The Grammaticalisation Process of Mandarin Topic Markers

Doctoral thesis for obtaining the
academic degree
Doctor of Philosophy (Dr. Phil)

submitted by
Tian Li

at the

Faculty of Humanities
Department of Linguistics

Konstanz, 2023
Tag der mündlichen Prüfung: 14.11.2023
Die Namen der Referenten:

1. Prof. Dr. George Walkden
2. Prof. Dr. Regine Eckardt
3. Prof. Dr. Christoph Anderl
Table of Contents

Zusammenfassung ............................................................................................................................... ii
Abstract ............................................................................................................................................... iv
Acknowledgements ............................................................................................................................. vi
List of Tables ..................................................................................................................................... vii
List of Abbreviations ........................................................................................................................ ix
1. Introduction ...................................................................................................................................... 1
   1.1 Topics in the Linguistic World ................................................................................................. 1
   1.2 Periodisation in the study of Chinese Linguistics .................................................................... 2
      1.2.1 Periodisation of Archaic Chinese ....................................................................................... 3
      1.2.2 Periodisation of Middle Chinese ....................................................................................... 7
      1.2.3 Periodisation Model Used in the Thesis ............................................................................ 7
   1.3 Literary Chinese (LC) and vernacular varieties ...................................................................... 8
   1.4 Aim for Researching ............................................................................................................... 10
   1.5 Structure of the Thesis ............................................................................................................ 10
2. Topic definition and classification of Mandarin Chinese ............................................................ 13
   2.1 What is a topic? ...................................................................................................................... 13
   2.2 Studies on Mandarin Topics .................................................................................................... 16
      2.2.1 Studies on modern Chinese topic constituents ............................................................... 16
      2.2.2 Studies on the Archaic Chinese topic constituents .......................................................... 25
   2.3 Left-Dislocated Sentences and Predicate Sentences ............................................................... 30
      2.3.1 Left-Dislocated Sentences and Topic-prominent Sentences ........................................... 30
      2.3.2 Chinese Predicate Sentences ........................................................................................... 31
      2.3.3 Are predicate sentences left-dislocated sentences in archaic Chinese? ......................... 38
      2.3.4 Section Summary ............................................................................................................. 39
   2.4 Classifications of Topic in Mandarin Chinese ........................................................................ 40
      2.4.1 Aboutness Topics ............................................................................................................. 40
      2.4.2 Contrastive Topics ........................................................................................................... 44
      2.4.3 Topic Chain ..................................................................................................................... 46
      2.4.4 Section Summary ............................................................................................................. 48
   2.5 Topic Marker ........................................................................................................................... 49
      2.5.1 What is a topic marker? ................................................................................................... 49
      2.5.2 Classifications of Topic Markers ...................................................................................... 52
      2.5.3 Topic Markers in Chinese ................................................................................................ 57
         2.5.3.1 Topic Markers in Modern Chinese ............................................................................. 58
         2.5.3.2 Topic Markers in Dialects .......................................................................................... 69
2.5.3.3 Topic Markers in Archaic Chinese ................................................................. 70

2.6 Summary ................................................................................................................. 75

3. Methodology .............................................................................................................. 76

3.1 What is Corpus Linguistics ...................................................................................... 76

3.2 Why corpus linguistics matters in our study? .......................................................... 79

3.3 Corpora used for this thesis ...................................................................................... 80

3.3.1 Corpus comparison ............................................................................................... 81

3.3.2 Why Buddhist Quotes Matter .............................................................................. 82

3.4 Books used for this thesis from the Corpora ........................................................... 84

3.5 Queries and visualisation ......................................................................................... 89

3.5.1 Command and example search ............................................................................ 89

3.5.2 Tools used for Visualisation ................................................................................ 89

3.5.3 Analysing the Graphs .......................................................................................... 90

3.6 Section Summary ..................................................................................................... 91

4. Topic markers in Mandarin ........................................................................................ 92

4.1 Ye “也” ..................................................................................................................... 92

4.1.1 Functions of Ye “也” ............................................................................................ 93

4.1.1.1 Ye “也” in the middle of the sentence ............................................................... 93

4.1.1.2 Ye “也” at the end of the clause .................................................................... 94

4.1.2 The trend of using ye “也” as a topic marker over time ..................................... 97

4.2 Zhi “之” .................................................................................................................. 100

4.2.1 Functions of zhi “之” .......................................................................................... 100

4.2.1.1 zhi “之” as a verb and a sentence final particle .............................................. 100

4.2.1.2 zhi “之” in topic sentences ......................................................................... 101

4.2.2 The trend of using zhi “之” as a topic marker over time ................................... 103

4.3 Ne “呢” .................................................................................................................. 107

4.3.1 The function word ne “呢” ............................................................................... 107

4.3.3 Historical change model of ne “呢” ................................................................ 109

4.3.4 The trend of using ne “呢” as a topic marker over time .................................... 110

4.3.5 Section Conclusion ............................................................................................. 115

4.4 Zhe “者” ............................................................................................................... 115

4.4.1 functions of zhe .................................................................................................. 115
4.4.2 The trend of using zhe as a topic marker over time.......................................................117
4.6 Section Summary ..................................................................................................................121
5. Grammaticalisation Process of topic markers in Mandarin Chinese.................................123
  5.1 Grammaticalisation and degrammaticalisation of Zhe in the history..........................123
    5.1.1 Grammaticalisation, linguistic cycle and degrammaticalisation..........................123
    5.1.2 Why zhe matters? .....................................................................................................125
    5.1.3 Methodology and Results .......................................................................................126
    5.1.4 Is there a linguistic cycle of zhe in Archaic Chinese? ............................................127
    5.1.5 Conclusion .................................................................................................................129
  5.2 The Grammaticalisation process of Ne “呢” as a topic marker.....................................130
    5.2.1 The grammaticalisation model of ne<sub>CT</sub> ......................................................130
    5.2.2 Features of topic marker ne “呢” ...........................................................................132
    5.2.3 Other findings ..........................................................................................................134
    5.3.1 Base-generated topics in Archaic Chinese ...............................................................135
    5.3.2 Moved topics in archaic Chinese .............................................................................140
      5.3.2.1 The syntactic position of moved topics .............................................................140
      5.3.2.2 Do moved topics really exist in archaic Chinese? ............................................143
    5.3.3 The mechanism of forming a topic constituent in archaic Chinese .......................147
    5.3.4 Section Conclusion ..................................................................................................151
  5.4 The interaction of cliticisation and topicalisation in Archaic Chinese............................152
    5.4.1 Introduction ...............................................................................................................152
    5.4.2 Null-copula phenomenon in Archaic Chinese ..........................................................153
      5.4.2.1 What is a copula? ...............................................................................................153
      5.4.2.2 Archaic Chinese copula ....................................................................................154
    5.4.3 Corpus analysis of the archaic Chinese copula in topic sentences .........................157
    5.4.4 Discussion ...............................................................................................................159
      5.4.4.1 Syntactic position of copula in topic sentences ...............................................160
      5.4.4.2 The interaction between cliticisation and topicalisation in topic sentences ......163
    5.4.4 Conclusion .................................................................................................................167
  5.5 Why do dedicated topic markers not exist in Mandarin Chinese? ...............................169
    5.5.1 Dedicated topic markers in different languages and dialects ..................................169
    5.5.2 Different Origins of Forming a Topic Marker ..........................................................171
      5.5.2.1 Locative Markers as Topic Markers ................................................................172
      5.5.2.2 Interrogative Markers as Topic Markers ............................................................173
      5.5.2.3 Time Adverbs as Topic Markers .......................................................................174
    5.5.3 The Grammaticalisation Cline Topic Marker in Mandarin Chinese .....................174
Declaration

I have read and understood the University of Konstanz guidelines on plagiarism and declare that this dissertation is all my own work except where I indicate otherwise by proper use of quotes and references.
Zusammenfassung


In dieser Studie werden unterschiedliche Formen der Thematisierung im archaischen Chinesisch untersucht und der Grammatikalisierungsprozess der Thematisierung anhand eines umfangreichen Korpus analysiert. Das Ziel dieser Arbeit besteht darin, die folgenden Fragen zu beantworten:

2. Der Prozess der Grammatikalisierung von ne “呢” als Thema-Marker: Die Verwendung von ne hat eine Vereinheitlichung der Form erfahren und es liegt ein funktionales Grammatikalisierungsverfahren von ne vor.


4. Die Wechselwirkung zwischen Klitisierung und Topikalisierung im archaischen Chinesisch: Die Kopula stellt eine phonologische Klitik dar, die einen Wirt zu ihrer Linken innerhalb derselben Intonationsphrase aufweisen muss.

5. Warum existieren keine spezifischen Themamarker im Mandarin-Chinesisch? Die chinesischen Themamarker leiten sich nicht von Lokativmarkern ab, die jederzeit weggelassen werden können.


Schlüsselwörter: Historische Syntax, Semantik, archaisches Chinesisch, themenzentriert, subjektzentriert.
Abstract

The notion of topic has been the subject of much debate in the field of Information Structure (IS). Different types of topic markers are found in different languages like Mandarin (Li & Thompson, 1976; 1981), English (Rizzi 1997; Krifka, 2008), Japanese (Kuroda, 2005), etc. Topic, as distinct from subject and focus, is another important grammatical role in a sentence. It can be placed sentence-initially and requires that the speaker and listener share the common ground (CG) of what they are talking about in the discourse (Li, 2005: 15).

Li & Thompson’s (1976) paper can be regarded as one of the most important papers discussing the differences between topic and subject. Li & Thompson (1976) use seven criteria to distinguish the differences between subject and topic and finally come to the conclusion that the topic is a notion in discourse while the subject is a “sentence-internal” one. In other words, subjects and topics behave differently in the discourse strategy. As argued by Rizzi (1997) and Krifka (2008), topic and focus each has a syntactic place in the sentence. Based on the mutual common ground (CG) speakers and hearers have, topic is the information which both hearer and speaker already know and focus is the new information in the discourse (Gundel & Fretheim, 2004).

Mandarin Chinese can be regarded as the most standardised and official language spoken in China. It is widely accepted that Mandarin is a topic-prominent language (Li & Thompson, 1976: 460; Huang, 2009). In my previous preliminary study of topic markers in Han Dynasty compared to Old English, a robust tendency of topic marker usage is shown. However, no scholar has summarised or analysed all kinds of topic markers that appear in different periods of Mandarin Chinese development.

This thesis discusses different kinds of topic markers that exist in archaic Chinese and analyses the grammaticalisation process of topic markers by using a large-scale corpus as methodology. This thesis is targeting the following questions for answering:

1. The Linguistic Cycle of “zhe” ‘者’ in Archaic and Middle Chinese: There is no linguistic cycle of “zhe” in archaic Chinese. The nominaliser “zhe” and the topic marker “zhe” are used from the beginning of history, and the demonstrative “zhe” is appearing in a very short time and soon dies out.

2. The Grammaticalisation process of “ne” ‘呢’ as a topic marker: The usage of “ne” ‘呢’ has experienced a form unification and there is a functional grammaticalisation procedure of “ne” ‘呢’.

3. Why archaic Chinese only has base-generated topics: previous studies suggest that there is a co-existence of base-generated topics and movement topics. It is worthwhile exploring into the fundamental mechanisms that contribute to the fact that only base-generated topics exist in Chinese and to discuss the reasons for their existence.

4. The interaction of cliticisation and topicalisation in Archaic Chinese: The copula is a phonological clitic that must have a host to its left within the same intonational phrase.
5. Why do dedicated topic markers not exist in Mandarin Chinese? Chinese topic markers are not derived from locative markers, which can be omitted at any time.

This thesis is an attempt of researching the inner mechanisms of topic markers, including the relationship between topic markers and clitics, the formation of topic markers and so on. There is little research on the answered questions in section 5, thus suggestions are given in the last part of the dissertation to provide directions for further studies.

Keywords: Historical Syntax, Semantics, Archaic Chinese, Topic-prominent, Subject-prominent.
Acknowledgements

Throughout my doctoral voyage, I would like to extend my sincerest gratitude and appreciation to everyone who has provided me with support and direction. Uncountered individuals' encouragement and invaluable contributions were crucial in ensuring the successful completion of this doctoral dissertation.

Prior to anything else, I would like to extend my most sincere appreciation to my advisor, Professor Dr. George Walkden, for his steadfast support, erudition, perseverance, and insightful counsel throughout the entire research process. Having been acquainted for nearly a decade, Prof. Dr. George Walkden has provided me with invaluable guidance on becoming a proficient linguist and maintaining composure in life. It is a tremendous honour to have a supervisor of your calibre, and I appreciate everything he did for my doctorate in linguistics. Additionally, I express my gratitude to Prof. Dr. Christoph Anderl and Prof. Dr. Regine Eckardt for their insightful observations, constructive criticism, and thorough assessment of my research. Their specialised knowledge in the domain has played a crucial role in enhancing the calibre and rigour of my research. Sincerely, the faculty and research staff of the University of Konstanz are to be thanked for granting access to invaluable resources and fostering a stimulating academic environment. Their unwavering commitment to cultivating intellectual advancement and scholarly distinction has made a substantial impact on my progress as a researcher.

Additionally, it gives me tremendous pleasure to express my gratitude to everyone who contributed to the completion of this dissertation. On my thesis, Qi Yu, Yixiao Song, and Ningning Zheng provided invaluable dialectological, phonological, and theoretical insight. Their unwavering encouragement, tolerance, and comprehension have served as a reservoir of resilience during this arduous expedition.

Finally, I would like to extend my deepest appreciation to my mother, Ping Chen, and closest friend, Yuhang Zhang, Jiayao Shen, Yue Yang for their unwavering support, meticulous attention to detail, and patience during the course of this endeavour.

The thesis is always imperfect, I cherish every moment I have experienced.
List of Tables

Fig. 1 Views on topic and subject constituents
Fig. 2 syntactic structure of a base-generated topic
Fig. 3 Syntactic structures of moved topics
Fig. 4 Syntactic Structure of example (121)
Fig. 5 Syntactic Structure of example (122)
Fig. 6 Syntactic Structure of example (124)
Fig. 7 Syntactic Structure of example (125)
Fig. 8 Syntactic Structure of example (127)
Fig. 9 Syntactic Structure of example (129)
Fig. 10 Moved topics in archaic Chinese sentences
Fig. 11 Base-generated topics in archaic Chinese
Fig. 12 Syntactic positions of different constituents in archaic Chinese.
Fig. 13 Sentence Containing a Topic
Fig. 14 syntactic structure of (148)
Fig. 15 Syntactic structure of (149)
Fig. 16 Syntactic structure of (151)
Fig. 17 Syntactic structure of topic sentences containing a copula
Fig. 18 syntactic structure of topic sentences containing a copula
Diagram 1: Returned results of *zai* in BCC corpus (Beijing Language and Culture University Corpus)
Diagram 2. Modelling KWOC corpus (source from CCL corpus)

Table 1 Books chosen for corpus analysis at the first stage
Table 2 34 Books chosen for corpus analysis (final choice)
Table. 3 current research results of the first appearance of copula *shi*
Table 4. Different structures of [Topic marker, copula] and numbers of hits in the corpus

Graph 1: The trend of using *ye* as a topic marker in the history of Chinese (x=Year, Y=Percentage
*ye / Total ye-hits in each book* *100%*)
Graph 2: The trend of using *zhe* as a topic marker in the history (x=Year, Y=Percentage (zhe /
*Total zhe-hits in each book* *100%*))
Graph 3: The proportion of the *zhe...ye...* among all the examples containing *zhe* and *ye* (x=Year,
Y=Percentage (zheyeye / Total zhe+ye-hits in each book* *100%*))
Graph 4: The trend of using zhi as a topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{TpSum}{Total \ zhi-hits \ in \ each \ book} \times 100\%\))

Graph 5: The trend of using zhi as a contrastive topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{CTpM}{Total \ zhi-hits \ in \ each \ book} \times 100\%\))

Graph 6: The trend of using zhi as an aboutness topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{ATpM}{Total \ zhi-hits \ in \ each \ book} \times 100\%\))

Graph 7: The comparative trend of using zhi as a topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{ATpM \ or \ CTpM}{Total \ zhi-hits \ in \ each \ book} \times 100\%\))

Graph 8: Grammaticalisation of ne (self-summarised)

Graph 9: The trend of using zhe as a topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{Zhe \ Tp \ Sum}{Total \ * \ 100\%}\))

Graph 10: The trend of using zhe as an aboutness topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{ZheatpP}{Zheatp/Total \ * \ 100\%}\))

Graph 11: The trend of using zhe as a contrastive topic marker in the history of Chinese (x=Year, Y=Percentage
\(\frac{ZhectpP}{Zhectp/Total \ * \ 100\%}\))

Graph 12: Comparison of zhe as a contrastive topic marker (red line, x=Year, Y=Percentage
\(\frac{ZhectpP}{Zhectp/Total \ * \ 100\%}\)) and an aboutness topic marker (blue line, x=Year, Y=Percentage
\(\frac{ZheatpP}{Zheatp/Total \ * \ 100\%}\)) in the history of Chinese

Graph 13: Comparison of zhe as a topic marker (blue line, x=Year, Y=Percentage
\(\frac{Zhe \ Tp \ Sum}{Total \ * \ 100\%}\)), ye as a topic marker (orange line, x=Year, Y=Percentage
\(\frac{ye}{Total \ ye-hits\ in \ each \ book} \times 100\%\)), and zhi as a topic marker (red line, x=Year, Y=Percentage
\(\frac{TpSum}{Total \ zhi-hits \ in \ each \ book} \times 100\%\)) in the history of Chinese
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TpM</td>
<td>Topic Marker</td>
</tr>
<tr>
<td>SbjM</td>
<td>Subject Marker</td>
</tr>
<tr>
<td>ART</td>
<td>Article</td>
</tr>
<tr>
<td>CG</td>
<td>Common Ground</td>
</tr>
<tr>
<td>CL</td>
<td>Classifier</td>
</tr>
<tr>
<td>SBJ</td>
<td>Subject</td>
</tr>
<tr>
<td>OBJ</td>
<td>Object</td>
</tr>
<tr>
<td>TOP</td>
<td>Topic</td>
</tr>
<tr>
<td>DM</td>
<td>Discourse Marker</td>
</tr>
<tr>
<td>MP</td>
<td>Modal Particle</td>
</tr>
<tr>
<td>PT</td>
<td>Particle</td>
</tr>
<tr>
<td>RF</td>
<td>Rhetorical Final Particle</td>
</tr>
<tr>
<td>SFP</td>
<td>Sentence-Final Particle</td>
</tr>
<tr>
<td>IntP</td>
<td>Interrogative Particle</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Topics in the Linguistic World

The communicative functions of subjects comprise a variety of duties that are carried out by these linguistic components in order to facilitate effective communication and shape discourse. Subjects function as initial clause components, providing an introduction to the primary entity or action under discussion. By establishing a thematic focus, they direct the course of subsequent discourse and enhance coherence through the provision of referential clarity. Moreover, the emphasis that subjects place on particular information can have an effect on audience perception and the impact of rhetoric. On the other hand, topics pertain to the substance or overarching concept of a discourse, text, or conversation. Subjects are distinct linguistic components contained within sentences, whereas topics encompass more generalised concepts or ideas that may extend across multiple sentences or even entire discourse. Topics serve as the foundational structure for discourse, establishing the necessary circumstances and directing the conversation.

Although subjects and topics both hold significance in communication, discussions frequently prioritise topics due to the fact that they furnish the comprehensive framework and context necessary to comprehend the content of the communication. By providing audiences or readers with a focus on the central idea or objective of a conversation or text, topics enhance engagement and comprehension. Furthermore, the act of concentrating on specific topics facilitates the investigation of more extensive patterns and ideas, thereby facilitating a more profound synthesis and analysis of data. Through adept comprehension and impactful resolution of topics, communicators can guarantee that their message strikes a chord with the recipients and accomplishes its desired objective. Hence, although subjects play a vital role in establishing sentence structures and promoting linguistic coherence, discourse frequently places greater emphasis on specific topics in order to guarantee that statements are significant, pertinent, and influential. Thus, topic constituents are the main target of this thesis. The sentences contain topics usually have certain syntactic structures, as illustrated in example (1):

(1) a. 书，我 喜欢。
    shū, wǒ xǐhuān.
    “Books, I like.”

b. 私 は 雪 が 白い と 思う
    watashi wa yuki ga shiroi to omou.
Identifying the subject and other element in a mix of nouns and pronouns can be challenging. Therefore, during the 1950s, linguists initiated the introduction of the concept of topic constituents. The preliminary investigation focused on topics utilising Japanese as the target language of study. In the 1980s, Li & Thompson (1989) conducted research on typological topic-prominent languages and subject-prominent languages. These prior investigations have presented substantial evidence that certain elements may precede subjects.

Chinese, being a topic-prominent language, has experienced a surge in research on topics since the concept of topics was introduced, particularly in the 1990s. Researchers examine the topic at hand from several viewpoints, such as the role of topics the process of topics becoming grammaticalised, topic markers and other related aspects. However, these studies have not progressed consistently and uniformly. There is a greater emphasis on discussing topic constituents and a significantly lesser focus on topic markers. Synchronic research has been conducted to a much greater extent than diachronic study. Hence, an additional component to the study of topic markers throughout time is necessary.

1.2 Periodisation in the study of Chinese Linguistics

In the preceding section, we have found a need for more research of the topic marker, notably the absence of studies conducted in a chronological manner. In order to fill this gap, we shall analyse the difficulties related to chronological research. Within this section and the subsequent one, we shall address significant challenges: one about categorising historical Chinese dynasties, and another concerning the distinctions between Literary Chinese (LC) and vernacular forms.

The study of periodisation in Chinese linguistics involves a wide range of viewpoints, where scholars propose different models characterised by their distinct terminologies and approaches. This section involves a discussion of periodisation theories put out by renowned scholars including Karlgren, Wang Li, Peyraube, and Behr. The main focus is to comprehend the chronological division between “Archaic Chinese” (AC) and “Middle Chinese” (MC). The cryptic nature of these concepts, which reside at the confluence of historical linguistics and cultural evolution, is crucial to our investigation. Our objective is to determine the exact time bounds of “Archaic Chinese” and “Middle Chinese” by thoroughly examining their nuances within the broader framework of Chinese linguistic development.
It is highly important to discuss the periodisation of Chinese history and analyse the differences between Literary Chinese (LC) and vernacular variations. It is essential to differentiate and address these matters, especially in the field of Chinese diachronic corpus linguistics, while choosing or using texts. Through comprehending these differences, researchers can make well-informed choices regarding the texts to use in their analyses, how to precisely define literary genres, and how to selectively exclude sections of texts that are crucial for tracking language shifts across time.

To further discuss on this point, having a comprehensive understanding of the chronology of Chinese history enables academics to situate linguistic transformations within distinct historical periods. Different dynasties and historical periods often correspond to distinct linguistic features and usage conventions. Therefore, through the categorization of texts according to their specific historical periods, scholars can better trace the evolution of the Chinese language and identify consistent patterns in linguistic changes throughout history.

Moreover, it is crucial to distinguish between literary Chinese (LC) and vernacular variants in order to fully understand the linguistic context of different periods in Chinese history. Literary Chinese, known for its sophisticated and conventional style, was mainly used in written documents and official records for a substantial period of Chinese history. Conversely, vernacular variants refer to the language used in daily conversations by individuals and provide a more accurate portrayal of regional dialects and informal expressions.

When conducting diachronic corpus linguistics research, scholars must carefully select texts that offer a comprehensive portrayal of language usage across different historical periods. Researchers can thoroughly record the diverse range of languages and their changes over time by analysing a broad selection of texts, including both classical Chinese and everyday spoken forms. This methodology facilitates a more comprehensive analysis of language evolution in Chinese history and improves our understanding of linguistic diversity and development.

1.2.1 Periodisation of Archaic Chinese
Karlgren (1923) presented an original viewpoint about the categorization of ancient Chinese history, taking into account the dynamic transformations that occurred. In his influential publication “The Grammar of Old Chinese” (1923), Karlgren proposed the classification of ancient Chinese into discrete epochs, distinguished by distinctive linguistic attributes that align with particular
historical dynasties. Instead of rigidly following chronological divides, his approach prioritised the language morphology and structure of Chinese throughout various historical epochs.

Karlgren placed particular emphasis on the phonological system of ancient Chinese, contending that alterations in these systems may mirror the linguistic advancements of distinct eras. While not providing specific times, it is probable that he delineated more general phases based on language attributes and historical circumstances: The Early Ancient Chinese Stage focuses on the Shang and Western Zhou periods, exploring the early development of the Chinese language and its phonetic characteristics. The Middle Ancient Chinese Stage refers to the period spanning from the Spring and Autumn period to the Warring States period, and continuing through the Qin and Han dynasties. It is highly probable that this phase witnessed notable phonological and grammatical transformations, including the emergence of tonal systems and the establishment of standardised writing. The Late Ancient Chinese Stage refers to the period that includes the Han, Jin, Southern and Northern Dynasties, and Sui and Tang dynasties. During this time, there were potential advancements in the phonological, grammatical, and lexical aspects of ancient Chinese, which were influenced by historical events.

It is important to emphasise that the classification of these stages was determined by language features and historical settings, rather than rigorous adherence to precise historical dates. Hence, the demarcations of these stages may possess a degree of ambiguity and be open to discrepancies among various experts. Karlgren’s periodisation concept (1923) sought to offer insights into the language evolution patterns of ancient Chinese, rather than rigorously classifying historical events.

Wang Li (1956) holds a distinct viewpoint from Karlgren regarding the categorization of Archaic Chinese into periods. He largely concentrates on the grammatical development of Archaic Chinese, with comparatively less attention given to phonetic alterations. According to Wang Li, the crucial factor in dividing Archaic Chinese into periods is the alterations in grammatical structures. He highlights the significance of syntactic modifications in comprehending Archaic Chinese, namely those that may be observed in the linguistic progression evident in different historical texts. Unlike Karlgren, Wang Li places greater importance on textual evidence as the main foundation, rather than focusing on phonological systems. He promotes the examination of grammatical structures and word usage in ancient Chinese writings as a means to enhance comprehension of the linguistic attributes of various time periods. Contrary to Karlgren’s precise categorization, Wang Li perceives the periodisation of Archaic Chinese as being more uncertain. He posits that the language progression of Archaic Chinese is a slow phenomenon, rendering it difficult to precisely delineate
distinct epochs. When discussing the periodisation of Archaic Chinese, Wang Li also highlights the contrast between literary Chinese and vernacular Chinese. According to him, these two literary styles display distinct linguistic features over different historical times, necessitating distinct methodologies for researching Archaic Chinese. Wang Li’s approach primarily focuses on the grammatical development and textual examination of Archaic Chinese. He favours a methodology that prioritises grammar rather than phonetics.

François Peyraube (1996) gives a unique viewpoint on the periodisation of Archaic Chinese, emphasising the impact of socio-cultural elements on linguistic evolution. Peyraube underscores the importance of societal and historical factors in influencing the development of language, promoting a nuanced comprehension of linguistic changes within the wider historical framework. He emphasises the connection between the development of language and historical occurrences, calling for an adaptable method of dividing time periods that takes into consideration socio-political changes. Unlike Karlgren and Wang Li, who primarily concentrate on linguistic structures and textual analysis, Peyraube places greater emphasis on the investigation of societal influences and historical evidence to clarify the linguistic development of Archaic Chinese.

William Behr (1971) offers an alternate perspective on the categorization of Archaic Chinese, with a specific emphasis on the development of written communication, in contrast to the approaches used by Karlgren and Wang Li. Behr’s research largely focuses on the evolution of written language in Archaic Chinese, with a special emphasis on the transformations in writing systems and literary compositions. According to him, the process of establishing a uniform written form of language has had a significant impact on the structure and words used in Archaic Chinese throughout various historical eras. Behr’s periodisation concept is founded upon the evolution of written language, categorising Archaic Chinese into discrete linguistic epochs. Unlike Karlgren and Wang Li, who prioritise phonetics, grammar, and textual analysis, Behr’s approach focuses on the importance of historical documents and literary works for comprehending the linguistic features and cultural background of Archaic Chinese.

Aside from these notable scholars, Chinese linguists are working on periodisation in ancient Chinese: Comparing to the previous scholars, certain Chinese academics seek to categorise the periodisation of Archaic Chinese solely based on chronological time. Luo Changpei’s periodisation model (1957) for ancient Chinese, as presented in his publication “Outline of the History of the Chinese Language” (1957), is founded upon the succession of dynasties in China. The author categorises archaic Chinese into separate time periods that align with various dynasties, each of
which is distinguished by unique linguistic attributes. Luo categorises the Shang Dynasty (1600 BCE - 1046 BCE) as the oldest phase of ancient Chinese, distinguished by oracle bone inscriptions that offer valuable information about early phonological and lexical aspects. The Zhou Dynasty (1046 BCE - 256 BCE) marks a notable era of linguistic advancement, as demonstrated by the appearance of bronze inscriptions that indicate shifts in writing systems and phonetic patterns. Political fragmentation throughout the Spring and Autumn Period (770 BCE - 476 BCE) and Warring States Period (475 BCE - 221 BCE) caused linguistic changes to occur at a faster pace, resulting in regional differences in language usage. Luo underscores the significance of literary works from this period, such as the “Shijing” and “Chunqiu”, in comprehending the evolution of language. The Qin Dynasty (221 BCE - 206 BCE) and Han Dynasty (206 BCE - 220 CE) played a significant role in the standardisation of written Chinese by adopting the clerical script. This led to the development of classical Chinese as a literary language, characterised by specific syntactic and lexical patterns. Luo Changpei’s periodisation model provides a systematic framework for comprehending the progression of archaic Chinese, emphasising the impact of political and cultural transformations on the language advancement across several dynastic eras.

Zhu Dexi’s periodisation concept for archaic Chinese, as outlined in his 1963 book “History of the Chinese Language”, focuses on the shifts in dynasties within China. He categorises archaic Chinese into specific epochs, with each epoch aligning with various ruling dynasties and distinguished by distinctive linguistic attributes. Zhu recognises the Shang Dynasty (1600 BCE - 1046 BCE) as the oldest era of ancient Chinese. He specifically examines linguistic evidence from oracle bone inscriptions to gain insights into the phonetic and lexical features of the language. Zhu highlights the importance of bronze inscriptions throughout the Zhou Dynasty (1046 BCE - 256 BCE). These inscriptions provide insights into the evolution of writing systems and phonetic patterns, as well as the development of regional linguistic differences. Zhu examines the linguistic advancements that occurred during the Spring and Autumn Period (770 BCE - 476 BCE) and Warring States Period (475 BCE - 221 BCE). He emphasises the growing variety of regional dialects and the influence of literary works on the evolution of language. Zhu examines the standardisation of written Chinese and the development of classical Chinese as a literary language during the Qin Dynasty (221 BCE - 206 BCE) and Han Dynasty (206 BCE - 220 CE). This period saw the establishment of specific syntactic and lexical rules in classical Chinese.

Zhu Dexi’s periodisation approach offers unique insights into the development of ancient Chinese language, highlighting the complex interaction between political, cultural, and linguistic elements across many dynastic eras.
1.2.2 Periodisation of Middle Chinese

During the Middle Chinese period, a few of the above scholars have extensively studied it, although not all have focused exclusively on this specific era.

Although Karlgren primarily concentrates on the initial phases of Ancient Chinese, such as the oracle bone and bronze inscription periods, he also explores Middle Chinese to a certain degree. He focused on studying the sounds and structure of Middle Chinese, but his most significant discoveries were in the field of earlier linguistic stages. Peyraube’s research generally centres on the ancient phases of the Chinese language, specifically the Shang and Zhou periods. While his investigation of Middle Chinese is not as deep as other periods, some of his works touch upon topics relating to Middle Chinese. Behr’s research generally focuses on the study of Chinese philology and ancient literature, with comparatively less attention on the field of Middle Chinese studies. Nevertheless, a portion of his literary creations do incorporate analyses of Middle Chinese phonetics and writing. In general, although the degree to which these academics have studied Middle Chinese differs, their research as a whole enhances our comprehension of the overall progression of Chinese language evolution.

Wang Li, not only provided significant contributions to the study of Ancient Chinese, but also offered extraordinary insights into the study of Middle Chinese. Wang Li presented a comprehensive framework in his influential book “Grammar and Vocabulary” to categorise Middle Chinese in a structured manner. He separated Middle Chinese into two distinct periods: “Early Middle Chinese” and “Late Middle Chinese”, spanning roughly from the 6th to the 10th centuries AD. Wang Li stated that there were noteworthy phonetic and grammatical changes occurring between these two periods, warranting their separate category. His scholarly insights regarding the evolution of Middle Chinese have had a major impact on subsequent research in the discipline, acting as a cornerstone of Chinese linguistics studies.

1.2.3 Periodisation Model Used in the Thesis

In the previous sections, we have explored the discussion surrounding periodisation presented by different scholars, along with the claims they have made. The existence of divergent research criteria among researchers has resulted in a multitude of varied results. To choose a periodisation framework that is appropriate for the corpus being studied in this study, we plan to use the assertion put out by Aldridge (2013) as the fundamental premise for defining classic Chinese periods.
The two articles by Aldridge (2013) offer a succinct summary of the syntax of Chinese historical evolution, spanning from the Pre-Archaic era to Middle Chinese. The papers largely focus on parts of grammar that existed before the mediaeval era and demonstrate significant distinctions from contemporary Chinese variations. These include the movement of object noun phrases to a position before the verb, the difference in how subject and object relative clauses are formed, and how changes in argument structure are represented, such as the active and passive voice. Each of these linguistic characteristics is associated with morphological differences observed in nouns, verbs, or pronouns. These differences were either clearly represented in the logographic writing system of Archaic Chinese or have been reconstructed for (Pre-)Archaic Chinese. In the second paper, the author analyses the alterations that took place in Middle Chinese and establishes a connection between these advancements and the vanishing of (Pre-)Archaic Chinese morphology. The main goal of these articles is to emphasise the importance of morphology in comprehending pre-medieval Chinese and the changes that resulted in the noticeable differences noticed in Middle Chinese and subsequent periods.

Based on the mentioned analysis, Aldridge (2013:40) divided Pre-Archaic era to Middle Chinese into the following periods:

- Pre-Archaic Chinese: 14\textsuperscript{th} century BCE - 11\textsuperscript{th} century BCE (Shang)
- Early Archaic Chinese: 10\textsuperscript{th} century BCE - 6\textsuperscript{th} century BCE (Zhou)
- Late Archaic Chinese: 5\textsuperscript{th} century BCE - 3\textsuperscript{rd} century BCE (Warring States)
- Early Middle Chinese: 2\textsuperscript{nd} century BCE - 2\textsuperscript{nd} century CE (Han)
- Middle Chinese: 3\textsuperscript{rd} century CE - 6\textsuperscript{th} century CE (Six Dynasties)
- Late Middle Chinese: 7\textsuperscript{th} century CE - 10\textsuperscript{th} century CE (Tang)

After 10\textsuperscript{th} century CE, from Tang to Qing dynasty, although Aldridge (2013) did not provide more discussion on the period after Tang dynasty, in this paper, we will classify 10\textsuperscript{th} century to 20\textsuperscript{th} as Post-Late Middle Chinese period.

1.3 Literary Chinese (LC) and vernacular varieties

Literary Chinese (LC) and vernacular variants are two separate linguistic manifestations within the Chinese language. During the Han Dynasty, LC underwent a degree of standardisation as a written form, which remained mostly unchanged until the early 20th century. On the other hand, vernacular variants refer to the spoken language utilised by the general population during different historical eras. It is essential to comprehend the distinctions between these two linguistic forms in order to choose and examine texts within the framework of Chinese diachronic corpus linguistics. This
research seeks to examine and contrast LC and vernacular variants, with a focus on their functions in tracking linguistic developments across history.

Literary Chinese, sometimes referred to as Classical Chinese, originated as a written form of communication during the Han Dynasty (206 BCE – 220 CE) and persisted as the predominant language for literary works for more than two thousand years. LC is distinguished by its precise syntax, limited grammatical markers, and abundant utilisation of classical lexicon. It functioned as the primary language for literature, official documents, and intellectual discussions during the period of Imperial China. LC, although considered a written standard, did experience evolution, specifically in terms of language and style, which were influenced by shifts in literary trends and philosophical ideas.

Vernacular variants, in contrast, cover the heterogeneous spoken languages and dialects utilised by the general populace throughout various geographies and historical epochs. These variants developed naturally from ancient Chinese and were shaped by regional cultures, historical events, and interactions with neighbouring linguistic communities. Vernacular variants, in contrast to LC, have enhanced adaptability in terms of grammar, vocabulary, and syntax, which mirrors the ever-changing character of spoken language development. Common dialects encompass Mandarin, Cantonese, Wu, Min, and other others, each exhibiting distinct characteristics and specific contexts of usage.

Comparison and Significance in Diachronic Corpus Linguistics: In the field of Chinese diachronic corpus linguistics, it is crucial to differentiate between Literary Chinese (LC) and vernacular varieties in order to ensure precise interpretation and comprehension when analysing historical texts. LC texts offer unique insights into the formal writing styles, literary traditions, and intellectual conversation of previous eras. Nevertheless, they might not faithfully depict the colloquial language used by regular individuals. Vernacular literature provide insights into daily conversation, regional language variations, and social circumstances. Researchers can analyse language changes, detect linguistic patterns, and investigate society dynamics throughout time by comparing LC and vernacular literature.

To summarise, Literary Chinese (LC) and vernacular variations are separate linguistic manifestations within the Chinese language, each possessing unique traits and historical importance. Comprehending the distinctions between LC (Literary Chinese) and vernacular variants is essential for accurately choosing texts, defining genres, and tracing language changes in Chinese
diachronic corpus linguistics. Through the analysis of both variants, scholars can acquire a thorough comprehension of the evolution of the Chinese language and the societal advancements that occurred in various historical periods. When selecting novels from the corpus, careful consideration will be given to both Literary Chinese (LC) and vernacular varieties.

1.4 Aim for Researching

While acknowledging the need for progress in diachronic research on topic markers, this thesis aims to fill the observed gap. This study aims to focus on two main aspects: firstly, employing corpus studies as the major approach to substantiate the research findings, and secondly, investigating the historical evolution of topic markers in archaic Chinese across time.

Although topic components are currently a popular subject of discussion in Chinese syntactic linguistics, the research on this topic is currently facing several limitations. These constraints encompass, but are not restricted to, the lack of a precise description of the topic notion, insufficient historical data from the corpus, and the absence of a marker shifting model. The objective of this thesis is to consolidate previous ideas and present a widely accepted concept regarding the issue, in order to enhance the unified definition of the topic conception within the academic community. In this thesis, a thorough corpus analysis is being conducted to address the lack of historical linguistic corpus data. The conclusion is strengthened and persuasive when it is based on the two most extensive Chinese corpora, which encompass the ancient Chinese data and are supported by a significant volume of data. Given the recent advancement of computers, the use of corpora as a research approach is still relatively new. However, this corpus study is helping to address the existing research gap to some extent. Prior studies have predominantly involved linguists conducting limited-scale corpus analyses, primarily centred around the thematic aspects of a single book. This thesis functions as an addendum to the previous corpus study and provides supplementary data that is beneficial to the existing research on the topic marker. Furthermore, the objective of this thesis is to examine the significance that is hidden within these facts and to ascertain the types of implications that can be derived from this data. This part will function as a promotion and an addition to the ongoing discussion on the issue. Furthermore, it will provide novel insights and explore unexplored phenomena pertaining to the topic matter.

1.5 Structure of the Thesis

The thesis comprises six sections, beginning with the introduction. This section provides an overview of the research backdrop, the study’s objective, a concise summary of my research goals,
and the research questions. The second section, known as the literature review, primarily focuses on providing a clear definition of the topic and presenting the various categories of topic markers in Mandarin Chinese.

The second section is made up of six primary divisions. The initial division presents the fundamental concept that will be explored in this thesis. The second division examines the research conducted on both contemporary and ancient Chinese topic constituents. The third division involves a comparison between the left-dislocated topicalisation phenomenon and predicate sentences found in Mandarin Chinese. We compare these two occurrences because they exhibit very comparable structures. In this section, we will examine the underlying mechanisms involved in the formation of these two syntactic structures and analyse the similarities and differences between left-dislocated topicalisation and Chinese predicate sentences. In the fourth section, our thesis places a strong emphasis on the constituents of the topic. Specifically, we focus on the classifications of topics in Mandarin Chinese. Drawing from the research conducted by Frascarelli & Hinterhölzl (2007), we have included the classifications of aboutness topics and contrastive topics in Chinese. However, we have omitted the familiar topics due to the lack of available information. In addition, we explored the concept of the subject chain in this section, as it is widely employed in Chinese discourse. The fifth component of the content provides a summary of prior studies on the issue marker. These studies are categorised under the following directions: The first route focuses on defining the subject marker. The second direction explores several categories of topic markers, as Chinese linguists have distinct methods for segmenting them. The final section presents a comprehensive overview of all the topic markers that have been studied thus far. The second portion of the literature review is concluding after a summary.

The third portion, known as the methodology section, is dedicated to providing a comprehensive and extensive account of the methodologies employed in the thesis. We provided a concise overview of corpus linguistics and its significance in our research. Next, we presented the corpora utilised in this thesis, as well as the selected books for analysis derived from these corpora. This section of the technique focuses on presenting the queries and visualisations.

The fourth section of this thesis examines the role of the topic markers we are examining, using the subject definition and topic marker definition we established in the second section. We have selected four common topic markers in Mandarin: ye “也”, zhi “之”, ne “呃”, and zhe “者”. Within each section, we examined the individual roles of each topic marker and the evolving usage patterns of the specific topic marker over time. In the fifth subsection of the fourth section, we provided a
comprehensive overview of the overall utilisation of all the topic markers and conducted a horizontal comparison between them. The final component is the summary of the section.

The fifth section of this thesis is dedicated to exploring the roles of topic markers and their grammaticalization process in Mandarin Chinese, which is the most crucial aspect of the debate. The initial investigation focuses on the presence of any linguistic cycle in archaic and middle Chinese concerning the topic marker zhe “者”. In the second section, we are simulating the process of grammaticalization of ne “呢” as a marker of topics. The final section comprises a theoretical analysis that examines several syntactic trees in order to determine if base-generated topics and relocated topics can coexist in archaic Chinese. In the fourth section, which is also a groundbreaking section, we are examining the interplay between cliticisation and topicalisation in archaic Chinese. The fifth section presents an additional theoretical argument explaining the absence of solo topic markers in Mandarin Chinese. Through the utilisation of case studies and models, we want to venture into a study domain that has been scarcely explored by individuals. In this section, our objective is to address a previously unanswered question regarding ancient Chinese subject markers, a topic that has eluded scholars thus far.

In the concluding section, we will consolidate all the findings, address the limitations of this thesis, identify unresolved issues, and propose recommendations for future research.
2. Topic definition and classification of Mandarin Chinese

We will provide a relatively comprehensive discussion of all concerns pertaining to the “topic constituent” in this section. What constitutes a topic? is the initial inquiry that will be addressed in section 2.1. Determining the meaning of the term “topic” is crucial for forming the following judgement regarding what qualifies as a topic and what the topic marker in Chinese denotes as the topic constituent in a thesis. In the second section, we shall provide a summary of the research on Mandarin-related topics in accordance with the initial inquiry. As left-dislocated sentences are frequently compared to topic-prominent sentences, we will examine the similarities and differences between left-dislocation sentences and predicate sentences in the third section. The classifications of topic constituents in Mandarin Chinese are the main theme of the following section, and this section concludes with a discussion of the prevalent and significant debates on topic markers that currently exist in the field of topic marker research.

2.1 What is a topic?

When conversing, individuals employ diverse methods to effectively transmit information, and there are numerous approaches to express various types of information. For example, distinctions in prosody can serve to differentiate whether the information being presented is new or old. Additionally, distinct morphological markers can serve to remind the conversation partners whether this information is assumed or not within the discourse. In order to enhance communication efficiency, the data is condensed into one or more packets within a sentence. Information Structure is the study of how information is compressed within a sentence in a discourse.

Alternatively stated, information structure is limited to the nouns and does not extend beyond the level of the individual sentence. According to Chafe (1976), nouns that can constitute packaging can serve as the subject or topic of the sentence, be specific or general, function as a point of contrast, or take on other forms that satisfy the requirement.

Similar to how the Information Structure evaluates the compressed data pragmatically, the Common Ground (CG) evaluates the propositions similarly. The Common Ground, as described by Krifka (2007), possesses two characteristics: Prior to communicating, the parties involved should each have a comparable set of propositions in their minds. Secondly, identical entities should be included in Common Ground. In contrast to the transitory nature of Information Structure, Common Ground remains constant throughout sentences. The minds of the interlocutors are in a constant state of flux regarding the Common Ground while conversing. The entities mentioned in the Common Ground,
which encompass Information Structures, are selected and discarded during the discourse. Put simply, the correlation between the Information Structure and the Common Ground is akin to two individuals entering a library together (representing the discourse in which both parties are engaged) and subsequently traversing distinct sections collectively (representing various Common Grounds). Upon encountering a captivating book within a particular domain, they shall extract it and reposition it (Information Structure) once they have completed their perusing of said book. They then proceed in the opposite direction until they exit the library. The concept of “topic” is fundamental to many of these captivating books.

Numerous scholarly investigations have been conducted on the topic matter (Lambrecht, 1996; Krifka, 2007; Chafe, 1976; Gundel & Fretheim, 2004). At its core, topic denotes the subject matter of the sentence (Gundel & Fretheim, 2004: 1). However, this concept is extremely imprecise, and the notion of topic encompasses much more. “The entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content” is what Krifka defines as the “topic” (Krifka, 2007: 41). Krifka (2007) also argues, in accordance with the established notion, that the topic's comment should also be preserved in the CG context. The topic of the discourse should be some information that both interlocutors were already aware of. Nevertheless, topic is not solely a pragmatic-semantic concept; it can also be denoted through morphosyntactic mechanisms, which merit further examination. The Spec CP position is occupied by the subject in English, as illustrated in the tree in (2):

(2)

The topic may be movement-related in English. Preceding the subject in a sentence can establish a topic sentence, as illustrated in (3):

(3) a. I like books.


When "books" are presented in the initial place of (3a), which can be considered as the canonical sentence, (2b) transforms into a topic sentence. This is the tree structure of (3b):
There are two movements in (3). The first movement is the movement of specifying that this is the subject "I": the movement of I is driven by the EPP feature, which indicates that the T head carries the uninterpretable D feature causing the subject to move to the upper DP position (Adger, 2003: 172). The second movement of books is driven by the information structure principle. Therefore, the topic constituent should be placed in the initial place of the sentence. There still exist other ways to introduce topics like cleft sentences (Rizzi, 1997); however, this thesis is only focusing on the sentence-initial topics, and the reason why we are only focusing on the sentence-initial topics will be answered later in this section. These topics share similarities: Firstly, the syntactic position of these topics is sentence-initial. Secondly, these topics are DPs. Mandarin Chinese has been claimed to be a topic-prominent language, as discussed by Li & Thompson (1976). However, do Mandarin Chinese topics differ from English topics? If yes, how do they differ?

Topics in Mandarin Chinese exhibit both similarities and differences in comparison to English topics. In a pragmatic sense, Mandarin Topics do not necessarily have to be old information, which is comparable to topic components in English. The ancient Chinese language includes a specific kind of topic sentence known as the judgement sentence (Liu, 2016; Shi, 2000; Xu & Langendoen, 1985), which is also referred to as the predicate sentence. This sort of sentence features noun or noun phrase predicates, which serve to attribute a property to an item located at the beginning of the sentence. Thus, the crucial aspect of the sentence lies in identifying the quality, regardless of whether the fronted noun represents new or old knowledge. According to Chafe (1976), topic in Mandarin Chinese serves as the framework that contains the sentence and restricts the applicability
of the major predication to a specific domain. Additionally, it establishes a spatial, chronological, or individual framework in which the fundamental assertion is valid (Chafe, 1976: 51). Thus, from a pragmatic standpoint, it may be concluded that the concept of topic in Mandarin Chinese does not necessarily refer to previously shared material on the Common Ground. Continuing with the library metaphor, the interlocutors can introduce a new book to each other and place it on the bookshelf collectively. A frame is a structure that serves to limit the domain of a predicate and determine its function.

In terms of syntax, Chinese themes exhibit a combination of base-generation and movement strategies. Huang, Li & Li (2009) argue that the topic's structure without gaps is created by a base-generation strategy, whereas the gapped topic is derived through A'-movement. The subsequent part will present an in-depth analysis of the classification of topics in Mandarin Chinese.

In summary, the concept of topic is fundamental in the field of Information Structure, and the English and Chinese languages exhibit distinct behaviours when it comes to topics. In Chinese, the topic does not necessarily have to be old information, which is analogous to the topic in English from a pragmatic perspective. In English, the topic element is syntactically displaced to the beginning of the sentence. However, the topic element in Mandarin Chinese can be created through either base-generation or A'-movement.

2.2 Studies on Mandarin Topics

In the preceding section, we explored the concept of topic and the distinctions between English and Chinese topics. Within this part, we will examine the research conducted on subjects related to Mandarin. This section comprises two parts: the first half focuses on the examination of contemporary Chinese topic constituents, while the second part delves into the analysis of ancient Chinese topic markers. The final portion of this thesis will introduce the fundamental theory derived from these findings.

2.2.1 Studies on modern Chinese topic constituents

The definition of subjects in Mandarin Chinese has been a subject of extensive debate over the years (Zhao, 1979; Lu, 1986; Li & Thompson, 1989; Hu & Fan, 1993). The disputes over themes revolve around whether the concept should be understood from a pragmatic perspective or a semantic perspective. In which position within a phrase can a topic element be located? In order to address these inquiries, linguists are categorised into four distinct types:
The first group of specialists, lead by Yuanren Zhao, concur that "in Chinese, it is appropriate to consider subject and predicate as topic and comment" (Zhao, 1979: 45), which is also one of the earliest examinations of the resemblances and distinctions between topic and subject. In line with Zhao’s (1979) findings, Liu (2001) likewise asserts that the Mandarin subject lacks nominal properties and has a broad reach. Thus, it is more suitable to consider the subject as “topics” (Liu, 2001). This scholar holds the belief that the concept of the subject is nonexistent in Mandarin, and only topic elements are there. At first glance, this argument appears highly rational: the concept of subject and topic exhibits numerous similarities. Based on the example provided in (4) below, it is not feasible to determine whether the emphasised element is a topic or a subject:

(4) 小 王 来 自 广州。

xiǎo wáng lái zì guǎngzhōu.

Little Wang come from Guangzhou.

“Little Wang comes from Guangzhou.”

When comparing the English translation "Little Wang comes from Guangzhou" with sentence (4), we consider "Little Wang" to be the subject in English. Can we also consider *xiaowang* in sentence (4) to be the subject? Affirmative. Given that the constituent *xiaowang* is in the same syntactic position as the subject constituent, is it possible for *xiaowang* to be the topic of the sentence? Affirmative, the answer is likewise in the affirmative. Considering the subject and predicate as the topic and comment respectively can elucidate why subjects and topics can align seamlessly. Nevertheless, when the subject and topic appear together in the same sentence, as seen in example (5) below, this argument lacks persuasiveness:

(5) 这 本 书 [NP1] [NP2] 很 多。

[zhè běn shū] [yè shù] hěn duō.

This book has many pages.”

In (5), there are two noun phrases (NP) present, including *zhebenshu* (这本书) meaning “this book” and *yeshu* (页数) meaning “page number”. As per Zhao's (1979) findings, both NP1 and NP2 are considered as themes. Nevertheless, if we interchange the positions of NP1 and NP2 in (6), what causes the alteration in the sentence's meaning?

(6) 这 本 书 [NP1] [NP2] 很 多。

[zhè běn shū] [zhebenshu] hěn duō.

This book has many pages.”

[page number]NP1 [this CL book]NP2 very many.
“(As for) Pages, this book has many (pages).”

In relation to this issue, Zhao (1979) only asserts the presence of a topic without adequately addressing certain inquiries, such as whether both noun phrases NP1 and NP2 in a sentence like (5) can be considered as topics. If we change the order of two noun phrases, as shown in example (6), why does the syntactic meaning change? Are there any limitations or restrictions between these two topic constituents? If the answer is affirmative, what is the hierarchical relationship between these two constituents? Yuanren Zhao’s exploration of the relationship between Mandarin themes and subjects has prompted linguists to debate whether they should coexist in Mandarin Chinese. This has led to the emergence of a new group of linguists.

The second type of researchers maintains that subject and topic coexist in Mandarin Chinese, albeit they pertain to distinct “levels” and necessitate distinct treatment. Under the guidance of Jianming Lu, they hold the belief that the “subject” is a syntactic notion that may elucidate the link between various words inside a sentence. According to Lu (1986: 81-82), the concept of topic is a practical idea that is beneficial for communication. In contrast to Zhao’s assertion in 1979, this argument possesses its own merits: Firstly, this assertion acknowledges the distinctions between topics and subjects, recognising that they are indeed separate ideas in Mandarin Chinese. Additionally, Lu elaborates on the distinction between the two concepts at different linguistic level: The concept of subject is observed in the syntax level, while the concept of topic is observed in the pragmatic level. In the first iteration of Yuanren Zhao’s work, he conflates two concepts and asserts that Mandarin lacks a subject, only possessing a topic. This sophisticated discourse acknowledges the coexistence and potential intersection of both the subject and topic in Mandarin.

Looking into details, when having a sentence as shown in example (7):

(7) 小 王 , 来 自 广州。

Little Wang, come from Guangzhou.

“In sentence (7), the NP 湘湘 is syntactically functioning as the subject: it is a noun phrase that is connected to the verb. Thus, 湘湘 functions as the subject of the phrase from a syntactic perspective. From a pragmatic standpoint, this noun phrase is positioned at the beginning of the sentence and is followed by a prosodic boundary. The reason for considering it a topic is that Mandarin topics have a potential phonological phrase edge, which means that every so-called subject is a topic constituent (Geng, 2008: 164-165). Thus, in accordance with Lu's (1986) theory, the constituent 湘湘 exhibits a functional overlap between the subject and the topic.
Nevertheless, does this categorization truly expose the fundamental distinction between these two concepts? Regrettably, the response is negative. An unresolved issue is to the distinction between the subject and topic functions when they potentially overlap. Are there any specific criteria that can differentiate whether a noun phrase at the beginning of a sentence serves as a subject or a topic? Although Lu differentiates between these two concepts and suggests that topics are primarily employed pragmatically, with speakers using them in discourse to convey new information to the listener, if we solely consider the syntax, we can only identify the subject rather than the topic, as per his assertion. From a grammatical standpoint, the sentence solely has subjects, which distinguishes it from Zhao’s (1979) assertion. Regardless of the syntactic existence of any constituent, their assertions are still insufficient to address the question of how we should analyse the constituents of NP1 and NP2 in example (6). While the topic and subject are distinct concepts, their relationship is not always straightforward. When two noun phrases appear together in a sentence, it is still unclear how the constituents of these noun phrases interact. Hence, the presence of either merely a subject or a topic in Mandarin is inadequate for doing research.

According to Lu’s (1986) analysis, linguists have come to understand that even when the topic and subject of a sentence have similar functions, they are still distinct concepts when they appear together. This pertains to the categorization of scholars into three distinct types. The third category of researchers posits that topic and subject can both coincide and exist autonomously. The perspectives presented in the works of Li & Thompson (1976, 1989) are commonly cited as evidence in favour of this viewpoint.

The core assertion in Li & Thompson’s (1976, 1989) work is that both the subject and the topic may be identified at the syntactic level. They argue that these two concepts are distinct from one other. Li & Thompson’s study on the subject and the topic can be categorised into two distinct phases. In Li & Thompson’s (1976) initial phase, they establish a definitive differentiation between the concepts of topic and subject. They concur that Chinese is a topic-prominent language, in contrast to English, which is a subject-prominent language. There are seven criteria for diagnosing an NP constituent. The key and essential distinction between topic and subject lies in their agreement with the verb, as seen in example (8).

(8) a. **He** wants the **books**
   
   SBJ          OBJ

   b. It is the **books** that **he** wants.
   
   Topic          SBJ
Sentence (8a) and (8b) convey the same meaning, but in the second sentence, the phrase "the books" has undergone grammatical transformation and become the topic. It is evident that the verb continues to choose the subject "he" instead of opting for the topic. The selection of the topic is not influenced by the verb. In other words, verbs are always connected to the subject rather than the topic. The subject of a sentence determines the perspective or viewpoint of the action, experience, state, etc., indicated by the verb (Li & Thompson, 1976: 464). The alignment between the topic and predicate is seldom encountered. An additional effective method for differentiating between topic and subject is to analyse the definiteness of the noun phrase. In a Tp language, the topic phrase must be explicitly specified. Regarding the subject, it does not always require specificity. The presence of a subject is mandatory in Spanish, which results in the use of dummy subjects in Spanish languages like English. However, the subject is not necessarily present in certain Tp languages.

Nevertheless, Li & Thompson’s (1976) article still exhibits several issues: The primary issue lies in the fact that Chinese does not follow a verb-final structure. Li & Thompson argue that Chinese is undergoing a transition towards becoming a verb-final language in the context of topic-prominent languages. Chinese, being a topic-prominent language, permits the flexibility of word order, allowing for variations of SOV and OSV in addition to the typical SVO order. This phenomenon has been examined by Huang, Li, and Li (2009), however the occurrence of SOV order is limited by certain constraints. Unlike the SVO order, it is not universally applicable, as demonstrated in example (9):

(9)

a. 我 喜欢 书。 (SVO order)

Wō xīhuān shū.
I like books.
“I like books.”

b. 书, 我 喜欢。 (OSV order)

Shū, wó xīhuān.
books, I like.
“Books, I like.”

c.*我, 书, 喜欢。 (SOV order)

Wō, shū, xīhuān.
I, books, like.
“I, books, like”
Li & Thompson (1976) argue that passive construction in Chinese lacks a “marginal construction” and is infrequently utilised in conversation. Nevertheless, whether in everyday conversation or written language, the passive form, which may be categorised into long passive and short passive (Huang, Li & Li 2009), is an essential sentence structure in Chinese. As the passive structure is not the focus of the thesis, the arguments will be addressed in future research.

In Li & Thompson’s (1989) book, the writers shift their attention to examining the distinct features of Mandarin in various domains during the second phase. Chapter 4 will look into the characteristics of topic and subject in Mandarin, as explained by the writers. In addition to comparing the topic and subject in Mandarin, they also analyse several sentence forms that include both subject and topic, sentences where the subject and topic are the same, sentences without a subject, and sentences without a topic. The topic of double subject phrases is now being debated. The formulation of these sentences necessitates a more thorough examination of the distinctions between the subject and the issue.

In addition, the authors provided a rationale for their belief that Mandarin is undergoing a shift towards becoming a verb-final language. This can be seen as an elucidation of Li & Thompson’s (1976) explanation. According to Li & Thompson (1989), determining the word order of Mandarin is challenging due to its combination of elements from both SOV and SVO languages, with a greater influence from the former (Li & Thompson, 1989: 23). However, they argue that Mandarin has a greater prevalence of the Subject-Object-Verb (SOV) trait when compared to other Chinese dialects. Consequently, they believe that Mandarin is prone to evolve into a verb-final language. Ultimately, the researchers determine that Mandarin is a topic-prominent language and it does not fit into any of the word-order patterns commonly described by universal linguists (Li & Thompson, 1989: 26). Nevertheless, when encountering phrases containing dual subjects, it becomes challenging to determine whether the noun phrase functions as the subject or topic. To address this drawback, a T, which stands for topic, component is introduced. This concept is then used with S(subject), O(object), and V(erb) to analyse the word order. Therefore, Mandarin can be classified as a tonal, subject-verb-object (TSVO) language.
The viewpoint presented by Rizzi (1997) and Krifka (2007) likewise falls under the category of the third type. According to them, both topic and focus have distinct syntactic positions within the sentence. The topic of a sentence refers to the subject being discussed, whereas the focus is the new information being presented in the conversation, based on the shared understanding of the speaker and the listener. Rizzi establishes five fundamental differences between focus and topic. According to Rizzi (1997: 289), the topic can include a resumptive clitic in the comment. Furthermore, it is important to note that the issue does not exhibit any weak cross-over effect. Furthermore, emphasis may be quantified whereas topic cannot, and attention occupies a distinct and exclusive structural place. Finally, a clause can consist of a single focus and multiple topics. According to Rizzi (1997: 291), a wh-operator in main questions can be used with a Topic in a specific order, but it cannot be combined with a Focus. These differentiations can be included into Li & Thompson's (1976) work, which can discern the subject and the topic constituent more explicitly.

The third category of scholars possesses both benefits and drawbacks: Regarding the benefits: Chinese scholars mostly prioritise studying the impact of Tp phenomena in pragmatics, rather than focusing on syntax or typology. The investigation of the Tp phenomenon should encompass not just pragmatics, but also syntax and typology, as these three facets are intricately linked. It is unjustifiable to segregate pragmatics from syntax or semantics in their investigation. This research integrates syntax and pragmatics to demonstrate that both the topic and the subject can be present in both aspects. It aims to uncover the underlying relationship between syntax and pragmatics, which is a novel contribution. One drawback is their exclusive emphasis on the noun phrase topic elements at the beginning of sentences. If different forms of topic constituents exist, their analysis will be limited and require further effort to substantiate the new assertion. Additionally, they do not prioritise the examination of the historical structure of the elements related to the topic. By examining the historical manifestations of topic elements, the idea could potentially provide further breakthroughs.

The final category of researchers, akin to the third category, also acknowledges that the topic and the subject are distinct concepts that can exhibit some degree of overlap in some activities. Contrarily, the fourth category argues that the subject matter is not a fundamental element, but rather a supplementary component that can be attached to the syntactic elements. Xu (2003, 2004, 2010) proposes that when a syntactic ingredient serves as the topic, it is labelled with [T+]. This sort of study appears to be a synthesis of the viewpoints presented by Lu (1986) and Li & Thompson (1976). Both scholars concur that the topic is a concept within pragmatics, while the
subject pertains to syntax. The [T+] label serves as a mediator between syntax and pragmatics. The [T+] trait is dependent on the constituents and cannot be marked independently. If a component in a sentence is marked with the [T+] feature, it will be allowed to bear the topic feature. Thus, in the presence of a sentence such as (10), the subject constituent bears the [T+] feature:

(10) 小 王 [T+], 来 自 广州。

xiǎo wáng [T+], lái zì guǎngzhōu.

Little Wang [T+], come from Guangzhou.

“Little Wang, he comes from Guangzhou.”

Comparing with the second type of researchers, this claim can be regarded as the advanced and combined version of the second and the third type. In this claim, when having a sentence like (10), the property of “Xiaowang” is obvious: it is a constituent carrying both the subject and the topic function. At the meanwhile, when having a sentence like (11), when the topic and the subject co-occur in one sentence, by using the [T+] feature, it is easy to distinguish whether the NP constituents is the topic or the subject:

(11) [页 数][T+] [这 本 书] 很 多。

[yè shù][T+] [zhè běn shū] hěn duō.

[page number][T+] [this CL book] very many.

“(As for) Pages, this book has many (pages).”

By adding a [T+] feature to the “Page number”, it indicates that the “page number” is the topic and “this book” is the subject of the sentence. Similar to the third type, this type of scholars gives a solution to the sentence containing both a topic and a subject.

By incorporating a [T+] attribute into the "Page number" constituent, it clarifies that the “page number” is the focal point and “this book” is the primary subject of the phrase. Like the third category, this group of scholars provides a resolution to a sentence that includes both a topic and a subject.

In contrast to the preceding three categories, the final category of scholars does not engage in a discussion over whether the topic constituent is syntactic or pragmatic in nature. The topic function is a versatile label that can be employed in both syntactic and pragmatic contexts. Nevertheless, a significant issue with Xu’s (2013) work is his inability to avoid addressing the distinction between subject and topic. Specifically, when a [T+] label is added to a constituent, the component bears the topic feature and is manifested at the syntactic level. Curiously, the constituent that carries the [T+] characteristic is referred to as the “big subject” because it emerges at the syntactic level. Hence,
despite the inclusion of the [T+] label, the determination of whether the subject matter pertains to syntax or pragmatics remains an issue of ongoing debate, leading to a more intricate analysis.

The table below summarises the views of four types of linguists towards the topic constituent:

<table>
<thead>
<tr>
<th></th>
<th>Representative Works</th>
<th>Views on Subject &amp; Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st type</td>
<td>Zhao, 1979</td>
<td>Only topic and comment exist in Chinese rather than subject and predicate.</td>
</tr>
<tr>
<td>2nd type</td>
<td>Lu, 1982</td>
<td>Subject and topic co-exists: subject exists on the syntactic level and topic exists on the pragmatic level.</td>
</tr>
<tr>
<td>3rd type</td>
<td>Li &amp; Thompson, 1976, 1989</td>
<td>Subject and topic co-exists in the syntactic level: the functions of these two constituents can be either identical or overlap.</td>
</tr>
<tr>
<td>4th type</td>
<td>Xu, 2003, 2004, 2010</td>
<td>Topic is a function rather than being a constituent: when the constituent has a topic function, a [T+] label is added onto this constituent. The new constituent is called as a “big subject”.</td>
</tr>
</tbody>
</table>

Fig. 1 Views on topic and subject constituents

After evaluating all four analyses about the distinction between topic and subject, this thesis will endorse Li & Thompson's claim as the foundational theory. The reasons are as follows: To begin with, it is widely acknowledged that the topic and the subject are distinct concepts that may be examined from both a syntactic and a pragmatic perspective. Subject and topic are terms that are present in both syntax and pragmatics. They have distinct syntactic positions and there are certain constraints about their sequential arrangement inside a phrase. The topic typically precedes the subject, suggesting the presence of a hierarchical restriction between the topic and subject elements. Another rationale for selecting Li & Thompson's assertion is their exclusive emphasis on the NP elements positioned at the onset of the phrase. According to Chinese linguists, sentence-initial noun phrases (NPs) are a common and primary type of topic sentences. It is imperative to prioritise the historical evolution of the Tp phenomena rather than solely concentrating on present Mandarin. The approach proposed by Li & Thompson (1976, 1989) is highly compatible. As previously stated, archaic themes typically take the form of a noun phrase at the beginning of a sentence, with or without a punctuation mark. Hence, their approach is highly appropriate for conducting study on the archaic Chinese topic.

Nevertheless, the foundational theory also possesses its drawbacks and presents unresolved inquiries: Initially, the conclusion they provide regarding word order is inaccurate. Based on the “Order of Subject, Object and Verb (81A)” characteristic in WALS
Mandarin can be classified as a predominantly Subject-Verb-Object (SVO) language. Li & Thompson (1976) have provided examples of scenarios in which the subject-object-verb (SOV) order can occur. However, it is not enough to solely consider modifiers without also evaluating the actual verb and object in the phrase. Hence, in the thesis, we exclude any word order that follows the Subject-Object-Verb (SOV) pattern, and the impact of this erroneous conclusion can be mitigated. Furthermore, while Li & Thompson (1976) have established seven criteria for distinguishing whether a noun phrase (NP) at the beginning of a sentence functions as a topic or a subject, they have not included any syntactic trees to illustrate the distinction between these two concepts. The benefits of creating these syntactic trees are that they allow for visualising whether a constituent is base-generated or transferred. If there is a motion present in the sentence, the process of inducing the motion can be explicitly elucidated through the trees. Hence, this thesis will offer a more profound examination of the syntactic distinction between these two concepts, as depicted by the syntactic trees.

In summary, despite the various perspectives on the definition and distinctions between topic and subject in Mandarin, it is crucial to consolidate these viewpoints and establish a robust standard. In this thesis, we employ the Li & Thompson (1976, 1989) theory to delve into the intricacies and evolution of the archaic topic constituent.

2.2.2 Studies on the Archaic Chinese topic constituents

The study of modern Chinese subject elements commenced fifty years earlier than the study of archaic Chinese topics. The study of ancient Chinese subject constituents is heavily influenced by Li & Thompson (1976). In this section, we shall go into various facets to investigate these constituents in greater depth.

The initial scholarly discourse on the distinctions between subject and topic emerged in the book *National Grammar Draft*. Following Mashi Wentong, published in 1896, this book is written in classical Chinese and completed in 1922. It can be seen as a highly significant grammar book that encompasses the syntax, semantics, and lexicology of classical Chinese during that time. In Chen's 1982 work, he introduces two concepts: “indicator” and “explanator”. The indicator is the element that the speaker wishes to emphasise and identify in the sentence, whereas the explanation provides

---

1 《国文法草创》, 1982
2 “标语”
3 “说明语”
the facts that clarify the indication. The concepts being discussed bear a strong resemblance to the
categories of “topic” and “comment”, more precisely, the issue that is being discussed and the
accompanying comment. Chen’s study (1982) might be seen as one of the early works to analyse
the topic element in classical Chinese. However, due to the book’s limited scope as a concise
introduction without further investigation, this argument did not receive significant attention at the
time, despite its highly constructive nature.

Following Chen’s study in 1982, Shen did research on several sorts of narrative sentences in Zuo
Zhuan, a work completed during the Spring and Autumn period. According to Shen (1986), archaic
Chinese consists of two sorts of subject-predicate sentences: those that include agents and those that
include themes. The sentences that include themes consistently follow the structure of “theme +
comment”. In Shen’s (1989) study, a more comprehensive analysis is conducted on the structure of
“theme + comment” sentences. According to him, the Zuo Zhuan thematic sentences can be
categorised into five types of relations: hypothesis-inference, purpose relationship, concession-
adversative, plain-adversative, and causal relationship thematic sentences (Shen, 1989: 1-3). Shen
(1989) has attempted to analyse the connection between the theme and the comment among these
five sentence patterns. Shen (1989) has two limitations. Firstly, he does not provide any additional
clarification on these sentence types. Secondly, certain sentence types such as concession-
adversative thematic sentences and plain-adversative thematic sentences resemble adverbial
sentences rather than thematic sentences. Despite several shortcomings, this study remains highly
important in its analysis of the logical correlation between the theme and the commentary.

In the early 21st century, Yang & He (2001) build upon the research conducted by Shen (1989) to
examine several categories of themes. The subjects are categorised into four types: agent subject,
recipient subject, existence subject, and thematic subject. The theme subject is closely related to the
issue we previously examined. According to Yang & He (2001), thematic subjects are not verb-
related, and the predicate in these sentences aims to specify the thematic subjects. The thematic
subjects explored by Yang & He (2001) align closely with the aboutness issues, while the predicate
element pertains to the comment.

In 2002, Yin (2002) also examines the narrative sentences found in the book The Annals of
Lvbuwei. In this study, the author categorises narrative sentences as remark sentences, which
provide explanations about the topic, and plain-narrative sentences, which describe deeds and
events (Yin, 2002: 165). The remark sentences follow a structure known as “topic-commentary
components”, whereas the plain-narrative sentences follow a structure known as “agent-narrative
components”. The work conducted by Yin (2002) has demonstrated advancements in comparison to the previously listed studies for two specific reasons: Shen (1989) included adverbial links in his five classifications of “theme-comment”, however Yin’s classification (2002) removed them. These two forms of narrative sentences can be classified as subject-predicate sentences and topic-comment sentences. Yin’s conclusion (2002) strongly supports the existence of topic sentences in early Chinese history.

Following Yin’s (2002) research, the examination of archaic Chinese topic elements has been divided into two primary approaches: The initial form of discussion centres around the sentence-beginning element and debates whether it functions as a subject or a topic. The notable works are Yang’s publication from 2008, Wan’s publication from 2007, and the joint publication by Luo & Ma from 2007.

In Luo & Ma (2007), they make detailed research on the co-referential relationship between the pronoun “qí (其)” and the referred sentence-initial noun phrase topic constituent in the Shijing, the earliest song album in Chinese history. From their classification of topic and subject, they claim that since the topic and the subject are usually situated at the initial place of the sentence, and they are both noun phrases, whether the constituent is the topic or the subject should be distinguished from a pragmatic view. Their hypothesis is that the function of the subject and the topic may overlap and they are usually relying on some specific indicators to mark this topic constituent in order to gain the hearer’s attention. Based on this hypothesis, they have classified the topic indicators into two types: topics with actual indicators and topics with phonological indicators⁴ (Luo & Ma, 2007: 11). The actual topic indicators refer to the indicators that can be seen in the surface structure, and these indicators include “demonstrative pronoun topic indicators, adverbial topic indicators, and modal particle indicators⁵” (Luo & Ma, 2007: 11). As for the phonological indicators, they are the phonological edge. Their research only focuses on the topics with the demonstrative pronoun topic indicator “qí (其)”, so they do not give any further interpretation of other types of indicators.

Although the rest of the paper is discussing the referential relationship between the demonstrative pronoun and the referred topic, their analysis of the difference between topic and subject is a contribution of great value.

---

⁴ “有书面的形式标志和无书面的形式标志”。
⁵ “用指示代词表示、用副词表示、用提顿词表示”。
Different from Luo & Ma (2007), Yang (2008) discusses the differences between the topic and the subject constituent in archaic Chinese essentially from a pragmatic level. He argues that pragmatically, each sentence has its implicit structure. In other words, each constituent in a sentence has its own place. The emergence of the topic sentence is because archaic people can “break the implicit structure of the sentence and highlight the constituent, making this constitute a sentence-initial topic” (Yang, 2008: 285). “Precisely because the speaker is aiming at emphasising some element in a sentence, this behaviour is the main reason causing the differences between a syntax sentence and a sentence in a discourse in the generation procedure of an implicit structure changing to an explicit structure” (Yang, 2008: 285). His study on archaic Chinese topics is strong proof of indicating topic originally exists in the history of Chinese, and the pragmatic skill we are still using nowadays is inherited from history. Wan (2007), targets the so-called “thematic subject” and claims that the thematic subject is the topic rather than the subject. This notion is also distinct from all kinds of subjects including agent subject, recipient subject and any other kinds of subjects. Her work, same as Yang (2008), focuses on the constituent which can be either analysed as a subject or a topic. The conclusion is made that in archaic Chinese, the topic constituent has its uniqueness.

These studies have made a valuable addition to the examination of archaic topic structure, but they do have some limitations: Ma and Luo (2007) conclude that there is overlap between the subject and topic in the Shijing. However, they acknowledge the presence of sentences where the subjects and topics are identical, but they do not provide any more analysis on this particular sort of topics. Exploring this form of sentence in archaic Chinese is highly valuable for identifying the distinctions between topic and subject. According to Yang (2008), his invention involves evaluating archaic topic constituents from a pragmatic perspective. Similarly to Luo & Ma (2007), he is investigating the noun phrase that appears at the beginning of a sentence. However, in his studies, he solely focuses on the themes that have been relocated. What about phrases that lack momentum and have a theme placed at the beginning? Does archaic Chinese permit the existence of this pattern? If the answer is affirmative, what is the pragmatic analysis of this particular sentence? The solution to this query is still pending. Lastly, in Wan's (2007) study, she focuses on elements that exhibit both subject and topic functions. Research solely dedicated to investigating the fundamental components of a given issue is still pending.

6 “打破原来的句子隐形结构，把这种成分提前，使之成为句首话题”
7 “在从隐性结构到显性结构的生成过程中，正是为了强调句中某些成分，才会产生语言和言语在序列上的某些差异。”
Another area of research in the study of archaic Chinese topic constituents centres on the categorization of theme subjects and topic phrases. This particular type, as demonstrated by the research conducted by Tang in 2008, has categorized the thematic topic phrases into 13 distinct forms. The classification method is a prominent study approach in the field of Chinese linguistics.

Additional discussions pertaining to the topic constituent encompass its historical evolution (Han, 2008), the anaphoric connection between the initial constituent of a sentence and the pronoun (which may also be a null pronoun) in the comment (Yin & Liu, 2009; He, 2008; Dong, 1998), and the topic constituent within a predicate sentence (Zhang, 2005). These talks contribute to a more thorough examination of the archaic issue component.

However, the majority of researches primarily concentrate on a certain book or dynasty (Shen, 1989; Luo & Ma, 2007; Zhang, 2005), while just a few scholars examine the chronological evolution of Chinese theme elements. Although Han (2008) and Wan (2007) make an effort to describe the emergence and historical evolution of subject constituents, their timespan is very little in comparison to the extensive history of Chinese. The majority of linguists are currently engaged in discussions regarding the archaic Chinese topic phenomenon, focusing on one or more specific books. These findings are highly valuable contributions to the study. If these findings are consolidated and expanded upon, it would represent a significant advancement in the research of archaic Chinese topic constituents. Furthermore, the investigation of the ancient Chinese topic phenomenon is confronted with a crucial issue: the lack of systematic research in this field. There is no universally accepted definition for the concept of a topic, nor is there a clear consensus on the characteristics of an old Chinese topic. This uncertainty has led to numerous debates, including whether topic constituents should include clauses. Should adverbial clauses be included in the syntactic topic position? Implementing a universally accepted definition can successfully address these issues.
2.3 Left-Dislocated Sentences and Predicate Sentences

Left-dislocated sentences share comparable structures with predicate sentences, however it is crucial to clarify that they are not equivalent to topic sentences. Within this part, we shall distinguish the subsequent concepts: What are the distinguishing characteristics between left-dislocated sentences and topic-prominent sentences? Secondly, what is the predicate sentence in Chinese and how do left-dislocated sentences differ functionally from predicate sentences? In this section, we are interested in understanding the functional differences between left-dislocated sentences and predicate sentences. The objective of this section is to assert that in Mandarin Chinese, predicate sentences and left-dislocated sentences exhibit distinct characteristics despite their superficial similarities. Left-dislocation exclusively generates sentence-initial subjects rather than topics, while predicate sentences establish sentence-initial topic constituents. This section will also establish criteria for differentiating between a left-dislocated sentence and a predicate.

2.3.1 Left-Dislocated Sentences and Topic-prominent Sentences

By rearranging elements in various syntactic places, one can create sentences of different types. Left-dislocation is a syntactic structure where a noun phrase is placed at the start of a sentence and subsequently repeated with a pronoun in the main clause. This construction is frequently employed to accentuate the noun phrase or to furnish more details about it. As demonstrated in example (12):

(12) a. [That son], I have known him for five years.
   b. I have known [that son] for five years.

In left-dislocation sentences, the fronted constituent has a referential relationship to the pronoun. The verb in the sentence must agree with the fronted constituent and should agree with the pronoun. The fronted constituent’s position is called a dislocated position, and the pronoun is called a co-referential pronoun.

Although topicalisation is very similar to the left-dislocated sentences, the sentence's syntactic structure slightly changes:

(13) a. [That book], I have read for five hours.
   b. [That son], I have known him for five years.

(13a) is an example of topicalisation, and (13b) is the left-dislocation example mentioned in (13a). The difference between (13a) and (13b) is the omission of the co-referential pronoun.

The referential pronoun is not necessarily appearing in topic sentences. Take Mandarin Chinese as an example:

(14) a. 小王， 陈总 很 赏识。
xiǎowáng, chén zǒnghěn shǎngshí.
Xiaowang, Chen head very appreciate.
“Xiaowang, Head Chen appreciates him very much.”

b. [小王], 陈 总 非 赏识 [他]。

xiǎowáng, chén zǒng hěn shǎngshí [tā]。
Xiaowang, Chen head very appreciate [he].
“Xiaowang, Head Chen appreciates him very much.”

Both (14a) and (14b) are grammatically correct in Mandarin Chinese. However, the sentence-initial constituent in each sentence has different syntactic constituents. In (14a), Xiaowang functions as the topic, while in (14b), Xiaowang is the left-dislocated subject.

According to Prince (1984), left-dislocation sentences and topicalisation sentences have different prosodic peaks: in topicalisation sentences, the prosodic peak falls within the sentence-initial NP while left-dislocation sentences have a prosodic peak in the predicate expression. Topicalisation and left-dislocation sentences have functional oppositions: in the topic sentences, there exists a gap in the sentence due to the movement. In left-dislocation sentences, there does not contain any gap in the sentence. Prince (1981) has provided an insightful discussion on the difference between topicalisation and left-dislocation: topicalisation sentences are “indexed to the discourse context in ways that left-dislocation sentences are not” (Gregory & Michaelis, 2000: 1668). In Prince’s claim, the topicalisation sentences must have a content connection with the previous sentences in the discourse. However, a left-dislocation sentence does not necessarily be included inside a discourse, and it does not necessarily need to have a content relationship with the previous discourse. Taking sentence (14) as an example, sentence (14a) must appear in the discourse and it is a contrastive topic, which may indicate someone apart from Xiaowang might not be appreciated by Head Chen. In sentence (14b), the sentence can appear out of the blue.

In summary, topicalisation and left-dislocation sentences share a similar syntactic structure but differ in their function. Sentences with a topic are required in the discourse and the topic element must be connected to the preceding context. Nevertheless, left-dislocated phrases can function independently and do not necessarily need to be part of a larger discourse.

2.3.2 Chinese Predicate Sentences
In Chinese, predicate sentences are categorised based on the characteristics of the predicate. Typically, predicates in sentences use nouns and noun phrases to make judgements about the parts that come at the beginning of the sentence. These noun phrases try to provide facts about the
element that appears at the beginning of the sentence or to identify a property of that constituent.

Chinese predicate sentences can be categorised into three types: archaic Chinese predicate sentences, middle Chinese predicate sentences, and modern Chinese predicate phrases. This classification is illustrated in the example provided in (15):

(15) a. 南冥 者，天 池 也。
   nánmíng zhě, tiān chí yě.
   Nanming ZHE, sky lake YE.
   “Nanming, (is a) lake in the heaven.”
   (Zhuang Zi: Section of Inner)

b. 泉 者，地 之 血；石 者，地 之 骨 也。
   quán zhě, dì zhī xiě; shí zhě, dì zhī gǔ yě.
   spring ZHE, earth ZHI blood; stone ZHE, earth ZHI bone YE.
   “Spring, (is the) blood of earth; stone, (is the) bone of earth.”
   (Taiping Jing, Vol. 45:61)

c. 妪，先 大 母 婢 也。
   yù, xiān dà mǔ bì yě.
   Woman, before big mother servant YE.
   “The woman, (was) originally grandmother’s servant.”
   (Xiang Ji Xuan Zhi)

The examples in (15) consist of predicate sentences spanning from archaic to modern Chinese. The research on the Chinese predicate sentences emerges practically at the same period as the initial research on archaic Chinese topic constituents. In 1898, Mashi Wentong is the first book to introduce the concept of predicate constituents in Chinese predicate sentences to the general audience. Instead of focusing on predicate sentences, it is more logical to assert that Ma (1983) assigns labels to various syntactic parts within the sentence. For example, he uses “expressing words” to indicate the nouns or noun phrases being the predicate. He also uses “judging words” to refer the connection words between the subject and the “expressing words”. The “judging words” bear a striking resemblance to the copula that is currently under discussion. Ma (1983) introduces the fundamental concept of predicate elements without elaborating on the distinct predicate

---

8 表词
9 断词
sentences. Regardless, this book has offered an excellent foundation and essential framework for linguists to conduct future study on Chinese predicate sentences.

According to Ma (1983), an increasing number of Chinese linguists have directed their attention towards this particular syntactic structure. The study of predicate sentences has undergone two distinct periods: In the initial phase, linguists (Lv, 2002; Wang, 1985; Zhou, 1961) are attempting to identify the many types of phrases, including predicate sentences, that are present in Chinese declarative sentences.

Chinese grammar research is entering an innovative age in the mid-20th century. During this period, numerous scholars are proposing new constructive ideas, while also revisiting concepts that have not been thoroughly researched previously. The concept of a predicate sentence is one of the issues being discussed. In Lv's (2002) book *Brief Introduction of Chinese Grammar*, Lv (2002) divides declarative sentences into four types which are narrative sentences, epistemic sentences, predicate sentences and existence sentences (Lv, 2002: 25-69). Predicate sentences are a type of sentences which “explains what a thing is about or define whether two things are the same or not” (Lv, 2002: 55), as shown in the example in (16):

(16) a. 项脊轩，旧南阁子也。

    xiàngjǐxuān, jiù nángézǐ yě.
    Xiangjixuan, old Nangezi YE.

    “Xiangjixuan, is the (name of) old Nangezi.”

b. 鲸鱼非鱼。

    jīngyú fēi yú.
    Whale not fish.

    “Whale is not a fish.”

(Lv, 2002: 55)

According to Lv (2002), the statements in (16) are both predicate sentences. The sentence structure, exemplified by (16a), is commonly observed in old Chinese predicate sentences. The constituent following the comma provides an explanation of the subject matter of the constituent at the beginning of the sentence. In (16b), the parts preceding and after the negative *fēi* (非, not) do not

---

10 《中国文法要略》
11 “叙事句、表态句、判断句和有无句”
12 “解释事物的涵义或判辨事物的同异”
pertain to the same attribute. Lv (2002) states that the beginning part of the sentence should be considered the subject, while the latter half should be evaluated as the predicate (Lv, 2002: 55).

In the book *Modern Chinese Grammar*\(^\text{13}\), Wang (1943) categorises Chinese sentences into narrative sentences, predicate sentences, and descriptive sentences, similar to Lv's (2002) classification. In contrast to Lv's (2002) approach, Wang (1943) examines the predicate phrases from a different standpoint. He presents five categories of predicate sentences: Firstly, narrative sentences can be converted into predicate sentences by inserting a copula between the subject and the predicate and appending a *de (的)* at the conclusion of the phrase (Wang, 1943: 52). For instance:

(17) a. 这 一 所 房子 很 大。
zhè yì suǒ fángzi hěn dà.
this one CL house very big.

“This house is very big.”

b. 这 一 所 房子 是 很 大 的。
zhè yì suǒ fángzi shì hěn dà de.
this one CL house is very big DE.

“This house is a very big one.”

(Wang, 1943: 52-53)

The distinction between (17a) and (17b) lies in the fact that (17a) is the declarative sentence. This sentence can be transformed into a predicate sentence by appending the corpus *shi (是)* and *de (的)* at the end. In the year 17, as stated by Wang in 1943, the predicate phrase comprises three components: the subject, the copula, and the adverbial constituent.

The second category of descriptive sentences serves the role of a predicate sentence. "Some sentences, despite having verbs as predicates and appearing to be descriptive sentences, are actually predicate sentences" (Wang, 1943: 53), as seen in example (18):

(18) a. 马 有 四 蹄。
mǎ yǒu sì tí.
horse have four hoof.

“Horses have four hoofs.”

b. 马 是 有 四 蹄 的。

---

\(^{13}\) 《中国现代语法》
mǎ shì yǒu sì tí de.
horse is have four hoof DE.

“Horses, (they) are having four hoofs.”

(Wang, 1943: 53)

According to Wang (1943), both (18a) and (18b) can be classified as predicate sentences. While lacking a copula, (18a) nonetheless fulfills predicate functions by introducing the attribute of an object.

The third category, analogous to the preceding one but distinct, comprises sentences transitioning from descriptive phrases to predicate sentences, as exemplified in example (19):

(19) a. 他 是 留 了 胡 子 的。

tā shì liú le húzi de.

he is grow LE beard DE.

“He is (the one who is) growing a beard.”

b. 张 先生 带 着 一 个 小孩 来: 这 小孩 是 张 先生 带 来 的。

zhāngxiānsheng dài zhe yí gè xiǎohái lái: zhè xiǎohái shì zhāng xiānsheng dài lái de.

Zhang mister takeZHE one CL child come: This child is Zhang mister take come DE.

“(Mr. Zhang is bringing a child here): This child is (the one) Mr. Zhang brings (here).”

(Wang, 1943: 54)

In (19a), the original sentence is “他留了胡子 (he grows a beard)”. By incorporating a 是…的 (is…DE) structure, the statement (19a) undergoes transformation from a descriptive sentence to predicate sentences, signifying the presence of a beard on a man. In contrast to (19a), the sentence preceding the colon functions as the descriptive sentence, while the sentence after the colon serves as the predicate sentence. By placing the theme in the subject position and using a 是…的 (is…DE) construction, the descriptive sentence is converted into a predicate sentence (Wang, 1943: 54).

The fourth form of predicate sentence lacks both a subject and a copula, as exemplified by (20):

(20) 李纨 道: “好 主意!”

liwán dào: “hǎo zhǔyì!”

Liwan say: “good idea!”

“Liwan says ‘good idea’!”

(Wang, 1943: 54)
In the fourth type, the omission of the subject and copula occurs when the speaker is responding to inquiries from others. Besides, “in the sentences full of emotions, subjects and copula can also be omitted” (Wang, 1943: 54). Example (20) is an exemplary sentence that exhibits complete concurrence.

In the last type, “sometimes, when you want to emphasise some information, you can add 是…的 (is…DE) in narrative sentences. Under this situation, although the sentence seems to be a predicate sentence, it is still a narrative sentence” (Wang, 1943: 54). Nevertheless, this type of statement does not qualify as a genuine predicate sentence. They are merely adopting the structure of predicate sentences.

In 1961, Fagao Zhou categorised sentence patterns into two distinct categories: predicate sentences and declaration sentences. Zhou (1961) asserts that a sentence is considered a predicate sentence if the subject and the predicate have a semantically equivalent relationship. Semantically equal means that both the subject and the predicate are referring to the same entity.

These three studies widely agree that a predicate sentence is a significant sort of phrase in Chinese. They have a similar understanding of what a predicate sentence is: it is a statement that explains the subject of the sentence. In preliminary investigations, the initial part of a sentence is sometimes referred to as the subject. Although these are preliminary investigations, the conventional structure of predicate phrases is not standardised. Sentences (16) and (19) are not predicate sentences; rather, they are narrative and exclamatory statements. Furthermore, Wang's (1943) study lacks evidence to support the "changing" technique. If a sentence is undergoing a transformation from one type to another, there must be a rule in existence within the historical context. Nevertheless, Wang (1943) lacks an in-depth examination of this transformation.

Following Zhou’s (1961) work, research on predicate sentences has entered a new phase and has been separated into two distinct directions: One direction is examining the predicate sentences in contemporary Chinese. The contemporary Chinese predicate sentences have undergone a lengthy historical evolution and transformation, resulting in significant differences from archaic predicate phrases in both structural and semantic aspects. Yang & He (2001) have explicitly presented an

---

14 “凡在富于情感的句子里，主语和系词也都可以省略”

15 “有时候，为了要加重语意，可以在叙述句里加上‘是…的’。这种说法，在形式上是变了判断句了，但在意义上仍是叙述的性质。”
argument highlighting these disparities. Yang & He (2001) categorise predicate sentences into two types: wide-scope predicate sentences and narrow-scope predicate sentences. A statement that contains a comment or judgement about the property of the subject is referred to as a wide-scope predicate sentence (Yang & He, 2001). Regarding the narrow-scope predicate sentences, they must meet three conditions: “Firstly, there must exist an equation between the subject and the predicate; secondly, the subject should be either nouns or pronouns, and the predicate should be nouns or noun phrases; last but not least, copula should mostly be found between the subject and the predicate” (Yang & He, 2001: 705-726).

Another approach involves thoroughly examining archaic predicate statements. The research on archaic Chinese predicate sentences is more diverse and encompasses a wider range of fields compared to the research on modern Mandarin predicate phrases. Linguists primarily focus on determining whether a predicate sentence falls into the category of "noun-predicate" sentences, as classified from a sentential standpoint (Zhu, 1980; Zhang, 2002; Li, 2004). A noun-predicate sentence is a type of sentence that uses nouns or nominal structures as predicates, without the usage of verbs. An example of a noun-predicate sentence is shown in (21):

(21) a. 老 张 北京 人。
   lǎo zhāng běijīng rén.
   Old Zhang Beijing people.
   “Old Zhang is a Beijinger.”

b. 他 老师，我 学生。
   tā lǎoshī, wǒ xuéshēng.
   he teacher, I student.
   “He (is a) teacher, I (am a) student.”

The two instances in (21) pertain to distinct categories of noun-predicate phrases. (21a) is a sentence in which a noun is used as the predicate to indicate a person's place of origin. Regarding (21b), it is a statement that denotes employment. Due to the striking resemblance between the structure of (21) and (22), there is ongoing debate on whether the predicate phrase may be classified as a noun-predicate sentence:

(22) 张骞， 汉中 人 也。
    zhāngqiān, hànzhōng rén yě.

16 “（一）主谓两部分有等同关系；（二）主语多由名词、代词充任，谓语由名词或其短语充任；（三）在主谓之间多有系词联系。”
Zhangqian, Hanzhong people YE.
“Zhangqian, (is a) Hanzhong people. ”

*(Han Shu: Anecdote of Zhangqian and Liguangli, Vol. 31)*

According to several scholars, predicate sentences are synonymous with noun-predicate sentences due to the distinct definition of "predicate". When discussing a predicate, we are mostly referring to the verb in a sentence or clause. This verb typically expresses information or attributes to the topic. In Chinese traditional grammar, the term "predicate" refers to the component of the sentence that is not the subject and is in a declarative connection with the subject. Chinese predicates have a broader scope than Western predicates.

In Zhu's (1980) analysis, he argues that ancient Chinese predicate sentences serve the purpose of identifying the subject or object being referred to. Therefore, he suggests that these predicate phrases should be classified as noun-predicate sentences. Another category of academic believes that a predicate sentence is a constituent of a noun-predicate sentence. Put simply, a predicate sentence is a specific type of noun-predicate sentence.

Despite the numerous debates around archaic Chinese predicate sentences and noun-predicate phrases, they fail to address the crucial inquiry: What is the significance of the phonological stop in (22)? The confirmation of the property of predicate sentences is still pending, pending an answer to this question.

Linguists are currently placing their emphasis on the predicate in sentences from a lexical standpoint. The predicate within predicate phrases is commonly referred to as a "judgement word". The purpose of these words is to indicate that the statement in question is a predicate sentence. Adjudicative terms can be formed by using copula and actual verbs. Throughout history, linguists have engaged in ongoing debates on the presence of a copula in the Chinese language, both in ancient times and the present era. Section 5.4 will provide a comprehensive introduction to the research conducted on the archaic Chinese copula.

2.3.3 Are predicate sentences left-dislocated sentences in archaic Chinese?
Language structures that are left-dislocated contain a distinct feature, which the subjects are placed at the beginning of the sentence. A referential relationship must exist between the sentence-initial constituent and its corresponding pronominal element in the rest part of the sentence for this construction to function. The subject of the initial clause is of utmost importance in ensuring
thematic continuity and coherence; it also fulfils a fundamental syntactic function. It is crucial to acknowledge, nevertheless, that although this subject is listed at the outset of the sentence, it may not comprehensively represent the main theme of the discussion.

On the contrary, the operation of predicate sentences in Chinese is distinct. In contrast to left-dislocated sentences, predicate sentences do not require explicit referential markers or pronouns. As a result, the lack of these pronouns renders direct command relationships between the subject and any pronouns contained in the sentence unnecessary. Conditional sentences are distinguished from left-dislocated sentences by the lack of a direct correlation between the subject and pronouns. Furthermore, in the discipline of Chinese sentence construction, predicate sentences afford a degree of adaptability by permitting the independent introduction of topics, free from any particular subject or pronoun in the initial position. This phenomenon is diametrically opposed to the mandatory reference requirement observed in left-dislocated sentences, where the subject is the sentence-initial element.

To summarise, left-dislocated sentences are characterised by subjects placed at the beginning of the sentence, whereas predicate sentences contain topics in the same position as the subjects. A comprehension of these differentiations illuminates the syntactic structures and discourse organisation in both languages, emphasising the fundamental disparities in predicate and left-dislocation sentence construction.

2.3.4 Section Summary
We have reached a conclusion after a thorough examination of two sentence structures in this section: predicate sentences and left-dislocation sentences. Left-dislocation is characterised by the necessity of a referential relationship, in which the subject is represented by the sentence-initial element and not the topic. On the contrary, pronouns are not required in predicate sentences in Chinese, which eliminates the need for c-command relationships between topics and pronouns. Left-dislocation therefore generates subjects rather than topics in the majority of sentences, whereas predicate sentences establish constituents of topics in the beginning of the sentences. In this section, distinct criteria for distinguishing a predicate sentence from a left-dislocated sentence will be further established.
2.4 Classifications of Topic in Mandarin Chinese

In the preceding section, we assert that this study primarily concentrates on topics found in predicate sentences, as these types of topics are the most characteristic and authentic form among all the topic forms in ancient Chinese. Now, we will shift our attention to the categorizations of topics in Mandarin Chinese.

There are three types of topics as discussed in Frascarelli & Hinterhölzl (2007) which are contrastive topics, aboutness topics and familiar topics. Among these three types, both aboutness topics and contrastive topics can be found in Mandarin Chinese. A familiar topic is defined as “a given, d-linked constituent, which is destressed and realised in a pronominal form” (Frascarelli & Hinterhölzl, 2007: 1; Pesetsky, 1987). Since there is no distinct formal marking, either syntactic or phonological, of familiar topics in Mandarin, in this thesis we only discuss aboutness topics and contrastive topics. In Huang Li & Li (2009), as mentioned before, the topic constituent can be formed either by A’-movement or by base-generation; therefore, the syntactic properties of aboutness topics and contrastive topics are also introduced. In addition to the discussion of these two sorts of topics, this section also introduces the concept of topic chain, which is a widely utilised phenomena in Mandarin.

When examining different topics, it is equally significant to explore the antiquated versions of each topic. By examining these historical forms, we may analyse the parallels and disparities in sentence topic patterns between the old and present. As we are simulating the historical progression of topic development, the "information from the past" can offer valuable syntactic insights and aid in reconstructing the pathway of topic construction in Mandarin Chinese.

2.4.1 Aboutness Topics

Aboutness topic is a kind of topic that indicates the main theme of a sentence, and it mainly introduces contents in a sentence that is “newly introduced, newly changed or newly returned to” (Givón, 1983: 8). The comment of this sentence is related to the topic, as in the example shown in (23):

(23) 那场 大火，幸亏 消防队员 来 的 快。

Nà-chang dàhuo, xìngkuī xiāofángyuán lái de kuài.

That-CL big fire, fortunately firefighter come DE fast.

In (23), the remainder of the sentence is the comment of na chang dahu (那场大火, that fire), so na chang dahu (那场大火, that fire) is the aboutness topic of this sentence, and it is appearing in
sentence-initial position. The rest of the sentence is the comment, acting as the explanation of the topic “that fire”.

Huang, Li & Li (2009) argue that topic sentences which consist of a comment that is a complete independent clause can also be referred to as “‘gapless’ topic structures” (Huang, Li & Li, 2009: 203). In all examples (23) and (24), the comment is a grammatically correct sentence and the topic constituents remain stationary in the sentence without undergoing any movement. In sentence (24), there is a subsequent deletion, while in sentence (23), the topic of matter is the fire. Thus, both (23) and (24) possess gapless aboutness topics that are formed at their basis. A phrase with a base-generated topic follows the Top-SVO order and does not include any long-distance movement for the topic.

(24) 水果, 我 最 喜欢 [(水果 中 的) 香蕉]。

shuíguǒ, wǒ zuì xǐhuān [(shuíguǒ zhōng de) xiāngjiāo].

fruit, I most like fruit among DE banana

‘(As for) fruits, I like bananas (among fruits) most.’

(Huang, Li & Li, 2009: 203)

In archaic Chinese, aboutness topics exists, as shown in the example (25):

(25) 道 者, [万 物 之 奥 也] COMMENT。

dào zhě, [wàn wù zhī ào yě] COMMENT。

Dao ZHE, [many thing ZHI mystery YE] COMMENT.

“Dao, it is the profound mystery of many things.”

(Dao De Jing, Chapter 62)

In example (25), dao is the topic, and the comment “the profound mystery of many things” is the explanation of the topic constituent. In archaic Chinese, it is easy to find out whether the topic is an aboutness topic or not: when the comment is explaining the topic, then the topic is an aboutness topic. In example (25), there is an ellipsis in the sentence, as shown in (26):

(26) 道, 者, [(其 为) 万 物 之 奥 也] COMMENT。

dào, zhě, [(qí wéi) wàn wù zhī ào yě] COMMENT。

Dao, ZHE, [(it is) many thing ZHI mystery YE] COMMENT.

“Dao, it is the profound mystery of many things.”

(Dao De Jing, Chapter 62)
The comment part is not a full sentence and it omits it is. The it refers to the Dao in the topic. However, the aboutness topic does not always have a gap in the sentence in archaic Chinese. In sentence (27), there is no gap:

(27) 五谷之者，[种之之美者也]

wǔ gǔ zhě，[zhòng zhī měi zhě yě]

Five cereal, ZHE, [plant it glorious ZHE YE]

“Five kinds of cereals, planting these are the most glorious thing (comparing to other cereals).”

(26) 不谷，不也。

bù gǔ，bù yě.

Not ZHE, not ye.

“Not, is (as a kind of) not.”

(Meng Zi, Gaozi: part one)

Different from (26), there is no gap in (27). The comment in (27) is explaining the topic “five kinds of cereals”. In (27), there is a referential relationship between the anaphor zhì and the antecedent “five cereals”. To conclude, in archaic Chinese, either existing a gap or does not exist any gap in the sentence both can form a topic-comment structure.

The aboutness topic has a very early and long history: the earliest aboutness topic structure can be traced back to Shang dynasty (from 1046BCE), as in the example shown in (28):

(28) 佑者，助也。

yòu zhě，zhù yě.

Bless ZHE, help ye.

“Bless, is (as a kind of) help.”

(Zhouyi: Ci Xi Shang Zhuan, Vol. 12)

Early in 1046BCE, sentences containing aboutness topics can be found from Zhouyi, the earliest written document found in archaic Chinese. Since the book is introducing many new information on Chinese era, the functions of containing these aboutness topics are: by fronting these newly invented words to the initial place of the sentence and adding a break between the topic and comment constituent, the reader will notice that the topic constituent before the break is the newly-introduced important information. From Zhou Dynasty, the aboutness sentence structure is one of the most useful way to introduce new information.

As for frame setting, it is also a related notion to aboutness topics. Syntactically, they are sharing the same position in the sentence; however, there are differences between these two notions. Frame setting is aims to “limit[ing] the applicability of the main predication to a certain restricted domain” (Krifka, 2007: 46). Different from aboutness topic, frame setting tends to add a restriction to the existing sentence, because “if there is no alternative perspective to be considered, then there is no need for an explicit frame setter either” (Krifka, 2007: 46). Consider frame setting examples in (29):
(29) A: How is business going for Daimler-Chrysler?
B: [In GERmany]Frame the prospects are [GOOD]Focus,
     but [in AMErica]Frame they are [losing MOney]Focus.

(Krifka, 2007: 45)

In (29), both [in Germany] and [in America] are frame setters, which have the syntactic status of adjuncts. If the frame setters are omitted, the rest of the sentence is still grammatical and it does not lack any meaning. However, as for aboutness topics, if the topic is omitted, the rest of the sentence is still grammatical but the semantic meaning is not complete because the core part is missing. Gasde (1999) is classifying the Mandarin topic constituent as aboutness topics and frame-setting topics, in this thesis, we are not going to adopt this classification because of the restrictions on frame-setters.

In Mandarin, there also exist frame setting sentences, as in the example shown in (30):

(30) A:  肯德基 在 中国 的 业绩 怎么样?
kēndéjī zài zhōngguó de yèjì zěnmeyàng?
     KFC at China DE business how?
     “How is the business going for KFC?”
B: 北京 肯德基 业绩 很好, 长春 肯德基 在 赔 钱。
běijīng kēndéjī yèjì hěn hǎo, chángchūn kēndéjī zài péi qián.
     Beijing KFC business very good, Changchun KFC is losing money.
     “In Beijing, the KFC business is good, but in Changchun, it is losing money.”

(29) and (30) share much similarity because Beijing and Changchun are frame setters, just like the setters in (29). These setters add geographical restrictions to the existing sentence, and when omitting these setters, the rest of the sentence is grammatical. However, the frame-setting phenomenon is not found in archaic Chinese, it is newly introduced to Chinese syntactic structure.

Therefore, to conclude, the relation between frame setting and the aboutness topics are: Frame setters are able to be omitted and they aim to indicate alternatives of the sentence, the newly-introduced aboutness topics cannot be omitted and it does not need to indicate the alternative of the sentence. When the aboutness topic appears, as long as the comment is giving the explanation or introducing the topic, the topic is the aboutness topic.
2.4.2 Contrastive Topics

Contrastive topic is a kind of topic that shows contrast with the content mentioned before. In discourse, contrastive topic either has the pitch accent of L* + H, which has a low accent before rise, or L + H*, which has a low accent but ending with a high accent, depending on where the topic constituent is in a sentence. This prosodic feature is not specific to Italian and German, but also in Mandarin Chinese (Féry, Ishihara, Chen, Lee & Pan, 2016). In Mandarin, contrastive topics can be found, but are not necessarily presented in sentences. There is no fixed position in the sentence for contrastive topics, as in the example shown in (31):

(31) 

a. **他, 我 都 不 想 理。**
   
   tā, wǒ dōu bù xiǎng lǐ.

   He, I even not want care.

   “(As for) Him, I do not even want to care about (him)”

b. — 现在 的 演员 都 穿 什么?
   
   Xiànzài de yǎnyuán dōu chuān shénme?

   Now DE actors all wear what?

   — 女演员 都 穿 晚礼服。
   
   Nǚyǎnyuán dōu chuān wǎnlǐfú.

   Actress all wear evening dress.

-“What do all the actors wear now?”

-“Actresses all wear evening dresses.”

(31a) and (31b) shows different kinds of in-sentence contrastive topics in Mandarin. In (31a), the topic is he (‘他’). When saying “Him, I do not even want to care about”, there is a contrast between “him” and “other people apart from him”. However, both “him” and “other people”, they share a similarity of “I want to care about”. Therefore, he here is providing a contrast. In (31b), when the hearer answers “all the female actors are wearing evening dress“, this reply contains another layer of meaning of “do not have more information on what male actors wear“, which also constructs a contrast. Frascarelli and Hinterhölzl (2007) point out that contrastive topic has “no impact on the focus value and creates oppositional pairs with respect to other topics“ (Frascarelli & Hinterhölzl, 2007: 1).

In archaic Chinese, the contrastive topic exists, as in the example shown in (32):

(32) 云荣 者, 与 屋 为 荣 饰; 言翼 者, 与 屋 为 翅翼 也。
In the sentence above, the topic constituents are *yunrong* and *yanyi*. The comments appearing after the topics aim at explaining the functions of both *yunrong* and *yanyi*. Both these constituents has form a comparative pair: these two parts belongs to a part of the room and they have different functions in the room decoration. By using the contrastive topics, the functions of these two items can be compared in a very obvious way.

The history of contrastive topic structure also can be traced back from 1046BCE from the same book: *Zhouyi*, as in the example shown in (33):

(33) 元 者, 善 之 長 也; 亨 者, 嘉 之 会 也; 利 者, 义 之 和 也; 贞 者, 事 之 干 也。

“Yuan, is the ancestor of the goodness; Heng, is the meeting of all the glorious things; Benefit, is the sum of the justice; Firmness, is the basic of all the things.”

(34) 天 地 之道, 贞 观 者 也; 日 月 之道, 贞 明 者 也…

Sky earth ZHI rule, justice see ZHE YE; sun moon ZHI rule, justice bright ZHE YE…
“The (development) rule of sky and earth, is the observable justice; the (development) rule of sun and moon, is the observable bright…”

(Zhouyi: Ci Xi Xia Zhuan)

Different from (33), although (34) can also be separated into different sentences, each contains an aboutness topic, the two topics “the rule of sky and earth” and “the rule of sun and moon” are making a contrast: They all belong to a natural phenomenon which has their innate features — the sky and earth follows a certain timing rule to change, from day to night, from spring to winter; and the rule of the sun and the moon refers to the rise and fall, which creates day and night. Both of these contrasts are obvious and undeniable, and the contrastive topics aim at indicating the observable feature of these natural elements. However, the syntactic topic structure in (34) is a NP+ZHI+NP constituent rather than the sole NP constituent, therefore, even if in Shang Dynasty, sentence (8) exists, it is not meeting the topic constituent we are researching in this thesis, this kind of sentence is not taken into consideration.

Therefore, all the contrastive topics in Zhouyi are the simple overlay of sentences containing aboutness topics such as example (33). In the later period of the history, the feature of contrastive topics gradually changes: the constituents being compared must have some features in common.

2.4.3 Topic Chain

Topic chain is a kind of discourse structure in which the same topic appears in a set of sentences. There are two types of topic chains: anaphoric and cataphoric. As for the anaphoric topic chains, a topic constituent can be found in sentence-initial position. The form of this sentence can be shown below:

\[
[\text{Top}\ldots ]\text{clause containing topic, } [\emptyset\ldots ]\text{null topic clause 1, } [\emptyset\ldots ]\text{null topic clause 2, }\ldots
\]

In this structure, \(\emptyset\) represents the null topic in the clause. This topic is coreferential to the following \(\emptyset\) constituents. Li (2004) brought out three features about topic chains in the previous studies: firstly, the topic can always be found at the beginning of the sentence; secondly, zero anaphora in the following clauses which have coreference to the topic in the main clause can be found; third, well-formed sentences can participate in topic chains (Li, 2004: 29), which means that the sentence with topic chains is grammatical and does not lack any semantic meaning. However, his findings show that there exists another type of topic chain, which is cataphoric.

A cataphoric topic chain has the feature of not containing the topic in the first clause, “they can occur in the middle of a chain, with a coreferential cataphor in each of the preceding links” (Li, 2004: 30), as in the example shown in (35):
(35) 说到这，她不禁流下泪来，

shuō dào zhè, tā bù jīn liú xià léi lái,
Say until here, she not control drop down tear come,

想起了以前的事情。

xiǎng qǐ le yǐqián de shìqíng.
Think up LE past DE things.

“When saying so, her tears began to fall down and she started to remember the past.”

In example (35), the first clause does not contain any topic, but the topic appears in the next clause
(she ‘她’), and the last clause still follows the previous clause, using she (‘她’), as the topic to
continue the sentence. The existence of this type of sentence is in contradiction to the features
mentioned before: topics can appear in the middle of the sentence. Therefore, the form of this type
of sentence can be shown below:

\[ [\emptyset \ldots ] \text{null topic clause 1, } [\text{Top}\ldots ] \text{clause containing topic, } [\emptyset \ldots ] \text{null topic clause 2, } \ldots \]

It is worth mentioning that there is not a standard restriction on where the topic may appear, it can
appear anywhere in the middle of the sentence. This thesis is only focusing on clause-initial topics,
therefore this form is not analysed.

When containing double topics in the sentence, the coreferential system seems very ambiguous, as
in the example shown in (36), taken from Li (2004):

(36) 那辆车价钱太贵，颜色也不好，我不喜欢 Øj,

Nà-liàng chē jiàqián tài guì, Øj yánse yè bù hǎo, wǒj bù xǐhuān Øj,
That-CL car price too high, color also not good, I not like,
Øj不想买Øi. Øj昨天去看了一下Øi, Øj还开了会 Øi,
Øj bù xiǎng mǎi Øi. Øj Zuótiān qù kàn le yìxià Øi, Øj hái kāi le yīhuìer Øi,
not want buy. Yesterday go see PRT-a-bit, also drive PRT-a-while,
Øj还是不喜欢 Øi.
Øj hǎishi bù xǐhuān Øi.

“那辆车太贵，颜色不好。我不喜欢它。昨天我去看了下它，我开了会。但我还是不喜欢它。“

(Li, 2004: 1)
A sentence like (36) is a combination of anaphoric and cataphoric topic chains. *Naliangche* (那辆车) is acting as the anaphora sentence-initial topic while *wo* (我) is the cataphoric topic. This double coreference phenomenon requires the hearer to put the correct topic in the corresponding null place.

Topic chain can be regarded as an another kind of aboutness topic structure because all the comment sentences are describing the topic which is put at the initial place of the sentence. Therefore, the topic constituent in topic chain sentences are also base-generated.

Interestingly, topic chain structure is not found in archaic Chinese. One of the main explanations might be: the information density in archaic Chinese is very high, in other words, in archaic Chinese, using a few words can express many meanings. The high density makes the sentence very short in archaic Chinese. If adding on the omission of topic constituent and form a topic chain, it will cause more ambiguity in the sentence. Therefore the topic chain structure is not found from the early period of history.

2.4.4 Section Summary

To sum up, in this part, we classify the topics based on the criteria of Frascarelli & Hinterhölzl (2007). There are two types of topics exist in mandarin Chinese, which are contrastive and aboutness topics. The existence of aboutness topics can be found to the very earliest stage — Shang Dynasty (1046BCE). Although the contrastive topic can also be found at the same stage in the history, the so-called “contrastive” is adding up several aboutness topic sentence together to form a contrast phenomenon. Throughout development, the contrastive topic sentence begin to use constituents which have inner connections to make contrast. As for the topic chain, this is a newly emerged phenomenon in the later part of the history.
2.5 Topic Marker

After discussing the notion of the topic and left-dislocated topic constituents, in this section, we are going to turn our focus on the topic markers. A very interesting coincidence is that both Western and Oriental linguists started the analysis of topic markers by studying the sentence-initial discourse markers. Due to the different syntactic properties of these languages in these two different areas, the research on the discourse markers went in two different directions throughout the research process. To discuss the formation and development of the topic marker, this section is divided into the following parts: The first part is giving the definition of the topic marker and briefly introduce the important research on this notion. In the second part, we are going to discuss the classifications of topic markers. In the last part of this section, we are going to introduce different topic markers in Chinese, either in Modern Chinese, dialects or archaic Chinese.

2.5.1 What is a topic marker?

Topic marker, a kind of marker which indicates that the constituent before it is a topic, has experienced different stages of study from the late 20th century. In the early stage, when the notion of topic marker is not separated from the discourse markers, linguists (Levinson, 1983; Schourup, 1985; Schiffrin, 1987; Blakemore, 1987) are discussing the notion of discourse markers from a pragmatic perspective: Levinson (1983) first systematically discussed the existence of topic marker. In his work, he regarded topic marker as “the relationship between an utterance and the prior discourse” (Levinson 1983: 87). The discourse markers he mentions are the sentence-initial markers including words like “all in all, however, anyway” and so on. The following studies including Schourup (1985), Schiffrin (1987), Blakemore (1987) and Fraser (1990) are all targeting on the sentence-initial discourse markers and provide further analysis of the properties of these markers, but rarely of them discuss the topic marker in English. Why the topic marker is being less analysed through history has its own reasons.

English, being a subject-prominent language, has abandoned the topic marker throughout history. In the early stage of English, Old English, compared to Modern English, is more inclined to be a subject-prominent language with more topic-prominent features. According to Chapter 4 of *The Cambridge history of the English language Vol.1*, there are three ways to distinguish a topic in an old English sentence: First, the most observed method for topicalising a component is to prepose the constituents to the initial place of the sentence, as shown in (37).

(37) Seo ilce burg Babylonia, seo þe maest wæs & ærest ealra burga, seo is nu læst

That same city Babylon, REL PT greatest was and first of-all cities, it is now least & westast.
and most-deserted.

“As for that same city of Babylon, which was the greatest and first among all cities, it is now the least and the most deserted.”

(Hogg 1992: 279 (Or 2 4.74.22))

In (37), different from archaic Chinese, there is a resumptive pronoun seo ‘it’, referring to Babylonia, as shown in (38). This kind of resumptive pronoun can hardly be discovered in archaic Chinese, or even in Mandarin we use nowadays.

(38) Seoilce burg Babylonia, seo þe mæst wæs & ærest ealra burga, seo, is nu læst & westast.

Second, Old English has a freer word order compared to Modern English because the constituents are marked by different cases. The way how the speaker wants to illustrate the topic is to change the proposed constituents into the nominative case, as shown in (39). But this method is not frequently observed. There is also a generic subject pronoun man in (39), which appears twice in the sentence. Due to the demonstrative of ea ‘the river’ is in the form of dative, therefore it is not proper to say that this sentence is similar to Modern English which has a contrastive topic which has a contrastive topic phenomenon.

(39) Þa land (NOM) þe man hæt Gallia Bellica, be eastan þæm (DAT) is sio

Those lands PRO calls Gaul Belgic, at east to-them is ART

ea þe man hæt Rin.

river PT one calls Rhine.

“Those lands which are called Belgic Gaul, east of them is the river called Rhine.”

(Hogg 1992: 280 (Or 1 1.22.22))

Third, in Old English, there is also a sentence structure “(It) is...Ƿœt”, which is very similar to the modern English structure “It is...that”. This structure is also used in Old English but not very frequently, as shown in (40).

(40) Þæt is laðlic lif þæt hi swa maciað.

That is loathsome life that they thus make.

“It is a loathsome life that they thus create.”

(Hogg 1992: 280 (WPol 2.1.1 (Jost) 183))

Since English is not a topic-prominent language from the very beginning, the study of topic marker begins from the study of other languages’ topic markers. From the 1960s, studies of topic markers emerge. In 1969, Hoffer (1969) makes a comparison between the Okinawan and Japanese, later in 1970, Song (1970) studied Korean and discusses the topic marker in such language. Until Li & Thompson (1976), the basic properties of languages are finally being discussed systematically.
Languages can be separated into two types: topic-prominent and subject-prominent. Subject-prominent means that a language place the subject at the initial position of the sentence and the word order of the language usually follows the SVO order (Li & Thompson, 1976), as in the example shown in (41):

(41) I love linguistics.

SBJ

In (41), the constituent “I” is a subject, and it is usually placed at the very front of the sentence. English is a typical subject-prominent language. Different from subject-prominent languages, topic-prominent languages refer to a kind of language which topic is found at the initial place of the sentences, as in the example shown in (42):

(42) a. 书， 我 喜欢。

shū， wǒ xiān.  

Books, I like.  

TOP

“Books, I like.”

b. 私 は 雪 が 白い と 思う

watashi wa {yuki ga shiroi} to omou.

I TpM snow SbjM white when think.

“I, think that the snow is white.”

The constituent “books” and “Watashi” are the topics in the example (42) and it is put at the initial place of the sentence. Discussed by Li & Thompson (1976), Chinese is a topic-prominent language, so are Japanese and Korean. The difference between Chinese and Japanese is that in Japanese, there usually exists a certain kind of particle word immediately following the topic constituent, in (6b), “wa” is one of these particle words. These kinds of particle words are called topic markers. Li & Thompson (1976) can be regarded as the conclusion and promotion of all the previous studies.

When western linguists are exploring the differences between topic-prominent languages and subject-prominent languages, Chinese linguists are also studying different types of topic makers at the same pace. Similar to the research pathway of discourse markers in English, starting from the 20th century 60s, the study of Mandarin topic markers also starts from the discourse markers including guanyu (关于, “about”) and duiyu (对于, “as for”). Lv & Zhu (1952) and Ding (1961) are the earliest scholars who study the function of these two discourse markers. However, in their analysis, they put their focus on the binding relationship between the verb and the preposition in the
sentences containing *guanyu* and *duiyu*. Their work has started a new era of observing the relationship among the sentence-initial discourse markers, verbs and the preposition behind the verb. In Liao (1986), different from the previous studies, he first discusses the relationship between these discourse markers and the topic constituent. Liao (1986) claims that *zhiyu* (至于, “as for”) has the function of transferring a topic to another different topic. Later in 1987, Zhang (1987) points out that the preposition phrase *guanyu* (关于, “about”), *zhiyu* (至于, “as for”) and *duiyu* (对于, “as for”) can be regarded as the topic markers.

A topic marker is a syntactic constituent which is aiming at indicating the proceeding constituent is the topic constituent. The existence of the topic marker is especially marking the topic constituent, as in the example shown in (43):

(43) 我呢，今天吃米饭。

wǒ ne, jīntiān chī mǐfàn.
I NE, today eat rice.
“(As for) I, eat rice today.”

The function word *ne* is the topic marker in this sentence, indicating that the constituent before that is the topic rather than the subject in the sentence. It acts as a prosodic edge and separates the topic and the comment apart. However, not all topic-prominent languages have topic markers: Japanese and Korean have the specific topic markers but Mandarin Chinese does not (see section 5.5.1). Hence, regarding this difference, the classification of topic-prominent languages can be divided into “topic-prominent languages with overt topic markers” and “topic-prominent languages with covert topic markers” (Zheng, 2017). Mandarin can be regarded as language with overt topic markers. In the next section, previous studies on Mandarin topic markers will be introduced.

2.5.2 Classifications of Topic Markers

Discussed in the previous part, the initial research on topic markers are targeting on the sentence-initial discourse markers which are *guanyu* (关于, “about”), *zhiyu* (至于, “as for”) and *duiyu* (对于, “as for”):

(44) a. 关于这件事，我有话要说。

guānyú zhè jiàn shì, wǒ yǒu huà shuō.
DM this CL thing, I have speech say.
“About this thing, I have some words to say.”
b. 至于 这 件 事, 我 处理 完 了。
zhìyú zhè jiàn shì, wǒ chǔ lǐ wán le.
As for this thing, I have already dealt with it.

c. 对于 她, 我 无 话 可 说。
duìyú tā, wǒ wú huà kě shuō.
As regards her, I am speechless.

The topic marker 《关于》(“about”), 《至于》(“as for”) and 《对于》(“as for”) in example (44) are all located at the sentence-initial position, and they can only be situated at the sentence-initial position. Sentences in (45) should be judged as ungrammatical:

(45) a. *这 件 事 关于, 我 有 话 说。
*zhè jiàn shì guānyú, wǒ yǒu huà shuō.
*About this thing, I have some words to say.

b. *这 件 事 至于, 我 处理 完 了。
*zhè jiàn shì zhì yú, wǒ chǔ lǐ wán le.
*As for this thing, I have already dealt with it.

c. *她 对于, 我 无 话 可 说。
*tā duì yú, wǒ wú huà kě shuō.
*For her, I am speechless.

Towards this kind of topic marker, linguists (Liao, 1986; Zhang, 1987; Wang, 1992) name them “prepositive topic markers”, or more specifically, “prepositive prepositional topic markers”, because these are prepositions. Once these topic markers are found and shown to the public, different opinions begin to arise. Supporters like Wang (1992) are researching into details about these prepositive topic markers: In his paper, he makes a very thorough study about the 《duiyu》 marker and believes that 《duiyu》 has the function of comparing the topics. Yuan (2002) and Zhang & Chen (2006) all believe that the prepositive topic marker should be regarded as topic marker and made a brief discussion towards these function words. However, there exist opposite voices. Scholars like (Yuan, 2002; Xu & Liu, 1998) strongly believes that sentences like (44) can only be
interpreted as prepositional-fronting sentences, they do not have the property of being a typical topic sentence.

Lei (2001) is the first one who attempts to classify the types of all the topic markers. In Lei (2001), she briefly classifies Mandarin topic markers into six types, which are: modal particles, prepositions, verbs, pronouns, conjunction words and pause words (Lei, 2001: 225): Modal particle topic markers include *ne* “呢”, *ma* “嘛”, *ya* “呀”, *li* “哩”, *na* “哪”, *a* “啊”, and *me* “么” (Lei, 2001). They appear immediately behind the topic constituent, as shown in the example in (46):

(46) a. 这 件 事 嘛，不 好 解决。

zhè jiàn shì ma, bù hǎo jiějué.  
This CL thing MA, not good solve.
“As for this thing, (it is) not easy to solve.”

b. 他 呢， 明天 去 纽约。

tā ne, míngtiān qù niǔyuē.  
He NE, tomorrow to New York.
“As for him, (he is going) to New York tomorrow.”

As for prepositions, topic markers include *xiang* (“像”, like), *zhìyu* (“至于”, as for), *zai* (“在”, at), etc. (Lei, 2001). Different from the modal particle topic markers, the preposition topic markers appears before the topic constituents, as shown in the example in (47):

(47) a. 像 这 件 事，不 好 解决。

xiàng zhè jiàn shì, bù hǎo jiějué.  
Like this CL thing, not good solve.
“Like this thing, (it is) not easy to solve.”

b. 至于 他， 明天 去 纽约。

zhìyú tā, míngtiān qù niǔyuē.  
As for he, tomorrow to New York.
“As for him, (he is going) to New York tomorrow.”

Verbs being topic markers include *shuodao* (“说到”, talk about), *tándao* (“谈到”, discuss about), *tígí* (“提起”, mentioning), *you* (“有” have), etc. (Lei, 2001). The syntactic position of these verbal topic markers are the same as prepositional topic markers: they can be found at the initial place of the sentence, before the topic constituent, as shown in (48):
(48) a. 说到这件事，可不好解决。

shuōdào zhè jiàn shì, kě bù hǎo jiějué.
**Talk about** this CL thing, *may not* good solve.

“Talking about this thing, is not easy to solve.”

b. 提起他，我记得他明天去纽约。

tíqǐ tā, wǒ jìde tā míngtiān qù niǔyuē.
**Mention** he, I remember him tomorrow to New York.

“Mentioning him, I remembered that he is going to New York tomorrow.”

Pronouns can also be topic markers, as discussed by Lei (2001). The pronouns which can be topic markers are: *shenme* (“什么” what), *zhege* (“这个”, this) (Lei, 2001).

(49) 这个事，可不好解决。

zhè gè shì, kě bù hǎo jiějué.
**this CL** thing, *may not* good solve.

“This thing, is not easy to solve.”

In example (49), the constituent “thing” after *zhege* is the topic.

The last type, pausing, has a strong relationship with punctuation marks and prosody edges. Example (50) shows the pausing topic marker in Chinese:

(50) 小王，明天去纽约。

xiǎo wáng, míngtiān qù niǔyuē.
Little Wang, tomorrow go NewYork.

“Wang, is going to New York tomorrow.”

In example (50), the punctuation marker acts as the phonological stop and topic marker of this sentence. The existence of such a pause indicates the constituent before the marker is the thing speaker wants to illustrate. Using the pause in topic sentences is one of the popular ways of creating a topic sentence in Mandarin Chinese.

The work of Lei (2001) can be evaluated as the most objective topic marker classification. However, she does not provide a detailed discussion on why these words can be the topic marker. This classification can be regarded as a summary of previous studies and development of the existing claims and received much support among Chinese linguists.
However, it is hard to strictly define the notion and the types of “topic markers” because Chinese does not have a strict definition of inflection and affixes. Thus, how to define the notion of a “topic marker” is still under much discussion. In a study by Deng (2009), he argues that topics are marked by two types of markers: lexical markers and phonological markers by studying different types of topic markers in TV talking shows.

In lexical topic markers, prepositions (or prepositional phrases), verbs and pronouns are three types of topic markers:

(51) 对于那些打牌什么我没意见，但对于看球我还是反感...

duìyú nàxiē dǎ pái shénme wǒ méiyìjiàn, dàn duìyú kàn qiú wǒ háishi fānggǎn…

As for those play cards what I no opinion, but as for watch football I still dislike…

“As for those who play cards, I do not have any opinion, but for those who watch football, I dislike them…”  

(Deng, 2009: 45)

Prepositions like duiyu (“对 于”, as for) in example (51) are prepositions which acts like topic markers. The functions of the prepositions are introducing the topic, in Deng (2009), he names this kind of introduction as “preposition introduction” (Deng, 2009: 45). As for the verbal topic markers, Deng (2009) discusses two types of verbs which are discourse verbs and feeling verbs:

(52) a. 说到王麻子刀剪的 命运，我认为和他创新 能力 的

shuōdào wángmázi dāo jiǎn de mìngyùn, wǒ rènwéi hé tā chuàngxīn nénglì de

Talking about Wangmazi knives scissors DE destiny, I think and he innovative ability DE

减弱 直接 有关 的。

jiǎnruò zhǐjiē yǒuguān de.

decrease directly correlate DE.

“Talking about the destiny of Wangmazi knives and scissors, I think it is directly correlating to the decreasing innovative ability.”

b. 所以我 觉得日本国民 啊，如果他有一些人是有良知的…

suǒyǐ wǒ juéde rìběn guómín a, rúguǒ tā yǒu yīxiē rén shì yǒu liángzhī de…

So I think Japan people A, if he have some people is have intuition DE

---

17 介引 (Deng, 2009: 45)
18 言谈类动词 (Deng, 2009: 46)
19 感知类动词 (Deng, 2009: 46)
“So I think Japanese people, if some people has the intuition...”

(Deng, 2009: 46)

In example (52a), the verb *shuodao* (“说到”, talking about) is the discourse verbs and the *juede* (“觉得”, think about) is the feeling verbs. Deng (2009) does not provide the judgement of why these verbs should belongs to these types but he mentions that these topic markers are typical in the discourse rather than in the written texts.

In the pronoun topic markers, Deng (2009) puts up a combination of [pronoun topic marker + topic constituent], as in the example shown in (53):

(53) 你 看 这个 学生， 这个 研究生…

nǐ kàn zhège xuéshēng, zhège yánjiūshēng…

you see this student, this master student…

“You see this student, this master student…”

(Deng, 2009: 47)

In example (53), zhège (“这个”, this) is the topic marker and Deng (2009) discusses it as an oral topic marker: in the written texts, the usage of zhège is forbidden because the written texts should be precise in words and being rigorous.

Regarding the phonological topic markers, Deng (2009) classifies them into three types: modal particles, phonological stops and repeating the topic constituents. The modal particles he discusses are *a* (“啊”) *ba* (“吧”) and *ne* (“呢”). These topic markers can be regarded as the high-frequency usage of topic markers because in the live shows, there exists many on-the-spot performances. Deng (2009) can be regarded as the development of Lei (2001) because he brings a more finalised and a more comprehensive study on the types of topic markers and distinguishes the topic markers from oral markers and written markers, which can be used as discussing the functions of topic marker in the later part of this section.

2.5.3 Topic Markers in Chinese

In the previous sections, we have argued that the research on Modern Chinese topic markers relies on the properties of punctuation, word order and stress. Although these are important methods of creating a topic sentence, topic markers also exist in forming topic-prominent sentences. To summarise the typical topic markers linguists have researched and analysed in Chinese, this section
is divided into the following parts: topic markers in modern Chinese; topic markers in dialects; and topic markers in archaic Chinese. Afterwards, the common problems of the topic marker research are raised in the fourth section. The last part is the conclusion.

2.5.3.1 Topic Markers in Modern Chinese
In modern Chinese, topic markers can mainly be divided into three types, according to different studies (Zhao, 1968; Zhang & Fang, 1996; Li & Thompson, 1989; Qu, 2006; etc.): sentence-initial discourse markers, modal particle topic markers, and prepositional topic markers (or prepositional phrases). The following topic markers will be introduced in detail in this section: sentence-initial discourse markers include duiyu (“对于”, as for), zhiyu (“至于”, as for), and shuo (“说”, say); modal particle topic markers include a “啊”, ba “吧”, ma “嘛”, and ne “呢”; prepositional topic markers include zai (“在”, at) and xiang (“像”, like).

1) Sentence-initial Discourse Markers
Since sentence-initial discourse markers are an important way to form topic markers, they are receiving much attention in topic marker research. In research, these discourse markers are known as “preposition introductory function words”. With the existence of these prepositions, the topics are introduced and the hearers will know that the following contents are the main theme in the discourse. With the inner research of these sentence-initial topic markers, researchers (Deng, 2015; Zhang & Chen, 2006; Deng, 2009) further divide these sentence-initial markers into different types including prepositions, verbs, and modal particles. In this section, the following markers are introduced: duiyu (“对于”), zhiyu (“至于”), and shuo (“说”).

i. duiyu (About, “对于”)
The prepositional marker duiyu is one of the earliest prepositional markers analysed as a topic marker. In Wang (1992), he made discussion on the four pragmatic features of duiyu, and one of the features is the comparative feature. Based on this study, researchers (Zhang, 2012; Fei, 2010, Deng, 2015) agree that duiyu is a topic marker which can introduce contrastive topics:

(54) 对于这事，我有话想说。
   duiyú zhè jiàn shì, wǒ yǒu huà xiǎng shuō.
   About this CL thing, I have word want say.
“About this thing, I have something to say.”

The reason why *duiyu* can become a topic marker is that it can introduce either old or new information to the sentence (Deng, 2015: 97), and the existence of this marker can let the hearer know that the discourse has not finished yet. It has the function of “textual cohesion” (Deng, 2015: 97).

ii. *zhiyu* (As for, “至于”)

The similarities between *duiyu* and *zhiyu* are they both can appear at the initial place of the sentence, and they all have the textual cohesion function. In other words, when having these markers at the initial place of the sentence, the hearer knows that the discourse has not finished yet and the topic is continuing. Different from *duiyu*, the usage of *zhiyu* is indicating that there is a turning point of the topic in the discourse:

(55) *zhìyú zhè jiàn shì, wǒ yǒu huà xiǎng shuō.*

As for this CL thing, I have word want say.

“As for this thing, I have something to say.”

Different from example (54), if example (55) appears in the discourse, it means that “this thing” was not the initial topic between the speaker and the hearer and the speaker chose to discuss “this thing” as a new topic in the discourse. However, it is worth mentioning that the initial topic must have a relationship to the new topic, otherwise, it is flouting the “textual cohesion” feature.

The differences between *zhiyu* and *duiyu* as topic markers are: *zhiyu* can “indicate the compare relationship”, and *duiyu* can prominent the relationship between the topic and the comment in a sentence (Zhou, 2012).

iii. *shuo* (Say, “说”)

Among all the verbal topic markers, *shuo* is one of the most typical markers under analysis, because any *shuo* appearing in the sentence-initial place can act as a topic marker function. The syntactic position of the topic-marker *shuo* is very different from the verbal *shuo*.

---

21 “语篇衔接”
Researchers (Liu, 2005; Liu, 2007; Qiang, 2009) conclude five types of *shuo* are topic markers: the first type of *shuo* refers to the sole character *shuo*. The existence of sentence-initial *shuo* is containing the function of a topic marker, as shown in the example (56):

(56) 说 小王，他 还 没 来。

shuō xiǎowáng, tā hái méi lái.

Say Little Wang, he yet not come.

“Say Little Wang, he has not came yet.”

To make sentence (56) reasonable and understandable, the topic marker *shuo* must exist in the discourse and the previous content must contain *Xiaowang*. Otherwise, if the previous content does not contain the topic constituent, it is not acceptable to contain the “*shuo* + topic” structure in the following sentence.

The second type of *shuo* is “verb + *shuo*”. The most typical marker under analyse is *birushuo* (“比如说”, for example say):

(57) 比如 说 小王，他 还 没 来。

bǐrú shuō xiǎowáng, tā hái méi lái.

For example say Little Wang, he yet not come.

“For example, say Little Wang, he has not came yet.”

In sentence (57), the combination of “for example” and *shuo* forms a topic marker expression. This expression contains two functions which are topic marking and making a hypothesis.

The third type of *shuo* is “adv + *shuo*”. The typical marker is *zaishuo* (再说):

(58) 再说 这 件 事，本来 就 是 小 王 的 责 任。

zàishuō zhè jiàn shì, běnlái jiù shì xiǎowáng de zérèn.

Besides, this CL thing, originally just is Xiaowang DE responsibility.

“Besides, this thing, it is originally Xiaowang’s responsibility.”

In sentence (58), the topic marker *zaishuo* is retrieving the previous contents. In other words, *this thing* should and must have appeared in previous contents. Otherwise, this sentence is not acceptable in the discourse. The “adv + *shuo*” combination is considered as one of the topic markers including the verb *shuo*.

The fourth type of *shuo* is “conjunction + *shuo*”, for example *ruguoshuo* (如果说, “if say”):
The usage of *ruguoshuo* is similar to the previous topic marker. The topic marker *ruguoshuo* is aiming at making a contrast to the previous content. In sentence (59), the topic *Xiaowang* is not “this kind of person”, which has a contrast with the person who has a certain bad personality. The existence of *ruguoshuo* is an assumption and making a hypothesis of the topic constituent’s property.

The last type of *shuo* is “*shuo + preposition*”, and the most typical marker is *shuodao* (说到, talking about). Similar to *shuo*, the verb *shuodao* is used as a sentence-initial topic discourse marker, indicating the component after the speaker stresses this verb:

(60) 说到这件事, 本来就是小王的责任。

*shuōdào zhè jiàn shì, běnlái jiù shì xiǎowáng de zérèn.*

“Talking about this thing, it is originally Xiaowang’s responsibility.”

In sentence (60), when adding the *shuodao*, the topic constituent must and should appear in the discourse ahead of time, and the content should be acknowledged by both the speaker and the hearer. The topic of the sentence must be stressed in the sentence.

The phonological stress of the topic constituent is not uniquely found in the *shuodao* marker but is also found in any topic markers with *shuo*. In the discourse, when stressing the constituent after this verb, it is “realised by phonology” (Deng, 2009). However, other verbs apart from *shuo* are rarely found as topic markers.

2) Modal Particle Topic Markers

i.  *a* ("啊")

The research on the topic marker *a* ("啊") starts from Zhao (1968). In his work, he has summarised the functions of *a* and proposed ten functions of *a*. The function related to the topic marker function is that *a* is a modal particle indicating exclamation, as shown in the example in (61):

(61) 小王啊！你还 没 上 床 啊？
xiǎowáng a! nǐ hái méishàng chuáng a?
LittleWang MP! You still no up bed MP?
“Little Wang, you haven’t gone to bed yet?”

(Zhao, 1968: 361)

According to Zhao (1968), the main function of a is about expressing an exclamation mood of the hearer. In sentences like (61), the existence of a is aiming at creating a phonological stop and giving the hearer some time to reflect on the speaker’s words (Zhao, 1968). Comparing to sentences like (62):

(62) 小王啊，一辈子都没有出息。
xiǎowáng a, yī bèizi dōu méiyǒu chūxi.
LittleWang TpM, one lifetime still no achievement.
“Little Wang, never has any achievement in his lifetime.”

Interestingly, Zhao (1968) does not discuss the topic marker a in (62). His discussion turns to the strength of the exclamation function of a. According to Zhao (1968), the first a in (61) is a “rising exclamation”\textsuperscript{22}, while the a in (62) is a “falling tone exclamation”\textsuperscript{23} (Zhao, 1968: 361-362). Combining the prosodic feature with the syntactic position of these modal particles is an innovation at that stage, however, he does not create any discussion on the topic marker function of a.

In Li & Thompson (1989), they have summarised the ten functions of Zhao (1968) and provided a more in-depth discussion on this modal particle. Not only do they claim that some functions of Zhao (1968) are repetitive, but they also unified these repetitive functions as “reduced forcefulness (RF)” (Li & Thompson, 1989: 313). The meaning of reduced forcefulness is that “it has the semantic effect of softening the query” (Li & Thompson, 1989: 313):

(63) 你 喜欢 不 喜欢 这个 车子 啊?
nǐ xǐhuān bù xǐhuān zhè-gè chēzi a?
you like no like this-CL car RF?
“Do you like this car?”

(Li & Thompson: 1989: 313)

When omitting the a in sentence (63), it will influence the attitude the hearer is receiving. In other words, when the sentence is without a, “the first impression one has is that the questions with the

\textsuperscript{22}“升调感叹”

\textsuperscript{23}“降调感叹”
particle are much softer and thus tend to suggest kindness on the part of the speaker” (Li & Thompson: 1989: 313).

Based on the study of Zhao (1968) and Li & Thompson (1989), in 1996, Zhang & Fang (1996) added that a can be a marker indicating the constituent in the sentence-initial position when appearing in the middle of the sentence. The constituent in the sentence-initial position can refer to either the subject or the topic, thus, the new topic-marker function is added to this modal particle.

ii. ba (“吧”)

Similar to a, the discussion on ba also experienced a stage of researching the sentence-final discourse markers to the topic marker ba. In Zhao (1968), he summarised four functions of ba which are suggestions, the hypothesis under a dilemma, yes-no sentence-final particle and questionable statement, as shown in (64):

(64) a. 咱们 就 这么 办 吧！
    zánmen jiù zhème bān ba!
    we just this do MP!
    “Let’s do it!”

b. 不 给 钱 吧，不 好 意思 白 拿；给 钱 吧，又 给 不 起。
    bù gěi qián ba, bù hǎo yìsī bái ná; gěi qián ba, yòu gěi bù qǐ.
    no give money MP, no good feel empty take; give money MP, but give no up.
    “(If I) don’t give money, I feel shamed of taking it; (If I) give money, I cannot afford it.”

c. 你 知道 吧？
    nǐ zhīdào ba?
    you know MP?
    “You know that, right?”

d. 我 告诉 过 你 了 吧？
    wǒ gàosù guò nǐ le ba?
    I tell over you LE MP?
    “I have already told you isn’t it?”

    (Zhao, 168: 362)
The sentences in (64a) to (64d) correspond to four different functions raised by Zhao (1968), however, it can be easily observed that the topic marker *ba* is not being discussed in the initial research.

In Qu (2006), he studies the *ne* in declarative sentences, following Li & Thompson (1989). He thinks that *ba* is the hedging of the speaker (Qu, 2006: 106) and it has the meaning of “I am not very sure”:

(65) 好 吧，就 这么 办 吧。

    hǎo ba, jiù zhème bàn ba.

    well BA, just this do BA.

    “We, let’s just do this.”

The sentence-final *ba* can be interpreted as “I am not sure but we can try to do like this” (Qu, 2006). Qu (2006) also mentions the sentence-middle *ba*, and he made the claim that the sentence-middle *ba* is derived from the sentence-final *ba* and it maintains the function of the hedging and unconfident status in the discourse. The constituents before *ba* can be regarded as topics and it is the contrastive topic marker (Qu, 2006). He is not the first one who makes the claim that *ba* is the topic marker, in Zhang & Fang (1996), they made a brief discussion on the discourse marker *ba* and they believe that *ba* has similar functions as *a* and apart from being a topic marker, it does not have any other functions in the narrative sentences. It is just a marker “marking nominative constituents” (Zhang & Fang, 1996: 43).

iii. *ma* (“嘛”)

Li & Thompson (1989) and Zhao (1968) do not make any discussions on *ma*, but Qu (2006) and Zhang & Fang (1996) briefly discussed this marker. Being an interrogative marker, *Ma*, both studies have researched *ma* from two perspectives: the first *ma* is the marker making nominative constituents, and the second *ma* is the sentence-final particle. In Qu (2006), he concludes that the fiction of sentence-final *ma* is to “express the truth condition and more importantly, it contains a persistence of the speaker”:

(66) 这 件 事 就 是 这 样 的 嘛。

    zhè jiàn shì jiù shì zhè yàng de ma.

    this CL thing just be this look DE SFP.

---

24 “主位标记” (Zhang & Fang, 1996: 43)

25 “表示事实，但更重要的是它表示执着” (Qu, 2006: 113)
“This thing is just like this.”

As for interrogative *ma*, Qu (2006) indicates that the function of *ma* as an interrogative marker must “have a strong correlation to the person and event that previously has mentioned”\(^{26}\).

(67) 这 件 事 是 这 样 的 嘛?

zhè jiàn shì shì zhè yàng de ma?

this CL thing be this look DE IntP?

“Is this the way it is?”

In the work by Li (2008), he summarises the functions of *ma* and concludes that there are three functions of *ma*, which are: reduced forcefulness, topic marker, and interrogative marker.

iv. *ne* (“呢”)

*Ne* being one of the important function words is under much discussion by linguists. In Zhao (1968: 358), he divided the functions of *ne* into the following functions:

First, *ne* can appear in the sentences which are involved in discourse\(^{27}\), as shown in the example (68):

(68) 现在 咱们 干 点 什么 呢?

xiànzàizánmèn gàn diǎn shénme ne?

now we do some what NE?

“What should we do now?”

Second, *ne* can appear in interrogative sentences which are indicating features\(^{28}\) (Zhao, 1968: 38):

(69) 他 学习 好, 你 呢?

tā xuéxí hǎo, nǐ ne?

he study good, you NE?

“He can study well, what about you?”

Third, *ne* can appear in either interrogative or narrative sentences which have meaningful pauses\(^{29}\) (Zhao, 1968: 38):

(70) a. 这 种 事情 呢, 是 需要 解决 的。

<table>
<thead>
<tr>
<th>Source Note</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>“必须与上文已出现或者隐含的人、事、物有密切的关系” (Qu, 2006: 113)</td>
</tr>
<tr>
<td>27</td>
<td>有上下文的对话 (Zhao, 1968: 38)</td>
</tr>
<tr>
<td>28</td>
<td>有指特点的文化 (Zhao, 1968: 38)</td>
</tr>
<tr>
<td>29</td>
<td>有意停顿 (Zhao, 1968: 38)</td>
</tr>
</tbody>
</table>
Fourth, *ne* can appear in interrogative sentences which have mild warnings\(^{30}\) (Zhao, 1968: 38):

\[(71)\]  \(\text{这倒很危险呢！}\)  
\(\text{zhè dào hěn wēixiǎn ne!}\)  
\(\text{This instead very dangerous NE!}\)  
\(\text{“This is very dangerous, indeed!”}\)

Being one of the earliest research on the functions of *ne*, Zhao’s work (1968) reveals several fundamental functions of *ne*. However, the functions he raised are not covering all the functions, the topic function is not included inside.

Later, Qu (2006) claims that *ne* is a conjunction word between parallel sentences. The usage of *ne* is to connect the rest of the sentence to the content before *ne*. This function is very similar to the third function mentioned in Zhao (1968):

\[(72)\]  \(\text{这种事情呢，是需要解决的。}\)  
\(\text{zhè zhǒngshìqíng ne, shì xūyào jiějué de.}\)  
\(\text{This kind thing NE, is need solve DE.}\)  
\(\text{“This kind of thing, is waiting to be solved.”}\)

In sentence (72), Qu (2006) indicates that the fundamental function of *ne* is a correlation. To explain in detail, if the content is not related to the topic (frame), the existence of *ne* is to add the connections between the content and the topic frame (Qu, 2006: 127).

Different from Qu (2006), Li & Thompson (1989) claim that *ne* is a response to expectation. Since the content before *ne* is mentioned previously in the discourse, the existence of *ne* is to ask the hearer to pay attention to the speaker’s words. The response to the expectation function is similar to the topic frame relation mentioned by Qu (2006).

---

\(^{30}\)温和的警告 (Zhao, 1968: 38)
To conclude, the function of *ne* is to be a topic marker, marking the initial content of the sentence as a topic constituent. The other two functions of *ne* are a modal word, adding the connections between the topic and the topic frame, and indicating meaningful pauses.

3) Prepositional Topic Markers

Prepositional topic markers are one of the important topic markers in Chinese. In the research (Zhang, 2006; Qu & Shi, 2006), Chinese linguistics usually call this type of topic marker “prepositional guiding words”31”. With the existence of these prepositional words, the topic constituent will be guided and introduced after these guiding words in a sentence. In Mandarin Chinese, there are three typical and well-studied prepositional topic markers, which are *duiyu* (“对 于”, with regard to), *zhiyu* (“至 于”, as for), and *guanyu* (“关 于”, as regards). In this part, the functions and previous studies of these topic markers are introduced.

*Duiyu, zhiyu, and guanyu* are usually studied together due to their similarities in syntactic usage and semantic meaning. However, there still exist subtle differences: Different from *guanyu* and *duiyu*, *zhiyu* is an adversative which indicates that there is a turning point in the discourse:

(73) a. 至于 这 件 事, 你 想 怎么 解决?

        zhìyú zhè jiàn shì, nǐ xiǎng zěnme jiějué?

    “As for this thing, how do you want to solve it?”

b. 关于 这 件 事, 你 想 怎么 解决?

        zhìyú zhè jiàn shì, nǐ xiǎng zěnme jiějué?

    “As regards this thing, how do you want to solve it?”

c. 对于 这 件 事, 你 想 怎么 解决?

        duìyú zhè jiàn shì, nǐ xiǎng zěnme jiějué?

    “With regard to this thing, how do you want to solve it?”

31 “介 引”
In example (73), apart from the prepositional topic markers, the rest of the sentences are the same. Compared to (73b) and (73c), (73a) is containing a turning point in the discourse. Example (73a) is valid if and only if there is a turning logic in the discourse and the topic constituent is the one that the speaker and the hearer talk about something previously in the discourse. However, (73b) and (73c) does not have this restriction.

Although these markers have a close meaning in semantics, they also share differences: Because these three markers have semantic differences, they can only appear in different types of sentences. Guanyu and zhiyu can be replaced inside a sentence, but after alternation, the meaning of the sentence slightly changes:

(74) a. 至于 我 的 书, 没 人 看 了。
    zhìyú wǒ de shū, méi rén kàn le.
    As for my DE book, no people read SFP.
    “As for my book, no one read it.”

b. 关于 我 的 书, 没 人 看 了。
    zhìyú wǒ de shū, méi rén kàn le.
    As regards my DE book, no people read SFP.
    “As regards my book, no one read it.”

As mentioned before, when containing a zhiyu as a topic marker inside a sentence, there must have a turning point. Therefore zhiyu is indicating the connection relationship between the current topic and the previous topic, which means the current topic must have a connection to the previous topic and they must either form a comparison relationship or have an opposite relationship in the meaning. As for guanyu, the topic constituent need not have any relationship with the previous topics. Guanyu is more focusing on the topic-comment relationship inside one sentence.

Together being a topic marker, guanyu and duiyu can freely replace and does not cause any semantic ambiguity:

(75) a. 关于 这 件 事, 你 想 怎么 解决?
    zhìyú zhè jiàn shì, nǐ xiǎng zěnme jiějué?
    As regards this CL thing, you want how solve?
    “As regards this thing, how do you want to solve it?”

b. 对于 这 件 事, 你 想 怎么 解决?
duìyú zhè jiàn shì, nǐ xiǎng zěnme jiějué?

With regard to this CL thing, you want how solve?

“With regard to this thing, how do you want to solve it?”

Although the replacement of these two markers does not result in any change in the meaning of the sentences, there are some subtle differences between the two markers: when using guanyu in a sentence, the hearer has the intention that the topic constituent has a “parallel” relationship to the previous topic. However, when using duiyu in a sentence, the intention is that there is a progressive relationship between the topics mentioned in the previous discourse (Zhang, 2006; Qu & Shi, 2006).

2.5.3.2 Topic Markers in Dialects

Apart from investigating the topic markers in Mandarin Chinese, the topic markers in different dialects are also being studied by various researchers (Sun, 2021; Wu, 2021; Jiang, 2022). In this part, we are going to introduce three topic markers studied in different dialects: [kæ²¹] in the Mizhi dialect, [ȵiε] in the Ruian dialect, and [nε³³] in the Yi dialect.

1) [kæ²¹]

[kæ²¹] is one of the most frequently used topic markers in the Mizhi dialect, which is mainly spoken in the north part of Shaanxi province. The character of [kæ²¹] is “敢”. Being a verb, it is pronounced as [kæ²¹³], and when being a topic marker, it is pronounced as [kæ²¹]. When containing this topic marker in the sentence, there must contain a phonological stop after the marker, and “the timing should be less than the pausing time of a comma³²” (Sun, 2021: 23). In this paper, Sun (2021) briefly introduces the development pathway of [kæ²¹]: It derives from the auxiliary verb function and the modal adverb. It has experienced a phonological change when functioning differently: when being an auxiliary verb and a modal adverb, it is pronounced as a [kæ²¹³], when being a topic marker, it is pronounced as [kæ²¹].

2) [ȵiε⁰]

[ȵiε⁰] being one of the topic markers of Ruian dialect, spoken in the Ruian city, located in the South part of Wenzhou city, is appearing after the noun or noun phrase inside a sentence. In Wu (2021), he indicates that the topic marker [ȵiε⁰] is derived from the interrogative marker [ȵiε³¹]. When [ȵiε³¹] becomes to [ȵiε⁰], it experienced a reanalysis procedure (Wu, 2021: 312). During this reanalysis process, the [ȵiε⁰] has experienced a phonological, semantical and syntactic change. This paper

³² “时值要小于句中一个逗号的通常停顿” (Sun, 2021: 23)
does not provide any detailed discussion of the grammatical process, therefore in this thesis, we are providing a brief summary of the current study on the topic marker [ŋε³³].

3) [ŋε³³]

The topic marker of the Yi ethnic minority language is [ŋε³³], the most typical topic marker in the dialect. It is often used after the topic constituent inside the sentence, as shown in the example (76):

(76) a²¹ŋε³³ ne³³ mi³³ to³³ bo¹⁵ hu¹⁴² tə²¹ khu²¹ ma²¹ gε³³ li³³.

   now TPM Jishanjing inside every year sacrifice go.

   “Right now, everyone goes to Jishanjing to offer sacrifice every year.”

   (Jiang, 2022:67)

However, in example (76), [a²¹ŋε³³] is a timing adverb, therefore it is hard to say that [ŋε³³] is a topic marker. In the following example, there exists more support that [ŋε³³] is a topic marker:

(77) m³³tu³³ ŋɛ²¹ tiε³³ dʑi⁴⁴ fu³³ dia²¹.

   bamboo TPM I AGT chop PERF TRANS.

   “The bamboo, I have already chopped them.”

   (Jiang, 2022:68)

In sentence (77), [ŋε³³] is the typical marker which locates after the noun [m³³tu³³], indicating the topic constituent “bamboo”.

Yi being a topic-prominent language is relying on topic markers to introduce the topic constituent in the sentence, although the topic marker does not necessarily exist in the sentence. When discussing the origin of [ŋε³³] as a topic marker, this marker is derived from a conjunction word to an adverb and finally becomes a topic marker (Jiang, 2022: 73).

2.5.3.3 Topic Markers in Archaic Chinese

Compared to the research on topic markers in Modern Chinese, the research on archaic Chinese topic markers are under little research. Another debatable idea is that what kind of constituent can be regarded as a topic still remains unclear.

1) zhi “之”

This marker has two functions: the first function is that it is a connection between topic and comment in the sentence. The existence of the marker zhi is to rescind the independence of the sentence, which means, by adding this marker between the topic and comment, the merged clause is no longer an independent sentence. In other words, the original combination of topic and comment is a well-formed sentence, by adding zhi marker between the subject and predicate, the sentence
then has turned into a phrase in the sentence with incomplete meanings (Li, 1999). Besides, if this kind of clause is not followed by the explanation sentence, the clause will not exist on its own.

Therefore, the common structure of this clause is:

SBJ+ZHI+Predicate+explanation sentences.

The function of zhi in the sentence is to make the construction of subject and predicate become more “objectively and evently” (Li, 2012, 75). There are differences between these two words. According to Li (2012), “‘evently’ is to regard the original SBJ-OBJ as an event while ‘objectively’ is to turn the original SBJ-OBJ structure into a thing which subject represents and the predicate then turns into a modifier of subject.”

The second function of this marker is that this marker is a structural auxiliary word. The existence of zhi can also enhance the indication of the topic in the sentence, where the constituent after zhi is the topic of the sentence. More details of ye being a topic marker will be further introduced in section 4.2.

2) ye (“也”)

Ye is considered one of the most used topic markers from the Early Qin Dynasty (Liu, 2016:2). Since there are two different places where ye can locate, the topic marker which linguists target are sentence-middle ye markers:

(78) a. 过 也，人 皆 见 之。

guò yě, rén jiē jiàn zhī.

Guo TPM, people all see him.

“Guo, people all met him.”

(Lun Yu: Zi Zhang)

b. 臣 之 壮 也，犹 不 如 人。

chén zhī zhuàng yě, yóu bù rú rén.

I TPM strong YE, still no compare people.

“When I was strong, I cannot be compared with people.”

(Zuo Zhuan: Xi Gong Year 30)

33 “‘事件化’指将原来的主谓结构当作一个事件来看待；而‘事物化’则是指将原来的主谓结构转化成主语所代表的事物，谓语则变成了主语的修饰成分。”
In example (78), ye in both sentences are topic markers. Different researchers have discussed the function of topic marker ye and they agreed with the claim that when ye is appearing in the middle of the sentence, it is a phonological pause and aims at introducing the following contents (Wang, 1999; Zhang & Song, 2002). More details of ye being a topic marker will be further introduced in section 4.1.

3) *fu* ("夫")

Fu is a homograph. It can be read as both “fū” and “fú” and these two pronunciations have different functions. When having the pronunciation of fū, it is a noun with the basic meanings of “men, young adults” and “husband”. It also has the extended meaning of “teacher” and “wife of a Duke”. Examples are shown in (80) below:

(79) a. 一夫不耕，或受之饥。
     yī fū bù gēng, huò shòu zhī jī.
     One man not till, somebody endure ZHI hunger.
     “(If) One man does not till, somebody will endure hunger (because of this).”
     *(Lun Ji Zhu Shu)*

b. 子见夫子乎?
   zǐ jiàn fūzǐ hū?
   You meet teacher HU?
   “Have you met the teacher?”
   *(Lun Yu: Wei Zì)*

c. 晋穆侯有夫子姜氏。
   Jìn Mùhóu yǒu fūzǐ Jiāng shì.
   Jin Duke Mu has wife Jiang name.
   *(Zuo Zhuan: Second Year of Emperor Huan)*

In example (79a), this fu means young adults. In this sentence it can refer to farmers. As for (79b), this fuzi refers to the teacher Confucius. In (79c), this fuzi contains the meaning of the wife of a duke, different from either (79a) or (79b).

When this function word is pronounced as fú, it has three functions. First of all, it is a sentence-final particle, as in the example shown in (80):

(80) 而今而后，吾知免夫！

ér jīn ér hòu, wú zhī miǎn fū!

“From now on, I know how to avoid (illness)!”

(Lun Yu: Tai Bo)

In sentence (80), this  is served to “express exclamation” (Guo, 1989:79). Different from other sentence-final particles like  (“矣”) or  (“也”), it is not used as frequently as them and  (“矣”) and  (“也”) can be found in a variety of sentence types. However,  can only be found when the speaker is expressing pity, exclamation or singing. When it is a sentence-final particle, it is obviously situated at the end of the sentence. The second function of  is to be a demonstrative pronoun, and it is found in the middle of the sentence, for example:

(81) 则 夫 二 人 者, 鲁 国 社 稷 之 臣 也。

“Thus, these two people are country Lu’s important courtiers.”

(Zuo Zhuan: The Sixteenth Year of Cheng Emperor)

In example (82), the contents after  are the discussion part. The theme of this discussion is “people who have talent but don’t know”, and the speaker’s attitude is “the first unlucky thing”.

(82) 夫 有 贤 而 不 知, 一 不 祥 也。

“You have talent (people) but you don’t know, is the first unlucky (thing).”

(Yan Zi Chun Qiu: Jean Xia Chapter 10)

In example (82), the contents after  are the discussion part. The theme of this discussion is “people who have talent but don’t know”, and the speaker’s attitude is “the first unlucky thing”.

2.5.3.4 Common Problems of the Topic Marker Research
In the above several links, we have summarised the topic markers that have been studied in Chinese. Due to the different definitions of topics in these studies, there are different criteria in the topic marker research. So in this part, we are putting our focus on what kind of problems the Chinese topic marker research is currently facing, and how linguists are attempting to solve them.

One of the most crucial problems of the topic marker research is: due to the definition of the topic constituent is still under much debate, the criteria for identifying whether a marker is a topic marker or a normal function marker still remains unclear over the years. Regarding this issue, although linguists have put forward many different opinions, the debate will continue in the coming days.

The second problem that exists in the current topic marker research is: there exists an imbalance between the modern Chinese topic marker research and the archaic Chinese topic marker research. The topic research of modern Chinese has had a prosperous development since the beginning of the topic study, the study of the modern Chinese topic markers is sophisticated and the functions of these markers are under well-research. Under this comparison, the research on archaic topic markers exists the following problems:

1) The study of topic markers in ancient Chinese is more about the topic marker used in a Dynasty or inside a book: Understanding the development of topic markers can help us understand the way how ancient people spoke. This kind of research is an indispensable part of historical linguistics, and it is also an important part for us to simulate the syntactic and semantic features of ancient Chinese by building models. However, most of the current studies on topic markers are aimed at a certain Dynasty or a certain book; few scholars are conducting a large time span study, so this kind of study is necessary and urgent.

2) In the existing chronological studies of topic markers, the time span is usually short: Although the short timespan can show the usage of topic markers in a Dynasty in detail, it is difficult to simulate the changes of these markers from the past to the present. So it is necessary to conduct this kind of research over a long time span.

3) Whether all the topic markers being analysed at the moment real topic markers? Although many markers have been researched as topic markers, these topic markers cannot be pure topic markers because they have other functions. So one of our questions waiting to be solved is: Are these topic markers really topic markers? Or are they just function words with topic marker functions?
Because of these drawbacks, this paper aims to fill in the gap that has not been studied by previous linguistic researchers. This paper is aiming at summarising and analysing all kinds of topic markers that appeared in different periods of time in Chinese history.

2.6 Summary
In this section, we provided a relatively detailed discussion on the topic definition and the classifications of Mandarin Chinese topic markers. In the first part, we discussed the definition of the topic constituent. Followed by the studies on Mandarin topic constituents, we aim to provide a broad picture of the current studies on both the modern and archaic Chinese topic constituents. In the third part, we make a distinction between left-dislocation and predicate sentences. We conclude that these two concepts have distinct differences. In section 2.4, we classify the topic constituent in Mandarin Chinese and segment them into two types, which are aboutness and contrastive topics. In the last part of section two, we discussed the notion of topic markers. Although there are many studies on topic markers in Chinese, there is still lacking systematic research on these topic markers. From these researches, there may exist a cline that topic markers are mostly derived from interrogative markers, in the latter part of this thesis, we are going to provide a more detailed analysis on this phenomenon.
3. Methodology
Following the setting up of the concept of topic, it is vital to contemplate the means by which the grammaticalisation process of topic markers in archaic Chinese can be examined, as well as the appropriate analytical methodology to be employed. Corpus linguistics, a prominent research methodology within the field of linguistics, seeks to ascertain the evolution and current state of a linguistic phenomena through the analysis of corpora. A corpus is comprised of diverse texts and genres, and its outcomes can vary based on the queries employed. The corpus approach will be used in our investigation to gather data. In this section, the subsequent inquiries are addressed in a comprehensive manner: Corpus linguistics refers to a research methodology that involves the systematic analysis of large collections of written or spoken texts, known as corpora, in order to gain insights about language patterns, usage, and structure. The utilisation of corpora in the analysis of topic markers serves several purposes. Firstly, a corpus can provide a large and diverse collection of authentic language data, allowing researchers to examine the usage patterns of topic markers across different genres, registers, and contexts. This enables a more comprehensive understanding of how topic markers function in natural language. Furthermore, corpus studies offer several advantages for the analysis of topic markers. Firstly, they provide a systematic and empirical approach to investigating linguistic phenomena, ensuring that findings are based on objective evidence rather than subjective interpretations. Additionally, corpus allow for quantitative analysis, enabling researchers to identify frequency patterns and statistical associations related to the use of topic markers. Moreover, corpus studies facilitate the identification of collocational patterns and co-occurrence tendencies between topic markers and other linguistic features. This allows researchers to explore the syntactic and semantic relationships that topic markers have with other elements in discourse. Which corpus is utilised for analysis? In addition to the aforementioned inquiries, we also provide an introduction to the tools utilised for visualization purposes, as well as guidance on how to interpret and analyse graphical representations.

3.1 What is Corpus Linguistics
Corpus linguistics is defined as “the study of language data on a large scale – the computer-aided analysis of very extensive collections of transcribed utterances or written texts” (McEnery & Hardie, 2012: 1). The difference between corpus study and other kinds of linguistic study like questionnaires or field study is that corpus study involves a larger amount of linguistic examples collected by the computer. Corpus studies can show the distribution of a linguistic phenomenon either chronologically or synchronically.
A corpus is a kind of online resource containing a large number of texts, audio files or media. Corpora containing texts are also known as text corpora, and a corpus containing audio files or media is called a spoken corpus. In a text corpus, a large quantity of electronic texts is imported, structured and stored. Text corpus can be of different types. McEnery & Hardie (2012) discuss two kinds of corpus: annotated and unannotated corpora. Before discussing the differences between these two kinds of corpora, we should ask what is an annotation? Annotation means that the constituents are tagged with different syntactic categories or semantic notations. By annotating the corpus, the users will know that the element they are looking at is the target constituent and the results will be more precise. Another advantage of annotation is that the annotated corpus is more convenient for searching. When searching an annotated corpus, you can use a command like "NP + V + ADV" to find out the examples like "Nick eats quickly" in the corpus. If a corpus is not annotated, the searching of these examples will be tedious and time-consuming and one may need many commands to meet one’s research needs.

However, this does not mean that if a corpus is annotated, the corpus itself is suitable for your analysis. For a good data analysis, choosing a suitable corpus is relatively important. How to choose a suitable corpus depends on what type of analysis you are researching: are you conducting a chronological or a synchronic analysis? If one is conducting a chronological analysis, a corpus containing all the year 2020 Twitter examples is not suitable for analysis, unless the time span of interest is extremely narrow. If one is analysing a 21st-century synchronic phenomenon, then a corpus that has a time range from 100BCE to 1500AD is not suitable for research.

Apart from annotated and unannotated corpus, an alternative method of categorising a corpus is to determine whether it is a KWIC corpus or a KWOC corpus. A KWIC corpus is commonly referred to as a Key Word in Context, as first introduced by Luhn in 1959. Within this kind of corpus, the source text of the target language is emphasised and positioned in a concordance, as illustrated in Diagram 1:
Diagram 1: Returned results of "zai" in BCC corpus (Beijing Language and Culture University Corpus)

In contrast to KWIC, the KWOC corpus exhibits a distinct type of concordance. KWOC is an acronym that stands for Key Word Out of Context. This corpus is characterised by the placement of the target word list on the left side. Nowadays, the KWOC corpus is hardly seen because of its inconvenient concordance, Figure 2 illustrates the structure of the KWOC corpus:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Text</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>的</td>
<td>第一节教育学的研究对象</td>
<td>32.1</td>
</tr>
<tr>
<td>的</td>
<td>每一门学科都有自己的研究对象，都有自己特有的研究领域。科学就是按照研究对象的特定领域所具有的特殊矛盾而分成各种门类的。那么，什么是教育学要研究的特殊矛盾？</td>
<td>33.1</td>
</tr>
<tr>
<td>的</td>
<td>每一门学科都有自己的研究对象，都有自己特有的研究领域。科学就是按照研究对象的特定领域所具有的特殊矛盾而分成各种门类的。那么，什么是教育学要研究的特殊矛盾？</td>
<td>33.2</td>
</tr>
<tr>
<td>的</td>
<td>...自己的研究对象，都有自己特有的研究领域。科学就是按照研究对象的特定领域所具有的特殊矛盾而分成各种门类的。那么，什么是教育学要研究的特殊矛盾？</td>
<td>33.3</td>
</tr>
<tr>
<td>的</td>
<td>...都有自己特有的研究领域。科学就是按照研究对象的特定领域所具有的特殊矛盾而分成各种门类的。那么，什么是教育学要研究的特殊矛盾？</td>
<td>33.4</td>
</tr>
</tbody>
</table>
KWIC (Keyword in Context) and KWOC (Keyword Out of Context) are two methodologies employed in corpus searches and information retrieval: KWIC generates a display of occurrences of a specific keyword within its surrounding context, presenting the adjacent text in the form of phrases or sentences. This feature assists users in comprehending the usage of the keyword inside the given text. In contrast, KWOC isolates the term and associated content from the source text, displaying it as a standalone list or summary devoid of contextual details. Both strategies improve the effectiveness of corpus searches, enabling users to quickly find specific material of interest and play a vital role in corpus analysis and research efforts. Hence, the selection of an appropriate corpus holds great significance for our research, regardless of whether it is annotated or unannotated, or whether it is a KWIC corpus or a KWOC corpus.

3.2 Why corpus linguistics matters in our study?
To examine the historical usage of the topic marker, it is necessary to determine the frequency of its occurrence in various publications of diverse genres. Conducting a book-by-book analysis is feasible but highly inefficient due to the likelihood of overlooking phrases that include topic markers when viewing a page with several characters. In addition, the process of thoroughly reading a book and documenting all the instances is likewise highly laborious and time-intensive. Hence, utilising a corpus is the optimal approach to assist in our investigation. The benefits of utilising corpora for our investigation are: The research commands are shown prominently in the centre of the corpus return page and are visually distinct from the material. Furthermore, the corpora selected for the thesis comprise an extensive collection of electronic archaic Chinese writings. By examining a substantial quantity of texts, it enables a more comprehensive analysis of the linguistic issue at hand.
However, it is crucial to consider the following aspects while choosing the suitable corpus (or corpora) for our research. What is the ideal size of the corpus that we should use to meet our study requirements? Moreover, is it imperative to require annotation for the corpus? To ensure the selection of a suitable corpus for our research, we have identified multiple Chinese corpora. In order to address these two questions, the following part will conduct a comparison examination of many Chinese corpora and present the two specific corpora utilised in this thesis.

3.3 Corpora used for this thesis

As previously mentioned, corpora can be categorised into different classifications. They have the potential to exist in either an annotated or unannotated format, and they can be either synchronic (occurring at a specific point in time) or diachronic (occurring throughout a period of time). There are several Chinese corpora available, including the Zhonghuayuwen corpus (Mainland Compilation Committee for Cross-Strait Cooperative Compilation of Chinese Language Reference Books, 2012), BCC corpus from Beijing Languages and Culture University (Xun, et. al., 2016), Sinica Corpus from Academia Sinica in Taiwan (Chen, Huang, Chang & Hsu, 1996), chiParc corpus (Li, 2017), Sheffield corpus (Hu, et. al., 2005; Hu, et. al., 2007), and CCL corpus from Peking University (Zhan, et. al., 2019; Zhan, et. al., 2003). Nevertheless, every corpus possesses its own unique limitations. The Zhonghuayuwen corpus is not suited for historical research due to its lack of archaic sections. When evaluating the chiParc and Sheffield corpora, it is crucial to emphasis their constraints. Both datasets are rather diminutive in magnitude, so limiting their comprehensiveness and capacity to accurately depict language usage. Moreover, they encounter difficulties as a result of inadequate tagging, which hinders the effective identification and categorization of linguistic aspects in the data. Therefore, these restrictions make them inappropriate for thorough study and extensive research in the topic.

The tagging mechanism in the Sinica corpus is rather sufficient, however the archaic section within the corpus is not extensive enough for research purposes. As a result, we have chosen not to include the Sinica corpus in our analysis. The BCC and CCL corpora were selected for this study based on their appropriateness. In order to enable a thorough comparison of existing Chinese corpora, this section will meticulously examine the specific characteristics of the corpora mentioned. Moreover, it will furnish a comprehensive justification for choosing either the BCC or CCL corpus as the principal subject of investigation for the corpus research of the thesis. In light of the profound impact that Buddhist scriptures had on the cultural evolution of China as a whole, this section will also address the repercussions of these scriptures. Furthermore, we shall delineate our methodology
for managing Buddhist scripture data in order to guarantee the analytical rigour and accuracy of our investigation.

3.3.1 Corpus comparison

There are currently few extensive Chinese corpora accessible for scholarly investigation. Out of the various corpora considered, we have identified four that are exceptionally utilised and stand out for their substantial usage: the Zhonghuayuwen corpus, the BCC corpus from Beijing Languages and Culture University, the CCL corpus from Peking University, and the Sinica Corpus. The chiParc and Sheffield corpus are excluded from discussion in this section due to their small corpus size (less than 15 million characters) or unsuitable tagging system for corpus research.

With one hundred million characters, the Zhonghuayuwen corpus provides a substantial and diverse dataset for language studies. There are two primary sections comprising the corpus: annotated and unannotated. By annotating fifty percent of the characters, an in-depth linguistic analysis is possible. The information in this annotation pertains to word segmentation and part-of-speech.

Designed for linguistic analysis and research, the Beijing Language and Culture University CCL Corpus, alternatively referred to as the Centre for Chinese Linguistics Peking University corpus, comprises an extensive compilation of Chinese language texts. The corpus is remarkably large, consisting of 700 billion characters in total, of which 577 million are believed to have been created prior to 1949 AD. In the absence of annotations, the CCL corpus is restricted to character searches only.

The BCC corpus, which is the Beijing Language and Culture University Centre Corpus, contains a total of 15 billion characters. The archaic Chinese half contains two billion characters, spanning from 2070 BCE to 1912 AD, whereas the modern Chinese section contains thirteen billion characters, covering the years after 1912 AD. The section on modern Chinese is not only annotated, but the section on archaic Chinese is not annotated. When exploring the archaic section, the only supported search method is character search. The BCC corpus contains further details on the syntactic structure of both Chinese and English sentences.

The Sinica Corpus has 18 million characters and consists of texts sourced from the Academia Sinica in Taiwan. The corpus is not only annotated, but also contains forty-six lexical tags. Furthermore, the corpus includes many distinctive syntactic annotations, such as specific forms of nominalization structures and insertion phenomena. It is important to mention that this corpus does not contain any
ancient Chinese section. Each of the above corpora contains a wide range of literary works, Buddhist quotations, novels, legal works, scientific works, artistic works, and other types of works.

Based on the comparison, the BCC corpus and the CCL Corpus are presently the two most extensive Chinese corpora that exist. Both of these collections encompass a diverse range of literary styles and forms. Due to the absence of annotated archaic Chinese in any of the corpora we are using for our research, it is not possible to utilise annotation for our analysis. To address the problem of tagging, our objective is to utilise corpora that encompass a wide range of genres and encompass an extensive array of characters. Therefore, the BCC corpus and the CCL corpus are chosen for analysis. Ongoing study is being carried out on the ancient Chinese portion of the BCC corpus, as well as the works in CCL that can be traced back to the year 1949 AD. When one of the corpora lacked punctuation marks, the other version of the corpus was used instead.

3.3.2 Why Buddhist Quotes Matter

Buddhist quotations are of major significance in our analysis; however, due to their complicated influence, we have intentionally chosen to downplay their role in order to facilitate concentrated research discussions. The intentional evasion, which is reflected in subsequent literary selections, originates from a recognition of the complex character of Buddhist scriptures. Notwithstanding this circumspect stance, it is critical to emphasise the profound influence that Buddhist scriptures had on the development of the Chinese language.

Why are we continuing to debate the significance of Buddhist scriptures when we have decided not to discuss their content? The introduction of Buddhist Quotes to China initially led to the emergence of several innovative forms and conceptions, while also influencing the contemporary society of that era. Fang (2020) asserts that the proliferation of Chinese literature throughout the Wei and Jin Dynasties was a direct result of the incorporation of Buddhist Quotes. The YongMingTi poetry style originated in the Southern Dynasty and is regarded as an innovative kind of poetry. The cited work by Fang (2020: 6) asserts that this particular event marked a crucial moment in Chinese poetry, serving as the basis for the subsequent development of phonological linguistics and the flourishing of poetry during the Tang and Song dynasties. YongMingTi is derived from Buddhist Quotes, and its visual representation marks a significant milestone. The translation of Buddhist quotes introduces various philosophical ideas into Chinese philosophy, so influencing the lives of individuals (Fang, 2020). During the Eastern Han Dynasty, Confucianism held a position of supremacy as the prevailing ideology, and it was acknowledged and embraced by the ruling class. Subsequently, upon the introduction of Buddhist scriptures to China, the general
The translation of Buddhist Quotes has introduced numerous novel linguistic phenomena to the realm of archaic Chinese linguistics. The East Han Dynasty marks the historical era when Buddhist quotations were initially introduced to China. The Sutra in Forty-two Sections is the first documented work containing translated Buddhist quotations. However, the identity of the individual responsible for translating this work is being withheld. China had a larger influx of Buddhist quotations during the East Han Dynasty. The translations of these quotations were undertaken by translators such as Dharmaraksha, Lokaksema, An Shigao, and various others. During the Tang Dynasty, the distribution of Buddhist quotations had a period of significant growth and popularity, leading to a widespread acceptance of these quotations.

Given that Buddhist quotations are rendered from Sanskrit, each translation adheres to a distinct system. Xianlin Ji provided the following evaluation of the translation of Buddhist quotations: "Throughout Chinese history, two infusions of fresh water have contributed to the flourishing of Chinese culture. One is Western water, while the other originates in India. "Translation is necessary for these two infusions of fresh water" (Wang, 2004). The reference to "water from India" pertains to the temporal span spanning from the early Song Dynasty to East Han (Fang, 2013:1). The East Han Dynasty introduced numerous exotic words to Chinese (e.g., 刹那  chànà, which is translated from Sanskrit "Ksana"). The translation process has the potential to enhance the meaning and incorporate additional visual representations alongside the original Chinese characters (Fang, 2020).

In addition to the recently added vocabulary, there is also a presence of syntactic influence. Using the Mūlamadhyamakakārikā (MK) as an example, which is a fundamental scripture in Mahayana Buddhism written by Nāgārjuna in the 2nd century, as in the example shown in (84):

\[
\text{(84) dJ}^{35}: bhīvās \text{ tāvan na nirvāṇam jarā-marāṇa-lakṣaṇam} \mid 25.4a
\]

\[
\text{K}^{36}: \text{涅槃 不 名有, 有 则 老 死 相。}
\]

\[
\text{nirvana NEG ming existent be-opt.3.sg old age death feature}
\]

---

35 The Sanskrit version is mainly based on Jan Willem de Jong (1977), referencing to Lindtner (1982), Saito (1985), MacDonald (2000, 2007), and was arranged by Ye (2011).

36 The Chinese version is based on Chinese Tripitaka.
S&K\textsuperscript{37}: Nirvana is not, on the one hand, an existent; if it were, it’s having the characteristics of old age and death would follow.

(Mūlamadhyamakakārikā 25.4a)

In the above example, there exists a disparity in the grammatical structure between Sanskrit and the English rendition. In contrast to Mandarin Chinese, Sanskrit is a language that places the head of a phrase or sentence at the end. The syntactic translation between Sanskrit and Mandarin differs, as shown in the example above. The translator may occasionally alter the word order and eliminate extraneous words, thereby demonstrating the impact of Sanskrit (Fang, 2013). Consequently, because of the aforementioned causes, Buddhist quotes exhibit distinct characteristics compared to other genres, making them a topic of special analysis in our study.

In summary, while refraining from looking into the influence of Buddhist scriptures on language contact, the translation of these texts into Chinese has been instrumental in shaping China's literary, philosophical, and linguistic traditions. Subsequent investigations will prioritise the impact of language contact, which is facilitated by Buddhist scriptures, on the topic marker.

3.4 Books used for this thesis from the Corpora
Due to the diverse range of text genres in the corpus, it is not possible to guarantee an equal distribution of topics during the data collection procedure. Hence, we chose 44 novels from the corpus provided in section 3.3 as the primary material for collecting topic markers. In this section, we will provide a detailed argumentation on how we select books and the reason why we select these books for our corpus research.

It is critical to conduct research on topic markers in historical syntax by meticulously selecting pertinent historical books from the corpus. This is crucial due to three primary factors: To begin with, historical literature serves as an abundant reservoir of genuine linguistic usage, facilitating an exhaustive examination of syntactic structures throughout various epochs. Additionally, insight into the progression of topic markers over time can be gained by examining the evolution of language patterns and syntactic modifications through the selection of pertinent historical texts. Finally, the incorporation of a wide range of linguistic contexts is guaranteed by a carefully curated corpus, which enables an exhaustive examination of topic marker variations across different historical periods. As a result, meticulous attention is devoted to the selection of historical works for the corpus in order to augment the scope and profundity of the investigation. Our selection of books for

\textsuperscript{37} The English version is based on Nāgārjuna's Middle Way: Mūlamadhyamakakārikā (Mark Siderits & Shōryū Katsura, 2013).
research was predicated on the aforementioned three factors: publication date considerations, literary conventions, and word count.

The careful and deliberate selection of books from various temporal contexts within the corpus is critical when conducting an analysis of topic markers in archaic Chinese. An examination of the historical underpinnings of the Chinese language, with an emphasis on the development of topic markers, necessitates a discerning adherence to temporal considerations. Through the deliberate selection of texts from different time periods, one can shed light on the complex changes that occurred in vocabulary, syntactical structures, and pragmatic subtleties that were linked to the utilisation of topic markers. The researcher is able to traverse the linguistic terrain of ancient Chinese discourse by employing the temporal dimension as an objective perspective. An exclusive time record is provided by each selected time period, which provides valuable insights into the dynamic characteristics of topic identifiers and their application in context. By purposefully altering the chronological order, the researcher gains the ability to decipher the nuanced linguistic complexities that are intrinsic in the evolution of topic markers. By conducting a thorough examination that encompasses various historical intervals, scholars can acquire a sophisticated comprehension of the ways in which topic markers have evolved, diversified, and influenced modes of communication in the Chinese language. By adopting this methodology, the academic conversation surrounding topic markers is not only enhanced, but also a thorough and focused examination of the historical evolution of this linguistic phenomenon in ancient Chinese texts is conducted.

The careful and systematic selection of historical book genres from the corpus is also crucial in syntactic research, particularly when examining the function of topic markers in ancient Chinese texts. Such accuracy is crucial for a number of reasons: To begin with, classical literary works offer a wealth of reliable linguistic usage examples, which facilitate an exhaustive examination of syntactic structures throughout different historical periods. The texts’ intrinsic syntactic intricacy provides abundant material for comprehending the progression of topic markers. Furthermore, distinct syntactic characteristics are present in various literary genres, which offers a more profound understanding of the syntactic manifestations of topic markers. By selecting works from a variety of literary genres—including prose, poetry, and historical documents—it is possible to gain a more comprehensive comprehension of the syntactic variations in topic markers that have occurred during various historical periods. Finally, the rigorous curation of the corpus guarantees the incorporation of a wide range of syntactic contexts, which facilitates an exhaustive examination of the syntactic evolution of topic markers throughout various historical epochs. Therefore, the process
of selecting corpus items from various genres serves a dual purpose: to reveal the syntactic diversity present in the expression of topic markers in ancient Chinese texts, and to ensure that the selection is not solely based on content.

Linguistically, investigating the phenomenon of topic markers requires that the word count of target analysis texts in the ancient section of a Chinese corpus be carefully weighed. This careful consideration is motivated by several important linguistic factors: To begin with, ensuring that the word count remains balanced facilitates a representative sample of linguistic usage, thereby enabling a thorough analysis of syntactic structures spanning multiple historical periods. The nuanced comprehension of how topic markers appear in various textual contexts is possible due to the linguistic richness inherent in varying word counts. Furthermore, an accurate representation of the changes in topic markers can be achieved through the examination of syntactic modifications and language patterns. Additionally, a balanced selection of word counts enhances the ability to trace the evolution of these markers over time. The inclusion of topic markers in historical texts with varying word counts provides valuable insights into the syntactic complexities involved, thereby enhancing the comprehensiveness of linguistic analysis. In the first section, we discussed the two forms of the Chinese language: Literary Chinese (LC) and vernacular variants. We also examine these two aspects when choosing books for this section.

Based on the previous criteria, we initially chose the following books at the first stage. The initial selection of these works for examination is based on two reasons: Firstly, they cover practically all dynasties, and the compilation times of these publications are chosen to prevent any chronological gaps observed in prior studies. Secondly, these books belong solely to either the Literary Chinese (LC) or vernacular variants classifications. We will first examine the sample size before further refining the selection of target books for study. Along with their estimated publication dates, these books are:

<table>
<thead>
<tr>
<th>Book</th>
<th>Character (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yi Jing (1046BCE)</td>
<td>30,970</td>
</tr>
<tr>
<td>Chun Qiu (481BCE)</td>
<td>252,534</td>
</tr>
<tr>
<td>Shang Shu (475BCE)</td>
<td>33,444</td>
</tr>
<tr>
<td>Guo Yu (450BCE)</td>
<td>147,711</td>
</tr>
<tr>
<td>Lun Yu (475-450BCE)</td>
<td>21,746</td>
</tr>
<tr>
<td>Zhong Yong (475-211BCE)</td>
<td>4,465</td>
</tr>
<tr>
<td>Han Fei Zi (475-211BCE)</td>
<td>128,831</td>
</tr>
<tr>
<td>Zuo Zhan (400-300BCE)</td>
<td>252,534</td>
</tr>
<tr>
<td>Yi Zhou Shu (&lt;425BCE)</td>
<td>44,402</td>
</tr>
<tr>
<td>Meng Zi (350-300BCE)</td>
<td>45,069</td>
</tr>
<tr>
<td>San Guo Zhi (215-185BCE)</td>
<td>686,342</td>
</tr>
<tr>
<td>Zhao Fei Yan Bie Zhan (221BCE-300AD)</td>
<td>2,481</td>
</tr>
<tr>
<td>Mo Zhi (200BCE)</td>
<td>103,245</td>
</tr>
<tr>
<td>Huai Nan Zi (150BCE)</td>
<td>158,669</td>
</tr>
<tr>
<td>Shi Ji (104-91BCE)</td>
<td>592,368</td>
</tr>
<tr>
<td>Book Title</td>
<td>Dynasty</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Lun Heng</td>
<td>(86AD)</td>
</tr>
<tr>
<td>Xian Di Chun Qiu</td>
<td>(153-212AD)</td>
</tr>
<tr>
<td>Qian Han Ji</td>
<td>(200AD)</td>
</tr>
<tr>
<td>Sou Shen Ji</td>
<td>(317-342AD)</td>
</tr>
<tr>
<td>Hua Yang Guo Shi</td>
<td>(348-354AD)</td>
</tr>
<tr>
<td>Hou Han Shu</td>
<td>(398-445AD)</td>
</tr>
<tr>
<td>Wen Xin Diao Long</td>
<td>(501-502AD)</td>
</tr>
<tr>
<td>Shui Jing Zhu</td>
<td>(515-527AD)</td>
</tr>
<tr>
<td>Qi Min Yao Shu</td>
<td>(533-544AD)</td>
</tr>
<tr>
<td>Tang Guo Shi Bu</td>
<td>(821-824AD)</td>
</tr>
<tr>
<td>Jiu Tang Shu</td>
<td>(945AD)</td>
</tr>
<tr>
<td>Zu Tang Ji</td>
<td>(952AD)</td>
</tr>
<tr>
<td>Tai Ping Guang Ji</td>
<td>(977-978AD)</td>
</tr>
<tr>
<td>Xin Tang Shu</td>
<td>(1043-1060AD)</td>
</tr>
<tr>
<td>Zi Zhi Tong Jian</td>
<td>(1065-1084AD)</td>
</tr>
<tr>
<td>Meng Xi Bi Tan</td>
<td>(1086-1093AD)</td>
</tr>
<tr>
<td>Xi You Ji</td>
<td>(1295-1307AD)</td>
</tr>
<tr>
<td>Da Song Xuan He Yi Shi</td>
<td>(1300AD)</td>
</tr>
<tr>
<td>Liao Shi</td>
<td>(1344AD)</td>
</tr>
<tr>
<td>Shui Hu Zhuan</td>
<td>(1360-1400AD)</td>
</tr>
<tr>
<td>Yuan Shi</td>
<td>(1370AD)</td>
</tr>
<tr>
<td>Mu Dan Ting</td>
<td>(1595-1600AD)</td>
</tr>
<tr>
<td>Tian Gong Kai Wu</td>
<td>(1637AD)</td>
</tr>
<tr>
<td>Quan Yuan San Qu</td>
<td>(1801AD)</td>
</tr>
<tr>
<td>Yin Shui Ci</td>
<td>(1655-1685AD)</td>
</tr>
<tr>
<td>Ming Shi</td>
<td>(1739AD)</td>
</tr>
<tr>
<td>Hong Lou Meng</td>
<td>(1773AD)</td>
</tr>
<tr>
<td>Wen Shi Tong Yi</td>
<td>(1850AD)</td>
</tr>
<tr>
<td>Zeng Guo Fan Jia Shu</td>
<td>(1850AD)</td>
</tr>
</tbody>
</table>

Table 1: Books chosen for corpus analysis at the first stage

The above classification is aimed at ensuring that the selected books cover as many dynastic periods as possible. If only books with literary Chinese are chosen for analysis, building upon Table 1, we further identify the following books for investigation in this thesis:

<table>
<thead>
<tr>
<th>Book Title</th>
<th>Dynasty</th>
<th>Character Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yi Jing</td>
<td>(1046BCE)</td>
<td>30,970</td>
</tr>
<tr>
<td>Chun Qiu</td>
<td>(481BCE)</td>
<td>252,534</td>
</tr>
<tr>
<td>Shang Shu</td>
<td>(475BCE)</td>
<td>33,444</td>
</tr>
<tr>
<td>Guo Yu</td>
<td>(450BCE)</td>
<td>147,711</td>
</tr>
<tr>
<td>Lun Yu</td>
<td>(475-450BCE)</td>
<td>21,746</td>
</tr>
<tr>
<td>Zhong Yong</td>
<td>(475-211BCE)</td>
<td>4,465</td>
</tr>
<tr>
<td>Han Fei Zi</td>
<td>(475-211BCE)</td>
<td>128,831</td>
</tr>
<tr>
<td>Zuo Zhuan</td>
<td>(400-300BCE)</td>
<td>252,534</td>
</tr>
<tr>
<td>Yi Zhou Shu</td>
<td>(&lt;425BCE)</td>
<td>44,402</td>
</tr>
<tr>
<td>Meng Zi</td>
<td>(350-300BCE)</td>
<td>45,069</td>
</tr>
<tr>
<td>San Guo Zhi</td>
<td>(215-185BCE)</td>
<td>686,342</td>
</tr>
<tr>
<td>Mo Zi</td>
<td>(200BCE)</td>
<td>103,245</td>
</tr>
<tr>
<td>Huai Nan Zi</td>
<td>(150BCE)</td>
<td>158,669</td>
</tr>
<tr>
<td>Shi Ji</td>
<td>(104-91BCE)</td>
<td>592,368</td>
</tr>
<tr>
<td>Lun Heng</td>
<td>(86AD)</td>
<td>246,956</td>
</tr>
<tr>
<td>Xian Di Chun Qiu</td>
<td>(153-212AD)</td>
<td>5,531</td>
</tr>
<tr>
<td>Qian Han Ji</td>
<td>(200AD)</td>
<td>212,086</td>
</tr>
<tr>
<td>Hua Yang Guo Shi</td>
<td>(348-354AD)</td>
<td>453,039</td>
</tr>
<tr>
<td>Hou Han Shu</td>
<td>(398-445AD)</td>
<td>910,107</td>
</tr>
<tr>
<td>Wen Xin Diao Long</td>
<td>(501-502AD)</td>
<td>50,000</td>
</tr>
</tbody>
</table>
A total number of 18,761,284 characters are analysed in the final-chosen books. The advantage for choosing these books for analysis is: Examining ancient Chinese topicalisation in literary Chinese texts in a doctoral dissertation provides notable benefits. This approach enables a concentrated examination of the utilisation and development of topicalisation within a comprehensive historical framework. Through the analysis of the usage of topicalisation in ancient Chinese texts, researchers can obtain valuable insights into the linguistic norms and conventions that were specific to earlier periods of the language. Furthermore, employing a uniform collection of written Chinese texts guarantees consistency and dependability when analysing the patterns and usage of topicalization. The rigorous methodology employed in this study reduces the influence of genre differences, thus enhancing the dependability and accuracy of the findings regarding topicalisation in ancient Chinese discourse. Furthermore, examining the phenomenon of topicalisation in the context of Chinese literacy offers a chance to investigate genre-specific characteristics and the ways in which linguistic decisions intersect with the communicative objectives and established conventions found in ancient Chinese literary works. This research not only enhances our comprehension of language usage and evolution, but also provides cultural insights into the communicative practices of ancient Chinese societies. Using Chinese literary texts to study ancient Chinese topicalisation in a doctoral dissertation enhances scholarly comprehension of language usage, historical linguistics, and cultural aspects in ancient Chinese discourse.

<table>
<thead>
<tr>
<th>Title</th>
<th>Character</th>
<th>Title</th>
<th>Character</th>
<th>Title</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shui Jing Zhu (515-527AD)</td>
<td>379,150</td>
<td>Qi Min Yao Shu (533-544AD)</td>
<td>141,567</td>
<td>Tang Guo Shi Bu (821-824AD)</td>
<td>23,829</td>
</tr>
<tr>
<td>Jiu Tang Shu (945AD)</td>
<td>2,389,053</td>
<td>Zu Tang Ji (952AD)</td>
<td>249,664</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xin Tang Shu (1043-1060AD)</td>
<td>1,999,737</td>
<td>Zi Zhi Tong Jian (1065-1084AD)</td>
<td>3,127,070</td>
<td>Meng Xi Bi Tan (1086-1093AD)</td>
<td>126,459</td>
</tr>
<tr>
<td>Da Song Xuan He Yi Shi (1300AD)</td>
<td>66,401</td>
<td>Liao Shi (1344AD)</td>
<td>425,341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yuan Shi (1370AD)</td>
<td>1,869,407</td>
<td>Ji Xiao Xin Shu (1560AD)</td>
<td>74,180</td>
<td>Ming Shi (1739AD)</td>
<td>3,303,829</td>
</tr>
<tr>
<td>Wen Shi Tong Yi (1801AD)</td>
<td>205,548</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table. 2: 34 Books chosen for corpus analysis (final choice)
3.5 Queries and visualisation

Following the presentation of the corpora that were utilised for this thesis, which will be the BCC corpus and the CCL corpus, this section will then show the queries that will be used to search the corpora and visualise the results.

3.5.1 Command and example search

Locating samples containing topic markers in the archaic section of the CCL corpus and the BCC corpus is more complex due to the absence of annotations, compared to using the annotated corpora. To ensure the accuracy of the results while also minimising the time required, the instructions used for searching the topic markers are divided into three stages: To initiate the process, we utilise the "customise books" command to explore the corpus and locate the specific books of our interest. After entering the search area, the subsequent action is to input the target topic marker, such as ye. Due to the syntactic property of the topic markers, the topic markers in the sentence should not be followed by a full stop because the topic constituent can never exist at the end of the sentence. This occurs because the topic markers possess a grammatical characteristic that hinders their completion. Hence, the final command is accountable for eliminating all samples that have the expression "topic marker + full stop." The command "topic marker + full stop" signifies that the full stop should immediately follow the topic marker. After carrying out these two instructions, we will proceed to systematically evaluate each of the remaining occurrences.

It is important to note that each result is thoroughly reviewed twice within a month to ensure the accuracy of the chosen instances. There are instances where the assessment of a sentence that includes a topic marker can be controversial. The second phase of analysis primarily concentrates on the disputed cases to guarantee that all the recorded examples have identical topic contents. The data from the corpus is recorded and shown using Excel in R. Excel is used for both objectives.

3.5.2 Tools used for Visualisation

R is an exceptionally potent and adaptable tool, rendering it the optimal choice for visualisation.

The ggplot2 package is referenced in this thesis (Wickham H, 2016). We use this programme for the purpose of visualisation. With the use of the software called R, derived from Wilkinson's Grammar of Graphics (a study conducted in 1999), it is feasible to create graphs that are both comprehensible and aesthetically pleasing. Although ggplot2 is a powerful tool, it imposes significant limitations on the structure of the raw data. There is no need for us to retain all of the
null data in order to accurately replicate the historical pattern of applying topic markers. Null data indicates the absence of an example for a specific topic marker in the book during research.

3.5.3 Analysing the Graphs

After gathering all the data from the corpus, we are currently considering the optimal method for presenting our data and the most suitable approach for explaining the use of topic markers. To generate graphs with smoothed conditional means and 95% confidence intervals, it is suitable to utilise the `geom_smooth()` function provided in the `ggplot2` package. The presence of numerous scattered plots throughout the dataset is the reason behind this. This section focuses on addressing the following three questions: The term “smoothed conditional mean” refers to the calculated average value of a variable, taking into account the smoothing technique applied to the data. Moreover, what justifies the suitability of the model incorporating the smoothed conditional mean for our hypothesis? Regarding the third question, what is the rationale behind incorporating a 95% confidence interval in the graph, and what insights may be gained from it?

The conditional mean is the anticipated value of a stochastic variable, and it is calculated by considering a significant number of data points from the total population. In other words, having the conditional mean enables one to construct a rough calculation of the probability of an event using a substantial amount of training data. The formula for determining the conditional mean of Y, given that X equals x, is as follows (Steyer & Nagel, 2017):

$$
\mu_{Y|x} = E[Y|x] = \sum_y y h(y|x)
$$

When utilising the `ggplot2` tool for visualisation, the package generates a graph that exhibits the smoothed conditional mean. The smoothed conditional mean refers to the estimated average value of a variable, taking into account past observations and incorporating a smoothing technique to reduce noise and variability in the data. The conditional mean is implemented to ascertain the smoothed conditional mean, which is thereafter accepted. The smoothed conditional mean refers to the average value of neighbouring values while sliding with a time series. By adopting a smoothing technique, the calculated averages are able to mitigate any erratic fluctuations and emphasise the gradual progression of a research topic. By employing these mean values, one can effectively mitigate unpredictable fluctuations.
Throughout our research, we employed a significant amount of data obtained from the corpora. By employing the smoothed conditional mean, we successfully made a precise forecast on the expected percentage of the usage of the topic marker in ancient Chinese. The graphs displaying smoothed conditional means are shown in the subsequent section, illustrating the suitability of this model for our research outcome.

In addition to the smoothed conditional mean, a confidence interval with a 95% degree of certainty is also presented. The confidence interval for 95% is a statistical range that has a 95% probability of containing the true mean. By incorporating the 95% confidence interval into our graph, we can not only draw conclusions about estimating the evolution of the topic marker in different dynasties, but also acknowledge the probable range of uncertainty in our forecasts. The inclusion of the 95% confidence interval in our graph is justified for this rationale.

3.6 Section Summary

In this section, we present the methodology employed in this study, encompassing the selection of corpus books, the tools utilized for modeling, and the techniques applied for data visualization. These aspects collectively enable a scientifically rigorous investigation into the topic marker. We then employed advanced modeling tools to analyze the corpus data, allowing us to identify and characterize patterns related to the topic marker. Additionally, we utilized sophisticated data visualization methods to present our findings effectively and intuitively, enhancing the interpretation and understanding of the research outcomes. This methodological framework ensures a comprehensive and systematic exploration of the topic marker within the context of our study.
4. Topic markers in Mandarin

In this section, four different Mandarin topic markers are analysed, which are ye “也”, zhi “之”, ne “呢” and zhe “者”. The markers are analysed for their individual functions and their distribution throughout the history of Chinese. Comprehending the current distribution of each marker is essential as it illuminates their frequency and usage patterns throughout Chinese history.

Examining the distribution of these markers allows us to gain a deeper understanding of their historical and contextual significance in the Chinese language. Through the analysis of their frequency patterns over time, we can acquire a more profound comprehension of the evolutionary and adaptive processes of these markers in diverse linguistic contexts. Analysing the distribution of these topic markers allows us to understand the subtle ways in which they have been used in various historical periods and types of texts. This examination can also uncover more extensive linguistic patterns and cultural advancements that have influenced the utilisation of these topic markers throughout history.

4.1 Ye “也”

Ye (也) is one of the most frequently used function words in Mandarin. The usage of ye can be traced back to Shang dynasty, around 1600BCE. Mostly, during this period, ye is used as a sentence final particle (SFP) and the aim of using this SFP is to “add the speaker’s emotion to this sentence” 38(Ma, 1989: 92), which can be either narrative or exclamation, and it is “a natural stop” 39 (Ma, 1989: 92). Around 400 years later, ye acquires a new function: to mark the topic in a sentence. In other words, ye not only functions as a SFP, but also as a topic marker. Later in Ming dynasty, ye is then used as an adverb but both this function and ye as a SFP will not be discussed in detail here.

Ye as a topic marker is usually found in judgement sentences (判断句), and this type of sentence is also known as the predicate sentence (Cheng, 2018; Sun, 2016). This type of sentence contains noun or noun phrase predicates and these predicates are used to identify the property of a thing referred to in the subject/topic position. In this type of sentence, ye has two fixed positions: either it can be found in the middle of the sentence, or at the end of the sentence.

38 增加语句的情调或感情色彩
39 自然的停顿和延宕
4.1.1 Functions of Ye “也”

4.1.1.1 Ye in the middle of the sentence

When *ye* is found in the middle of the sentence, it looks like the example in (83):

(83) 臣闻国之兴也以福，其亡也以祸。

chén wén guó zhī xīng yě yǐ fú , qí wáng yě yǐ huò.

I hear country ZHI prosperous YE because luck, its die YE because misfortune.

“I heard that if a country is prosperous, it is because this country has luck; when the country has perished, this country has misfortune.”

(Zuo Zhuan: The First Year of Duke, 350BCE)

In example (83), the meaning of *ye* has the same meaning as the copula “is” in English, aiming to give an explanation or express further details of the subject. Thus, *ye yi fu* (也以福, because of luck) is explaining the status of a country, and so is *ye yi huo* (也以祸, because of misfortune). When situated in the middle of the sentence, the component after *ye* is the comment aiming to explain the topic before *ye*. As a topic marker, the topic can be an adjective phrase (*guo zhi xing*, 国之兴) as in (83), or a verb phrase, as in the example shown in (84):

(84) 其行己也恭，其事上也敬。

qí xíng jǐ yě gōng, qí shì shàng yě jìng.

its behaviour oneself YE polite, its serve king YE respect.

“(He) behaves himself in a polite way, and serves the king in a respectful way.”

(Lun Yu: Gong Ye Chang, 475BCE)

In example (84), *xing ji* (行己, behave oneself) and *shi shang* (事上, serve the king) are verb phrases and *gong* (恭, polite) and *jing* (敬, respect) are comments modifying the verb phrases. Besides, *ye* can also have a noun as its topic, as in the example shown in (85):

(85) 求也退，故进之。

Qiú yě tuì, gù jìn zhī.

Qiu YE flinch, so progress him.

“Qiu is timid so I am encouraging him.”

(Lun Yu: Xian Jin, 475BCE)

To conclude, when having *ye* as a topic marker in the middle of the sentence, *ye* can take nouns (including noun phrases), verb phrases and adjective phrases as topics, and the component after *ye* is the comment indicating the status or the property of the topic. The syntactic position is shown below. The topic marker *ye* occupies the head of TopP position and forms a TopP with the DP Qiu.
In addition to its use in a sentence, *ye* can also combine with *zhe* (“者”) and forms the expression *yezhe* (“也者”). *Zhe* has the function of ‘giving hint’, and it is used to gain reader’s attention\(^{40}\) (Ma, 1989: 91). *Ye*, as discussed previously, has an affective/emotive function here and it is a natural stop in a sentence; combining *zhe* and *ye* together highlights the topic in a sentence and they reinforce each other as the example (86) shows:

(86) 夫 祭 也 者, 必 夫 妇 亲 之。

夫 祭 也 者, 必 夫 妇 亲 之。

FUsacrifice YE ZHE, must husband wife in person ZHI.

“Sacrifice is (the thing that) husband and wife should do by themselves.”

*(Li ji: All about Sacrifice, 30AD)*

When translating this type of sentence like (86), the meaning of *zhey* falls onto *zhe* which is a nominalized marker meaning “the thing like” or “the person like”. The comment part is the explanation of the topic thus the topic is an aboutness topic. When coming across *yezhe* in a sentence, the translation will always be “something like + topic” or “somebody like + topic”.

To sum up, there are two circumstances when *ye* is situated in the middle of the sentence. Firstly, when *ye* is solely used in the sentence, it follows the pattern of “Topic (N/NP/AdjP/VP…) + ye + comment”. Secondly, when *ye* is combined with *zhe* forming *yezhe*, the meaning of this expression falls onto *zhe* and has the meaning of “something/somebody like + topic”.

### 4.1.1.2 Ye “也” at the end of the clause

When *ye* is found at the end of a clause, the sentence either has the pattern like (87a) or (87b):

(87) a. XXX ye, XXX.

b. XXX, XXX ye.

Example (90) below has the pattern of (89a):

---

\(^{40}\) “者”有“提示”的作用，以引起读者的注意。
When a man is crowning (becoming an adult), his father is instructing him; when a girl is marrying, her mother is instructing her.

(Mengzi: Duke Teng Wen Volume Two, 325 BCE)

In this sentence, *fumingzhi* (父命之, father instruct him) and *mumingzhi* (母命之, mother instruct her) are explanations to the topics *zhangfuzhi* (丈夫之冠, man crowns) and *nvzijia* (女子之嫁, girl marries). These two topics reveal different important events in men and women’s lives, so the comment is the further explanation of these two topics. Thus, these two topics are contrastive topics.

Example (89) below has the pattern of (87b). The first part of the sentence contains the topic of the whole sentence and the second part of the sentence is the comment. Scholars (Liu, 2016; Shi, 2000; Xu & Langendoen, 1985) argue that the structure *zhe...ye...* is the predicate sentence marker and both *zhe* and *ye* here are topic markers. In other words, *jidi* (吉禘, lucky worship) is the topic and *bujizhe* (不吉者, not a lucky thing) is the comment of lucky worship, indicating that a lucky worship is not a lucky thing.

(89) 吉禘者，不吉者也。

*jidi* ZHE, not lucky thing YE.

“A unhindered worship ceremony is not a lucky thing.”

(Chunqiu: The Story of Guliang, 470 BCE)

However, this paper argues that the *ye* here is not a topic marker, it is a comment marker instead. A sentence cannot have two topic markers, one situated in the middle of the sentence and the other at the end of the sentence. The structure of this sentence is given below:

In this structure, *zhe* is a nominalized marker meaning “something like…”, so *jidi* and *zhe* both form a TopP constituent, and *zhe* is the topic marker. The criteria of distinguishing *zhe* being a topic marker or a nominaliser will be shown in section 5.1. As for the latter part of the sentence, *ye* is the marker of the whole sentence and it occupies the position of Spec,C and C-commands the rest of the sentence. Therefore, *ye* cannot refer to the *jidi* in the first part of the sentence and it is surely not a topic marker of the sentence. Since the latter part of the sentence is the comment explaining the topic, *ye* here is a comment marker.
As discussed in 4.1, ye can also combine with zhe to form an expression. Different from ye zhe, which can be put in the middle of the sentence, zhe ye is the fixed expression when these two characters are combined and placed at the end of the sentence. Similar to (88), (90a) and (90b) are two different types of zhe ye found in the sentence.

(90) a. XXX zhe ye, XXX.
    b. XXX, XXX zhe ye.

When zhe ye is situated at the end of the first half of the sentence like example (91), zhe ye here can be explained separately: zhe here has the meaning of “people like…” and ye here is the topic marker indicating people like the scholars are the people worth respecting. The latter part of the sentence is the comment explaining the topic. It is worth mentioning that there is another ye at the end of the whole sentence, and this ye is the comment marker, different from the first ye, which is the topic marker. From another perspective, it shows that in a sentence, both ye as a topic marker and a comment marker can co-exist.

(91) 士之尊贤者也, 非王公之尊贤也。
    shì zhī zūn xián zhě yě, fēi wánggōng zhī zūn xián yě.
    Scholar ZHI respect worthy people YE, not nobility ZHI respect worthy YE.
    “(The way) scholars respect those worthy people (is) not (the way that) nobilities respect those worthy people.”

    (Meng Zi: Wan Chapter, 325BCE)

As for example (92), the translation of zhe ye is the same as (91); however, here ye has the same function as the second ye in (91), which is the comment marker. The topic marker in this sentence is the first zhe.

(92) 五谷者, 种之美者也。
    wǔ gǔ zhě, zhòng zhī měi zhě yě.
Five grain ZHE, grow ZHI best thing YE.

（Meng Zi: Gao Zi, 325BCE）

To conclude, when the expression zheye is found at the end of the sentence, when it is ending the first half of the sentence, ye here is the topic marker and zhe refers to “the thing/man like…”. When this expression is found at the end of the whole sentence, the meaning of zhe remains the same and ye here is the comment marker. The notion of comment marker has not been proposed so far by Chinese linguists. The comment marker is a clause-final particle that marks that the whole structure has topic-comment form. In sentence (94), the constituent before zhe is the topic, and the latter part of the sentence is the comment and ye is acting as the comment marker of the clause.

4.1.2 The trend of using ye “也” as a topic marker over time

In order to investigate the change of ye throughout history, we adopted 34 books (shown in methodology section 3.3.3) for analysis. According to the position of ye, the following variables are counted: number of ye as a topic marker (named ye); number of all the examples containing ye in 34 books (named ye-total); percentage of ye being a topic marker among all the examples (named ye-percentage=ye/ye-total*100%). Previously, we mentioned that zhe can also be a topic marker and it can combine with ye to form the topic-comment structure. Therefore, in our analysis, the number of zhe as a topic marker is also counted (named zhe); the total number of hits containing zhe is counted (named zhe-total); and the percentage of zhe being a topic marker is calculated (named zhe-percentage=zhe/zhe-total*100%). Apart from calculating the individual topic marker zhe and ye, the combination of zhe…ye... structure is also calculated in the corpus (named zhe-ye). Two other numbers are also presented, which are zhe+ye (= zhe + ye, which is the total number of the sentences containing either zhe and ye); and zheye/zhe+ye*100%, which calculates the percentage of zheye in all the sentences includes either zhe or ye.

Graph 1 below shows the percentage of ye as a topic marker in 34 books. The size of the black dots represents the total number of characters that the books contain, and the blue line is the smoothed conditional mean of the usage of ye as a topic marker. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence of ye being a topic marker among all the sentences containing ye in each book (Percentage = ye / ye-total * 100%). The developing trend of ye as a topic marker has experienced a rise-and-fall process: Starting from 0%, the trend gradually rises up to 13% at 300AD and falls to 0% by 1822AD.
There are two books that are particularly responsible for the rise of the trend which are the *Xian di Chunqiu* (29.196%) and the *Qian Han Ji* (23.482%). The general trajectory of the topic marker *ye* demonstrates an initial upward trend followed by a subsequent decrease. The marker usage begins at a modest 2% in 1046 BCE and gradually increases over the centuries. It reaches its highest point at 11.5% around 510 AD, during the Eastern Han dynasty. Following that, the usage undergoes a substantial decrease, dropping to around 0% by 1801, signifying the conclusion of the 19th century. Upon further analysis of the data, we are able to pinpoint distinct periods of variation within this overarching pattern. During 1046 BCE to 221 BCE, the topicalisation rate experiences a slow increase from 2% to 7%; during the period from 221 BCE to 510 AD, there is a significant increase in the usage of topic marker *ye*, with the percentage rising from 7% to 11.5%. This surge aligns with the Han dynasty, a time characterised by notable progress in governance, culture, and education. After 510 AD, the literacy rate experiences a long-lasting decrease.

Graph 2 below shows the percentage of *zhe* as a topic marker in 34 books. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence of *zhe* being a topic marker among all the sentences containing *zhe* in each book (*Percentage = zhe / zhe-total * 100%). Different from the trend of *ye*, the developing trend of *zhe*
as a topic marker has experienced a decreasing process: The percentage of zhe begins at 40% and falls to 3% from 1064BCE to 100AD. Then the usage increased to 9% in 900AD and then decreased to near 0% percent in 1822AD.

Graph 2: The trend of using zhe as a topic marker in the history (x=Year, Y=Percentage (zhe / Total zhe-hits in each book* 100%))

Similar to graph 1, in graph 2 there are two dots occupying a high percentage among all the dots using zhe as a topic marker: the Yijing (37.99%) and the Chunqiu (33.33%). From the results, the conclusion that the usage of zhe has a long history can be drawn. Yijing is one of the earliest documents that can be found so far, thus zhe carries the topic marker function from a very early age. However, after the Spring and Autumn period, the usage of zhe as a topic marker sharply decreases. Although there is a small rise later in the history, the average of using zhe remains around 6.5%. From the trend, we can draw the conclusion that the topic marker of zhe is not derived from other functions, like being a nominaliser or a demonstrative. The topic marker function exists from the very beginning and lasts to the end of the history.

Graph 3 below shows the proportion of zhe...ye... in all the sentences containing either zhe or ye. The overall percentage is around 50%, which means that nearly half of the sentences with zhe or ye tend to use zhe...ye... to introduce the topic constituent. Since the denominator is different from the previous graphs, the trend is also different. There are three near-zero plots which are the Lun Yu (2.00%), the Story of Zhao Feiyan (0%) and the Miscellany in the Northern and Southern Dynasties
(0%). Apart from these three books, the collocation of zhe and ye is very common in the history of Chinese and the combination of zhe...ye... has an average of 40%. Therefore we can conclude that among all the sentences either has zhe or ye as a topic marker, the combination of zhe...ye... is always preferred by the users.

Graph 3: The proportion of the zhe...ye... among all the examples containing zhe and ye (x=Year, Y=Percentage (zheye / Total zhe+ye-hits in each book* 100%))

4.2 Zhi “之”

4.2.1 Functions of zhi “之”

4.2.1.1 zhi “之” as a verb and a sentence final particle

Zhi can be a verb, a pronoun, and a genitive particle, as in the examples shown in (93) to (95).

(93) 沛公引兵之薛。
Pèigōng yǐn bīng zhī Xuē.
“Emperor Pei led the army to Xue”

(Han Shu: Record of Emperor Gào)

(94) 愿车马衣轻裘，与朋友共，敝之而无憾。

100
yuán chē mǎ yī qīng qiú, yǔ péngyǒu gòng, bì zhī ér wú hàn.
Wish car horse clothes light fur, with friend share, break it however no pity.
“I wish to share my car, horse, clothes and my fur with my friends. Even if they are ruined, I don’t feel guilty.”

(95) 三 年 之 丧 ， 期 已 久 矣。
sān nián zhī sàng , qī yǐ jiǔ yǐ.
Three year ZHI keep vigil, time already long YI.
“Three year’s keeping vigil beside the coffin is already a very long time.”

In example (93), zhi is the verb in the sentence meaning “go, to”. From an English perspective, it looks like a preposition; however, it has the same meaning as zhi (至) and zhi (之) can be regarded as the variant character of zhi (至). In (94), the zhi character here refers to “the car, horse, and clothes shared with friends”. It can be translated to “those (things)”. In example (97), zhi has the meaning of de (的) and it is put between the modifier three years (‘三年’) and the modified part sang (‘丧’).

4.2.1.2 zhi “之” in topic sentences
There are many discussions on “Subject + zhi + Predicate” construction in Chinese (see He, 2004; Li, 2017; Shen & Wan, 2009; Sun, 2016; Xu & Liu, 2007) because in ancient Chinese, this syntactic structure can be widely found in many historical references. This structure can be found in the subject position, the object position, or in the compound clause (Huang, 1998; He, 2004). This structure often indicates emphasis or exaggeration. Some (Sun, 2016) argue that “NP + zhi + VP” cannot be regarded as “Subject + zhi + Predicate” and some (Deng, 2015) reject this claim. However, it would be too perfunctory to just link “NP + zhi + VP” and “Subject + zhi + Predicate”. Mandarin Chinese has been claimed to be a topic-prominent language, as discussed by Li & Thompson (1976). Could “NP + zhi + VP” be a “Topic + zhi + Comment” structure?

Li (2017) states that “NP + zhi + VP” structure can derive three types, which are “NP (Topic) + zhi + VP (Comment)”, “NP + zhi + VP (Topic)” or the whole structure “NP + zhi + VP” being the topic
in the sentence. Due to the topic definition mentioned before, this paper only focuses on the first type, which is the “NP (Topic) + zhi + VP (Comment)” structure.

There are three types of topics as discussed in Frascarelli & Hinterhölzl (2007) which are contrastive topics, aboutness topics and familiar topics. Among these three types, both aboutness topics and contrastive topics can be found in Mandarin Chinese. Zhi can be found both in aboutness topic sentences and contrastive topic sentences, as in examples shown in (96):

(96) 士 之 失 位 也, 犹 诸侯 之 失 国家 也。

shì zhī shī wèi yě, yóu zhūhóu zhī shī guójiā yě.
Knight ZHI lost nobility YE, like duke ZHI lost country YE.
“(If) A knight loses his nobility, just like a duke loses his country.”

(Analects of Mencius)

In (96), this sentence has a syntactic pattern of “NP (Topic1) + zhi + VP (Comment1), NP (Topic2) + zhi + VP (Comment2), …, Comment.” There is a contrast between Topic1 and Topic2, strongly suggesting that these are contrastive topics, and the comments after each zhi make explanations or statements.

However, apart from “NP (Topic) + zhi + VP (Comment)”, “NP (Topic) + zhi + AdjP (Comment)” can also form a topic structure, as in example shown in (97):

(97) 德 之 不 修, 学 之 不 讲, …, 是 吾 忧 也。

dé zhī bù xiū , xué zhī bù jiǎng, …, shì wú yōu yě.
Morality ZHI no cultivate, learn ZHI no lecture, …, is my concern YE.
“(People) Do not cultivate their moralities; and they do not study, …, these are my concerns.”

(Lun Yu)

However, unlike (97), although (98) seems to have a similar structure to (97), this is not a topic sentence:

(98) 霸者 之 民 , 欢 虞 如 也; 王者 之 民, 頌 頌 如 也。

bàzhě zhī mín , huān yú rú yě; wángzhě zhī mín, gāogāo rú yě.
Overlord ZHI people, happy amuse like YE; emperor ZHI people, free like YE.
“The people under an overlord feel happy and amused; the people under the emperor feel free.”

(Analects of Mencius)
There is a difference between (97) and (98): in (97), the sentence has the form of “NP (Topic) + zhi + AdjP (Comment)”. However, in (98), the form changes to “NP₁ + zhi + NP₂”. In this structure, NP₂ is modified by NP₁, therefore, NP₁ is not the topic in the sentence. Zhi has the function of *de* in Mandarin Chinese, i.e. to mark possession.

In aboutness topic sentences, as shown in (99),

(99) 周公之封於魯,為方百里耶…

Zhōugōng zhī fēng yú lǔ, wéi fāng bǎi lǐ yē…

Emperor Zhou ZHIfeofferat Lu, is square hundred miles YE…

“(When) Emperor Zhou was sealed at Lu, (the area) is several hundred square meters…”

(*Analects of Mencius*)

(99) has the form of “NP (Topic) + zhi + VP (Comment)”. To conclude, zhi can be found both in aboutness topic sentences and contrastive topic sentences.

4.2.2 The trend of using zhi “之” as a topic marker over time

In order to find out the position of zhi in the history, we adopted a small-scale corpus study: 34 books (shown in the methodology section 3.3.3) are chosen for analysis. In order to find out the percentages of topic marker zhi in different dynasties, the following variables are analysed: The total number of zhi in each book (Total); the number of zhi as an aboutness topic marker in each book (ATpM); the number of zhi as a contrastive marker in each book (CTpM); the total number of zhi as the topic marker in each book (Tp Sum = ATpM + CTpM); the percentage of zhi being an aboutness topic marker in each book (ATpMP = ATpM/Total * 100%); the percentage of zhi being a contrastive topic marker in each book (CTpMP = CTpM/Total * 100%); and the percentage of all the zhi as a topic marker in each book (Percentage = TpSum / Total * 100%).

Graph 4 below shows the percentages of using zhi “之” as a topic marker in 34 books. The size of the black dots represents the total number of characters that the books contain, and the blue line is the smoothed conditional mean of the usage of zhi as a topic marker. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence of zhi being a topic marker among all the sentences containing zhi in each book (Percentage = TpSum / Total * 100%). Overall the trend of Tp-zhi is decreasing from 1064B to 1850AD from 2.85% to 0.5% from 1064BCE to 1850AD. From the results, the proportion of using zhi as a topic marker is very low throughout history. Although there is a small rise (from 0.9% to 1.1%) from 300AD to
900AD, the ratio is extremely small and the confidence intervals are quite wide in the graph so that this small increase is not very meaningful.

Graph 4: The trend of using zhi as a topic marker in the history of Chinese (x=Year, Y=Percentage (TpSum / Total zhi-hits in each book* 100%))

However, the trend in graph 4 only shows the trend of zhi as a topic marker; we need to find out the proportion of contrastive and aboutness topic marker among all the hits containing zhi. In graph 5 shown below, the trend of using zhi as a contrastive topic marker is presented. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence of zhi being a contrastive topic marker among all the sentences containing zhi in each book (CTpMP = CTpM/Total * 100%). If there is a compared constituent in the latter part of the sentence, the zhi is a contrastive marker. Otherwise, it is an aboutness marker. From the graph, the average usage of zhi being a contrastive topic marker is around 0.45%. The usage of contrastive zhi was 1% in 1046BCE and experienced a fall to 0.25% in around 300AD. After 300AD, the usage of zhi experienced a small rise to 0.36% from 300AD to 900AD, and then fall to 0.1% until 1900AD. Same as graph 1, the small rise in graph 2 is meaningless because the overall existence of the contrastive topic marker is very low: the existence of zhi as a contrastive topic marker reaches 0 around 400AD.
Graph 5: The trend of using zhi as a contrastive topic marker in the history of Chinese (x=Year, Y=Percentage (CTpM / Total zhi-hits in each book * 100%))

When being an aboutness topic marker, graph 6 below shows the trend throughout history. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence of zhi being an aboutness topic marker among all the sentences containing zhi in each book (ATpMP = ATpM / Total * 100%). The average usage of zhi as an aboutness topic marker is around 0.6% in a single text. The usage of contrastive zhi was 1.7% in 1046BCE and experienced a fall to 0.5% in around 200BCE. After 200BCE, the usage of zhi experienced a small rise to 0.6% from 200BCE to 900AD, and then falls to 0.3% by 1900AD. Similar to zhi as a contrastive topic marker, this rise is barely perceptible. The confidence intervals also suggest that zhi being an aboutness topic marker has been stable the whole time from 200BCE onwards.
Graph 6: The trend of using zhi as an aboutness topic marker in the history of Chinese (x=Year, Y=Percentage (ATpM/Total zhi-hits in each book*100%))

Graph 7 combines the trend of zhi as an aboutness topic marker and contrastive topic marker. The blue line is the aboutness topic marker line, and the orange line is the contrastive topic marker line. From graph 4, it shows that zhi as a topic marker, either as an aboutness topic marker or being a contrastive marker, has experienced a fall from 1046BCE to 1900AD. The highest percentages are found in the Book of Changes (aboutness topic marker: 1.7%; contrastive topic marker: 1.2%). What can be conveyed from graph 7 is that Zhi being an aboutness topic marker mainly has a higher usage percentage than as a contrastive topic marker, apart from 630BCE to 4BCE, and here the confidence intervals overlap entirely.
Graph 7: The comparative trend of using zhi as a topic marker in the history of Chinese (x=Year, Y=Percentage 
(ATpM or CTpM / Total zhi-hits in each book* 100%))

4.3 Ne “呢”

4.3.1 The function word ne “呢”

Being a multi-functional word in Chinese, ne can be an interrogative marker, a contrastive marker and an aspectual marker (Jiang, 1986, Qi, 2002, Constant, 2011). There are several positions of ne in the sentence: as a topic marker, this particle is found in the sentence-initial position behind the topic component; as a sentence final particle in declarative sentences, this ne is used as a mood particle (Jiang, 1986; Shao, 1989).

In interrogative sentences, ne is usually found at the end of the sentence, as shown in example (100):
(100) 你吃饭呢？

nǐ chīfàn ne?
you eat NE?

“Are (you) eating or not?”

Ne marker can be found in two positions: either in the middle or at the end of the sentence. When found in the middle of the sentence, according to Constant’s theory (2011), the function of ne is to
mark contrastive topics, and this type of *ne* is marked as *neCT* (Constant, 2011). “NP + *ne*” structure is always found and it is usually found after the subject, forming a pre-subject topic. It has the meaning of “as for…”, indicating a contrast to the constituent mentioned before. This marker can be omitted but does not change the meaning of the sentence, as example shown in (101):

(101) 我 会 弹 钢琴，她（呢），会 拉 小提琴。
    wǒ huì tán gāngqín, tā ne, huì lā xiǎotíqín.
    I can play piano, she (NE), can play violin.
    “I can play the piano, as for her, she can play the violin.”

Another type of *ne* can be regarded as *neASP* (Constant, 2011); under this type, “*ne* serves to intercept a situation between its inception and termination, without focusing on any particular part of the situation’s actualisation” (Constant, 2011: 21), as in the example shown in (102):

(102) 这个 城市 可 大 呢！
    zhègè chéngshì kě dà ne!
    this city very big ne!
    “This is a very big city!”

There are various studies on this function word, especially on the grammaticalisation process of *ne*. The first linguist who proposed a relationship between *er* and *ne* was Li Wang in 1956. He believes that it is very hard to find out the ancestral form of *ne*. In *Han You Shi Gao* (Wang, 2004), Wang argues that *er* is able to be used as an interrogative marker, and that the sound of *er* [njelʔ] is very similar to the pronunciation of present-day *ne*. Therefore it is reasonable to say that there are connections between these two characters. However, there is nearly a 1000-year gap between the appearance of *er* and *ne*; therefore, there is little chance to build up a connection between these two characters (Wang, 2004). In order to find out the further connections between *er* and *ne*, by conducting a corpus study, Jiang (1986) proposed a possible model of the development of *ne*. He points out that *er* is used in Wei and Jin Dynasty, and he found some examples to fill in the 1000-year gap. He also points out that in Song dynasty, *ni* “你” and *na* “那” appear and gradually develop into *ni* “妳”, *ni* “尼” and *li* “裏”. Finally, in Qing Dynasty, *ne* “呢” is frequently used and becomes the function word still used in Chinese today.
4.3.3 Historical change model of *ne “呢”*

Studying the origin and historical development of *ne* is not easy because this function word has gone through a complex historical process of change. Scholars argue that there are two types of *ne*: *ne*$_1$ and *ne*$_2$ (Qi, 2002; Jiang, 1986; Wang, 2004). The difference between these two *nes* is that *ne*$_1$ represents an interrogative force while *ne*$_2$ is used in non-interrogative sentences. *Ne*$_2$ contains the functions of exaggeration and marking the topic constituent in the sentence. However, as discussed in the previous section, there are three types of *ne*, therefore, only classifying *ne* into two categories is not enough. Some scholars combine the aspectual function and the topic marker function together, but when discovering the historical pathway of the development of *ne*, these two functions may point to different origins. To find out the grammaticalisation progress of *ne* more precisely, in this paper, *ne* is separated into three types which are *ne*$_{INT}$, *ne*$_{ASP}$ and *ne*$_{CT}$.

*Ne*$_{INT}$, *ne*$_{ASP}$ and *ne*$_{CT}$ are originally derived from *er “尔”* in ancient times (Zheng-Zhang, 2003; Qi, 2002; Jiang, 1986): Phonologically, according to the old Chinese phonology chart (Zheng-Zhang, 2003: 311), *er (尔)* was pronounced as [njelʔ], which has the same onset as *ne*. Syntactically, *er (尔)* can be used either in the middle or at the end of the sentence, as in the examples shown in (103):

(103) a. 漏子 之 为 善 也, 躬 足 以 亡 尔。
luòzi zhī wéi shàn yě, gōng zú yǐ wáng ěr.
“LuZi loves to do good things, which can cause the country to perish.”

(Gongyang Zhuan: Xuan Gong 15 Year)

b. 功效 卓 尔, 自 左 内 使 初 置 以 来 未 尝 有 也。
gōngxiào zhuó ěr, zì zuǒnèishǐ chū zhì yǐ lái wèicháng yǒu yě.
“The result is outstanding, and had never been seen since the reassignment of ZuoNeiShi“

(Full Han’s Text)

After Han Dynasty, scholars (Qi, 2002; Jiang, 1986) believe that *ne*$_{INT}$ derives from *ni “聻”* and *na “那”* while *li “哩”* is the etymon of *ne*$_{ASP}$ and *ne*$_{CT}$ because written texts were mostly from the South part of China where there was a dialectal difference between */l/ and */n/*. In Tang and Five Dynasties, *ni “聻”* can be found in Zu Tang Ji. It is worth mentioning that *ni “聻”* is also written as *ni “尼”* or *ni “你”*. These words are mainly used in wh-questions. As for *na “那”*, this word is used earlier than
First found in Wei and Jin Dynasties, *na “那”* is used mostly in *yes-no* questions in late Tang and Five Dynasties. There is also a new character *li “裏”* used as an interrogative marker. Until Song Dynasty, *na “那”* and *li “里/哩”* are frequently used in *yes-no* questions and rhetorical questions, they can also be found in some aspectual sentences. In Jin and Yuan Dynasties, *na “那”* can be found in *A-not-A* questions. Interestingly, in Yuan drama, the position of *na “那”* has moved from the middle of the sentence to the end. In conclusion, *neINT* words were found originally in Five Dynasties, and reached a peak in Song Dynasty mainly in Buddhist Quotes. In Ming Dynasty, all the characters mentioned previously had merged to *ne*. In other words, *neINT* began to be used frequently. To show the grammaticalisation of *ne* more straightforwardly, graph 8 below is a model I constructed myself:

![Graph 8: Grammaticalisation of *ne* (self-summarised)](image)

4.3.4 The trend of using *ne “呢”* as a topic marker over time

To find out the grammaticalisation process of the topic marker *ne*, a large-scale corpus study is conducted in two corpora, the BCC corpus and CCL corpus. As in the examples shown above, since topic markers are always followed by a comma, [N + the topic marker + “,”] is used for the query in both corpora. Based on the previous studies, the following characters are under analysis: *ni “你/聻”*, *li “里/哩/裏”*, *na “那”*.

<table>
<thead>
<tr>
<th></th>
<th><em>ni “你”</em></th>
<th><em>ni “聻”</em></th>
<th><em>li “里”</em></th>
<th><em>li “哩”</em></th>
<th><em>li “裏”</em></th>
<th><em>na “那”</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw numbers</td>
<td>19483</td>
<td>16</td>
<td>203401</td>
<td>3105</td>
<td>82</td>
<td>3846</td>
</tr>
</tbody>
</table>

It is worth mentioning that although there are advanced queries in the BCC corpus, the queries do not cover the archaic section. Therefore all the hits are analysed and sifted manually for the second
time. There are nine examples out of 229,933 hits. *Ni “聻”, na “那” and li “哩” are found to be used as topic markers.*

1) Marker *ni “聻”*

The usage of *ne* can be traced back to East Han Dynasty. *Ni “聻”* is usually used as an interrogative marker, as in the example shown in (104):

(104) 师 云: “ 王 老师 聻?” 黄 无 对。

         shī yún: “wáng lǎoshī nī?” huáng wú duì.

teacher say: “Wang teacher NI?” Huang no answer
“The teacher asks: where is Mr.Wang? Huang didn’t answer.”

(*Wu Deng Quan Shu, Qing Dynasty*)

This interrogative marker is often found in the Buddhist quotes. In example (107), being a topic marker, *ni “聻”* in this sentence is also found from a Buddhist quote:

(105) 复 云: “ 佛 聻, 切 忌 话 堕 时。

         fù yún: “ fó nī, qiè ji huà duò shí.

Fu say “Buddhist NI, real abstain say bad time
“Fu says, ‘Being a buddhist, you should avoid saying something bad.”

(*Lv Yanhe Chanshi Yulu Juan, Qing Dynasty*)

In example (105), *ni* is a contrastive topic marker, and the topic constituent is the ‘Buddhist’. The topic constituent sets up a contrast with other people apart from the ‘Buddhist’. The comment part is a further explanation of the topic. This sentence is the only example containing *ni “聻”* as a topic marker.

2) Marker *na “那”*

*Na “那”* has different functions in archaic Chinese. It can be either an interrogative marker or an aspectual marker. As an interrogative marker, *na “那”* is often found at the end of the sentence, as in example shown in (108):
(106) 说：“六哥，他这里还有多少钱那？给他拿了去吧！”

shuō: “liù gé, tā zhèlǐ háiyǒu duōshǎo qián na? gěi tā ná le qù ba!”
Say: “Six Brother, he here still how much money NA? Give him take LE away BA!”

“Say: ‘Six Brother, how much money does he have? Let him take away!'”

(Shi Gong An, Qing Dynasty)

Being an aspectual marker, na “那” has similar functions of ne “呢” in the sentence, and it also appears at the final position of the sentence, as shown in (107):

(107) 延昭 听 罢, 乃 曰: “事 不 偶然 , 此 贼 害 吾 一 家,

yánzhāo tīng bà, nǎi yuē: “shì bù ǒurán , cǐ zéi hài wú yì jiā,
Yanzhao hear over, so say: “thing no coincidence, this thief harm I one family
今 又 来 谋 我 那。”
jīn yòu lái móu wǒ na.”

today again come murder I NA.”

“When Yanzhao finished hearing these words, he says: ‘There is no coincidence in these happenings. This thief harmed my family, today he comes to murder me again.'”

(the Generals of the Yang Family, Qing Dynasty)

In the corpus study, there are two examples of na “那” as a topic marker, as shown in (108):

(108) a. [有牛璘在店肆中算帐…]

[having Niulin settling accounts in the shop] niúlín na, gēn pín dào chū jiā qūlái.
[meaning: come and become a monk with me.)’

(Quan Yuan San Qu, Yuan Dynasty)

b. 夫人 那, “靡 不 有 初, 鲜 克 有 终。”

fūrén na, “mí bù yǒu chū, xiǎn kè yǒu zhōng.”
Madam NA, “without no have initial, rare can have end.”

“Madam, everything has an initial, but rarely has an end.”

(Quan Yuan San Qu, Yuan Dynasty)
Both a and b in example (108) are sentences with contrastive topics. In (108a), the topic constituent ‘Niulin” shows a contrast with other people. The comment sentence of (108a) is the complimentary content of the topic constituent. In (108b), there are three people in the garden, which are the woman, the young lady, and a servant. The topic constituent points to the woman rather than other people. The comment sentence is the advice to the woman.

3) Marker *li* “哩”

Compared to *ni* “聻” and *na* “那”, there are six examples of *li* “哩” being a topic marker. Same as *ni* “聻” and *na* “那”, *li* “哩” can be both an interrogative marker and an aspectual marker, as shown in (109):

(109) a. 那 神 说道: “ 漪水 县 城隍 姓 甚么 哩?”

    nà shén shuōdào: “lìshuǐ xiàn chénghuáng xìng shěnme li?”

    that god say: “Lishui county town-god name what LI?”

    “That god said, ‘What is the god name of the Lishui county?’”

b. 长老 道: “ 原来 你 是 个 城隍 菩萨 哩! ”

    zhǎnglǎo dào: “yuánlái nǐ shìge chénghuáng púsà li!”

    Old man say: “ So you is a town-god Bodhisattva LI!”

    “The old man said, ‘So you are a town Bodhisattva!’”

    *(Sanbao Eunuch's Journey To The West Ocean, Ming Dynasty)*

Being a topic marker, *li* “哩” is put after the nouns, as shown in example (110) below:

(110) a. 古 谱 哩, 不 过 小 色 样 多些; 今 谱 小 色 样 少些。

    gǔ pǔ li, bú guò xiǎo sè yàng duōxiē; jīn pǔ xiǎo sè yàng shǎoxiē.

    Old recipe LI, nothing but little colour style more, now recipe little colour style less.

    “Old recipe, nothing but having more colour schemes. The recipe right now has less colour schemes.”

    *(Jing Hua Yuan, Qing Dynasty)*
b. 茅撒哩，伯也、叔也。

Maosa LI, uncle YE, uncle YE.

“Maosa, a father’s elder brother, an uncle.”

(Rebuild the Taiwan Province, Qing Dynasty)

c. 悟果哩，你也说句话叫奶奶听。

Wuguo LI, you also speak out sentence let grandma hear.

“Wuguo, speak out something to let your grandma listen.”

(Qi Lu Deng, Qing Dynasty)

d. 王中哩，你大爷他原不是惜费的人，但叫他出

Wangzhong LI, your uncle he originally not save money DE person, but call him spend

这宗银子打点书办，他那板直性情…

this part silver bribe scribe, he that straight manner…

“Wangzhong, your uncle is not the person who likes to save money. However, if you tell

他那板直性情…

him to spend money bribing the scribes, his straight temper…”

(Qi Lu Deng, Qing Dynasty)

e. 翠姐哩，孩子是什么病?

Cui sister LI, child is what illness?

“Sister Cui, what’s the child’s problem?”

(Qi Lu Deng, Qing Dynasty)

f. 盘子哩，六大盘也不高兴去吃。

Panzi LI, six big dishes also no happy to eat.

“Panzi, (although there are) six big dishes, he is also not happy to eat.”

(Jingjiang Stories, Qing Dynasty)
Examples a to f in (110) show different kinds of topic. In (110a), the topic marker *li* is a contrastive topic marker: the topic constituent “old recipe” is a contrast with ‘the recipe nowadays’. In (110b), “Maosa” is the aboutness topic, and the comment is the introduction of “Maosa”, showing that “Maosa” is the uncle and an elder brother. In (110c), the topic constituent “Wuguo” is the aboutness topic, and the comment is the explanation of the topic constituent. In (110d), the topic marker *li* is an aboutness topic marker, the comment sentence is aiming to explain the topic “Wangzhong”. Same as (110d), the topic constituents in (110e) and (110f) are aboutness topics.

4.3.5 Section Conclusion
In summary, previous scholars have explored the origins of *ne*, although their studies have been somewhat limited in depth. Through the analysis of a corpus, we have discovered multiple instances of the historical ancestral form of *ne*. However, its usage as a topic marker is extremely uncommon, with only a few scarce examples identified.

4.4 Zhe “者”

4.4.1 functions of *zhe*
In the history of archaic Chinese, the appearance of *zhe* has a very high frequency. Due to its appearance, Chinese linguists discuss the function of *zhe* from a very early age. In East Han dynasty, Xu (1963: 74) first analyses the function of *zhe*: “*Zhe*, is the word indicating people, object and event” 41. In Qing Dynasty, the book *Ma Shi Wen Tong* make further analysis on the function of *zhe*: There are two functions of *zhe*: being a pronoun and being an auxiliary word. There are seven functions of pronoun *zhe*, which are “the first (function) is the initial word of the sentence; the second is the punctuation word; the third is the copula; the fourth is the object of the preposition; the fifth is indicating the first noun (if there are several nouns parallel in one clause); the sixth is the adjunct, the seventh is the hypothesis word” 42. As for the auxiliary *zhe*, it acts as the punctuation marker (Ma, 2010). In recent years, more scholars (Yang, 1928; Wang, 1956; He, 2004; Yang, 1981) have discussed the functions of *zhe*, and they each hold different opinions. No matter how they differ in minor aspects, there is some consensus: The sentence-final *zhe* always acts as a modal

41 “者，别事辞也”。（Xu, 1963: 74）
42 “一为句之起词，二为止词，三为表词，四为司词，五居偏次者，六用若加语者，七有假设词气者。” (Ma, 2010: 50)
particle, as for the sentence-middle zhe, there are two basic functions: being a nominaliser, and being a demonstrative.

Being the nominaliser is the main function of zhe among all the functions. When combining with nouns, verbs, adjectives or numbers, it refers to someone, something or some time, as in the example shown in (111):

(111) 逝者如斯夫！不舍昼夜。

shì zhě rú sī fū! bù shě zhòu yè.

go ZHE like this FU! No discard day night.

“(The time) goes like this and it (goes) disregards day and night!”

(Confucius: Zi Han)

The nominaliser function is the same as the analysis of Xu (1963). In (111), the zhe refer to “the thing”. There is an argument of whether this zhe is a pronoun or a nominaliser, since we are only targeting on the topic marker function of zhe, we are not going to discuss the debate of this function in detail.

The second function is that zhe, being a demonstrative, has the meaning of “this”. This meaning is mostly found in poetry, as in example shown in (112):

(112) 那边走，者边走，莫厌金杯酒。

nà biān zǒu, zhě biān zǒu, mò yàn jīn bēi jiǔ.

that way go, this way go, not dislike gold glass wine.

“Either you go this way or that way, do not be fed up with the alcohol.”

(Zui Zhuang Ci)

The example in (112) is not usually found in archaic Chinese. This function has some overlap with the first function, therefore some linguists (Xiang, 1900; Yang, 1928) get the conclusion that zhe is a demonstrative. This claim is also facing many challenges, with the same reason of the first function, we are not going to discuss into details.

The last function of zhe is that it is a function word indicating the main topic of the sentence:

(113) 北山愚公者，年且九十。

běi shān YúGōng zhě, nián qiě jiǔshí.
North mountain Yu man ZHE, age yet ninety.

“At the North Mountain, there is an old man named Yu, he is ninety years old.”

(*Lie Zi: Tang Wen*)

Being a topic marker, which is admitted by most of the Chinese linguists, it is usually found in the *zhe*...*ye*... construction. The topic marker *zhe* is usually following a people’s name or a proper noun, and the constituents after the phonological stop is the comment, explaining the topic.

Sentence (113) indicates that *zhe* can be an aboutness topic marker. It can also be a contrastive topic marker, as in the example shown in (114):

(114) 象 者, 言 乎 象 者 也; 矢 者, 言 乎 变 者 也。

*tāo zhě, yán hū xiàng zhě yě; yáo zhě, yán hū biàn zhě yě.*

Tao ZHE, say HU phenomenon ZHE YE; Yao ZHE, say HU change ZHE YE.

“Tao, (is the thing) indicating a phenomenon; Yao, (is the thing) indicating change.”

(*Book of Change: Xi Ci*)

In (114), there is a comparison between “Tao” and “Yao”, and the topic marker *zhe* acts as a contrastive topic marker. The example in (114) also reveals an interesting phenomenon: Two different kinds of *zhes* can co-exist in one sentence, one is the topic marker and the other one is the *zhe* indicating a phonological edge.

4.4.2 The trend of using *zhe* as a topic marker over time

In order to investigate the type of topic marker *zhe* throughout history, we adopted 34 books (introduced in methodology section 3.3.3) for analysis. The following variables are analysed: The total number of *zhe* in each book (*Total*); the number of *zhe* as an aboutness topic marker in each book (*Zheatp*); the number of *zhe* as a contrastive marker in each book (*Zhectp*); the total number of *zhe* as the topic marker in each book (*ZheTp Sum = Zheatp + Zhectp*); the percentage of *zhe* being an aboutness topic marker in each book (*ZheatpP = Zheatp/Total * 100%*); the percentage of *zhe* being a contrastive topic marker in each book (*ZhectpP = Zhectp/Total * 100%*); and the percentage of all the *zhe* as a topic marker in each book (*Percentage = ZheTp Sum / Total * 100%*).

Graph 9 below shows the percentages of using *zhe* as a topic marker in 34 books. The blue line is the smoothed conditional mean of the usage of *zhe* as a topic marker. The x-axis is the year when the book was completed (from 1046BCE to 1822AD) and the y-axis is the frequency of occurrence.
of *zhe* being a topic marker among all the sentences containing *zhe* in each book (*Percentage = Zhe Tp Sum / Total * 100%*). From the graph, the overall trend of *Tp-zhe* is decreasing from 1064B to 1850AD from 57% to 2% from 1064BCE to 1850AD. The proportion of using *zhe* as a topic marker throughout history has the average of 37%, which indicates that *zhe* is frequently used as a topic marker. From 1064BCE to around 220BCE, the usage of *zhe* is keeping on decreasing, but from 220BCE to around 1100AD, there is a rise of using *zhe* as a topic marker. It then decreases to near zero from 1100AD to 1822AD, which indicates that *zhe* being a topic marker is gradually dying out in the history.

![Graph 9](image)

**Graph 9:** The trend of using *zhe* as a topic marker in the history of Chinese (x=Year, Y=Percentage (Zhe Tp Sum / Total * 100%))

To look into details, the graph 10 below shows the percentage of *zhe* as an aboutness topic marker (*ZheatpP = Zheatp/Total * 100%*). The trend of aboutness topic marker *zhe* is similar to graph 1, which has a decrease from 38% to 5% from 1046BCE to around 200BCE. It follows a rise from 5% to 12% from 200BCE to 1100AD and then falls again from 12% to 2% from 1100AD to 1822AD.
Different from aboutness topic marker *zhe*, the contrastive topic marker *zhe* has a comparatively stable trend \( (Zhectp = Zhectp/Total \times 100\%) \). In 1046BCE, the usage of *zhe* starts from 13% and gradually decrease to 2% near 80BCE. Then it has a small rise from 2% to 4% from 80BCE to 1000AD and then decrease from 4% to 1%.
Graph 11: The trend of using zhe as a contrastive topic marker in the history of Chinese (x=Year, Y=Percentage

\( ZhectpP = \frac{Zhectp}{Total} \times 100\% \))

When combining graph 10 and graph 11 together, graph 12 below shows the comparison between zhe being a aboutness topic marker and a contrastive topic marker. The graph indicates that the usage of zhe as an aboutness topic marker (blue line) is always having a higher percentage than the usage as a contrastive topic marker (orange line).

Graph 12: Comparison of zhe as a contrastive topic marker (red line, x=Year, Y=Percentage \( ZhectpP = \frac{Zhectp}{Total} \times 100\% \)) and an aboutness topic marker (blue line, x=Year, Y=Percentage \( ZheatpP = \frac{Zheatp}{Total} \times 100\% \)) in the history of Chinese

Which is worth mentioning that in the 34 books, only the functions of zhe being a topic marker and being a nominaliser are found. In section 5.1, why this phenomenon happens is going to be discussed in detail.

4.5 Comparison of the Usage of Different Topic Markers

The usage patterns of the Chinese topic markers zhe, ye, and zhi provide a story of the development of language transformations over thousands of years. Upon analysing the data presented in Graph 13, we can clearly observe major changes in the frequency of these markers from 300 BCE to the 19th century: Before 300 BCE, the topic marker zhe is the most commonly used. Nevertheless, its prominence experiences a substantial decline over time, dropping from 37% to 5% by the beginning
of the third century BCE. At the same time, the usage of ye increases slightly from 2% to 5%, while the usage of zhi decreases from 2% to only 0.2%.

From 300 BCE to 1200 AD, the usage of ye gradually replaces zhe as the predominant topic marker. The frequency of the marker ye experiences a significant increase from 5% in 300 BCE to a peak of 12% around 500 AD, indicating its growing linguistic significance. However, starting from the year 500 AD until 1200 AD, the usage of ye undergoes a decline, decreasing to a mere 6%. During the period from 300 BCE to 100 BCE, the usage of zhe experiences a small decline from 5% to 3%. However, it then gradually increases and reaches approximately 8% by 1200 AD. Zhi, on the other hand, consistently maintains the lowest usage percentage, staying around 0.2% throughout this period.

After the year 1200 AD, the topic marker zhe regains its dominant position, with a usage rate of around 6%. Significantly, both ye and zhi gradually become less well-known and eventually become completely unused during this time period.

Graph 13: Comparison of zhe as a topic marker (blue line, x=Year, Y=Percentage (Zhe Tp Sum / Total * 100%)), ye as a topic marker (orange line, x=Year, Y=Percentage (ye / Total ye-hits in each book* 100%)), and zhi as a topic marker (red line, x=Year, Y=Percentage (TpSum / Total zhi-hits in each book* 100%)) in the history of Chinese

4.6 Section Summary
In this section, we discussed four typical topic markers in Chinese, which are zhi, ye, ne and zhe. The functions of these topic markers are firstly introduced and the usage percentages of these topic markers are presented. From the results, the frequency of using these markers in the history of
Chinese is zhe > ye > zhi > ne. Besides, the topic function of these markers exists from the beginning of the history and is not derived from other functions.
5. Grammaticalisation Process of topic markers in Mandarin Chinese

This section will examine various notable linguistic phenomena in the development of Mandarin Chinese, with a specific emphasis on syntax and the evolution of language. We will focus on the lack of dedicated topic markers in Mandarin Chinese. Our investigation will involve studying how cliticisation and topicalisation interacted in Archaic Chinese. Additionally, we will explore the reasons behind the prevalent use of base-generated topics in Archaic Chinese. In addition, we will examine the process of grammaticalisation of *ne* as a marker of topic and investigate the historical grammaticalisation and degrammaticalisation of *zhe*. These investigations will yield valuable insights into the evolution of the syntax and structure of Mandarin Chinese.

5.1 Grammaticalisation and degrammaticalisation of Zhe in the history

We have provided a concise overview of the grammaticalisation of topic markers in history in section 2.5. *Zhe* stands out among all the topic markers in archaic Chinese due to the fact that, as a referential nominalizer, its origin is a subject of considerable debate, regardless of whether it denotes a topic or not. Since the function of *zhe* has already been described in section 4.4, the focus of this section is on the historical evolution of *zhe*.

5.1.1 Grammaticalisation, linguistic cycle and degrammaticalisation

One approach to language development is grammaticalisation, among others. Grammaticalisation refers to the process through which a lexical item transitions from a symbolic to a practical function. Meillet (1912) introduced the concept of grammaticalisation and provided its definition as “l’attribution du caractère grammaticale a un mot jadis autonome” (Meillet, 1926: 131). Grammaticalisation, to put it another way, is the process of attributing grammatical characteristics to a word that was previously independent. While the terminology employed to depict a language-changing phenomenon may be limited in scope, the process of language development is protracted and gradual. Detecting a language change, on the other hand, requires an enormous quantity of linguistic data. Various categories of linguistic transformations have occurred throughout history. During the past half-century, linguists have been actively engaged in the development of the theory of grammaticalisation and the search for additional linguistic evidence to support the formulation of a more comprehensive definition of grammaticalisation. Kuryłowicz provided the following definition of grammaticalisation: “Grammaticalisation is the progression of a morpheme from a lexical to a grammatical status, or from a grammatical to a more grammatical status” (Kuryłowicz, 1975: 52). They introduced the notion of a “cline of grammaticality” in Hopper & Traugott (2003: 7).
In this cline, Hopper & Traugott (2003) has made a detailed claim of what is happening under the surface changing: The identical item gradually “loses in autonomy by becoming more subject to constraints of the linguistic system” (Lehmann, 2004: 155) and it finally becomes an inflectional affix and loses its dependent position in the word string. Reanalysis and actualisation are two grammatical procedures inside grammaticalisation which help underlying structures convert to the surface level and make it easy to observe. The cline of grammaticality is happening in many languages, including Chinese. This thesis is very curious about what kind of grammaticalisation, especially in topic markers, is taking place, and how will the topic marker develop in the future.

When talking about language development, it is worth pointing out that language is not always developing in a string, in other words, language development is not always a one-way direction, it can have other forms, including linguistic cycle and degrammaticalisation.

A linguistic cycle is an important notion in syntax and it means that “languages do not reverse earlier change but may end up in a stage typologically similar to an earlier one” (van Gelderen, 2011: 7-8). In other words, when the evolved usage is very similar to the original usage of a given word, a phenomenon thus has gone through a linguistic cycle. However, the evolved usage and the original usage are not the same and there still exist differences. Jespersen (1922: 428-429) use “spiral” to describe this cycle, which is very reasonable, because the new usage is located above the original usage, which shows that although the position is similar, the usages are not the same. The notion of the linguistic cycle is under the concept of grammaticalisation. Not all the grammaticalisation examples include linguistic cycles but all linguistic cycles involve grammaticalisation. From this relationship, it is obvious to see that the linguistic cycle is included within the grammaticalisation.

The process of degrammaticalisation is defined as the unidirectional shift from right to left on the cline mentioned above. Through the degrammaticalisation process, an affix is less bound and turns into a lexical word. There are three types of degrammaticalisation which are degrammation, deinflectionalisation and debonding (Norde, 2009). Degrammation is the degrammaticalisation changing from a grammatical word to a content item, and it is appearing in the content level. Deinflectionalisation is the shift from “more grammatical to less grammatical” (Norde, 2009: 133) and it happens at the content-syntactic level. Debonding is the change from a bound morpheme to a free morpheme and it happens at the morphosyntactic level (Norde, 2009: 133). Within these three types, degrammation is mostly found in degrammaticalisation.
Based on an examination of the linguistic cycle and degrammaticalisation, it is evident that language change can result in a variety of outcomes: a unidirectional development, a spiral development, or a regression from an affix to a word, which can be considered an antigression against development. Regarding the linguistic cycle or degrammaticalization of the topic markers, linguists refrain from offering any commentary or analysis during the discourse surrounding the development of Mandarin Chinese topic markers. The objective of this thesis is to determine whether topic markers in Mandarin Chinese have undergone any additional grammaticalization developments through the use of corpus analysis.

5.1.2 Why zhe matters?

As stated in the preceding section, the development of a language can occur in a multitude of ways, and it is exceedingly difficult to forecast the manner in which a language phenomenon will emerge, unless one conducts an in-depth corpus analysis of pertinent linguistic examples. We have analysed the trend of zhe as a topic marker in the history of Chinese in section 4.4. It is quite intriguing to learn that the percentage of occurrences of the topic marker zhe fluctuated from high to low throughout history, and that it gradually disappeared from the historical stage after 1100CE. There are three functions of zhe, which includes serving as the sole nominaliser among all topic markers in section 4 and providing a concise summary of the material covered in section 4.4: To begin with, it functions as a nominalizer. The second function is that zhe, being a demonstrative pronoun, has the meaning of “this”. The last function of zhe is that it is a function word indicating the main topic of the sentence.

In history, zhe being a topic marker is dying out from history but being a nominaliser and a demonstrative pronoun, these two functions are still heavily used. In this part, we are going to raise a hypothesis that whether zhe is having a grammaticalisation process and the predicted semantic changing of zhe is as follows:

Zhe (nominaliser) —> zhe (nominaliser, topic marker) —> zhe (nominaliser, demonstrative pronoun)

This thesis predicts that zhe in the beginning was appearing as a nominaliser. With the development in history, the topic marker function began to emerge and zhe begins to be multi-functional. In the third stage, the topic marker function dies out and zhe being a demonstrative pronoun gradually takes place.
5.1.3 Methodology and Results
In order to test the grammaticalisation process of *zhe*, a three-phase small-scale corpus study was conducted. In the first phase, 13 books were chosen from the corpus:

<table>
<thead>
<tr>
<th>Book of Change (1046BCE)</th>
<th>Lun Yu (475-450BCE)</th>
<th>Zhan Guo Ce (215-185BCE)</th>
<th>Yan Zi Chun Qiu (390-300BCE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang Zi Fa Yan (35BCE-18CE)</td>
<td>Qian Han Ji (200CE)</td>
<td>Sou Shen Ji (317-342CE)</td>
<td>A New Account of Tales of the World (415-440CE)</td>
</tr>
<tr>
<td>Shui Jing Zhu (510CE)</td>
<td>Zhen Guan Zheng Yao (627-649)</td>
<td>Zu Tang Ji (952CE)</td>
<td>Meng Xi Bi Tan (1086-1093CE)</td>
</tr>
<tr>
<td>Hunan Yilao Ji (1194-1240CE)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We selected the materials in a manner that accounted for the completion time of the book and a variety of disciplines. Despite the fact that we are performing a corpus study on a modest scale, these measures ensured that the data returned remains balanced.

When analysing the three functions of *zhe* in the previous books, the result shows that the first two functions of *zhe* are taking over a large proportion of the three functions: The average of nominaliser *zhe* is taking up 93% while *zhe* being a topic marker is used by 7%. *Zhe* being a demonstrative pronoun is rarely seen in the chosen books.

In the second phase, three poetry books are under analysis:

<table>
<thead>
<tr>
<th>The Shijing (11BCE-7BCE)</th>
<th>Quan Tangshi (1705)</th>
<th>Quan Songci (1937)</th>
</tr>
</thead>
</table>

The chosen books are three important poetry books in history: *The Shijing* is the first and oldest poetry book collecting all the poetry used by people from all social classes in their daily life. Dating from the 11th BCE to the 7th BCE, the types of poetry are divided into three types: *Feng*, *ya* and *song*. *Feng* refers to folk songs and areal music, and they are collected from 15 places under the territory. *Ya* refers to the songs which are used by the nobility, these songs are usually used in palace life when they are attending parties and get-togethers. As for *song*, these songs are often heard in the sacrifice ceremony. The content of *The Shijing* covers all the facets of society at that time, thus being a perfect source for analysis. The other two books, *Quan Tangshi* and *Quan Songci*, are collections of different types of poetry from the Tang and Song dynasties. During these
two dynasties, the creation of poetry is prosperous. The choice of these three books can represent the poetry development throughout history.

When analysing three functions of *zhe* in the poetry collections, the result shows that *zhe* is not acting as a topic marker in the poetry:

<table>
<thead>
<tr>
<th></th>
<th>Shijing</th>
<th>Quan Tang Shi</th>
<th>Quan Song Ci</th>
</tr>
</thead>
<tbody>
<tr>
<td>zhe 1</td>
<td>66</td>
<td>2519</td>
<td>586</td>
</tr>
<tr>
<td>zhe 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>zhe 3</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>66</td>
<td>2523</td>
<td>592</td>
</tr>
</tbody>
</table>

The first function of *zhe* is still taking up a large proportion of the three functions, and *zhe* being a demonstrative only has four examples each in the poetry of the Tang and Song dynasties. The result shows that in the poetries in Chinese history, *zhe* is not used as a topic marker, and it is largely used as a nominaliser.

In the last phase, two Buddhist Quotes are under analysis. The two books chosen for analysis are *Sutra of Forty-two Chapters* and *Diamond Sutra*. The reason for choosing these two books is that these two books have the most official translation from Sanskrit to Chinese and the usage of *zhe* can reveal the language usage at that time. The results of the usages of *zhe* are shown in table 3:

<table>
<thead>
<tr>
<th></th>
<th>Sutra of Forty-two Chapters (322-385CE)</th>
<th>Diamond Sutra (868CE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>zhe 1</td>
<td>43</td>
<td>14</td>
</tr>
<tr>
<td>zhe 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>zhe 3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>43</td>
<td>14</td>
</tr>
</tbody>
</table>

From the table above, only the nominaliser *zhe* can be found in the Buddhist quotes, the other two functions are not found in the two books.

5.1.4 Is there a linguistic cycle of *zhe* in Archaic Chinese?
From the results, it is obvious to be observed that there does not exist any linguistic cycle of *zhe* in archaic Chinese: there does not contain any function replacement in the graph and the three functions of *zhe* are kept under usage by generations throughout history. We have the following findings through analysis:
1) Main functions of zhe is being a nominaliser

Through the corpus study, zhe being a nominaliser is the main function of the usage of zhe. In a sentence, zhe is frequently used to refer to a thing or a people, as in the example shown in (115):

(115) 孔子 曰: “益 者 三 乐, 损 者 三 乐。”

kǒngzǐ yuē: “yì zhě sān lè, sǔn zhě sān lè.”

Confucius say: “benefit ZHE three happiness, harm ZHE three happiness”

“Confucius says: There are three kinds of beneficial happiness, and there also exist three kinds of harmful happiness.”

(Lun Yu: Jishi)

(116) 子 曰: “知 者 不 惑, 仁 者 不 忧, 勇 者 不 惧。”

zǐ yuē: “zhī zhě bù huò, rén zhě bù yōu, yǒng zhě bù jù.”

Confucius say: “wisdom ZHE no seduce, kind ZHE no worry, brave ZHE no fear.”

“Confucius says: Wisdom people cannot be seduced, kind people do not worry, brave people do not fear.”

(Lun Yu: Zihan)

In (115), zhe refers to “the thing” and zhe in (116) refers to “the people”. From the previous section, the conclusion can be made that whether in ordinary books or poetry or Buddhist quotes, the nominaliser zhe is the main function throughout history.

2) The topic function of zhe appears at the beginning of the history

From the tables, in the first phase, it is obvious to find out the topic marker zhe appears at the very beginning of history, which is, it is found in the Book of Changes from 1046 BCE. The topic marker zhe is taking up 43.8% of all the examples containing zhe in this book, which is far more than in other books. It is still remaining unclear why the topic marker zhe is taking up such a large proportion at the beginning of history, however, it is undoubted that the topic marker zhe can be found in early history.

3) Being a demonstrative, zhe is rarely appearing in history and it disappears at a very rapid speed

In the table, the demonstrative zhe is only appearing from 415CE-952CE, and in the poetry collection, it only appears in Tang and Song dynasties. Therefore, the conclusion can be drawn that the demonstrative zhe is only appearing for a very short period of time and soon dies out.

4) Different genres have different restrictions on the function of zhe
From the result of the previous section, the poetry genre has a restriction on the topic marker zhe. The reason for the restriction is that there should usually contain a phonological gap between the topic constituent and the topic marker, however, in poetry, due to the strict format, it is unreasonable to contain any gaps inside. Therefore, the poetry genre cannot and should not contain any topic zhe inside.

As for the demonstrative zhe, it cannot be found in Buddhist quotes. The possible reason is that when translating from Sanskrit to archaic Chinese, the demonstrative zhe is not used by ordinary people, thus there is no reason that it can be found in the Buddhist quotes.

5) There is no linguistic cycle of zhe in Archaic Chinese

In the original assumption, we raised the hypothesis that zhe acting as a topic marker appears later than the nominaliser zhe. Due to the number of topic markers zhe decreasing in history, the question that “whether there is a linguistic cycle of zhe in archaic Chinese” is raised. However, from the results, the data show that among the three functions of zhe, there does not exist the phenomenon that a function is emerging and taking over another function. Therefore there is no linguistic cycle of zhe in archaic Chinese. The nominaliser zhe and the topic marker zhe are used from the beginning of the history, and the demonstrative zhe is appearing in a very short time and soon dies out. Since we are only conducting a small-scale study, the conclusion may not be very rigorous. If in the future, a larger-scale corpus study is conducted, hoping the new findings can amend and supplying the conclusion in this part.

5.1.5 Conclusion

This section is discussing whether zhe contains a linguistic cycle through development. By conducting a small-scale corpus study, results show that among all the functions of zhe, they do not show a linguistic cycle tendency: The main usage of zhe is acting as a nominaliser, which takes up over 90% of the history. Zhe is also carrying a topic marker function. The topic marker function can be found from the beginning of history. As for the demonstrative function, it only appears several times and soon dies out after usage.

As for the genre, different genres have a “filtering mechanism” on the functions of zhe: in poetry, the topic and the demonstrative function of zhe cannot be found. In Buddhist quotes, the demonstrative function cannot be found. The linguistic cycle cannot be found in archaic Chinese.
5.2 The Grammaticalisation process of Ne “呢” as a topic marker

Being a multi-functional word in Chinese, *ne* can be an interrogative marker, a contrastive marker and an aspectual marker (Jiang, 1986; Qi, 2002; Constant, 2011). There are various studies on this function word, especially on the grammaticalisation process of *ne*.

The first linguist who proposed a relationship between *er* and *ne* was Li Wang in 1956. He believes that it is very hard to find out the ancestral form of *ne*. In *Han You Shi Gao* (2004), Wang argues that *er* is able to be used as an interrogative marker, and that the sound of *er* [njelʔ] is very similar to the pronunciation of present-day *ne*. Therefore it is reasonable to say that there are connections between these two characters. However, there is nearly a 1000-year gap between the appearance of *er* and *ne*; therefore, there is little chance to build up a connection between these two characters (Wang, 2004). In order to find out the further connections between *er* and *ne*, by conducting a corpus study, Jiang (1986) proposed a possible model of the development of *ne*. He points out that *er* is used in Wei and Jin Dynasty, and he found some examples to fill in the 1000-year gap. He also points out that in Song dynasty, *ni* “你” and *na* “那” appear and gradually develop into *ni* “聻”, *ni* “尼” and *li* “裏”. Finally, in Qing Dynasty, *ne* “呢” is frequently used and becomes the function word still used in Chinese today.

However, from a large-scale corpus study, as shown in section 4.3, the results show that *er* is not the ancestral form of *ne*. Therefore, what is the “real” ancestor of *ne* in the history? In this section, we are going to discuss the grammaticalisation process of topic marker *ne*. Two main questions are waiting to be solved. Firstly, what is the grammaticalisation process of *ne* being a topic marker? Secondly, what are the features of the newly-formed topic markers in the history of Chinese? Apart from discussing two main points, other findings from the result section will be discussed.

5.2.1 The grammaticalisation model of *ne*CT

From the results, *na* “那” being a topic marker is firstly found in Yuan Dynasty in *Loose Songs of Yuan Dynasty*. The exact date cannot be traced but the approximate time of emerging is around 13 to 14 century. *Li* “哩” appears later than *na* “那” and it is mostly found in Qing Dynasty. There is also one example of *ni* “聻” in Qing Dynasty. Based on the previous grammaticalisation model, the new grammaticalisation process is shown in graph 13 below:
Li“哩” being a topic marker in Qing Dynasty is the derived form of li“哩” in the North Song Dynasty. As for na“那”，the topic marker na“那” is newly found in Yuan Dynasty, therefore in Yuan Dynasty, na“那” can be either an interrogative marker, an aspectual marker or a topic marker. Then na“那” becomes ne“呢” in Qing Dynasty. However, there is no obvious evidence of showing that there are connections between na“那” and li“哩”，or li“哩” and ne“呢”，and therefore there are two question marks remaining in the Graph 14.

According to Jiang (1986), all forms of ne“呢” is unified in The Dream of Red Chambers in Qing Dynasty. However, the existence of li“哩” and ni“聻” flouts this claim. It is still remaining unclear whether there is a developmental connection between li“哩” and ne“呢” historically, in other words, it is hard to draw the conclusion that ne“呢” is derived from ni“聻” as a topic marker, therefore, there is a question mark between li“哩” and ne“呢” in graph 14.

The existence of ni“聻” is also worth discussing. In graph 14, there is a connection between ni“聻” in the Five Dynasties and in Qing Dynasty because there is one example found in Qing Dynasty, as in the example shown in (117), however, the connection between ni“聻” being an interrogative marker and a topic marker is very tenuous:

(117) 复 云:“ 佛 聻，切 忌 话 堕 时。

fù yún: “fó ní, qiè jì huà duò shí.
Fu say "Buddhist NI, real abstain say bad time
"Fu says, ‘Being a buddhist, you should avoid saying something bad.””
(Quotes of Lv Yanhe Zen Master, Qing Dynasty)

Different from other examples in the result section, (117) is different since it is an imperative sentence. The comment part of this sentence omits “you should” and the “you” refers to the Buddhist. Besides, ni “聻” does not have the aspectual function throughout the history, the appearance of ni “聻” in (117) is different from li “哩” and na “那”. Therefore, it is reasonable to suspect that ni “聻” in example (117) could be a typo or there is a mistype in this sentence.

5.2.2 Features of topic marker ne “呢”

The newly-emerged topic markers share similarities. Firstly, the position of ne “呢” has experienced a syntactic change. Both li “哩” and na “那” can be both an aspectual marker and an interrogative marker. Discussed by Pan & Paul (2016), the syntactic position of sentences final particle ne “呢” is the same as the interrogative ne “呢”:

```
CP CP AttP
  TP C' 呢
```

However, when the marker is a topic marker, it is appearing after the NP. Therefore, the position of the topic marker is the head of TopP, as shown below:
Unlike the interrogative marker and aspectual marker, the topic marker ne “呢” only c-commands the DP rather than c-command the CP/TP. The projection scope of topic marker is narrower than the other two markers.

The topics in the result section shows that these topics can be formed either by base-generated rather and moved. Example (118) is a base-generated topic. This topic exists a co-reference between the topic and the subject in the sentence. Example (119) is a moved topic from the null co-referred subject in the DP to the topic position:

(118) 悟果 哪，你 也 说 句 话 叫 奶奶 听。

wúguǒ lì, nǐ yě shuō jù huà jiào nǎinai tīng.

“Wuguo, speak out something to let your grandma listen.”

(119) 牛璘 那，∅ 跟 贫 道 出 家 去 来。

niúlín na, ∅ gēn pín dào chū jiā qùlái.

“Niulin, go out of your home with this poor Taoist. (meaning: come and become a monk with me.)”

Therefore, when the topic marker appears in Yuan Dynasty, the marker allows base-generated topics and moved topics. The hypothesis of moved topics appear first is made because in the Loose
Songs of Yuan Dynasty, both examples (119) and (120) contain moved topics rather than base-generated topics:

(120) 夫人 那，“靡 不 有 初，鲜 克 有 终。”

fùrén na,“ mí bù yǒu chū, xiǎn kè yǒu zhōng.’’
Madam NA, “without no have initial, rare can have end.”

“Madam, everything has an initial, but rarely has an end.”

(Loose Songs of Yuan Dynasty, Yuan Dynasty)
These examples are found in 13-14 century while the other examples in the result section are found after 17 century. From the timing perspective, the aspectual function appears later than the interrogative function, therefore the functional grammaticalisation procedure of ne “呢” is:

\[
\text{ne “呢” INT} > \text{ne “呢” ASP} > \text{ne “呢” TopP}
\]

5.2.3 Other findings
The findings also flout some hypotheses and hypothesis put up by scholars: Firstly, Jiang (1989) argues that there is a north-south dialectal distinction between the usage of ne “呢” and li “哩”. In the result section of marker li “哩”, the author of Rebuild the Taiwan Province, Xian Fan is from Zhejiang, the south part of China; the author of Qi Lu Deng, Haiguan Li is from Henan Province, the north part of China; the author of The Marriage of Flowers in the Mirror, Ruzhen Li is from Beijing, the north part of China. In Mandarin we are using nowadays, Southerners prefer to use li “哩” while Northerners prefer to use ne “呢”. However, in the result section, there is no distinction between the usage of li “哩” and ne “呢” between Northerners and Southerners, so the north-south dialectal distinction is not obvious.

Secondly, Jiang (1989) argues that in the Dream of the Red Chambers, ne “呢” is unifying all the ancestral forms of ne “呢”, however, li “哩” is not unified and used nowadays. This finding also shows that this claim is not correct as well. From the results, there is a grey line between li “哩” and ne “呢”, however, there is no evidence showing that there is a character merge between li “哩” and ne “呢”. Thus this claim is still waiting to be proven.
5.3 Why Archaic Chinese Only Has Base-Generated Topics

Section 5.3 will provide a comprehensive analysis of base-generated subjects and moved topics in archaic Chinese. In section 2.4, we have addressed the findings of Huang Li & Li (2009), who concluded that they were unable to determine the specific conditions under which a moved-topic sentence is formed. In an effort to elucidate the forming reason why only base-generated topics are present in archaic Chinese, this section will discuss the matter. In accordance with the discourse presented by Huang, Li & Li (2009), we will address the subsequent inquiries: What are the characteristics of a base-generated topic in archaic Chinese? Furthermore, akin to the initial inquiry, what characteristics should a moved topic possess in the context of archaic Chinese? Furthermore, what particular topic does ancient Chinese encompass? Does archaic Chinese have both base-generated topics and moved topics, or is there just one type in archaic Chinese? Is there a regulation that restricts the type of topic that should be present in the sentence?

5.3.1 Base-generated topics in Archaic Chinese

Prior to examining base-generated topics in archaic Chinese, it is necessary to provide a definition of the term base-generation. Base-generation means that a topic constituent is not moved into Spec,CP from a lower structural position but rather is first Merged or generated there, the syntactic structure of a base-generated topic is shown in Fig.2:

![Fig.2 syntactic structure of a base-generated topic](image)

The syntactic position of a base-generated topic is based on the Spec,CP position and there contains no gap in the TP, which is different from a moved topic:
Usually, in a topic sentence, if there is no gap in a sentence, as in the example shown in (126), the topic is a base-generated topic.

(121) 果，我最喜香蕉。

shuǐguǒ, wǒ zuì xǐhuān xiāngjiāo.

fruit, I most like banana.

‘(As for) fruits, I like bananas most.”

(Huang, Li & Li, 2009: 202)

In (121), the topic constituent shuīguǒ (fruit) is usually considered to be base-generated, since there is no gap in the sentence. The syntactic structure of (121) is shown in Fig. 4 below:
The topic constituent *shuiguo* (fruit) is located in the Spec, CP position and it is not derived by movement. However, whether the sentence in (121) contains movement or not is still under discussion. In Shi (1992), he claims that the sentence in (121) contains movement, but that there is also a subsequent deletion existing in (121) at the same time, and the original sentence according to this approach should like (122) below.

(122) 水果，我 最 喜欢 [( 水果 中 的) 香蕉]。

    shuǐguǒ, wǒ zuì xǐhuān [(shuǐguǒ zhōng de) xiāngjiāo].

    fruit, I most like [( fruit among de) banana].

    “(As for) fruits, I like bananas (among fruits) most.”

    (Huang, Li & Li, 2009: 203)

The syntactic structure proposed by Shi (1992) for (122) is shown in Fig. 5:

![Fig. 5 Syntactic Structure of example (122)](image)

If sentence (122) is derived by movement, the constituent *shuiguo* (fruit) is move to the initial place of the sentence from the Spec, DP position. Since the sentence in (123) is ungrammatical, according to a subsequent deletion operation deletion, zhong de (in + DE) is deleted.

---

43 The syntactic structure of Adv ‘zhong’ is irrelevant to our study therefore will not be discussed here.
(123) *水果，我最喜欷新 [(e中de)香蕉]。

shuǐguǒ, wǒ zuì xǐhuān [(e, zhōng de) xiāngjiāo].

fruit, I most like [(e, among de) banana].

‘(As for) fruits, I like bananas (among) most.’

However, Shi’s (1992) assertion is illogical due to the redundancy introduced by including the movement in sentences such as (121) and the violation of the economic principle by this movement (sentence such as (121) is unacceptable in Mandarin Chinese).

Following the contention that no-gap topic sentences do not utilise the movement method to construct topic constituents, our subsequent focus will be on the base-generated topics found in ancient Chinese. The presence of base-generated topics in archaic Chinese is illustrated in (124) as an example:

(124) 五谷者，[种之美者也]COMMENT。

wǔ gǔ zhě, [zhòng zhī měi zhě yě]COMMENT.

Five cereal, planting these are the most glorious thing (comparing to other cereals).

*(Analects of Mencius, Gaozi part one)*

Example (124) is a clause containing an anaphor that pertains to the object zhi and is an aboutness topic. Additionally, the subject may restate the topic constituent within the comment clause, as illustrated in (125):
The Spec, CP position is occupied by the topic constituent, and the topic marker zhe serves as the head of the TopP. In the case where the topic is generated from the base, its position remains constant, as illustrated in Figure. 6. This typical method of forming a topic constituent, a predicate sentence, is frequently observed in archaic Chinese.

Example (124) is a clause containing an anaphor that pertains to the object zhi and is an aboutness topic. Additionally, the subject may restate the topic constituent within the comment clause, as illustrated in (125):

(125) 回i 也，其i 心 三 月 不 违 仁。

Hui, ye, his heart three month no obey kind.

“Hui, his heart can stay three months without obeying kindness.”

(Lun Yu)

Similar to sentence (124), sentence (125) also comprises an aboutness topic. The comment sentence contains a co-referential indication pertaining to the topic Hui. The position of the topic is identical to that of the example (124).

Apart from aboutness topic, contrastive topic can also be base-generated, as shown in (126):

(126) a. 过i 也，人 皆 见 之i; 更j 也，人 皆 仰 之j.

guòi ye, ren jiē jiàn zhī; gèngi ye, ren jiē yǎng zhī.
Guo, Ye, people all see him; Geng, Ye, people all respect him.

“Guo, is the person people all can notice; Geng, is the person people all respect.”

(Lun Yu)

b. 人不堪其忧, 回也不改其乐。

rén bù kān qí yōu, huí yě bù gǎi qí lè.

People no bear his worry, Hui no change his happiness.

“People cannot bear their worries, Hui cannot change his happiness.”

(Lun Yu)

In contrastive topics, the position of the topic constituent is also situated at the Spec,CP position.

Although the positions of contrastive topics and aboutness topics are the same, the co-referential relationship between the anaphor and the antecedent indicates that in archaic Chinese, there must exist a semantic relationship between the anaphor and the antecedent. In the third part, we are going to discuss the antecedent (topic) - anaphor (pronoun) agreement in detail.

5.3.2 Moved topics in archaic Chinese

The existence of moved topics in ancient Chinese is a subject of considerable debate. Within this segment, we shall present syntactic trees that visually represent the movement of topics within a sentence. We can ascertain the constituent components of each element through a more profound syntactic analysis using these trees.

5.3.2.1 The syntactic position of moved topics

In contrast to base-generated topics, movement was also utilised to generate topic constituents in ancient Chinese. When a pause occurs in a sentence, the topic constituent should be moved, as illustrated in example (127):

(127) 书，我 喜欢。

shū, wǒ xǐhuān.

Book, I like.

In sentence (127), there is a movement from the lower DP to the TopP, as shown in Fig. 8:
Although similar to the examples in 5.3.1, the topic constituent occupies the TopP position, the mechanism of forming the topic is not the same. The movement of the topic constituents involves two mechanisms which are the Extended Projection Principle (EPP) feature and the information structure theory. The EPP feature demands that a clause must contain a determiner phrase (DP) or a noun phrase (NP) in the subject position, and the phrase is checked by the uninterpretable D feature (Adger, 2003).

As previously mentioned, there are various types of moved topics in archaic Chinese. For instance, aboutness topics have the ability to incorporate sentence gaps, as illustrated in example (128):

(128) 地 者, 万 物 之 母 也。
    dì zhě, wàn wù zhī mǔ yě.
    "Earth, is the mother of thousands of things."

(Meng Zì)

In example (128), there is an omission of the subject 'di (earth)”and the clitic ‘wei (being)’, as shown in the example (129):

(129) 地 者, (地 为) 万 物 之 母 也。
    dì zhě, (dì wéi) wàn wù zhī mǔ yě.
    "Earth, (earth is) thousand thing ZHI mother YE.
    "Earth, is the mother of thousands of things.”
The omission constituent is the subject and the predicate, and the TopP must obligatorily move from Spec,TP to Spec,CP, as shown in Fig. 9:

![Syntactic Structure of example (129)](image)

Fig. 9 Syntactic Structure of example (129)

Apart from omitting the subject and the predicate in the sentence, only the subject can also be moved, as in the example shown in (134):

(130) 士之尊賢者也。

shì zhī zūn xián zhě yě.

“Knights, (are the people who) respect the talented people.”

(131) 士之(士)尊賢者也。

shì zhī (shì) zūn xián zhě yě.

“Knights, (are the people who) respect the talent people.”

In contrast to example (128), which retains the subject, example (130) omits the predicate while retaining the subject. Similar to the syntactic tree in Fig. 6, the topic constituent also locates on the DP of the Spec TopP position, and it is also moved from the subject position in the lower DP.

In contrastive topic sentences, when there is a movement in the sentence, the omission of the sequence is also the same as the examples shown in (128) and (130):
(132) a. 天者，高之级也；地者，下之级也。

tiān zhě, gāo zhī jí yě; dì zhě, xià zhī jí yě.

Sky ZHE, high ZHI extreme YE; earth ZHE, low ZHI extreme YE.

“Sky, is the highest thing; earth is the lowest thing.”

(Shi Ji)

b. 天者，(天为)高之级也；地者，(地为)下之级也。

tiān zhě, (tiān wéi) gāo zhī jí yě; dì zhě, (dì wéi) xià zhī jí yě.

Sky ZHE, (sky is) high ZHI extreme YE; earth ZHE, (earth is) low ZHI extreme YE.

“Sky, is the highest thing; earth is the lowest thing.”

(133) a. 鱼，我所欲也；熊掌，亦我所欲也。

yú, wǒ suǒ yù yě; xióngzhǎng, yì wǒ suǒ yù yě.

Fish, I SUO want YE; bear palm, also I SUO want YE.

“Fish, is the thing I want; bear palm, is also the thing I want.”

(Meng Zi)

b. 鱼，(鱼为)我所欲也；熊掌，(熊掌)亦(为)我所欲也。

yú, (yú wéi) wǒ suǒ yù yě; xióngzhǎng, (xióngzhǎng) yì (wéi) wǒ suǒ yù yě.

Fish, (fish is) I SUO want YE; bear palm, (bear palm) also (is) I SUO want YE.

“Fish, is the thing I want; bear palm, is also the thing I want.”

The syntactic position of the topic constituent is the same as the position in Figure 4. To conclude, In all the examples containing a topic constituent, whether these is a gap in the sentence or not, the topic constituent is always situated at the Spec,CP position. If the topic has a topic marker, the topic marker is located at the head of the TopP, and it can be omitted.

5.3.2.2 Do moved topics really exist in archaic Chinese?

After investigating the syntactic structure of both base-generated topics and moved topics, one important question is waiting to be solved: Do moved topics really exist in archaic Chinese? Why there is not a null subject or object in those moved-topic examples, with the topic being base-generated after all? Sadly, there does not exist any criteria to test whether a constituent is moved or base-generated, therefore, whether a topic constituent is base-generated or moved is still remaining unknown. However, we can still try to investigate the syntactic structure to infer the topic type in archaic Chinese. We claim that there is no moved topics exist in archaic Chinese, only base-generated topic exist.
The positioning of the topic constituent at the beginning of a sentence in Chinese syntax functions to introduce the main theme or focal point of the conversation, subsequently followed by the latter part of the sentence that offers supplementary information or commentary. The utilisation of the topic-comment structure holds a crucial place in Chinese communication, as it enables speakers to effectively and efficiently convey intricate concepts.

When there is a moved topic in a sentence, the syntactic structure is shown in Fig. 10:

![Fig. 10 Moved topics in archaic Chinese](image)

However, is it possible that the subject in the sentence “di” is omitted, as in the figure shown in Fig. 11:
In Fig. 11, if there is no movement in the sentence, there must exist a null-subject $e$ in the Spec,TP position.

In Chinese syntax, the phenomenon of null elements, where certain syntactic positions can be empty while still being grammatically valid, plays a crucial role in shaping sentence structure. Null elements, such as null subjects and objects, are often implied by the context of the sentence rather than overtly expressed. This phenomenon has implications for the positioning of topic constituents within Chinese sentences, particularly in topic-comment structures. Huang (1984, 1989) has made many discussions on the Mandarin Chinese null-subject and null-object parameter, also known as the pro-drop phenomenon. This parameter argues that “a pronoun may drop from a given sentence only if certain important aspects of its reference can be recovered from other parts of the sentence” (Huang, 1989: 185). In this parameter, the following sentences are eligible:

(134)

Speaker A: Zhangsan kanjian Lisi le ma?
Zhangsan See Lisi LE Q
“Did Zhangsan see Lisi?”

Speaker B: a. Ta kanjian ta le.
He see he LE.
‘He saw him.’
b. $e$ kanjian ta le.
‘[He] saw him.’
c. ta kanjian $e$ le.
‘He saw [him].’
d.  e kanjian e le.
    ‘[He] saw [him].’
e. wo cai [e kanjian e le].
    ‘I guess see LE’.
    ‘I guess [he] saw [him].’
f. Zhangsanshuo [e kanjian e le].
    Zhangsan say see LE
    ‘Zhangsan said that [he] saw [him].’

Null-constituents are permissible in specific discourse contexts; they can only be identified when the referent is present in the sentence. However, the property of the topic constituent cannot be defined using the null-constituent parameter alone, as the sentences in (134) are permissible in a discourse context. When the sentence is a topic sentence and the discourse environment is absent, the pro-drop parameter is modified:

(135)

书 i, 我 读 了 e i.
shū i, wǒ dú le e i.
book, I read LE ei.

Sentence (135) contains an omitted object denoted as e, which is currently positioned as the object. While both example (134) and example (135) include null-constituents, the manner in which these constituents are formed differs. The sentence (134) contains a predicate sentence. With respect to a topic sentence that is dislocated to the left, Fig. 12 illustrates the syntactic position of every constituent:

Fig. 12 Syntactic positions of different constituents in archaic Chinese.
In this figure, the topic should have a co-referential relationship with either the subject or the object, in other words, either the subject and the object can be replaced by a $e$. The TopP is $c$-commanding the subject, which is situated at the Spec,TP position. Since the TopP is $c$-commanding the subject in a sentence, it is also $c$-commanding the object place which is situated at the lowest place of the tree. Therefore, the hierarchy of projection is shown as follows: $CP > TopP > TP > T'\mapsto VP > DP$ (where the object stays).

The Minimal Movement Principle posits that language generation processes should minimize movement operations to achieve optimal syntactic structures. This principle is particularly relevant in Chinese syntax, where topic constituents play a crucial role in discourse organization and thematic focus. According to the projection mentioned above, the topic constituents in Chinese typically hold a high degree of information structural prominence, serving as markers for introducing topics and focusing discourse. According to the Minimal Movement Principle, if a topic constituent can fulfill its informational requirements in its base-generated position, there is no need for additional movement operations to position it elsewhere in the sentence. By applying the Minimal Movement Principle to Chinese syntax, we can elucidate the base-generated nature of topic constituents. Instead of being moved to achieve their correct positions, topic constituents are inherently placed in their appropriate syntactic locations during language generation. This underscores the efficiency and economy of Chinese sentence structures, as they optimize linguistic expression by minimizing unnecessary movement operations.

In conclusion, it is impossible to find moved topics in Mandarin Chinese. Under the discourse level, the pro-drop phenomenon may exist in Chinese; however, in the case of an unexpectedly constructed sentence, both the subject and topic must be present. Following this, we will examine the limitations associated with formulating topics in Mandarin Chinese. There are additional proofs in section 5.4 that suggest the moved topic should not be present in Mandarin Chinese.

5.3.3 The mechanism of forming a topic constituent in archaic Chinese

In the previous sections, we discussed the formation of base-generated and moved topics in archaic Chinese. We also claim that in archaic Chinese, moved topics do not exist and only the base-generated topics exist. In this section, we will try to find out the rules of forming a topic constituent.

The Extended Projection Principle is first put forward by Chomsky (1981) to constrain the requirement of a subject in a sentence and it is a supplement of the projection principle. This theory
is stated in Minimalist terms and substantially postdates Chomsky (1981) in this form. As discussed in the previous section, this principle defines: The sentence must contain a DP in its Spec,TP position and it should be checked by an uninterpretable D feature [uD*] on the T level. However, when dealing with a topic-prominent language like Mandarin Chinese, this theory seems invalid because the subject does not need to be overt in a sentence: under some cases, either the topic or the subject can exist in a sentence; under some cases, both the topic and the subject can co-exist in a sentence, so this principle is not powerful enough to cover all the linguistic situations. Under this circumstance, can we upgrade the EPP feature to a stronger feature which can constrain both constituents in the sentence? In this section, we are going to shed some light on a modified version of the existing EPP feature to enable it to account for topic-prominent languages.

The original EPP principle (Chomsky, 1993; Adger, 2003) is only arguing the uD* feature on T (the head). To fulfil this principle with the topic constituent, we try to upgrade to a version of EPP feature targeting on Mandarin Chinese:

(136)  
CEPP feature (stands for: Chinese EPP feature): A clause must contain either  
a determiner phrase (DP) in the subject position, checked by uD*  
or a determiner phrase (DP) in the topic position (the topic marker can be omitted in TopP)  
If the topic position and the subject position both exist in one sentence, The TopP occupies the Spec,CP position.

To test the CEPP feature, we are going to look at the following examples:

(137) a. 父母者，人之本也。  
fùmǔ zhě, rén zhī běn yě.  
Parents ZHE, people ZHI root YE.  
“Parents, (they are) people’s roots.”  
(Shi Ji)

b. 娄者，呼万物且内之也。  
lóu zhě, hū wàn wù qiě nèi zhī yě.  
Lou ZHE, call all things and pull them inside.  
“Lou, (it is) calling everything and pulling them inside.”  
(Shi Ji)
The sentences in (137) illustrate different types of topic sentences: (137a) and (137b) only contain topics, and in (137c), both the topic and the subject can be found in the sentence. Any of (137a), (137b) or (137c) must contain either a topic or a subject in the sentence to make the semantic meaning of the sentence complete. In (137c), when both the topic and the subject are present, the topic position is higher than the subject position, because the topic constituent yue "music" appears before the subject constituent yin "sound". And, crucially, specifiers are to the left.

So far the CEPP feature seems reasonable. However, there is another question appearing in this CEPP principle: in Mandarin, there also exist sentences where both the topic and the subject cannot be found, as in the example shown in (138):

(138)
取平阳，得魏王母妻子，尽定魏地，凡五十二城。
qǔ PíngYáng, dé Wèi wáng mǔ qī zǐ, jìndìng Wèi dì, fán wǔshíèr chéng.
"(Cao Can) Gets Pingyang, obtains the family of the Wei emperor and pacifies the Wei land, containing all together fifty-two cities."

In example (138), there is no subject or topic in the sentence, and the reason is that in the very beginning of the discourse, there is one sentence:

(139) 平阳侯曹参者，沛人也。
píngyáng hóu cáocān zhě, pèi rén yě.
"The Pingyang Duke Caocan, is a person from Pei."

Since this sentence is the introduction sentence of the Cao Can anecdote, either the subject and the topic is omitted in some sentences (like (139)) in the rest of the article. Reanalysing the syntactic structure of the sentence (138), there are two omissions taking place: the first omission is the eliding
of the topic constituent, and the second omission is the pro-drop of the subject. When both the topic and the subject can be covert in a sentence, then the CEPP feature should be revised. Before revising the CEPP feature, we should solve several questions: Under what circumstance can a deletion take place? And which deletion comes first, the topic deletion or the subject deletion?

The first question is very easy to answer: just like the topic chain sentences, when there is a specified and defined DP at the very beginning of the discourse; and when either the subject and the topic belong to this and only this specified DP, they can be omitted. From this explanation, the omission of either the topic or the subject constituent is more from a pragmatic level rather than a syntactic level because the omission of the sentence-initial DP can make the communication more effective: the referent of the specified DP is stored in the CG between both the hearer and the speaker, so they can rapidly refer to the DP when the subject or topic is null. As for the second question, which deletion comes first? The topic omission or the subject omission? This is not very easy to answer.

To find out which omission comes first, we should look into the formation of both the subject and the topic constituents. In archaic Chinese, to form a subject is always easier than to form a topic constituent: the topic constituent needs a DP + topic marker (can be omitted), and a gap in a sentence. Then it can be followed by a subject and the rest of VP. As for the subject constituent, the VP usually immediately follows it. Therefore, creating a topic sentence usually needs more effort. The minimality principle infers that “given a choice between two comparable operations, the smallest is chosen” (Baltin & Collins, 2008: 51). Besides, since the topic constituent and its head can be covert, according to the Procrastinate principle (Chomsky, 1993) within the Timing Principle (Baltin & Collins, 2008): the covert movement is less costly than overt movement (Baltin & Collins, 2008: 55). Consequently, it is justifiable to deduce that the subject constituent ought to come first in the order of omission followed by the topic constituent. As a result, a more refined iteration of the CEPP is illustrated in (136), which is further detailed in (140) below:

(140)  
CEPP feature revised version: A clause can contain (but not necessarily contain) either  
       a determiner phrase (DP) or a noun phrase (NP) in the subject position, checked by $uD^*$  
       or a determiner phrase (DP) or a noun phrase (NP) in the topic position (the topic marker can be omitted in TopP)  
If the topic position and the subject position both exist in one sentence, The TopP occupies the Spec,CP position.
When there is an omission of both the topic constituent and the subject constituent, the topic omission comes earlier than the subject omission.

Positions are assigned to both subjects and topics within syntactic structures. The migration or displacement of a subject to assume the role of a topic is therefore not possible according to conventional sentence construction. Subjects and topics maintain their fixed syntactic positions, which eliminates the potential for subject-to-topic migration. The CEPP feature aims at ensuring the positions of subjects and topics in ancient Chinese: Subjects and topics are assigned distinct and predetermined slots in the syntactic principles that regulate sentence construction. This serves to reinforce their respective roles within the structure of the sentence and prevents any transposition or interchange between these designated positions.

5.3.4 Section Conclusion

We reviewed sentences containing aboutness topics and contrastive topics in this section. An examination of the syntactic trees reveals that the placement of these subjects is comparatively consistent: they are all positioned at the Spec,CP level. Furthermore, we asserted that the sentences in ancient Chinese adhere completely to the principle of Binding theory, with the exception of one rule that is present in the relocated topics: the topic constituent restricts the subsequent predicate. When the subject is changed, the predicate will be omitted if it is a clitic; otherwise, it should remain in the comment clause. In the following section, the relationship between cliticisation and topicalisation in archaic Chinese will be examined.
5.4 The interaction of cliticisation and topicalisation in Archaic Chinese

An additional noteworthy phenomenon emerges upon examination of the topic sentences. Our research reveals that the topic component of a sentence restricts the number of copula that can be included. A verb is constrained when a topic constituent is present in the TopP position and that topic constituent was previously the subject of the comment clause. If a copula follows the initial topic constituent, the copula is also omitted when the topic constituent is moved to the Spec, TopP position. In the event that a predicate follows the initial topic constituent, it shall remain unomitted. This section will look into the phenomenon of copula deletion in ancient Chinese.

5.4.1 Introduction

When discussing the topicalisation in Mandarin Chinese in the previous sections, one type of sentence structure cannot be neglected: the cleft sentences. As discussed in the second section, using a copula is a typical way to form cleft sentences. In English, there are two types of cleft sentences: *it*-cleft sentences and *wh*-cleft sentences:

(141) a. It is September that I started my new semester.

   b. What he likes is cooking.

In *it*-cleft sentences, the constituent after either *it* or *what* is emphasised. In (141a), “September” is the information under stress, and the construction of (141a) is [It is DP that XP]. Different from (141a), in *wh*-cleft sentences, the constituent after *is* is emphasised. Thus, in (141b), the speaker wants the hearer to know a male person’s preference is cooking rather than the other things, and the syntactic construction of (141b) is [What XP is DP]. In Chinese, predicate sentences are similar to the *it*-cleft sentences: the emphasised constituent is situated at the initial place of the sentence, as shown in the example in (142) below:

(142) 蚩尤 者，炎 帝 之 后。

   chīyóu zhě, yán dì zhī hòu.

   Chiyou ZHE, Yan emperor ZHI after.

   “Chiyou, (is) the descendant of Emperor Yan.”

   *(Classic of Mountains and Seas: Abroad West Classic)*

In sentence (142), *chīyóu* is the topic of the sentence and followed by a topic marker *zhe*. Sentence (142) has the structure of [DP Top (phonological pause) XP]. The translation of (142) has omitted a copula “is”. In such kind of predicate sentences, when there is an aboutness topic followed by no notional verbs in the sentence, the copula is usually omitted. When there is a notional verb in the sentence, as in the example shown in (143), the verb is not and should not be omitted.

(143) 士 之 尊 贤 者 也。

152
In (143), there is no phonological gap between the TopicP and the comment, which proves that the structure of archaic Chinese predicate sentences can have the structure of [DP Top XP] without a phonological pause.

Regarding the copula phenomenon, different linguists have different opinions: some linguists believe that there is no copula in archaic Chinese (Wang, 1940; Xiang, 1993), but some believe that words like 为“为”, 是“是” are copula or carries the copula function (Tian, 2008; Xu, 1981; Li, 1995). In this part, we aim to solve the following questions: Does archaic Chinese have copula? If so, what is the syntactic position of copula? Why can the copula be omitted? What is the mechanism of this omission?

To answer these questions in a more fine-grained way, in the second part, the null-copula phenomenon is going to be discussed, including the definition of copula and the analysis of archaic copula. In the third part, to analyse the copula-omission phenomenon, a small-scale corpus study is conducted. In the discussion section, the following questions are waiting to be answered: What is the syntactic position of the copula in the sentence? What are the interactions between the topic and the copula? And what causes the omission of these copulae in the sentence.

5.4.2. Null-copula phenomenon in Archaic Chinese

5.4.2.1 What is a copula?

Copula, derives from a Latin word, means the “link” or “connection” of two different things. In linguistics, the existence of a copula ties the subject and the subject complement together in a sentence. A typical copula in English is “to be”, and any equivalent forms of “to be” in other languages can also be regarded as a copula. In Mandarin Chinese, 是“是” is widely accepted as the copula, as shown in the example in (144):

(144) 我是老师。
    wǒ shì lǎoshī.
    I be teacher.
    “I am a teacher.”
In (144), the word *shi* acts as a linking verb between the subject *wo* “I” and the complement *laoshi* “teacher”. Usually, when a copula exists in the sentence, the sentence usually has the form of [DP Cop XP]. XP refers to the predicate position, and it can be filled by either a DP or a non-DP constituent, as in the example shown in (145):

(145) a. 我 的 爱好 是 读书。
    wǒ de àihào shì dúshū.
    “My hobby is to read books.”

b. 这 个 苹果 是 爸爸 给 我 的。
    zhè gè píngguǒ shì bàbà gěi wǒ de.
    “This apple is (the one) father gives to me.”

In (145a), the XP constituent is a VP or other verbal constituent *dushu* “read” and in (145b), the XP constituent is a clause. From (145), we can get the conclusion that “a copular construction can be understood as form and meaning pair that entails a proposition with the nominal semantics of specificational and predicational” (Zhan & Sun, 2013: 755).

To conclude, it is uncontroversial that present-day Mandarin exists a copula, and the copula in Mandarin only has the meaning of “to be” compared to other notional verbs: the function of these copulae is to identify the functional pro, and it is also a connection of the subject and the complement in the sentence.

5.4.2.2 Archaic Chinese copula
The study of archaic Chinese copula has experienced several stages. The first stage of the study of the Chinese copula mainly focuses on the copula *shi* “是”, which is the widely accepted copula used from archaic Chinese to modern Chinese nowadays. Being a widely-accepted copula in modern Chinese, as mentioned in the previous section, Chinese linguists have already conducted a thorough study of this function word. Then linguists begin curious about what the copula looked like in the archaic Chinese. Taking *shi* as a starting point, linguists begin to find out the historical form and try to model the grammaticalisation process of *shi* being a copula nowadays. In the second stage, linguists begin to expand their research on finding other topic markers in different dynasties. From the second stage, the Chinese linguists”research is separated into two different directions: The first direction focuses on the historical origin and the grammaticalisation process of a specific copula.
The second direction targets identifying whether the found function word is the copula or not. In this part, we are not only going to introduce these studies but also provide an introduction to the null-copula phenomenon.

Whether there exists any copula in archaic Chinese is under much debate: The first research on archaic Chinese copula starts from the book of *Mashi Wentong* which was published in 1896, and the research only focuses on the copula *shi*. This initial discovery has provided a great start to copula research. However, Wang (1940) argues that there does not exist any copula in archaic Chinese, which strongly contradicts the initial claim. Wang (1940) indicates that the original predicate sentences do not contain any copula and this is demonstrated by many historical materials. Besides, the copula originated in the Six Dynasties, discussed by Wang (1940). However, his claim was overturned by himself in 1957: He reclaimed *shi* origins from the Han dynasty. However, he still did not change the claim that archaic Chinese does not contain any copula. With this claim being credited by some linguists, different voices begin to emerge during the same period: In 1957, Hong (1957) claims that archaic Chinese must have a copula and *shi* was the most typical copula. The original time of finding such a copula was in the Han Dynasty. Also in the year of 1957, Liu (1957) was supporting Hong’s claim that copula did exist in the early Chinese period. However, he made a new assumption that the copula *shi* could be found earlier than Han Dynasty: the first copula *shi* could be found after Qin Dynasty. However, he did not provide any specific time range of its appearance. Liu (1957) is not the only one who thought *shi* was appearing around Qin Dynasty. In Zhao (1957, 1959), he also agreed with the claim that archaic Chinese copula existed. He first claims that *Shi* originally appeared between Qin and Han Dynasty, in 1959, he changed his claim that *shi* being a copula was first appearing in the *Classic of Poetry*, the first poetry book in Chinese history in the West Zhou Dynasty. In 1975, Fagao Zhou revisited the *shi* copula, and he claimed that this copula first appeared in Han Dynasty. Same as Zhao (1959), Xu (1981) also agreed with the claim that the copula *shi* first appeared in the *Classic of Poetry*. The table below summarises the current research results of the first appearance of copula *shi*:

<table>
<thead>
<tr>
<th>Year</th>
<th>Linguist</th>
<th>First appearance of copula <em>shi</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1937</td>
<td>Li Wang</td>
<td>Six Dynasties</td>
</tr>
<tr>
<td>1956</td>
<td></td>
<td>Han Dynasty</td>
</tr>
<tr>
<td>1957</td>
<td>Cheng Hong</td>
<td>Han Dynasty</td>
</tr>
<tr>
<td>1957</td>
<td>Shiru Liu</td>
<td>After Qin</td>
</tr>
<tr>
<td>1957</td>
<td>Lizhe Zhao</td>
<td>Between Qing and Han</td>
</tr>
<tr>
<td>1959</td>
<td></td>
<td>Classic of Poetry</td>
</tr>
</tbody>
</table>
Table. 3 current research results of the first appearance of copula *shi*

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Era</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>Fagao Zhou</td>
<td>Han Dynasty</td>
</tr>
<tr>
<td>1981</td>
<td>De’an Xu</td>
<td>Classic of Poetry</td>
</tr>
<tr>
<td>1991</td>
<td>Yuming Tang</td>
<td>Late Warring States Period</td>
</tr>
<tr>
<td>2007</td>
<td>Xie</td>
<td>Late Han Dynasty</td>
</tr>
<tr>
<td>2019</td>
<td>Zhang</td>
<td>Late Warring States Period</td>
</tr>
</tbody>
</table>

In table 3, the copula *shi* first appears in Han Dynasty has the most supporters, following by appearing in the late Warring States Period and in the *Classic of Poetry*.

When one type of linguists are researching the first copula appearance of *shi*, another type of linguists are putting their focus on whether the copula function is derived from any other functions. The most acceptable analysis is that the copula function of *shi* derives from the demonstrative function to the copula. “In the Early Qin period, subject usually use *shi* to refer them and add the predicate complement behind the copula […] *Shi* is usually situated between the subject and the predicate (complement), which is gradually creating the copula function44” (Wang, 2011). The change from demonstrative to copula is a well-evidenced grammaticalization pathway, as shown by work by Diessel (1999), van Gelderen (2011), and others.

Apart from defining in which historical point *shi* being a copula begins to appear and used by people, scholars like Zhou (1959) think that copula do exist in archaic Chinese, and there does not exist only one copula. There is also another copula being discussed by Chinese linguists: *wei* “为”.

Comparing to *shi*, whether *wei* is a copula under much debate. In 1937, Wang claims that *wei* is only carrying the copula function, it is not occasionally used in the predicates sentences. The main function of *wei* is a verb, indicating “function as”. However, in the study of Hong (1957), he rejects this claim and believes that *wei* and *shi* are both archaic Chinese copulas. *Wei* being a copula does not only have one supporter, in Yang & He (2001), they also make the claim that *wei* is situated between the subject and the predicate comment, and it should be a copula. In (Xie, 2007), he provided the examples and reasons why *wei* should be a copula:

44 “在先秦时期，主语后往往用’是’字复指，然后加上判断语 […] ‘是’字经常处于主语和谓语之间，这样就逐渐产生出系词的性质来。”
According to the grammar rules in Early Qin Dynasty, if the interrogative pronoun is being the object, it should be moved before the predicate verb. Before Han Dynasty it is a relatively restricted rule. If *wei* is a verb in the example sentence above, the interrogative pronoun *shui* should be moved before the verb to form a ‘somebody *shui wei*’ construction. The reason why there is no fronting indicates that *wei* is not a notion verb but a copula” (Xie, 2007: 54)”

This paper is also adopting this explanation and believes that *wei* is acting as a copula in the archaic Chinese.

However, the problem found in the previous studies is that from Wang (1956), he claims that Chinese topic sentences usually omit the copula. Once this claim was reported, it gained many supporters and no one has further investigated why the copula and the topic constituent cannot co-exist in archaic Chinese topic sentences. However, in Wang (1956), he does not discuss whether the copula is obligatory or not since it is perfectly possible in principle for a language to have a non-obligatory copula. That is, if a language often omits the copula, that doesn’t imply that the language has no copula. To fill in this gap, we assume that there must exist a constraint between the topic and the copula constituent. Such constraint means that the topic and copula cannot co-exist and the copula must be bound to the host to its left. In the next section, we are going to conduct a small-scale corpus study to test this claim.

### 5.4.3 Corpus analysis of the archaic Chinese copula in topic sentences

To analyse the omission of the copula in archaic Chinese, a corpus study is conducted. This corpus study includes the combination of all the topic markers mentioned in section 4 and the two copulae *wei* and *shi* mentioned above. By using the same dataset shown in section 3.3.3, a total number of 54,789 hits are found. From these hits, 495 hits have the combination of [topic marker, *wei*...] and 342 hits have the combination of [topic marker, *shi*...]. Table 3 below shows different combinations of topic markers and copula and how many hits are found in the corpus:

<table>
<thead>
<tr>
<th>Target Structure</th>
<th>Numbers containing the target structure</th>
<th>Target Structure</th>
<th>Numbers containing the target structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>Zhi, wei... 之，为...</code></td>
<td>380</td>
<td><code>Li, shi... 哩，是...</code></td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4. Different structures of [Topic marker, copula] and numbers of hits in the corpus

However, this combination is a mixture of different structures, take the [zhe, wei...] construction as an example:

(146) a. 天 地 者，为 万 物 父母。

Sky earth ZHE, is myriad thing parent.
“The sky and earth, is the parent of everything.”

(Wen Xin Diao Long: Yi Proof, Vol. 1)

b. 称「丈」者，为 丈夫。

Call ‘Zhang’person, is husband.
“The one who called ‘zhang’ is the husband.”

(Zhu Zi Yu Lei: Confucian Story)

Although both the examples in (146) contain the zhe, wei...structure, the interpretation of zhe is not the same: in (146a), the zhe is a topic marker. However, in (146b), the zhe is a nominaliser. Thus, a total number of 837 examples will be sifted again manually to obtain the results. From the 1148 examples, our aim is to find out how many [Topic marker, copula] structures exist in history.

The result indicates that only 2 out of 837 examples have the structure of [Topic marker, copula] and both of the examples exist in the same book: The Popular Stories in the Three Kingdoms (year 1321):
In (147), both of the topic markers zhe are aboutness topic markers, and each of them is followed by a copula after the phonological phrase edge. The existence of the copula is to link the comment to explain what the topic is about.

The result reveals two findings: First, the existence of [Topic marker, copula] exists but occupies an extremely small percentage of 0.239% (= 2/837*100%) in the dataset. This result indicates that the omission of copula is largely found in the history. Secondly, only the combination of [zhe, shi...] and [zhe, wei...] are found among all the examples. This finding reveals that there must be some restrictions between the copula and the topic marker. Lastly, the only two examples are from the same book, which is The Popular Stories in the Three Kingdoms (year 1321). Why there are no other examples in the same dynasty? What can we know from this result? Many questions are waiting to be answered. In the next discussion part, the following questions will be solved: Why these two examples are found in one book? Is it a coincidence? Besides, why cannot the [Topic marker, copula] structure be easily found in history? Is there any restriction between the copula and the topic marker?

5.4.4 Discussion

In the previous section, the result has shown a robust tendency of the copula-omission in archaic Chinese: when there is a topic situated at the initial place of the sentence, the copula is usually omitted. In this section, the reason why this phenomenon can be found is discussed. To fully answer this question, we should start by looking into the syntactic position of the copula in the topic sentences. The next step is the interaction between cliticisation and topicalisation in topic sentences is discussed. Several questions are waiting to be answered: When the copula is acting as a phonological clitic, what characteristics should it meet?
5.4.4.1 Syntactic position of copula in topic sentences

To find out why the copular can be omitted in the topic sentences, looking into the syntactic position of these copula is always necessary and fundamental: From the syntactic position, the hierarchy of different constituents is very obvious. If movements are found, the pathway of this movement will also be clearly shown in the tree.

If there is no copula, a normal topic sentence should have the structure in Fig. 13:

Fig. 13 Sentence Containing a Topic

The topic constituent is situated at the Spec,CP position and it includes the DP and the topic marker. The subject constituent is lower than the topic and can be found in the Spec,TP position. If having a sentence like (148), the syntactic tree should be like in Fig. 14:

(148) 士 之 尊 贤 者 也。
    shì zhī zūn xián zhě yě.
    Knight ZHI respect talented people YE
    “Knights (are people who) respect talented people.”

(Analects of Mencius)
However, when there is a copula in a topic sentence, as in the example (148) shown below, where should this copula locate?

(149) 天地者，为万物父母。

tiān dì zhě, wéi wàn wù fùmǔ.

Sky earth ZHE, is myriad thing parent.
“The sky and earth, is the parent of everything.”

Placing the copula in the V position seems reasonable, since the copula can be regarded as a type of linking verb. However, when facing sentences like (154), the place of copula should be reconsidered:
(150) 三爷者，为仆所役使之人。

sānyé zhě, wèi pú suǒ yìshǐ zhī rén.
Sanye ZHE, is servant SUO enslave ZHI people.
“Sanye, is the people who is enslaved by the servants.”

(Qing Bai Lei Chao: Zi Zang)

If we still put the *wei* in the VP position as shown in Fig 14, the sentence will have duplicated constituents on the VP position, which is not acceptable. Since the *wei...suo...* construction is a passive construction, which will bring other constraints on different constituents into the syntactic structure, and no other proper examples can be found from the BCC corpus, we are going to analyse the position of the copula from a Mandarin Chinese’s view:

(151) 他，是吃过饭的。

tā, shì chī guò fàn de.
He, is eat pass meal DE.
“He, has already ate.”

In example (151), if *shi* is located at the Spec,VP position, *eat* will lost its Spec,VP position, which is the place it should be situated at. Therefore, combining the fact that *shi* appears at the left of *chi*, it is reasonable to make the hypothesis that the position of this copula should be higher than the VP and is located at the CopP (stands for Copula P), as in the Fig. 16 shown below:

Fig. 16 Syntactic structure of (151)

Therefore, to conclude, the syntactic position of the copula should be c-commanding the lexical verb, as shown in Fig. 17.
5.4.4.2 The interaction between cliticisation and topicalisation in topic sentences

In the previous section, we discussed the syntactic position of several constituents: the topic marker, the copula and the notion verb. We drew the conclusion that the copula should be c-commanding the notional verb. The difference of these syntactic positions is showing that in Chinese sentences, copula has its specific position and notion verb cannot be placed at the copula place. In this part, we are going to move forward a step and discuss the interaction between cliticisation and the topicalisation in a topic sentence.

As shown in the result section, the copula shows a very strong tendency on omission when appearing in the topic sentences in archaic Chinese, sentences like (149) has an extremely low rate for appearance. However, as discussed in the second part, previous Chinese studies do not provide any reason on why this omission is taking place and can frequently be found in the predicate sentences. Comparing sentences containing topics with containing [subject + copula] structure:

(152) a. 夫心为万事主，动而无节则乱。

fú xīn wéi wàn shì zhǔ, dòng ér wú jié zé luàn.
FU heart is myriad thing host, move but no constraint thus chaotic.
“Heart is the host of everything, if you do things without constraint, thus (you will be) chaotic.”

(New Books of Tang, History Collection)

b. 地者，万物之母也。
dì zhě, wàn wù zhī mǔ yě.
Earth ZHE, thousand thing ZHI mother YE.
“Earth, is the mother of thousands of things.”

(Analects of Mencius)

The omission of the copula in (152b) is crucially because of the suprasegmental. Accent, stress, rhythm, tone, pitch and intonation are the six main feature of suprasegmental (Fox, 2002). The existence of these features can assign words extra meanings and emotions in a discourse in order to keep the listeners engaged. In sentences like (152b), there is a stress at the topic position and a phonological break between the TP and the comment. However, in (152a), since there is no phonological break, the copula appears. This phenomenon indicates that there must exist a constraint between the topic and the copula. Based on this phenomenon, we propose that the copula in the archaic Chinese also has the phonological feature and it must be bound to a host to its left with the same intonational phrase.

In the previous section, we discussed the syntactic structure of topic sentences containing a copula:
When observing the binding constraint between the subject and the copula, the tree above should add the feature of [+COP] between the copula and the subject. In the study of Wang (2011), he mentioned that the copula is placed between the subject and the predicate complement, we are going to use this claim as a starting point.

When assuming that the copula is following the subject rather than the topic, it is reasonable to answer the question of why in predicate sentences, copula should not exist. When there exists a topic and a prosodic boundary, the copula is omitted. Therefore, we can infer that the copula and the topic constituent do not have the same feature. If the copula contains the interpretable [+cop] feature, only the subject can take the [+cop] feature in order to check the same feature. In the previous sections, we claim that the topic is prior to the subject, thus when an NP is situated at the initial place of the predicate sentence, it first falls to the subject position and automatically gains a [+cop] feature:
Under this formation rule, the normal subject-predicate sentences are created, such as in (10a). When there exists a topic and this NP is moved from the subject position to the topic position in the sentence, only the copula contains the [+cop] feature and the phonological boundary is put between the topic and the copula. With this prosodic edge, the [+cop] feature cannot be moved with the topic, thus the copula is out of alignment.

When both the topic and the subject exist in the sentence, the [+cop] function is only related to the nearest subject constituent. Although in Chinese, the copula does not have any phi-features including number, gender or case, the feature of the copula is still being assigned by the subject position.
By this rule, why the copula and the topic cannot co-exist can be explained, and this can be considered as a supplementary proof of why moved topics cannot exist in archaic Chinese.

It is worth mentioning that in some sentences like (153), why topic and copula can both exist? This is because *wei* in this sentence has duplicate functions: In archaic Chinese, the *wei...suo...* construction is a typical passive construction. Although the *wei* in the sentence (153) seems like a copula, it is containing other functions and contains a little copula function. This meets the claim of Wang (1937) that *wei* is being a copula in only certain type of sentences.

(153) 三爷 者，为 仆 所 役使 之 人。

sānyé zhě, wéi pú suǒ yìshǐ zhī rén.
Sanye ZHE, is servant SUO enslave ZHI people.
“Sanye, is the people who is enslaved by the servants.”

(Qing Bai Lei Chao: Zi Zang)

5.4.4 Conclusion

We have thoroughly examined the null-copula phenomenon in this section. Since the existence of copula in archaic Chinese is the subject of considerable debate, we have demonstrated that they do in fact exist. The findings of the corpus analysis demonstrate that the omission of copula is a common occurrence throughout history. In light of this discovery, we proceeded to examine the syntactic placement of the copula in an effort to provide an explanation for the omission. A phonological clitic, the copula *wei* or *shi* in Archaic Chinese, requires a host to the left of it within the same intonational phrase. Due to the intonational break that separates topics in Spec,CP from
the remainder of the clause, they cannot function as hosts for the copula; thus, the clause becomes illogical when Spec,TP is vacant.
5.5 Why do dedicated topic markers not exist in Mandarin Chinese?
In section 2.5.3, inspired by Liu’s work (2004): “Where does the Topic Marker Come From? A Continuation of the Commonality and Individuality of Grammaticalization”45, why dedicated topic markers do not exist in Mandarin Chinese interested me very much. Although no scholars have carried out research on this topic either because in order to analyse the synchronic part of the topic marker development, it is important to know how the topic marker arose in order to study a thorough way. However, it is still worth investigating because, among all the topic-prominent languages in the world, many of them contain dedicated topic markers, i.e. the markers are only used in the topic sentences and do not appear in any other types of sentences including interrogative sentences, negative sentences and so on. In this part, beginning with presenting different types of topic markers in different languages and dialects, from the second part, we are going to find out the universal developmental features of forming the dedicated topic marker by conducting a case study. The next part is about the grammaticalisation cline of Mandarin Chinese topic marker development, and the last part will discuss where Chinese topic markers should go. In this part, we will focus on discussing current linguists’ predictions on Chinese topic markers and the pros and cons of these predictions. Besides, this part also contains a discussion on why Mandarin Chinese does not contain any dedicated topic markers like Japanese and what has caused this phenomenon to happen.

5.5.1 Dedicated topic markers in different languages and dialects
Being a topic-prominent language, it is necessary to have a specific position for the topic constituents, even if the topic marker is unnecessarily appearing in the sentence. Languages like Mandarin Chinese, Japanese and Korean have their topic markers to indicate the topic constituents, but the functions of these markers vary.

In section 4, we introduced the different functions contained in archaic Chinese topic markers. This section also shows that Chinese topic markers are not pure topic markers. They can also appear in other sentences, such as interrogative sentences or narrative sentences:

(154) a. 晚 饭 吃 什么 呢?
   wǎn fàn chī shénme ne?
   late meal eat what NE?
   “What's for dinner?”

45 《话题标记从何而来？——语法化的共性与个性续论》
b. 知 之 为 知 之，不 知 为 不 知，是 知 也。
zhī zhī wéi zhī zhī, bù zhī wéi bù zhī, shì zhī yě.
know it is know it, no know is no know, is intelligence YE.

“When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it, and this is knowledge”

(154a) is a sentence containing the topic marker *ne* as a sentence-final interrogative marker, and in (154b), *zhī* being a demonstrative is not the topic marker and it appears after the verb, indicating the “thing” one knows or does not know.

Japanese, which is also another topic-prominent language, has two types of markers: One overtly marks the topic constituent and the other one marking the subject, as shown in the example in (155):

(155) コンスタンツ は 大学 が 良い。
Konstanz wa daigaku ga ii.
Konstanz TpM university SbjM good.

“Konstanz, a good university.”

In (155), the topic marker *wa* is a marker which only indicates the overt topic constituent which is situated at the initial place of the sentence, and *ga* is marking the subject, which is *daigaku*, the university. It is worth indicating that the Japanese topic marker *wa*, which is spelt as *ha* but should be pronounced as *wa*/wa/* when being the topic marker, is a pure topic marker, and it can appear in different kinds of sentences containing the topic constituents:

(156) a. あなた は バナナ が 好き です か?
Anata wa banana ga suki desu ka?
You TpM Banana SbjM like is KA?

“As for you, do you like bananas?”

b. 私 は 雪 が 白い と 思う
watashi wa {yuki ga shiroi} to omou.
I TpM snow SbjM white when think.

“As for me, I think that the snow is white.”

c. 先生 は 誰 ですか?
sensei wa dare desu ka?
teacher TpM who is KA?
“The teacher is who?”

In examples in (156), the Japanese topic marker *wa* can appear either in interrogative sentences, like (156a), or subordinate clauses, like in (156b), or it can appear with the presupposition and mark them, as in the example shown in (156c). In (156c), the *sensei* is already known and must be given in the discourse. Thus it can be marked by *wa*.

Apart from Japanese, different Chinese dialects also have topic markers. We borrow some examples from Liu (2004) of both the Shanghai dialect and the Wu dialect. The former is only in the Shanghai district and the latter is distributed in southern Jiangsu, Shanghai, most of Zhejiang, southeastern and southern Anhui, northeast Jiangxi and northwest Fujian area:

(157) a. 有 个 厂 末 关脱，有 个 主人 末 逃 走，弄 得 一 败 堕 地。

      you geq cang  me  guaite, you geq zinin  me  do  zoe, nong de  i  bai  do  di.
      have CL factory TpM close, have CL host TpM runaway, make DE one fail fall ground.
      “One factory is closed, one host ran away, (they are) suffering a crushing defeat.”
      (Liu, 2004: 292, revised translation)

b. 依 啊，假使买 小 菜 的话 一定 要 早 一点 去。

     nong a, jiasima xio cei eu iding nya  zeo inyan qie.
     you TpM, if buy little vegetable say must should early some go.
     “You, if you go to buy vegetables, you should go early.”

However, different from the topic marker *wa* in Japanese, the topic markers in Chinese dialects, just like the topic markers in Mandarin Chinese, can appear in other sentences like interrogative sentences or as a sentence-final modal particle. In other words, Chinese dialects, like Mandarin, do not have dedicated topic markers. So why is there such a phenomenon? In order to answer this question, firstly we should investigate the origin of these topic markers.

5.5.2 Different Origins of Forming a Topic Marker

Although there are many topic-prominent languages in the world, these languages basically have topic markers to mark topic constituents, but we know little about the sources of these markers. Research on topic markers in Japanese has so far largely focused on synchronic research, so linguists know very little about where this topic marker came from and what kind of grammaticalisation it went through. In Kuteva et al. (2019), there is very little discussion of topic markers. Other topic-prominent languages, because they are particularly niche, such as Lahu (a
Sino-Tibetan language, mainly spoken in Thailand), Formosan (mainly spoken in Taiwan District), Tauya (mainly spoken in Papua New Guinea) and so on, no tradition of topic marker research in these languages and in lack of the historical records. Because there is no tradition of topic marker research in these languages and lacking the historical records also increase the difficulty of analysing these markers. These gaps in the level of language knowledge also add a lot of difficulties to the research of topic markers. But such research is very important because, through the sources of these markers, we can find out the commonality of the development of these topic-prominent language topic markers. Besides, we can roughly infer what kind of words, especially function words will develop into topic markers. If the topic markers in some languages have different origins, we can also compare the topic markers in different languages horizontally to analyse the similarities and differences between these development paths. However, since there are not many studies on historical topic markers, these conclusions cannot be obtained for the time being. This part aims to introduce the sources of topic markers in several representative topic-prominent languages, based on the research of Radetzky (2002), Liu (2004) and Deng (2015). By looking at these sources, we can infer that not all linguistic topic markers developed in the same way. Some of them come from locative words, and some come from interrogative words. The difference in these origins also leads to the different usages of topic markers in different languages.

5.5.2.1 Locative Markers as Topic Markers
Radetzky (2002) has proposed that through a cross-linguistic study, the topic-prominent languages all follow a grammaticalisation cline of:

locative marker > contrastive marker > topic marker

This change can be found in Quechua, Lahu and some Formosan languages. In the Huallaga Quechua, which is spoken in Peru, the topic marker -qa originated from the locative marker -qa:

(158) a. hana-qa-a

upper-qa-1. Pos
“above me”

(Radetzky, 2002: 10)

b. Chaynaw yaykuptinpis wakin runa-qa fiyu…

like.that though.he.enter some men-qa bad…

Wakin runa-qa huklaapa qshipipaykun.
other men-qa to.some.other.place they.escape.on.them

Wakin runakuna-qa alli.
other men-qa good
Chay kaq-qa hamachin alli parlan.

that which.is qa they.seat.them good they.converse

“Even though he enters like that (in the proper way), some men are bad.... Other men escape to some other place. Other men are good. Those seat him and speak nicely to him.”

(Radetzky, 2002: 11)

c. Mayuyaqshi chayaykun. Y mayu-qa chayashashi aywakkuykaanaq
to.the.river he.arrived and river qa full it.was.going

“(The fox) arrived at the river. And the river was swollen.”

(Radetzky, 2002: 12)

In example (158), (158a) is the locative -ga. In (158b) and (158c), (158b) is the contrastive marker example and in (158c), the -ga marker is used as a topic marker, more precisely, an aboutness topic marker.

Apart from Huallaga Quechua, languages like Eastern Kayah Li, mainly spoken in Thailand, and Middle Korean all follow this cline of locative marker > contrastive marker > topic marker.

5.5.2.2 Interrogative Markers as Topic Markers
As for languages like Mandarin Chinese and Chinese dialects, Xu (2004) indicates that interrogative markers are the most important origins of developing into a topic marker. The interrogative markers transforming into topic markers can be found either in Mandarin Chinese and Chinese dialects like Wu and Shanghai dialects. Take Mandarin Chinese as an example:

(159) a. nǐ chīfàn ne?
you eat NE?

“Are (you) eating or not?”

b. tā ne, zài lā xiǎotíqín.
she NE, is play violin.

“She, is playing the violin.”

In the work of Xu (2004), he argues that in dialects and Mandarin Chinese, all the interrogative markers that can develop into topic markers are definitely not coincidences. Another reason is that some of the sentence-final modal particles which do not contain the interrogative functions cannot be topic markers. At the same time, the development of topic markers is a one-way development direction, that is, interrogative particles can become topic markers, not vice versa. Because there
exist topic markers that cannot be used as interrogative particles in other varieties., there are other sources of topic markers in Chinese dialects (Xu, 2004). In other words, topic markers can be interrogatives in Mandarin Chinese but not in other dialects.

5.5.2.3 Time Adverbs as Topic Markers
Time nouns serving as topic markers mainly appear in Chinese dialects. According to Xu (2004), “The time component is naturally suitable as a topic, especially a framing topic46” (Xu, 2004: 297). Gasde (2001) has also come to similar conclusions. In Mandarin Chinese, the timing adverb is shi “时，time”. During the development of Chinese, shi first developed from a time marker to a hypothetical conditional marker and then developed into other more functional usages including topic markers (Xu, 2004; Jiang, 2002):

(160) 臣妾 饮 时，号 曰 发装 酒…
chenqie yin shi, hao yue Fazhuang jiù…
Female subject drink time, name call Fazhuang wine…
“When I (the female subject) was drinking, (the wine) is named as Fazhuang wine…”
(Xu, 2004: 298)

The shi in the Liancheng Hakka dialect is the most grammaticalised topic marker:

(161) 春兰 时 系 老 婢子 $#55。
Chunlan TpM is always servant.
“Chunlan is a servant.”
(Xu, 2004: 298)

5.5.3 The Grammaticalisation Cline Topic Marker in Mandarin Chinese
In the first two parts, we discussed the topic markers contained in different languages and the universal development of topic markers in other languages. In this part, we will discuss the grammaticalisation pathway of topic markers and why Chinese does not have a dedicated topic marker.

Different from the grammaticalisation cline put up by Radetzky (2002), the development of Chinese topic markers seems to be completely different from the language he studies: In Chinese, the topic marker development follows the overall grammaticalisation cline of:

46 “时间成分是天然适合充当话题的成分，特别是做框架性话题” (Xu, 2004: 297)
The reason for putting up this cline is because in section four, the data presenting section, it is shown that the Chinese aboutness topic marker sentences are significantly earlier and much more than the contrastive topic marker sentences. Then the next question waiting to be answered is which function words are the function words that develop into Mandarin topic markers?

Liu (2004) has concluded that all the Chinese topic markers are developed from interrogative words. But are all the topic markers in Chinese really developed from interrogative markers? In the fourth section, we looked at different topic markers and explored their functions. Through discussion, we can clearly find that they are not all used in interrogative sentences: only ye "也" and ne “呢” are interrogative words, while zhe “者” and zhi “之” are not. Even though the advanced intellectuals carried out a Literary Revolution during the New Culture Movement, they changed written Chinese from classical Chinese to vernacular Chinese in the late 1910s. in vernacular Chinese, zhe and zhi are no longer used. Although there are differences in language usage, it does not mean that there exists a gap in language development. During the development of language usage from the 20th century, topic markers were left over, and some are still used today. Therefore, it is debatable to conclude that topic markers are transformed from interrogative words only by studying language development in the past hundred years. If we look at the grammaticalisation process of topic markers in Mandarin Chinese and analyse from a perspective of grammaticalisation, Chinese topic markers are not only converted from interrogative words but also from nominalisers and pronouns:

It is a pity that there is very little research on topic markers in Japan, otherwise, we can improve the development process of Chinese topic markers through the development of topic markers in archaic Japanese: Because in the Tang Dynasty and later, there were many cultural exchanges between China and Japan, and there were also much cross-cultural communication and influence. The nature of topic markers in Japan can reflect the nature of topic markers in archaic Chinese. Therefore, the topic markers in both languages are influenced by language contact.

It is reasonable for interrogative words to become topic markers because in ancient Chinese, interrogative words that have become topic markers already existed. But this is not the only development path for Chinese topic markers. Research shows that pronouns and nominalisers are also sources of topic markers, due to the existence of zhe “者” and zhi “之”, although the topic
markers still being used in Modern Chinese is no longer using these topic markers. Thus, the grammaticalisation cline of Mandarin topic markers is:

function word (origin)
interrogative markers ➔ aboutness topic markers ➔ contrastive topic markers
nominaliser
demonstrative pronouns (gradually leaving out the nominaliser and demonstrative pronoun)

5.5.4 Where should the Chinese Topic Marker go?
The concluding segment will delve into the future direction of the Chinese topic. The trajectory of language in the future is unpredictably determined due to the impossibility of forecasting the evolution of human language patterns (Sánchez-Stockhammer, 2015). Therefore, it is impossible to predict whether additional varieties of Chinese topic markers will emerge in the future. Presently, there appears to be widespread consensus regarding the following statement: the Chinese topic marker is an incomplete form that is undergoing further development and may be simplified further or combined with other function terms. This claim is initially presented by Liu (2007), whose work continues from Liu (2004).

Liu (2004) indicates that “Due to the frequent and widespread use of topic markers in topic-prominent languages, subsequent bleaching usually occurs, and it is mainly causing the lexicalisation in which the topic markers are joined with the preceding topic constituents. Lexicalisation usually happens on function words^47^” (Liu, 2004: 309). In this paper, he raised three possible future developments of the topic marker, which are: Firstly, Demonstrative + topic marker —> resultative sentence conjunctions, as in the example shown in (162):

(162) a. (我听说你要去) 如果 那样, 我 也 去。  
(I heard you are going there) rúguǒ nàyàng, wǒ yě qù.  
(I heard you are going there) If such, I also go.  
(I heard you are going there) “If so, I will also go (there).”  
(Liu, 2004: 310)

b. (听说你要去) 那(样)么, 我 也 去。  
(I heard you are going there) nà(yàng) me, wǒ yě qù.  
(I heard you are going there) Such TpM, I also go.

---

^47^ “由于话题标记在话题优先语言中的高频、广泛使用，因此也常发生后续虚化，而且主要表现为与其前面的话题融合，而导致的词汇化所构成的主要是虚词。” Liu (2004: 309)
(I heard you are going there) “If so, I will also go (there).”

(Liu, 2004: 310)

In (162a), which is a hypothesis sentence, if not using the demonstrative like ruguo nayang “If so”, it can also be combined with the topic marker me “么” and become sentence (162b). The name “If so” in sentence (162b) is becoming a resultative sentence conjecture because “I will also go” indicates a decisive result of the speaker. This ongoing development pathway can be found both in Mandarin Chinese and in dialects like Wu dialect.

The second predicted way of topic marker development is X + topic marker —> Compound sentence conjunction, as shown in the example in (163):

(163) 普通人要么种田，要么做生意。

pǔtōng rén yàome zhòngtián, yàome zuò shēngyì.

Normal people either farm, either do business.

“Normal people either farm or do business.”

(Liu, 2004: 315)

In sentence (163), because the topic marker has the contrastive function, by combining with the hypothesis function word yao “要”, the combination of yaome not only can express a hypothesis function from yao, but can also highlight its contrastive function brought by the topic marker me. The me in (163) is not a topic marker and its topic marker function is totally lost in this sentence.

The last future direction of topic marking is: topic marker+modal particle —> compound modal particle. This is usually found in dialects like Wu:

(164) 难末好哉！三个局还勿曾去，老旗昌咿来叫哉！

This MO good ZAI, three CL party still no yet go, Laoqichang YI come call ZAI.

“Great, I haven’t go to three parties, Laoqichang is coming to call me.”

(Liu, 2004: 318)

In (164), the combination is the “mo…zai” structure. In this structure, zai is the modal particle, adding the exaggerative force to the sentence. In this way, the topic marker marking function is totally lost in (164).

However, according to his conclusion, there is no longer a topic marker in Chinese, but it is further fused with other components in the sentence to become a function word. That is to say, in the
process of further development of Chinese, topic markers are about to disappear, and the disappearance of topic markers has caused Chinese to lose a shred of strong evidence as a topic-prominent language. So will Chinese really become a non-topic-prominent language in the future? The answer is no. In the second section, we mentioned that Chinese topics and subjects have their own positions in syntax. And in the fourth section, we also found that Chinese is a topic-prominent language from the very beginning because topic markers can be found in the earliest manuscript Book of Changes. These obvious pieces of evidence point to the fact that Chinese will definitely not become a non-topic-prominent language in the near future, the process of grammaticalisation never stops and new topic markers will exist to indicate the topic constituent. Therefore, the theory of topic markers changing into function words will not be true.

5.5.5 Why there is no dedicated Topic Marker in Chinese?
The last question waiting to be solved in this section is, will Mandarin Chinese produce a dedicated topic marker like Japanese? The answer is also no. According to the work of Radetzky (2002), he claimed that Japanese topic markers were originally developed from locative markers. The advantage of developing from locative markers is that locative phrases can be used as adjunct phrases, so these locative markers do not necessarily exist in sentences. If these markers do not necessarily exist in the sentence, these markers can be omitted at any time. However, the situation between the Chinese and Japanese is not the same: being a nominaliser or a demonstrative pronoun, compared to adjuncts, they are more likely to appear in a sentence and are an essential element. Even if functions as an interrogative marker, some of them like ma “吗” are needed in the sentence in order to indicate what type of interrogative sentence is. The different attributes in the sentence mean that the developing trend of the topic marker in Chinese is completely different from that in Japanese: Since the topic marker in Japanese is an adjunct and can be omitted, even if this locative marker becomes a topic marker, there still exist other types of locative markers filling in the “gap” and they will not affect the main structure of the sentence. Becoming a dedicated topic marker thus is very efficient for communication because it does not cost any meaning change in the sentence. However, since the original forms of Chinese topic markers are very important components in sentences, it is impossible for a Chinese topic marker to become a dedicated topic marker and only indicate the topic constituent. This leads to the phenomenon of why the topic marker also carries other functions apart from the topic marker function.
6. Conclusion

6.1 Findings

Over the course of the last five decades, the scholarly investigation pertaining to Chinese topic markers has seen a transformation, progressing from its nascent stages to a more comprehensive and thorough examination. The initial focus of the study pertained to a pragmatic notion, and as a result of its distinctive syntactic nature, investigations into syntactic topic markers were also conducted in conjunction with the pragmatic research. Within the field of pragmatics, a topic serves as a communicative mechanism employed to capture the listener's attention towards the content being expressed. In discourse, it is customary to position the topic constituent at the beginning of a sentence. This allows the listener to efficiently comprehend the most crucial information, regardless of whether it is new or previously known, in the shortest possible time. As a result of this characteristic of sentence-initial position, scholars in syntax initiated investigations into the syntactic distinctiveness of topic constituents.

The definition of the topic marker utilized in this thesis aligns with the generally accepted understanding among linguists and does not engender controversy. For a sentence containing a topic marker, its initial form is:

\[
\text{NP + [Topic Marker (can be omitted)], comment}
\]

Sentences conforming to this structure have been identified inside the \textit{Yi Jing} dating back to 1046 BCE, a correlation that aligns with our own study discoveries. There are several benefits associated with the utilization of this particular style of topic marker phrases for research purposes. The oldest topic sentence pattern, which can be observed by analyzing the historical frequency of such sentence distributions, indicates that Chinese has maintained a subject-prominent structure since its inception. The significance of this discovery lies in its relevance to topic research, as it addresses the ongoing debate among linguists over the presence of a subject element in Chinese syntax.

There are several benefits associated with the utilization of this particular style of topic marker structure in research: The earliest topic sentence pattern, which can be observed by analyzing the historical frequency of such sentences, suggests that Chinese has maintained a subject-prominent structure since its inception. The significance of this discovery lies in its relevance to topic research, as it addresses the ongoing debate among linguists over the presence of a subject element in Chinese syntax.

Based on the preceding contextual information, this thesis presents the subsequent discoveries:

a. Mandarin Chinese exhibits co-occurrence of the topic and subject constituents, with the topic constituent only forming through base-generated approach and not through movement.
How topic markers are generated in history is one of the questions waiting to be answered by linguists who are interested in historical Chinese syntax. The reason why they disagree about whether the topic components are relocated or base-generated is that they need to find the syntactic location of the topic and the subject position. Through our investigation, we made the argument that the subject and the topic each have their own syntactic position: the topic is placed in the Spec,CP position, while the subject is in the Spec,TP position. These positions indicate that if there is a lack of either the subject or the topic in the sentence, the corresponding position is replaced by a null constituent Ø rather than being replaced by any other elements. To be more specific, we argue that only base-generated topics exist in archaic Chinese, and moved topics cannot be found under any circumstance. We have put up a completed version of the EPP feature: CEPP, which is, a clause can contain (but not always contain) either

- a determiner phrase (DP) or a noun phrase (NP) in the subject position, checked by ¿D*
- or a determiner phrase (DP) or a noun phrase (NP) in the topic position (the topic marker can be omitted in TopP)

If the topic position and the subject position both exist in one sentence, the TopP occupies the Spec,CP position.

When there is an omission of both the topic constituent and the subject constituent, the topic omission comes earlier than the subject omission.

This CEPP feature provides the priority relationship between topic and subject, which is strong theoretical support to the claim that only base-generated topics exist in archaic Chinese.

a. The topic markers in Mandarin Chinese have undergone a process of semantic bleaching, distinguishing them from topic markers in other languages. The function of the topic marker in Mandarin Chinese is derived from its interrogative function. This suggests that, in comparison to topic markers in other languages, Chinese topic markers cannot easily develop into dedicated topic markers. Instead, they serve as topic markers while carrying out other functions.

b. The clitics and the topic constituent have a constraint on each other and this constraint indicates that the copula in archaic Chinese must have a host to its left within the same intonational phrase. The clitic may also be referred to as a phonological clitic that exhibits a left-leaning behavior.

A small-scale corpus study has been conducted on various clitics in archaic Chinese, which has led to the identification of a novel generalization on the interplay between topicalisation and cliticization in Archaic Chinese, as illustrated by example (165):

(165)女 为 君子 儒，无 为 小人 儒。
rǔ wèi jūnzǐ rú, wù wèi xiāorén rú.
You be gentleman man, not be villain man.
“You should be regarded highly by a gentleman rather than a villain.”

(Lun Yu)
The copula in question can be used after a subject, but it is never observed directly following a topic that is clearly identified by a topic marker. Extensive research has been conducted to explore the underlying reasons for this phenomena. Based on an examination of several instances, it is posited that non-topicalized subjects in Archaic Chinese typically occupy the specifier of TP, whereas topics are situated in the specifier of CP, in accordance with established conventions in the scholarly literature (e.g., Huang, Li, & Li, 2009; Shi, 1992). Additionally, we put up the proposition that the copula in Archaic Chinese can be classified as phonological clitics, which necessitate the presence of hosts to their left within the confines of the same intonational phrase. Given that topics in Spec,CP are marked by an intonational break, they are unable to function as hosts for the clitic in the phrase. Consequently, the sentence is deemed ungrammatical in cases where Spec,TP is devoid of any content.

6.2 Innovation, Limitations and Problems waiting to be solved

6.2.1 Innovations

This thesis has the following innovations:
This thesis contains various original contributions, including the following:
There aren't many instances of syntactic trees being included in Chinese syntactic works. Nevertheless, a sizeable number of syntactic trees have been incorporated into this thesis. Even if at first look this theory does not appear to be particularly novel, it actually includes a wide range of extremely significant consequences. The requirement to provide responses to multiple questions is the key element that slows down the formation of a syntactic tree in Chinese syntax during the analysis phase. Is there any reference to the concept in the Chinese language? Is there such a thing as a subject? In the event that there is a connection between the topic constituent and the subject constituent, what exactly is the nature of that connection?

It is not very typical for the syntactic tree that is associated with a particular concept to become apparent in situations in which the subject matter relates to a pragmatic notion. This is because the syntactic tree predominantly bears a feature that is based on the subject. Within the context of the Chinese language, the distinction between subject and topic as independent syntactic conceptions is investigated within the scope of this thesis. Both the subject and the topic occupy separate positions,
with the topic's position being given priority over the subject's position in any given discussion. In those cases where a sentence does not have a clear topic but does have a subject, the topic position is indicated by the symbol, which indicates that it is present even though it does not have any precise substance. While the topic is located in the complement position of the specifier of the specifier, the subject is found in the tense phrase position of the specifier of the specifier. The novelty of this thesis resides in the fact that it summarizes and improves upon past research, as well as correctly integrating these theories into the creation of each syntactic tree for archaic Chinese sentences. This is what makes the thesis so special.

The contribution that this thesis has made toward the development of the topic marker *ne* argument is the subject of the second innovation that this thesis has to provide. The majority of the current scholarly research on the subject of *ne* focuses on its functional usage within a given context, more especially as either an interrogative marker or as a sentence-final modal particle. Despite the fact that it performs the role of a subject marker, linguists have paid very little attention to the topic marker *ne*. This is especially true in terms of the temporal analysis of the marker. This thesis utilizes a corpus analysis to investigate the developmental path of the topic marker *ne*, simulates the process of grammaticalization of *ne*, and supports certain assumptions regarding *ne* put forward by the community of Chinese linguists. This methodology has the ability to partially overcome the limits in modeling the historical evolution of this function word and bring unique insights to the study of *ne*. In other words, it has the capacity to address the limitations in modeling the historical evolution of *ne*.

6.2.2 Limitations

In spite of the fact that the thesis goes through continual revision and improvement while it is being finished, with the goal of being both compelling and innovative, the research technique will always involve certain subjective and objective restrictions, including the following:

i. In the methodology section, the CCL corpus and the BCC corpus were chosen in the first place to be analyzed as primary sources of data. Because the corpus contains works from a wide variety of genres, a screening process that is overly simplistic might not be adequate to ensure that the corpus is well-balanced. As a result, for the sake of our research, we have decided to carry out a screening process that focuses on particularly targeting representative works. Between the events recorded in the Book of Changes and those recorded in the Spring and Autumn Annals, there was a gap in time equal to five centuries. The repercussions of wars and other forms of societal upheaval can be seen
in the lack of written evidence that has been handed down to later generations. This thesis is incapable of providing an answer to the problem that was presented earlier. It is only capable of minimizing the subsequent selection of the remaining materials in the corpus, with the goal of reducing the amount of time that elapses between steps. Due to the fact that I have a poor comprehension of archaic Chinese, the analysis of the corpus may be susceptible to inaccurate or erroneous phenomena. It is unclear why certain statements don't have more in-depth explanations and analyses, which raises issues about the existence of other types of phenomena. There are several ideas that require additional rethinking and polishing before they can be implemented.

ii. The examples in archaic Chinese are susceptible to manual analysis, which opens the door to the potential of inaccuracies that may have an effect on the final result. When conducting case analyses, extreme vigilance is required due to the presence of duplicate data inside the corpora. This is because there is a high probability of coming across redundant data while conducting case analyses. In addition, it is recommended that topic markers be omitted whenever possible, particularly when the instances include the terms but it is unclear whether or not the terms belong to the topic of the sentence. This will guarantee that the selected examples consist of sentences with topical structures that are clearly established and established clearly. In this particular situation, the final numerical values will be impacted, and there will be a subset of sentences that are thematically stressed but are not included in the tabulation. These sentences will be excluded from the analysis.

iii. The innovative facets that are covered in the work demand further examination, which should include both data and textual analysis. The variables that prevent topic markers in Chinese from evolving into specialized subject markers are one of the aspects that are under examination. Also under consideration are the dynamics at play between cliticization and topicalisation.

6.3 Suggestions for further studies
At this point in time, the majority of academic research concerning Chinese topic markers focus on classification studies and the functional analysis of various topic markers. The existing body of academic scholarship on topic markers in historical Chinese focuses, for the most part, on diachronic investigations that are restricted to a constrained range of chronological periods. In order to make progress in a particular area of research, it would be beneficial to see a rise in the number of corpus-based investigations that take into account comprehensive data sets spanning a considerable amount of time. Using the data that has been provided, one is able to design a more comprehensive model for replicating the historical development and changes that have occurred in Chinese subject markers.
In addition, it is of the utmost importance that the process of creating the corpus be improved upon in the work that comes after it. Both the CCL and the BCC are plagued by a number of issues, some of which include a lack of punctuation marks, duplicate entries in the corpus, and transcribing mistakes, amongst other things. The existence of these defects will have an adverse effect on the accuracy and rigor of the research because of the impact it will have. As a result, it is essential for future attempts to continue working toward the development of an excellent corpus.

The final and most important factor to take into account is the necessity of increasing the rigor of the topic's definition, with the end aim of producing a unified and authoritative definition in the near future. This is the most important factor to take into account. The construction of a thorough definition will have a considerable impact on the evaluative criteria utilized by linguists, which highlights the essential requirement for an authoritative and conclusive definition.
Reference


Li, M. (2017). Chinese Parsed Historical Corpus (ChiParHC)


Peng, J. (2011). *Lunyu Hua Ti Jiegou Yanjiu [A Study on the Topic Structure of the Analects]* (Doctoral dissertation, Huazhong University of Science and Technology). Retrieved from https://kns.cnki.net/kcms2/article/abstract?v=ZJxhFRRmSlijNkzjQ2G-r6ezjFU7XBql3GhJQtznDIWsL9dau6hPoCBM1x0vqQrwpekaQ9YA637KqJ4dTORb3uPeVgHz3yrNi_l8lFOZpYTdx9SpoG41ytkofn_yfwtC6sxGqQ6VSD5zzpVCu_CP8A==uniplatform=NZKPT1language=CHS


Tang, J. (2018). "Chao Ye Qian Zai" Hua Ti Ju Yanjiu [A Study on Topic Sentences in "Chao Ye Qian Zai"] (Master's thesis, Xinyang Normal University). Retrieved from https://kns.cnki.net/kcms2/article/abstract?v=ZJxhFRRmSIgdK4qrpkW2xNkl1yZNh_UH4eYwZXPPBrHBl5VGLvESR8xjFO7EenOEJjXCq5mAluXQIXbsFtWH28-PDuL9gBSujRBavoC2SstEKjnD2Ilt8S4svZZd8U4z8zgjDYDXfRvH9MNWeLV5gQ==uniplatform=m=NZKPTlanguage=CHS


Yu, Q. (2017). Ming-Qing Changpian Baihua Xiaoshuo Huayu Biaoji Yanjiu [Study on Discourse Markers in Ming-Qing Vernacular Novels] (Master's thesis, Yunnan Normal University). Retrieved from https://kns.cnki.net/kcms2/article/abstract?v=ZJxhFRRmSIj20XPS6Zr4xCbrqIV0hrJ9TxqOHd47IS e53BArWN-vV7xdG1DYsotPTsnRoz5sPf_foIZ9MhvQav1thN5-on_KMdx7eqG2ilciQ343ogZD_YqSI_L1Hqn62YP7ZewtRg0jnoBIIUCttA==uniplatform=NZKPT
language=CHS


