A Study of Chinese Internet Homophones from the Systemic Functional Perspective

by

Qian Tang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of East Asian Studies

University of Alberta

Abstract

This thesis explores the phonetic features, morphological classification, and discourse function of Chinese Internet homophones based on the Leiden Weibo Corpus (a corpus of Chinese micro-blogging). The analyses show that more than half of the Chinese Internet homophones in the data retain the same pronunciation as their base words, while the rest differ in initials, finals, or tones. Based on their morphological structure, Chinese Internet homophones are categorized into Chinese character homophones, alphabet homophones, numeric letter homophones, and blending homophones. An examination of the data shows that Chinese character homophones and alphabet homophones are used much more frequently than numeric letter homophones and blending homophones. This study also explores the discourse function of Chinese Internet homophones from the perspective of systemic functional grammar. Chinese Internet homophones are used to achieve grammatical cohesion (conjunction, reference, ellipsis, and substitution) and lexical cohesion (repetition, synonymy/antonymy, hyponymy/meronymy, and collocation) in micro-blogging. Finally, the motivations for the use of Chinese Internet homophones are discussed. This study contributes to our understanding of the dynamic development and functions of homophonic puns in Chinese Internet language.

Key words: Chinese Internet homophones, cohesion, phonetic features, morphological classification, mirco-blogging/*Weibo*

Table of Contents

List of Tables

Chapter 1 Introduction

As the fastest-growing communication medium, the Internet has exerted a great influence on language use. Internet language is a variant of social dialect. Its lexical system is tied to everyday talk. Lexicon is the most active and fast-changing part of Internet language (Li, 2007, p.8). The social factors of characterization of Chinese dialects and Chinese Government censorship facilitate the emergence of the distinctive features of Chinese Internet language (Liu, 2011; Inkster, 2010). This thesis focuses on the linguistic features and discourse function of Chinese Internet homophones. In this chapter, the notions of computer-mediated communication, Internet language and micro-blogging (*Weibo* in Chinese), and Chinese Internet homophones will be introduced respectively. This chapter will provide the background information for the thesis.

1.1 Computer-Mediated Communication

Computer-mediated communication (CMC) is defined as the "communication that takes place between human beings via the instrumentality of computers" (Herring, 1996). The Internet was put into use in 1969[1] to share and exchange data. Through more than 40 years of development, the Internet has become an indispensable communication medium for people of all walks of life.

From the 1970's to the early 1990's, the primary computer-mediated communication tool was e-mail, with most of its users from academia, science, and information technology. In the 1990's, the Internet experienced a period of rapid growth which resulted in a significant increase in the use of CMC among North American language educators (Thorne, 2008, p.2). E-mail remained

[1] It was originally developed by the American Advanced Research Projects Agency. "The first group of networked computers [is] called ARPANET" (Thorne, 2008, p.1).

the primary communication tool, but Internet Relay Chat and instant messaging tools became popular. Now, due to the development of communication technology and networks, the use of cell phones and tablets is becoming more and more popular in communication. Newer communication tools such as blogging and micro-blogging came into existence. Traditional computer-mediated communication has been extended to digital-mediated communication.

The development of network technology along with the increase of Internet users facilitates the use of Internet language. The following section will introduce Internet Language and a new communication tool *Weibo* in China.

1.2 Internet Language and *Weibo*

Internet language is a type of language used in computer-mediated communication. It has been widely used in BBS (bulletin board system), chat rooms, e-mail, on-line magazines and literature, blogs, micro-blogs, etc. It is a social dialect in cyber space influenced by the Internet and its users.

Yu (2001) argues that there are two types of Internet language utilized by Internet users. The first type includes Internet terminologies and proper nouns. For example, 调制解调器 (*tiao*2*zhi*4*jie*3*tiao*2*qi*4 'modem'), 宽带 (*kuan*1*dai*4 'broadband'), and 传输控制协议 (*chuan*2*shu*1*kong*4*zhi*4*xie*2*yi*4 'Transmission Control Protocol/Internet Protocol') belong to Internet terminologies; 网虫 (*wang*3*chong*2 'Internet geek'), 黑客 (*hei*1*ke*4 'hacker'), and 虚拟空间 (*xu*1*ni*3*kong*1*jian*1 'cyberspace') are Internet proper nouns. The second type of Internet

language is the everyday language used by users on the Internet (Yu, 2001, p.2). The interaction between Internet language and daily communication has been accelerated, due to the advancement of new technology and the increase of Internet users. This thesis focuses on the second type of Internet language. Among its linguistic features, the lexical features of Internet language make it most distinctive from other types of social dialects.

Chinese scholars have studied Internet language as a variety of language, typically featuring variability and individuation (Yan, 2005). Internet language shows uniqueness, simplicity, causality, creativity, and vulgarity (Li, 2004). Although the social features of Internet users (e.g. career, age, gender, education degree, etc.) are hidden, the style of Chinese Internet language reveals that the majority of Internet users are students and young middle class with middle or higher education. The communication between Internet users has formed a discourse community, though it is not very stable. The content and pattern of Chinese Internet language has the following features: irrationality, diversification, randomness, etc. (Xia, 2005; Liu & Qian, 2005; Chen, 2008). The Internet in China also adapts to the needs of grassroots movements, and Chinese Internet language has distinctive grassroots characteristics. It shows the creativeness of grassroots movements. In addition, their on-line expressions reveal their opposition toward authorities (Shi, 2009, p.2).

Crystal (2006) classified Internet language into six types based on different "Internet situations": the language of e-mail, the language of chat groups, the language of the virtual world, the language of the World Wide Web, the language of instant messaging, and the language of blogging. Since 2006, a new communication method has emerged known as micro-blogging.

Micro-blogging, a shortened broadcast medium, shares information to other subscribers in real time through web and Wireless Application Protocol (WAP). Micro-blogs "allow users to exchange small elements of content such as short sentences, individual images, or video links" (Kaplan & Haenlei, 2011). The earliest and most popular micro-blogging service is Twitter. In 2009, one of the largest four portals in China, Sina, founded the earliest Chinese micro-blogging service, Sina *Weibo*[2]. Following Sina *Weibo* there also appeared Sohu *Weibo*, 163 *Weibo*, Tencent *Weibo*, and many others.

According to China Internet Network Information Center (CINIC), as of the end of 2012 the total population of Internet users in China reached 564 million. Among the Internet users, there were 309 million micro-blogging users. 65% of the users used cell phone terminals to visit micro-blogs[3]. The majority of micro-blog users were from Sina *Weibo*[4].

Micro-blogging has become one of the most important ways of communication on the Internet, influencing people's lifestyles, how they transfer information and even public opinions, leading China to a new "micro era" (Zhu et. al, 2013, p.204). People from all walks of life, whether government officials or ordinary Internet users, can micro-blog information regarding politics and government. The Chinese public security authorities have even opened a "*Weibo* 110[5]" on the Internet where people can express "micro complaints". This reflects the powerful force of micro-blogging on public opinions.

[2] *Weibo* (微博, *wei*1*bo*2) refers to Chinese micro-blogging services.
[3] Source: http://companies.caixin.com/2013-01-16/100483154.html
[4] Source: http://www.cq.xinhuanet.com/2013-02/21/c_114749460.htm
[5] 110 is an emergency phone number in China.

Compared with traditional communication media, micro-blogging is more open, more convenient, and more interactive. First, its promptness ensures rapid dissemination of information and timely feedback. Second, the constraint of background context is relatively weak. The connection between preceding context and following context is poor. Third, there is extensive use of homophones, abbreviations, other symbols, word conversions, and special sentence patterns in micro-blogging communication (Xiang, 2012, p.13). As the extensive use of homophones stands out in the language used in micro-blogging, this thesis will focus on Chinese Internet homophones and their discourse function.

1.3 Chinese Internet Homophones

According to the *Longman Dictionary of Applied Linguistics*, homophones are defined as "words which sound alike but are written differently and often have different meanings" (Platt, Richards, & Weber, 1985, p.130). In this thesis, Chinese Internet homophone is defined as "the language phenomenon resulting from intentional uses of words" which are "identical or similar in sound for the purpose of achieving certain special effect" in Internet communication (Chen, 2011, p.iii). The relation between the two identical-sounding words is called homonymy. Homophones can be spelled the same (i.e. homographs and homonyms; e.g. *ground* 'floor' and *ground* 'past tense and past perfect tense of *grind*'); they can also be spelled differently (i.e. heterographs; e.g. *dye* 'a colored substance' and *die* 'cease life'). Homophones may also apply to phrases (e.g. the longer string 2*EZ* 'too easy'), which includes two or more lexical items. In this thesis, the study of homophones does not consider homophonic phrases.

Chinese homophones, *xieyin ci*, are considered to include the words that are identical or similar in sound. The use of homophones always involves two words linked by phonetic similarity. In Chinese, the original word is called the base word (Chen, 2011), and the derived word having the same or close pronunciation as the base word is usually called homophone. The base word is the sound source of the derived word. A word with the same sound as the base word is named a narrow-homophone in this paper and a word with similar sound to the base word is named a broad-homophone. For example, the word 脖友 (*bo*2*you*3, 脖 'neck', 友 'friend', 脖友 'friends made through blogging') is from the base word 博友 (*bo*2*you*3, 博 'blog', 友 'friend', 博友 'friends made through blogging'). 脖友 is a narrow-homophone of 博友, as they have the same pronunciation. 脖友 literally means 'neck friend,' but it is used to present 'blogging friends' with a humorous effect created by the original meaning of 脖 (*bo*2 'neck') that has the same sound as 博 (*bo2* 'blog'). In Chinese, the meaning of the homophonic word is related to the base word, as the example of 博友 and 脖友 shows. A homophonic word can be used in communication with or without the base word.

Chinese is full of homophones, which can be found in literary works and folk adages. Homonymy is an important means to improve the effect of their expression. For example, in Cao Xueqin's *Dream of the Red Chamber*[6], a character's name 贾雨村 (*Jia*3 *yu*3*cun*1, 贾 'a Chinese family name', 雨 'rain', 村 'village', 贾雨村 'a person's name'; 贾雨村 is a narrow-homophone of 假语村 [言][7]*jia*3*yu*3*cun*1(*yan*2); 假 'false', 语 'language', 村 'village', 言 'speech', 假语村 [言] 'rumor') implies that this person is not honest (Ning, 1975). Another character's name, 甄

[6] A masterpiece of Chinese literature, its first 80 sections are considered written by Mr. Cao Xueqin.
[7] Only the pronunciation of the first three characters were used as the base for the creation of the homophone.

士隐 (*Zhen*1*shi*4*yin*3, 甄 'a Chinese family name', 士 'bachelor ', 隐 'hide', 甄士隐 'a person's name'; 甄士隐 is a narrow-homophone of 真事隐 *zhen*1*shi*4*yin*3; 真 'true', 事 'thing', 隐 'hide', 真事隐 'true things are hidden') implies that he may be hiding some truth. In folk culture, there are a large number of homophones known as "lucky words" (Fong 2000). For instance, in the Chinese (lunar) New Year there must be fish on the table. This is because that 鱼 (*yu*2 'fish') has the same sound as 余 (*yu*2 'surplus'), expressing a wish to have more than needed in the New Year.

Normally, Chinese homophony is introduced as a rhetorical device, using a word that has the same or similar pronunciation as another word (i.e. base word) to express the meaning of the base word and create certain effect (e.g. humour or satire). Chinese homophones are dynamic uses of words that sound alike (Chen, 2011). When the word is in actual use, it can become a homophonic pun of the base word. For example, 鱼 (*yu*2 'fish') and 余 (*yu*2 'surplus') are two semantically unrelated words. Only under certain circumstances, such as at a Chinese New Year table, 鱼 'fish' connotes the meaning of 余 'surplus'.

As a result of the popularity of the Internet, homophones have been widely used in digital-mediated communication through all types of tools. Chinese Internet homophones are homophones used in the context of the Internet in China. The Internet seems to be a breeding ground for Chinese Internet homophones. There is a reflexive relationship between Chinese Internet homophones and language used in daily communication. For example, the Chinese Internet homophone 神马 (*shen*2*ma*3, 神 'magic', 马 'horse', 神马 'what'; a homophone of 什

么 *shen*3*me* ‘what’) originates from an on-line novel and later became popular in daily communication. Chinese Internet homophones can also come from daily communication, such as, 闹太套(*nao*4*tai*4*tao*4, 闹‘noisy’, 太 ‘too’, 套 ‘cover’, 闹太套 ‘a homophone of English *not at all*’). This homophone came from a Chinese singer Huang Xiaoming. When he was singing the song *One Word One Dream*, his pronunciation of *not at all* was very close to 闹太套. People made fun of him, calling him 闹太套. Then the word 闹太套 became popular on the Internet. The lexical system of Internet homophones is not very stable. Some of them might only be used on the Internet for a short period of time and then disappear, while others could be long lasting and gradually become part of the common lexicon.

Based on the background of computer-mediated communication and the popularity of a new communication tool, micro-blogging/*Weibo*, Chinese Internet homophones reveal certain new features, such as the new morphological structures and phonetic changes. In addition to Chinese characters, alphabets and numeric letters can also be used to create homophones. The base words can be from other languages, such as English and Japanese. In this thesis, I will explore the use of Chinese Internet homophones and their base words in the data and phonetic and morphological classification of Chinese Internet homophones (Chapter 3), and their discourse function (Chapter 4). The next chapter will provide the theoretical and methodological background of the study.

Chapter 2 Theoretical and Methodological Background of the Study

The past two decades have seen rapid development of systemic functional grammar (SFG) in China and it has exerted great influence on Chinese linguistics (Zhang, D. J., 2006). However, there is not much research on homophones in Chinese computer-mediated communication (CMC) from the SFG perspective. In this study, I will use systemic functional grammar to examine Chinese Internet homophones in a Chinese micro-blogging corpus. A brief discussion of the theory of cohesion in SFG and the data used for this thesis will provide the background for the analyses in the following two chapters. In Section 2.1, I will briefly introduce systemic functional grammar. In Section 2.2, I will present the methodology of this study. In Section 2.3, an overview of past studies of Chinese homophones will be provided.

2.1 Theoretical Background

The analysis of discourse function of Chinese Internet homophones in this thesis is based on Hu et al. (2005)'s framework of cohesion developed from Halliday's systemic functional grammar. In this section, I will introduce the theory and development of SFG and the classification of cohesion.

2.1.1 Systemic Functional Grammar

SFG is a theory of grammar, established by Michael Alexander Kirkwood Halliday (1985). Guided by semantics, he proposed that language be considered as a regular resource that describes system rather than structure. Systemic theory is "a theory of meaning as choice, by

which a language… is interpreted as networks of interlocking options" (Halliday, 1994, p.xiv). Halliday describes SFG as a way in which "a language is interpreted as a system of meanings, accompanied by forms through which the meaning can be realized" (1994, p.xiv). SFG is an analysis of synthesis grammar based on the study of semantics through systems of lexical and grammatical choices that serve functions within social environment (Halliday, 1985).

Halliday (1968) proposed that in all languages, the system has four functions: experiential, logical, speech, and interpersonal. Halliday (1972) later revised the four functions: experiential and logical functions were combined into ideational function (i.e. to understand the environment); speech function was named as textual function; while the interpersonal function remained the same (i.e. to act on the other in it). These three functions, known as the general functions of language, were named metafunctions. Cohesion is included in textual function. Halliday & Hasan (1976) stated that the fundamental difference of text and non-text is whether it has texture, and whether this texture is formed by cohesive relation. Therefore, cohesion has become an important area of research in discourse studies.

SFG is useful in explaining the functions of language. It enhances our understanding of the functions of language in different types of discourse, including discourse in computer-mediated communication. Halliday (1985 &1994) pointed out that the purpose of SFG is to provide a theoretical framework for the analysis of both spoken and written discourse.

Halliday's work is the most important development of the 'London School', which was founded by J. R. Firth" (Butler, 1985, p. 1).[8] Halliday's experience studying in China with two well-known Chinese linguists, Luo Changpei and Wang Li, also influenced the formation of his

[8] J. R. Firth is regarded as Halliday's "main inspiration". Firth studied language from a formal linguistics perspective, specializing in contextual theories of meaning and prosodic analysis.

linguistic theory. Halliday's SFG "is in line with the tradition of Chinese linguistics, which is 'meaning-oriented'" (Huang, 2002, p. 281 & 285). SFG has become one of the most influential theories in Chinese linguistics (Zhang, D. J., 2006). Since 1980s, Hu started to introduce SFG in China, and since then SFG has been the most popular and well-received linguistic theory in China (Huang, 2011, p. 45). In China, SFG has been used to study different types of discourse, such as spoken discourse (e.g. Zhang & Fang, 1996), written literature (e.g. Shi & Sun, 2001), advertisement (e.g. Tang, 2005) and second language teaching (e.g. McDonald, 1992). Though SFG has been used to examine the Chinese media discourse, it has not been widely used in studying Chinese Internet language. The advantage of using Halliday's theory in this thesis is that it allows for the analysis of cohesion function of Chinese Internet homophones, and the exploration of the semantic, social, and cultural factors underlying the use of homophones. SFG will provide a new perspective to the study of Chinese Internet homophones.

2.1.2 Cohesion Theory

SFG provides the theoretical background for the study of cohesion function of Chinese homophones in computer-mediated communication. Through years of development, the theory of cohesion has been widely used to analyze written discourse. This section will discuss the concept of cohesion, the development of the cohesion theory, and its classification.

2.1.2.1 Definition of Cohesion

Cohesion refers to "relations of meaning that exist within the text, and that define it as a text" (Halliday & Hasan, 1976, p.4). It is "part of the system of a language" (Halliday & Hasan, 1976, p.5). The construction of cohesion is stated as follows:

> Cohesion occurs where the interpretation of some elements in the discourse is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby at least potentially integrated into text.
>
> (Halliday & Hasan, 1976, p. 4)

Halliday & Hasan (1976) proposed cohesion functions of English discourse in detail while pointing out the significance of cohesion in language systems in general. It is generally agreed that *Cohesion in English* (Halliday & Hasan, 1976) marked the establishment of cohesion theory and serves as the classical theory for cohesion studies. Before Halliday & Hasan (1976), studies of cohesion in English were conducted by scholars such as Jakobson (1960), Halliday (1964), and Hasan (1968). These studies provide the foundation of cohesion theory.

After the publication of *Cohesion in English* (1976), Halliday and Hasan, along with other scholars, further developed the cohesion theory. For instance, in *Language, Context, and Text: Aspect of Language in a Social-semiotic Perspective* (1985), Halliday and Hasan stated that

semantic relations are the basis of cohesion, and that there is a cohesive tie which connects different textual units of a message. They pointed out that cohesion specifically refers to co-referentiality, co-classification, and co-extension in internal components of discourse.
The following examples will demonstrate Halliday and Hasan (1976 & 1985)'s classification and concept of cohesion.

Halliday and Hasan (1976) classified cohesion into reference, substitution, ellipsis, conjunction, and lexical cohesion. Reference, substitution, ellipse, and conjunction are grammatical cohesions. Lexical cohesion includes reiteration and collocation.

Reference occurs when a participant or circumstantial element serves as another element's reference point to make the information signalled for retrieval.
For example,

[2.1]
"I had a little nut tree
Nothing would it bear
But a sliver nutmeg
And a golden pear" (Halliday & Hasan, 1985, p. 73)

In this excerpt, *it* in the second line refers to *a little nut tree* in the first line. The information of *it* is retrieved from the proceeding text *little nut tree* and the cohesion lies in the continuity of

reference. “This relationship of situational identity of reference is known as co-referentiality” (Halliday & Hasan, 1985, p. 73).

Substitution occurs when one item is used to replace another.

For example,

[2.2]

“I play the cello. My husband does, too.” (Halliday & Hasan, 1985, p. 73)

In this excerpt, *does* in the second sentence substitutes *play the cello*, serving as a cohesive device. *Does* and *play the cello* belong to an identical class, and this kind of relationship is also referred as co-classification (Halliday & Hasan, 1985).

Ellipsis is “something left unsaid,” but implies “understood nevertheless,” where there is there is a corresponding completed item which presents the unsaid information (Halliday & Hasan,1976, p.142). For example:

[2.3]

“Joan brought some carnations, and Catherine some sweet peas.” (Halliday & Hasan, 1976, p.143)

The two clauses in this text are in a parallel structure, and the second one can be interpreted as *Catherine bought some sweet peas,* while in the text, *bought* is omitted. *Bought* serves as the source of an elliptical item and provides a connection within the text.

Conjunction covers the use of adjunct-like elements to express certain meanings that presuppose the presence of other components in the discourse. For example:

[2.4]

"They fought a battle. Afterwards, it snowed." (Halliday & Hassan, 1976, p. 228)

Afterwards is acting as a cohesive agent to connect the two sentences by time sequence. Lexical cohesion was treated as a parallel cohesion type as a subcategory of grammatical cohesion. Cohesion is "the cohesive effect achieved by the selection of vocabulary" (Halliday & Hasan, 1976, p. 274).

Lexical cohesion includes reiteration and collocation. They define reiteration as follows:

> …[It] involves the repetition of a lexical item, at one end of the scale; the use of a general word to refer back to a lexical item, at the other end of the scale; and a number of things in between- the use of a synonym, near synonym, or superordinate.
>
> (Halliday & Hassan, 1976, p. 278)

Below is an example of reiteration:

[2.5]

"Accordingly…I took leave, and turned to the ascent of the peak. The climb is perfect easy…" (Halliday & Hassan, 1976, p. 278)

In this example, the use of two synonyms, *ascent* and *climb* falls within the lexical cohesion. The use of two synonyms repeated the meaning "uprise" and supplemented to the preceding content to link the two sentences.

Collocation refers to those lexical items with a particular association (i.e. having co-occurrence tendency) and are used together in the text.

Here is an example of collocation:

[2.6]

The king was in his counting-house, counting out his money,
The queen was in the parlour, eating bread and honey,
The maid was in the garden, hanging out the clothes.
Along came a blackbird and pecked off her nose. (Halliday & Hassan, 1976, p. 292)

Each pair of lexical items is in a recognizable lexico-semantic relation: *king* and *queen*; *parlour* and *garden*; *dish* and *eat*. Their co-occurrence serves as collocational cohesion. Also, in this

example, *King* and *queen* are exclusive of each other but under the same superordinate *monarch.* This type of relationship is named as co-extension (Halliday & Hasan, 1985).

In a later publication, Halliday (1985) re-categorized lexical cohesion into repetition, synonymy, and collocation. [9] Halliday & Matthiessen (2004) proposed the classification of cohesion into conjunction, reference, substitution & ellipsis, and lexical cohesion. Lexical cohesion includes elaboration, extension, and collocation. Elaboration comprises of repetition, synonymy, and hyponymy.

In sum, cohesion is regarded as a type of semantic relations (Halliday & Hasan, 1976; Halliday & Hasan, 1985; Nunan, 1993; & Berry, 1994). Cohesion indicates semantic relation between one discourse element and another, creating semantic consistency. Some researchers also consider cohesion as a lexical or grammatical feature of discourse (e.g., Hoey, 1991; Huang, 1988 & 2001). It is an approach to connecting different textual units within discourse. Halliday & Hasan (1976)'s reference, substitution, ellipsis, conjunction, and lexical cohesion are the specific lexico-syntactic resources to achieve semantic cohesive relation. These textual items that create cohesion in discourse are called cohesive devices.

2.1.2.2 The development of cohesion theory in China

The study of cohesion theory in China began in the 1980s. During this period, many scholars published papers and monographs introducing cohesion from different perspectives (e.g. Wang, 1981; Xu, 1982; Hu, 1984; etc.).For example, Hu, Zhu, and Zhang (1989) published their book *A*

[9] There is no change to the classification in Halliday (1994).

Survey of Systemic Functional Grammar to introduce Halliday and Hasan's cohesion theory in China. In this book, they used examples from Chinese to explain the theory.

Since the 1990s, cohesion theory has been further developed. Hu (1993) discussed the connection between intonation and cohesion. Hu (1994) further developed cohesion theory by expanding the scope of structural cohesion. He argued that structural cohesion is the semantic linkage that connects a word, phrase or clause whose meaning doesn't clearly appear in a structure. The semantic meaning of these words, phrases or clauses will be retrieved by comparing the syntactical structure with the presupposed structure (Hu, 1994, p.69). Zhu (1995) pointed out that understanding the context and the reader's background knowledge contribute to discourse cohesion, especially for discourse without explicit cohesive devices. However, in some other studies, these factors tend to be ignored. Most studies only focus on cohesive devices. Zhang (1999, 2000, & 2001) put forward a more detailed theoretical framework of cohesion. Zhang (2001) argued that cohesion should be defined as "the textual meaning that organizes meaning realized by formal features within language", and also the relation "that connects the text with the context of situation" (p. 23). For text-internal cohesion, Zhang (2001) proposed 11 general cohesive principles. Aside from the development from Halliday & Hasan (1976), the principles include mental distance (the distance of cohesion) and natural language's order. His contributions to cohesion theory also include introducing the cross-category cohesive mechanism, exophora cohesion, recessive cohesion, and other types of cohesion.

2.1.2.3 Hu et al. (2005)'s Classification of Cohesion

Based on Chinese discourse, Hu et al. (2005) classified cohesion into grammatical cohesion and lexical cohesion. Grammatical cohesion includes conjunction, reference, ellipsis, and substitution. Lexical cohesion includes repetition, synonymy/antonymy, hyponymy/meronymy, and collocation. Hu et al. (2005)'s definition of cohesion along with conjunction, reference, ellipsis, substitution, and lexical cohesion are from Halliday & Hasan (1976). However, Hu et al. (2005) treated grammatical cohesion and lexical cohesion as parallel subcategories of cohesion, which is different from Halliday & Hasan (1976), Halliday (1985), and Halliday & Matthiessen (2004). This thesis will adopt Hu et al. (2005)'s classification.

The reasons for using Hu et al. (2005)'s classification of cohesion are three-fold. First, Hu et al. (2005)'s categorization keeps the traditional classification of cohesion based on Halliday & Hasan (1976): grammatical cohesion and lexical cohesion. However, they put grammatical and lexical cohesion at the same level. Halliday & Hasan (1976), Halliday (1985), and Halliday & Matthiessen (2004) consider lexical cohesion as a subcategory of grammatical cohesion. In other words, grammatical and lexical cohesion are at different levels. However, there is no clear evidence that they belong to different levels. Second, based on the principal of economy, the less the better. Hu et al. (2005)'s classification is the most economical, while it also includes all types of cohesion in discourse. Third, Hu et al. (2005)'s classification of cohesion is based on and is adapted to Chinese discourse. It is the most suitable for the analysis of discourse function of Chinese Internet homophones. Thus, Hu et al. (2005)'s classification is adopted in this study.

2.2 Methodological Background

This section will introduce the data used in this thesis. The data are from the Leiden Weibo Corpus (LWC, http://lwc.daanvanesch.nl/). The corpus contains messages from a Chinese micro-blogging service Sina *Weibo*. The LWC is created by Daan van Esch. It contains 5.1 million messages from Sina *Weibo* that were posted between January 8th, 2012 and January 30th, 2012. All messages in this database were processed using the Penn Chinese Treebank standard for Chinese word segmentation and part-of-speech tagging from the Stanford Natural Language Processing Group. These messages come from real communication rather than being constructed. Each message in the LWC has a unique message ID. Each message contains the user's screen name, location, and gender, provided by the users of Sina *Weibo*. At the home page http://lwc.daanvanesch.nl/, messages can be searched by lexical items. From each message, the searching result also shows the number of occurrences, frequency, and frequency rank of a specific target word. Because the LWC relies heavily on computer natural-language processing, the best ones can reach approximately 95% accuracy rates (Daan van Esch, 2012). Considering the large quantity of the data, "these accuracy rates are similar to or better than the inter-annotator accuracy rate for human annotators" (Daan van Esch, 2012). Because of this, the LWC is an ideal corpus to study Chinese *Weibo* language.

In order to examine a large number of messages within a reasonable time, this thesis uses a dataset of 18,791,530 words taken from the Leiden Weibo Corpus based on a random sampling method. As homophonic pans (homophones and their base words) are hard to identify in the corpus without reading each message from the dataset, an initial word list of Chinese internet homophones was constructed based on the *Green Paper on Language Situation in China*

(GPLSC, 2006-2013), Chinese Internet language dictionaries, *Sina Weibo Niandu Pandian* (*Sina Weibo Annual Report*, 2011-2012). The GPLSC has been published annually by the Ministry of Education and National Language Committee of China since 2006. GPLSCs are official reports on contemporary Chinese language. These reports include statistical data based on their nation-wide database, lists of new words with their meanings, and commonly-used words from the Internet. These sources list all new Internet words or/and commonly-used Internet words each year. The Chinese Internet language dictionaries that I have consulted for this study include *New Internet Dictionary* and *Chinese Internet Language Dictionary*. I manually identified Internet homophones from all the word lists from the aforementioned sources. The GPLSCs provide reliable information on Chinese Internet language. The dictionaries and the *Sina Weibo Annual Reports* provide information from non-governmental perspectives. Thus, arguably the initial list constructed from these different sources is a relatively accurate representation of contemporary Chinese Internet homophones. The initial list of 150 different homophones was imported to CasualConc 1.9.7[10] to count each homophone's token frequency from the dataset that was taken from Leiden Weibo Corpus. Ranked by token frequency, a final word list that contains 117 Chinese *Weibo* homophones based on the dataset was built[11].

In the analysis of phonetic features of Chinese Internet homophones, I used IPA to notate the Chinese words (with the exception of Section 3.1, Chinese words in other parts of this thesis are notated with Chinese Pinyin[12]). In the analysis of the morphological classification of Chinese Internet homophones, an overview of each type of homophone with their frequency and

[10] CasualConc 1.9.7 is a concordance software program which can generate concordance lines, word clusters, collocation analysis, and word counts. It works on the English language, UTF-8 text, etc. Details can be found at http://casualconc.blogspot.ca/.
[11] There are 33 Chinese Internet homophones from the initial word list that didn't appear in the dataset.
[12] Chinese Pinyin refers to the Romanized phonetic system which transcribes the pronunciation of Chinese characters using Latin script. It was officially published by the Chinese government. This system is widely used for teaching Mandarin Chinese, as a method of inputting Chinese characters, and so on.

percentage is presented. Under each category, the most frequently used homophones from the final word list are provided with Chinese Pinyin notation, the corresponding base words, and the percentage of each homophone's occurrences in the total occurrences of the homophones and its base word(s).

The excerpts containing homophones in Chapter 4 are all from the final word list, and each excerpt comes with a unique message ID which can be looked up at the LWC (http://lwc.daanvanesch.nl/). In the data, some messages contain emoticons (i.e. pictorial representations of facial and emotional expressions) and hyperlinks. I chose to only look at the text of each message. This thesis keeps the original version of the messages which include mistakes made by micro-blog users in words, sentences, and punctuation. For each excerpt, a gross translation is provided: the first line shows the original data in Chinese characters; the second line is the Pinyin notation; the third line shows the English meaning of each word; and the fourth line provides the English translation for the entire line.

2.3 Previous Studies of Chinese Homophones

Since the introduction of systemic functional grammar in China, the theory of cohesion has been largely applied to study different types of Chinese discourse. However, the theory of cohesion is not widely used to study Chinese homophones in computer-mediated communication (CMC). The study of Chinese Internet homophones will provide new insight into cohesion in Chinese CMC. This section will offer a brief overview of past studies of Chinese homophones.

The discourse function of Chinese homophones is examined in different types of discourse. For example, Zheng (2007) explored discourse function and register features of cell phone text messages. Zheng (2007) pointed out that homophones can be agents of cohesion in text messages and they create intertextuality. Intertextuality refers to the connection and interaction between two or more texts. In Zheng (2007), it refers to the connection and interaction of a chain of cell phone text messages. By exploring the textual mechanism of Chinese homophones in the discourse of text messaging, Li (2009) studied the phonetic, morphological, discourse, and semantic features of Chinese homophones. She also explored lexical cohesion and grammatical cohesion created by homophones. She stated that there are mainly two types of grammatical cohesion used in text messages: ellipsis and reference. Wang (2005) studied means of cohesion based on grammar and sound, and pointed out that the correlation of semantic meanings in discourse is essential. The use of homophones in allegorical sayings,[13] poems, and numeric letter messages can achieve textual cohesion. Chen (2007) examined Chinese cross talks[14] from the data *Xiangsheng mingzuo yu xingshang* (*Masterpiece of cross talks and appreciation*). He argued that the use of homophones can realize cohesion in cross talks and achieve humorous effect. Wang (2010) studied cohesion of Chinese antithetical couplets (i.e., two parts of a pair of lines of poetry written on a pair of scrolls) from the perspective of lexis and sound. Wang (2010) argued that homophony can be a device to realize textual cohesion. By exploring the phonetic features of homophonic puns, Bo (2007) analyzed cohesion function of homophonic puns in Chinese classical literature. Bo (2007) put forward that the use of homophonic puns depends on the user's understanding of the context and the purpose of expression. Zhang (2007) studied the lexical and

[13] Allegorical sayings, i.e. *xiehouyu* in Chinese, which in the *Oxford Language Dictionaries Online* (http://www.oxfordlanguagedictionaries.com) is defined as "a two-part allegorical saying, of which the first part, always stated, is descriptive, while the second part, sometimes unstated, clinches the point".

[14] Chinese cross talks, i.e., comic dialogues, usually involving two persons.

grammatical features of new Internet words. He pointed out that new Internet homophones can be created by borrowing phonetic sounds from English, Chinese dialects, Chinese characters, numeric letters, and Chinese acronyms.

In addition, Chinese homophones have been studied from other perspectives, such as culture (e.g. Zhao, 1987 & Fong, 2000), rhetoric (e.g. Zhang, 2005; Lu & Li, 2003), and psychology (e.g. Siok & Fletcher, 2001). From a culture perspective, Zhao (1987) examined the relation of homophones and Chinese culture, and pointed out that Chinese homophones are related to Han peoples' cultural tradition and the ethnic group's psychology.[15] Zhang (2005) explored Internet language from a perspective of rhetoric. He indicated that homophony, a type of rhetoric of the Internet language, can be more expressive for the purpose of realizing humorous and pleasant effects. Lu & Li (2013) explored the rhetorical effect of Chinese homophonic puns by means of semantic field theory, and explicated the factors that affect rhetorical effects. From a psychology perspective, Siok & Fletcher (2001) studied phonological awareness and visual-orthographic skills in Chinese reading comprehension. They found that among their students, the success of reading for students from Grade 2, 3, and 5 is affected by the ability to distinguish homophonic characters.

Through the overview of the past studies of Chinese homophones, we can see that homophones are widely used in all types of discourse to achieve textual cohesion. Homophony is common in cell phone text messages, allegorical saying, poems, cross talks, antithetical couplets, written

[15] Han is the majority ethnic group in mainland China. Similarly, Fong (2000) explored Chinese New Year "luck talk" that consists of speech acts related to luck, and discussed it with the principle of Yin and Yang and the theory of relative relativism. She stated that New Year food and the use of "luck talk" to counteract bad luck are homophones of the lucky words in Chinese, symbolizing luckiness, abundance, wealth, etc.

literature, etc. However, these studies did not provide clear classification of cohesion used in their analyses. For example, there is no clear statement about the classification of cohesion in Li (2009). Wang (2010) treated homophony as a type of cohesion. Chen (2007) studied homophones under the category of structural cohesion. However, their analyses of cohesion in discourse do not seem to be thorough enough. In addition, most of the papers did not indicate how the data were collected (e.g. Li, 2009), which makes their analyses less convincing. Thus, in this study, I use the data from a large established corpus based on a Chinese micro-blogging service. I will systematically explore the discourse function of Chinese Internet homophones in this corpus, by adopting Hu et al. (2005)'s classification of cohesion.

In the next chapter, I will explore the use of Chinese Internet homophones and their base word (Section 3.1), the classification of Chinese Internet homophones based on their phonetic features (Section 3.2), and the morphological structures (Section 3.3). Chapter 4 discusses the discourse function of Chinese Internet homophones.

Chapter 3 Classification of Chinese Internet Homophones

The use of homophones is a prominent feature of Chinese Internet discourse. When a homophone is formed, its pronunciation and its morphological structure may differ from its base word. This chapter will explore the classifications of Chinese Internet homophones based on their phonetic and morphological features. In Section 3.1, an overview of the use of Chinese Internet homophones and their base words in the data will be provided. In Section 3.2, I will examine the classification of Chinese homophones based on their phonetic features. The phonetic features of narrow-homophones and broad-homophones will also be explored. In Section 3.3, I will investigate the morphological classification of Chinese Internet homophones. This classification includes Chinese character homophones, alphabet homophones, numeric letter homophones, and blending homophones. In the second section of this chapter, the details of the phonetic features of Chinese Internet homophones will be discussed. These homophones are presented with IPA phonetic notation, while in other parts of this chapter, they are presented with Chinese Pinyin.

3.1 An Overview of the Use of Chinese Internet Homophones and Their Base Words in the Data

As mentioned in Chapter 1, Chinese homophones are always derived from their base words, and they are linked by their phonetic similarities. When faced with the two kinds of written forms, i.e. the homophones and the base words, some Internet users choose to use homophones, while others still use the original forms. The following table provides an overview of the use of the top 10 most frequently used homophones and their base words in the randomly sampled dataset.

Table 3.1.1 Overview of the Top 10 Most Frequently Used Homophones and Their Base Words in the Data

Homophone (H)	**Homopho-nic meaning**	**Occur-rences of H**	**H in the total occurrences (%)[16]**	**Base word (B)**	**Occur-rences of B**	**B in the total occur-rences (%)**
SB *sha2bi1*	idiot, asshole	17, 469	92.46%	傻逼 *sha2bi1* 'idiot /asshole'	1, 424	7.54%
神马 *shen3ma3* 'magic horse'	what	9, 139	11.97%	什么 *shen3me1* 'what'	67, 219	88.03%
围脖 *wei2bo2* 'wrap neck'	mirco-blog	5, 152	30.98%	微博 *wei2bo2* 'micro blog'	11, 480	69.02%
尼玛 *ni2ma3* 'priestess agate'	fuck off	4, 534	99.85%	你妈 *ni3ma1* 'your mom'	7	0.15%
粉丝 *fen3si1* 'vermicelli'	fans	4, 495	92.30%	fans 'enthusiastic followers of a sport, celebrity, etc.'	375	7.70%
RP *ren2pin3* 'personality'	personality, generalized for luck	3, 706	67.60%	人品 *ren2pin3* 'personality'	1,776	32.40%
88 *ba1ba1* 'bye-bye'	bye-bye	3, 334	65.17%	拜拜 *bai4bai4* 'bye-bye'	1,782	34.83%
表 *biao3* 'watch'	don't	3, 081	91.48%	不要 *bu2yao4* 'not want'	287	8.52%

[16] The percentage is calculated based on the formula: [occurrences of the homophone/ (occurrences of the homophone + occurrences of its base word(s))]*100%.

肿么 *zhong*3*me*1 ‘swelling/swell *me* (particle)	how	2, 837	7.14%	怎么 *zen*3*me1* ‘how’	36, 874	92.86%
囧 *jiong*3 ‘expressing the look of gloom or embarrass’	glooming, embarrass-ing	1, 716	100%	囧(ancient Chinese) *jiong*3 ‘bright’	0	0%

As is shown in the Table 3.1.1, among the top 10 most frequently used homophones, 7 of them are used more frequently than their base words in the data. For example, the Chinese Internet users never use the base word 囧 (ancient Chinese) *jiong*3 ‘bright’, but always use its homophone 囧 *jiong*3 ‘glooming or embarrassing’. The homophones 尼玛 *ni*2*ma*3 ‘fuck off’, SB *sha*2*bi*1 ‘idiot, asshole’, 粉丝 *fen*3*si*1 ‘fans’, and 表 *biao*3 ‘don’t’ are also used much more frequently (99.85%, 92.46%, 92.30%, and 91.48%) than their base words. However, 肿么 *zhong*3*me*1 ‘how’, 神马 *shen*2*ma*3 ‘what’, and 围脖 *wei*2*bo*2 ‘micro-blog’ seem to be used less frequently (7.14%, 11.97%, and 30.98%) than their base words. Table 3.1.1 shows that among the top 10 most frequently used homophones, the homophones are generally used rather frequently compared to their base words. However, Chinese Internet homophones are not very stable in the Chinese lexical system. Some of the Internet homophones may replace the base words and come into the Chinese common lexicon, while others may disappear.

The following sections will explore the phonetic features and morphological features of Chinese Internet homophones.

3.2 Phonetic Features of Different Types of Chinese Internet Homophones

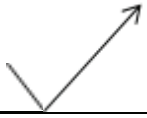
Chinese language has a writing system which is character-based with each graph corresponding to one syllable. In Old Chinese, the majority lexical items are monosyllabic (i.e., each word is one syllable in length). Through the development of disyllabification from Old Chinese to Modern Chinese, there has been an increase of disyllabic words or multi-syllabic words in the lexicon system (Pi, 1992, p. 56). Mandarin Chinese syllables are composed of initials, finals, and tones. [17] In the Mandarin phonological system, there are 21 initials, 39 finals, and 5 tones (4 basic and 1 neutral).

Given the small number of initials and finals, the Mandarin phonological rules, and the convention of spelling,[18] there can be only 398 types of initial-final combinations, and

[17] In Mandarin syllable structure, syllables are divided into initials, finals and tones. Initials are the optional consonants that can occupy the first position (Sun, 2006, p.35). The other part of the syllable is called a final. For example, 他 *ta*^{55} [ta^{55}] 'he' is constituted by the initial *t* [t], the final *a* [a] and a flat tone (the Mandarin tones will be explained at Table 3.1.1); 灯 *deng*^{55} [təŋ^{55}] 'light' has the initial *d* [t], the final *eng* [əŋ] and a flat tone. All Mandarin initials are consonants, but the consonant *ng* cannot function as an initial. Essentially, a Mandarin final consists of vowel(s) and optional nasals (please see the graph of Mandarin syllabus structure in 3.2.3.1).

Chinese tones change the pitch of syllables and make them lexically contrastive. The 4 basic tones in Mandarin Chinese are the high-level tone (the first tone), the high-rising tone (the second tone), the low-rising tone (the third tone) and the falling tone (the fourth tone). There is also a short and weak neutral tone, in addition to the 4 basic tones. As the neutral tone is short and weak and the pitch level is decided by the previous syllable, it is not described in this paper. The scale of five pitch levels of Mandrin tones is set forth by Yuen Ren Chao (1930). The pitch values of the 4 basic tones in Mandarin are 55, 35, 214 and 51 respectively as shown in table 3.2.1 with examples.

Table 3.2.1 The Four Basic Tones in the Scale of Five Pitch Levels

Pitch	1st Tone High-level tone 55/HH	2nd Tone High-rising tone 35/MH	3rd Tone Low-rising tone 214/LH	4th Tone Falling tone 51/HL
5/ High 4/High 3/Mid 2/Low 1/Low				
Example	妈 [ma ^{55}] 'mather'	麻 [ma ^{35}] 'hemp'	马 [ma ^{214}] 'horse'	骂 [ma ^{51}] 'scold'

[18] In Mandarin Chinese pinyin spelling rules, some initials and finals cannot be combined. For example, the initial *f* [f] cannot be spelled with *i* [i] or any final that begins with *i* [i]. Some initials and finals can be spelled together according to the conventional

approximately 1,200 distinct tonal syllable types (Suen, 1979). However, according to *Modern Chinese Frequency Dictionary* (1986), there are 4,574 characters in common usage. The limited number of syllables in Mandarin is conducive to the generation of homophones. The pronunciation of Chinese homophones is drawn from Mandarin, Chinese dialects, and occasionally other languages. The next section will offer a classification of Chinese homophones based on their phonetic features.

3.2.1 Types of Chinese Internet Homophones Based on Their Phonetic Features

Through homophonization, Chinese homophones may have the same or similar pronunciation as their base words. Based on Chinese linguistic traditions, Chinese homophones are classified into two types: homophones with the same sound as their base words and homophones with a similar sound to their base words (Chen, 2011). In this chapter, homophones with the same pronunciation as their base words are named narrow-homophones, whereas homophones with a similar pronunciation to their base words are called broad-homophones. Within the category of narrow-homophones, there are homographic homophones (homophones with the same spelling as their base words) and heterographic homophones (homophones that have a different spelling than their base words). All broad-homophones are heterographic homophones.

In the dataset, there are 117 homophones. Among these homophones, 64 are narrow-homophones and 53 are broad-homophones. Table 3.2.2 provides an overview of the two types of homophones.

---

phonology rules, but there are no such corresponding words in Mandrin. For example, the initial *f* [f] can be spelled with the final *ai* [ai] theoretically, but it doesn't occur in Mandarin.

Table 3.2.2 Overview of Two Types of Chinese Internet homophones

Type	Numbers	%
Narrow-homophones	64	54.70%
Broad-homophones	53	45.30%
Total	117	100%

Table 3.2.2 shows that, in the data, Chinese Internet homophones with the same pronunciation as base words (narrow-homophones) are slightly more frequent than those with similar pronunciation (broad-homophones). The following sections will elaborate on the phonetic features of narrow-homophones and broad-homophones in Chinese Internet discourse. The presentation of Mandarin Chinese initials and finals is transcribed using the International Phonetic Alphabet (IPA). The 4 basic tones are marked based on Chao (1930)'s scale of five pitch levels (see Footnote 17).

3.2.2 Narrow-Homophones

As mentioned earlier, narrow-homophones are homophones which have the same pronunciation as their base words. Within narrow-homophones there are homographic homophones (homophones with the same spelling as their base words) and heterographic homophones (homophones that have a different spelling than their base words. The following sections will explore the two types of narrow-homophones: homographic narrow-homophones and heterographic narrow-homophones.

3.2.2.1 Homographic Narrow-homophones

Homographic narrow-homophones possess the same orthographic form and pronunciation as their base words. For instance, the homophone 囧 [tɕyŋ] ^{214} has the same appearance and pronunciation as its base word, the ancient written form 囧 [tɕyŋ^{214}] ‘bright’. Based on the look of this character, the homophone 囧 has been used to express the look of gloom or embarrassment by internet users. However, the meaning of the homophone 囧 [tɕyŋ] ^{214} ‘gloom or embarrassment’ is completely different from the meaning of the ancient word 囧 [tɕyŋ^{214}] ‘bright’. Another example is the homophone 槑 [mei ^{35}] ‘very silly’ which has the same written form as base word–the ancient character 槑 [mei ^{35}] ‘plum’ (equivalent to the simplified character 梅 [mei ^{35}] ‘plum’). However, the meaning of the homophone 槑 comes from the meaning of each of the two components 呆 [tai^{55}] ‘silly’ rather than from the meaning of the ancient character 槑.

3.2.2.2 Heterographic Narrow-homophones

Heterographic narrow-homophones refer to words having the same pronunciation but different orthographic forms from base words. Based on the data, the homophonic pronunciation of the syllable(s) of heterographic homophones can be found in Chinese characters, alphabets, and numeric letters. The following section will elaborate on the phonetic features of character, alphabet, and numeric letter heterographic narrow-homophones.

First, character heterographic narrow-homophones are heterographic homophones with the same sounds as Chinese characters. In the dataset, the heterographic homophone 杯具 (杯[pei 55] 'cup', 具[tɕy51] 'tool', 杯具 'cup') which has an original meaning of 'cup' and a homophonic meaning of 'tragedy', is read as [pei 55 tɕy51], the same pronunciation as its base word 悲剧 [pei 55 tɕy51] 'tragedy', but they are written in different Chinese characters. The homophone 脖友 [po53 iou214] 'friends made through blogging' (脖 [po53] 'neck', 友 [iou214] 'friend') and the base word 博友 [po53 iou214] 'friends made through blogging' are a pair of homophonic puns. The character 脖 [po35] 'neck' in 脖友 has the same pronunciation as the character 博 [po35] 'blog' in the base word 博友. Although they have the same pronunciation, character heterographic narrow-homophones have totally or partially different written forms from base words.

Alphabet heterographic narrow-homophones refer to the alphabet(s) or letters in heterographic narrow-homophones that represents the same sound as the base word. Most alphabets in homophones represent the same pronunciation of the Chinese characters found in the base words. In the data, there are two specific types of alphabet heterographic narrow-homophones. The first type is that alphabets are used to represent the sound of the base word. The second type is that alphabets are used as acronyms to represent the sound of the base word. As a full Chinese Pinyin form, *KAKA* [k^ha^{55} k^ha^{55}] has the same sound as its base word 咔咔 [k^ha^{55} k^ha^{55}], an onomatopoeic word for laughter. *DING* [t iŋ214] also has the same pronunciation as its base word 顶[t iŋ214] 'to support or to push up the ranking of a posting'. Other examples show that acronyms represent the full pronunciation of the base words, such as *LX* [lou35 ɕia51] (base word 楼下[lou35 ɕia51] 'downstairs'). On BBS (bulletin board system), it usually refers to the comment posted

after the original post or the previous comment in a series of discussion. The alphabetic letters in *ZF* [tʂ əŋ^{51} fu^{214}] (base word 政府 [tʂ əŋ^{51} fu^{214}] 'government') and 傻*B* [ʂa^{214} pi^{55}] (base word 傻逼 [ʂa^{214} pi^{55}] 'fool') also represent the full pronunciation of their base words.

Finally, numeric letter heterographic narrow-homophones are homophones in which the numeric letters represent the same pronunciation as the base words. [19] For example, 2货 [ɚ^{51}xuo^{51}] 'idiot' is the heterographic narrow-homophones of 二货 [ɚ^{51}xuo^{51}] 'idiot', and *2B* [ɚ^{51}pi^{55}] 'stupid' is the heterographic narrwo-homophones of 二逼 [ɚ^{51}pi^{55}] 'stupid'.

3.2.3 Broad-homophones

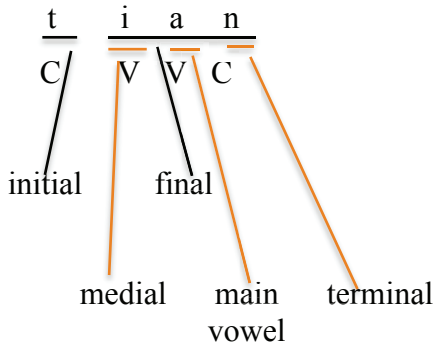
Broad-homophones are heterographic homophones which have a similar pronunciation to their base words. Broad-homophones can be further divided into three types: Chinese character, alphabet, and numeric letter broad-homophones. The following sections will provide detailed information on each type of broad-homophones.

3.2.3.1 Chinese Character Broad-homophones

In this section, as the discussion involves Mandarin Chinese syllable structure, an example with a Mandarin initial and final structure diagram is presented as follows.[20] Mandarin finals are composed of optional medials, main vowels, and optional syllabic terminals. Medials, also

[19] Conventional written forms in Arabic numbers are excluded, such as 1 in 第 1 名 'rank on the first place', and 98 in 98 年 'the year of 1998', etc.
[20] For the definition of Mandarin initials and finals, please see footnote 16.

known as on-glides, are optional sounds in the first position of a final.[21] The main vowel is the essential vowel in a final.[22] A terminal, also known as off-glide, is an optional sound in the last position of a final.[23] For example, in the syllable *tian* [tʰiɛn] 天 'sky', *t* is a initial, *ian* is a final; among the segments of the final, *i* serves as a medial, *a* a main vowel, and *n* a terminal.



Of the three components of a Mandarin syllable (i.e., initials, finals and tones), at least two are the same as their base words in Chinese character broad-homophones (Li, 2007). There are three types of the different phonetic features compared with the base words. The first one is that the initial and final of a homophone are the same as its base word, while the tone differs. This type is very common. The second type is that a homophone retains the same tone, and the same main vowel and terminal (of the final) as its base word. This is also very common. The third type is that a homophone may have a similar initial and tone to its base word, but a different final. It is because that a final is the most important part in a Chinese syllable and it is unlikely to change (Zhou, 2005). This type of phonetic feature is relatively rare. Homophones generally have the same main vowels and terminals as their base words. The following sections will summarize the phonetic features of initials, finals, and tones of Chinese character broad-homophones compared with their base words. In addition, the phonetic features of Chinese transliteration homophones

[21] Only [i], [u],[y] in Mandarin Chinese may be used as medials (e.g., the segment [i] in the syllable *tian* 天 'sky').
[22] Main vowels in Mandarin are [a], [ə], [ɛ], [o], [i], [y] (e.g., the segment [a] in the syllable *tian* 天 'sky').
[23] Only [i], [u], [o], [n], [ŋ] may be used as Mandarin terminals (e.g., the segment [n] in the syllable *tian* 天 'sky').

will be addressed at the end, as their base words are from other languages, which is out of the scope of the Mandarin syllable system.

3.2.3.1.1 Initials of Chinese Internet Homophone

First, summarizing the phonetic features of homophone initials involves explaining both their places and manners of articulation. Homophones tend to have the same manner, but different places of articulation from their base words. The most common cases are outlined below.

(1) *Post-alveolar affricates/fricatives in homophones corresponding to dental affricates/fricatives in their base words and* vice-versa

It is common that post-alveolar affricates/fricatives in homophones correspond to dental affricates/fricatives in their base words. For example, the homophone 肿么 [tʂuŋ^{214} me^{51}] ‘how’ and its base word 怎么[tsən^{214} me^{51}] ‘ how’ have the same manner of articulation (i.e., they share the manner of affrication, but they differ in articulators). The base sound of [ts] is a dental affricate, while the homophonic sound [tʂ] is a post-alveolar affricate. The same feature of initials is also found in the homophone 妹纸 [mei^{55} tʂ ʅ^{214}] ‘sister’ or ‘girl’, and its base word 妹子 [mei^{55} tsʅ^{214}] ‘sister’ or ‘girl’, as well as the homophone 汉纸 [xan^{55} tʂ ʅ^{214}] ‘man’ or ‘true man’ and its base word 汉子[xan^{55} tsʅ^{214}] ‘man’ or ‘true man’, etc. In addition, a dental sound in homophones may also correspond to a post-alveolar sound in the base words. This correspondence is also seen in the dataset. For example, the dental fricative [s] in the homophone 桑心 [saŋ^{55}ɕ in^{55}] ‘sad’ corresponds to the post-alveolar fricative [ʂ] in the base word 伤心 [ʂaŋ^{55}ɕ in^{55}] ‘sad’.

Here, the initials of Mandarin homophones are usually related to those of their base words. Specifically, when they have the same manner of articulation, the place of articular of the homophone initials is usually different from that of the base words. This type of difference may be influenced by local dialects, where there may not be a differentiation between sounds that have the same place of articulation, or the same manner of articulation. For example, [n] and [l], [f] and [x] are not differentiated in Xiang dialect.

(2) *A labiodental fricative [f]in homophones corresponding to a velar fricative [x] in their base words and vice-versa*

The sounds [f] and [x] have the same manner of articulation(i.e., fricative), but different places of articulation. [f] is a labiodental fricative, but [x] is a velar fricative. For instance, the labiodental fricative [f] in the homophone 稀饭 [ɕi^{55} fan^{51}] 'porridge' corresponds to the velar fricative [x] in the base word 喜欢 [ɕi^{214} xuan^{55}] 'like'. However, the velar fricative [x] in the homophone 灰常 [xueɪ^{55} tʂʰɑŋ^{35}] 'very' can also correspond to the labiodental [f] in the base word 非常 [fei^{55} tʂʰɑŋ^{35}] 'very'.

(3) *A nasal [n] in homophones corresponding to a lateral [l] in their base words*

The sounds [l] and [n] have the same place of articulation; both are alveolars. But they have different manners of articulation. [l] is an alveolar lateral, whereas [n] is an alveolar nasal. For example, the initials [n] of the first two Chinese characters of the homophone 内牛满面 [nei^{51} niou^{35} man^{214} miɛn^{51}] 'a face bathed in tears' (内 'inner', 牛 'ox', 满'full,' 面 'face') correspond

to [l] in the base word 泪流满面 [lei^{51} liou^{35} man^{214} miɛn^{51}] ‘a face bathed in tears’ (泪 ‘tear’, 流 ‘flow’).

(4) *A lateral l [l] or a zero-initial in homophones corresponding to a post-alveolar approximant r[ʐ] in their base words*

A lateral *l* [l] or a zero-initial in homophones corresponds to a post-alveolar approximant *r*[ʐ] in the base words. For example, the lateral [l] in the homophone 伦家 [luən ^{35} tɕia^{55}] ‘I’ corresponds to the post-alveolar approximant [ʐ] in the base word 人家 [ʐuən ^{35} tɕia^{55}] ‘I’. Yet the zero-initial in the homophone 男淫 [nan^{35} in^{35}] ‘man’ corresponds to the sound [ʐ] in the base word 男人 [nan^{35} ʐ̩en^{35}] ‘man’.[24]

3.2.3.1.2 Finals of Chinese Internet Homophone

Mandarin finals include medials, main vowels, and syllabic terminals.[25] Comparing the finals of homophones and those of their base words, there seems to be more regular patterns. The medials of homophones tend to be unrounded, while those of the base words are rounded.[26] The phonetic features of main vowels involves the degree of openness (vowel height), the location of the active part of the tongue (vowel backness), and lip rounding (vowel rounding) (Lin, 2007, p.63). In the data, generally, the main vowels of homophones tend to be higher and unrounded. In addition, for the phoneme /e/ in the base words, a high front unrounded [i] in homophones corresponds to a mid-central unrounded vowel [ə] in the base words, and a low central vowel [a]

[24] The homophone 淫[in^{35}] only has the main vowel and the terminal.
[25] For the diagram of Mandarin Chinese final structure, please see Page 35.
[26] Glides refer to transitional sounds.

in homophones corresponds to a mid-high front unrounded vowel [e] in the base words. The phonetic features of homophones also involve the addition or deletion of main vowels compared to their base words. As for terminals, alveolar nasal and velar nasal can interchange, but there doesn't seem to be regular rules in place. Some typical cases of finals in the data are outlined below.

(1) *A high front unrounded medial [i] in homophones corresponding to a high front rounded medial [y] in their base words*

Through homophonization, a high front unrounded medial [i] in homophones corresponds to a high front rounded medial [y] in the base words. For example, [i] in the homophone 童鞋[tʰuŋ³⁵ ɕ iɛ ³⁵] 'classmate' corresponds to the medial [y] in the base word 同学[tʰuŋ³⁵ ɕyɛ ³⁵] 'classmate'.

(2) *The phoneme /e/ in the base words*

In the data, the phoneme /e/ in the base words can be realized as a mid-central unrounded vowel [ə] and a mid-high front unrounded vowel [e].

A high front unrounded [i] in homophones corresponding to a mid-central unrounded vowel [ə] in the base words. For example, [i] in the homophone 男淫[nən ³⁵ in³⁵] 'man' corresponds to the main vowel [ə] in the base word男人[nan ³⁵ ʐ ən³⁵] 'man'.

A low central vowel [a] in homophones corresponds to a mid-high front unrounded vowel [e] in the base words, and vice-versa. For instance, [a] in the homophone 神马 [ʂən^{35} ma ^{214}] 'what' corresponds to [e] in the base word 什么[ʂən ^{21} me^{55}] 'what', whereas [e] in the homophone 特么[t^{h}e^{51} me^{55}] 'damn it' or 'fuck' corresponds to [a] in the base word 他妈 [t^{h}a ^{55} ma ^{55}] 'damn it' or 'fuck'. In addition, a high front unrounded vowel [i] in homophones corresponds to [e] in the base words. For example, [i] in the homophone 虾米 [ɕia ^{55} mi^{214}] 'what' corresponds to [e] in the base word 什么[ʂən ^{21} me^{55}] 'what'. A back rounded diphthong [uo] corresponds to [e] in the base words as well. For example, [uo] in the homophone 帅锅 [ʂuai ^{51} kuo^{55}] 'handsome man' corresponds to [e] in the base word 帅哥[ʂuai ^{51} ke^{55}] 'handsome man'.

(3) *Addition or deletion of main vowels*

When considering the phonetic features of homophones' finals, there can be addition or deletion of vowels. Addition of vowels can make a diphthong in homophones correspond to a monophthong in the base words. For example, the diphthong [uə] in the homophone 伦家 [l uən ^{35} tɕia^{55}] 'I' corresponds to the single vowel [ə] in the base word 人家[ʐən^{35} tɕia^{55}] 'I' (the initial [l] corresponds to [ʐ]). The diphthong [ou] in the homophone 偶棉 [ou^{214} miæn^{35}] 'we' corresponds to the main vowel-monophthong [o] in the base word 我们[uo^{55} m ən ^{35}][27] 'we'. The deletion of vowels may make a monophthong in homophones correspond to a diphthong in the base words. For example, the monophthong [i] in the homophone 米国[mi^{214} kuo^{25}] 'USA' corresponds to the diphthong [ei] in the base word 美国[mei^{214} kuo^{35}] 'USA'.

[27] The sound [u] in 我 [uo^{55}] is the medial, and it disappeared in the homophone.

(4) *A velar nasal terminal [ŋ] in homophones corresponding to an alveolar nasal terminal [n] in their base words and vice-versa*

A velar nasal terminal [ŋ] in homophones corresponds to an alveolar nasal terminal [n] in their base words, and vice-versa. For instance, the velar nasal terminal [ŋ] in the homophone 肿么 [tsuŋ ^{214} me^{55}] 'how' corresponds to the alveolar nasal terminal [n] in the base word 怎么 [tsən ^{214} me^{55}] 'how'. The alveolar nasal terminal [n] in the homophone 盆友 [pʰ ən^{53} iou ^{214}] 'friend' corresponds to the velar nasal terminal [ŋ] in the base word 朋友 [pʰəŋ^{53} iou ^{214}] 'friend'.

3.2.3.1.3 Tones of Chinese Internet Homophones

In the data, there is no fixed pattern related to the tones of Chinese Internet homophones compared with their base words. The tones appear to be somewhat random, mainly due to the way that words are typed on a computer keyboard. Chinese is a tonal language, and keyboard input cannot reflect tones. The frequency of the words popped up on the screen is based on common lexicon. Often instead of choosing the correct word, Internet users pick a homophone, which appears first among the many alternative forms with the same initial and final on the computer keyboard. Sometimes, they are new lexical items, such as 矮油'oh', 'wow', 'my God' [ai^{214} iou^{35}] (base word 哎呦[ai^{55} io^{55}] 'oh', 'wow', 'my God'), or 怪蜀黍'weird uncle'[k uai^{51} ʂu^{214} ʂu^{35}] (base word 怪叔叔[k uai^{51} ʂu^{55} ʂu^{55}] 'weird uncle'). In other cases, homophones are based on pre-existing words in Chinese, such as 斑竹[pan^{55} tʂu^{35}]'mottled bamboo' (base word 版主[pan^{214} tʂu^{214}] 'board moderator'), 鸭梨[ia^{55} li^{35}] 'pear' (base word 压力[ia^{55} li^{51}] 'pressure'),

and 河蟹[xe^{35} ɕiɛ^{51}] ‘river crab’(base word 和谐[xe^{35} ɕiɛ^{51}] ‘harmonious’, a political concept of the Chinese government). However, these cases are not commonly seen in the current data.

In addition, as homophones may also have base words originating from other languages, the Chinese transliteration homophones undergo certain sound changes from the pronunciation found in the original language. As these transliteration words use Chinese pronunciation, the forms are not entirely the same as those of their base words. This type of homophone often borrows similar sounds from Chinese to express certain meanings. Some changes that can occur during transliteration include addition of tone, and addition or subtraction of syllables. Some transliteration words have a significant sound change, while others undergo slight sound changes. For example, 粉丝 [fən ^{214} sɿ^{55}], transliterated from the English word *fans* [fæns], is changed from one syllable to two syllables in addition to tone and vowel changes. The added vowel [ɿ] adds one syllable to the homophone. 卢瑟 [lu^{35} se^{51}] is mapped from the English pronunciation *loser* [ˈluzɚ]. Although there is no change to the number of syllables, there is an addition of Chinese tones (this automatically leads to a change in pitch). The latter syllable [zɚ] in the base word *loser* is changed to [se] in the homophone. Usually the changes involve close consonants or vowels changes and tone addition. In the database, there are only minor sound changes in Chinese transliteration words.

3.2.3.2 Alphabets and Numeric Letters

Most of the alphabet homophones retain the pronunciation of Chinese characters. The letters in alphabet homophones are usually acronyms of the Pinyin form of the base words and therefore

represent the pronunciation of their base words. For example, the homophone *ZF* [tʂ əŋ^{51} fu^{214}] 'government' comes from the base word 政府 [tʂ əŋ^{51} fu^{214}] 'government'. However, there are some exceptions that are more complex and more random. For example, the word string *3Q* referring to the English word 'thank you' comprises of the Chinese pronunciation of the numeric letter 3 [san^{55}] and the English pronunciation of *Q* [kju:]. As for *V5* [vei^{55} u^{214}] (base word 威武 *wei*1*wu*3[uei^{55} u^{214}] 'mighty') , the letter *V* (a labiodental fricative) replaces the first letter *W* of 威's Pinyin form.[28] It is likely that this spelling and sound change is influenced by an existing Chinese dialect. The sound change from [u] to [v] can be found in the Chinese Ningxia dialect and the Gulin dialect (Huang & Liao, 2008, p.46).

As for numeric letters, a single numeric letter can represent different homophonic sounds. For instance, 3 [san^{55}] in 3166 [san^{55} i^{55} liou^{51} liou^{51}] (base word Japanese さようなら[ˌsaɪəˈnɑːrə], 'goodbye') is a homophonic sound to さ[sa], 3 in 1314 [i^{55} san^{55} i^{55} sɿ^{51}] (base word 一生一世[i^{51} ʂəŋ^{55} i^{35} ʂʅ^{51}] 'in all one's life') is a homophonic sound of 生 and 3 in *3Q* [san^{55} kju:](base word *thank you*) is a homophonic sound of *than* [θæŋ]. These examples show that numeric letter homophones have certain variability. In addition, the pronunciation of numeric letters in some dialects co-exists with that of Mandarin in the Chinese Internet homophones. For example in 1314 (一生一世) and *B4* [pi^{214} sɿ^{4}] (base word 鄙视 [pi^{214} ʂʅ^{51}] 'to despise or disdain'), 4 [sɿ^{51}] presents the homophonic sound of 世 [ʂʅ^{51}] and 视[ʂʅ^{51}], which originates from some southern dialects. In these southern dialects, there is no distinction between dental and post-alveolar initials. So in this case, people use the dental [s] to replace the post-alveolar [ʂ].

[28] In Mandarin Pinyin spelling, when the final is a single *u*, *w* will be added in front of *u*; if the final begins with *u*, then *u* will be changed to *w*.

3.2.4 Interim Summary

The special phonetic features make homophony an interesting linguistic phenomenon in Chinese. Chinese Internet homophones are formed with identical pronunciation (narrow-homophones) or similar pronunciation to their base words (broad-homophones). The data show that the number of narrow-homophones is slightly higher than that of broad-homophones. In this section, I have explored the phonetic features of narrow-homophones and broad-homophones. Narrow-homophones can have the same spelling as their base words (homographic) or different spelling from their base words (heterographic). Most of the alphabet homophones are narrow-homophones. Broad-homophones, which are heterographic homophones, can be divided into homophones with similar sounds to Chinese characters, alphabets, and numeric letters. Specifically, the analysis of the phonetic features of Chinese character broad-homophones is based on the changes to initials, finals, and tones from their base words. To sum up, more than half of the Chinese Internet homophones have the same pronunciation as their base words, while comparing with the base words, the initials of homophones tend to differ in the place of articulation, while the manner of articulation remains the same. Vowels in homophones tend to be unrounded and are higher than the vowels in the base words. Chinese dialects and even foreign languages can influence the phonetic features of Chinese Internet homophones.

3.3 Morphological Features of Different Types of Chinese Internet Homophones

Based on their morphological features, homophones in Chinese Internet discourse can be (Chinese) character homophones, alphabet homophones, numeric letter homophones, or a blend of characters, alphabets and numeric letters. However, after a homophone is formed, its

morphological type may be changed. This section will offer a classification of homophones based on their morphological structures.

3.3.1 Types of Chinese Internet Homophones Based on Their Morphological Features

Traditionally, Chinese morphological classification is solely based on Chinese characters. However, in addition to Chinese characters, Chinese Internet homophones are comprised of alphabets and numeric letters. Based on their morphological structure, Zhou (2008) categorized Chinese Internet lexicon into four types: Chinese characters, alphabets, numeric letters, and blends (p. 434). This study adopts Zhou's categorization.

Among the 117 Chinese Internet homophones identified in the data, there are 55 character homophones (47%), 46 alphabet homophones (39.32%), 4 numeric letter homophones (3.42%), and 12 blending homophones (10.26%). Here I also provide an overview of the occurrence and distribution of each type of homophone in the data. The occurrences of character homophones, alphabet homophones, numeric letter homophones, and blending homophones are shown in Table 3.3.1 below.

Table 3.3.1 Types of Homophones With Their Token Frequency and Distribution[29] in the Data

Type	Token Frequency	%
Chinese character homophones	41,641	55.08%
Alphabet homophones	29,223	38.66%
Numeric letter homophones	3,763	4.98%
Blending homophones	965	1.28%
Total	75,592	100%

Table 3.3.1 shows that the Chinese character homophones occurred the most frequently (55.08%) followed by the alphabet homophones (38.66%), the numeric letter homophones (4.98%) and the blending homophones (1.28%). It can be seen that character homophones and letter homophones are used significantly more frequently than numeric letter homophones and blending homophones. Character homophones are the most frequently used type. This result is consistent with Zhou (2008a)'s finding that Chinese character homophones are the most frequently used in Internet communication, compared to other types of homophones. The following sections will elaborate on the morphological features of each type of homophones.

3.3.2 Chinese Character Homophones

A Chinese word, *ci*, is defined as the smallest meaningful language unit which can be used independently (Hu, 1992, p.230). Morphemes are the minimal units of meaning. A Chinese word

[29] The distribution is calculated based on the formula: $F_i = \frac{n_i}{N} * 100\%$, where F_i is the target i's frequency, n_i is the occurrences of target i, and N is the total occurrences of all targets.

consists of one or more morphemes. Based on the number of morphemes that the words have, they are classified into single-morphemic words and multi-morphemic words. In Chinese, the number of morphemes and that of characters (one character corresponds to one syllable) are not in a corresponding one-to-one relationship. In this section, the examples in Chinese will be noted with Chinese Pinyin.

Firstly, a single-morphemic word only has one morpheme, including monosyllabic and polysyllabic single-morphemic words. For instance, 败 *bai*4 'defeat' (base word English *buy*) is a single-morphemic word, and it only has one syllable. However, 特么 *te*4*me*1 'damn it/fuck' (特, 'special', 么, '*me* (particle)') with the corresponding base word 他妈 *ta*1*ma*1 'damn it/ fuck' (他 'his', 妈 'mother') is a single-morphemic word composed of two syllables/characters. Single-morphemic homophones with two or more syllables are considered to be polysyllabic single-morphemic words, such as 特么.

Secondly, multi-morphemic words consisting of more than one morpheme include compounds, reduplications, and derivations (Gu, Pan, & Liu, 2001, p.10). Most words in modern Chinese are compounds. Compounding is the most productive method of word-formation in Chinese (Huang & Liao, 2003; Hu, 1992; etc.). For instance, 杯具 *bei*1*ju*4 'cup /tragedy' (base word 悲剧 *bei*1*ju*4 'tragedy') has two morphemes, 杯 *bei*1 'cup' and 具 *ju*4 'tool'. The morpheme 洗 *xi*3 'wash' and another morpheme 具 *ju*4 'tool' can be combined together, forming the compound word 洗具 *xi*3*ju*4 'cleaning set/comedy' (base word 喜剧 *xi*3*ju*4 'comedy').

"Reduplication is a morphological process in which a root or stem or part of it is repeated" (Hartmann & Stork, 1972, p. 193). Some scholars categorize reduplicated words into multi-morphemic words (Huang & Liao, 2002), while others consider them as single-morphemic words (Sun, 2006). In this study, reduplicated words are treated as multi-morphemic words. There are no reduplicated homophones found in the data.[30]

Derivation is the formation of a new word from another word or stem (Hartmann & Stork, 1972, p. 62). For instance, the Chinese suffix 子 *zi*3 'zi' is usually attached to morphemes and denotes small or round things, kinship terms or persons. 筒子 *tong*3*zi* 'thick tube-shaped thing/ comrade' (base word 同志 *tong*2*zhi*4 'comrade') is constituted by the root morpheme 筒 and the suffix 子.

A summary of the morphological structure of Chinese Internet homophones in the data is shown in Table 3.3.2.

[30] They are only examples of reduplicated words found in the base words. If we look at the example of the morpheme 妹 *mei*4 'sister', it can be repeated to create the base word 妹妹 *mei*4*mei* 'sister' (homophone 美眉 *mei*3*mei*2, 美'beautiful', 眉 'eyebrow', 美眉 'beautiful girls').

Table 3.3.2 Morphological Structure of Chinese Character Homophones

Single-morphemic words	Monosyllabic single-morphemic words
	Polysyllabic single-morphemic words
Multi-morphemic words	Compounds
	Reduplications
	Derivations

Table 3.3.3 provides an overview of the top 10 most frequently used character homophones and their morphological structure found the data.

Table 3.3.3 Top 10 Most Frequently Used Chinese Character Homophones

Homophone	**Homophonic meaning**	**Base word**	**Morphological structure**	**Occurrences**
神马 *shen*3*ma*3 ‘magic horse’	what	什么 *shen*3*me* ‘what’	polysyllabic single-morphemic word	9,139
围脖 *wei*2*bo*2 ‘wrap neck’	mirco-blog	微博 *wei*2*bo*2 ‘micro blog’	compound word	5,152
尼玛 ni2ma3 ‘priestess agate’	fuck off	你妈 *ni*3*ma*1 ‘your mom’	polysyllabic single-morphemic word	4,534
粉丝 *fen*3*si*1 ‘vermicelli’	fans	fans Enthusiastic followers of a sport, celebrity, etc	polysyllabic single-morphemic word	4,495
表 *biao*3 ‘watch’	don’t	不要 *bu*2*yao*4 ‘not want’	monosyllabic single-morphemic word	3,081
肿么 *zhong*3*me*1 ‘swelling/swell *me* (particle)	how	怎么 *zen*3*me1* ‘how’	polysyllabic single-morphemic word	2,837
囧 *jiong*3 ‘expressing the look of gloom or embarrass’	glooming, embarrassing	囧 (ancient Chinese) *jiong*3 ‘bright’	monosyllabic single-morphemic word	1,716
虾米 *xia*1*mi*3 ‘shrimp rice or little shrimp’	what	什么 *shen*3*me*1 ‘what’	compound word	1,575
擦 *ca*1 ‘wipe’	fuck	操 *cao*4 ‘fuck’	monosyllabic single-morphemic word	961
特么 *te*4*me*1 ‘special *me* (particle)’	damn it/ fuck	他妈 *ta*1*ma*1 ‘his mother’	polysyllabic single-morphemic word	927

Within the category of character homophones, single-morphemic words are used more frequently than multi-morphemic words. This is because homophones are created based on pronunciation, and the newly created words (homophones) do not tend to reflect the number of morphemes in the base word. For example, in some southern dialects, the multi-morphemic word 不要 (*bu*2 *yao*4 'don't') sounds like a monosyllabic word *biao*3. As a result, some internet users started to type the single-morphemic word 表 (*biao*3 'watch') as a homophone of 不要.

3.3.3 Alphabet Homophones

Alphabet homophones are written in alphabetic letters, which derive from the Pinyin form of each Chinese base word. Some alphabet homophones are written in Pinyin or Pinyin with some variations. For example, *Ding* is a homophone of the base word 顶 (*ding*3 'to support' or 'to push up the ranking of a posting'), while *HOHO* is a homophonic pun for the base word 吼吼 (*hou*3*hou*, 'an indicator of emotion representing happiness or excitement'). Acronym homophones are a major type within the category of alphabet homophones. For example, 人品 *ren*2*pin*3 'luck' is the base word for the homophone *RP*. When Internet users see *PR*, they read it as *ren2pin3,* rather than pronouncing the alphabets.

Table 3.3.4 provides a summary of the top 10 most frequently used alphabet homophones in the data. As alphabet homophones are not case-sensitive, lower case, upper case, and combination of the two are counted in the data. For instance *PR*, *Pr*, and *pr* are all homophones of 人品

*ren*2*pin*3 ‘luck’, but in the table only the most frequent form, *PR*, is used to represent all the spellings.

Table 3.3.4 Top 10 Most Frequently Used Forms of Alphabet Homophones

Homophone	Homophonic meaning	Base word	Occurrences
SB	idiot, asshole	傻逼 *sha*2*bi*1 ‘idiot /asshole’	17,469
RP	personality, generalized for luck	人品 *ren*2*pin*3 ‘personality’	3,706
MM	sister, generalized for girls	妹妹 *mei*4*mei* ‘sister’	1,703
	girls, especially beautiful girls	美眉 *mei*3*mei*2 ‘girls’	
YY	unrealistic imagination	意淫 *yi*4*yin*2 ‘*unrealistic imagination*’	832
ZZ	to forward others' posting	转载 *zhuan*3*zai*3 ‘to forward others' posting’	460
LG	husband	老公 *lao*3*gong*1 ‘husband’	452
BT	abnormal	变态 *bian*4*tai*4 ‘abnormal’	447
PP	butt	屁屁 *pi*4*pi* ‘butt’	384
	picture	片片 *pian*1*pian* ‘picture’	
HOHO	indicator of emotion, showing happiness or excitement	吼吼 *hou*3*hou* ‘modal particle’	350

Generally, alphabets are easy to type. This appears to make alphabet homophones one of the most frequently used types in the data. However, the pronunciation of alphabet homophones is

different from English abbreviations.[31] For instance, *WTO* stands for *World Trade Organization* in English and is pronounced as the three alphabets, which is different from what we saw above. However, polysemous alphabet homophones (alphabet homophones with two or more different meanings) may cause ambiguity. For most of the alphabet homophones in Chinese, each of them corresponds to one base word. But some alphabet homophones correspond to two or more base words. For instance, *PP* is a homophone of 屁屁 (*pi*4*pi* 'butt') and 片片 *(pian*1*pian* 'picture'). In these particular cases, the reader needs to rely on the context to decide which the base word is.

3.3.4 Numeric Letter Homophones

Numeric letter homophones consist of Arabic numbers based on two types of pronunciations. Numeric letters in the first type present the same or close pronunciation in Chinese. As to the second type, numeric letters present the same or close pronunciation of other languages, such as Japanese. For example, Internet users use the numeric letter homophone 88 (*ba*1*ba*1 'bye-bye'; base word 拜拜 *bai*4*bai*4 'bye-bye'), which comes from the English word *byebye*, when they say goodbye to each other over the Internet. The frequency of numeric letter homophones is relatively low. The total occurrences in the data are 3,763 (4.98 % of all Chinese Internet homophones). [32] Table 3.3.5 shows the numeric letter homophones found in the data.

[31] In addition to their pronunciation being different than English abbreviations, Chinese alphabet letter homophones are non-case sensitive, while English abbreviations need to be capitalized.

[32] In the current study, phrases with numeric letters are excluded, such as 520 *wu*3*er*4*ling*2 'I love you' (base phrase 我爱你 *wo*3*ai*4*ni*3 'I love you'), 530 *wu*3*san*1*ling*2 'I miss you' (base phrase 我想你 *wo*3*xiang*3*ni*3 'I miss you'), 7456 *qi*4*si*4*wu*3*liu*4 'piss me off' (base phrase 气死我了 *qi*4*si*3*wo*3*le*1 'piss me off'). Their base expressions are longer than lexical units, which are beyond the scope of this study.

Table 3.3.5 Most Frequently Used Forms of Numeric Letter Homophones

Homophone	**Homophonic meaning**	**Base word**	**Occurrences**
88 *ba*1*ba*1	bye-bye	拜拜 *bai*4*bai*4 'bye-bye'	3,334
555 *wu*3*wu*3*wu*3	crying	呜呜呜 *wu*1*wu*1*wu* 'imitative word representing crying'	352
1314 *yi*1*san*1*yi*1*si*4	in all one's life	一生一世 *yi*4*sheng*2*yi*2*shi*4 'in all one's life'	66
3166 *san*1*yi*1*liu*4*liu*4	bye-bye	さようなら (Japanese) Sayonara 'bye-bye'	11

Sometimes, one numeric letter can represent different characters and thus has different pronunciations. As is shown in table 3.3.5, 3 in 1314 represents the characters 生 *sheng*1, whereas 3 in 3166 represents the syllable さ *sa* in Japanese.

3.3.5 Blending Homophones

Blending homophones are a mix of characters, alphabets and/or numeric letters. The frequency of blending homophones is the lowest among the four morphological types, totaling 965 in the corpus (1.28% of the total occurrences). Table 3.3.6 lists top 10 most frequently used forms of blending homophones.

Table 3.3.6 Top 10 Most Frequently Used Forms of Blending Homophones

Homophone	**Homophonic meaning**	**Base word**	**Occurrences**
V5 *vei*1*wu*3	mighty and domineering	威武 *wei*1*wu*3 ‘power military’	442
2B *er*4*bi*1	stupid	二逼 *er*4*bi*1 ‘two female-genitalia’	412
B4 *bi*3*si*4	to despise	鄙视 *bi*3*shi*4 ‘despise look-on’	63
3Q *san*1*Q*	thank you	thank you	21
傻 B *sha*3*bi*1	fool	傻逼 *sha*3*bi*1 ‘stupid female-genitalia’	14
小 P 孩 *xiao*3*pi*4*hai*2	young man (pejorative)	小屁孩 *xiao*3*pi*4*hai*2 ‘little butt child’	5
2 货 *er*4*huo*4	idiot	二货 *er*4*huo*4 ‘two goods’	4
P 民 *pi*4*min*2	shitizen	屁民 *pi*4*min*2 ‘fart citizen’	1
装 B *zhuang*1*bi*1	cocky	装逼 *zhuang*1*bi*1 ‘pretend female-genitalia’	1
牛 B *niu*2*bi*1	freaking awesome	牛逼 *niu*2*bi*1 ‘ox female-genitalia’	1

Here we can see that blending homophones are constituted by a blend of characters and numeric letters, such as 2 货 *er*4*huo*4 ‘idiot’ (base word 二货 *er*4*huo*4 二 ‘two’, 货‘goods’, 二货 ‘idiot’). We also find combinations of characters and alphabetic letters, such as 装 B *zhuang*1*bi*1 ‘cocky’ (base word 装逼 z*huang*1*bi*1 装 ‘pretend’, 逼 ‘female-genitalia’, 装逼 ‘cocky’). Blending homophones can also be formed by the combination of alphabetic letters and numeric

letters. For example, *B*4 *bi*3*si*4 ‘despise’, ‘look-on’, or ‘to despise’ is based on the Chinese pronunciation of the base word 鄙视 *bi*3*shi*4 ‘despise’, ‘look-on’, or ‘to despise’. Unlike the homophones in Section 3.2.3, where alphabets are based on the Chinese pronunciation, here we see another type. Some alphabets in blending homophones have a similar pronunciation to English. For instance, 3*Q* *san*1*q* ‘thank you’ is based on the Chinese pronunciation of the numeric letter 3 *san1* and the English pronunciation of the letter *q*.

3.3.6 Interim Summary

The morphological classification of Chinese homophones from the data reveals the diversity in the methods used to form Chinese Internet homophones. Internet homophones are composed of not only Chinese characters, but also alphabets and numeric letters. Alphabet homophones, numeric letter homophones, and blending homophones are not formed through the traditional methods of word-formation in Chinese. This shows the creativity and vitality of Chinese Internet language.

From the statistical overview of each morphological type of Chinese Internet homophones in the data, the most frequently used homophones are character homophones, with a frequency of 55.08%. The second most frequently used homophones are alphabet homophones, with a frequency of 38.66%.

3.4 Summary

In this chapter, the phonetic features and morphological structures of Chinese Internet

homophones have been studied. Section 3.1 provides an overview of the use of the Chinese Internet homophones and their base words. Among the Top 10 most frequently used homophones in the data, the homophones are generally used rather frequently compared to their base words. Section 3.2 examines the phonetic features of Chinese Internet homophones and classifies them into narrow-homophones and broad-homophones. More than half of the homophones have the same pronunciation as their base words. In broad-homophones, the initials of homophones tend to differ from their base words, when the place of articulation or the manner of articulation remains the same. Vowels in broad-homophones tend to be unrounded and higher than those of the base words. Section 3.3 explores the morphological features of Chinese Internet homophones. They are classified into four types, Chinese character homophones, alphabet homophones, numeric letter homophones, and blending homophones. In the data, character homophones and alphabet homophones are used much more frequently than numeric letter homophones and blending homophones.

Chapter 4 The Discourse Function of Chinese Internet Homophones

This chapter will analyze the discourse function of Chinese Internet homophones based on the classification of cohesion proposed by Hu et al. (2005) within the theoretical framework of systemic functional grammar (SFG) (Halliday & Hassan, 1976; Halliday, 1985). Cohesion is the connection within a text that makes it a semantic unit (Halliday & Hasan, 1976, p.4). Chinese Internet homophones have certain phonetic and morphological features (Chapter 3), and are related to the lexical meaning of their base words. This chapter will examine the cohesion function of Chinese Internet homophones based on the Leiden Weibo Corpus (LWC). We can see that the lexical units are connected through the use of Chinese Internet homophones as cohesive devices. The analysis in this chapter will also show the semantic relation between different units within the Chinese Internet discourse.

Chapter 2 summarized the development of cohesion theory in the English and Chinese literature. In this study, I use the framework of cohesion proposed by Hu et al. (2005) which is developed from Halliday & Hasan (1976). The reasons for adopting this classification have been discussed in Chapter 2. This classification includes grammatical cohesion and lexical cohesion. Grammatical cohesion includes reference, ellipsis, substitution, and conjunction. Lexical cohesion is comprised of repetition, synonymy/antonymy, hyponymy/meronymy, and collocation. Table 4.1 provides a detailed summary of Hu et al. (2005)'s classification of cohesion. It is used in the data analysis in Section 4.1 and Section 4.2.

Table 4.1 Hu et al. (2005)'s Classification of Cohesion

Category		Subcategory
Grammatical Cohesion	Reference	Personals Demonstratives Comparatives
		Endophoric:Anaphoric, cataphoric Exophoric
	Ellipsis	Nominal ellipsis Verbal ellipsis Clausal ellipsis
	Substitution	Nominal substitution Verbal substitution Clausal substitution
	Conjunction	Transitional words
Lexical Cohesion	Repetition	
	Synonymy/Antonymy	
	Hyponymy/ Meronymy	
	Collocation	

In this chapter, the analysis will focus on homophones that serve as lexical and grammatical cohesive devices. In what follows, Section 4.1 will examine the grammatical cohesion function of Chinese Internet homophones, and Section 4.2 will investigate the lexical cohesion function of Chinese Internet homophones.

4. 1 Grammatical Cohesion

According to Halliday & Hassan (1976) and Hu et al. (2005), grammatical cohesion includes reference, ellipsis, substitution, and conjunction. The following section will explore the different types of grammatical cohesion function of homophones in Chinese Internet discourse. Following

the definition of each type of cohesion, examples from the data Leiden Weibo Corpus will be provided.

4.2.1 Reference

Reference means that a participant or circumstantial element serves as another element's referent. If an element serves as a reference, it makes the information signaled for retrieval. When the referential meaning is retrieved, "the cohesion lies in the continuity of reference, whereby the same thing enters into the discourse a second time" (Halliday & Hasan, 1976, p. 31). Reference includes personal reference, demonstrative reference, and comparative reference. Usually, personal reference words refer to personal pronouns, such as *you* and *him*. Comparative references refer to words with "comparative" meanings, such as *same as* and *different*. Demonstrative references include definite articles (e.g. *that*) along with place and time demonstrative words (e.g. *there* and *now*). In Chinese, they are usually proximals, such as 这 (*zhe*4 'this'), 此 (*ci*3 'this'), and 这里 (*zhe*4*li*3 'here') and distal-deixis, such as 那 (*na*4 'that'), 彼 (*bi*3 'that'), and 那里 (*na*4*li*3 'there').

Reference can be divided into exophora and endophora depending on the relationship between the referent, context, and discourse. The referent of an exophoric reference is found in the context of situation. The referent of an endophoric reference lies in the surrounding text. Furthermore, endophoric reference can be subcategorized into anaphora and cataphora, based on the position of the referent within discourse. The referent of an anaphoric reference is in the preceding text, while that of a cataphoric reference is in the following text.

The following examples are drawn from the data for analysis. Gloss translation is provided to the Chinese examples.[33] Each example contains a unique message ID which can be found in the Leiden Weibo Corpus.

The following example is about a leading man in a drama, which involves both demonstrative and anaphoric reference.

[4: 1] (LWC, Message ID 3401883045887457)

1. 难怪　　个个　都　喜欢上　　李　大仁，

nan2gui4　ge4ge4 dou1 xi3huan1shang4　**Li3 Da4ren3**,

no doubt　everyone　all　　like　　**Li Daren**,

‘There is no doubt that everyone likes **Li Daren**.’

2. 介个　火星　好　男人，也　只有　电视剧　会　出现，

jie4ge4 huo3xing1 hao3 nan2ren2, ye3　zhi2you3　dian4shi4ju4　hui2　chu1xian4,

this　Mars　good　man,　yet　only　television　will　appear,

‘ Yet **this** good Martian guy only appears in television,’

3. 因为　本　故事　纯粹　虚构. ...

yin1wei4　ben3　gu4shi4　chun2cui3　xu1gou4

because　this　story　indeed　fabricate

‘as　this story is indeed fabricated…’

[33] For abbreviations used in the glosses of the Chinese examples, please see Appendix.

In line 2, 介个 (*jie*4*ge*4, 介 'interpose', 个 'CL', 介个'this'), a homophone of 这个 (*zhe*4*ge*4 'this /these'), derived from the Tianjin vernacular, a variation of the Northern Mandarin dialect. The homophone 介个 refers to 李大仁 (*Li*3 *Da*4*ren*3, 'a leading man in a Taiwan idol drama'). 介个 serves as a demonstrative reference and also an endophoric reference. The referent of 介个 is 李大仁. As the referent is in the preceding text, 介个 can be further categorized as an anaphoric reference. The reference of the homophone 介个 creates cohesion, as it links the two clauses into a cohesive unity.

Example [4:2] is a complaint from a bank clerk resulting from many people coming to his branch before the Chinese New Year. It exemplifies exophoric, demonstrative, and personal reference.

[4:2] (LWC, Message ID 3402444914763617)

1. 人山 人海！！！ 人山 人海！！！ 人山 人海 啊！！！

ren2shan1 ren2hai3! ren2shan1 ren2hai3! ren2shan1 ren2hai3 a1!

people mountain people sea! people mountain people sea! people mountain people sea PRT!

'A sea of people! A sea of people! A sea of people!'

2.<u>介个</u> 不 会 按 密码， <u>类个</u> 不 会 填 单子，

<u>jie4ge4</u> bu2 hui4 an4 mi4ma3, **<u>lei2ge4</u>** bu2 hui4 tian2 dan1zi,

<u>this</u> NEG can press password, **<u>that</u>** NEG can fill form,

'<u>This</u> [person] doesn't know how to key in [his] password, **<u>that</u>** [person] doesn't know how to fill out the form,'

3. 还有　　一批　抱怨　　排队　太久　　的　人，

hai2you3　yi4pi1　bao4yuan4　pai2dui4　tai4jiu3　de1　ren2,

and　　a　CL　　complain　　line up　too long　ASSC　people,

‘and there is a group of people complaining that they are waiting in line too long.’

4.乃们　　把　　我　　劈 成　　几半　　好么，

nai3men2　ba3　wo3　pi1cheng2　ji3ban4　hao3 me,

2nd PL　　BA　1stSG　chop into　several half　OK,

‘Would **you** please chop me into several pieces?’

5. 谁　　规定　　过年　　前　　灰得[34]　来　　银行　　凑　　一脚　　哒？

shui2 gui1ding4 guo4nian2 qian2　hui1de3 lai2　yin2hang2　cou4　yi1jiao3　da1?

who　　set　　new year　before　have to　come　　bank　　add　　a feet　PRT?

‘Who told [you] that [you] must come to the bank before the [Chinese] New Year?’

6. 抱头　　痛哭　啊　我

bao4 tou2　tong4 ku1　a1　wo3.

hug head　cry hard　PRT　1stSG.

‘I [am] hugging [my] head and crying sadly.’

In Line 2, 介个 (*jie*4*ge*4, 介 ‘interpose’, 个 ‘CL’, 介个‘this’) is a homophone of 这个 (*zhe*4*ge*4 ‘this/these’) and 类个(*lei*4*ge*4, 类 ‘category’, 个 ‘CL’, 类个‘that’) is a homophone of 那个 (*na*4*ge*4 ‘that/those’). In Lin 4, 乃们 (*nai*3*men*2, 乃 ‘you’ in ancient Chinese, 们 ‘suffix for plural personal pronoun’, 乃们 ‘you’) is a homophone of 你们 (*ni*3*men*2 ‘you’). 介个, 类个,

[34]灰得(*hui*1*de*3, 灰 ‘grey’, 得 ‘have to’ , 灰得 here it means ‘have to’) is a homophone of 非得 (*fei*1*de*3 ‘have to’).

and 乃们 all serve as reference in the discourse. From the message, no referent can be found in the discourse itself. By reading the message, this micro-blog user works at a bank and he was complaining about his customers who came to the bank the day before the Chinese New Year. Retrieving information from the context, 介个, 类个, and 乃们 all refer to this micro-blog user's customers. As the referent is in the context of situation, they all serve as exophoric reference.介个 and 类个 are also demonstrative reference, and 乃们 is a personal reference. The use of 介个, 类个, and 乃们 specify their referents from the context and connect the text to form a unified message.

The following example talks about "that thing" which is helpful at work, but there is no context showing the referential meaning. This message displays exophoric and demonstrative reference.

[4:3] (LWC, Message ID 3402333038113555)

1.如果 我 上班 有 **嘞个** 东东 的话·· 不 晓得 好 给力 哟··

ru2guo3 wo3 shang4ban1 you3 **lei4ge4** dong1dong de1hua4, bu4 xiao3de2 hao3 gei2li4 yo1.

if 1stSG work have **that** thing if, NEG know very great PRT.

'If I had **that** thing at work, I can't tell [you] how great it would be.'

2.可以 保暖 还 能 防 噪音· 一箭双雕 哈·

ke2yi2 bao3nuan3 hai2 neng2 fang2 zao4yin1. Yi2jian4shuang1diao1 ha1.

can keep warm and can prevent noise. One arrow two birds PRT.

'[It] could keep me warm and prevent noise. [It would] kill two birds with one stone.'

嘞个(*le*1*ge*4, 嘞'PRT', 个 'CL', 嘞个 'that') is a homophone of 那个(*na*4*ge*4 'that'). The referential meaning of 嘞个 has to be found outside of the text.[35] 嘞个 servers as exophoric and demonstrative reference. In this case, by reading one piece of the message, it is unlikely to retrieve the specific information of what 嘞个 refers to. More information is needed from the context of situation.

In summary, reference includes personal reference, demonstrative reference, and comparative reference. It can also be classified into exophora and endophora. In an exophoric reference, the referent is in the context of situation. In an endophoric reference, the referent can be either found in the preceding text (anaphora) or the following text (cataphora). In the data, the majority of the Chinese Internet homophones display personal, demonstrative, anaphoric, and exophoric references. Chinese Internet homophones can serve as cohesive devices, linking the referential meaning of the reference to the referent, and create cohesion within discourse.

4.2.2 Ellipsis

Ellipsis is "something left unsaid", but "understood nevertheless"(Halliday & Hasan, 1976, p.142). "Where there is ellipsis, there is a presupposition, in the structure, that something is to be supplied, or 'understood' " (Halliday & Hasan, 1976, p.144). This phenomenon is very common in Chinese. Ellipsis is characterized as having syntactic recoverability, that there is a corresponding completed item which presents the unsaid information. The elliptical item and the corresponding item are semantically the same. The elliptical item leaves a specific structural slot

[35] For this message, there is no clear context showing the exact referent.

to be filled by the corresponding item (Halliday & Hasan, 1976, p. 143). Moreover, this syntactic recoverability and semantic consistency can be achieved through a certain pragmatic environment. Generally, ellipsis in Chinese is more widely used than that in English (Zuo, 1995, p.41).

Ellipsis, by omitting some components in discourse, reduces redundancy and still represents the semantic information of the elliptical items. It is a grammatical method to make discourse compact and cohesive. There are three types of ellipsis, namely nominal ellipsis, verbal ellipsis, and clausal ellipsis. The elliptical item is consistent with the presupposition, and they are in the same grammatical class. Usually, the elliptical item is easy to recover from the previous text. In Chinese, the subject is commonly omitted (Zuo, 1995, p.41).

The following three examples are all from the LWC. Example [4:4], shown below, shows a micro-blog user's feelings about their travel in Australia ending soon. It presents nominal ellipsis, or specifically subject ellipsis.

[4:4][36] (LWC, Message ID 3399871598226546)

[36] This is an excerpt of the original message.

1. **偶** 的 墨尔本 之旅 很 快 就 结束 了 哇。

ou3 de1 Mo3er3ben3 zhi1lv3 hen3kuai4 jiu4 jie2su4 le1 wa1.

1st SG POSS Melbourne ASSO trip very quick just end CRS PRT.

'My trip to Melbourne will end soon.'

2. _____ 舍 不 得 舍 不 得。

_____she3 bu4 de2 she3 bu4 de2.

abandon NEG CSC abandon NEG CSC.

'[**I**] am very reluctant to go.'

偶 (*ou*3 'even', it means 'I' here) is the homophone of 我 (*wo*3 'I'). As the subject, 偶 is omitted in the second line. The unsaid subject in Line 2 can be recovered from "偶" in the preceding text. The homophone 偶 serves as the source to supply the missing information.

In example [4:5], a blogger "cold jest" talks about a little girl who was eating noodles. It contains nominal (subject) ellipsis.

[4:5] (LWC, Message ID 3399769395789797)

1. #冷笑话#　　对面　　的　　**小 MM**　　吃　面　　真　　浪费。

leng3 xiao4hua4: dui3mian4　de1　**xiao3 mei3mei2** chi1 mian4 zhen2 lang4 fei4.

cold jest:　　opposite　　ASSC　**little girl**　　eat　noodle　very　wasteful.

'A cold jest: **the little girl** sitting opposite to me really wasted her noodles.'

2. _____面条　　扒拉　　两口　　就　　走人。

____mian4tiao2　ba1la1　liang3kou3 jiu4　zou3 ren2.

noodles　stir　　two CL　then　　left.

'[**The little girl**] only had a few bites and then left.'

3.于是 我　正义　的 把　那　碗　　面　　倒给　　路边　　看 起来　很 饿 的　野猫。

yu2shi4 wo3 zheng4yi4 de1 ba3 na4 wan3 mian4 dao4gei3 lu4bian1 kan3qi3lai3 hen3e4 de1 ye3mao1.

so 1st SG　righteous ADV BA　that　CL　noodle　pour　street　look like very hungry ASSC stray cat.

'So righteously, I poured the bowl of noodles on the side of the road for the hungry-looking stray cat.

4. 一会儿，**小 MM**　　回来　了，　手里　　拿着　一　瓶　水。

yi2 hui4r,　**xiao3 mei3mei2** hui2lai2 le1,　shou3li3　na2zhe　yi4　ping2 shui3.

soon,　　**little girl**　　come back CRS,　hand　　hold　　a　　CL　water.

'After a little while, **the little girl** came back with a bottle of water in her hand.'

5. _____一 脸　　狐疑　　的　　看着　　那 个　　空碗…

_____ yi4 lian3 hu2yi2　de1　kan4zhe　na4ge4　kong1wan3.

a face suspicious　ADV　look　　that　CL　empty bowl.

'[**The little girl**] looked at the empty bowl with a suspicious face.'

6.那 一刻，　　我　　只想　　当 个　　埋头　　吃面　　的　　路人甲…

na4 yi2ke4, wo3 zhi3xiang3 dang1 ge4 mai2tou2 chi1mian4 de1 lu4ren2jia3.

that moment, 1stSG only want be CL face down eat noodle ASSC passer-by.

'At that moment, I just wanted to be [a] passer-by eating noodles with my face down.'

MM (*mei*3*mei*2 'girl') is a homophone of 美眉 (*mei*3*mei*2 'girl/beautiful girl'). 小 *MM* (*xiao*3 *mei*3*mei*2 'little girl') means 'a little girl' here. As the two empty structural slots in Line 2 and Line 5 shown above, there are two places where 小 *MM* is omitted. Subject ellipsis is also a type of nominal ellipsis. Subject ellipsis is very commonly used in Chinese to make discourse compact and cohesive.

In example [4:6] the mirco-blog user talks about writing micro-blogs to his friend "Le-Shu-Yan". It presents nominal ellipsis after numbers and a classifier (CL) *tiao*.

[4:6] (LWC, Message ID 3399682279759812)

1. 对@Le-淑-Yan 说:

dui4 Le4 Shu1yan shuo1:

to Le Shuyan speak:

'[I] speak to Le Shuyan:'

2. 第 99 条 **围脖** 给 你,

di4 99 tiao2 **wei2bo2** gei3 ni3,

99th CL **Weibo** for 2nd SG,

'the 99th **Weibo** is for you.'

3. 以后 第 199. 299. 399. 499. 599. 699. 799. 899. 999 条_____都 给你

yi3hou4 di4 199 299 399 499 599 699 799 899 999 tiao_____dou1 gei3 ni3.

in the future 199th, 299 th, 399 th, 499 th, 599 th, 699 th, 799 th, 899 th, 999th, CL all for 2nd SG.

'In the future the 199 th, 299 th, 399 th, 499 th, 599 th, 699 th, 799 th, 899 th, 999th [**Weibo**] are all for you'

This micro-blog user posted a *Weibo* (micro-blog) message to his friend Le Shuyan to show his affection toward her.[37] 围脖 (*wei*2*bo*2 'scarf', here means 'micro-blog') is a homophone of 微博 (*wei*1*bo*2 'micro-blog'). In Line 2, 第 99 条围脖 (*di*4 *jiu*3*shi*2*jiu*4 *tiao*2 *wei*2*bo*2 'the 99th CL Weibo message') appears in the pattern of "number + CL+noun". In Line 3, the clause 第 199, 299, 399, 499, 599, 699, 799, 899, 999 条(*di*4 *yi*4*bai*3*jiu*3*shi*2*jiu*3, *er*4*bai*3*jiu*3*shi*2*jiu*3, *san*1*bai*3*jiu*3*shi*2*jiu*3, *si*4*bai*3*jiu*3*shi*2*jiu*3, *wu*3*bai*3*jiu*3*shi*2*jiu*3, *liu*4*bai*3*jiu*3*shi*2*jiu*3, *qi*1*bai*3*jiu*3*shi*2*jiu*3, *ba*1*bai*3*jiu*3*shi*2*jiu*3, *jiu*3*bai*3*jiu*3*shi*2*jiu*3 *tiao*2 'the 199th, 299th, 399th, 499th, 599th, 699th, 799th, 899th, 999th CL') follows the same pattern. As the noun 围脖 has been mentioned in Line 2 'the 99th CL Weibo message', it is omitted at Line 3, but the "unsaid" information is easy to be retrieved.

[37] The numeric letter 9 in Chinese has the meaning of 'forever'. From the words 99, 199, 299…in this message, it reveals that this micro-blog user wishes the relationship between him and his friend Le Shuyan will last forever.

Ellipsis includes nominal ellipsis, verbal ellipsis, and clausal ellipsis. In this section, I focused on nominal ellipsis, as it is the only ellipsis type found in my data. I have also found that most of the nominal elliptical items are subjects, which is consistent with Zuo (1995)'s conclusion that subjects are more commonly omitted in Chinese. These examples show that the elliptical items can be retrieved from the preceding homophones. Homophones serve as the source of elliptical items and contribute to cohesion within discourse.

4.2.3 Substitution

Substitution means to use one item to replace another. The meaning of the substitution item is drawn from the item that has been replaced. Substitution includes nominal substitution, verbal substitution, and clausal substitution. Hu et al. (2005) argued that it is controversial whether there are items in Chinese that are similar to those in English, which can substitute nouns, verbs, or clauses. For nominal substitution in English, if the same noun appears again, it will be substituted by a related syntactic item, such as *one*. For instance, in the example "This **shirt** is too dark. Would you please find me a yellow **one**?", *one* substitutes *shirt*. Usually, in Chinese, 的 (*de*1 'nominalizer') can be used as a nominalizer. In the pattern "noun/personal pronoun/adjective + (的)+ noun", 的 can be used to substitute the noun after 的. "Noun/personal pronoun/adjective +的" is equivalent to a noun phrase in the "*de* (的) structure".

To give an example of the "*de* (的) structure":

[4:7]

这 件 **衬衫** 颜色 太 深 了。能 给 我 找 一件 黄 **的** 吗?

zhe4 jian4 **chen4shan1** yan2se4 tai4 shen1 le1. neng2 gei3 wo3 zhao3 yi2 jian4 huang2 **de** ma?

this CL **shirt** color too deep CRS. can for 1stSG find one CL yellow NOM Q?

'The color of this **shirt** is too dark. Would you find me a yellow **one**?'

衬衫 (*chen*4*shan*1 'shirt') is mentioned in the first sentence. In the second one, after the adjective word 黄 (*huang*2 'yellow'), 的 substitutes 衬衫 from the proceeding text. 的 acts as a cohesive device through nominal substitution.

Substitution is not very common in the data and the only type of usage of homophones as substitution is 滴. As in Examples [4:8], the homophone 滴 serves as a nominal substitution device.

[4:8] (LWC, Message ID 3399828346705296)

1. **Chanel** 最新 紫色，银联 的 timeless classic flap 喔， 喜欢 的 联系 吧，

Chanel zui4xin2 zi3se4, yin2lian2 de1 timeless classfic flap wo1, xi2huan1 de1 lian2xi4 ba1;

Chanel newest purple, Union Pay ASSC timeless classic flap PRT, like NOM contact PRT;

'**A Chanel [bag]**, in the latest purple color [comes with a] timeless classic flap from Union Pay;

Please contact [us] if you like it;'

2. 有 喜欢 银色 **滴**___也 可以 联系 喔

you2 xi3huan1 yin2se4 **di1**___ye3 ke3 yi3 lian2xi4 wo1.

has like silver **NOM** also can contact PRT.

"those who like silver **ones** can also contact [us]."

滴(*di*1 'drop', here it is used as a nominalizer) is a homophone of 的(*de*1 'nominalizer'). After 银色 (*yin*2*se*4 'silver'), 滴 replaces *Chanel* (bag). The homophone 滴 serves as a cohesive device, linking the two clauses.

In English, if the same verb is used again, but replaced with a related syntactic item, such as *do* in the text, it is verbal substitution. In Chinese, certain items, such as 是(*shi*4 'so') and 做 (*zuo*4 'do'), can substitute verbs. However, there is no such homophone found in the data. Clausal substitution refers to the phenomenon that a repeating clause is substituted by a related syntactic item, such as *so*. In Chinese, there is no equivalent usage of *so* as in English. Only in certain cases, items such as 是(*shi*4 'so') can substitute a clause. As there is no such usage of verbal substitution or clausal substitution of Chinese Internet homophones found in the data, no specific examples will be given in this thesis.

To summarize, substitution includes nominal substitution, verbal substitution, and clausal substitution. Nominal substitution by using 的 as a nominalizer is commonly found in Chinese. The homophone 滴 can occur as a substitution item making clauses linked to each other.

4.2.4 Conjunction

"Conjunctive elements are cohesive not in themselves but indirectly by virtue of their specific meanings," and "they express certain meanings which presuppose the presence of other components in the discourse" (Halliday & Hasan, 1976, p.226). Usually, conjunctive items are transitional words, indicating logical relations between different clauses. These logical relations include temporal, causal, conditional, and so on. Conjunctive items include conjunctions, correlative adverbs, and other types of words playing the role of conjunction. In discourse, sometimes there is no specific item that indicates a specific logical relation. This type of conjunction is called an implicit conjunction, or parataxis conjunction. All the examples of conjunction in this section are explicit.

An example [4:9] from Hu et al. (2005) showing conjunction that indicates a progressive relation is provided below.

[4: 9] (Hu et al., 2005, p.194)

1. "……我　从没　见过　他　办过　一 件　事　要　花　半天　工夫!

…wo3 cong2 mei2 jian4guo4 ta1 ban3guo4 yi2 jian4 shi4 yao4 hua1 ban4tian1 gong1fu!

…1st SG even NEG seen 3rd SG deal one CL thing would take half day effort!

'… I have never seen him take half a day to get anything done,

2. **何况**　是　那么 一点　小事,

he2kuang4 shi4 na4me yi4dian3 xiao3shi4,

moreover be that a bit little thing,

‘[and] **moreover**, for such a minor issue.’

3. 他 只要 眉头 一 皱， 办法 就 全 有 了！……”

ta1 zhi3yao4 mei2tou2 yi2 zhou4, ban4fa3 jiu4 quan2 you3 le! …

3rdSG only eyebrow frown, solution then all have CRS! …

‘Only if he frowns, solutions will arrive! ...’

The example tells that the man can solve problems very quickly. 何况 (*he2kuang4* ‘moreover’) connects 那么一点小事 (*na3me yi4dian3 xiao3shi4* ‘such a minor issue’) and add the information that he can come up with a solution in a second. The conjunctive item 何况 expresses a progressive relation.

The following example is drawn from the LWC. There were not many examples found of this type. There is only one homophone 可素 showing a transitive relation in [4:10].

[4:10] (LWC, Message ID 3402332436848113)

1. 腊月 二十三 啊！ 小年 啊！

la4yue4 er4shi2san1 a1! xiao3nian2 a1!

lunar year’s December 23rd PRT! Spring Festival Eve PRT!

‘December 23rd [is] the Chinese Lunar New Year! Spring Festival Eve!’

2. **可素** 我们 不 放假！！！ 介[38] 都 是 神马[39] 世道 啊！！！

ke2su4 wo3men bu2 fang4jia4! jie4 dou1 shi4 shen2ma3 shi4dao4 a1!

but 1st PL NEG have holiday! this all be what world PRT!

'**But** we don't have a holiday! What kind of a world is this!'

可素(*ke*3*su*4, 可'but', 素 'plain', 可素 means 'but' here) is a homophone of 可是(*ke*3*shi*4 'but'). It links the message with an explicit transition. 可素 serves as a cohesive device connecting its preceding and following texts.

In this section, through the temporal, causal, conditional and other relations, Chinese Internet homophones connect elements in a micro-blogging message and serve as cohesive devices. But the usage of Chinese Internet homophones as conjunctions is not very common compared to other types of grammatical cohesion in the data.

4.3 Lexical Cohesion

Lexical cohesion is the effect achieved through the choice of lexical items and operates within the lexis (Halliday & Matthiessen, 2004, p. 535). The following sections will analyze lexical cohesion of Chinese Internet homophones, namely from repetition, synonymy/antonymy, hyponymy/meronymy, and collocation.

[38] 介(*jie*4 'interpose'), the homophone of 这 *zhe*4 'this', means 'this'.
[39] 神马(*shen*2*ma*3, 神 'magic', 马 'horse'), a homophone of 什么 *shen*3*me*1 'what', means 'what'.

4.3.1 Repetition

Repetition refers to repeating a lexical item that has appeared in preceding text. Each repeated item contains substantially the same ideational meaning[40]. Repetition is the simplest way to achieve lexical cohesion. Repeated lexical items include simple repetition and inflectional repetition (e.g., *eat*, *eating*, *ate*, and *eaten*; *apple* and *apples*). Chinese lacks morphological changes and therefore inflectional repetition is very rare. These repeated words are usually the keywords in a sentence or discourse.

Lexical repetition plays a very important role within discourse. It functions as a method of cohesion and makes information exchange more efficiently. When it comes to communication, there cannot always be new information, and this sometimes necessitates repetition. Additionally, from a rhetorical consideration, repetition keeps the meaning consistent, and it will also attract readers' attention. Therefore, readers will remember repeated content more easily and be more strongly influenced by the writing (Liao, 2004, p. 117).

The following three examples are all from the LWC. They display cohesion by repetition of the homophones.

Example [4:11] talks about a creepy man by repeating the homophone *ws* three times.

[4:11] (LWC, Message ID 3401488479999787)

[40] For ideational meaning, please see Chapter 2.

1. **ws** 男 看 **ws** 的 图片 竟然 要 **ws** 大学 男 解释，

wei3suo3 nan2 kan4 **wei3suo3** de1 tu2pian4 jing4ran2 yao4 **wei3suo3** da4xue2 nan2 jie3shi4,

creepy man look **creepy** ASSC picture unexpectedly want **creepy** university man explain,

'The **creepy** man looked at the **creepy** picture and unexpectedly, he wanted the **creepy** university man to explain it,'

2. 然后 **ws** 的 笑 了.

ran2 hou4 **wei3suo3** de1 xiao4 le1.

then **creepy** ADV smile PFV.

'and then [he] laughed **creepily**.'

3.可恨 的 是 旁边 还有 几 个 特 清纯 的 妹纸[41]

ke3hen4 de1 shi4 pang2bian1 hai2 you3 ji3 ge4 te4 qing1chun2 de1 mei4 zhi3.

hateful MON be besides and have several CL extremely innocent ASSC girls.

'The terrible thing is that there were several extremely innocent girls beside [him].'

Ws (*wei*3*suo*3 'creepy') is a homophone of 猥琐 (*wei*3*suo*3 'creepy'). The word *ws* appears four times in this message and shows the man is very creepy. The cohesive device *ws* makes the theme clear and the discourse cohesive.

Example [4:12] talks about the party and the government in China. It shows repetition of the homophones *d* and *ZF*.

[4:12] (LWC, Message ID 3403229884397184)

[41]妹纸(*mei*4 *zhi*3,妹'sister',纸'paper'), a homophone of 妹子 *mei*4*zi*3 'sister', means 'sister'.

1. 以前 一直 呼吁 希望 **d** 和 **ZF** 为 老百姓 做 点 什么，

yi3qian2 yi4zhi2 hu1yu4 xi1wang4 **dang3** he2 **zheng4fu3** wei4 lao3bai3xing4 zuo4 dian3 shen3me,

before always advocate hope **party** and **government** for civilians do a bit what,

'Before, [I] always hoped and advocated that the **party** and the **government** would do something [good] for the people.'

2. 现在 后悔 了，

xian4zai4 hou4hui3 le1.

now regret PFV.

'Now [I] have regretted [it].'

3. 只要 **d** 和 **ZF** 别 再 对 老百姓 做 什么 了

zhi3yao4 **dang3** he3 **zheng4fu3** bie2 zai4 dui4 lao3bai3xing4 zuo4 shen3me le,

only **party** and **government** NEG again toward civilians do what CRS,

"If only the **party** and the **government** wouldn't do anything more to the people,'

4. 就 谢天谢地 了

jiu4 xie4tian1xie4di4 le1.

then thank goodness CRS.

'I would be grateful.'

In this message, *d* (*dang*3 'party') is a homophone of 党 (*dang*3 'party') and *ZF* (*zheng*4*fu*3 'government') is a homophone of 政府(*zheng*4*fu*3 'government'). Because of Internet censorship in China, the actual words of 党 and 政府 sometimes cannot be posted online. As a result,

Internet users use *d* and *ZF* as replacement words. Through the repetition of *d* and *ZF* makes the message lexically cohesive.

Example [4:13] is a New Year greeting message, showing repetition of the homophone 兔 in several idioms and allusions, indicating the Year of the Rabbit.

[4:13] (LWC: Message ID 3404670880432789)

1. 姜 你 军[42]，蒜 你 狠[43]， 我 **兔** 故 纳 新 算 得 准。

jiang2 ni3 jun1, suan4 ni3 hen3, wo2 **tu4** gu4 na4 xin1 suan4 de2 zhun3.

ginger 2ndSG general , onion 2ndSG ruthless, 1stSG **rabbit** old take new calculate CSC accurate.

'The prices of ginger and onions are killing me. "**Out** with the old, in with the new", and I plan everything well.'

2. 糖 高宗[44]， 豆 你 玩[45]，我 大 展 宏 **兔** 铁 臂 腕！

tang2 gao1zong1, dou4 ni3 wan2, wo3 da4 zhan3 hong2 **tu4** tie2 bi4 wan3!

sugar Gaozong, bean 2ndSG play, 1stSG big spread big **rabbit** steel arm wrist!

'The prices of sugar and beans are crazy; I [can] **succeed** **[**if] I am strong!'

3. 谁 说 肉 的 理想 只 有 白菜 的 命？

shui2 shuo1 rou4 de1 li3xiang2 zhi3 you2 bai2cai4 de1 ming4?

who say meat ASSC dream only have cabbage ASSC destiny?

---

[42]姜你军 (*jiang*1 *ni*3 *jun*1 'ginger your general'), a homophonic phrase of 将你军(*jiang*1 *ni*3 *jun*1 'to kill your General (a term from the Chinese chase game)'), means that the price of ginger is killing people.

[43]蒜你狠 (*suan*4 *ni*3 *hen*3 'union you ruthless'), a homophonic phrase of 算你狠 (*suan*4 *ni*3 *hen*3 'you are malicious'), means that the price of unions is killing people.

[44]糖高宗(*tang*2 *gao*1*zong*1 'sugar Gaozong'), a homophonic phrase of 唐高宗 (*tang*2 *gao*1*zong*1 'Emperor Gaozong of the Tang Dynasty'), means that the price of sugar is very high.

[45]豆你玩(*dou*4 *ni*3 *wan*2 'bean you play'), a homophonic phrase of 逗你玩(*dou*4 *ni*3 *wan*2 'it is a joke'), means that the price of beans is crazy.

‘Who says dreams can’t come true?’

4. 只要 “**兔兔**” 是 道[46]，就 能 扬 眉 “**兔**” 气，

zhi3yao4 “**tu4 tu4”** shi4 dao4, jiu4 neng2 yang2 mei2 **tu4** qi4.

only if “**rabbit rabbit**” be Dao, then can raise eyebrow **rabbit** air.

‘Only by finding the Dao **everywhere**, can you become proud and **elated**.’

5. 恭祝： 龙 年 吉 祥，

gong1zhu4: long2 nian2 ji2 xiang2.

sincerely wish: dragon year lucky.

‘[I] sincerely wish [you] a good luck in the Year of the Dragon.’

6. 不 求 腾 云 驾雾， 只 求 赛过 金龙！

bu4 qiu2 teng2 yun2 jia4 wu4, zhi3 qiu2 sai4 guo4 jin1long2!

NEG hope mount cloud ride mist, only hope surpass golden dragon!

‘Do not try to mount the clouds and ride the mist [like a god], but I hope that you can surpass the golden dragon [and become wealthy]!’

This New Year greeting message uses 兔 (*tu*4 ‘rabbit’) as a homophone to replace the words in the idioms. 兔故纳新(*tu*4 *gu*4 *na*4 *xin*1, 兔 ‘rabbit’, 故 ‘old’, 纳 ‘take’, 新 ‘new’) a homophone of 吐故纳新(*tu*3 *gu*4 *na*4 *xin*1 吐 ‘exhale’, 故 ‘old’, 纳 ‘take in’, 新 ‘new’, 吐故纳新 ‘out with the old, in with the new’), means ‘out with the old, in with the new’. 大展宏兔 (*da*4 *zhan*3 *hong*2 *tu*4, 大‘big’, 展 ‘spread’, 宏 ‘big’, 兔‘rabbit’), a homophone of 大展宏图 (*da*4 *zhan*3 *hong*2 *tu*2, 大 ‘big’, 展 ‘spread’, 宏 ‘big’, 图 ‘picture’, 大展宏图‘to succeed’), means ‘to

[46]道 (*dao*4 ‘Dao’) here refers to 道场 (*dao*4*chang*3 ‘Bodhimanda’), a term from Buddhism, meaning the “position of awakening”.

succeed'. 兔兔是道(*tu*4 *tu*4 *shi*4 *dao*4, 兔 'rabbit',兔 'rabbit', 是 'be', 道 'Dao'), a homophone of 处处是道(*chu*4 *chu*4 *shi*4 *dao*4, 处处 'everywhere', 是 'is', 道 'Dao', 处处是道 'Dao is everywhere'), means 'Dao is everywhere'. 扬眉兔气(*yang*2 *mei*2 *tu*4 *qi*4, 扬 'raise', 眉 'eyebrow', 兔 'rabbit', 气 'air') a homophone of 扬眉吐气(*yang*2 *mei*4 *tu*3 *qi*4, 扬 'raise', 眉 'eyebrow', 吐 'exhale', 气 'air', 扬眉吐气 'to feel proud and elated') means ' to feel proud and elated'. Through repetition of 兔 four times， this message is cohesively linked. Thus the message is clear and prominent, looking back upon the Year of the Rabbit and then making wishes for the Year of the Dragon (the coming New Year). The lexical repetition is used for cohesion in discourse.

4.3.2 Synonymy/Antonymy

Synonymy and antonymy also play an important role in cohesion. Synonyms are morphologically different, but with the same or a close meaning. In discourse, synonyms can be repeated in a unit of semantic meanings and supplement the preceding content. Synonymy makes different components semantically linked to each other. In addition, from a rhetorical point of view, the use of synonyms can avoid repetition and prevent a monotonous style in the discourse. It makes the expression varied and vivid. Antonymy is established between two lexical or more items with opposite meanings. The use of antonyms generates a contrastive effect, links discourse, and achieves cohesion.

The following examples are from the LWC, showing cohesion created by synonymy or antonymy.

Example [4:14] talks about the Chinese Communist Party's policy towards Hong Kong. It shows cohesion created by synonymy.

[4:14] (LWC, Message ID 3404225311063790)[47]

1. 对于　　香港　　这个　　地方　　的　　优越感，

dui4 yu2 xiang1gang3 zhe4ge4 di4fang2 de1 you1yue4gan3,

as to Hong Kong this CL place ASSC superiority feeling,

'Regarding Hong Kong, this place's sense of superiority'

2. 完全　　是　　**中共**　　自己　　惯　　出来　的。

wan2quan2 shi4 **Zhong1gong4** zi4ji3 guan4 chu1lai2 de1.

total be **Chinese communist party** itself spoil out PRT.

'has been totally spoiled by the **Chinese Communist Party**.'

3. 78 年　以前　　是　只有　　这个　　对外　　的　　窗口，

78 nian2 yi3qian2 shi4 zhi2you3 zhe4g4 dui4wai4 de1 chuang1kou3,

78 year ago be only have this CL external ASSC window,

'Before 1978, it was the only window out.'

4.改革　后　　香港　　的　　作用　　和　　地位　　肯定　　要　　下降，

gai3ge2 hou4 Xiang1gang3 de1 zuo4yong4 he2 di4wei4 ken3ding4 yao4 xia4 jiang4,

reform later Hong Kong ACCS role and status definite would decrease,

[47] This is an excerpt of the original message.

‘after the reform, Hong Kong’s role and status will definitely decrease.’

5.但 **TG** 认为

dan4 **tian1gong4/tu3gong4** ren4wei2

but **Celestial Kingdom of the Communist Party/ Communist Bandits** think

‘But the **Celestial Kingdom of the Communist Party/ Communist Bandits** thinks that’

6. 要是 在 自己 手里 下降 会 显得 没 面子，

yao4shi4 zai4 zi4ji3 shou3li3 xia4jiang4 hui4 xian3de2 mei2 mian4zi,

if at itself hand decrease would seem NEG face,

‘if a recession happens under their supervision, they will lose face.’

7. 所以 资源 尽量 倾斜 港澳， 造就 出 这么 一 个 极端 地方。

suo3yi3 zi1yuan3 jin4liang4 qing1xie3 Gang3ao4, zao4jiu4 chu1zhe4me yi2ge4 ji2duan1 di4fang1.

so resource try to lean Hong kong & Macao, build out such a CL extreme place.

‘So they tried their best to invest lots of resources in Hong Kong and Macao to build such a special place.’

中共(*zhong*1*gong*4) refers to ‘the Chinese communist party’. *TG* (*tian*1*gong*4/*tu*3*gong*4) is a homophone of 天共 (*tian*1*gong*4 ‘Celestial Kingdom of the Communist Party’) or 土共 (*tu*3*gong*4 ‘Communist Bandits’). The micro-blog user used 中共 and *TG* as a pair of synonyms, referring to the Chinese Communist Party. 中共 and *TG* function as synonymy to connect the message.

Example [4:15] uses “tragedy” and “comedy” to talk about life, showing cohesion created by antonymy.

[4:15] (LWC, Message ID 3400563930754870)[48]

1. 生活 是 自己 的 **杯具**, 别人 眼里 的 **洗具**。

sheng1huo2 shi4 zi4ji3 de1 **bei1ju4** bie2ren2 yan3li3 de1 xi3ju4.

life be your own ASSC **tragedy,** other people eye ASSC **comedy.**

“Life is a **tragedy** to you, but a **comedy** to others.”

In the message,杯具 (*bei*1 *ju*4, 杯‘cup’, 具 ‘tool’, 杯具 ‘cup’, here it refers to ‘tragedy’), a homophone of 悲剧 (*bei*1*ju*4 ‘tragedy’), means ‘tragedy’. 洗具 (*xi*3*ju*4, 洗 ‘clean’, 具 ‘tool’, 洗具 ‘cleaning set’, here it refers to ‘comedy’), a homophone of 喜剧 (*xi*3*ju*4 ‘comedy’), means ‘comedy’. The two homophones 杯具 and 洗具 are a pair of antonyms. The antonymy created by 杯具 and 洗具 makes the text cohesive.

In example [4:16], the micro-blog user talks about finding beautiful girls on the Internet. It displays cohesion through synonymy and antonymy.

[4:16] (LWC, Message ID 3402490597102154)

1. **美眉** 几 时 有， 上机 问 网友，

mei3mei2 ji3 shi2 you3, shang4ji1 wen4 wang3you3,

[48] This is an excerpt of the original message.

beautiful eyebrow when have, use computer ask net-friend,

'When there will be **beautiful girls**? I am using a computer to ask friends online.'

2. 不知 网上 **佳丽**， 此 **妹** 是 何人？

bu4 zhi1 wang3shang4 **jia1li4,** ci3 **mei4** shi4 he2ren2?

NEG know Internet **beautiful girls,** this **girl** be who?

'I don't know [any] **beautiful girls** online. Who is this **girl**?'

3. 我 欲 下线 归 去，惟 恐 天 南 海 北， 远处 不 堪 寻，

wo3 yu4 xia4xian4 gui1qu4, wei2 kong3 tian1nan2 hai3 bei3, yuan3chu4 bu4kan1 xun2.

1st SG want off line go back, only afraid sky south sea north, far NEG can search.

'I want to be offline and go home, but [I am] afraid that I am not able to search very far or all over the country.'

4. 查找 选 芳名， 何似 在 网吧.

cha2zhao3 xuan3 fang1ming2, he4si4 zai4 wang3ba1.

search and find select girls' name, NEG compare to at Internet bar.

'In terms of finding girls, it can't compare to [being] in an Internet bar.'

5. 打 电话， 记 传呼， 单 照 付， 不 应 心 疼，

da3 dian4hua4, ji4 chuan2hu1, dan1 zhao4fu4, bu4 ying1 xin1teng2.

make phone call, write down pager, bill still pay, NEG should care.

'[I] make phone calls, write down the pager messages, pay bills, [and] shouldn't care [about money].'

5. 何时 不 想 聚 时 缘？

he2shi2 bu4 xiang3 ju4 shi2 yuan2 ?

when NEG think get together time destiny?

'When [do you] not think about destiny when [you think about] getting together?'

6. 月 有 阴 晴 圆 缺， 网 有 **恐龙** 青蛙[49]，

yue4 you3 yin1 qing2 yuan2 que2, wang3 you3 **kong3 long2** qing1wa1.

moon has cloudy sunny round unround, Internet has **dragon** frogs.

' The moon can be bright or dim, full or waning, and there are **ugly-looking girls** and boys on the Internet.'

7. 此 事 古 难 全。

ci3 shi4 gu3 nan2 quan2.

this thing ancient hardly perfect.

'There has been nothing perfect since ancient times.'

8. 但愿 好 **美眉**， 千 里[50] 共 聊 天。

dan4yuan4 hao3 **mei3mei2**, qian1 li3 gong4 liao2 tian1.

Wish good **beautiful girl**, thousand Li together chat.

'[I] only wish to have a nice **beautiful girl** chatting with me, [even if we are] thousands of miles away.'

美眉 (*mei*3*mei*2 美 'beautiful', 眉 'eyebrows', 美眉 'beautiful girl') is a homophone of 妹妹 (*mei*4 *mei*4 'sister/girl'). The synonymy of 美眉, 佳丽(*jia*1*li*4 'beautiful girl'), and 妹 (*mei*4 'sister/girl') not only shows the user's eagerness to find a beautiful girl on the Internet, but it also avoid monotonous duplication. In addition, 美眉 and 恐龙 (*kong*3*long*2 'dragon', it refers to

[49]青蛙 (*qing*1*wa*2 'frog') is used to refer ugly-looking boys by Chinese Internet users.

[50]里(*li*3 'Li') is a Chinese measure word for distance. 1 Li equals 500 Meters.

'ugly-looking girls') are antonyms, showing a contrastive effect. The use of synonymy and antonymy makes the message more cohesive.

4.3.3 Hyponymy/Meronymy

Hyponymy is the relation between specific lexical items and their superordinate. Meronyms and their superordinate have a relation of "parts and whole": A belongs to B, and A is a component of B. For example, *head*, *neck*, and *leg* are meronyms of the superordinate *body*. The relation of hyponyms and their superordinate is: the semantic field of A is included in that of B. Normally, the meaning of a superordinate is general and abstract, whereas that of its hyponyms is specific. For instance, *apple*, *pear*, and *banana* are hyponyms of the superordinate *fruit*. In discourse, cohesion can be achieved by using either hyponyms/meronyms, or hyponyms/meronyms with their superordinate.

The following example from Hu et al. (2005) shows meronymy in a Chinese novel. Cohesion is achieved through different words of human body parts.

[4:17] (Hu et al., 2005, p. 194)

1. 她 的 **脸** 色 现在 又 飞 红了, 她 的 **眼** 光 迷乱,

ta1 de1 **lian3** se4 xian4zai4 you4 fei1 hong2 le1, ta1 de1 **yan3** guang1 mi2luan4,

3rdSG POSS **face** now again fly red CRS, 3rdSG POSS **eye** sight confuse,

'Her **face** was blushing, her **eyes** were confused,'

2. 她的 **胸部** 很 剧烈 一起一伏。

ta1de1 **xiong1bu4** hen3 ju4 lie4 yi4qi3yi4fu2.

3rdSG POSS **chest** very extreme rise and fall.

'and her **chest** rose and fell heavily.'

3. 突然 她 伸开 了 **两臂**。

tu1ran2 ta1 shen1kai1 le1 **liang3 bi4**.

all of a sudden 3rdSG open PFV **two arms**.

'All of a sudden, she opened her **arms**.'

4. 雷 参谋 抢上 一 步，

Lei2 Can1mou2 qiang3shang4 yi 2bu4;

Lei Chief-of-staff take a step;

'Chief-of-staff Lei took a step;'

5. 吴 少奶奶 便 像 醉 迷 似的 扑在 雷参谋 **胸** 前，

Wu2 Shao4nai3nai3 bian4 xiang4 zui4mi2 si4de pu1zai4 Lei2 Can1mou2 **xiong1** *qian2*,

Wu lady then seem drunk like fall on at Lei Chief-of-staff **chest** front,

'Lady Wu fell on the **chest** of Chief-of-staff Lei like a drunken person'

6. 她 的 **脸** 恰 靠 在 雷 参谋 **肩**头。

ta1 de1 **lian3** qia4 kao4 zai4 Lei2 Can1mou2 **jian1** tou2.

3rdSG POSS **face** just rest on Lei Chief-of-staff **shoulder**.

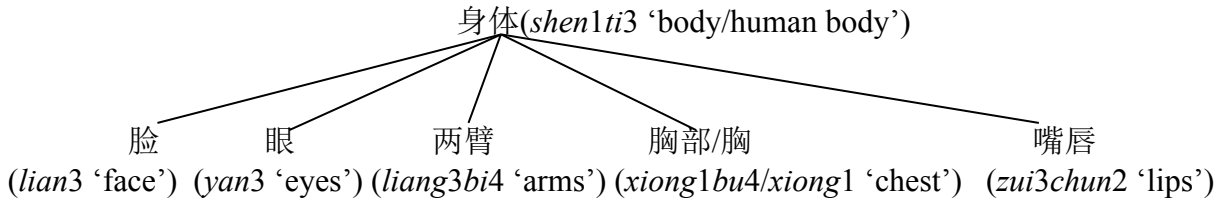
'with her **face** just resting on Chief-of-staff Lei's **shoulder**.'

7. 雷 参谋 俯下 头去， 两 个 **嘴唇** 接在 一起。

Lei2 Can2mou2 fu3xia4 tou2qu4, liang3ge4 **zui3chun2** jie1zai4 yi4qi3.

Lei Chief-of-staff lower head, tow CL **lips** connect together.

‘Chief-of-staff Lei lowered his head, and then the two [people’s] **lips** came together.’



As shown above, 脸 (*lian*3 ‘face’), 眼 (*yan*3 ‘eyes’), 两臂 (*liang*3*bi*4 ‘arms’), 胸部/胸 (*xiong*1*bu*4/*xiong*1 ‘chest’), and 嘴唇 (*zui*3*chun*2 ‘lips’) are part of the human body. Cohesion in this message is achieved by using five meronyms at the same time, but no superordinate. Through using different parts of the human body, the meronyms make the description of Chief-of-staff Lei and Lady Wu’s actions cohesive.

In the data, for this type of cohesion, the cases only show cohesion created by hyponymy. The following three examples show cohesion through hyponymy.

In example [4:18], the micro-blog user talks about the gender imbalance of people working in the government.

[4:18] (LWC, Message ID 3399982109657868)

1. 网友 建议 想 当官 的 **男淫** 做 变性 手术，

wang3you3 jian4yi4 xiang3 dang1guan1 de1 **nan2yin2** zuo4 bian4xing4 shou3shu4.

Internet friend suggest want be leaders ASSC **men** take trans-gender surgery.

‘An Internet user suggested that **men** who want to be leaders should have transgender surgery.’

2. 这 是 不错 的 主意。

zhe4 shi4 bu2cuo4 de1 zhu3yi1.

this be good ASSC idea.

‘This is a really good idea.’

3. 若干 年 后，

ruo4 gan2 nian2 hou4,

several year after,

‘After several years,’

4. 当 主席台 上 的 女 领导 坐 满 了 的 时候，

dang1 zhu3xi2tai2 shang4 de1 nv3 ling3dao3 zuo4 man3 le1 de1 shi4hou4,

when platform on ASSC female leader sit full PFV ASSC time,

‘when the platform is full of female leaders,’

5. 当官 的 机会 基本上 都 是 **男淫** 的 了，

dang1guan1 de1 ji1hui4 ji1ben3shang4 dou1 shi4 **nan3yin2** de1 le1.

be leader ACCS opportunity basical all be **men** POSS CRS.

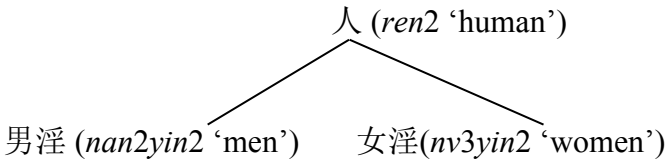
‘**men** would have the most opportunities to become leaders.’

6. 那 时候 又 轮到 **女淫** 做 变性 手术 了。

na4 shi3hou4 you4 lun2dao4 **nv3yin2** zuo4 bian4xing4 shou3shou4 le1.

that time again turn **women** take trans-gender surgery CRS.

‘Then it would be **women**’s turn to have transgender surgery.’



In this example, 男淫 (*nan*2*yin*2, 男 ‘male’, 淫 ‘obscene’, 男淫 ‘men’) is a homophone of 男人 (*nan*2ren2 ‘men’). 女淫(*nv*3*yin*2, 女 ‘female’, 淫 ‘obscene’, 女淫 ‘women’) is a homophone of 女人 (*nv*3*ren*2 ‘women’). They are both hyponyms of 人 (*ren*2 ‘human’). The two Chinese Internet homophones 男淫 and 女淫 link the theme to gender imbalance. Cohesion within this message is achieved through the two hyponyms.

In example [4:19], the user talks about the change of their expectation on the baby’s gender. The use of hyponyms makes the information of the gender change cohesive.

[4:19] (LWC, Message ID 3402514318079658)

1. 从 今天 起,

cong2 jin1tian1 qi3,

from today on,

‘From now on,’

2. 我 要 把 迎接 DD 的 心态 调试 成 迎接 MM 的，

wo3 yao4 ba3 ying2jie1 **di4di** de1 xin1tai4 tiao2shi4 cheng2 ying2jie1 **mei4mei** de1,

1st SG want BA welcome **boy** ASSC mentality change into welcome **girl** NOM,

‘I have to change my mentality of welcoming [a] **boy** to welcoming [a] **girl**.’

3. 之前　　的　　心理　　暗示　　式　　强大　　了，

zhi1qian3　de1　xin1li3　　an4shi4　　tui1 qiang2da4　le,

before　　ASSC　mental　　suggestion　too　　strong　CRS,

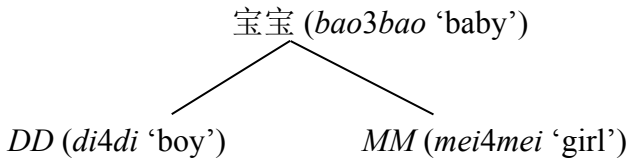
‘[As] the previous mental suggestion is too strong,’

4. 要　　一下　改　　过来，　　不　　容易　　啊

yao4　　yi1xia4 gai3 guo4lai2,　bu4 rong2yi4　　a1.

want　　once　change　　back,　NEG　　easy　　PRT.

‘it is not easy to change [my mentality] quickly.’



As shown above, *DD* (*di*4*di* ‘boy’) is a homophone of 弟弟 (*di*4*di* ‘brother/young boy’) and *MM* (*mei*4*mei* ‘girl’) is a homophone of 妹妹 (*mei*4*mei* ‘sister/young girl’). They are both hyponyms of 宝宝 (*bao*3*bao* ‘baby’). The two hyponyms *DD* and *MM* serve as lexical cohesive devices to express the user’s expectation of the baby’s gender has changed.

Example [4:20] is about the scene at a car show. Cohesion in this message is achieved through the superordinate and its hyponyms.

[4:20] (LWC, Message ID 3401766368879998)[51]

1. 不　来　一 发　吗　**骚年**！！！

bu4　lai3　yi4 fa1　ma1 **sao1nian2**!

NEG have　a shot　Q　**young person**!

'Why don't [you] have a shot, y**oung person**!'

2. 我　想到　了　飞 天夜翔[52]　那 句　名言，　人　瘦　哔-----大。

wo3 xiang3dao4 le1　Fei1tian1ye4xiang2 na4 ju4 ming2yan2,　ren2 shou4 bi4　da4.

1st SG　recall　PFV　Feitianyexiang　that　CL famous saying, people slim then　better.

'I recalled the famous saying of Feitianyexiang, slim people look better.'

3. 现场　**妹纸**　的　脸上　都　洋溢　着　特殊　的　笑容，

xian4chang3 **mei3zhi3**　de1 lian3shang4 dou1 yang2yi4 zhe2　te4su4　de1 xiao4 rong2.

on the spot　**girls**　POSS　face　all　fill　with　special　ASSC　smile.

'All of the g**irls** here have special smiles on their face.'

4. 拿着　单反　不　停　亮　闪光灯　的　**汉纸**　们　你们　赢　了

na2zhe2 dan1fan3 bu4 ting2　liang4　shan3guang1deng1 de1 **han4zhi3** men2 ni3men2 ying2 le.

hold　SLR[53]　NEG stop　light　flash　ASSC　**boys**　2nd PL　win PFV.

'The boys with non-stop flashing SLRs [cameras] have won.'

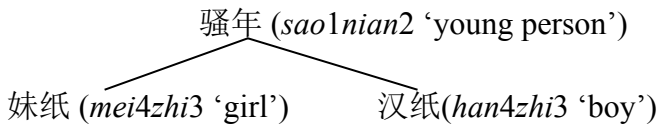
5. orz。

wo3 ren4 zai1.

I　confess fall.

'I confess.'

[51] This is an excerpt of the original message.
[52] 飞天夜翔(*Fei*1*tian*1 *Ye*4*xiang*2) is a pen name of a Chinese online novelist.
[53] SRLs is the short form of single-lens reflex.



In this message, 骚年 (*sao*1*nian*2, 骚 'coquettish', 年 'year', 骚年'young person') is a homophone of 少年 (*shao*4*nian*2 'young person'). 妹纸(*mei*4*zhi*3, 妹 'sister', 纸 'paper', 妹纸 'girl') is a homophone of 妹子 (*mei*4*zi*3 'girl'). 汉纸(*han*4*zhi*3, 汉 'sweat', 纸 'paper', 汉纸 'boy') is a homophone of 汉子 (*han*4*zi*3 'boy'). 妹纸 and 汉纸 are both hyponyms of the superordinate 骚年. The homophonic superordinate and hyponyms create lexical cohesion within the discourse and make the description of the scene cohesive.

4.3.4 Collocation

Collocation refers to the phenomenon that lexical items with a tendency to co-occur are used together in discourse. These co-occurred words are related on a semantic level. Collocation can appear in a same clause or different clauses. The lexical items that create collocation are semantically associated. When one word occurs, other associated words will appear. For instance, *doctor*, *patient,* and *hospital*, and *rain* and *umbrella* are lexical devices that help achieve cohesion within discourse.

This type of cohesion is not very common in the data. The following example shows collocation presented by a Chinese Internet homophone from the data.

[4:21] (LWC, Message ID 3399857782088538)

1. 我 参与 了 新浪 微博 沃 3G[54] 随 想 随 行 活动,

wo3 can1yu3 le1 xin1lang4 wei1bo4 Wo4 3G sui3 xiang3 sui3 xing3 huo4dong4.

1[nd]SG participate PFV Sina Weibo Wo 3G whenever think whenever act activity.

'I have participated in the Sina Weibo *Wo 3G Mind with Action*.'

2. 参与 证言人 话题 讨论, 获得 手机 大奖 抽奖 机会,

can1yu3 zheng4yan3ren2 hua4ti2 tao3lun4, huo4de2 shou3ji1 da4 jiang3 chou1jiang3 ji1 hui4.

join in witness topic discussion, get cell phone big prize draw chance.

'Join in the discussion of *witness*, and you will have the chance to win a cell phone.'

3. 坚持 每天 随身 随心 **织** **围脖** 就 有 更多 好礼 等 你 拿!

jian1chi2 mei3tian1 sui2shen1sui2xin1 **zhi1** **wei2bo2** jiu3 you3 geng4duo1 hao3li3 deng3 ni3 na2!

keep everyday carry satisfied **knit** **micro-blog** then have more good prize wait2[nd]PL take!

'Keep "**knitting"** [your] **micro-blog** everyday and everywhere follow your inclinations, [and] there will be more prizes waiting for you.'

围脖 (*wei*2*bo*2 'scarf', here it refers to micro-blog') is a homophone of 微博(*wei*1*bo*2 'Weibo, micro-blog'). 织 (*zhi*1 'knitting') in the context of this message, means to 'write or use' (micro-blogs). The cohesion of collocation is achieved by 围脖 and 织. The two words 围脖 and 织 tend to appear in the same lexical environment. The collocation of 围脖 and 织 make the theme of promoting using Sina *Weibo* cohesive.

[54]沃 3G (*wo*4 *san*3 *ji*4 'Wo 3G') is a type of 3G service provided by China Union (a Chinese telecommunication company).

In this section, through the examination of Chinese Internet homophones in repetition, synonymy/antonymy, hyponymy/meronymy, and collocation, I have shown that Chinese Internet homophones can be used as lexical cohesive devices and enhance cohesion of the micro-blogs in Chinese Internet discourse.

4.3 Summary

This chapter studied the cohesion function of Chinese Internet homophones by analyzing the micro-blog data from the LWC. I have explored the function of Chinese Internet homophones from two aspects: grammatical cohesion and lexical cohesion. When Chinese Internet homophones serve as agents of reference, ellipsis, substitution, and conjunction, they function as grammatical cohesive devices. Chinese Internet homophones can also function as lexical cohesive devices when they are used in repetition, synonymy/antonymy, hyponymy/meronymy, and collocation.

Chapter 5 Conclusions

The use of new social media, such as micro-blogging, provides a platform for the acceleration of the spread of Internet homophones. In previous chapters, I discussed the phonetic features and morphological classification of Chinese Internet homophones, as well as their discourse function. In this chapter I will summarize the findings of this thesis, and discuss the incentives for the creation and use of Chinese Internet homophones. This will be followed by a discussion of the limitations of this study and implications of the findings.

5.1 Summary of Findings

The research presented in this thesis combined a quantitative and qualitative analysis of Chinese Internet homophones in the Leiden Weibo Corpus. In the data, 117 Chinese homophones were identified. Among the top 10 most frequently used homophones in the data, the homophones are generally used rather frequently compared to their base words. Based on their phonetic feature, Chinese Internet homophones can be classified into two types: narrow-homophones which have the same pronunciation as their base words, and broad-homophones which have a similar pronunciation to their base words. In the dataset, 64 homophones (54.70%) are narrow-homophones, while 53 homophones (45.30%) are broad-homophones. Narrow-homophones can be further classified into homographic homophones (those that have the same spelling as their base words or borrow ancient written forms) and heterographic homophones (those that have a different spelling than their base words, including Chinese character heterographic homophones, alphabet heterographic homophones, and numeric letter heterographic homophones). As for

broad-homophones, this thesis examines Chinese character, alphabet, and numeric letter broad-homophones. Compared with their base words, character broad-homophones tend to have the same manner but different place of articulation as their base words. They may also display a nasal [n] in homophones corresponding to a lateral [l] in the base words, or a lateral [l] or zero-initials in homophones corresponding to a post-alveolar approximant r[ʐ] in the base words. The differences in finals of a syllable involve differing in medials, main vowels, and syllabic terminals. The medials of homophones tend to be unrounded while that of the base words are rounded. As for differences of the main vowels, the vowels of most homophones tend to be higher and unrounded (e.g., a high front unrounded medial [i] in homophones corresponds to a high front rounded medial [y] in the base words). Comparing the phoneme /e/ in the base words, a high front unrounded [i] in homophones corresponds to a mid-central unrounded vowel [ə] in the base words and a low central vowel [a] in homophones corresponds to a mid-high front unrounded vowel [e] in the base words. In addition, the phonetic differences of homophones involve the addition or deletion of vowels. As for the terminals, an alveolar nasal [n] and a velar nasal [ŋ] can be interchanged sometimes. There is no fixed pattern for the tones of homophones in the data compared with the tones of the base words. Chinese transliteration homophones undergo minor sound changes compared to their pronunciations in the original language. Most of the alphabet homophones retain the pronunciation of the Chinese characters from the base words. The alphabet homophones are usually acronyms of the Pinyin of the base words, representing the pronunciation of the base words. As for numeric letters, a single numeric letter can represent different homophonic sounds, such as sounds from Mandarin, sounds from other dialects, and sounds from other languages.

Based on their morphological features, Chinese homophones can be classified into (Chinese) character homophones, alphabet homophones, numeric letter homophones, and a blend of characters, alphabets, and numeric letters (Zhou, 2008, p. 434). Among the 117 Chinese Internet homophones identified in the data, there are 55 character homophones (47%), 46 alphabet homophones (39.32%), 4 numeric letter homophones (3.42%), and 12 blending homophones (10.26%). There are in total 75,592 occurrences of the 117 homophones in the dataset, among which character homophones occurred the most frequently (55.08%), followed by alphabet homophones (38.66%), numeric letter homophones (4.98%), and blending homophones (1.28%). Character homophones and letter homophones are used much more frequently than numeric letter homophones and blending homophones. Chinese character homophones are the most frequently used type, which is consistent with Zhou (2008a)'s finding. As for Chinese character homophones, they are classified into single-morphemic words (including monosyllabic single-morphemic words and polysyllabic single-morphemic words) and multi-morphemic words (including compounds, reduplications, and derivations). Single-morphemic words are used more frequently than multi-morphemic words. Alphabet homophones are derived from the Pinyin of each Chinese base word. Alphabet homophones are written in Pinyin, Pinyin acronym, or Pinyin with some variations. Acronym homophones are the most prominent in the category of alphabet homophones. Numeric letter homophones consist of Arabic numbers. Numeric letters either present the same or close sounds from Chinese, or the same or close sounds from other languages, such as Japanese. Blending homophones are a mix of characters, alphabets, and numeric letters. The frequency of blending homophones is the lowest among the four morphological types of homophones. Some alphabets in blending homophones have similar

pronunciations to those in English, unlike the letters in alphabet homophones, where they are pronounced based on the Chinese sounds.

The qualitative analysis of homophones is based on the theoretical framework of systemic functional grammar (Halliday & Hasan, 1976; Halliday, 1985; Halliday & Matthiessen, 2004). It mainly studies the cohesion function of homophones in Chinese Internet discourse. Cohesion is the connection within text that makes it a semantic unit (Halliday et al., 1976, p.4). In this study, Hu et al. (2005)'s classification of cohesion was adopted, which includes grammatical cohesion and lexical cohesion. Grammatical cohesion is further divided into four subcategories: reference, ellipsis, substitution, and conjunction. The subcategories of lexical cohesion include repetition, synonymy/antonymy, hyponymy/meronymy, and collocation. From the excerpts drawn from the Leiden Weibo Corpus, homophones can function as grammatical and lexical cohesive devices in Internet discourse. They make each message form a unified text, rather than a disconnected sequence of clauses. However, the homophones that are used as conjunction (a subcategory of grammatical cohesion), and collocation (a subcategory of lexical cohesion) devices are not very common in the data.

5.2 Discussion

The emergence of Chinese Internet homophones is inseparable from their own characteristics and the linguistic, social, and cultural environment in which they are used. The following section will briefly discuss the reasons for the creation of Chinese homophones. They are phonology of Mandarin Chinese, the use of typing method for Mandarin, the influence of Chinese dialects, and

the Internet censorship of the Chinese government. These factors facilitate the creation and use of Chinese homophones in computer-mediated communication.

5.2.1 Phonology of Mandarin Chinese

Mandarin Chinese is a tonal language, with 21 initials, 39 finals, and 4 basic tones in its phonological system. Given the small set of initials and finals, the number of syllables that can be formed is limited. According to the Mandarin phonological rules and conventions, there can be only 398 sets of initial-final combinations, and approximately 1,200 distinct tonal syllabic sets (Suen, 1979). However, there are 4,574 Chinese characters in common usage (*Modern Chinese Frequency Dictionary*, 1986). As a result, there exists a great amount of homophonic words with the same or close sound as their base words.

5.2.2 Typing Method

Many people believe that Chinese, with its character-based orthography, could not serve for electronic use. But by 2006, it had become the second most commonly used language on the web (Gottlieb, 2009). Most Chinese Internet users rely on the Pinyin typing method. First, they type out the Pinyin of a word through the keyboard, and then they select from a pop-up word list that contains words with the same initial and final. Finally, they choose the desired word, which is displayed on the screen. Because the words in the typing method system are arranged based on the frequency of common lexicon. In fast-paced Internet communication, Internet users may choose to use homophonic words and the most concise script to express their intended meaning

and to speed up the typing and the exchange of information. Some Internet users use homophonic puns, disregarding the accuracy of their typing and the language norm. This is because, for some words, their homophonic puns are faster and more convenient to type. For example, if Internet users want to type the word 版主 *ban*3*zhu3* ‘board moderator’ by Pinyin, the word 斑竹 *ban*1*zhu*2 ‘mottled bamboo’ comes first on the list. To save time, more and more Internet users use 斑竹 instead of 版主, and gradually 斑竹 has become a homophonic pun of 版主. This is also because that the Internet word 版主 ranked relative low in the frequency of the common lexicon when this word was stared to use on the Internet.

For most alphabet homophones, their emergence is also related to the Pinyin typing method. Instead of scrolling down the word list and finding the corresponding word, Internet users press the Capital Lock key on the keyboard, and type the full pinyin form or acronym, such as *Ding ding*3 *‘to* support or to push up the ranking of a posting’ (base word 顶 *ding*3 ‘to support’) and *ZZ zhuan*3*zai*3 ‘to forward others’ posting’ (base word 转载 ‘to forward’).

5.2.3 The Influence of Chinese Dialects

In Mainland China, there are 7 main dialect groups distinguished by their phonological and phonetic differences, but they share the same writing system (i.e. using simplified Chinese characters). With the prevalence of local on-line chat rooms and local BBS, and the popularity of youth culture, various spoken words in Chinese dialects have acquired written forms, with phonetic borrowing from Chinese characters. These new words have become popular among

other Internet users across the country. For example, the use of 虾米 *xia*1*mi*3 'little shrimp' as a homophone of 什么 *shen*3*me* 'what' started from Southern Fujian dialect speakers. Later it was gradually adopted and frequently used by other young Internet users nationwide.

5.2.4 Internet Censorship by the Chinese Government

Mainland China is considered to have the most pervasive and sophisticated regime of Internet filtering and information control in the world (John, Palfrey, Jonathan, & Zittrain, 2011). Originally, the censorship of the Internet carried out by the Chinese government was to "purify" the language environment (i.e., to filter obscene and erotic words). Since the early 2000's, more and more "politically sensitive" words have been censored in order to maintain social stability. For example, 党 *dang*3 'party' and 政府 *zheng*4*fu*3 'government,' which are frequently used in comments that are regarded as threatening to social stability, have been censored. As a result, the Internet censorship restricts Internet users' online postings. As an alternative, Internet users started to use Chinese character homophones and acronym alphabet homophones in order to pass the filter and publish their words online. For example, 河蟹 *he*2*xie*4 'river crab' is used as a homophone of 和谐 *he*2*xie*3 'harmonious' (a political concept of the government) and GCD *gong*4*chan*3*dang*3 'the Communist Party' is used as a homophone of 共产党 *gong*4*chan*3*dang*3 'the Communist Party'.

In addition, other factors, such as the influence of foreign languages and multilingual youth culture, also facilitate the creation and use of Chinese homophones on the Internet.

5.3 Limitations of the Study

Internet language, known as a type of social dialect, is changing along with social development. The lexicon is one of the most sensitive parts in Internet discourse, and it is affected by technology, politics, society, culture, and the language environment. Internet users' gender, age, and educational background also play an important role in computer-mediated communication.

The present study has several limitations. The first drawback is its word list of Chinese Internet homophones. As constrained by the large size of the corpus, the initial word list was constructed from the *Green Paper on Language Situation in China* (2006-2013), Chinese Internet language dictionaries, and *Sina Weibo Niandu Pandian* (*Sina Weibo Annual Report*, 2011-2012). This may have caused certain homophones to be excluded. Second, the calculation of the total occurrences of each homophone found within the corpus is done by the software CasualConc. Because the data is too large (i.e. 18,791,530 words), only part of the data that contained target homophones was checked manually. As a result, there may be some minor errors in the frequency statistics. Finally, as this study started in the spring of 2012, the data were collected in January 2012. Therefore any homophones that came into use after January of 2012, along with any new features, are not included in this study.

5.4 Implications of the Study

This study contributes to the fields of Chinese linguistics, systemic functional linguistics, language variation and planning, and second language teaching.

Some of the more frequently used homophones have already entered the lexicon of computer-mediated communication. With the interaction of computer-mediated communication and daily communication, some words may become common words within the Chinese vocabulary system, while other words may disappear. From the phonetic and morphological studies of Chinese Internet homophones, we can see the development of a trend within Chinese Internet language, i.e. more than half of the homophones retain the same pronunciation as their base words, and Chinese character homophones are the most frequently used type. It is also interesting to see the changes and variations of Chinese Internet language, which can be used for future language planning.

From a systemic functional grammar perspective, homophones can be used as lexical and grammatical cohesive devices. Homophones serve as cohesive devices to make the discourse a complete semantic unit. This paper shows that Hu et al. (2005)'s model can be applied to different styles of written texts, including Internet discourse. The study of the cohesion function of Chinese Internet homophones allows us to see the semantic meaning of whole text and the connections between textual units within the discourse. The analysis of the cohesion function also contributes to our understanding of the relationship between discourse and context. The findings of the cohesion analysis of Internet homophones can also be applied to second language teaching and learning to help Chinese language learners to understand Chinese Internet language and to communicate with Chinese native speakers more efficiently.

Reference

"Annual Inventory Check of 2011 Sina Mirco-blog." (n.d.). In *Sina Tech*. Retrieved from http://tech.sina.com.cn/z/2011pandian/index.shtml

"Annual Inventory Check of 2010 Sina Mirco-blog." (n.d.). In *Sina Tech*. Retrieved from http://tech.sina.com.cn/i/2010-12-29/07505040121.shtml

Bo, J. L. 2007. "Homophobic Puns and the Purpose of Expression." *Science & Technology Information* 30: 195 & 205.

Brown, G. & Yule, G. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.

Butler, C. 1985. *Statistics in Linguistics.* Oxford: Basil Blackwell.

Campell, K. S. 1995. *Continuity and Cohesion: Theoretical Foundation of Document Design*. New Jersey: Lawrence Erlbaum Association.

Cao, W. 2004. *The Study of Modern Chinese Lexicons*. Beijing: Peking University Press.

Chen, D. W. 2007. "Analysis on Discourse Cohesion in Chinese Cross Talk." M.A. Dissertation, South China Normal University.

Chen, R. 2004. *Contemporary Chinese Rhetoric*. Beijing: Peking University Press.

Chen, P. 2011. "A Comparative Study of *Xieyin* in Chinese and English from the Perspective of Cognition." M.A. dissertation, Central China Normal University.

Collot, M. & Belmore, N. 1996. "Electronic Language: A New Variety of English." Pp. 13-28 in *Computer-Mediated Communication: Linguistic, Social, and Cross-culture Perspectives* edited by Herring S. John Benjamins Publishing Company.

Crystal, D. 2006. *Language and the Internet*. Cambridge: Cambridge University Press.

Enkvist, N.E. 1978. "Coherence, Pseudo-Coherence, and Non-coherence." Pp. 109-128 in *Report on Text Linguistics: Cohesion and Semantics* edited by Ostman, J. O. Turku-Abo, Finland: Research Institute of the Abo Academy Foundation.

Fang, H. F. & Li, Y. B. 2011. "Literature Review on Text Cohesion and Coherence Theory in China." *Data of Culture and Education* 5: 37-39.

Fang, Y. 2005. “Systemic Functional Grammar and Discourse Analysis.” *Foreign Language Education* 5: 1-5.

Fang, Z. 2005. “On the Internet Vocabulary Formation.” *Rhetoric Learning* 6: 39.

Feng, J. (ed.). 2012. *New Internet Dictionary*. Beijing: New World Press.

Fong, M. 2000. “ ‘Luck Talk ’ in Celebrating the Chinese New Year.” *Journal of Pragmatics* 32: 219-237.

Fu, H.Q. 1985. *Modern Chinese Lexicon*. Beijing: Peking University Press.

Ge, B. Y. 1985. On Chinese Word-formation and Morphology. *Journal of Literature, History and Philosophy* 4: 28-33.

Ge, B.Y.2001. *Modern Chinese Lexicology*. Shandong: Shandong People’s Publishing House.

Georgakopoulou A. 2011. “Computer-mediated Communication.” Pp. 93-106 in *Handbook of Pragmatics Hightlights 9: Pragmatics in Practice* edited by Fried, M., Jan-Ola Östman, & Verschueren, J. John Benjamins Publishing Company.

Gottlieb, N. 2009. “Language on the Internet in Japan.” Pp. 65-78 in *Internationalizing Internet Studies: Beyond Anglophone Paradigms* edited by Goggin, G. & McLelland, M. New York & London: Routledge.

Gu, W, Pan, W. Y., & Liu, Y. H. 2001. *Practical Modern Chinese Grammar*. Beijing: The Commercial Press.

Guo, H.J. 2002. “English Loan Words in Contemporary Chinese and Their Impact on Chinese Grammar: the Period from 1980s to the Present.” *Journal of PLA University of Foreign Language* 05: 1-4.

Guo, Z.R. 2001. “On Homophonic Rhetoric.” *Chinese Character Culture* 4: 50-52.

Guo, Y. K. 1999. “The Changes of Homophonic Pun.” *Rhetoric Learning* 5: 6-7.

Gutwinski, W. 1976. *Cohesion in Literary Texts: A Study of some Grammatical and Lexical Features of English Discourse*. The Hague: Mouton Linguarum Series Minor 204.

Halliday, M. A. K. 1962. “Descriptive Linguistics in Literary Studies.” Pp. 26-39 in *English Studies Today*, 3rd Series edited by Duthie, G. I. Edinburgh University Press.

Halliday, M.A.K. 1964. “The Linguistic Study of Literary Texts.” Pp. 302-307 in *Proceedings of the Ninth International Congress of Linguists* edited by Lunt, H. G. The Hague: Mouton. [reprinted in Chatman , S. & Levin, S. R.1967. *Essays on the Language of Literature*. Boston: Houghton Mifflin. 217-223.]

Halliday, M. A. K. 1968. “Notes on Transitivity and Theme in English, Part 3.” *Journal of Linguistics* 4:179-215.

Halliday, M. A. K. 1970. “Language Structure and Language Function.” Pp. 140-165 in *New Horizons in Linguistics*, edited by Lyons, J. Harmondsworth: Penguin.

Halliday, M.A.K. & Hasan, R.1976. *Cohesion in English*. London: Longman.

Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M. A. K. & Hasan, R. 1985. *Language, Context and Text: Aspects of Language in a Social Semiotic Perspective*. Victoria, Australia: Deakin University Press.

Halliday, M.A. K. 1994. *An Introduction to Functional Grammar (2nd Edition)*. London: Arnold.

Halliday, M.A. K. & Matthiessen, C. 2004. *An Introduction to Functional Grammar*. New York: Oxford University Press.

Hartmann, R.R.K., & Stork, F.C.1972. *Dictionary of Language and Linguistics*. London: Applied Science.

Herring, S. 1996. “Introduction.” Pp. 1-12 in *Computer-Mediated Communication: Linguistic, Social, and Cross-culture Perspectives* edited by Herring, S. John Benjamins Publishing Company.

Hoey, M. 1991 *Patterns of Lexis in Texts*. Oxford: Oxford University Press.

Hu, G. M. 2006. “The Study of Chinese Harmonic Phenomenon”, M.A. dissertation, Changchun University of Science and Technology.

Hu, Z. L. 1984. “Differences in Mode.” *Journal of Pragmatics* 8: 595-606.

Hu, Z. L. 1993. “The Cohesive Function of the Phonological System in Producing English Texts.” *Foreign Language Teaching and Research* 2: 1-8.

Hu, Z. L. 1994. *Discourse Cohesion and Coherence*. Shanghai: Shanghai Foreign Language Education Press.

Hu, Z. L. 1996. “Thought on Multiple Levels of Textual Cohesive Theory.” *Journal of Foreign Languages*1: 1-8.

Hu, Z.L., Zhu, Y. S., Zhang, D. L., & Li, Z.Y. 2005. *Introduction to System-Functional Linguistics*. Beijing: Peking University Press.

Hu, Y.S. 1992. *Modern Chinese*. Hong Kong: Joint Publishing.

Huang, B. R., & Liao, X. D. (ed.) 2008. *Modern Chinese*. (3rd Edition) Beijing: Higher Education Press.

Huang, G.W. 1988. *Essentials of Text Analysis*. Hunan, China: Hunan Education Press.

Huang, G. W. 2000. "Systemic Functional Linguistics: Forty Years on." *Foreign Language Teaching and Research* 1: 15-21.

Huang, G.W. 2001. *Theory and Practice of Discourse Analysis*. Shanghai Foreign Language Education Press.

Huang, G. W. 2002. "Hallidayan Linguistics in China." *World Englishes* 21(2): 281-290.

Huang, G. W. 2011. "Systemic Functional Grammar Research in China: Its Development and Future." Pp. 42-78 in *A General Introduction to Systemic Functional Linguistics* edited by Huang, G. W. & Xin, Z. Y. Beijing: Foreign Language Teaching and Research Press.

Institute of Linguistic Studies, Beijing Language and Culture University (ed.). *Modern Chinese Frequency Dictionary*. 1986. Beijing Language and Culture University Press.

Jiang, Y. 2003. "Summary on the Network Language Derived Principle." *Rhetorical learning* 2: 32.

Joel B. 2004. "Second Language Cyber Rhetoric: a Study of Chinese L2 Writers in an Online Usenet Group." *Language Learning & Technology* 8(3): 66-82.

Li, C. L. 2009. "On Discourse Mechanism of Homophonic SMS." M.A. Dissertation, Hunan Normal University.

Li, C. Y. 2009. "The Formation and Features of Internet Catchphrases". *Guangxi Social Sciences* 2: 114-116.

Li, L. X. 2007. "Network Language Research." M.A Dissertation, Shaanxi Normal University.

Li, M. 2004. "The Types, Characters and Criterion of Words in the Internet Language." *Journal of Women's Academy at Shandong* 3: 48-50.

Li, J. 2003. "'Cyber language' and Its Space." *Beijing Morning Post* 19 (5).

Li, E. S.2007. *A Systemic Functional Grammar of Chinese*. London: Continuum International Publishing Group.

Liao, Y. 2004. *Linguistic Study on News Report: Cohesion and Coherence of News Text*. PHD Dissertation, Hunan Normal University.

Lin, J. C. 1986. "Exploration of Lexical Devices in English Texts." *Journal of Foreign Languages* 5: 20-26.

Lin, G. 2002. "The Types and Characteristics of the Network Term." *Rhetoric Learning*1: 26.

Liu, C.M. 2002. "'Network Token' Group Is Reasonable, According to the Analysis." *Language Teaching and Research* 6: 31-35.

Liu, H. R. 2010. "Overview of Chinese Internet Language Research from 2000 to 2009." M.A. Dissertation, Northeast Normal University.

Liu, J. 2011. "Deviant Wring and Youth Identity: Representation of Dialects with Chinese Characters on the Internet." *Chinese Language and Discourse* 2: 58-79.

Liu, Q. Y. 2004. "Cohesion and Coherence in Text Translation." *Shangdong Foreign Languages Journal* 3: 5-9.

Liu, X. M. 2006. "Homophonic Puns in Internet Language." *Journal of Language and Literature Studies* 9: 129-131.

Lu, H.2003. "On Research of Voice Variation Phenomenon of Internet Language." *Rhetoric Learning* 1: 38.

Luo, X. M. & Liu, Q. R. 2003. "The Coherent Function of Lexical Chains in Discourse." *Foreign Language Education* 2: 7-10.

Lü, S. X. 1984. *Papers in Chinese Syntax*. Beijing: The Commercial Press.

Martin, J. 1992. *English Text: System and Structure*. Amsterdam: John Benjamins.

Martin, J. 2001. "Cohesion and Texture." Pp. 35-53 in *The Handbook of Discourse Analysis* edited by Schiffrin, D., Tannen, D., & Hamilton, H. E. Malden, Mass: Blackwell Publishers.

McCarthy, M. & Carter, R. 1994. *Language as Discourse*. London and New York: Longman.

McDonald, E.1992. "Outline of a Functional Grammar of Chinese for Teaching Purposes." *Language Sciences* 14 (4): 435-458.

Miao, X. W. 1998. "The Relationship between Cohesion and Coherence." *Journal of Foreign Languages* 4: 44-49.

Morgan, J. L. & Sellner, B.M. 1980. “Discourse and Linguistic Theory.” Pp. 165-200 in *Theoretical Issues in Reading Comprehension* edited by Spiro, R. J., Bruce, B.C. & Brewer, F.W. New Jersey: Lawrence Erlbaum.

Nigel, I. 2010. “China in Cyberspace.” *Survival: Global Politics and Strategy* 52(4): 55-66.

Ning, S. W. 1975. “ ‘Jiang zhenshi yinqu’ ‘yong jiayu cunyan’ [to Hide Truth and Use Rumor]”. *Journal of Nanjing Normal University* (Social Science Edition) 3: 70-75.

Oxford Language Dictionaries Online, http://www.oxfordlanguagedictionaries.com

Packard, J.L. 2000. *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge University Press.

Pan, G.W., Ye, B.Q., & Han, Y.2004. *Research on Chinese Morphology*. Shanghai: East China Normal University Press.

Parivash, M. & Bahram, L. 2011. “Introducing Categorization of Homophonous Words in English Language.” *Procedia-Social and Behavioral Sciences* 31: 602-606.

Parsons, G. 1990. *Cohesion and Coherence: Scientific Texts: A comparative Study*. Department of English Studies, University of Nottingham.

Partington, A. 2009. “A Linguistic Account of Wordplay: the Lexical Grammar of Punning.” *Journal of Pragmatics* 4: 1794-1809.

Paula L. R. 2007. “Teaching L2 Vocabulary through SMS Language: Some Didactic Guidelines.” *Estudios de linguistic ingleca aplicada* 7: 165-188.

Platt, J., Richards, J., & Webber, H. 1985. *Longman Dictionary of Applied Linguistics*. Great Britain: The Richard Clay (The Chaucer Press) Ltd.

Pi, H.M. 1992. “The Features and Significance of Bi-syllablization in Chinese.” *Research in Ancient Chinese Language* 01: 56-59.

Pollack, J. 2011. *The Pun Also Rises*. New York: Gotham Press.

Qi, X. J. 2002. “On the Semantics of Homophones in Internet Language.” *Studies of Rhetoric* 3: 21.

Redfern, W.1984. *Puns*. London: Blackwell.

Ronald J., P. Edibert, J. Rafal, & Z. Rohozinski 2011. *Access Contested: Security, Identity, and Resistance in Asian Cyberspace.* Massachusetts: The MIT press.

Shi, C.H. 2009. “The Language Value and Linguistic Value of Web Language.” Proceedings from the *6th Academic Forum on Chinese Language Application*. Retrieved from http://www.dwhyyjzx.com/www/news.asp?id=22&newsid=346.

Shi, Y. L. & Sun, Z. X. 2011. “A Comparative Study on the Cohesion in *Treasury Island* and Its Chinese Version.” *Writer Magazine* 3: 193-194.

Song, Y.Z.1992. *Basic Knowledge of Modern Chinese*. Beijing: Language & Culture Press.

Suen , C. Y. 1979. *Computational Analysis of Mandarin*. Basel, Switzerland: Birkhauser.

Sun, C.C. 2011. “From the Rapid Population of *beiju* to See the Development Trend of Internet Homophones.” M.A.Dissertation, Jilin University.

Sun, C.F. 2006. *Chinese: A Linguistic Introduction*. Cambridge University Press.

Sun, Y. 1995. “Commentary on Hu Zhuanglin’s Discourse Cohesion and Coherence.” *Journal of Foreign Languages* 4: 78-79.

Sun, Y. 1997. “On the Origin, Nature and Relation of Cohesion and Coherence.” *Journal of Foreign Languages* 1: 31-35.

Sun, Y. 2011. “Research on Systemic Functional Grammar and Russian Textual Coherence.” PHD Dissertation, Heilongjiang University.

Tang, Q. Y. 2005. “Exploration of Cohesion in Medical Discourse: Discourse Cohesion Analysis on an Eye Drop Advertisement.” *Journal of Jinggagnshan Medical College* 2 (4): 18-19.

Thompson, G. 2004. *Introducing Functional Grammar*. London: Arnold.

Thorne, S. 2008. “Computer-Mediated Communication.” Pp. 1-12 in *Encyclopedia of Language and Education*, 2nd Edition, Volume 4: *Second and Foreign Language Education* edited by Deusen-Scholl, V & N. Hornberger (eds). Spring Science+ Busniess Media LLC.

Wang, C. L. 2000. “On the Anaphoric Function of The Personal Pronoun *ta*.” *Studies of the Chinese Language* 3: 228-237.

Wang, H. Y. 2010. “On Grammatical and Phonetic Cohesion of Antithetical Couplets.” *Literature and Arts* 4: 15-16.

Wang, J. 2005. “A Study on the Cohesion Means Based on Grammar and Sound”. M.A. Dissertation, Nanchang University.

Wang, P. (2000). “Objective Basis for Chinese homonym.” *Rhetoric Learning* 2: 8-9.

Wang, Z. Q. 2002. “Another Discussion on Cohesion, Coherence and Correlation.” *Foreign Language Research* 2: 28-31.

Wang, Z. Y. 1981. “Review of Halliday’s ‘Grammatical Categories in Modern Chinese.” *Linguistic abroad* 2:55-62.

Wei, N.S. 2011. “Detailed Analysis of Chinese Character Homophones and Homophones with Borrowed Base words In the Internet.” *Modern Chinese* 08: 61-62.

Widdowson, H. C. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.

Wu, D. Y. 2001. “On Englishization of Chinese Morphology.” *Contemporary Linguistics* 2: 81-89.

Wu, W. Z. & Tian, X. L. 2000. *Clusters of Chinese Sentences*. Beijing: The Commercial Press.

Xiang, Z. 2012. “The Research of Verbal Communication Characteristics in MicroBlog.” MA Dissertation, Jilin University, China.

Xu, S. H. 1982. On Theme and Rheme. *Foreign Language Teaching and Research* 2: 1-8.

Xu, T.J. 1993. *Language*. Changchun: Northeast Normal University Press.

Xu, J. 2004. “Cohesion, Texture, and Coherence.” PHD Dissertation, Fudan University.

Xu, Y. C. 1996. “Comparative Studies on Cohesive Devices in Chinese and English.” *Shangdong Foreign Language Journal* 4: 1-6.

Xu. Y. C. 2003 a. “Comparative Analysis on English Essay Cohesive Devices and Errors.” *Foreign Language Teaching Abroad* 3: 54-60.

Xu. Y. C. 2003b. “Textual Cohesion and Mental Coherence.” *Shangdong Foreign Language Journal* 3: 12-16.

Yan, B. 2005. “On Partial Tone Use in Network Language.” *Journal of Changdu University of Technology Social* 13(3): 32-34.

Yan, X. 2010. “Cohesion Studies in the Past 30 Years: Development, Application and Chaos.” *The International Journal - Language Society and Culture* 31: 139-147.

Ye, H. 2005. “Analysis of the Impact of Internet Language to Chinese and Its Future.” August 12, http://www.hpe.sh.cn.

You, Y.X. 2012. "Characteristics of Neologism and Its Cognitive Motivations: A Study of Chinese Neologisms from 2006 to 2009." PhD Dissertation, Shanghai International Studies University.

Yu, G. Y. (ed.) 2001. *Chinese Internet Language Dictionary*. Beijing: China Economic Publishing House.

Yu, Y. 2009. "On the Internet Word *jiong*." *Studies of Rhetoric* 1: 93-95.

Yukiko N.2003. "Linguistic Innovations and Interactional Features of Casual On-line Communication in Japanese." *Journal of Computer-Mediated Communication* 9 (1), http://jcmc.indiana.edu/vol9/issue1/nishimura.html

Zhang, B. J. & Fang, M. 1996. *Functional Studies of Chinese Grammar*. Beijing: Commercial Press.

Zhang, D.L. 1999. "A Survey of Studies on Text Coherence." *Journal of Foreign Languages* 6: 24-31.

Zhang, D. L. 2001a. Cohesive Force and Text Coherence. *Foreign Languages and Their Teaching* 1: 9-15.

Zhang, D. L. 2001b. "On Cohesion." *Journal of Foreign Languages* 2: 23-28.

Zhang, D. L. 2001c. "On the Principles of Internal Cohesion of Text." *Journal of PLA University of Foreign Languages* 11: 26-32.

Zhang, D. L. 2002. "Cohesion and Style: The Stylistic Features of Reference and Lexical Cohesion." *Foreign Language and Their Teaching* 10: 1-7.

Zhang, D. L. & Liu, R. S. 2003. *Progress in Studies on Text Cohesion and Coherence*. Shanghai: Shanghai Foreign Language Education Press.

Zhang, D. L. 2004. "New Developments in Systemic Functional Linguistics." *Contemporary Linguistics* 6: 57-65.

Zhang, D. L. 2005. "Form and Meaning in Text Cohesion." *Journal of Foreign Languages* 5: 32-38.

Zhang, D. L. 2006. "Systemic Functional Linguistics in China." *Foreign Languages in China* 02: 27-32.

Zhang, Y. H. 2007. "On Grammatical Features of Internet Lexicon." *Studies of the Chinese Language* 6：531-535.

Zhang, Y.Y. 1982. *Brief Introduction of Lexicology*. Wuhan: Huazhong University of Science and Technology Press.

Zhao, H. L. 2005. “On the Rhetoric of Internet language.” *Applied Linguistics* 9: 123-125.

Zhao, J. M. 1987. “Homophony and Culture.” *Language Teaching and Linguistic Studies* 1: 40-56.

Zheng, Q. J. 2007. “A Study of Intertextual Cell Phone Messages and Their Texture.” *Language Teaching and Linguistic Studies* 05: 82-89.

Zhou, C. L. 2005. “On Homophony’s Role, Word-Formation, and Rhetoric Effect”. *Economic and Social Development*, 2: 4−8.

Zhou, H. Y. 2006. “The Application of Cohesion in Text Functional Analysis.” *Foreign Languages and Their Teaching* 10: 22-24.

Zhou, Q.S. (ed.) 2006 a. *Language Situation in China: 2005 Part 1*. Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2006 b. *Language Situation in China: 2005 Part 2*. Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2007 a. *Language Situation in China: 2006 Part 1*. Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2007 b. *Language Situation in China: 2006 Part 2.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2008 a. *Language Situation in China: 2007 Part 1.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2008 b. *Language Situation in China: 2007 Part 2.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2009 a. *Language Situation in China: 2008 Part 1.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2009 b. *Language Situation in China: 2008 Part 2.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2010 a. *Language Situation in China: 2009 Part 1.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2010 b. *Language Situation in China: 2009 Part 2.* Beijing: The Commercial Press.

Zhou, Q.S. (ed.) 2011. *Language Situation in China: 2010.* Beijing: The Commercial Press.

Zhou, Q.S., & Hou, M. (ed.) 2012. *Language Situation in China: 2011*. Beijing: The Commercial Press.

Zhu, D.X. 1982. *Lecture Notes of Grammar*. Beijing: The Commercial Press.

Zhu, Y. S. 1995. "The Development and Improvement of Cohesion Theory." *Journal of Foreign Languages* 3: 36-41.

Zhu, Y. S., Yan, S. Q., & Miao, X. W. 2001. *Contrastive Study of Cohesion in English and Chinese*. Shanghai Foreign Language Education Press.

Zhu, Y. S. 2002. "Semantic Indirectness in Chinese Compounds and the Discrimination between Compounds and Phrases." *Chinese Teaching in the World* 05: 50-59.

Zuo, Y.1995. "Comparing Cohesive Device in Chinese and English." *Foreign Language Teaching and Research* 3: 37-42.

Appendix

Abbreviations used for the glosses in the Chinese examples:

ADV	an adverbial maker (*de*)
ASSC	associative (*de*, *zhi*)
BA	a pretransitive maker (*ba*)
CL	classifier
CRS	currently relevant state (*le*)
CSC	complex stative construction (*de*)
NEG	negatives (*bu*, *mei*, *bie*)
NOM	nominalizer
PFV	perfective aspect (*le*)
PL	plural personal pronoun
POSS	possessive (*-de*)
PRT	particle
Q	question marker (*ma*)
SG	singular personal pronoun