological inflection, they are yet key for accessing whether a particular idiom occurs first or later in the corpus. The fact that 安贫乐道 (*ān-pín-lè-dào*, accept-poverty-happy-spiritual, "be content with poverty and devoted to spiritual things") and 乐道安贫 (*lè-dào-ān-pín*, happy-spiritual-accept-poverty, "devoted to spiritual things and be content with

poverty") are hapax legomena and that both of them first appear during the Tang dynasty (618-907 A.D.) makes it impossible to distinguish them when deciding which form appears first. Then, I had to check the token frequencies in the dynasty (Song 960-1,279 A.D.) following that of the hapax legomena, and found 17 occurrences of 安贫乐道 (*ān-pín-lè-dào*, accept-poverty-happy-spiritual) and 1 occurrence of 乐道安贫 (*lè-dào-ān-pín*, happy-spiritual-accept-poverty). This suggests that the former is more widespread than the latter. For this case, 安贫乐道 (*ān-pín-lè-dào*, accept-poverty-happy-spiritual) is categorized as AABB, and is labelled as [[V N] [V N]]1, whereas 乐道安贫 (*lè-dào-ān-pín*, happy-spiritual-accept-poverty) is categorized as BBAA, therefore, labelled as [[V N] [V N]]2 based on Method 2 (token frequency-based). Method 2 has the caveat of being retrospectively inferential, but is perhaps the most rigorous way possible to assume which form is older when diachronic data are not reflected in the corpus.

6.2.3 Method 3: A record-based method for distinguishing AABB from BBAA

The third method for classifying AABB and BBAA is what I called the record-based method, as in cases where only one form has been recorded, but not the other in the CCL. In these cases it is impossible to distinguish AABB from BBAA based on the previous two methods. To address this issue, I categorised the one found in the CCL as AABB, while the other one (which is present in the Xinhua dictionary, but not in the corpus) as BBAA. It is also to be noted that when two forms cannot be compared in terms of timeline, the record-based method is also a relatively more sensible method to distinguish AABB from BBAA. For example, 百孔千疮 (*bǎi-kǒng-qiān-chuāng*, a hundred-wound-ten thousand-hole) and 千疮百孔 (*qiān-chuāng-bǎi-kǒng*, ten thousand-hole-a hundred- wound) are the two forms of the same Type 1 interchangeable ([NP NP]). There is one occurrence for 百孔千疮 (*bǎi-kǒng-qiān-chuāng*) as shown in example (5), but no occurrence for 千疮百孔 (*qiān-chuāng-bǎi-kǒng*).

(20) 已经　弄得　　百孔千疮,　　　　　　　　　　　　　　　背了　一身　亏累

yǐjīng　nòng-dé bǎi-kǒng-qiān-chuāng,　　　　　　　　bèi-le yī-shēn kuī-lěi

already make　a hundred wounds and ten thousand holes, burden a body exhausted

"Someone has already been through more than he or she can afford, and is deeply in debts."

The strange situation witnessed in 20 years (Qing, 1,644-1,912 A.D.)

Thus, the idiom 百孔千疮 (*bǎi-kǒng-qiān-chuāng*) is labelled as AABB ([NP NP]1), while the idiom 千疮百孔 (*qiān-chuāng-bǎi-kǒng*) as BBAA ([NP NP]1). All of the above exemplars and methods that are used for distinguishing AABB from BBAA can be illustrated in the following table.

Idioms	Types	Methods
明眸皓齿 (*míng-móu-hào-chǐ*, bright-eyes-white-teeth)	[NP NP]1 (AABB)	Method 1 time-based
皓齿明眸 (*hào-chǐ-míng-móu*, white-teeth-bright-eyes)	[NP NP]2 (BBAA)	Method 1 time-based
万水千山 (*wàn-shuǐ- qiān-shān*, ten thousand-river-a thousand-mountain)	[NP NP]1 (AABB)	Method 2 token frequency-based
千山万水 (qiān-shān-wàn-shuǐ, a thousand-mountain-ten thousand-river)	[NP NP]2 (BBAA)	Method 2 token frequency-based
安贫乐道 (*ān-pín-lè-dào*, "be content with poverty and devoted to things spiritual")	[[V N] [V N]]1 (AABB)	Method 2 token frequency-based
乐道安贫 (*lè-dào-ān-pín*, "devoted to things spiritual and be content with poverty")	[[V N] [V N]]2 (BBAA)	Method 2 token frequency-based
百孔千疮 (*bǎi-kǒng-qiān-chuāng*, a hundred-wound-ten thousand-hole)	[NP NP]1 (AABB)	Method 3 record-based
千疮百孔 (*qiān-chuāng-bǎi-kǒng*, ten thousand-hole-a hundred-wound)	[NP NP]2 (BBAA)	Method 3 record-based

Table 11 Exemplars of AABB and BBAA in Type 1 interchangeables

These three methods apply to all the symmetrical interchangeables (Types 1, 2, 3 ,4, 5). There are 69 [NP NP]1 idioms and 69 [NP NP]2 idioms in all 138 Type 1 interchangeables, 31 [AP AP]1 and 31 [AP AP]2 in all 62 Type 2 interchangeables, 17 [VP VP]1 and 17 [VP VP]2 in all 34 Type 3 interchangeables, 33 [[N V] [N V]]1 and 33 [[N V] [N V]]2 in all 66 Type 4 interchangeables, and 53 [[V N] [V N]]1 and 53 [[V N] [V N]]2 in all 106 Type 5.

However, the three methods will not be applied to the asymmetrical interchangeables (Types 6, 7) due to the fact that inter-switching the two units of such interchangeable types shall result in the change of the constituents' word classes, leading to further subsets not to yield accurate results. For example, 黑白混淆 (*hēi-bái-hùn-xiáo*, "black and white" + "mistake with") and 混淆黑白 (*hùn-xiáo-hēi-bái*, "mistake with" + "black and white"), where 黑白混淆 (*hēi-bái-hùn-xiáo*) is the combination of NP (黑白: *hēi-bái*, black and white) and VP (混淆: *hùn-xiáo*, mistake with), while 混淆黑白 (*hùn-xiáo-hēi-bái)* is one of VP (混淆: *hùn-xiáo*, mistake with) plus NP (黑白: *hēi-bái*, black and white). If I applied the above three methods to asymmetrical interchangeables, 黑白混淆 (*hēi-bái-hùn-xiáo*) would be labelled as [NP VP] 1 as it appears first in my dataset, but the other form 混淆黑白 (hùn-xiáo-hēi-bái) would not be labelled as [NP VP] 2 due to the exchange of word classes. Furthermore, I shall not discuss which asymmetrical interchangeable comes before the other, but will focus on how the two types (types 6, 7) differ from each other in terms of type frequency, token frequency, and functionality.

To sum up, for the symmetrical interchangeables (Types 1, 2, 3, 4, 5), I aim to discuss the differences between the two forms (AABB and BBAA) in each type, but for the asymmetrical interchangeables (Types 6, 7), I only focus on how these two types differ from each other instead of finding out which asymmetrical interchangeable comes before the other.

6.3 Three mechanisms governing the change of AABB and BBAA

In regard to language development, functionally similar forms are often in competition with one another (De Smet et al., 2018). When it comes to Chinese FCIs, an example of functionally similar forms is the interchangeables AABB and BBAA. They can be identified as a case of "isomorphism" (Givón, 1991; Bolinger, 1977; Haiman, 1980; Wierzbicka, 1988). In linguistics, the concept of isomorphism is often associated with Kuryłowicz (1949) and mainly means "leparallélisme complet des deux plans du contenu et de l'expression" (Martinet 1957, p. 105), i.e. a unique mapping between form and meaning (see Dressler 1999). AABB and BBAA are, respectively two unique mappings between form and meaning. According to De Smet at al. (2018, pp. 198-199), ideally, isomorphism means that a single form should express a single function. However, Van de Velde (2014) argues that form-function relations are much more complicated because they are actually organized in many-to-many mappings instead of the simple one-to-one relationship. More specifically, when it comes to Chinese FCIs, [VP VP]1 is not fully synonymous with [VP VP]2 because the former mainly serve the predication function, while the latter the modification function (see Figure 10). Two forms (AABB and BBAA) of other types of interchangeables remain the same function.

This argument applies to the Chinese FCIs as different types of FCIs can all be mapped onto three different propositional acts. When the two forms (i.e. AABB and BBAA) of an interchangeable are competing with each other, the function change and token frequency change of the two forms can be regarded as the two important indicators for deciding which mechanism is at work with which type(s) of interchangeables. In terms of function change, there are two possibilities in which both forms are more often associated with a certain function or in which both forms have their own unique function. In terms of token frequency, there are also two possibilities in which both forms are frequently used or in which one form may be preferred over the other.

The three mechanisms governing the competition of AABB and BBAA have been defined as substitution, differentiation, and attraction (De Smet et al., 2018). Competition leads to only one form surviving (substitution) or each form finding its unique niche in functional space (differentiation) (De Smet et al., 2018). In substitution, the functional domain over which two forms compete comes to be occupied by a single form at the expense of all others. That is, there is no function shift between the two forms, but the token frequency of one form is higher than that of the other in substitution. Take for example the following two sentences: *I considered Neal as a killer* and *You consider the operation Ø a success?* The first sentence is regarded as an as-secondary predicate construction (as-SPC), while the second one zero-secondary predicate construction (zero-SPC). Both secondary predicate constructions are found to extend to mental verbs and perception verbs (D'hoedt, 2017), but over time [consider + zero-SPC] has been on the increase, whereas [consider + as-SPC] has declined (De Smet et al., 2018). However, in differentiation, the functional domain that was being competed over is split, with each expression taking on a distinct functional role (Berg, 2014). For example, the [begin + -ing-clause] construction is highly relevant to situations involving an agentive subject, while [begin + to-infinitive] occurs with both agentive and non-agentive subjects; however, when the subject is non-agentive, the to-infinitive is the expected form (De Smet et al., 2018). Substitution and differentiation are schematically represented in Figure 8 (De Smet et al., 2018, p. 198).
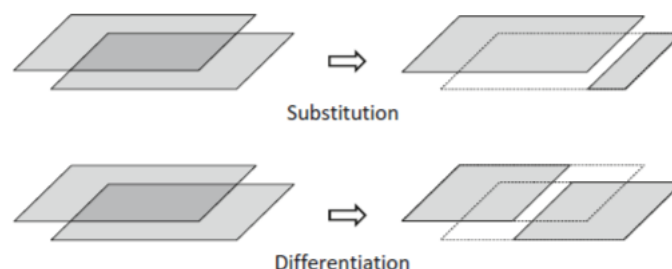

Figure 8 Mechanisms: Substitution and Differentiation

In the process of substitution, the two full grey solid rectangles stand for the two functionally similar forms (AABB and BBAA in interchangeables). When they are competing with each other, one form (the solid rectangle) will overlap with the other form (the dotted rectangle) in terms of function. Substitution often leads to only one form gaining a high token frequency, while the other form shows a decreased tendency in usage because of the language economy principle (Vicentini, 2003). However, in the differentiation process, the two full grey solid rectangles also stand for the two functionally similar expressions, but the functional overlap between them diminishes as the figure shows that each rectangle has been reduced in function (half solid and half dotted). In other words, each reduced rectangle must represent a certain different function. In terms of token frequency, both forms can coexist and survive.

On the one hand, there is similarity between substitution and differentiation as argued by De Smet (2018, pp. 198-199) that "in both substitution and differentiation, form-function pairings are reorganized in such a way that functional overlap is reduced". On the other hand, there are differences between the two mechanisms. One difference is that substitution must require the loss of a particular form-meaning pairing (one form prevails over the other in terms of token frequency), while differentiation does not. Another difference is that differentiation must require the assignment of new functions to existing forms (De Smet et al., 2018), while in substitution both forms keep the same function.

However, substitution and differentiation cannot easily explain why some functionally similar forms co-exist. De Smet (2018) proposed the notion of "attraction" as a third mechanism to explain how functionally similar forms maintain and increase functional overlap in language. When two expressions show functional overlap, they are in fact likely to become more similar, as if being attracted to each other. De Smet (2018) argues that attraction may be a natural consequence of analogy (Anttila, 2003; Fischer, 2007; Aaron, 2016; Itkonen, 2005; Wanner, 2011), which is especially likely to take place between variant expressions, causing them to switch syntactic patterns (from AABB to BBAA) and share similar functions. Attraction is schematically

represented in Figure 9 (De Smet et al., 2018, p. 204). When attraction applies to interchangeable FCIs, both forms coexist and share the similar function.
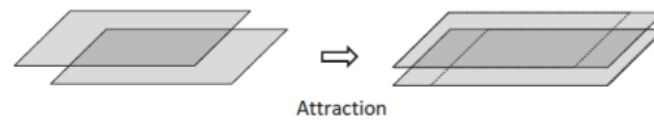
Figure 9 Mechanism: Attraction

In the process of attraction, the two full grey solid rectangles on the left stand for the two functionally similar constructions (AABB and BBAA in interchangeables). The competition between them ends up with both forms maintaining the same for a certain function and they coexist through time, as the figure shows that both rectangles are still full. This means that attraction does not result in the loss of either form; both forms mainly serve the same function as a result of analogy. For example, Rosenbach (2007, p. 168) argued that in the seventeenth century, there was "barely any semantic overlap between genitives and noun modifiers". Genitives were typically restricted to human nouns (*the man's new car*), while noun modifiers were usually confined to inanimate nouns (*a school bus*). However, in the eighteenth century, both genitives and noun modifiers were used with collective nouns (*the court/court's favourite*), showing some overlap. This overlap continued to grow, with genitives increasingly distributing to inanimate nouns, while noun modifiers to animates. Another case of attraction is the study of *though* and *although* by Hilpert (2013: Ch. 5). He showed that *though* and *although* had slightly different syntactic preferences in nineteenth century, but later on the preferences gradually converged.

My argument is that interchangeables (AABB and BBAA) are undergoing, in different forms, substitution, differentiation, and attraction, and that the said three mechanisms have governed the change of different types of interchangeables. More specifically, attraction can be found in Types 1 and 2 interchangeables ([NP NP] and AP+AP), differentiation in Types 3, 6, and 7 ([VP VP], [NP VP], and [VP NP]), substitution in Type 4 and 5 ([N+V] + [N+V], and [[V N] [V N]]) (see the following

chapters for a detailed discussion).

6.4 Type frequency, token frequency and function differences between AABB and BBAA

As I argued previously, different types of interchangeable idioms have undergone different kinds of change. It is thus important to investigate AABB and BBAA in each type in terms of type frequency, token frequency, and their functions. This way it will be possible to look at how the three mechanisms are applied to each type. I sorted the type frequencies and normalized the token frequencies of all the interchangeables according to their internal constituencies as shown in Table 12. This table is used to make comparison of the two forms (AABB and BBAA) in each interchangeables in terms of token frequency. The reason why the two different orders are presented in this why is that the first one is the older one, and the order one has the highest token frequency except for Type 3. This is because I used the token frequency-based method to I distinguish AABB from BBAA if the two forms appear in the same period. Although token frequency does not necessarily reflect earlier stages of development, but this can be a feasible method to distinguish AABB from BBAA.

Types	AABB or BBAA	Type Frequency		Occurrences	
Type 1	[NP NP]1	69	138	1453	2649
	[NP NP]2	69		1196	
Type 2	[AP AP]1	31	62	1033	1475
	[AP AP]2	31		442	
Type 3	[VP VP]1	17	34	237	744
	[VP VP]2	17		507	
Type 4	[[N V] [N V]]1	33	66	749	1408
	[[N V] [N V]]2	33		659	
Type 5	[[V N] [V N]]1	53	106	1248	2342
	[[V N] [V N]]2	53		1094	
Type 6	[NP VP]	11	22	300	447
Type 7	[VP NP]	11		147	

Table 12 Type frequency and normalized token frequency of interchangeables

The table shows that the [NP NP] construction has the highest type frequency (N = 138) in interchangeables, followed by [[V N] [V N]] (N = 106), [[N V] [N V]] (N = 66), [AP AP] (N = 62), [VP VP] (N = 34), [NP VP] (N = 11), and [VP NP] (N =11), and that the [NP NP] construction has the highest token frequency (N = 2,649), followed by [[V N] [V N]] (N = 2,342), [AP AP] (N = 1,475), [[N V] [N V]] (N = 1,408), [VP VP] (N= 744), [NP VP] (N = 300), and [VP NP] (N=147). Generally speaking, the frequency distribution of the two forms (AABB and BBAA) in each type reflects the overall difference in distribution (the first form has a higher token frequency than the second form), but the two forms of Type 3 may have skewed the overall distribution as [VP VP]1 has a lower token frequency than [VP VP]2. Besides, the distribution of any two forms in different types is generally in line with the overall distribution, but they are exceptions. The token frequency of [NP NP]1 佳人才子 (*jiā-rén-cái-zǐ*, gifted-woman-gifted-man, "an adorable couple of lovers") is 119, while the token frequency of [NP NP]2 才子佳人 (*cái-zǐ-jiā-rén*, gifted-man-gifted-woman, "an adorable couple of lovers"). However, the token frequency of [NP NP]1 万水千山 (*wàn-shuǐ-qiān-shān*, ten thousand-river-a thousand-mountain, "a long journey during which

numerous obstacles are encountered") is only 57, while the token frequency of [NP NP]2 千山万水 (qiān-shān-wàn-shuǐ, a thousand-mountain-ten thousand-river, "a long journey during which numerous obstacles are encountered") is 116.

One interesting element emerging the classification above is that a higher type frequency construction does not necessarily yield a higher token frequency as argued in Berg (2014) and Traugott & Trousdale (2013). For example, the type frequency of the [[N V] [N V]] construction (N = 66) is higher than that of the [AP AP] construction (N = 62), but the token frequency of the [[N V] [N V]] construction (N = 1408) is lower than that of the [AP AP] construction (N = 1475).

It should be noted that Type 1 ([NP NP]) registers the highest type frequency (N = 138), followed by Type 5 ([[V N] [V N]]) (N = 106), but in Type 1, the type frequencies of both [NP NP]1 and [NP NP]2 are 69. Similarly, both [[V N] [V N]]1 and [[V N] [V N]]2 share the same type frequency which is 53. This indicates that the main difference between AABB and BBAA does not lie in type frequency as the two forms in each type remain the same. Therefore, it is vital to find the differences between AABB and BBAA in terms of their function and token frequency, which shall shed light on the mechanisms that are applied to the different interchangeable types. Section 6.4.1 discusses how symmetrical interchangeables change over time and section 6.4.2 sheds light on asymmetrical interchangeables in terms of token frequency and function.

6.4.1 Symmetrical interchangeables change over time in terms of token frequency and function

This section discusses how two forms of a symmetrical interchangeable interact with one another. More specifically, it investigates token frequencies of the two forms in each type over three historical periods and explores the two forms' functional change over time.

Besides showing the type frequency and token frequency of AABB and BBAA in symmetrical interchangeables, this study aims to shed light on how the internal

constituency of respectively AABB and BBAA is structured and how it changes over time in connection with their propositional act functions. In order to achieve this, I plotted an MCA (see Section 5.5) by using the normalized token frequency from Table 5.

The interaction between symmetrical interchangeables and their functions can be captured on a two-dimensional space with an MCA (see Figure 10). In Figure 10, the two dimensions represent 100% of variation, which is considered an optimal approximation of internal variance for MCA visualization (Levshina, 2015). The label 'Internalcon' in red refers to the different internal types of constituency in interchangeables, while the one 'PPacts' in green refers to the functions that interchangeables serve in context.
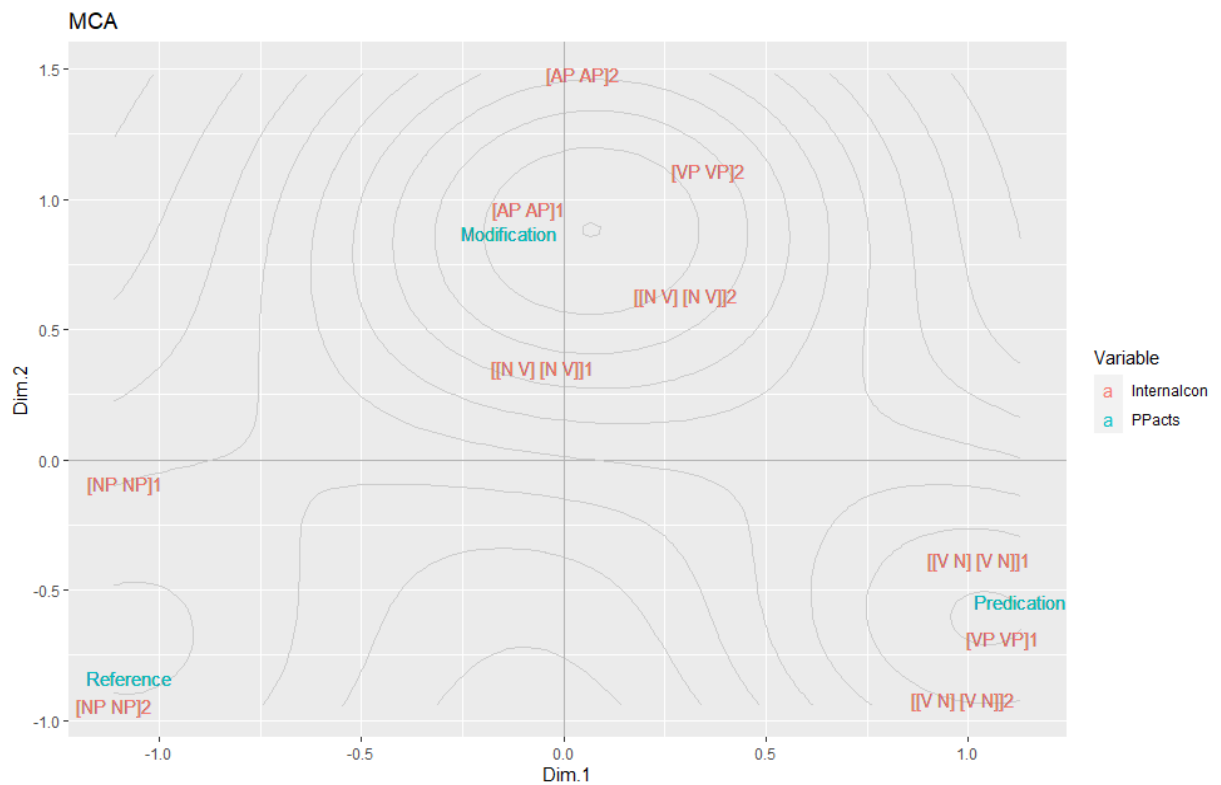
Figure 10 Two-dimensional correspondence of symmetrical interchangeables and propositional act functions

6.4.1.1 The competition between [NP NP]1 and [NP NP]2

I will now proceed to investigate the following:

a) the types of mechanisms that underpin Type 1 interchangeables ([NP NP]) and

b) the differences between [NP NP]1 and [NP NP]2 in terms of type frequency, token frequency, and their functions.

From Figure 10, it can be seen that there is a strong association among variables [NP NP]1 (Dim. 1: -1, Dim. 2: 0), [NP NP]2 (Dim. 1: -1, Dim. 2: -1) and Reference (Dim. 1: -1, Dim. 2: -1), namely showing that both [NP NP]1 and [NP NP]2 more prototypically serve the propositional act of reference (cf. Croft 2001).

For example, the [NP NP]1 idiom 佳人才子 (cái-zǐ-jiā-rén, gifted scholars) and [NP NP]2 idiom 才子佳人 (*jiā-rén-cái-zǐ*, gifted scholars) are both used as the subject in examples (21) and (22),

[NP NP]1

(21) 自古 及今, 佳人才子, 少 得 当年 双美

zì-gǔ-jí-jīn,   jiāréncáizǐ,   shǎo dé dàng-nián shuāng-měi

from.beginning till.now, excellent.people.intelligent.person, seldom.meet in.suitable.age two.beautiful

"<In history> Rarely has a gifted intellectual been able to come upon a beautiful lady who happened to admire him equally."

Liu Yong's poems (Song Dynasty, 960-1,279)

[NP NP]2

(22) 才子佳人   乘 酒力,   大家   今夜 好 降龙.

cái-zǐ-jiā-rén   chéng jiǔ-lì,   dàjiā   jīnyè hǎo xiáng-lóng

intelligent.person.excellent.people take.advantage.of wine, everyone tonight good defeat.the.dragon

"Intelligent people drank wine, and they all will have a happy and smooth night."

Wake-up Marriage Story (Ming Dynasty, 1,368-1,644)

Both 佳人才子 (*cái-zǐ-jiā-rén*) and 才子佳人 (*jiā-rén-cái-zǐ*) serve the propositional act of reference in the above examples. The main reasons for this is that subject and object functions usually are represented by a concrete entity or an abstract idea, and that [NP NP]1 and [NP NP]2 can be regarded as a single nominal which often refers to a concept. Therefore, there is a functional overlap between [NP NP]1 and [NP NP]2. Indeed, functional overlap seems to be common and sometimes it appears to be diachronically more or less stable (Torres Cacoullos & Walker, 2009). Attraction can maintain and increase functional overlap in language (De Smet et al., 2018). In other words, when functionally similar forms are experiencing attraction, either AABB or BBAA can serve the same function. Although these two forms maintain the same reference function and type frequency (N = 69) in the three historical periods, they yield different token frequencies in history. It is then important to look at the how the token frequencies of [NP NP]1 and [NP NP]2 change over time (see Figure 11).
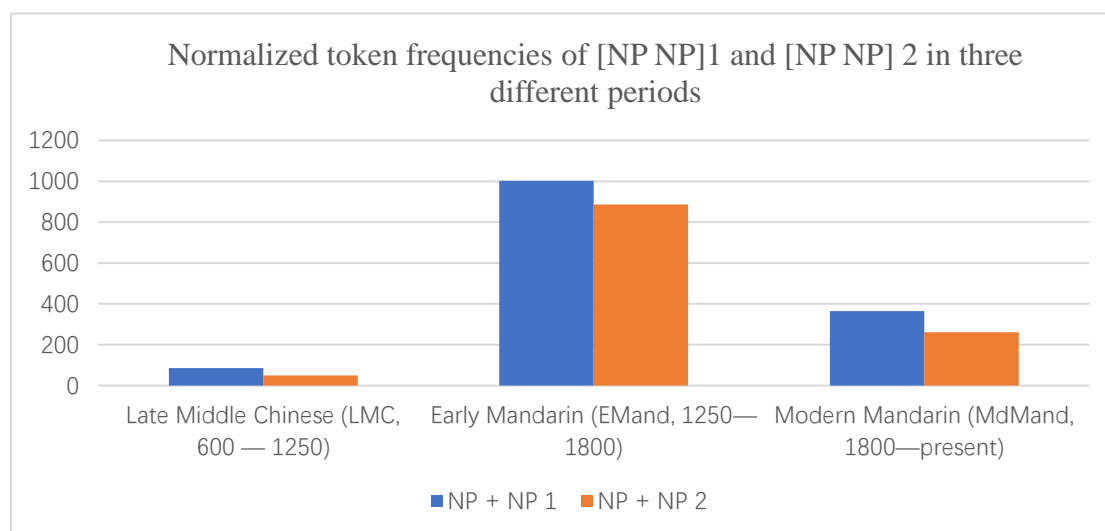
Figure 11 Normalized frequencies of [NP NP]1 and [NP NP]2 in three different periods

In Figure 11, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 1 interchangeables. It can be observed that the token frequency of [NP NP]1 is always higher than that of [NP NP]2 in three Chinese historical periods. However, this does not necessarily mean that [NP NP]1 prevails over [NP NP]2 just because the token frequency gap between the two

forms is not huge. That is to say, [NP NP]2 is relatively lower than [NP NP]1 in terms of token frequency, but the absolute token frequencies of both forms are similar. This means that both AABB and BBAA in the [NP NP] construction coexist in the three periods. The relationship between two functionally similar forms ([NP NP]1 and [NP NP]2) can be often seen in competing (construction-based) for long-term survival in language history (De Smet et al., 2018).

I argue that Early Mandarin is the period where [NP NP]1 and [NP NP]2 have entered this competition and coexisted as the normalized token frequency of both forms are largely increasing. It implies that both forms are widely used in this period. For example, the frequency of [NP NP]1 佳人才子 (*cái-zǐ-jiā-rén*, gifted scholars) has grown from 1 to 22 between Late Middle Chinese and Early Mandarin, while that of BBAA 才子佳人 (*jiā-rén-cái-zǐ*, gifted scholars) from 0 to 28 during the same period.

In Modern Mandarin, the same tendency (both forms feature a similar absolute token frequency) that has been found throughout Early Mandarin is also preserved. In fact, there are 4 occurrences of [NP NP]1 佳人才子 (*cái-zǐ-jiā-rén*, gifted scholar) in Modern Mandarin, and there are 6 for [NP NP]2 才子佳人 (*jiā-rén-cái-zǐ*, gifted scholar).

The two forms have been in competition for a long time, with both [NP NP]1 and [NP NP]2 being frequently used in Early Mandarin and Modern Mandarin, due to constructional analogy, where "the behaviour of one expression is modelled after the behaviour of another which it resembles" (De Smet, 2018, p.217).

To sum up, the type frequencies of [NP NP]1 and [NP NP]2 are both 69, while [NP NP]1 (N = 1453) yields a slightly higher token frequency than [NP NP]2 (N = 1196). The token frequency gap between the two forms is not wide during the three periods, where both forms are found to coexist and be frequently used. Moreover, functions that the two forms serve remain the same as the propositional act of reference. Thus, Type 1 interchangeables (NP+NP) are undergoing attraction.

6.4.1.2 The competition between [AP AP]1 and [AP AP]2

I will now proceed to investigate the following:

a) the types of mechanisms that underpin Type 2 interchangeables ([AP AP]) and

b) the differences between [AP AP]1 and [AP AP]2 in terms of type frequency, token frequency, and their functions.

In Figure 10, it can also be observed that there is a strong association among variables [AP AP]1 (Dim. 1: 0, Dim. 2: 1), [AP AP]2 (Dim. 1: 0, Dim. 2: 1.5), Modification (Dim. 1: 0, Dim. 2: 1), suggesting that both of the two forms often serve the "propositional act of modification" (Croft, 2001) as in examples (23), (24),

[AP AP]1

(23) 侯、童　二位　　　　是 光明磊落　　　　　　　　　　的　　侠士。

　　hóu, tóng èr-wèi　　shì guāng-míng-lěi-luò　　　　de　　xiá-shì

　　hou, Tong two persons be upright/open righteous and open-hearted PART Knight.

　　"Mr. Hou and Mr. Tong are two frank and straightforward heroes."

Yongzheng swordsman diagram (Minguo, 1,912-1,349 A.D.)

[AP AP]2

(24) 论 平日　　为人，　　是 我们　　素所　敬服　的

　　lùn píng-rì wéi-rén,　shì wǒ-men sù-suǒ jìng-fú de,

　　talk daily behaviour, be our 　　always respect PART,

　　磊落光明　　　　　　　　　　的　　好 学生

　　lěi-luò-guāng-míng　　　　　　de　　hào xué-shēng

　　upright/open righteous and open-hearted PART good student

　　"Judged by his daily behaviours, he is a righteous and kind-hearted student who is always respected by us."

The Sequel of Liudong's story (Minguo, 1,912-1949 A.D)

In (23), [AP AP]1 光明磊落 (*guāng-míng-lěi-luò*, frank and straightforward) is used as with attributive adjective function to modify the noun 侠士 (*xiáshì*, knight). Similarly, [AP AP]2 磊落光明 (*lěi-luò-guāng-míng*, straightforward and frank) is also used to modify the noun 学生 (*xuéshēng*, students). Both idioms share the same modification function because they are a combination of two AP units that are often used to modify an entity. Although both [AP AP]1 and [AP AP]2 are often associated with the modification function, the proportions of [AP AP]1 are always higher than those of AP +AP 2 during the three different periods (see Figure 12).
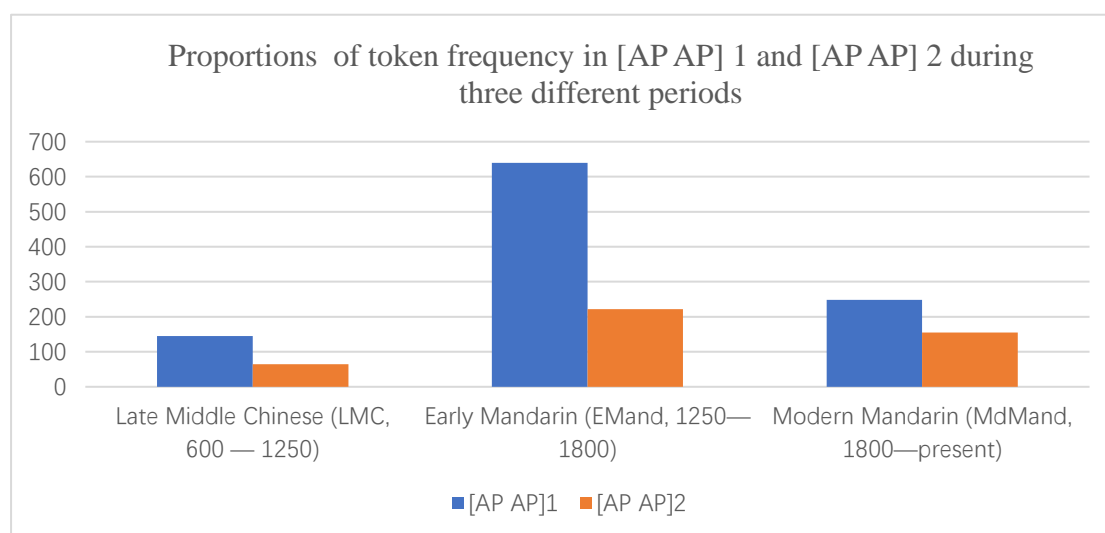
Figure 12 Normalized frequencies of [AP AP]1 and [AP AP]2 in three different periods

In Figure 12, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 2 interchangeables. More specifically, the token frequencies of [AP AP]1 are 145 in Late Middle Chinese, 640 in early Mandarin, and 248 in Modern Mandarin, while the token frequencies of [AP AP]2 are 65, 222, and 155. It can be observed that the token frequency of [AP AP]1 is always higher than that of [AP AP]2 in three Chinese historical periods. However, this does not necessarily mean that [AP AP]1 prevails over [AP AP]2 just because the token frequency gap between the two forms is reduced from 418 in Early Mandarin to 93 in Modern Mandarin. Although the token frequency of [AP AP]1 is always higher than

that of AP +AP 2 during the three periods, I nonetheless argue that they actually coexist in history based on the fact that the token frequency gap is decreased from 418 in Early Mandarin to 93 in Modern Mandarin. To put this simply, there is no function shift between the two expressions and they coexist all the way through. Therefore, Type 2 interchangeables ([AP AP]) are arguably also undergoing attraction.

6.4.1.3 The competition between [VP VP]1 and [VP VP]2

I will now proceed to investigate the following:

a) the types of mechanisms that underpin Type 3 interchangeables ([VP VP]) and

b) the differences between [VP VP]1 and [VP VP]2 in terms of type frequency, token frequency, and their functions.

In Figure 10, it can be observed that there is a strong association between variables [VP VP]1 (Dim. 1: 1, Dim. 2: -0.7) and Predication (Dim. 1: 1, Dim. 2: -0.5), and that there is another strong association between variables [VP VP]2 (Dim. 1: 0.3, Dim. 2: 1) and Modification (Dim. 1: 0, Dim. 2: 1), suggesting that [VP VP]1 is more often used as predicate, while [VP VP]2 more often serve the modification function. For example, [VP VP]1 半疑半信 (*bàn-yí-bàn-xìn*, half-doubt and half-believe) serves the predication function in example (26), while [VP VP]2 半信半疑 (*bàn-xìn-bàn-yí*, half-believe and half-doubt) serves the modification function in examples (27) and (28).

[VP VP]1

(26) 仰山　　半疑半信,　　　　　　却　也　不　能　性　　急

　　　yǎngshān bàn-yí-bàn-xìn,　　　　què yě bù néng xìng jí

　　　shan Yang half-doubt and half-believe, but also not can temper hurry

　　　"Shan Yang doubts it, but he cannot ask too much about it."

　　　　　　A Romance of the Three Hundred Years of Yan History (Minguo, 1,912-1,349 A.D.)

In (26), [VP VP]1 半疑半信 (*bàn-yí-bàn-xìn*) is a verb after the subject (仰山, *yǎngshān*), serving the predication function. The fact that [VP VP]1 correlates with the

predicate function can be ascribed to its VP units being holistically treated as a verb, and to the fact that verbs are more often used as the predicate in actual language use.

[VP VP]2

(27) 大家 半信半疑　　　　　地　守　等

dàjiā bànxìnbànyí　　　　　de　shǒu děng

everyone half-believe and half-doubt PART guard wait

"Everyone waits with doubt."

<div align="right">Humble Words of a Rustic Elder (Qing, 1,644-1,912 A.D.)</div>

In (27), [VP VP]2 半信半疑 (bàn-xìn-bàn-yí) is used as the adverb to modify the verb (等, děng, wait) , serving the modification function.

(28) 不能 不 有 些 半信半疑　　　　的　心思

bù-néng bù yǒu xiē bànxìnbànyí　　　　de　xīnsī

cannot　not have some half-believe and half-doubt PART thoughts

"You cannot but be critical in thinking."

<div align="right">The Sequel of Liudong's story (Minguo, 1,912-1949 A.D)</div>

In (28), [VP VP]2 半信半疑 (*bàn-xìn-bàn-yí*) is used adjectivally to modify the noun (心思, xīnsī, thoughts) , serving the modification function. Interestingly, based on its "[VP VP]" structure, however, [VP VP]2 is more often associated with the modification function rather than the predication function. This means that there is a function shift in interchangeable [VP VP] constructions. The first form in [VP VP] construction mainly functions as a predicate, while the second form is used as a modifier. Therefore, there is a difference between [VP VP]1 and [VP VP]2 in terms of function. Besides functional differences, it is important to further investigate the token frequency differences between them in order to figure out which mechanisms (substitution, differentiation, attraction) can be applied to Type 3 interchangeables (see Figure 13).
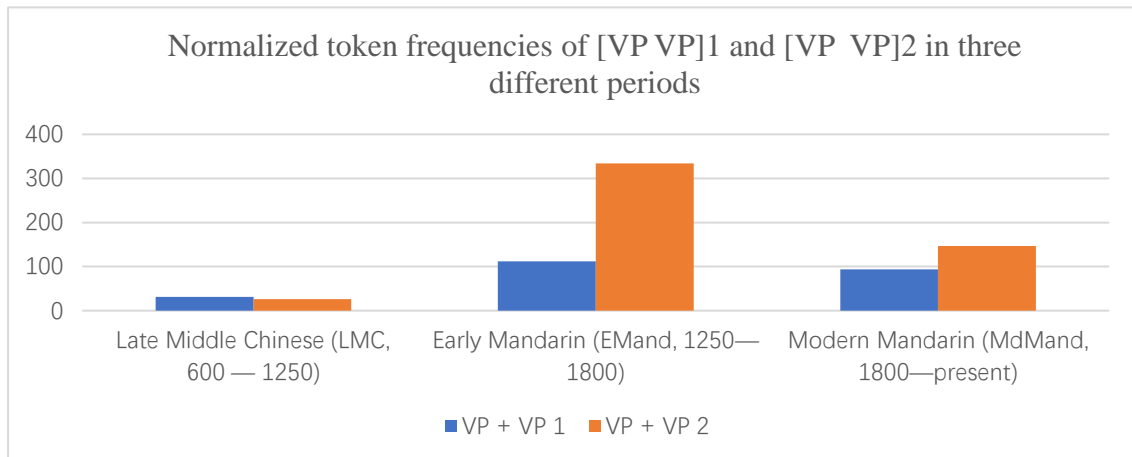
Figure 13 Normalized frequencies of [VP VP]1 and [VP VP]2 in three different periods

In Figure 13, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 3 interchangeables. It can be seen that the token frequency of VP +VP 1 is slightly higher than that of [VP VP]2 in Late Middle Chinese. In early Mandarin, the token frequency of [VP VP]2 is obviously rising from 26 to 334 occurrences, while the token frequency of [VP VP]1 from 31 to 112. In Modern Mandarin, the token frequency of [VP VP]2 (N = 147) is still higher than that of [VP VP]1 (N = 94). This shows that [VP VP]1 only prevails in the first period, and that when the two forms are competing, [VP VP]2 prevails in its following periods.

Unlike those in Type 1 ([NP NP]) and Type 2 ([AP AP]), Type 3 ([VP VP]) interchangeables, in turn, are undergoing substitution. In substitution, the competition between the two forms leads to one form taking the place of the other. That is, "substitution requires the loss of a particular form-meaning pairing" (De Smet et al., 2018, p. 199). There are two main reasons why one expression may gain more popularity than the other in substitution process. The first reason is that one expression has a higher social prestige than the other (Labov, 1972; Keller, 1990; Croft, 2000; Labov, 2001), while the second reason is that one expression has better adapted to its new (emergent) function (Haspelmath, 1999; De Smet, 2008; Petré, 2014). The second reason can be used to explain the function change of [VP VP]1 and [VP VP]2.

Specifically, when [VP VP]2 starts competing with its alternative form [VP VP]1, people tend to use [VP VP]1 for predication, while [VP VP]2 serves the modification function. Also, the token frequency change of Type 3 interchangeables is different from that of Type 1 and Type 2. That is, the two forms in Type 1 and Type 2 coexist, while [VP VP]2 prevails over [VP VP]1.

To sum up, there is function shift between the two expressions and [VP VP]2 prevails over [VP VP]1. Therefore, Type 3 interchangeables ([VP VP]) are arguably also undergoing substitution. Note that although the changes in Type 3 involve substitution, they simultaneously show attraction as the two forms in each type tend to behave more similarly in terms of function. That is to say, substitution may presuppose some extent of attraction because two variants must develop sufficient functional overlap to permit one variant to fully replace the other as argued by Traugott (2020, p. 550) that "attraction is a fundamental tendency in change".

6.4.1.4 The competition between [[N V] [N V]]1 and [[N V] [N V]]2

I will now proceed to investigate the following:
a) the types of mechanisms that underpin Type 4 interchangeables ([[N V] [N V]]) and
b) the differences between [[N V] [N V]]1 and [[N V] [N V]]2 in terms of type frequency, token frequency, and their functions.

It can also be seen that there is a strong association among variables [[N V] [N V]]1 (Dim. 1: 0, Dim. 2: 0.3), [[N V] [N V]]2 (Dim. 1: 0.3, Dim. 2: 0.6) and Modification (Dim. 1: -0.1, Dim. 2: 0.8), meaning that both these two forms serve the "propositional act of modification" (Croft 2001). For example, the [[N V] [N V]]1 idiom 鬼使神差 (*guǐ-shǐ-shén-chāi*, ghost orders and god requests) and [[N V] [N V]]2 idiom 神差鬼使 (*shén-chāi-guǐ-shǐ*, god requests and ghost orders) are a pair of Type 4 interchangeables, serving the modification function as in examples (29) and (30).

[[N V] [N V]]1

(29) 鬼使神差　　　　　地　使 我们　知道,

　　Guǐ-shǐ-shén-chāi　　　de　shǐ wǒmen zhīdào,

　　　ghost orders and god requests PART. let us　　know,

　　We were informed of the fact unexpectedly.

<div align="right">Secret History of the Qing (Minguo, 1,912-1949 A.D)</div>

[[N V] [N V]]2

(30) 一个　活 虎 神差鬼使　　　　　地　离了 深山

　　Yīgè huó hǔ, shén-chāi-guǐ-shǐ　　　de　líle　shēnshān

　　　A living tiger, god requests and ghost orders PART. leave the mountain

　　A tiger unexpectedly left the mountain

<div align="right">Wake-up Marriage Story (Ming Dynasty, 1,368-1,644 A.D.)</div>

In (29), 鬼使神差 (*guǐ-shǐ-shén-chāi*) is used to modify the verb (知道, *zhīdào*, inform) and in (30) 神差鬼使 (*shén-chāi-guǐ-shǐ*) is used to modify the verb (离, *lí*, leave). In other words, there is no function shift between the two forms as they share the same propositional act function of modification. Interestingly, although Type 4 interchangeables operate on the [[N V] [N V]] construction, the function that they serve is mainly modification which is often realized via Type 2 interchangeables ([AP AP]). In other words, [[N V] [N V]] and [AP AP] constructions are similar in terms of function despite the fact that they are composed of different word classes. That is to say, when Chinese speakers use FCIs to modify an entity, they are likely to treat [[N V] [N V]] adjectivally, or, in other words, as if [[N V] [N V]] constructions were [AP] ones. It is then also important to further investigate the token frequency differences between [[N V] [N V]]1 and [[N V] [N V]]2 in order to figure out which mechanisms (substitution, differentiation, attraction) can be applied to Type 4 interchangeables (see Figure 14).
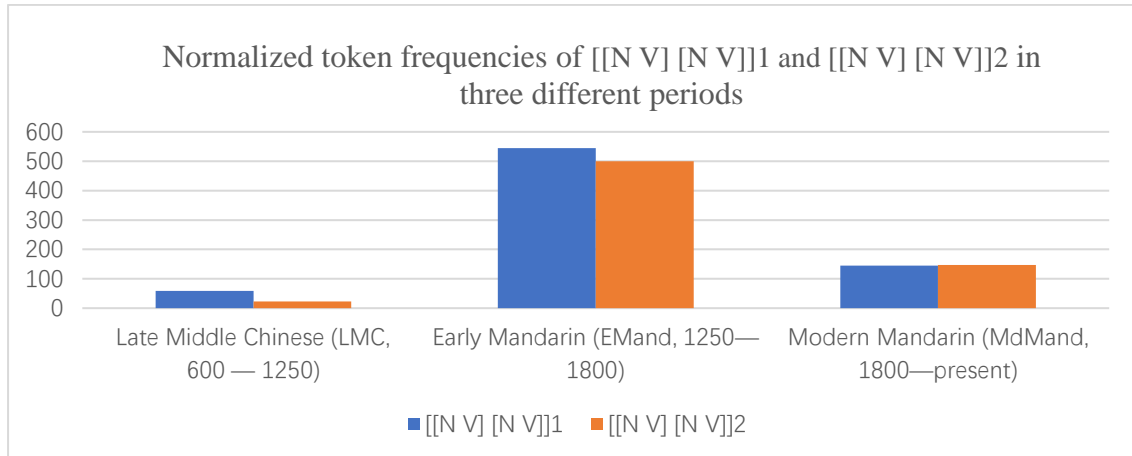
Figure 14 Normalized frequencies of [[N V] [N V]]1 and [[N V] [N V]]2 in three different periods

In Figure 14, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 4 interchangeables. It can be seen that the token frequency of [[N V] [N V]]1 is slightly higher than that of [[N V] [N V]]2 in Late Middle Chinese. More specifically, the token frequency of [[N V] [N V]]1 is 59, while that of [[N V] [N V]]2 is 23. The reason for this tendency might be that [[N V] [N V]]1 displays a high token frequency as it first appears in history. In early Mandarin, the token frequency of [[N V] [N V]]2 is obviously rising from 23 to 500 occurrences, while the token frequency of [[N V] [N V]]1 from 59 to 545. In Modern Mandarin, the token frequency of [[N V] [N V]]2 (N = 147) is even higher than that of [N V] [N V]]1 (N = 145), but this does not mean that [[N V] [N V]]1 prevails over [[N V] [N V]]2 because the token frequency gap between the two forms is not huge.

To sum up, there is no function shift between [[N V] [N V]]1 and [[N V] [N V]]2, and they coexist all the way through from Late Middle Chinese (600-1,250) to Modern Mandarin (1800-present). Thus, Type 4 interchangeables ([[N V] [N V]]) are undergoing attraction.

6.4.1.5 The competition between [[V N] [V N]]1 and [[V N] [V N]]2

I will now proceed to investigate the following:

a) the types of mechanisms that underpin Type 5 interchangeables ([[V N] [V N]]) and

b) the differences between [[V N] [V N]]1 and [[V N] [V N]]2 in terms of type frequency, token frequency, and their functions.

It can also be seen that there is a strong association among variables [[V N] [V N]]1 (Dim. 1: 1, Dim. 2: -0.3), [[V N] [V N]]2 (Dim. 1: 1, Dim. 2: -0.5) and Predication (Dim. 1: 1, Dim. 2: 0.5), suggesting that both these two forms serve the "propositional act of predication" (Croft 2001). For example, a pair of Type 5 interchangeables can be seen in (31) the [[V N] [V N]]1 idiom 顿足捶胸 (*dùn-zú-chuí-xiōng*, "to stamp one's foot and to beat one's breast") and (32) [[V N] [V N]]2 idiom 捶胸顿足 (*chuí-xiōng-dùn-zú*, "to beat one's breast and to stamp one's foot").

[[V N] [V N]]1

(31) 张旺　　　　顿足捶胸,　　　　　　　　　　　　唉声叹气

　　 zhāngwàng dùn-zú-chuí-xiōng,　　　　　　 āishēngtànqì

　　 zhangwang stamp one's foot and beat one's breast, moan and groan

　　 "Zhang Wang stamps his feet and beats his breast, sighing in despair."

<div align="right">Three Heroes Sword (Qing, 1,644-1,9112)</div>

[[V N] [V N]]2

(32) 邓辰　　　　捶胸顿足,　　　　　　　　　　　大　放　悲　声

　　 dèng chén chuí-xiōng-dùn-zú,　　　　　　 dà fàng bēi sheng

　　 deng chen beat one's breast and stamp one's foot, big out sad voice

　　 "Deng Chen beats his breast and stamps his feet, crying out loud."

<div align="right">Court Yanshi of the Han Dynasty (Minguo, 1,912-1,949)</div>

In (31), [[V N] [V N]]1 顿足捶胸 (*dùn-zú-chuí-xiōng*) is used as a verb which follows the subject 张旺 (*zhāng wàng*, a name), and similarly in (32), [[V N] [V N]]2 捶胸顿足 (*chuí-xiōng-dùn-zú*) is also used as a verb which follows the subject 邓辰 (*dèng chén*, a

name). Both expressions share the same propositional act function of predication, that is to say, there is no function shift between the two forms. Interestingly, although Type 5 interchangeables operate on the [[V N] [V N]] construction, the function that they serve is mainly predication which is often realized via [VP VP]1, despite the fact that they are composed of different phrasal categories. In other words, when Chinese speakers use FCIs to perform an action, the [[V N] [V N]] construction is often used as a predicate in a way that is similar to the [[VP VP]] construction.

Another important observation is that although Type 4 and Type 5 interchangeables both feature a combination of nouns and verbs, Type 4 ([[N V] [N V]]) is more often associated with the modification function, while Type 5 [[V N] [V N]] the predication. In other words, different combinatory patterns (either [N V] or [V N]) are crucial in achieving different functions of Chinese FCIs.

It is then also key to look at how the token frequencies of [[V N] [V N]]1 and [[V N] [V N]]2 change over time in order to figure out which mechanism is applied to Type 5 interchangeables (see Figure 15).
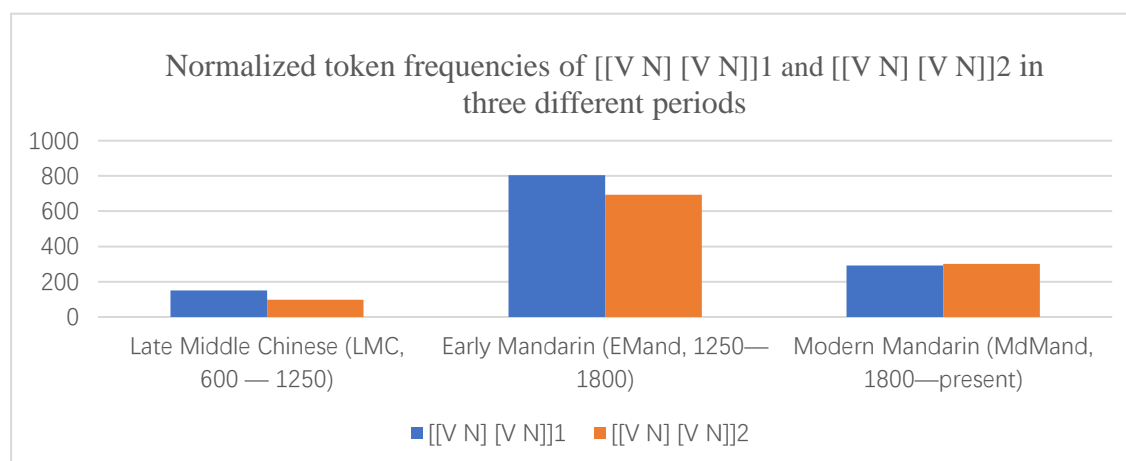

Figure 15 Normalized frequencies of [[V N] [V N]]1 and [[V N] [V N]]2 in three different periods

In Figure 15, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 5 interchangeables. It can be seen that the token frequency of [[V N] [V N]]1 is slightly higher than that of [[V N] [V N]]2

in Late Middle Chinese. More specifically, the token frequency of [[V N] [V N]]1 is 150, while that of [[V N] [V N]]2 is 98. The reason for this tendency might be that [[V N] [V N]]1 displays a high token frequency as it first appears in history. In early Mandarin, the token frequency of [[V N] [V N]]2 is rising from 98 to 694 occurrences, while the token frequency of [[V N] [V N]]1 from 150 to 805. In Modern Mandarin, the token frequency of [[V N] [V N]]2 (N = 302) is higher than that of [[V N] [V N]]1 (N = 293). This shows that the token frequency gap between [[V N] [V N]]1 and [[V N] [V N]]2 is diminished.

Put simply, there is no function shift between the two expressions and they coexist all the way through. Therefore, Type 5 interchangeables ([[V N] [V N]]), like Type 4 (([N+V] + [N+V]), are arguably also undergoing attraction.

All in all, attraction applies to Type 1 ([NP NP]), Type 2 ([AP AP]), Type 4 ([N+V] + [N+V]), and Type 5 ([[V N] [V N]]) interchangeables, and substitution to Type 3 ([VP VP]). The differences between the mechanisms can be summarized in Table 13.

Mechanisms	Idioms Type	Function Shift	Token frequency change
Attraction	Type 1 ([NP NP])	NO	Both forms coexist
Attraction	Type 2 ([AP AP])	NO	Both forms coexist
Substitution	Type 3 ([VP VP])	YES	BBAA prevails over AABB
Attraction	Type 4 ([[N V] [N V]])	NO	Both forms coexist
Attraction	Type 5 ([[V N] [V N]])	NO	Both forms coexist

Table 13 Three mechanisms and their indicators

As I discussed previously, function and token frequency are the two indicators entailing the kind of mechanism applied to distinct types of interchangeables. When there is no function shift and neither of the two forms displays a significantly higher token frequency than the other, the interchangeables (Types 1, 2, 4, and 5) undergo attraction. f there is a function shift between the two forms with either form displaying a significantly higher token frequency, the interchangeables (Type 3) undergo substitution. If any function shift is found between the two forms with both forms

coexisting, then the interchangeables (Types 6 and 7) undergo differentiation (see section 6.4.2).

6.4.2 Asymmetrical interchangeables change over time in terms of token frequency and function

After looking at how the symmetrical interchangeables (Types 1, 2, 3, 4, 5) operate on the three mechanisms (attraction, substitution, and differentiation), I will now investigate how different mechanisms are applied to asymmetrical interchangeables (Types 6, 7). More specifically, I will discuss the differences between Type 6 and Type 7 in terms of token frequency and function. Such functional differences between Type 6 and Type 7 can be captured by Figure 17.
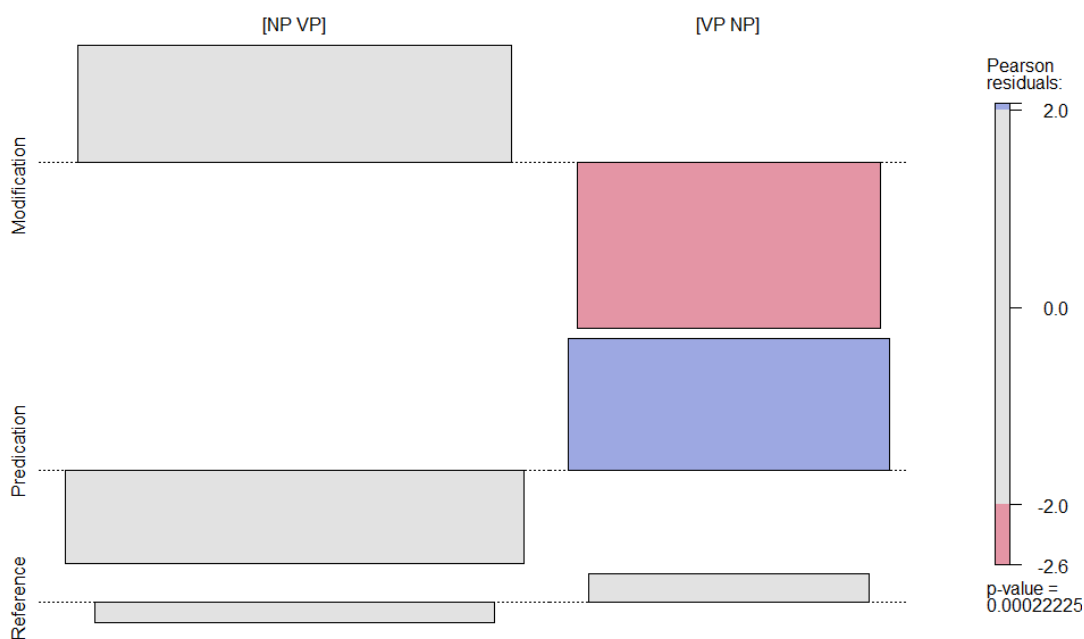
Figure 16 Association plot of residuals: the different usages of [NP VP] and [VP NP] constructions in three functions

Figure 16 shows the Pearson residuals based on the chi-square differences between observed and predicted frequencies, with a significant mismatch between [NP VP] and

[VP NP] constructions in three different propositional acts (X-squared = 16.823, df = 2, p-value = 0.0002225).

The first thing to observe is that [NP VP] shows significantly "positive" residuals in modification in contrast with [VP NP]. This illustrates that the [NP VP] construction is comparatively more attracted to the modification function. For example, the [NP VP] idiom 一钱不值 (yī-qián-bùzhí, a penny not worth) serves the modification function in (33),

[NP VP]

(33) 偏把　　　　　这　一钱不值　　　　　的　　卖身字　　　丢了.

　　piān-bǎ　　　zhè yī-qián-bù-zhí　　de　　mài-shēn-zì diū-le

　　unexpectedly this a-penny-not-worth PART indenture　　lost

　"Even the worthless indenture was unexpectedly lost."

The Sequel of Liudong's story (Minguo, 1,912-1,349 A.D.)

一钱不值 (*yī-qián-bùzhí*) is used as an adjective to modify the noun 卖身字 (*mài-shēn-zì*, indenture), fulfilling the modification function. Interestingly, [NP VP], [AP AP], and [[N V] [N V]] are all more often associated with the modification function although they display different internal constituencies. This implies that these three constructions share some similarity in terms of function in the construction of Chinese speakers.

It also can be observed that [VP NP] shows significantly "positive" residuals in predication in contrast with [NP VP]. This means that the [VP NP] construction is comparatively more associated with the predication function. For example, the [VP NP] idiom 不值一钱 (*bù-zhí-yī-qián*, not worth a penny) serves the predication function in example (34),

[VP NP]

(34) 要 说　这等　法子，　讲破　不 值一 钱

yào shuō zhè-děng fǎ-zi,　jiǎng-pò bù-zhí-yī-qián

If speaking this　approach, actually not worth a penny

"Speaking of this approach, it is actually not worth a penny."

The Eight Immortals (Qing, 1,644-1,912 A.D.)

不值一钱 (bù-zhí-yī-qián) is used as a verb following the subject 法子 (fǎzi, approach), serving the predication function. Note that [VP NP], [VP VP]1, and [[V N] [V N]] constructions display different internal constituencies, but they are more often associated with the predication function. In other words, these three constructions are similar in terms of function to Chinese speakers because they can be realized as a verb based on their internal constituency. Besides the differences in function between [NP VP] and [VP NP] constructions, it is also important to further investigate those in terms token frequency in order to figure out which mechanisms (substitution, differentiation, attraction) can be applied to Type 6 and 7 interchangeables (see Figure 17).
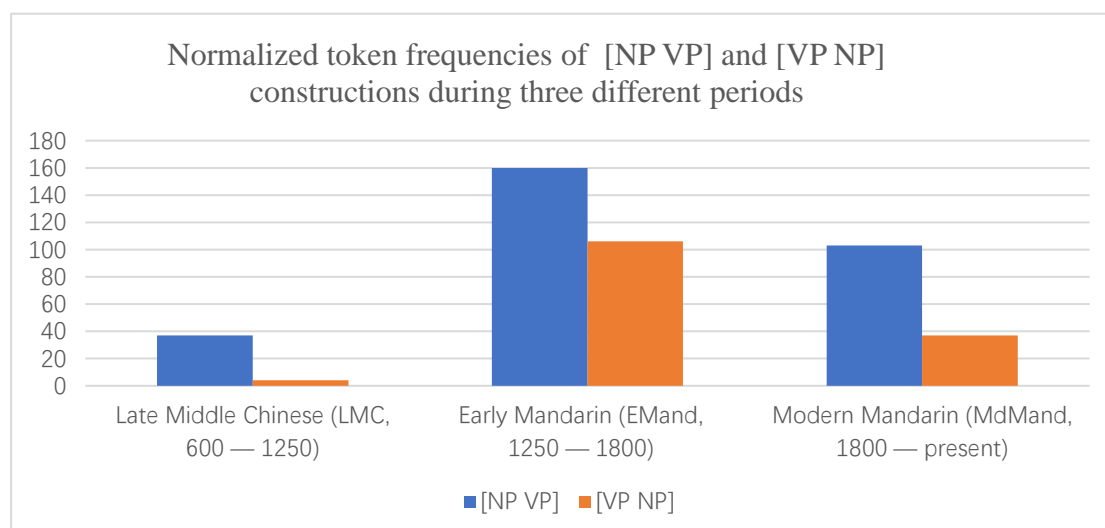
Figure 17 Proportions of [NP VP] and [VP NP] constructions during three different periods

In Figure 17, the X axis stands for the three historical periods in Chinese history, and the Y axis represents the token frequencies of Type 1 interchangeables. It can be observed that the token frequency of [NP VP] is always higher than that of [VP NP] in three Chinese historical periods. However, this does not necessarily mean that [NP VP] prevails over [VP NP] because the token frequency gap between the two forms nearly keeps the same. That is to say, neither [NP VP] nor [VP NP] shows an increased tendency in terms of token frequency, meaning they coexist in the three periods.

To sum up, there is function shift between the two expressions and they coexist. Therefore, Type 6 and Type 7 interchangeables arguably also undergo differentiation.

The motivations for such differentiation are typically based on the conflicting syntactic organization principles (De Smet et al., 2008), such as transparency vs. economy (Körtvélyessy, Štekauer & Zimmermann, 2015), end weight (Eitelmann, 2016) vs. adjacency (Matthews, 2014), and iconicity (Engler, 1995). When it comes to Type 6 and Type 7 interchangeables, the principle of iconicity might be a reason why Type 7 ([VP NP]) is more likely to be used after a subject due to its internal constituency sequence. For example, in (34), the [VP NP] construction (不值一钱, *bù-zhí-yī-qián*, not-worth-a-penny) is more frequently used than the [NP VP] construction 一钱不值 (*yī-qián-bùzhí*, a-penny-not-worth) when it is used after a noun because 不值一钱 (bù-zhí-yī-qián) is more suitable for the predication function due to its [VP NP] structure. It complies with the principle of iconicity. Thus, Type 7 is more often selected to serve the predication function by Chinese speakers in contrast with Type 6 ([NP VP]).

The differences between AABB and BBAA in terms of types, functions and mechanisms are summarised in table 14.

Types	Mechanisms	Forms	Functions	Type frequency	Token frequency Competition
Type 1	Attraction	[NP NP]1	Reference	69	coexist
		[NP NP]1	Reference	69	
Type 2	Attraction	[AP AP]1	Modification	31	coexist
		[AP AP]2	Modification	31	
Type 3	Substitution	[VP VP]1	Predication	17	[VP VP]2 prevails
		[VP VP]2	Modification	17	
Type 4	Attraction	[[N V] [N V]]1	Modification	33	coexist
		[[N V] [N V]]2	Modification	33	
Type 5	Attraction	[[V N] [V N]]1	Predication	53	coexist
		[[V N] [V N]]2	Predication	53	
Type 6	Differentiation	[NP VP]	Modification	11	coexist
Type 7		[VP NP]	Predication	11	

Table 14 Differences in AABB and BBAA in terms of types, functions, and mechanisms

All in all, Types 1 ([NP NP]), Type 2 ([AP AP]), Type 4 ([[N+V] + [N+V]]), and Type 5 ([[V N] [V N]]) interchangeables undergo attraction whereby AABB and BBAA are both frequently used, and both forms in each type mainly serve a similar function. In fact, both [NP NP]1 and [NP NP]2 are more often used for reference, [[N+V] + [N+V]]1 and [[N+V] + [N+V]]2 for modification, [[V N] [V N]]1 and [[V N] [V N]]2 for modification, while [AP AP]1 and [AP AP]2 for modification. Moreover, Type 6 ([NP VP]) and Type 7 ([VP NP]) interchangeables also undergo differentiation. That is, Type 6 more often serves the modification function, while Type 7 the predication function. Unlike Types 1, 2, 4, 5, 6, and 7, Types 3 ([VP VP]) interchangeables undergo substitution. That is, [VP VP]1 is more often used for predication, while [VP VP]2 for modification, and the token frequency of [VP VP]2 prevails over [VP VP]1. Simply put, different types of interchangeables have undergone different diachronic changes which are arguably governed by different mechanisms.

I argue that this is also determined by the internal constituency of Chinese FCIs.

The internal constituencies of Type 1, Type 2, Type 4 and Type 5 are symmetrical chunks of [NP], [AP], [N+V] or [V+N]. This leads to the result that two forms in Types 1, 2, 4, and 5 coexist as the internal constituency of these types are comparatively more stable. The internal constituencies of Type 6 and Type 7 are asymmetrical chunks of [NP] and [VP]. The internal constituencies of 6 and 7 are rather flexible and less stable because they are different combinations of [NP] and [VP], which implies that the function change and token frequency change are more complicated.

It also can be seen from Table 14 that Types 1 (N = 138), Type 2 (N = 62), Type 4 (N = 66), and Type 5 (N = 106) interchangeables are undergoing attraction, Type 3 (N = 34) substitution, while Type 6 (N = 11) and Type 7 (N = 11) differentiation. Thus, attraction is the most frequently found mechanism that applies to interchangeables as it governs the diachronic change of 372 interchangeables (type frequency), followed by substitution which in turn governs 34 interchangeables, and differentiation 22 interchangeables.

6.5 Summary

Attraction, differentiation, and substitution are the three mechanisms that govern the change of different types of interchangeables over time. That is, Types 1 ([NP NP]), Type 2 ([AP AP]), Type 4 ([[N V] [N V]]), and Type 5 ([[V N] [V N]]) interchangeables undergo attraction; Types 3 ([VP VP]) interchangeables undergo substitution; Type 6 ([NP VP]), and Type 7 ([VP NP]) interchangeables undergo differentiation. Attraction is a fundamental tendency in language change (Traugott, 2020), while substitution may presuppose some extent of attraction because two variant forms must share some degree of functional overlap before one could fully replace the other on the basis of token frequency. When it comes to the Chinese FCIs investigated, there are 372 (type frequency-based) interchangeables undergoing attraction, 34 interchangeables substitution, and 22 differentiation.

The change in the above-mentioned mechanisms may be determined by the internal

constituency of the Chinese FCIs. The internal constituencies of Types 1, 2, 4, and 5 are comparatively less flexible than those of Types 6 and 7 because the former are symmetrical [NP], [AP], [N+V] or [V+N], while the latter are different combinations of [NP] and [VP]. This leads to the result that two forms in Types 1, 2, 4, and 5 coexist as their internal constituency are comparatively more stable. On the contrary, the internal constituencies of 6 and 7 are rather flexible and less stable because they are different combinations of [NP] and [VP], which implies that such changes in function or token frequency are more complicated. Although the internal constituency of Types 3 ([VP VP]) FCIs comes in [VP] chunks, there remains a function shift from [VP VP]1 (predication) to [VP VP]2 (modification) (see Figure 10), indicating that it has undergone substitution.

Chapter 7 A synchronic study on the construction [不 (bù, not) A 不 (bù, not) B]

7.1 Outline

This chapter investigates the partly schematic 2+2 Mandarin construction [不 (*bù*, not) A 不 (*bù*, not) B] from a synchronic perspective. More specifically, Section 7.2 introduces the type frequencies of this form and explains the reasons why there are more non-interchangeables than interchangeables in the [*bù* A *bù* B] construction. Section 7.3 deals with the internal constituency of the construction. It also explores the distributions of token frequencies retrieved from CCL and shows how the [*bù* A *bù* B] idioms are used in contexts in terms of their propositional act functions. The relationships between morphemes A and B together with the four different constructional meanings of [*bù* A *bù* B] are explored in Section 7.4. Section 7.5 investigates the relationship between the semantic prosody and lexical distribution of [*bù* A *bù* B] and explains why different types of [*bù* A *bù* B] idioms display a different semantic prosody. Section 7.6 includes the summary and conclusion of this chapter.

7.2 Distributions of the [不 (*bù*, not) A 不 (*bù*, not) B] construction in terms of interchangeability

As I argued in Section 3.3, the Xinhua Dictionary of Idioms (henceforth, the dictionary) has been used as the benchmarking criterion for identifying FCIs in this research. There are 32 [*bù* A *bù* B] FCIs retrieved from the dictionary. To clarify, the 32 [*bù* A *bù* B] idioms are treated as 32 types. These 32 [*bù* A *bù* B] idioms (types) can be further divided into two groups: a) interchangeables (N = 10), and b) non-interchangeables (N = 22). That is, when it comes to the [*bù* A *bù* B] construction, interchangeables are prevalent over non-interchangeables. I will explore all the 32

idioms and speculate on why there are more non-interchangeables have such a distinctively higher frequency.

My hypothesis is that if components A and B do not form a compound word and if there is no iconicity involved in [*bù* A] the semantic relationship between and [*bù* B], then the idiom [*bù* A *bù* B] can be used as [*bù* B *bù* A], i.e., it can be considered as interchangeable. The twelve interchangeable [*bù* A *bù* B] constructions are listed in Table 15.

10 interchangeable [*bù* A *bù* B] idioms	Meaning
不闻不问 (*bù-wén-bù-wèn*, not-listen-not-ask)	showing no interest in something
不问不闻 (*bù-wèn-bù-wén*, not-ask-not-listen)	
不卑不亢 (*bù-bēi-bù-kàng*, not-low self-esteem-not-arrogant)	to be neither humble nor arrogant
不亢不卑 (*bù-kàng-bù-bēi*, not-arrogant-not-low self-esteem)	
不屈不挠 (*bù-qū-bù-náo*, not-surrender-not-yield)	indomitable
不挠不屈 (*bù-náo-bù-qū*, not-yield-not-surrender)	
不骄不躁 (*bù-jiāo-bù-zào*, not-conceited-not-rash)	free from arrogance and impetuosity
不躁不骄 (*bù-zào-bù-jiāo*, not-rash-not-conceited)	
不蔓不枝 (*bù-màn-bù-zhī*, no-creepers-no-branches)	neither spreading about nor branching out
不枝不蔓 (*bù-zhī-bù-màn*, no-branches-no-creepers)	

Table 15 The ten interchangeable [不 (*bù*, not) A 不 (*bù*, not) B] idioms (types)

Note that the expressions above are technically five interchangeable idioms. For example,不闻不问 (*bù-wén-bù-wèn*) and 不问不闻 (*bù-wèn-bù-wén*) are two forms of the same interchangeable idiom, which applies to the remaining four pairs of interchangeable idioms. However, I still count them as ten interchangeables because according to Goldberg (1995:4), a construction is a form-meaning pairing, with different forms which should be idiosyncratically regarded as different constructions. To be noted it also that the first four idioms are defined as interchangeables by the

dictionary. I then identified the remaining six idioms of this category after consulting the corpus zhTenTen (https://www.sketchengine.eu/zhtenten-chinese-corpus/), which is a Chinese corpus made up of texts collected from the Internet. This corpus comprises more than 15.9 billion words. The reason why the six idioms are included is that they qualify as interchangeable [*bù* A *bù* B] idioms based on their syntactic behaviour in the corpus. In the interchangeable [*bù* A *bù* B], the different orders of the interchangeable patterns are truly synonymous because they can be used in the exact same context and they can maintain the same function.

The reasons why these ten idioms are treated as an interchangeable construction are a) A and B in these four idioms do not form one single compound word, B) [*bù* A] and [*bù* B] are not in a temporal or sequential relationship (no iconicity involved). For example, 不闻不问 (*bù-wén-bù-wèn*, not-listen-not-ask) is composed of two VP units which are 不闻 (*bù-wén*, not-listen) and 不问 (*bù-wèn*, not-ask). 闻 (*wén*, listen) and 问 (*wèn*, ask) are not part of a single compound word. The units 不闻 (*bù-wén*, not- listen) and 不问 (*bù-wèn*, not-ask) are considered to be fully transparent constructions, and there remain no syntactic constrains between the units. These units denote two independent events, which means that inter-positioning the two units does not result in any change in meaning as either event can take place first. Thus, 不闻不问 (*bù-wén-bù-wèn*) can be re-organized as 不问不闻 (*bù-wèn-bù-wén*). Similarly, the idiom 不卑不亢 (*bù-bēi-bù-kàng*, not-low self-esteem-not-arrogant) is composed of two AP units which are 不卑 (*bù-bēi*, not-low self-esteem) and 不亢 (*bù-kàng*, not-arrogant). 卑 (*bēi*, low self-esteem) and 亢 (*kàng*, arrogant) are also not part of a single compound word. 不卑 (*bù-bēi*, not-low self-esteem) and 不亢 (*bù-kàng*, not-arrogant) are independent adjective phrases, each describing one's personality. They are not in a temporal or sequential relationship and are considered to be transparent. That is, changing the position of the two units does not alter the meaning of this idiom. So, 不卑不亢 (*bù-bēi-bù-kàng*) can also be re-constructed as 不亢不卑 (*bù-kàng-bù-bē*). The remaining 6 interchangeables can also be tested according to the same hypothesis.

In [*bù* A *bù* B] interchangeables, the units – e.g., 不闻 (*bù-wén*) vs. 不问 (*bù-wèn*), 不卑 (*bù-bēi*) vs. 不亢 (*bù-kàng*), 不屈 (*bù-qū*) vs. 不挠 (*bù-náo*), 不骄 (*bù-jiāo*) vs. 不躁 (*bù-zào*), and 不蔓 (*bù-màn*) vs. 不枝 (*bù-zhī*) – can be considered as fully transparent constructions. The notion of semantic transparency (Schäfer, 2018) is comparable to the one of compositionality (Werning, 2010). In lexical semantics, semantic transparency refers to the degree to which the meaning of a compound word or an idiom can be inferred from the meaning of its constituents (Auch et al., 2020). Compositionality (Pelletier, 1994) refers to the meaning of an expression that is determined by its structure and the meanings of its constituents. Semantic transparency and compositionality have been analysed in several different ways in terms of compound words (Sandra, 1990; Gagné et al., 2019; Libben, 2010) and form a continuum, ranging from fully transparent to fully opaque.

In English, for example, the compound word *snowball* is thought to be fully transparent as the meaning of *snowball* comes from both *snow* and *ball*. However, the compound *strawberry* is composed of one transparent constituent *berry* and one opaque constituent *straw,* while the compound *shindig* can be considered to be opaque because its meaning is not related to the meanings of the two constituents *shin* and *dig*. Therefore, an FCI behaves in ways like a compound word in terms of semantic transparency because they both have undergone some sort of lexicalization (Brinton & Traugott, 2005).

After examining the reasons which make the four [*bù* A *bù* B] idioms interchangeable, I also investigated the remaining non-interchangeable idioms (types) and enquired as to why they are non-interchangeable. Semantics and morphosyntax are the two factors with which to distinguish the 25 non-interchangeable [*bù* A *bù* B] idioms.

The first factor is related to semantic iconicity (see hypothesis 1), which is represented by both [*bù* A] and [*bù* B]. In cognitive semantics, iconicity refers to the similarity between a sign and its meaning. This underpins three principles – quantity principle, proximity principle, and sequential order principle (Croft, 1999). The quantity principle states that formal complexity corresponds to conceptual complexity,

the proximity principle that conceptual distance tends to match with linguistic distance, and the sequential order principle that the sequential order of events described is mirrored in the speech chain. These iconicity principles also apply to [*bù* A *bù* B] idioms. There are three different manifestations of the sequential order principle between the two units, that is top-down, small-to-large and sequential cause-effect relationships (see hypothesis 1).

Firstly, the top-down relationship can be illustrated by the idiom 不衫不履 (*bù-shān-bù-lǚ*, no-shirts-no-shoes, 'be not properly dressed as gentlemen should be'), which is a non-interchangeable idiom due to the compounding of 衫履 (*shān-lǚ*, shirts -shoes), not 履衫 (*lǚ-shān*, shoes-shirts). This implies that when Chinese people observe a person, they tend to profile him/her in a top-down manner, that is from 衫 (*shān*, shirts) to 履 (*lǚ*, shoes).

Secondly, the small-to-large order is reflected in the idiom 不三不四 (*bù-sān-bù-sì*, not-three-not-four, "dubious"), which cannot be re-organized as 不四不三 (*bù-sì-bù-sān*, not-four-not-three) due to the fact that numbers are traditionally counted from small to large. Therefore, when numbers 三 (*sān*, three) and 四 (*sì*, four) fit in the [*bù* A *bù* B] construction, they nonetheless comply with the small-to-large order.

Such compounds as 衫履 (*shān-lǚ*) and 三四 (*sān-sì*) indicate the top-down and the small-to-large relationships, which are characterised by the language iconicity (Engler, 1995; Croft, 2010; Downing & Stiebels, 2012; Dingemanse et al., 2015; Wu, 2017) shared among its speakers, reflecting ways in which people observe things in the real world.

Thirdly, the cause-effect sequential relationship can be illustrated by 不破不立 (*bù-pò-bù-lì*, not-eradicate-not-build). The idiom can be interpreted as "if there is no destruction, there can be no construction". Indeed, the first unit (不破, *bù-pò*, not-eradicate) is the cause of the second unit (不立, *bù-lì*, not-build). Thus, the first unit should take place before the second unit, which iconically dictates the word order. This semantic relation between these two units is a cause-effect one. Such cause-effect relationship can also be shown in the idioms 不塞不流 (*bù-sè-bù-liú*, not-clog-not-flow,

"if there is no damming, there is no flowing") and 不止不行 (*bù-zhǐ-bù-xíng*, not-stop-not-move, "if there is no rest, there is no motion").

Morphosyntactic order is the other factor that makes [*bù* A *bù* B] idioms non-interchangeable. This can be illustrated by the idiom 不清不白 (*bù-qīng-bù-bái*, not-clear-not-white, "unclear"). To explain, two characters that fit in A and B are by themselves a compound word, i.e., 清白 (*qīng-bái*, innocent); therefore, when 清白 (*qīng-bái*, innocent) is split and relocated to the positions of A and B, the morphemes of 清 (*qīng*, clear) and 白 (*bái*, white) retain the same collocational order that they have in the compound word. This entails that 清 (*qīng*) fits in the position of A, while 白 (*bái*) is relocated to the position of B (see 35).

(35) 没有　　　一些　不清不白　　　的　　事

méi-yǒu　　yīxiē bù-qīng-bù-bái　　de　　shì

not there are some not-clear-not-white PART things.

"There are no unclear things."

<div align="right">Haigong Little Red Robe Biography</div>

In (35), the idiom 不清不白 (*bù-qīng-bù-bái*) is used as an adjective to modify the noun (事, *shì*, "things"). Note that in CCL there is no occurrence of 不白不清 (*bù-bái-bù-qīng*, not-white-not-clear). Thus, the reason why 不清不白 (*bù-qīng-bù-bái*, not-clear-not-white) behaves as a non-interchangeable idiom is arguably due to the fact that 清白 (*qīng-bái*) is itself a compound word. When A and B are part of a single compound word, [*bù* A *bù* B] idioms are non-interchangeable since this construction follows the same morphosyntactic order of that compound word.

Similar compounds are 慌忙 (*huāng-máng*, hurried), 明白 (*míng-bái*, understand), 尴尬 (*gān-gà*, awkward), and so on. When these compounds fit in the [*bù* A *bù* B] construction, they also retain their morphosyntactic order. In other words, the first morpheme of each compound will be relocated to the position of A, while the second

morpheme to the position of B. Thus, the idioms 不慌不忙 (*bù-huāng-bù-máng*, "calm and unhurried"), 不明不白 (*bù-míng-bù-bái*, "dubious") and 不尴不尬 (*bù-gān-bù-gà*, "embarrassing") are accordingly classed as non-interchangeable constructions.

To sum up, when A and B are part of a compound word and there is iconicity involved in [*bù* A] and [*bù* B], the idiom [*bù* A *bù* B] cannot be re-constructed as [*bù* B *bù* A] (non-interchangeable).

However, there are some [*bù* A *bù* B] idioms that operate on the conditions as specified in my hypothesis (see 7.2), that is, these are supposed to be interchangeables, but are indeed non-interchangeable idioms. These could be seen as somewhat 'exceptional'. An example is 不伦不类 (*bù-lún-bù-lèi*, "neither fish nor fowl"), where the characters in the positions of A and B are 伦 (*lún*, type) and 类 (*lèi*, type). The two morphemes both mean "type", and 伦类 (*lún-lèi*) as a term is not a compound word itself and there are no iconicity relationships between 不伦 (*bù-lún*) and 不类 (*bù-lèi*). Based on my hypothesis, the idiom 不伦不类 (*bù-lún-bù-lèi*) should be interchangeable, however this idiom only operates non-interchangeably. That is to say, Chinese speakers only use the idiom 不伦不类 (*bù-lún-bù-lèi*) instead of 不类不伦 (*bù-lèi-bù-lún*). This can also be verified in CCL where no occurrence was retrieved for the latter (see 2).

(36) 怎 有　　许多　不伦不类　　　　的　　怪物

zěn yǒu　　xǔduō bù-lún-bù-lèi　　　de　　guàiwù

how there be many neither fish nor fowl PART monsters.

“How come there are so many nondescript monsters.”

The Eight Immortals

In (36), the idiom 不伦不类 (*bù-lún-bù-lèi*) is used as an adjective to modify the noun (怪物, *guài-wù*, "monsters"). Note that only 不伦不类 (*bù-lún-bù-lèi*) has been found, but there is no occurrence of 不类不伦 (*bù-lèi-bù-lún*). Thus, 不类不伦 (*bù-lèi-bù-lún*) is a non-interchangeable idiom as it is an exception.

In conclusion, semantic (iconicity) and morphosyntactic (compound words) considerations are key to account for the non-interchangeability of the [*bù* A *bù* B] construction, and some idioms (exceptions) that feed in my hypothesis remain non-interchangeables.

The type frequency of the [*bù* A *bù* B] construction in terms of interchangeability can be quantified by the Table 16.

[*bù* A *bù* B]	Factors	Type frequency	In total
Non-interchangeable	Semantic iconicity	8	22
	Morphosyntactic (compound words)	10	
	Exceptions	4	
Interchangeable	Not compound word and no iconicity	10	10

Table 16 Type frequency of the [*bù* A *bù* B] construction and their factors

In conclusion, there are 22 non-interchangeables and 10 interchangeables in the [*bù* A *bù* B] construction. The reasons why the 22 idioms are non-interchangeable are based on morphosyntactic (compound words) and semantic (iconicity) considerations (Bauer, 1992; Lehmann, 2002; Brinton & Traugott, 2005) despite some exceptions. From Table 16, it can be observed that there are 10 [*bù* A *bù* B] idioms (highest type frequency) that are non-interchangeables due to the morphosyntactic constraint, meaning that morphosyntax is the major factor that limits the interchangeability of the [*bù* A *bù* B] construction. Following this, semantic iconicity is the second factor which accounts for the non-interchangeability of the [*bù* A *bù* B] construction with 4 exceptions. These morphosyntactic and semantic considerations do not apply to the remaining 10 idioms; therefore, they are interchangeable.

7.3 Distributions of the [不 (*bù*, not) A 不 (*bù*, not) B] construction in terms of internal constituency and occurrences

This section discusses the internal constituency distributions of the [*bù* A *bù* B] construction, based on which seven different types of FCIs have been identified (cf. section 4.2). Also, it investigates the occurrence distributions of the [*bù* A *bù* B] idioms and how they are used in context. The seven types of 2+2 constructions are listed as follows:

2+2 construction types	[*bù* A *bù* B] construction
Type 1: [NP NP] construction	n = 10
Type 2: [AP AP] construction	n = 8
Type 3: [VP VP] construction	n = 11
Type 4: [N V] [N V] construction	N/A
Type 5: [V N] [V N] construction	N/A
Type 6: [NP VP] construction	N/A
Type 7: [VP NP] construction	N/A

Table 17 Seven types of FCIs and their application to [bù A bù B] construction

As discussed in chapter four, there are 7 different types of 2+2 constructions' internal constituency types. However, the specific construction [*bù* A *bù* B] does not operate on all the seven types identified; in fact, only [NP NP], [AP AP], and [VP VP] constructions (Types 1, 2, 3) are found in [*bù* A *bù* B], due to two main reasons. Firstly, [*bù* A *bù* B] is a symmetrical patterning, which means that it cannot be found in asymmetrical structures (Types 6 and 7). Secondly, the two units of the construction ([*bù* A] and [*bù* B]) cannot be categorised into [N V] or [V N] due to the fact that the character (不, *bù*, not) is a negator which cannot be characterized as a noun or a verb. Thus, only three out of seven types can be found in this [*bù* A *bù* B] construction.

Based on the classification of the 29 [*bù* A *bù* B] idioms, 10 [NP NP] constructions, 8 [AP AP] and 11 [VP VP] have been identified. An example of the [NP NP] construction is 不衫不履 (*bù-shān-bù-lǚ*, no-shirts-no-shoes, "be not properly dressed as gentlemen should be") where both 不衫 (*bù-shān*, no-shirts) and 不履 (*bù-lǚ*, no- shoes) are NP units. Similarly, the [AP AP] construction can be illustrated by the idiom 不卑不亢 (*bù-bēi-bù-kàng*, not-low self-esteem-not-arrogant) where the two units 不卑 (*bù-bēi*, not-low self-esteem) and 不亢 (*bù-kàng*, not-arrogant) are adjective phrases. Also, the [VP VP] construction can be shown in the idiom 不破不立 (*bù-pò-bù-lì*, not- eradicate-not-build, "if there is no destruction, there can be no construction") where 不破 (*bù-pò*, not-eradicate) and 不立 (*bù-lì*, not-build) are two verb phrases.

It can be found that the 29 [*bù* A *bù* B] idioms are nearly evenly distributed across the three patterns, and thus there is no salient tendency towards one particular type of internal constituency for the [*bù* A *bù* B] construction.

Aside from only looking at the internal constituency of [*bù* A *bù* B], I also further investigated the specific occurrences of the construction, as it is important to investigate how [*bù* A *bù* B] is used contextually. The 29 types have led to 472 occurrences. More specifically, 54 occurrences were found for 10 [NP NP] constructions, 170 occurrences for 8 [AP AP] constructions, and 248 occurrences for 11 [VP VP] constructions. The distributions of [*bù* A *bù* B] in terms of type and token frequencies can be shown in Table 18. Note that all the [*bù* A *bù* B] occurrences are retrieved from CCL in the Minguo period (1911-1949). The reason why I choose this period for my synchronic study is that it features a significant use of 白话文 (*bái-huà-wén*, vernacular Chinese) instead of 文言文 (*wén-yán-wén*, literary Chinese). Crucially, the language in this period is worth researching because it connects to both the old and new writing styles of the Chinese language.

[*bù* A *bù* B] idioms	Type frequency	Token frequency
[NP NP] construction	10	54
[AP AP] construction	8	170
[VP VP] construction	11	248

Table 18 Type and token frequencies of [*bù* A *bù* B] idioms

It can be found that the [VP VP] construction features the highest type frequency (N = 11), followed by the [NP NP] (N = 10) and [AP AP] (N = 8) constructions. The interesting finding is that, as argued by Berg (2014) and Traugott & Trousdale (2013), a higher type frequency construction does not necessarily yield a higher token frequency. For example, the type frequency of the [NP NP] construction (N = 10) is higher than that of the [AP AP] construction (N = 8), but the token frequency of the [NP NP] construction (N = 54) is lower than that of the [AP AP] construction (N = 170). To sum up, there is no salient tendency towards one specific type of internal constituency for the [*bù* A *bù* B] construction, and a higher type frequency may sometimes yield a lower token frequency in this construction. Among all the 29 [*bù* A *bù* B] idioms, there are 11 [VP VP] constructions, 10 [NP NP] constructions and 8 [AP AP] constructions.

Besides investigating the internal constituency distributions of the [*bù* A *bù* B] construction, I also further look at how those occurrences (tokens) are used in context. The [*bù* A *bù* B] idioms can serve the three propositional act functions (reference, predication, and modification) which are proposed by Croft's Radical Construction Grammar (2001).

(37) 一个　是　不衫不履

　　yī-gè　　shì　bù-shān-bù-lǚ

　　Someone be　no shirts no shoes

　　"Someone is not properly dressed as gentlemen should be."

<div align="right">Romance of the History of the Republic of China</div>

In (37), the idiom 不衫不履 (*bù-shān-bù-lǚn*) serves the object function as it follows the copula 是 (*shì*, is). When idioms act as an object, they serve the function of REFERENCE.

(38) 一切　　　外事,　　　　均　置之　不闻不问

yīqiè　　　wàishì　　　jūn　zhì-zhī bù-wén-bù-wèn

Everything external things, all　put　　not listen not ask

"Put all the external things aside, and show no interest in them."

<div align="right">History of the Republic of China</div>

The idiom 不闻不问 (*bù-wén-bù-wèn*) is used as the predicate in (38) as it means "show no interest in something". When idioms act as the predicate, they fulfil the function of PREDICATION.

(39) 这些　　不伦不类　　的　　畜生

zhè-xiē bù-lún-bù-lèi　de　　chùshēng

These　　no type no type PART bastards

"These nondescript bastards."

<div align="right">Court Yanshi of the Han Dynasty</div>

In (39), the idiom 不伦不类 (*bù-lún-bù-lèi*, no-type-no-type, "neither fish nor fowl") is used as the attributive adjective function to modify the noun 畜生 (*chù-shēng*, bastards). This idiom adds an additional feature to the referent 畜生 (*chù-shēng*). Thus, it can be viewed as the act of MODIFICATION.

The distributions of [*bù* A *bù* B]'s token frequencies and propositional act functions are quantified as in Table 19.

[*bù* A *bù* B] idioms	Propositional act functions			Token frequency
	modification	predication	reference	
[NP NP] construction	37	14	3	54
[AP AP] construction	147	19	4	170
[VP VP] construction	224	20	4	248
In total	408	53	11	472

Table 19 Token frequencies of [*bù* A *bù* B] idioms and their proposition act functions

It can be observed from Table 19 that [VP VP] is the most frequent pattern (N = 248), followed by [AP AP] (N = 170), and [NP NP] (N = 54). As mentioned in Section 3.3, I argued for a correspondence between internal constituency and usage in the Chinese idiomatic system. That is, [VP VP] is more often used for predication, [AP AP] for modification, and [NP NP] for reference. Therefore, predication could be reasonably expected to be the most frequently used propositional act function of [*bù* A *bù* B] idioms, modification the second, and reference the third. However, modification is indeed found to be the most frequently used function of [*bù* A *bù* B] idioms, although [VP VP] has the highest type frequency. This implies that there is a mismatch between propositional act functions and the internal constituencies of [*bù* A *bù* B] idioms. In fact, among the 472 occurrences, 408 have a modification function, 53 a predication function, and 11 a reference function. As shown in Table 19, there are two interesting shifts at play, which are:

(a) the (expected) reference function tends to shift to modification in [NP NP] idioms.

(b) the (expected) predication function tends to shift to modification in [VP VP] idioms.

More specifically, 37 out of 54 [NP NP] idioms are used for modification, whereas 224 out of 248 [VP VP] idioms are used for modification.

An example of function shift from reference to modification in [NP NP] idioms can be seen in (36). Although 不伦不类 (*bù-lún-bù-lèi*, nondescript) is an [NP NP] idiom and should be expected to be used more often for referencing (see 4.3), it acts for

modification in (36), because 不伦不类 (*bù-lún-bù-lèi*) modifies the following noun 畜生 (*chù-shēng*, bastards). Indeed, the [NP NP] idiom is composed of two [bù N] constructions (based on [*bù* + X]), that is, [[bù N] [bù N]]. Specifically, a [bù N] construction can acquire a modifying function when it is followed by the Chinese complementiser 的 / 之 (*de/zhī*). For example, 不法之徒 (*bù-fǎ-zhī-tú,* not-law-complementizer-person, someone who does not obey the law) is composed of a [bù N] construction and a noun (徒, *tú,* person), which is an instance of the Chinese [*bù* X *de* NP] construction. Crucially, the negator 不 (*bù,* not) does not prototypically occur before nouns in Chinese; instead, it is often followed by adjectives (漂亮, *piào-liang,* beautiful), adverbs (快, *kuài,* fast), or verbs (吃, *chī,* eat) — for instance, 她不漂亮 (*tā-bù-piào-liang,* She is not beautiful), 他跑得不快 (*tā-pǎo-dé-bù-kuài,* He does not run fast) or 我不吃面 (*wǒ-bù-chī-miàn,* I do not eat noodles). However, when the *bù* negator is followed by a noun, it can lead to a functional shift of the noun, for instance, from reference to modification. Therefore, the *bù* negator and the noun following it may chunk up as a complete unit (i.e. the [*bù* N] construction) distinctively behaving as a modifier. For example, when *bù* is followed by 法 (*fǎ,* law), the [*bù* N] construction (不法, *bù-fǎ,* not-law) tends to be used to modify another noun (徒, *tú,* person). Another way to put this is that [bù N] has the idiosyncratic property of modification as part of the [*bù* N de NP] construction, as seen in (36).

Simply put, the [*bù* N] construction coerces the noun to modify another noun instead of serving the reference function. Accordingly, the [*bù* N *bù* N] idioms are used as modifiers although [NP NP] idioms usually serve the reference function (see section 4.3). In fact, [*bù* N *bù* N] idioms being employed to modify a noun is a frequent phenomenon. There are two such collocations. For example, 不伦不类 (*bù-lún-bù-lèi,* no-type-no-type, "neither fish nor fowl") as an [NP NP] idiom is used to modify nouns such as 女子 (*nǚ-zǐ,* woman), 人 (*rén,* person), 事 (*shì,* things), and 话 (*huà,* words). On the other hand, 不三不四 (*bù-sān-bù-sì,* not-three-not-four, neither one thing nor the other, "dubious") is used to modify 女子 (*nǚ-zǐ,* woman), 面孔 (*miàn-kǒng,* faces), and 朋友 (*péng-yǒu,* friends). Such shifting from reference to modification functions in the [*bù*

N *bù* N] construction constitutes the first reason as to why modification is the most frequently used function that [*bù* A *bù* B] serves.

Secondly, the function shift from predication to modification in [VP VP] can also explain the mismatch between internal constituency and propositional act function. Just like the [*bù* N *bù* N] construction, the [*bù* V *bù* V] construction is often coerced to modify a noun. For example,

(40) 不屈不挠　　　　　之　精神

　　 bù-qū-bù-náo　　　　zhī　jīng-shén

　　 not-surrender-not-yield PART spirits

　　 "The indomitable spirit."

<div align="right">Romance of the Republic of China</div>

In (40), 不屈不挠 (*bù-qū-bù-náo*, not-surrender-not-yield, "indomitable") is a [VP VP] idiom which often functions as a predicate (see section 4.3). However, in (40) above it is indeed used to modify the noun (精神, *jīng-shén*, spirit). Both 屈 (*qū*, surrender) and 挠 (*náo*, yield) are verbs, and when they fit in [*bù* A *bù* B], the construction coerces these two verbs to modifying another noun in a same way it coerces nouns. As discussed previously, X is usually replaced with a noun in the [*bù* X *de* NP] construction, with [*bù* N *de*] partaking of the function of modification. My argument in this case is that when a verb fills in the X slot, the [*bù* V *de*] construction may also encode a propositional act function of modification based on structural analogy (constructional motivation) with the [*bù* N *de* NP] construction. This entails that the [*bù* V *bù* V *de*] construction, which is composed of the double [*bù* V] patterning and the *de* complementizer, may serve as modifier rather than a predicate, as so does the [*bù* N *bù* N *de*] construction. Thus, both [*bù* N *bù* N] and [*bù* V *bù* V] are capable of modifying a noun as attributive adjectives would prototypically do. Moreover, the [*bù* V *bù* V] construction can also be used to modify a verb, serving as an adverbial modifier. For example,

(41) 眼泪　便　不知不觉　　地 流了 下来

　　yǎn-lèi biàn bù-zhī-bù-jué de liúle xiàlá

　　tears then not-know-not-feel PART. flow down

　　"Tears are trickling down without me knowing it."

<div align="right">Ancient and Modern Love Sea</div>

　In (41), 不知不觉 (bù-zhī-bù-jué, *not-know-not-feel, without knowing*) is a [VP VP] idiom which is indeed used to modify the verb (流, *liú, flow*). Here, [*bù* V *bù* V] also has the property of modification. It can be used as attributive adjective as in (40), and also can be used as an adverbial modifier as in (41). Actually, 224 out of 248 [VP VP] idioms are used to modify a noun or a verb. The most frequent collocation is 不知不觉 (bù-zhī-bù-jué) used to modify other verbs such as 回 (*huí*, return), 落 (*luò*, fall) and 流 (*liú*, flow). Such shifting also makes modification the most frequently used propositional act function in [*bù* A *bù* B] idioms.

　To conclude, like other 2+2 FCIs, [*bù* A *bù* B] idioms can express different propositional act functions. There is a strong correspondence between propositional act functions and internal constituencies in 2+2 FCIs; however, there is a mismatch between propositional act functions and internal constituencies in the [*bù* A *bù* B] construction — that is, [VP VP] has the highest token frequency, while modification is its most frequently used propositional act function. The explanation for this phenomenon is that such functions may shift (i) from referencing to modification in the [[NP NP]] construction and (ii) from predication to modification in the [VP VP] construction. The "unexpected" mechanism behind this is the construction [*bù* X *de* NP], which causes [*bù* N] and [*bù* V] to behave like adjectival and adverbial modifiers. Thus, both [NP NP] and [VP VP] idioms do show a strong tendency towards modification.

7.4 Morphemes A and B in the [不 (*bù*, not) A 不 (*bù*, not) B] construction: relationships and constructional meanings

This section discusses the three relationships between morphemes A and B in the [*bù* A *bù* B] construction and its four different constructional meanings. Interestingly, the relationships between A and B comprise two different categories, i.e., antonyms and synonyms (Cruse, 1986). The category of antonyms can be sorted into two sub-types: complimentary antonyms and relational antonyms.

According to Cruse (1986, p. 198), "the essence of a pair of complementaries is that between them they exhaustively divide some conceptual domain into two mutually exclusive compartments, so that what does not fall into one of the compartments must necessarily fall into the other." For example, a student can either pass or fail in an exam, but not engage in both at the same time. In relational antonyms, A and B are relative concepts which "express a relationship between two entities by specifying the direction of one relative to the other along some axis" (Cruse, 1986, p.231). For instance, if object A is higher than object B, we can say "A is above B" or "B is below A" because we can take either A or B as the reference point. Thus, we can describe the relationship between objects A and B in two different ways where "above" and "below" are relational concepts as opposed to complementary antonyms (pass and fail). Another example of relational antonyms is student vs. teacher. If A is a student and B is a teacher, A could say "I am your student" or "You are my teacher" to B to indicate their relationship. Note that this relationship between A and B is rather relative than absolute as the student A may be C's teacher. In this scenario, A could say "I am your teacher" or "You are my student" to C to indicate the relationship. Thus, the difference between complimentary antonyms and relational antonyms is that the former are absolute concepts, while latter are relative concepts. This distinction also applies to the [*bù* A *bù* B] construction.

In some cases, A and B are in a relationship of complementarity. For example, the idiom 不死不活 (*bù-sǐ-bù-huó*, no-dead-no-alive, "neither dead nor alive") where the two

units 死 (*sǐ*, dead) and 活 (*huó*, alive) are complimentary antonyms. There is no possibility of a third term lying between 死 (*sǐ*, dead) and 活 (*huó*, alive). A person can only either be dead or alive. These two words are absolute concepts. In other cases, A and B are in a relationship of relativity. For instance, the idiom 不止不行 (*bù-zhǐ-bù-xíng*, not-stop-not-move, "if there is no rest, there is no motion"), where the two action verbs 止 (*zhǐ*, stop) and 行 (*xíng*, move) are relative concepts as either verb can be regarded as the reference point for the other.

Moreover, the relationship of synonyms can also be realized by A and B. For example, 不伦不类 (*bù-lún-bù-lèi*, no-type-no-type, "neither fish nor fowl"), where 伦 (*lún*, type) and 类 (*lèi*, type) are regarded as synonymous pairs. In conclusion, the relationship between A and B can vary from relational and complimentary antonyms to synonyms.

Asides from the three relationships between morphemes A and B, what is also worth investigating are the four different constructional meanings of [*bù* A *bù* B] as they are determined by the paradigmatic items (types of morphemes) that fit in the construction.

The first constructional meaning is "neither A nor B" which comprises double negatives, negating both A and B. For example, 不伦不类 (*bù-lún-bù-lèi,* no-type-no-type, "neither fish nor fowl") in (6) adds the additional negative feature (nondescript) to the referent 畜生 (*chùshēng*, bastards), which carries a negative semantic prosody due to such double negation. However, double negation does not always carry a negative semantic prosody. For example, 不屈不挠 (*bù-qū-bù-náo*, not-surrender-not-yield) is bound to the "neither A nor B" constructional meaning, but it carries a positive semantic prosody as in (42).

(42) 惟膺　　不 屈 不 挠

    wéiyīng bù-qū-bù-náo

    weiying not surrender not yield

    "Weiying is indomitable."

<div align="right">Later Han Dynasty Romance</div>

In (42), 不屈不挠 (*bù-qū-bù-náo*, indomitable) is used to add an additional positive feature to the subject (weiying, a person's name), which shows that "neither A nor B" can also be used to convey a positive semantic prosody. This is due to the fact that 屈 (*qū*, surrender) and 挠 (*náo*, yield) by themselves carry a negative frame or purport, and that when they fit in the double-negative construction (neither A nor B), the idiom is deprived of any negative reading. In other words, the semantic prosody cannot simply rely on the constructional meaning, but is determined by the semantics of the two elements (A and B) in the construction. Another pair of such negative words can be shown as 偏 (*piān*, be partial to) vs. 倚 (*yǐ*, be biased). Both of the two elements carry a negative frame, and when they fit in the "neither A nor B" construction, the idiom 不偏 不倚 (*bù-piān-bù-yǐ*, be-not-partial-to-and-be-not-biased, "show no partiality to either side") can also convey a positive semantic prosody.

In summary, the first constructional meaning "neither A nor B" contains double negatives, which more often aligns with a negative reading. However, it can also generate a positive semantic prosody when A and B are considered negative in meaning.

The second constructional meaning is "not AB" which only contains one negative. An instance is 不清不白 (*bù-qīng-bù-bái*, not-clear-not-white, "unclear") as seen in example 1. When A and B are by themselves a compound word, i.e., 清白 (*qīng-bái*, innocent) and 明白 (*míng-bái, understand*), the [*bù* A *bù* B] construction will produce the second constructional meaning, that is, "not AB".

The third constructional meaning of the [*bù* A *bù* B] construction is "if not A, then there is no B". A typical example is 不破不立 (*bù-pò-bù-lì*, not-eradicate-not-build). The

idiom can be interpreted as "if there is no destruction, there can be no construction" as in (43).

(43) 有　　句　话　　说　得　　好，叫作　　不破不立

Yǒu　　jù　huà　shuō dé　　hǎo, jiào-zuò bù-pò-bù-lì

There be one saying speak PART well, is called not eradicate not build

"Here is a good saying — No destruction, no construction."

<div align="right">Newspaper from Xinhua News Agency</div>

The interpretation of this idiom must not rely on the first and second constructional meanings because the idiom generates a different kind of meaning which is "if not A, then no B". When A and B are relational antonyms, i.e., 止 (*zhǐ*, stop) vs. 行 (*xíng*, move), and 破 (*pò*, eradicate) vs. 立 (*lì*, build), [*bù* A *bù* B] will coerce the third constructional meaning.

The fourth constructional meaning of [*bù* A *bù* B] construction is "A + B" which serves to affirm an event or situation. For example, the idiom 不尴不尬 (*bù-gān-bù-gà*) actually means 尴尬 (*gān-gà*, *embarrassing*) instead of 不尴尬 (*bù-gān-gà*, *not embarrassing*). The morphemes A and B (尴尬, *gān-gà*) are by themselves a compound word, but unlike 清白 (*qīng-bái*, innocent) or 明白 (*míng-bái*, understand), 不尴不尬 (*bù-gān-bù-gà*) is interpreted differently from the second construction meaning "not AB". The rationale behind this might be that the individual morphemes of 清 (*qīng*, clear), 明 (*míng*, bright), and 白 (*bái*, white) have obtained established semantic autonomy (meanings), while 尴 (*gān*) and 尬 (*gà*) have not. The four constructional meanings of [*bù* A *bù* B] can be summed up by Table 20.

Constructional meaning	Relationships between A and B	Negativity	Type frequency
neither A nor B	synonyms or complimentary antonyms	double negation	20
not AB	compound words and independent morphemes	single negation	8
if not A, then not B	relational antonyms	conditional	3
A+B	compound words, but not independent morphemes	affirmative	1
Total			32

Table 20 The four constructional meanings of [bù A bù B] and the relationships between morphemes A and B

In conclusion, internal constituency (A and B) influences the interpretation of the [*bù* A *bù* B] construction (see hypothesis 3), and further leads to different semantic prosodies. A detailed discussion of the mismatch between semantic prosody and internal constituency will be investigated in Section 6.4. There are four different constructional meanings of the [*bù* A *bù* B] idioms, that is (a) "neither A nor B" which is double negation, (b) "not AB" which is single negation, (c) "if not A, then not B" which signifies a conditional circumstance, and (d) "A+B" which is used to affirm an event. When morphemes A and B are synonyms or complimentary antonyms, the construction [*bù* A *bù* B] triggers the first constructional meaning (neither A nor B). When A and B constitute a compound word and when A and B are independent morphemes, [*bù* A *bù* B] triggers the second constructional meaning (not AB). When A and B are relational antonyms, the interpretation operates on the third constructional meaning (if not A, then not B). Lastly, when A and B constitute a compound word, but neither of them is an autonomous morpheme, the construction triggers the fourth constructional meaning (A+B).

It should also be noted that "neither A nor B" and "not AB" are the most frequent readings of [*bù* A *bù* B]. The construction's distributions largely depend on the nature

of the idioms' internal constituency, because different internal constituencies will trigger different constructional meanings. In table 20, it can be found that A and B are more likely found in the relationships of synonyms (e.g. 伦 (*lún*, type) vs. 类 (*lèi*, type)), complimentary antonyms (e.g. 死 (*sǐ*, dead) vs. 活 (*huó*, alive)), and compound words of independent morphemes (e.g. 明 (*míng*, bright) vs. 白 (*bái*, white)), while A and B are less likely found in the relationships of relational antonyms (e.g. 止 (*zhǐ*, stop) vs. 行 (*xíng*, move)) and compound words of dependent morphemes (e.g. 尴尬 (*gān-gà*, *embarrassing*). Thus, the first and second constructional meanings ("neither A nor B" and "not AB") will be triggered more frequently than the third and fourth constructional meanings ("if not A, then not B" and "A+B").

7.5 The relationship between semantic prosody and lexical distribution in [不 (bù, not) A 不 (bù, not) B]

This section discusses the mismatch between semantic prosody and the internal constituency of [*bù* A *bù* B]. Semantic prosody is "the spreading of connotational colouring beyond single word boundaries" (Partington 1998, p. 68). It refers to the situation where the usage of a word conveys a sense of attitudinal or pragmatic meaning. This means that a certain seemingly neutral word should be perceived in contexts to decide whether it is positive or negative rather than examining the word in isolation. This clearly also applies to the [*bù* A *bù* B] idioms. For example, 不言不语 (*bù-yán-bù-yǔ*, not-speak-not-say, "utter not a single word")) is a idiom which can carry a positive or negative semantic prosody based on different contexts in the following examples (43) and (44).

(44) 这个　不言不语,　　　干　活　顶　两个人　　　的　　小伙子

　　　zhè-ge bù-yán-bù-yǔ,　　gàn huó dǐng liǎng-gè-rén de　　xiǎo-huǒ-zi

　　　This　not-speak-not-say do work equal two people　PART young man.

　"This young man is energetic in working, and he does not talk too much nonsense."

<div style="text-align: right">Feng Deying's book Yingchunhua</div>

In (44), 不言不语 (*bù-yán-bù-yǔ*) is used to modify 小伙子 (*xiǎo-huǒ-zi*), adding the positive feature (not talking too much nonsense) to the character. However, this idiom sometimes can carry a negative semantic prosody in (45).

(45) 方晓东　　　　　知道　自己　　的　　眼睛　　失明　　以后,

　　　fāngxiǎodōng zhī-dào zì-jǐ　　de　　yǎn-jīng shī-míng yǐ-hòu,

　　　xiaodong fang knows himself PART eyes　　lose sight after,

　　就 不言不语,　　　　　不吃不喝

　　jiù bù-yán-bù-yǔ,　　　bù-chī-bù-hē

　　then not-speak-not-say, not eat not-drink

　"After knowing losing sight permanently, Xiaodong Fang does not talk to anybody and not
　　　eat at all."

<div style="text-align: right">Qiong Yao's book Smoke Lock Heavy Building</div>

In (45), 不言不语 (*bù-yán-bù-yǔ*) adds the negative semantic prosody (not willing to communicate with others) to the subject. Note that there are some occurrences where idioms cannot be identified with either a positive or negative semantic prosody as in (46).

(46) 他俩 不言不语　　　　地　　向　　连队　　　　走

tāliǎ bù-yán-bù-yǔ　　de　xiàng lián-duì　　　zǒu

they not-speak-not-say PART towards military camp walk

"They are walking towards the military camp without talking to each other."

<div align="right">Defend the city Yanan</div>

In (46), 不言不语 (*bù-yán-bù-yǔ*) does not convey a negative or positive semantic prosody as it only describes the manner (without talking to each other) as to how the subject (they) is walking. I would categorise this occurrence as undefinable, and label it as "neutral" for easy of reference.

In order to better investigate the mismatch between semantic prosody and the internal constituency of [*bù* A *bù* B], I fitted a conditional inference tree (Hothorn et al., 2006) by using the data from Table 19.
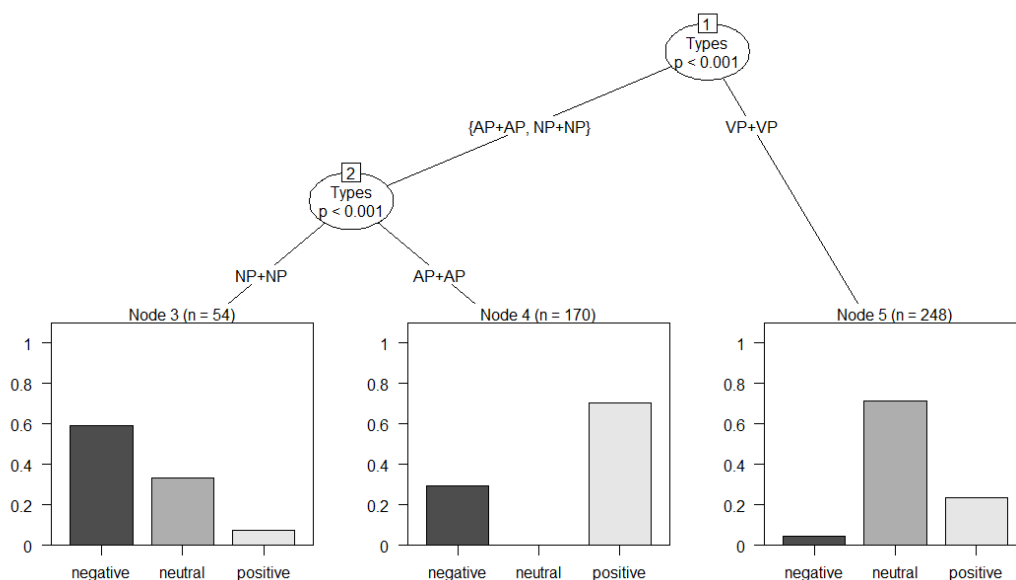

Figure 18 The mismatch between semantic prosody and the internal constituency of [*bù* A *bù* B]

It is clear that the most important predictor of propositional functions of [*bù* A *bù* B] is "type" (internal constituency), as it appears at the top of the conditional inference

tree. That is, in terms of [*bù* A *bù* B]'s functions in a given context, internal constituency is the most decisive variable.

The first illocutional concurrence (IC) worth noting runs from nodes 1 to 2 and then to 4. There are two interesting findings to be noted: there are no "neutral" occurrences for [AP AP] idioms and that [AP AP] idioms are used to convey more positive semantic prosody (70%). Firstly, the zero "neutral" occurrences can be explained in terms of "salience" (Giora, 2003). When polysemy is at play, salience has to do with some meaning being distinctively prominent and noticeable. Giora (2003) argues that contexts may affect the comprehension and production of language, but cognitively prominent salient meaning plays an even more important role for semantic disambiguation. For example, 不慌不忙 (*bù-huāng-bù-máng*, "calm and unhurried") is used to describe someone who is calm when dealing with tough situations, which serves to express a positive meaning, while 不清不白 (*bù-qīng-bù-bái*, not-clear-not-white, "unclear") obscure things that usually carry a negative meaning. These two [AP AP] idioms encode a strong context-independent meaning. Regardless of the contexts, they almost never express a "neutral" semantic prosody. Secondly, the lopsided distributions between positive and negative semantic prosodies of the [AP AP] construction (see nodes from 1 to 2, and to 4) are due to the fact that there are more positive [AP AP] idioms than negative ones in terms of salient meaning. In conclusion, the semantic prosody distributions of [AP AP] are less dependent on context than on the (default) salient meaning of the idioms.

However, [VP VP] idioms show a different distribution (nodes 1 to 5). The "neutral" occurrences take up the largest percentage among the three semantic prosodies. As discussed in section 6.3, [VP VP] idioms may transform to modify a noun or a verb, whose most frequent collocation is 不知不觉 (bù-zhī-bù-jué, unintentionally), which modifies other verbs such as 回 (*huí*, return), 落 (*luò*, fall) and 流 (*liú*, flow). Such expressions only describe the manner of action (i.e., lexical aspect or Aktionsart, see Vendler, 1967; Comrie, 1976; Moens & Steedman, 1988) without indicating a speaker's positive or negative evaluative attitude towards someone or something. This means that

when 不知不觉 (bù-zhī-bù-jué) is used for modification, the construction's semantic prosody is almost impossible to identify from its context although semantic prosody itself should emerge contextually. This shows that 不知不觉 (bù-zhī-bù-jué) is less context-dependent as the idiom means "unintentionally" — a word difficult for assessment of the semantic prosody. This widely used idiom accounts for the greater number of "neutral" [VP VP] occurrences.

The last IC that is worth noting runs from nodes 1 to 2 and then to 3. The semantic prosody distribution of [NP NP] is different from those of [VP VP] and [AP AP] as the negative semantic prosody takes up the largest percentage in [NP NP] idioms, which is largely determined by the construction's internal constituents' meaning. More specifically, in [*bù* A *bù* B], A and B are more likely to be found as synonyms (伦 (*lún*, type) vs. 类 (*lèi*, type)). Sections 7.3 and 7.4 have shown that when morphemes A and B are synonyms or complimentary antonyms, the [*bù* A *bù* B] construction may trigger the first constructional meaning (neither A nor B). This constructional meaning (double negatives) more often contributes to a negative reading, which accounts for the largest percentage of negative semantic prosody in [NP NP] idioms. To sum up, the internal constituents' semantics of [NP NP] affects its constructional meaning, and in turn, this constructional meaning affects the semantic prosody of [NP NP] as a whole.

In conclusion, different types of [*bù* A *bù* B] idioms show different semantic prosody distributions. The semantic prosody distribution of [AP AP] is determined by the (default) salient meanings of the idioms, while [VP VP] does not necessarily convey a semantic prosody as it is usually employed for describing manners of action without referencing a speaker's positive or negative evaluative attitude towards someone or something. In addition, the semantic prosody of [NP NP] as a whole is determined by the semantics of its internal constituent.

7.6 Summary

This chapter discusses the partly schematic 2+2 construction [不 (*bù*, not) A 不 (*bù*, not) B] from a synchronic perspective. There are more non-interchangeables than interchangeables in the [*bù* A *bù* B] construction. Morphosyntax (compound words) and semantic iconicity are two key factors that constrain the interchangeability of the [*bù* A *bù* B] construction. There are three iconicity relationships between [*bù* A] and [*bù* B], that is, top-down, small-to-large and cause-effect sequential relationships.

The [*bù* A *bù* B] construction is unique for three reasons. Firstly, unlike regular 2+2 idioms comprising all 7 different types — Type 1 [NP NP], Type 2 [AP AP], Type 3 [VP VP], Type 4 [N V] [N V], Type 5 [V N] [V N], Type 6 [NP VP], and Type 7 [[VP NP]], the [*bù* A *bù* B] construction features only Types 1, 2 and 3 due to the facts that [*bù* A *bù* B] is a symmetrical patterning and that neither [*bù* A] or [*bù* B] can be categorised into [N V] or [V N]. Secondly, it has been found that the 29 [*bù* A *bù* B] idioms are nearly evenly distributed across the three types, while there appears to be a salient distributional tendency towards one particular type of the regular 2+2 constructions. Thirdly, there is a correspondence between propositional act functions and the internal constituencies of regular 2+2 idioms, while there is a distributional mismatch between propositional act functions and the internal constituencies of the [*bù* A *bù* B] construction.

The relationships between A and B can be specified as synonyms, complimentary antonyms, relational antonyms, and compound words (either independent morphemes or non-independent morphemes). These different relationships will trigger four different constructional meanings which are (a) "neither A nor B", (b) "not AB", (c) "if not A, then not B", and (d) "A+B".

Different types of [bù A bù B] idioms display different semantic prosody patterns due to internal constituency, semantic meaning, and constructional meaning of the idioms. The delimitation of this study is that I only investigate [*bù* A *bù* B] from a synchronic way. Future studies may focus on the exploration of [bù A bù B] idioms

from a diachronic perspective so as to find out which constructional meaning emerges first. Also, it is of importance to investigate the schematicity and productivity of [*bù* A *bù* B] idioms, which reflects the mental representation of language processing.

Chapter 8 Conclusion

8.1 Summary of the findings

The present research is centred on a common but scarcely explored language phenomenon: the usage of Chinese FCIs. The findings of this work constitute a substantial contribution to the better understanding of Chinese FCIs as well as the applications of Construction Grammar to Chinese FCIs.

In Chapter 2, I argued that Chinese idioms can be regarded as constructions as they satisfy the Goldbergian definition of constructions as holistic pairings of form and meaning (1995; 2006). Elements of Croft's (2001) and Langacker's (2007) approaches were also included in the present work for inspecting Chinese FCIs. This allowed me to identify the similarities and differences between general idioms and Chinese idioms after looking at the patterns and classifications of Chinese FCIs. This thesis made the case that construction grammars can be applied to the study of Chinese idioms, and suggested that syntax, semantics, and pragmatics should all be taken into account to understand the internal constituency and usage of Chinese FCIs.

Based on this idea, in Chapter 3 I adopted a corpus-based approach to study the differences between interchangeables and non-interchangeables. Usually, idioms with symmetrical structure tend to be interchangeables, while idioms with asymmetrical structure tend to be non-interchangeables. However, iconicity strongly constrains the interchangeability of Chinese FCIs. This leads to the result that some 2+2 FCIs have a symmetrical structure, but they operate in fact as non-interchangeable FICs. The token frequencies of interchangeables and non-interchangeables have reached the highest point in Early Mandarin (1250 — 1800) before a clear decrease in Modern Mandarin (1800 — present). The trend can be explained by two socio-cultural factors: the development of literary works in Early Mandarin and the May Fourth Movement in Modern Mandarin, respectively. Interchangeables increase in productivity and schematicity in Early Mandarin (1250 — 1800) because of the new BBAA schema, while the token frequencies of the interchangeables decrease in Modern Mandarin

(1800 — present) due to the fact that the token frequencies of AABB and BBAA decreased when they were competing. Different types of FCIs serve different propositional act functions (Croft, 2001). More specifically, Type 1 FCIs tend to combine with the reference function, Types 3, 5, 6, and 7 correlate with the predicate function, and Types 2 and 4 the modification function. This means that the differences in actual language use of these seven types are determined by the internal constituency of the FCIs. What's more, the different internal constituencies FCIs affect the constructional changes of interchangeables and non-interchangeables, with the former showing a linear pathway of change, while the latter showed a non-linear pathway of change.

In Chapter 4, I have argued that AABB and BBAA may undergo attraction, differentiation, and substitution. Function and token frequency are the two indicators showing the kind of mechanism applied to distinct types of interchangeables. More specifically, Type 1 ([NP NP]), Type 2 ([AP AP]), Type 4 ([[N V] [N V]]), and Type 5 ([[V N] [V N]]) interchangeables undergo attraction; Types 3 ([VP VP]) interchangeables undergo substitution; Type 6 ([NP VP]) and Type 7 ([VP NP]) interchangeables undergo differentiation. Their different changes may be determined by their internal constituency (see 5.6).

In Chapter 7, I have provided evidence to show that the [*bù* A *bù* B] construction only features Type 1 [NP NP], Type 2 [AP AP], and Type 3 [VP VP] structures due to its symmetrical patterning and neither [*bù* A] nor [*bù* B] can be categorised into [N V] or [V N]. The [*bù* X *de* NP] construction causes the [*bù* N] and [*bù* V] constituents to behave like adjectival and adverbial modifiers. Moreover, the internal constituency of such components influences the interpretation of the [*bù* A *bù* B] construction and leads to differences in semantic prosody. In fact, different semantic relationships between A and B may trigger different constructional meanings; semantic prosody largely depends on the internal constituency of [*bù* A *bù* B]. More specifically, when morphemes A and B are synonyms or complimentary antonyms, [*bù* A *bù* B] means "neither A nor B". When A and B constitute a compound word and when A and B are independent

morphemes, [*bù* A *bù* B] denotes "not AB". When A and B are relational antonyms, [*bù* A *bù* B] signifies that "if not A, then not B". Lastly, when A and B are combined as a compound word, but neither of them is an autonomous morpheme, [*bù* A *bù* B] refers to "AB". The semantic prosody of [AP AP] is determined by the salient meanings of the idioms, while [VP VP] does not necessarily convey a positive nor a negative semantic prosody as it is usually employed for describing manners of action without referring a speaker's positive or negative evaluative attitude towards someone or something. In addition, the semantic prosody of [NP NP] as a whole is determined by the semantics of its internal constituency. Simply put, internal constituency affects the constructional meaning of the idiom, which in turn affects its semantic prosody.

8.2 Limitations

This research sheds new light on the realization of Chinese idioms based on construction grammar. There are three points that I need to clarify in this research.

The first issue regards the internal constituency of FCIs. As found in Section 2.3.3, there are seven different types of Chinese FCIs, namely, Type 1 ([NP NP]), Type 2 ([AP AP]), Type 3 ([VP VP]), Type 4 ([[N V] [N V]]), Type 5 ([[V N] [V N]]), Type 6 [NP VP], and Type 7 ([VP NP]). However, there remain a few FCIs that do not operate on these seven types. For example, the idiom 长此以往 (*cháng-cǐ-yǐ-wǎng*, long-this-by-past, "if things go on like this") which is consisted of 长此 (*cháng-cǐ*, all the time) and 以往 (*yǐ-wǎng*, previously) cannot categorized into any of the seven types. Its internal constituency is more complicated than and different from those found in this study. Note that, FCIs that do not fall in these seven types are rare cases.

Secondly, there might be a difference between the Xinhua Dictionary definition, corpus instances, and a speaker's usage in terms of interchangeability. For example, some idioms defined by the Xinhua Dictionary as an interchangeable type are acceptable to some native speakers, but are not found in the corpus. For example, 腹诽心谤 (*fù-fěi-xīn-bàng*, belly-defame-heart-slander, "silent curse or disagreement") can

alternate with 心谤腹诽 (*xīn-bàng-fù-fěi*, heart-slander-belly-defame, "silent curse or disagreement") according to the Xinhua Dictionary, but only the former can be found in the corpora while the latter yielded no results from either CCL or BCC. Also, there exist some interchangeable idioms acceptable to native speakers and found in the corpora, but they are not defined as interchangeable idioms by the Xinhua Dictionary, for instance, 甜言蜜语 (*tián-yán-mì-yǔ*, sweet-talk-sugar-word, "hypocritical flattery") and 蜜语甜言 (*mì-yǔ-tián-yán*, sugar-word-sweet-talk, "hypocritical flattery"). The two forms, indeed, are considered acceptable to native speakers and can be found in the corpora; however, the Xinhua Dictionary does not treat them in the same way. In addition, some idioms are defined as interchangeables in the Xinhua Dictionary supported by corpus findings, but today's Chinese speakers only use one form but not the other. For example, 堂堂正正 (*táng-táng-zhèng-zhèng*, fair-fair-square-square, "displaying strength and discipline") is used in today's Chinese speech community, but 正正堂堂 (*zhèng-zhèng-táng-táng*, square-square-fair-fair, "displaying strength and discipline") is not. This means that there exists variance among the Xinhua Dictionary, corpus instances, and a speaker's usage in terms of interchangeablity. The study capitalizes on a quantitative study using different corpora (BCC, CCL, and zhTenTen); however, although the usage-based dimensions are brought into consideration, no actual qualitative research has been carried out to substantiate the claims made based on aforementioned corpus findings. In order to examine the acceptability of the Chinese FCIs, interviews and questionnaires may be included in methodology.

Lastly, although I found that iconicity may limit the interchangeability of FCIs (e.g., top-down, small-to-large, and sequential cause-effect relationships), no analytical account was provided regarding the number of iconicity rules there could be because the (non)-interchangeability of FCIs is much more sophisticated (i.e., socio-cultural factors, collocationality in certain FCIs).

8.3 Future research

There is still much to be discovered regarding some of the issues I mentioned in Section 7.2. Firstly, future research can take into account 1+3, 3+1, and 1+1+1+1 constructions because they constituent the remaining 4% of all Chinese FCIs not investigated in this research.

Secondly, interviews or questionnaires may be conducted to help provide extra insight into how a speaker evaluates, believes or predicts when making use of a 2+2 construction. By doing qualitative surveys, the researcher will be able to examine the similarities and differences between a speaker's usage, corpus instances, and the examples in the Xinhua Dictionary in terms of interchangeablity and token frequency.

Thirdly, future research can inquire into other constructions that are partly schematic (as opposed to [*bù* A *bù* B] construction) in order to find out how each 2+2 construction may affect its propositional act functions and semantic prosody.

To sum up, suggestions for further research may include: (1) 1+3, 3+1, and 1+1+1+1 constructions, (2) interviews and questionnaires involving native speakers so as to see how idioms can be used differently, and (3) other constructions that are partly schematic.

References

Aaron, J. E. (2016). The road already traveled: Constructional analogy in lexico-syntactic change. *Studies in Language, 40*(1), 26–62.

Anttila, R. (2003). Analogy: The warp and woof of cognition. In Brian D. Joseph & Richard D. Janda (Eds.), *The handbook of historical linguistics* (pp. 425–440). Oxford: Blackwell.

Auch, L., Gagné, C. L., & Spalding, T. L. (2020). Conceptualizing semantic transparency: A systematic analysis of semantic transparency measures in English Compound words. *Methods in Psychology*, *3*, 100030.

Bauer, L. (1992). Lexicalization and level ordering. *Linguistics, 30,* 561–568.

Berg, T. (2014). Competition as a unifying concept for the study of language. *The Mental Lexicon 9*(2), *338–370.*

Bolinger, D. (1977). *Meaning and form.* London: Longman.

Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press.

Burton-Roberts, N. (2016). *Analysing sentences: An introduction to English syntax.* Abingdon: Routledge.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins Publishing.

Bybee, Joan L. and James L. McClelland. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *In Nancy A. Ritter, The Role of Linguistics in Cognitive Science. Special Issue of The Linguistic Review*, *22*, 381–410.

Cacoullos, R. T., & Walker, J. A. (2009). The present of the English future: Grammatical variation and collocations in discourse. *Language 85*(2), 321–354.

Cappelle, Bert. (2006). Particle placement and the case for "allostructions". *Constructions*, 1-28.

Cartoni, B. (2006). Constance et variabilité de l'incomplétude lexicale. *Proceedings*

of TALN/RECITAL, *Leuven, Belgium*, 661–669.

Chen, K.Y (2001). *A Construction Grammar Approach to Chinese Four-Character Idiomatic Expressions*. M.A. Thesis in Linguistics, Graduate Institute of Linguistics, National Taiwan University.

Chang, K. I. S., & Owen, S. (Eds.). (2010). *The Cambridge history of Chinese literature (Vol. 1)*. Cambridge: Cambridge University Press.

Corcoran, J. (2006). Schemata: the concept of schema in the history of logic. *Bulletin of Symbolic Logic*, *12*(2), 219–240.

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of pragmatics, 37*(10), 1510-1536.

Croft, William. 1998. Linguistic evidence and mental representations. *Cognitive Linguistics 9*(2):151-173

Croft, W. (2000). *Explaining language change: An evolutionary approach*. London: Longman.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Croft, W. (2010). Some contributions of typology to cognitive linguistics and vice versa. In T. Janssen & G. Redeker (Eds.), *Cognitive Linguistics* (pp. 61–94). Berlin: De Gruyter Mouton.

Cruse, A. (1986). Lexical semantics. Cambridge: Cambridge University Press.

Cruse, A. (2011). *Meaning in language: an introduction to semantics and pragmatics*. Oxford: Oxford University Press.

De Smet, H. (2008). Functional motivations in the development of nominal and verbal g erunds in Middle and Early Modern English1. *English Language & Linguistics*, *12*(1), 55–102.

De Smet, H., D'hoedt, F., Fonteyn, L., & Van Goethem, K. (2018). The changing functions of competing   forms: Attraction and differentiation. *Cognitive Linguistics*, *29*(2), 197–234.

Diesendruck, G. (2012). The scope and origins of children's assumption of conventionality. *Access to language and cognitive development* (pp. 116–134). Oxford: Oxford University Press.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: an empirical examination. Developmental Psychology, 41(3), 451–463.

Ding, S.S., & Lv, S.X. (1956). 现代汉语词典 （Xiandai Dictionary of Chinese Idioms). Beijing: The Commercial Press.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603–615.

Downing, L. J., & Stiebels, B. (2012). Iconicity. *The morphology and phonology of exponence*, 379–426.

Dressler, Wolfgang U. (1999). "On a semiotic theory of preferences in language". *The Peirce Seminar Papers: Essays in Semiotic Analysis* ed. by Michael Shapiro & Michael Haley, 389–415. New York: Berghahn.

Eitelmann, M. (2016). Support for end-weight as a determinant of linguistic variation and change. *English Language and Linguistics, 20*(3), 395–420

Engler, R. (1995). Iconicity and/or Arbitrariness. *Amsterdam studies in the theory and history of linguistic science series* 4, 39–40.

Fillmore, Charles J. (1985). Syntactic Intrusions and the Notion of Grammatical Construction. *Berkeley Linguistic Society*, 11, 73–86.

Fillmore, Charles J. (1988). The Mechanisms of 'Construction Grammar'. *Berkeley Linguistic Society*, 14, 35–55.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of let alone. *Language*, 501–538.

Fischer, O. (2007). *Morphosyntactic change: Functional and formal perspectives* (Vol. 2). Oxford: Oxford University Press

Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). LADEC: the large database of English compounds. *Behavior research methods*, 51(5), 2152–2179.

Goldberg, A. E. (1995). Constructions: *A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.

Givón, T. (1991). Isomorphism in the grammatical code: cognitive and biological considerations. *Studies in language, 15*(1), 85–114.

Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, 515–540.

Haspelmath, Martin. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18, 180–205.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hilpert, M. (2015). From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics*, 26(1), 113–147.

Hoffmann, T., & Trousdale, G. (2013). *The Oxford handbook of construction grammar.* Oxford: Oxford University Press.

Hofstadter, Douglas R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid.* 1st ed. New York: Basic Books.

Hollmann, W. B. (2012). Word classes: Towards a more comprehensive usage-based account. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *36*(3), 671-698.

Hollmann, W. B. (2013). Nouns and verbs in Cognitive Grammar: Where is the 'sound' evidence?. *Cognitive Linguistics*, *24*(2), 275-308.

Hopper, P. J., & Traugott, E. C. (2003). Grammaticalization. Cambridge: Cambridge University Press.

Itkonen, E. (2005). *Analogy as structure and process: Approaches in linguistics,*

cognitive psychology and philosophy of science (Vol. 14). Amsterdam: John Benjamins Publishing.

Jacobson, O. 2014. *Compositional semantics: an introduction to the syntax/semantics interface*. Oxford: Oxford University Press.

Jaszczolt, K.M. 2007. *Default semantics: foundations of a compositional theory of acts of communication*. Oxford: Oxford University Press.

Jiang, Z., Zhao, D., Zheng, J., & Chen, Y. (2020). A Study on Differences between Simplified and Traditional Chinese Based on Complex Network Analysis of the Word Co-Occurrence Networks. *Computational Intelligence and Neuroscience*, 2020.

Kay, P. (2001). *Pragmatic aspects of grammatical constructions*. Retrieved from https://www1.icsi.berkeley.edu/~kay/cg.prag.pdf.

Keller, Rudi. (1990). Sprachwandel von der unsichtbaren Hand in der Sprache. Tübingen: Francke.

Kong, Y. (2014). The study of English and Chinese numerical idioms and their translation. *Journal of Language Teaching and Research*, 5(2), 446-451.

Körtvélyessy L., Štekauer P., Zimmermann J. (2015). Word-Formation Strategies: Semantic Transparency vs. Formal Economy. *In Semantics of complex words* (pp. 85–113). Springer, Cham.

Kracht, M. (2014). *Interpreted languages and compositionality*. Dordrecht: Springer.

Kuryłowicz, Jerzy. (1949). "La notion de l'isomorphisme". *Travaux du Cercle Linguistique de Copenhague 5*(1), 48–60.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press

Labov, W. (2011). *Principles of linguistic change, volume 2: Social factors*. Oxford: Wiley-Blackwell

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago and London: The University of Chicago press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Descriptive*

applications. Stanford: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar: a basic introduction*. Oxford: Oxford University Press.

Langlotz, A. (2006). *Idiomatic creativity: a cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: John Benjamins.

Langlais, P., Gotti, F., & Cao, G. (2005). Nukti: English-inuktitut word alignment system description. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, 75–78.

Lardilleux A., Lepage Y. (2009). *Hapax Legomena*: Their Contribution in Number and Efficiency to Word Alignment (pp.440-450). In: Vetulani Z., Uszkoreit H. (eds) Human Language Technology. Challenges of the Information Society. LTC 2007. Subseries   Lecture Notes in Computer Science, vol 5603. Springer, Berlin, Heidelberg.

Lehmann, C. (2002). *New reflections on grammaticalization and lexicalization*. In Wischer and Diewald (Eds), 1–18.

Libben, Gary. (2010). Compound words, semantic transparency, and morphological transcendence. In Olson, S. (Eds), *New impulses in word-formation* (Linguistische Berichte Sonderheft 17), 212–232. Hamburg: Buske.

Luo, J. (2015). On the Interaction Between the Structure and the Prosody of Chinese 1+ 3 Idioms. In *Chinese Lexical Semantics*: 16th Workshop, CLSW 2015, Beijing, China, May 9-11, 2015, Revised Selected Papers 16 (pp. 340-351). Springer International Publishing.

MacWhinney, Brian. (2014). Conclusions: Competition across time. In Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (Eds.), *Competing motivations in grammar and usage*, 364–386. New York: Oxford University Press.

Makkai, A. (1972). *Idiom Structure in English*. Berlin: De Gruyter Mouton.

Martinet, André. (1957). "Arbitraire linguistique et double articulation". *Cahiers Ferdinand de Saussure* 15, 105–116.

Matthews, P. H. (2014). *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.

McGlone, M. S. (1996). Conceptual metaphors and figurative language interpretation: Food for thought?. *Journal of memory and language, 35*(4), 544-565

Menard, S. (2002). *Applied logistic regression analysis (Vol. 106).* Sage.

Meng, P. (2022). A Language and Culture Comparative Study of Three Primary Colors in English and Chinese Idioms and Usual Expressions from the Perspective of Metaphorical Cognition. *International Conference on Education, Language and Art (ICELA 2021)* (pp. 241-245). Atlantis Press.

Michaelis, Laura A. (1994). A Case of Constructional Polysemy in Latin. *Studies in Language*, 18, 45–70.

Michaelis, Laura A., and Lambrecht, Knud (1996). Toward a Construction-Based Model of Language Function: The Case of Nominal Extraposition. *Language*, 72, 215–247.

Michaelis, L. A. (2013). Sign-based construction grammar. *The Oxford handbook of construction grammar*, 133–152.

Nall, T. M. (2008). *Analysis of Chinese Four-character Idioms Containing Numbers: structural patterns and cultural significance*. Unpublished PhD thesis. Ball State University

Nesset, T. & Janda, L. (2023). A network of allostructions: quantified subject constructions in Russian. *Cognitive Linguistics, 34*(1), 67-97.

Nørgård-Sørensen, Jens, Lars Heltoft and Lene Schøsler. (2011). *Connecting Grammaticalisation. The Role of Paradigmatic Structure.* Amsterdam: Benjamins.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language 70*(3): 491-538.

Pan Y and Kádár DZ. (2011). *Politeness in Historical and Contemporary Chinese*. London: A&C Black.

Partington, A. 1998. *Patterns and meanings*. Amsterdam and Philadelphia: Benjamins.

Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi*, 13(1), 11–24.

Perek, F. (2018). Recent change in the productivity and schematicity of the way-construction: a distributional semantic analysis. *Corpus Linguistics and Linguistic Theory*, *14*(1), 65–97

Petré, P. (2014). *Constructions and environments: Copular, passive, and related constructions in Old and Middle English*. Oxford: Oxford University Press.

Pierrehumbert, J., & Granell, R. (2018). On hapax legomena and morphological productivity. In *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 125–130.

Pulleyblank EG. (1995). *Outline of Classical Chinese Grammar.* Vancouver: University of British Columbia Press.

Robinson, C. (2014). Knowing linguistic conventions. *South African Journal of Philosophy*, *33*(2).

Sag, Ivan A. (2012). Sign-Based Construction Grammar: An informal synopsis, in H. C. Boas and I. A. Sag, (eds), *Sign-Based Construction Grammar*. Stanford: CSLI Publications). 69–202

Sandra, Dominiek. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology* 42A, 529–567.

Schäfer, M. (2018). *The semantic transparency of English compound nouns.* Berlin: Language Science Press.

Schrader, B. (2006). How does morphological complexity translate? A cross-linguistic case study for word alignment. In *Proceedings of Linguistic Evidence Conference*, 189–191.

Scott, J. C., & Henderson, A. M. E. (2013). Language matters: Thirteen-month-olds understand that the language a speaker uses constrains conventionality. *Developmental Psychology*, 49(11), 2102–2111.

Shu, Xin-cheng (1936). 辞海（*Cihai Dictionary of the Chinese Language*). Beijing:

ZhongHua Book Company.

Simone, R. (Ed.). (1995). *Iconicity in language* (Vol. 110). John Benjamins Publishing.

Sommerer, L., & Smirnova, E. (Eds.). (2020). *Nodes and networks in diachronic construction grammar* (Vol. 27). John Benjamins Publishing Company.

Steels, L. (2011). Introducing Fluid Construction Grammar. In L. Steels (Ed.) *Design patterns in Fluid Construction Grammar* (pp. 3–30). Amsterdam: John Benjamins

Su, I-Wen. 2002. Why a Construction—That is the Question! *Studies in English Literature and Linguistics,* 28(2):27–42.

Sullivan, K. 2013. *Frames and constructions in metaphoric language*. Amsterdam: John Benjamins.

Tantucci, V., & Wang, A. (2018). Illocutional concurrences: The case of evaluative speech acts and face-work in spoken Mandarin and American English. *Journal of Pragmatics*, *138*, 60–76.

Tantucci, V., & Di Cristofaro, M. (2019). Entrenchment inhibition: Constructional change and repetitive behaviour can be in competition with large-scale "recompositional" creativity. *Corpus Linguistics and Linguistic Theory*, 1 (ahead-of-print).

Tantucci, V., & Wang, A. (2019). Diachronic change of rapport orientation and sentence-periphery in Mandarin. *Discourse Studies*, *22*(2), 146–173.

Tantucci, V., & Di Cristofaro, M. (2020). Entrenchment inhibition: Constructional change and repetitive behaviour can be in competition with large-scale "recompositional" creativity. *Corpus Linguistics and Linguistic Theory*, *16*(3), 547–579.

Tauli, V. (1945). Morphological analysis and synthesis. *Acta Linguistica*, *5* (1), 80–85.

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes (Vol. 6).* Oxford: Oxford University Press.

Traugott, E. C. (2020). The intertwining of differentiation and attraction as exemplified by the history of recipient transfer and benefactive alternations. *Cognitive Linguistics*, 1 (ahead-of-print).

Van de Velde, F. (2014). 6. Degeneracy: The maintenance of constructional networks. In *Extending the scope of construction grammar* (pp. 141–180). Berlin: De Gruyter Mouton.

Vicentini, A. (2003). The Economy Principle in language. Notes and Observations from   Early Modern English Grammars. *Mots, Palabras, Words*, *3*, 37–57.

Wang DD-W. (2010). Chinese literature from 1841 to 1937. In: Chang KS and Owen S (Eds), *The Cambridge History of Chinese Literature.* Cambridge: Cambridge University Press, pp. 413–564.

Wang, Lei, Yu, Shi-wen, Zhu, Xue-feng, Leo, Feng-shu, Shagang, Hezi, & Jiang, Bing-gui. (2013). Principles for and New Developments of Constructing the Base of Chinese Idiom Knowledge [汉语成语知识库的建构理念与新展].

Conference paper at The Fourth International Conference on the Digitalization of Chinese Classics. Beijing

Wanner, D. (2011). *The Power of Analogy*. Berlin, New York: De Gruyter Mouton.

Werning, M. (2010). Descartes discarded? Introspective self-awareness and the problems of transparency and compositionality. *Consciousness and cognition*, 19(3), 751761.

Wierzbicka, Anna. (1988). *The semantics of grammar.* Amsterdam: John Benjamins.

Wu, S. (2017). Iconicity and viewpoint: Antonym order in Chinese four-character patterns. *Language Sciences*, *59*, 117–134.

Wulff, S. (2010). *Rethinking idiomaticity: a usage-based approach.* London: Continuum.

Zimmermann, T. E., & Sternefeld, W. (2013). *Introduction to semantics: an essential guide to the composition of meaning*. Berlin: De Gruyter Mouton.

Zuo, Zhijun. (2006). *Acquisition of Chinese Idioms: from the perspective of cognition*.

MA thesis. Ocean University of China.