A Thesis

Entitled

Using Regional Dialects through Computer-Mediated Communication in China

By

Ying Wu

Submitted as partial fulfillment of the requirements for

The master of Arts Degree in English with a concentration in

English as a Second Language

____________________________________
Advisor: Dr. Melinda Reichelt

____________________________________
Dr. Douglas W. Coleman

____________________________________
Dr. AnChung Cheng

____________________________________
College of Graduate Studies

The University of Toledo

May 2009
Copyright © 2007

This document is copyrighted material. Under copyright law, no part of this document may be reproduced without the expressed permission of the author.
An Abstract of

Using Regional Dialects through Computer-Mediated Communication in China

Ying Wu

Submitted as partial fulfillment of the requirements for The Master of Arts Degree in English with a concentration in English as a Second Language

The University of Toledo

May 2009

This thesis explores how people in China write regional dialects for Computer-Mediated Communication. Two prestige dialects in China, Shanghainese and Cantonese, are the focus of the study. 30 observations were made on electronic postings for examining the dialect writings, and 40 dialect speakers were involved in a questionnaire survey about the use of the dialects online from the sociolinguistic perspective. The findings of the study indicated that for online communication in the dialects, people mainly used common characters of the existing modern written Chinese. To represent unique vocabulary in the dialects, the dialect speakers resorted to phonetically borrowed characters from modern written Chinese. The results also showed that people used the dialect writings more frequently in local electronic forums than forums for people from everywhere.
Acknowledgements

First and foremost, I would like to express profound gratitude to my thesis advisor, Dr. Melinda Reichelt for her invaluable support, encouragement, supervision and useful suggestions throughout the research. Without her encouragement on the topic of the study and continuous guidance and support, I could not have finished my thesis.

I am also highly thankful to my second reader, Dr. Douglas W. Coleman, director of MA ESL program, for his insightful comments on my thesis. I thank him for introducing me to a scientific way of thinking, challenging me all the way through the design of the research to the final report, and helping me find ways to improve the research. I am also very grateful to my third reader, Dr. An-Chung Cheng, for her careful review on my paper and moral understanding and support.

I would like to express my appreciations again to my thesis committee. I thank them for never accepting less than my best efforts and patiently guiding me through the thesis process.

I also like to extend my deepest appreciations to my data coders, Xiaoyue Wang in Shanghai University of Finance and Economics, Mei Zhang in Guangdong University of Foreign Studies, and Jing Luo in Huangzhou Normal University. They helped me greatly with coding the data and researched for relevant literature in the present study. Without their continuous support all the way through the coding process, I could never finish my study.

Last, I would like to thank my family and best friends in China, who forever stand by my side and cheer me up whenever I need them.
# Table of Contents

Abstract 

Acknowledgements 

Table of Contents 

List of Figures 

I. Research Overview and Literature Review  
   Key features of Cantonese and Shanghainese lexicon 
   The type of special dialect characters 
   The status of dialect writing in China 
   Dialect writing in Cantonese 
   Dialect writing in Shanghainese 
   Using Dialects: A Sociolinguistic Perspective 
   The Present Research 

II. Methodology  
   Data 
   Instruments 
   Data collection 
   Data analysis 

III. Findings  
   Dialect writing and lexicon 
   The sociolinguistic variables and the use of the dialects online 
   The reasons and attitudes of the use of the written dialects online 
   Conclusion 

IV. Discussion, Conclusion and Implications 
   Discussion 
   Conclusion and implications 

V. References 

VII. Appendix: Questionnaire
Lists of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The timeline of 30 internet observations at US East Coast Timezone</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>The observed use of Cantonese writing types</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>The observed use of Shanghainese writing types</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>The reported use of Cantonese writing types</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>The reported use of Shanghainese writing types</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>The types of vocabulary in Cantonese written posts</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>The types of vocabulary in Shanghainese written posts</td>
<td>56</td>
</tr>
<tr>
<td>8</td>
<td>The use of dialects online by age</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>The use of dialects online by gender</td>
<td>59</td>
</tr>
<tr>
<td>10</td>
<td>The use of dialects online by education level</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>The use of dialects online by occupation type</td>
<td>61</td>
</tr>
<tr>
<td>12</td>
<td>The use of dialects online by residency</td>
<td>62</td>
</tr>
<tr>
<td>13</td>
<td>Years living outside Shanghai or Guangzhou</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>The reported use of dialects online by topic type</td>
<td>64</td>
</tr>
<tr>
<td>15</td>
<td>The reported inclination of website type</td>
<td>65</td>
</tr>
<tr>
<td>16</td>
<td>The Percentage of Shanghainese written posts in three website types</td>
<td>67</td>
</tr>
<tr>
<td>17</td>
<td>The percentage of Cantonese written posts in three website types</td>
<td>69</td>
</tr>
<tr>
<td>18</td>
<td>The reported use of dialects online by people type</td>
<td>70</td>
</tr>
<tr>
<td>19</td>
<td>The reasons of using dialects online</td>
<td>71</td>
</tr>
</tbody>
</table>
Figure 20. Reasons for the use of dialects online by gender  
72

Figure 21. The vote for having Cantonese news report online  
73

Figure 22. The vote for having Shanghainese news report online  
74

Figure 23. The impact of internet on the promotion of Cantonese  
75

Figure 24. The impact of internet on the promotion of Shanghainese  
76
Chapter One

Research Overview and Literature Review

Regional dialect is a prominent feature of one’s language in communication. In China where the standard variety (Mandarin) and regional dialects exist side by side, people usually use regional dialects for daily communication, such as chatting with family members, friends, or neighbors, buying grocery, or ordering food in a local restaurant. As more and more people in China turn to the internet for communication, using regional dialects online becomes a popular type of net-speak among the native speakers of the regional dialects, especially those of prestige varieties. This study will explore people in China using regional dialects through Computer-Mediated Communication (CMC) — what they use, how they use and why they use.

Considering the large number of Chinese dialects, I limit my study to two prestige regional dialects in two big cities in mainland China, Guangzhou-Cantonese\(^1\) and Shanghainese. A questionnaire of 24 questions was distributed to 20 native speakers of Cantonese and 20 native speakers of Shanghainese. Approximately 40,000 words were drawn from the posts in four electronic forums involving informants from Guangzhou and Shanghai. Quantitative and qualitative analysis were performed to the data drawn from the questionnaires and the texts. There has been very little research done on this topic, and I hope
this thesis will provide some new avenues for consideration.

In this chapter, I will review previous studies to enumerate some distinctive lexical features of Guangzhou-Cantonese and Shanghainese as well as the dialect writing, so as to provide a base for my data coding in the methodology part. I will also review the sociolinguistic studies on the use of regional dialects in their spoken forms, which is strongly associated with the use of written dialects online. In Chapter Two, I will describe my methodology of conducting the research, including the description of data, the procedure of collecting the data, and the qualitative and quantitative analysis of the data. In Chapter Three, I will present the findings according to my research questions, what people use for writing dialects online, how they use the written dialects online and why they use the regional dialects for CMC. The results are consistent with previous studies in some cases while not in others. I leave Chapter Four for further discussion.

Before I start the literature review, I would like to clear grounds for my thesis. First, I would like to justify why I refer Cantonese and Shanghainese as regional dialects. A regional dialect is a variety of a language that is spoken by people in a certain region. Cantonese and Shanghainese have for long been considered by Chinese scholars as “regional dialects” of the Chinese language family since the publication of Fangyan (“regional dialects”) by Xiong Yang around 2000 years ago. Whether the varieties of the Chinese languages are dialects or languages only became a controversy in the 1890s when Fangyan was translated into “dialect” and introduced to western linguists (Chen, 1996). Western linguists found a dialect in the Chinese language family did not equate a dialect in English. For example, the disparities
between Wu-dialect and Cantonese are far away from the difference between Boston English and Toronto English. Given the mutual unintelligibility between Mandarin (the officially standard spoken Chinese since 1955) and Cantonese or Wu-dialect which includes Shanghainese, it had been argued that Cantonese was a separate “language.” However, in addition to the consideration of mutual intelligibility, socio-historical and political variables also correlate with the debates over “dialect” versus “language”. For example, China has long history of considering Cantonese and Shanghainese as regional dialects; the official governments of China have for long implemented “one language” policy in order to unify all the regions of the country; linguists in mainland China reached agreements on the status of regional varieties. Given what have been mentioned above, I will in the present study refer Cantonese and Shanghainese as “regional dialects.”

Second, I would like to define the key word “dialect writing” in the thesis. The main focus of the thesis is to explore how people write dialects online. Dialect writing in a broad sense refers to any written forms people employ to transcribe the spoken dialects. Basically there are two ways to transcribe a spoken variety, either by a Romanization system or by Chinese characters. The Romanization systems involve Pinyin (a Romanization system used to transcribe spoken Mandarin), English alphabets, IPA (International Phonetic Alphabets) and other Romanization systems such as Jyutping (a Romanization system used to transcribe spoken Cantonese). The way of transcribing spoken varieties in Chinese characters involves the modern simplified Chinese characters and the traditional Chinese characters. The dialect writing in the thesis only refers to the transcription in Chinese characters because Chinese
people have tradition of transcribing sound in characters for long and the practice is still prevalent at present. Details will be discussed later in this chapter.

Thirdly, I would like to clarify the rationale of the notation systems adopted in the thesis. For the three varieties, Cantonese, Shanghainese, and Mandarin, I will use IPA (represented in square brackets) to transcribe their pronunciation, English lexicon (represented in single quotes) to denote their meaning, and modern Chinese characters to represent the graphic forms. As one may argue, the standard form of transcribing the pronunciation of Mandarin is Pinyin, a more established and popular way to transcribe the speech of Cantonese is Jyutping, and a newly proposed way to transcribe spoken Shanghainese is IPA\(^1\). However, using three different systems to transcribe the pronunciations of the three varieties may cause inconsistency over the transcription system of pronunciations, which in turn poses hurdles on the readability of the thesis. In addition, the main focus of the thesis is dialect writing, and thus the transcription of pronunciations will be reasonably sacrificed to the slightest degree when using IPA for the three varieties\(^2\). In terms of the notation for meaning and graphic forms, since there are no standard writing systems for Cantonese and Shanghainese so far and the only standard form to write Mandarin in mainland China is modern Chinese characters, the thesis will employ English vocabulary to denote the meaning of words in the three varieties, and modern Chinese characters to represent the graphic forms if necessary.

Fourth, I would like to explain the difference among \( zi, ci, \) and \( duanyu \) in Chinese and morpheme, word and phrase in English. Generally speaking, \( zi \) in Chinese is translated to
‘Chinese character’ in English, ci to ‘word’ and duanyu to ‘phrase’. However, there are many controversies around these issues. First, zi can be a morpheme or a word in English. In a word [tʂɔn tʂuo] ’think over’, [tʂɔn] and [tʂuo] are considered as morphemes of the word because the meaning of the word should be inferred from the combination of the two parts. But in a phrase [tʂɔn tʂA] ’pour-tea’, [tʂɔn] is considered as a word because it can stand alone as a meaning. Second, ci can be a word or a phrase in English. For example, [tʂɔn tʂA] ‘pour-tea’ is usually considered as a verb phrase in a sentence such as “she poured me some tea.” But in a coordinating phrase [tʂɔn tʂA ti s’uei] ‘pour-tea-serve-water’, [tʂɔn tʂA] is considered as a word. The phrase as a whole is used as a noun phrase or a verb phrase in a sentence. In addition, it is not easy to tell the difference between word and phrase in English either. In all, what should be considered as a word and what should be considered as a phrase are hard to define when they stand alone. Therefore, in the present study, what makes a “word” in Chinese depends on the context where the “word” is used by informants.

Key features of Cantonese and Shanghainese lexicon

Spoken Cantonese and Shanghainese differ from Mandarin in many aspects, such as speech sound, vocabulary and sentence structure. But in written forms, dialects are characteristics of their lexicon. In the family of Chinese language, a sentence of a dialect cannot be identified by looking at its sentence structure alone, but probably can by looking at the word used and its written form. Since the purpose of the present study is to explore how people write their regional dialects online, I will focus on the lexical characteristics of the two dialects only. In the following paragraphs, I will review the systematically lexical
differences on Cantonese and Shanghainese so as to give background information on the data analysis for the texts of dialect writings in the present study.

As a variety which maintains a lot of features of Nanyue-yu, an ancient variety spoken in Southern China, Cantonese has some distinctive features in vocabulary. Many researchers, such as Wang (1960), Zheng (1973), and Zhan (1988), have conducted statistical surveys on the ratio of overlapping vocabulary between Cantonese and Mandarin. Their studies were based on dictionaries of Mandarin and Cantonese. The word which shared the same written form and similar meaning in Mandarin and Cantonese is considered as a common word no matter they have similar pronunciations or not. For example, 茶 [tʂĄ] is considered as a common word in Mandarin and Cantonese, even though [tʂĄ] in Mandarin can be used as ‘water’ as well as ‘tea’, whereas referring to ‘tea’ only in Cantonese.

According to the weighted average computing (You & Yang, 1998), Cantonese shares 48.24% of its lexicon with Mandarin. However, there are still quite a few lexical differences between Cantonese and Mandarin (Zhu, 2004). First, there are many one-syllable words in Cantonese while they are two-syllable in Mandarin. For example, a son and daughter is referred as [dʐʰi] ‘son’ and [nœy] ‘daughter’ respectively in Cantonese, whereas as [ʂɿ tsi] ‘son-child’ (‘son’) and [ny ʂɿ] ‘daughter-son’ (‘daughter’) respectively in Mandarin. Second, the modifier-noun phrases in Mandarin are usually in the order of modifiers plus nouns, while sometimes in the order of nouns plus modifiers in Cantonese. For example, a guest is referred as 客人 ‘guest-person’ in Mandarin, while 人客 ‘person-guest’ in Cantonese. Third, in Mandarin, verb-object phrases function as verbs in most cases and nouns in some
cases, while they may act as adjectives in Cantonese. For example, 生性 means ‘personality’ 
\((n.)\) in Mandarin, while it could be interpreted as ‘matured’ \((adj.)\) in Cantonese. Fourth, there 
are three-word coordinate phrases in Cantonese, in which the first two words parallel the 
third word in meaning; this is a rare case in Mandarin. For example, a lazy person with good 
appetite can be referred as ‘big-eat-lazy’ in Cantonese, but ‘love-eat-lazy-do’ in Mandarin. In 
addition to the systematic differences mentioned above, Cantonese exhibits its specialty in 
lexicon in terms of the feature words as well. Feature words are special words that are 
commonly used in the dialect spoken regions but rarely used outside the regions \((Li, 2001)\). 
There are some feature words that represent meanings in unique pronunciations in Cantonese, 
such as [dzep lep] (‘close a business’), [dzeu dzon] (‘or so’), and [dzon kʷəi] (‘hilarious’).

Compared to Cantonese, less research has been found on the Shanghainese lexicon. 
This may be due to the similarity Mandarin and Shanghainese share in vocabulary. 
According to You & Yang’s study in 1998, the ratio of overlapping vocabulary between 
Shanghainese and Mandarin is as high as 64.88%. However, Shanghainese still shows some 
distinctions from Mandarin in vocabulary \((Xu & Tang, 1981)\). First, the use of reduplication 
is much more commonly seen in Shanghainese than in Mandarin. Adjectives, verbs, nouns or 
classifiers (a word used to classify a noun according to its meaning, such as [ke] to quantify 
the amount of a tree) can be doubled to construct a word in Shanghainese. For example, the 
color of blood red is referred as ‘blood-blood-red’ in Shanghainese, while ‘blood-red’ or 
‘deep-red’ in Mandarin. Second, [tɕiɔ] or [nəŋ], two function words in Shanghainese with no 
meaning, can be used to modify a verb after double-adjjectives, while these are rare cases in
Mandarin. For example, an imperative sentence to ask somebody to walk slowly can be referred as ‘slow-slow-[tɕio]-walk’ or ‘slow-slow-[nəŋ]-walk’ in Shanghainese while ‘slow-slow-walk’ in Mandarin. Third, the prefix [a], a function word with no meaning, is usually placed before words of relations to address family members in Shanghainese, while the way functioned in Mandarin is to double the words of relations. For example, an elder brother is referred as ‘[a]-elder brother’ in Shanghainese, while ‘elder brother-elder brother’ in Mandarin. Fourth, the suffixes [tsʃ] and [dɤ], two function words with no meaning, are usually added to nouns of time and place in Shanghainese, while seldom seen in Mandarin. For example, this morning can be referred as ‘today-morning-[tsʃ]’ in Shanghainese, while only ‘today-morning’ in Mandarin; doorway can be referred as ‘doorway-[dɤ]’ in Shanghainese, while ‘doorway’ only in Mandarin. In addition, there are some feature words that are unique in Shanghainese, such as, [k’ɔ me pɨɲ] (‘deceive’), [mɔə zʃ] [mɔə ku] (‘slow’), and [tsɔ de fɨɲ] (‘show off’).

In this section, I reviewed the systematic difference of vocabulary between Mandarin and Cantonese, Mandarin and Shanghainese. These characteristics in vocabulary, as well as the feature words of the dialects, will serve as a basis for understanding the lexical features of the dialect-written texts, identified by the informants in the present study. In the following section, I will introduce the type of dialect characters in general.

*The type of special dialect characters*

Dialect character has for long been considered as the only option that can transcribe the dialects since Chinese people have a long history of transcribing the oral language in
ideographic form. The assumption didn’t change until the phonetic spelling was introduced by western missionaries to China around the 1890s. However, the use of phonetic spelling is not a prevalent practice; it is used in most cases for giving notations of pronunciation to a character. Therefore, nearly all the linguistic research on dialect writing is based on dialect characters. As members of the Chinese language family, Chinese dialects share a great amount of characters with Mandarin. The shared characters, namely common characters, are the dialect characters which have similar or the same meaning and pronunciations, as well as the same graphic forms with the characters in modern Chinese (the writing system for Mandarin) (Lin, 2003; Dong, 2005). In addition to the common characters, there are four types of special dialect characters: Ben-zi, Jiajie-zi, Xundu-zi, and zizhao-zi (Chen, 1996; Lin, 2003; Zhao, 2007; Matulewicz, 2002).

*Ben-zi* refers to native characters which can be tracked down in ancient literature of traditional Chinese. The association between the form of native characters and their meaning or sound was relatively consistent over time. However, these characters have fallen into disuse in modern Chinese writing but still remained in some regional dialects. In dictionaries of dialects, origins of these characters can always be dated to long time ago. For example, a commonly used Cantonese character 睇 (‘look’) was defined in *Shuo Wen*, a dictionary compiled around 100AD; the Shanghainese character 頃 (‘recline’) can be traced back to *Guangyun*, a dictionary compiled 900 years ago.

*Jiajie-zi* refers to the phonetically borrowed characters from modern Chinese characters. The pronunciations of these characters are usually similar to those in Mandarin in
terms of vowels and consonants, but not exactly the same since there are some distinctive phonetic features included in the dialects. The meanings of the phonetic borrowed characters are usually quite different from those in Mandarin. For example, 掃 refers to ‘press on’ or ‘wipe’ ([wən]) in Mandarin but refers to ‘earn’ ([wən]) in Cantonese; 拨 refers to ‘dial’ or ‘move with hands’ ([bo]) in Mandarin but refers to ‘to’ or ‘by’ ([bo]) in Shanghainese. The identification of phonetically borrowed characters lies in the collocation of the characters; whether the way the characters are combined is meaningful in Mandarin as in the dialects. If the combination renders meaning to the characters in the dialects but not in Mandarin, the character is considered as a phonetically borrowed character. For example, 仲 is pronounced as [tsuŋ] in both Mandarin and Cantonese, but means ‘the second’ (adjective) in Mandarin whereas ‘still’ (adverb) in Cantonese. In a sentence 仲未看完?, 未看完 means ‘not finish reading’ in both Mandarin and Cantonese, and 仲 can be only interpreted as ‘still’ rather than ‘the second’ to complete the meaning of the sentence. Therefore, 仲 in the sentence is used as a Cantonese character which carries the same pronunciation as in Mandarin but refers to a different meaning.

*Xundu-zi* refers to the dialect characters which are semantically borrowed from modern Chinese characters. They maintain the same or similar meanings and the same graphic forms as in the written Chinese while pronounced differently in terms of consonants and vowels in the dialects. For example, 看 refers to ‘read (books) or look at’ in both Cantonese and Mandarin, but is pronounced as [tʰai] in Cantonese while [kan] in Mandarin, which differ in both vowels and consonants. The identification of the semantically borrowed characters
cannot be made by looking at the characters alone. The characters can only be identified as semantically borrowed characters by examining the written context (e.g., the posting) in which they occur. If a character occurs in a sentence that has been identified as Cantonese by native speakers, then that character which has the same meaning and the same graphic form as in Mandarin but is assumed to be a semantically borrowed character in Cantonese. For example, in a sentence 仲未看完? (‘not finish reading yet?’) which is identified as a Cantonese sentence, 看 is pronounced as [tɔi], the pronunciation of which is quite different from [kan] in Mandarin but refers to the same meaning ‘read (books)’ as in Mandarin.

Zizhao-zi refers to the previously unattested forms of characters by the speakers of the dialects, which have never existed in written modern Chinese before. They cannot be traced for traditional use in ancient literature either. The speakers of the dialects usually combine a phonetic morpheme (usually an existing character) and a semantic morpheme to create a new character. For example, Cantonese speakers combine the semantic morpheme 口 ‘something related to mouth’ with the phonetic morpheme 左 [tsuo] to form a new character 咕 [tsuo], a functional word indicating the present perfect tense. The dialect speakers sometimes create a new character by combining two existing characters. For example, Shanghainese speakers put 勿 (‘no’) and 要 (‘want’) together to form a new character 鈷 (‘don’t want’), which is not included in modern written Chinese.

Researchers (Pawel, 2002; Zhao, 2007) have studied the percentage that each category of dialect characters accounts for in the written form of a dialect. In Cantonese, previously unattested forms of characters are the most common type while in Southern Min phonetically
or semantically borrowed characters accounts for the majority (Pawel, 2002). However, those studies are limited to the text analysis on dictionaries or literature works, hardly sourcing from the written dialects for daily use. In addition, there are two rubs under the previous studies of dialect writing. First, it is hard to categorize some dialect characters into a single type, because some dialect characters which have been disused in written Chinese were picked up for use again. For example, researchers are debating over whether 囝 ‘son’ is a native character or previously unattested form of character because it used to be included in traditional Chinese but disused for a while, and now is used frequently in writing Cantonese. Second, there are no standard or unanimously agreed criteria for the categorization of native characters. Up till now, there have not been an official writing scheme for spoken dialects and researchers have been working on confirming the sources of dialect characters in ancient literature. For example, some researchers consider 返 is the native character for ‘back’ in Cantonese while others argue 番 is the original graphic representation for the meaning (Zhan, 1995). The unfixed criteria increase the difficulty of categorizing characters into one type.

In this section, I introduced the type of dialect characters as a basis of understanding the characters identified by informants from online postings. In the following sections, I will review research on dialect writing in general, and on Cantonese writing and Shanghainese writing respectively.

The status of dialect writing in China

Due to the political concerns on the unity of Chinese people from different regions,
successive Chinese governments have for long discouraged or even prohibited the use of
dialect writing. Usually there was only one dialect and its writing system declared official
since Qin Dynasty (221 BC). Other dialect writings were proscribed in school systems,
public media and government documents. The dialect writings which have ever existed so far
were mostly found in dialect dictionaries or in literature works. Compiled by informant
researchers or government officers who were in charge of dialect transcriptions, the
dictionaries were hardly used by ordinary people for writing when the administrations tried
to put the country under the domination with one written language (Chen, 1996). Dialectal
literature works, such as novels, drama scripts, and folk songs, occasionally spotted in the
Chinese literature history. For example, *Min Du Bie Ji*, a novel in Fuzhou dialect, features the
invented dialect characters by the author (Lin, 2003). Compared to the eye dialect (a
non-standard spelling used for transcribing English dialects) extensively used in English
language literature, basically no Chinese literature works which were fully written in dialects
are as influential as “The Celebrated Jumping Frog of Calaveras County” by Mark Twain, or
the comic strip “Pogo” by Walt Kelly. And the amount of such works is far less as well.
However, there were classical literature works which adopted a few of dialect writings as a
manifestation of characters’ regional identities, such as Anhui Dialect used in *Ru Lin Wai Shi*,
and Shangdong Dialect used in *Jin Ping Mei* (Xiao, 2002).

In modern China, the current mainland government maintains the same strict language
policy as previous governments. In the *Law of the People’s Republic of China on the
Standard Spoken and Written Chinese Language* (Jiang, 2000), standard written Chinese is
required to be used in four areas: government, schools or other educational institutions, media industry such as publications, broadcasting, televisions, and public service industry. Using dialect writings in above-mentioned areas would be considered nonstandard and detrimental to the modernization of China, and users are at risk for administrative penalties except for a few special circumstances. However, three dialects of prestige, Cantonese (spoken in Guangzhou, Hong Kong and other southern regions of China), Southern Min (spoken in Fujian province, Taiwan and other southeastern regions of China), and Shanghainese (spoken in Shanghai), have aroused the interests of linguists and some political forces in the recent years for the development of their standardized written forms. Thanks to the promotion by Taiwan and Hong Kong administrations respectively and the historical independence from the mainland, the studies of Southern Min and Cantonese writings are in progress with regard to the dialect writings. But Shanghainese, which is under the strict control by Central Government of China, is still fighting for its room (Xu, 2005).

*Delect writing in Cantonese*

Written Cantonese characters, as mentioned above, have been studied in many research papers and have been compiled in dictionaries since a long time ago. Kong (1933) documented a large amount of popular Cantonese characters in *Studies on Popular Cantonese Words*; Luo (1960) verified 58 native characters in Cantonese by tracing back to ancient books in *Studies on the Origins of Guangzhou Dialect Words*; Bai (1980 & 1998) researched into Guangzhou dialect characters and compiled *Guangzhou Dialect Dictionary*; Rao, Ouyang, & Zhou collected many special Guangzhou dialect writings in *Guangzhou

In the 1990s, due to the needs to standardize Cantonese writing in official documents and writings in public media, Hong Kong administration made public the Government Common Character Set and increased the amount of standard Cantonese characters up to over four thousands in Hong Kong Supplementary Character Set. Two Cantonese Romanization systems are widely used for the notation and the input of spoken Cantonese into computer for education purposes. One is the Jyutping, a Cantonese Romanization scheme developed by the Linguistic Society of Hong Kong in 1993; the other is the Standard Cantonese Pinyin scheme, accepted by Education and Manpower Bureau of Hong Kong as well as Hong Kong Examinations and Assessment Authority. The former scheme features its undemanding learning and thus is more popular among ordinary people, and the latter is characteristic of its approximation to IPA (International Phonetic Alphabet).

However, the system of Cantonese characters has not been established in Guangzhou, although people in Guangzhou are profoundly influenced by the standard Cantonese character set proposed by the Hong Kong administration. This may be explained by two reasons. First, the use of the Jyutping or the Standard Cantonese Pinyin scheme of writing was not declared official or standard in Guangzhou because the local government has been implementing the one-language policy issued by the central government of China. As a result, people in Guangzhou do not have opportunities to learn them from formal education. Second, both Romanization schemes generate characters more of traditional Chinese, for which
people in mainland China are not educated and the generation of which requires a different font in computer.

_Dialect writing in Shanghainese_

Compared to Cantonese writing, written Shanghainese has not been intensively studied due to the strict political control over the writing system in Mainland China. However, there were a few research and dictionaries compiled for the Shanghainese words in history. Around the mid-1900s, missionaries from UK and France started to develop Shanghainese dictionaries for preaching purposes, such as _A Collection of Phrases in the Shanghai_ by John MacGowan in 1862, _A Vocabulary of the Shanghai Dialect_ by J. Edkins in 1870, _Dictionaire Francais-Chinois, Dialete de Shanghai_ by P. Rabouin in 1894 and a new version by Le P. A. Bourgeois (Office of Shanghai History, 2009). Researchers in modern China, such as Zhao (1928), Wang (1935), Xu & Tang (1960, 1981, & 1988), and Qian (1992, 2003, 2004, & 2007), have conducted studies on feature words of Shanghainese, but no research has been done to the dialect writing in its own right so far. In addition, the Shanghai government has not published any standard Shanghainese writing.

In the past ten years, the importance of Shanghai has been increasing in China due to its growing economic power and cultural influence. The loyalty to Shanghai and Shanghainese correspondingly becomes stronger. The popularization of the internet among Shanghainese people provides them a new venue to express their regional identity online by writing in Shanghainese. Linguists of Shanghainese, as well as internet users, have made great efforts to come up with a scheme to type Shanghainese in computer. In 2003, Xiaojun
Zheng, an ordinary internet user, developed a Shanghainese input scheme based on Latin spelling. In 2007, Qian, a linguist of Shanghainese in Fudan University, developed another input scheme based on IPA and published the *Shanghainese Dictionary* (2007), which includes over 14,000 words. However, neither of the schemes is well recognized by the government or the internet users. In all, there has been no standard transcription system or standard character set of Shanghainese for now.

In the previous sections, I examined the dialect lexicon and the dialect writings of Cantonese and Shanghainese from the linguistic perspective, and reviewed historical research on dialect writing. In the following section, I will explore the literature related to the use of the two dialects in social contexts.

*Using Dialects: A Sociolinguistic Perspective*

As a new avenue to study Chinese dialectology, the sociolinguistic perspective is not prevalent in the scholarship and there are few studies in this regard. Therefore, only a limited number of sociolinguistic studies of Cantonese and Shanghainese will be reviewed below. At the beginning, I would like to discuss the social functions of regional dialects in general. First, regional dialects are always associated with regional identity — one positioning himself or herself as a local resident. Through the use of the regional dialects, the speakers are able to reinforce their linguistic identities which in turn reinforce their regional identities. For example, in the Berlin Urban Vernacular (BUV) projects carried out in the early 1980s, Dittmar and Scholobinski found that East Berliner insisted using Berlinerisch (BUV) to highlight their regional identity of East Berliners as opposed to West Berliners after the
unification of the Federal Republic of Germany and the German Democratic Republic (as
cited in Mesthrie, Swann, Deumert, & Leap, 2000, p. 129-132). Second, regional dialects and
regional identities are often associated with social identities which correlate with variables
such as social status, education and gender. For example, in Hemnesberget, a village in
northern Norway, people (often non-locals) who had high social status in the community
spoke Bokmål, the standard variety, rather than Ranamål, the local dialect (Blom & Gumperz,
1972). In addition, by using regional dialects as “marked” choices, speakers can establish
solidarity with interlocutors or increase social distance against them as convenience in a
“linguistic market” (Mesthrie, Swann, Deumert, & Leap, 2000, p. 349). Bourdieu points out
“Linguistic interactions between speakers (in terms of content and, more so, style) depend
largely on the social relation between the speakers… Favored patterns of language (style,
discourse, accents) are conceived of as symbolic assets which can receive different values
depending on the market in which they are offered” (as cited in Mesthrie, Swann, Deumert,
& Leap, 2000, p. 344). Mazraani (1997) shows that Arabic politicians tend to use dialects
like Cairene, Baghdadi or Tripoli Arabic rather than Modern Standard Arabic when they give
speeches on their political programs or refer to their personal experience, because the former
ones can establish an equal personal relationship with the audience and get them involved.

Most places in China are diglossic where the standard spoken Chinese (Mandarin) and
regional dialects exist side by side. In most cases, Mandarin is considered the prestige variety,
although Shanghainese and Cantonese in Guangzhou are also considered prestige varieties
when compared to other regional dialects nationwide (Xiao, 2002; Xu, 2005). Qin (2007)
discussed the use of Shanghainese through examining the transcribed text of about 40,000 words from the recordings in public settings and mass media. She found that Shanghai people tended to use Shanghainese in informal settings while use Mandarin in formal settings such as on broadcasting or TV. She also noticed Shanghai people tended to talk with interlocutors who speaks Mandarin in Mandarin while talk with local people in Shanghainese. In addition, Shanghai people prefer to use Mandarin when talking about big issues such as the annual conference of Chinese congress, while use Shanghainese when talking about everyday-life issues such as buying groceries. Huang (2006) examined the use of Shanghainese and Mandarin by 50 female Shanghainese who aged from 20 to 25 through surveys and interviews. The findings are consistent with most of the findings in Qin’s research. Huang also explored the attitudes young Shanghai females hold to Mandarin and Shanghainese. She found that the young females considered the standard variety may increase social distance between native speakers of the dialect while using Shanghainese shows a close relationship. And it is noted that about 60% of the subjects reported to be more comfortable with using Shanghainese, but over 60% of the subjects reported they actually used Mandarin more often than Shanghainese. Aside from the gender issue, a report (2006) from the Department of Education in China shows the age may also correlate the use of dialects. The report found that only 37% of the local students in the 5th grade were able to speak Shanghainese, but 70% of the local college students were able to speak Shanghainese.

Due to historical, socio-economic and geographical reasons, the prestige of Cantonese in Guangzhou may differ from that of Shanghainese in Shanghai (Xiao, 2002). Tang (2006)
investigated the attitudes that 554 students in middle schools of Guangzhou held towards Mandarin and Cantonese. Through questionnaires and follow-up interviews and observations, Tang found that the subjects regarded Cantonese higher than Mandarin, no matter in terms of affection connections or social prestige. The subjects considered people who speak Cantonese are more friendly and approachable than people who speak Mandarin; they also assumed that people who speak Cantonese have received better education and get better pay than people who speak Mandarin. These attitudes contradict the common views which consider Mandarin the higher variety in China (Lin 1998; Zhan 1997 & 2000). Tang concluded that it was because many migrants in Guangzhou who speak the lingua franca (Mandarin) are low working class, while Cantonese people as a whole are wealthier than people from other regions. However, Tang also pointed out that the subjects from the community where most middle-class newcomers resided thought highly of Mandarin in terms of social prestige. The gender of the subjects and the occupations of the subjects’ parents were reported to be an irrelevant factor for the evaluations.

Another comparable survey was conducted by Wang & Ladegaard (2008). They examined the language attitudes and the reported use of Cantonese speaking students in Guangzhou. They asked 174 adolescents (age from 13-16) from a secondary school in Guangzhou to fill in a questionnaire about their language preferences and experiences. The results show that about half of the subjects whose native variety is Cantonese prefer to speak Cantonese only and the other half whose native variety is Cantonese as well prefer to speak both varieties. In consistent with previous studies, the “both” subjects stated that they would
prefer to use Mandarin in formal contexts such as in class, and they are more inclined to use Mandarin when talking with strangers or people who don’t speak Cantonese. In informal contexts, such as at home, after class and talking to friends who also speak Cantonese, the “both” subjects preferred the dialect. Although most of the subjects whose native variety is Cantonese reported speaking Cantonese at home, about 16% of them reported speaking Mandarin at home with their parents who are native speakers of Cantonese. The author attributed this to the invasion of Mandarin into local families because of the promotion of Mandarin by the government. The study also examined the correlation between gender and language use, finding that female students prefer to use the standard, high prestige variety (Mandarin), while male students prefer to use the vernacular (Cantonese). The result is consistent with the gender differences in previous studies (Eckert, 2003; Milroy, 1980; Trudgill, 1972).

The Present Research

In the previous sections, I reviewed the studies on Cantonese and Shanghainese. The two dialects show distinctive features from the standard variety (Mandarin) in terms of vocabulary, although they share words with Mandarin to a large extent. The previous literature shows that special dialect writings can be generally categorized into four types and mentions that the written forms of the two dialects are still under development. The sociolinguistic studies in this regard indicate the use of the dialects in the speech community correlates with many variables such as context, gender, education and social class. However, when a speech community is moved onto the non-physical internet, are there regional
characteristics still couched in the communications?

As an attempt to fill the gap in the study of using regional dialects on the internet, the present study examines the use of Cantonese and Shanghainese on BBS (Bulletin Board System) by doing a questionnaire survey and analyzing the texts drawn from the BBS posts.

Three research questions are addressed in general. The first question is how people write Cantonese and Shanghainese online when the phonetic cues are mute. I will break down the texts of dialect-written posts in terms of dialect writings and lexicon. My second question is whether the use of Cantonese and Shanghainese on the internet correlates with other variables. I will examine the observed data and the questionnaires for the social and discourse variables that may correlate with the use of dialect online. The social variables involve gender, age, residency, occupation, education level, and years outside the regions. The discourse variables consist of topic, forum type and discussion participants. Thirdly, I would like to know why the dialect speakers use dialect online. I will explore the reasons for the use of dialects and the attitudes they hold towards using dialects online.

In the following chapter, the methodology of the research is discussed, including a description of data, instruments, data collection, and data analysis.
Chapter Two
Methodology

This chapter outlines the methodology used in the present study. First, an overview of the data is presented, followed by a description of the research instrument and the rationale behind it. Second, there is a brief description of data collection. A justification of the method adopted in the present study is presented as well. Finally, the author analyzes the data according to the research questions outlined at the end of the first chapter. The problems which were encountered during the data analysis are also addressed.

Data

The raw data used in the present study consists of two parts: the data from internet observation and the self-reported data from a survey of internet users about the writing of dialects online. The decision of collecting two types of data lies in two reasons. First, there may be potential invalidity resulting from the self-reported survey data. What self-reporters report are information they are aware of, but they may ignore the unconscious production of their speech. To examine how the dialect speakers write their speech online requires direct internet observation on the real use. Secondly, only the observed data may not be able to address the research questions which ask about the speakers’ awareness of language use.

The data from internet observation includes the statistic figures from 30 internet
observations on the posts of four electronic forums over the time period from November 18th, 2008 to February 13th, 2009, as well as the dialect-written texts of 40,098 characters drawn from the internet observations. The statistic figures consist of the percentages of the dialect-written posts out of the most recent 100 posts (including new messages and replies) per day. The dialect-written texts are constituted of 20,011 characters in Shanghainese writing and 20,078 characters in Cantonese writing.

The data from survey are descriptive statistics extracted from 40 questionnaires, 20 collected from Cantonese speakers and 20 from Shanghainese speakers. The data covers social variables of gender, age, occupation, education level, residency, and years outside the regions, as well as the discourse variables of forum types, topic types and discussion participants. The way to write dialect online and the reasons and attitudes for the use of the dialects are also included in the data.

Instruments

The instrument employed in the study is a questionnaire about the use of dialect online. The questionnaire consists of 24 questions (see Appendix I). Questions 1-6 ask for the background information of the respondents, which are considered as sociolinguistic variables correlating with their use of dialect online. It includes age, gender, education, occupation, native dialect, residency, and years outside the regions. The reason why I include the seven variables in the present study lies in the findings of the significant correlation between the variables and the language choice from previous studies. The report issued by the Department of Education in China (2006) indicates that the number of the local students who
are able to speak the dialect dramatically increased from the 5\textsuperscript{th} grade to the college level. Wang & Ladegaard (2008) find that female prefer to use the standard variety (Mandarin) while male prefer to use the local vernacular (Cantonese). Studies on the language choice of migrants in big cities (Zhang, 2006; Cao, 2007; Fan, 2007; Qin, 2007; Tang, 2006) show that the higher the social status of migrants is, the longer the migrants stay in the cities of prestige dialects; the lower prestige the native dialect of the migrants is, the more likely the migrants would choose the prestige dialect (Cantonese or Shanghainese).

Questions 7-11 ask about the circumstance under which the respondents would use or not use dialect online. It explores three factors affecting the discourse of the net speak: the topics of the posts, the residency of interlocutors, and the type of the forum. Halliday et al. (1964) identifies three variables that correlate with discourse: field (the subject matter of the discourse), tenor (the participants and their relationships) and mode (the channel of communication, e.g. spoken or written). Adapted from the Halliday’s discourse model, I examine the regional characteristics of computer-mediated communication in terms of three dimensions: topics (namely “field”), the residency of interlocutors (namely “tenor” of regional characteristics), and the type of forums. The type of forums simulates the physical regions where people conduct daily communication in dialects. The forums or the subsections are titled with “region” or “people in (city)”, indicating the regional features of the discussion groups. Questions 12-14 examine the sociolinguistic awareness of net speakers, involving the reasons for the use of dialect online and the interaction among speakers. As I reviewed in the first chapter, the use of dialects may associate with local loyalty, social
identity, and the social functions in a “linguistic market” (Mesthrie, Swann, Deumert, & Leap, 2000).

Questions 15-19 look at the metalinguistic awareness of the respondents. It asks about the linguistic choice of the respondents to write dialect characters, how they identify the linguistic characteristics of the written dialect, and the venue they learn to write the dialects. Informants may write dialects with native characters, semantically borrowed characters, phonetically borrowed characters, or previously unattested forms of characters (Chen 1996, Lin 2003, Zhao 2007, & Matulewicz 2002). They may identify the speaking of dialects through the dialect writing, the featured lexicon of the dialects, the special sentence structures, or the content of the communication. Questions 20-22 in the questionnaire ask about the respondents’ code-switching behavior between the local dialect and the standard variety. Questions 23-24 ask about the views the respondents hold to the status of their dialect and the influence of internet on the dialects.

Data collection

I made 30 internet observations over posts from four forums every three days from November 18th, 2008 to February 13th, 2009. The four forums involved in the study are of two types: the forums for people nationwide in China and the forums for local people in Guangzhou and Shanghai only. The first forum is tieba.baidu.com, one of the most popular forums in China. The subsections of the main forum are organized by topics of economics, music, love, fashion, entertainment, TV shows, sports, computers, and regions. Under the “Regions” section, there are special forums for people in Guangzhou and people in Shanghai.
The second website is Mitbbs.com, the largest forum for Chinese people who stay overseas temporarily or permanently. The subsections of the forum are organized by topics of news, life, love, entertainment, sports, academics, and regions. Under the “Regions” section, there are special forums for Cantonese and Shanghainese. The third forum is Shanghaining.com/forum, one of the most popular forums among Shanghainese. The forum is based on topics too, but the topics only center on the life of Shanghainese, such as shopping, restaurants, computers, and flea market. The fourth forum is kugz.net, one of the most popular forums among Guangzhou people. The structure and the topics of this forum are similar to the third forum, except that it is for people in Guangzhou. The Internet observations were made over the 100 most recent posts from the “region” subsection under the two forums nationwide and the two local forums. The time when I viewed the posts was not fixed; it could be in the morning, in the afternoon or at night according to the US East Coast Timezone. But the internet observations on the four forums were made at nearly the same time in a day. Because of the practical constraints on my schedule, I did not do any observations from 0:00 to 8:00 a.m., which is an active period of time when people in China participated in forum discussions. The distribution of the times I made the internet observations is given in Figure 1.
Figure 1. The timeline of 30 internet observations at US East Coast Timezone.

A total of 18,000 posts were saved as plain text in my personal computer and dated. Posts only comprising facial icons or function words for mood expression, such as ‘Ah!’ 哈哈 (‘laugh online’) and 嗯 (‘Hmm...’), were not counted as a post. The posts of this type contained minimum information for analysis and thus were excluded. Posts of commercial advertising were not counted, either. I will analyze the raw data of 18,000 posts according to my research questions in the data analysis section.

To collect data from the questionnaire survey, I first emailed 20 speakers of Shanghainese and 20 speakers of Cantonese based on my internet observations on the discussions in the forums. The prospective subjects are those who actively participated in discussion in the dialect writing. The pool of prospective subjects covers a range of people in different age, gender, education, occupations, and current status of residency. In the email, I
asked for their consents on participating in my survey. However, only 8 Shanghainese and 5 Cantonese consented to do the survey in the first round of the data solicitation. I emailed my questionnaires to the respondents and kept recruiting more participants until I received consents from 20 Cantonese speakers and 20 Shanghainese speakers. The participants were asked to fill up the questionnaire I described in the section of “Instrument” and email back to me after they finish. In this way, I collected 20 completed questionnaires for the CMC in Shanghainese and 20 for the CMC in Cantonese.

Data analysis

The identification of dialect written posts. As described above, I have collected data from two sources: the internet observations on the discussion posts and the descriptive statistics from the questionnaires. Based on my research questions, I analyzed the two types of data in both the statistical and text-analytical ways. First, in order to see how people write dialects online, I separated the dialect-written posts from the pool of 18,000 posts. The identification of the dialect-written posts was conducted by two informants, one informant of Cantonese and one of Shanghainese. The two informants are the native speaker of Cantonese or Shanghainese and are able to read and write the dialects online. However, they cannot read or write each other’s dialect. For example, the Cantonese informant is able to read and write Cantonese but cannot understand Shanghainese no matter in writing or speaking. The two data coders are graduate students of Chinese linguistics in two universities in mainland China. They are literate in modern simplified Chinese but did not receive any formal schooling in dialect writing.
The two data coders at first were asked to review the 9,000 posts in their native dialects and identified the dialect-written posts. After they finished the screening, the two coders were asked to review the other 9,000 posts which might contain the dialect that is not their native dialect. They then identified the posts that sound strange to them. The purpose to include a native speaker of the dialects and a speaker who does not know the dialect in coding the data is to make sure the dialect-written posts are picked while the posts that do not contain the dialect writings are ruled out. The length of the posts was not relevant with the identification of dialect-written posts. In other words, no matter how short the post is, it would be considered as dialect-written posts as long as it contains the dialectal features of vocabulary or characters. The inter-rater agreement rate was computed for measuring the rate of time when two coders agree on the identification of a dialect-written post. The coders agreed 87.8% of the time on whether a post is Cantonese written or not, 71.4% of the time on whether a post is Shanghainese written or not. Coding disagreements were resolved through discussion. Examples of dialect-written posts are presented as follows.

Examples for Shanghainese-written posts:

(1)
个  则  店  哪能啊
one this shop  how [a]^3
‘How about this shop?’

(2)
半夜 额 sf

midnight [ŋə]\(^4\) sofa\(^5\)

‘(It’s already) midnight. I’m the first one to respond.’

Examples for Cantonese-written posts:

(1)

我之前 试过 翻 足 14 日 先 休息翻 一日

I before have tried work completely 14 days first take off one day

‘I have tried working for 14 days in a row and taking one day off.’

(2)

几 好 听 啊

Very good to listen [a]

‘(the music) is good.’

In the above examples of dialect-written posts, the first ones are those that contain feature words of the dialects. For example, 哪能 (‘how’) in the first sentence of the Shanghainese examples is not a word in Mandarin; the word how in Mandarin is written as 怎么样. 翻(工) (‘go to work’) in the first sentence of Cantonese examples is a nonsensical word in Mandarin; to go to work in Mandarin is written as 上班. The second sentences in the example posts are those that are partly intelligible for Mandarin speakers but involve special dialect writings. For example, the character 额 is a typical function word placed at the end of sentence in Shanghainese, indicating present perfect tense, while used as a noun in
Mandarin, meaning ‘forehead’ or ‘quantity’. The character 几 is a typical adverb in Cantonese, meaning ‘very’, while used as quantifier in Mandarin, meaning ‘several.’

The following examples are the posts which were excluded as non-dialect written posts since no dialect writing and no lexical features are involved. The non-dialect speakers who speak Mandarin proved to have no difficulty understanding these samples and did not find them sound strange.

Examples for posts which are not considered Shanghainese or Cantonese written:

(1)
哈哈 我喜欢那个 SONY 的 SZ 系列的一个顶级配置的
HAHA... I like that one Sony SZ series a top configuration
“HAHA...I like the one with top configuration in the Sony SZ series.”

(2)
是挺 有意思
is really interesting
“(This) is really interesting”

(3)
有钱 拿 就好 啊
have money to get just good [a]
“On the condition that you can be paid.”
The analysis of dialect characters. The pool of the dialect written posts served as a basis for the exploration of my three research questions. For the analysis of dialect writing and lexicon, I drew the posts each of which are over 50 characters out of the pooled dialect-written posts. I consider a post of over 50 characters is a reasonable length for text analysis. In total, 381 posts of 20,011 characters in Shanghainese writing and 316 posts of 20,078 characters in Cantonese writing were used as my corpus for the study of the linguistic dimensions of the online communication. According the previous study of this type (Qin, 2007), texts of about 40,000 characters are considered a reasonable amount for analysis. To examine the type of dialect writing people most frequently use online, the two informants who identified the dialect-written posts were asked to categorize the dialect writings into five types: native characters, semantically borrowed characters, phonetically borrowed characters, previously unattested forms of characters and common characters. The definitions of the first four types of dialect writing were adapted from the previous literature presented in Chapter One. The native characters are the dialect characters that can be explained for their origins in the dialectal dictionaries. Two dictionaries which provide detailed verifications of native characters in ancient literature were employed by the two coders as tools for identifying the native characters. The two dictionaries used are Shanghainese Dictionary (2007), which gives explanation of the Shanghainese native characters, and Native Characters in Cantonese (1995), which contains 1,405 native characters in Cantonese. To date, the two dictionaries are the most frequently used dictionaries by researchers on native characters in the regards.

The phonetically borrowed characters are the dialect characters that have similar or the
same pronunciations as in Mandarin but bear different meanings. The semantically borrowed characters are the dialect characters that share the similar meaning as in Mandarin but differ from Mandarin characters in terms of pronunciations. The dialect characters, which are included neither in modern Chinese characters nor in the dictionaries of the native characters, are considered the previously unattested forms of characters. Aside from the four types of dialect characters, there are common characters in dialect written posts. The common characters are the dialect characters which share the same or similar meaning and pronunciations with the characters in modern Chinese (Lin, 2003; Dong, 2005). The pool of 20,011 Shanghainese characters and 20,078 Cantonese characters were categorized into the five types, and the percentage of each type of dialect writing over the total amount of characters was counted and plotted. Examples of the five categories for the dialect writings are presented below. Dialect characters which were targeted for analysis were highlighted in squares.

Examples for Shanghainese writing:

<table>
<thead>
<tr>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>起来 [loə] in Shanghainese; N/A in Mandarin</td>
<td>‘get up’ in Shanghainese; N/A in Mandarin</td>
</tr>
<tr>
<td>Character</td>
<td>Pinyin in Shanghainese;</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>揮</td>
<td>[o]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>发嘅</td>
<td>[hɤ]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantically borrowed characters</td>
<td></td>
</tr>
<tr>
<td>吓人</td>
<td>[hɔ]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>鬼话</td>
<td>[tɕy]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>眼</td>
<td>[ŋe]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonetically borrowed characters</td>
<td></td>
</tr>
<tr>
<td>落起来</td>
<td>[loa]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>白相</td>
<td>[bɔ][ɕiaŋ]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>交关</td>
<td>[tɕia][kue]</td>
</tr>
</tbody>
</table>
Previously unattested forms of characters

| 嘔答答 | [gə] in Shanghainese; N/A in Mandarin | ‘apathetic’ in Shanghainese; N/A in Mandarin |
| 水淋溚渧 | [di] in Shanghainese; N/A in Mandarin | ‘dripping’ in Shanghainese; N/A in Mandarin |
| 剛 | [na] in Shanghainese; N/A in Mandarin | ‘you’ (pl.) in Shanghainese; N/A in Mandarin |

Common characters

| 中飯 | [tsɔŋ] [ve] in Shanghainese; [tson][tʰan] in Mandarin | ‘lunch’ in Shanghainese and Mandarin |
| 便宜 | [bi][ɲi] in Shanghainese; [bʰiæn][ji] in Mandarin | ‘cheap’ in Shanghainese and Mandarin |
| 办公 | [be][koŋ] in Shanghainese; [ban][kun] in Mandarin | ‘working in an office’ in Shanghainese and Mandarin |

Examples for Cantonese writing:

<p>| 虎羅 | [pʰə][lɔ] in Cantonese; | ‘most cherished things’ in |</p>
<table>
<thead>
<tr>
<th>晃</th>
<th>[tɕœk] in Cantonese; N/A in Mandarin</th>
<th>‘to cook’ in Cantonese; N/A in Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>立</td>
<td>[ɡei] in Cantonese; N/A in Mandarin</td>
<td>‘to stand’ in Cantonese; N/A in Mandarin</td>
</tr>
</tbody>
</table>

**Semantically borrowed characters**

| 看 | [tʰei] in Cantonese; [kʰan] in Mandarin | ‘look’ in Cantonese and Mandarin |
| 系 | [hei] in Cantonese; [ei] in Mandarin; | ‘is’ in Cantonese and Mandarin |
| 学校 | [ho : ] in Cantonese; [cyi] in Mandarin | ‘study’ in Cantonese and Mandarin |

**Phonetically borrowed characters**

Source: Beishi (Zhan, 1995, p. 25)
| 波罗 | [pʰɔ][ lɔ] in Cantonese; [bo][luo] in Mandarin | ‘most cherished things’ in Cantonese; N/A in Mandarin |
| 白灼 | [tɕoek] in Cantonese; [tʂuo] in Mandarin | ‘to cook’ in Cantonese; ‘burn’ or ‘bright’ in Mandarin |
| 仲未 | [tʂʊŋ] in Cantonese; [tʂun] in Mandarin | ‘still’ in Cantonese; ‘the second’ in Mandarin |

Previously unattested forms of characters

| 嘞 | [kei] in Cantonese; N/A in Mandarin | ‘of’ in Cantonese; N/A in Mandarin |
| 唔 | [hɔi] in Cantonese; N/A in Mandarin | ‘at’ in Cantonese; N/A in Mandarin |
| 冇 | [mou] in Cantonese; N/A in Mandarin | ‘not’ in Cantonese; N/A in Mandarin |

Common characters

| 法律 | [faːt][leːt] in Cantonese; [fA][ ly] in Mandarin | ‘law’ in Cantonese and Mandarin |
| 可以 | [hɔ][ji] in Cantonese; [kʰs][ ji] in Mandarin | ‘can’ in Cantonese and Mandarin |
To compare with the results from the internet observations, descriptive statistics from Question 16 and 18 in the questionnaire survey, which asked the participants how they write the dialects online, was analyzed for the dialect writing as well. The frequency of the reported use for each type of the dialect writings was computed and plotted.

*The analysis of dialect lexicon.* For the analysis of lexicon of the dialect-written posts, the two informants and one Mandarin speaker who knows nothing about Cantonese or Shanghainese were employed to read through the dialect-written posts that involved the dialects they don’t know. By reading through the data over and over and discussing with me, a coding scheme was agreed upon for the vocabulary. The vocabulary used in the posts was roughly categorized into three categories: common words, partly intelligible words and feature words. Examples of a Cantonese-written post and a Shanghainese-written post are presented in the following and analyzed for the three categories of lexicon. The two dialect-written posts are provided in four versions: the dialect writing, the translation in modern Chinese, the word-for-word translation in English and the sentence-for-sentence translation in English. Each of the posts includes eight sentences. Words, equivalent to *ci* in Chinese, are separated by slash.

The table:

<table>
<thead>
<tr>
<th>用品</th>
<th>[joŋ][p̂̚n] in Cantonese; ‘supplies’ in Cantonese and Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[jun][b̚in] in Mandarin</td>
</tr>
</tbody>
</table>

The Cantonese post:
Cantonese (C): 真系 / 五 / 知道 / 可以 / 讲 / D / 咩野 / 好

Mandarin (M): 真是 / 不 / 知道 / 可以 / 讲 / 点 / 什么 / 好

Really/ don’t/ know/ can /say/ a little/ what/good

‘Really don’t know what can be said in a little better way’

(2)

C: 毕竟 / 都 / 系 / 自己 / 咪 / 啊 / 妈

M: 毕竟 / 都 / 是 / 自己 / 的 / 妈 / 妈

anyway / still / is / my / mom

‘Anyway, (she is) still your mom’

(3)

C: 就算 / 几难 / 顶 / 都 / 要 / 顶

M: 就算 / 很难 / 受 / 都 / 要 / 受

Even / very hard / to take / still / need / take

‘Even (it is) hard to take (her as your mom), (you) still need to take’

(4)

C: 好彩 / 我 / 啊 / 妈 / 系 / D / 搞笑 / 人物

M: 幸好 / 我 / 妈 / 妈 / 是 / 个 / 搞笑 / 人物

Luckily / my / mom / is / an / amusing / person

‘Luckily, my mom is an easy-going person.’

(5)
C: 我/从来/都/五/当/佢/系/我/啊
M: 我/从来/都/不/当/她/是/我/妈妈

I/for ever/all/no/treat/ her/is/my/mom.

“I never treat her as ‘mother’ seriously.”

(6)

C: 我/当/佢/就/好似/知心朋友/甘样！
M: 我/当/她/就/好像/知心朋友/那样！

I/treat/ her/ just/as/close friend/ that

“I treat her just like my close friend.”

(7)

C: 有/咩/野/都/会/同/佢/倾,
M: 有/什么/事/都/会/同/她/聊,

There is/any/thing/all/can/with/her/talk

“If there is anything, I will talk with her.”

(8)

C: 所以/我地/感情/好好。
M: 所以/我们的/感情/很好。

So/our/relation/very good.

“So our relation goes very well.”

The Shanghainese post:
“I am not the little girl who is still looking for a short term relation.”

“So the guys who are not serious about relationship are asked to stop reading at this line.”

“I just want to find a long-term stable relationship.”

“I don’t want a guy who is too poor”
S: 发 /要 /特 /丑 /特 /矮，

M: 不/要 /特别/ 特 /丑 /特/ 矮的

No /want /too /ugly /too /short

‘I don’t want a guy who is too ugly or too short.’

(6)

S: 外地人 /我/ 不 /要的，

M: 外地人 /我/ 不 /要

Outsiders/ I/ don’t/ want

‘I don’t want outsiders.’

(7)

S: 其他 /么 /大/ 要求，

M: 其他 /没啥 /大/ 要求

Others /no/ big/ requirements

‘I don’t have other strict requirements.’

(8)

S: 最重要点 /就是/ 全心全意 /爱 /无饿。

M: 最重要的 /就是/ 全心全意 /爱 /我。

Most important /is / soul and heart/ love/ me

‘The most important thing is to love me soul and heart.’

Common words are the words shared by Mandarin and the dialects. The common
words in dialect writing bear the same forms and the same meaning as in Mandarin, such as 知道 (‘know’), 全心全意 (‘soul and heart’), 知心朋友 (‘close friends’), and 外地人 (‘outsiders’) in the post examples above. Partly intelligible words are the words of which the main parts are written in the same form and have the same meaning as in Mandarin, and yet partially include the lexical features of the dialects. For example, 几难 in the Cantonese written post includes the main part 难 (‘difficult’) as in Mandarin, but also employs a Cantonese featured quantifier 几 (‘very’) to modify 难, which makes itself partially understood. Feature words are the words that are only used by the speakers of the dialects. Even though they bear the same written forms as in Mandarin or not, the words do not represent anything in Mandarin. For example, 好彩 in the Cantonese post means ‘luckily’, but cannot be interpreted the same way as in Mandarin. It can only be interpreted as two separate words 好 (‘good’) and 彩 (‘colorful’) in Mandarin. 哈来来 in the Shanghainese post makes no sense as in Mandarin although it bears plausible written forms in modern Chinese. The Cantonese native character 佢 is also a feature word, which is not included in modern Chinese.

The data coders were asked to categorize the words into the three categories. A repeated word was counted only once as a type. The inter-rater agreement rate was computed for measuring the rate of time when two coders agree on the categorization of a dialectal word. The coders agreed 82.7% of the time on whether a word is a common word or a Cantonese word, 77.4% of the time on whether a word is a common word or a Shanghainese word. Coding disagreements were resolved through discussion.
The analysis of the use of dialects online in sociolinguistic contexts. When it comes to my second research question which asks about the correlation between the use of dialect online and the social and discourse variables, I analyzed the data from both the questionnaire and the internet observations. Since no social variables, such as gender, age, residency, occupation, education level, and years outside the regions, can be collected from online internet observation over the posts, I have to rely on the reported data from questionnaire only. Since all the participants recruited in the survey reported they used the dialects for computer-mediated communication, the frequency of the use of the dialect online by each social variable (Question 1 to 4 and 6) was computed through SPSS frequency analysis and plotted.

For the analysis of the correlation between the use of dialects online and the discourse variables such as topic, forum type and discussion participants, the descriptive statistics, such as frequency, of the three variables from the questionnaire were computed and plotted through SPSS procedures. The three variables were covered in Question 7 to Question 11 in the questionnaire.

In order to see the correlation between topic types and the use of dialect online based on the internet observations, I had intended to compute the percentage of dialect-written posts over 100 most recent posts under each type of topics. However, there was a rub underlying the data collection. The pool of posts, including dialect-written and non-dialect written posts, were categorized by different topics from forum to forum. For example, the forum for Chinese who live overseas categorizes posts under more general topics, such as
news, life, love, entertainment, sports, academics and regions, while the Shanghainese forum divides the topics into more specific aspects of life, such as shopping, restaurants, computers, and flea market. This may result in a wide range of topic types which would be difficult and meaningless to process. Therefore, I relied on the results from the questionnaire data only, in terms of the correlation between topic types and the use of dialect online.

In order to look at the correlation between forum type and the use of dialect online, I computed the percentage of the dialect-written posts over every 100 most recent posts per day in three different types of forum. For each dialect (Cantonese and Shanghainese), three datasets of 30 internet observations were set up, including the percentage of the dialect-written posts in the forum nationwide in China, the percentage of the dialect-written posts in the forum for Chinese people who live oversea, and the percentage of the dialect-written posts in the local forum. A one-way ANOVA analysis in the SPSS was conducted to see the difference among three types of forum.

For the analysis of the correlation between the variable of participants and the use of dialect online, I had planned to collect data about participants’ variables, such as, the residency or the dialect he/she speaks. However, the data cannot be collected from direct observations on the internet because I cannot tell the residency or the dialect a participant speaks only by looking at what he/she writes. For example, people who speak the dialect may not be able to write the dialect online and accordingly they present themselves in the same way the people who cannot speak the dialects do. Therefore, I will present the results from the questionnaire data only, which asked the respondents for their background information.
To address the third research question which asks for the reason why people write dialects online and their attitudes, I computed the descriptive statistics for the data from the questionnaire (Question 12, 23 and 24) and the results were plotted. Since the research question inquires about the sociolinguistic awareness of the participants, data cannot be collected from internet observations, and I can rely on the reported introspection only. As gender proves to be an important variable correlating with the use of dialects (Wang & Ladegaard, 2008; Huang, 2006), I also cross-examined the gender variable with the psychological reasons for the use of the dialects through post procedures.

In Chapter 1, the generative linguistic and sociolinguistic studies on the two dialects were reviewed. In Chapter 2, I described the data, the instrument, the data collection and the data analysis. In the next chapters, the findings from the data analysis will be presented and discussed.
Chapter Three

Findings

Based on my research questions, what, how and why dialect speakers use dialects for compute-mediated communication, I will present my findings in three sections. In the first section, the use of dialect writing and dialectal vocabulary will be discussed. The second section examines the sociolinguistic variables that correlate with the use of dialects in general and the discourses variables from both observed and reported data. In the third section, I analyze the reason why the respondents in the survey used the written dialects online, and the attitudes they held to the use of the dialects online.

Dialect writing and lexicon

_Dialect writing._ My first research question asks about how people write Cantonese and Shanghainese online when the phonetic cues are mute. 316 posts of 20,078 characters in Cantonese writing and 381 posts of 20,011 characters in Shanghainese writing were analyzed by two informants in the two linguistic dimensions: dialect writings and dialect lexicon. Relevant answers from the questionnaire survey which involved a group of 40 dialect speakers were computed correspondingly. For the analysis of the dialect writings in the present study, the characters in the dialectal posts were categorized into five types (common characters, phonetically borrowed characters, semantically borrowed characters, native characters,
characters and previously unattested forms of characters), and the percentage of each type of characters out of the total amount of characters in the posts was computed.

As Figure 2 showed for the Cantonese written posts in the present study, the common characters which have the same or similar pronunciation and meaning in Mandarin and Cantonese accounted for the largest proportion out of the total amount of the pool of characters (78.3%), while the native characters which can be explained for their origins in dialectal dictionaries accounted for the least proportion, only 0.01%. The other three types of dialect writings, phonetically borrowed characters, semantically borrowed characters, and previously unattested forms of characters, were of comparable amount, although the
phonetically borrowed characters were used slightly more than the previously unattested forms of characters and the semantically borrowed characters in the Cantonese written posts of the present study.

For the Shanghainese written posts (Figure 3), the common characters took the largest proportion of the pool of characters as the same case in the Cantonese written posts, but the common characters in the Cantonese written posts were of slightly higher percentage than in the Shanghainese written posts. In contrast with the comparable proportion among phonetically borrowed characters, semantically borrowed characters and previously
unattested forms of characters in the Cantonese written posts, the phonetically borrowed characters in Shanghainese showed significantly larger proportion than the other two types of dialect writing. In terms of the previously unattested form of writing, it was more frequently used in the Cantonese-written posts than the Shanghainese ones, 6.44% as opposed to 1%.

**Figure 4. The reported use of Cantonese writing types**

![Bar chart showing the reported use of different types of Cantonese writing.](image)

Drawing on answers to Question 16 to 18 in the questionnaire survey, the frequency of the reported use of each type of dialect writing by the respondents in the present study was computed. As shown in Figure 4, the phonetically borrowed characters and native characters were reported to be used by most of the Cantonese respondents. The observed use of phonetically borrowed characters was consistent with the self-reported use, while the
respondents in the survey reported a much higher percentage of using native characters than the Cantonese speakers who were observed online. In addition, 30% of the respondents reported using Pinyin alphabets to transcribe the spoken Cantonese, but the use of this type of writing is limited to a minimum level according to previous literature (Chen, 1996) and thus it was not included in the observation.

Figure 5 showed that among the Shanghainese-using respondents, the phonetically borrowed characters were reported to be used by every respondent in the survey, and the percentage of reportedly using the type of writing was far higher than the other four kinds of writings. The overwhelming reported preference for the phonetically borrowed characters
among the Shanghainese respondents fits very well with the observed preference for phonetically borrowed characters over other types. In contrast with the reported use of native characters by 75% of the Cantonese respondents, the use of native characters was only reported by 25% of the Shanghainese respondents. The reported use of semantically borrowed characters and previously unattested forms of characters by the Shanghainese respondents was as low as by the Cantonese respondents, and in line with the observations on the two types of dialect writings. In addition, there was a certain percentage of the Shanghainese respondents reporting the use of Pinyin alphabets as well as the Cantonese respondents in the survey.

In addition to the descriptive statistical analysis of the observation and the questionnaires in the present study, we can draw some generic conclusions upon the dialect writings from the qualitative analysis on the electronic postings. First, the use of the phonetically borrowed characters, which have similar or the same pronunciations as in Mandarin but indicate different meanings, was most frequent yet inconsistent in the dialect written postings. For a single representation of speech sound and semantic meaning in the dialects, there were usually found two or more variations of the written forms in the postings. For example, [le] (‘to come’) was found to be written as 黎 or 离 in the Cantonese postings, while in Mandarin 黎 [li] means ‘dawn’ and 离 [li] means ‘to leave’. [gә] (‘this’) was found to be written as 咯 or 各 in the Shanghainese postings, while in Mandarin 咯 [gә] imitates the sound of giggling and 各[gә] refers to ‘every.’

With regard to the native characters, this type of writing was seldom used by the online
writers in the study even if there were standard or established written forms available. For example, [da bjɛ lu] (‘eat around a hot pot’) in Cantonese was written as 打边炉 in most cases (打 for [da], 边 for [bjɛ], and 炉 for [lu]), and nobody in the present study used the original form 甂 for [bjɛ] (Zhan, 1995, p. 29). In the Shanghainese postings, [loʊ tɕ i le] (‘get up’) was usually written as 落起来 (落 for [loʊ], 起 for [tɕ i], and 来 for [le]), and nobody in the present study used the native character 落 for [loʊ] (Qian, 2007, p.133).

By observing the previously unattested forms of characters in the online postings, it was found that the characters were usually formed by attaching a radical which indicates meaning to a phonetic morpheme which indicates pronunciation. For example, the radicals such as 口 (‘mouth’), 亻 (‘person’), 氵 (‘water’) or 扌 (‘hand’) were added to many characters in written Chinese to form a new character in the dialects. For example, Cantonese speakers added 口 (‘mouth’) to 甘 [gan], 麦 [mai], and 太 [tai] in written Chinese, making new characters 啻([gəm], a functional word indicating doubt), 啪 ([mjɛ], a functional word indicating questions), and 啉 ([tɛ] ‘tie’).

*Dialect lexicon.* The Cantonese-written posts of 20,078 characters and the Shanghainese-written posts of 20,011 characters were also analyzed in terms of lexicon. They were categorized by coders into three types: common words, partly intelligible words and feature words. Common words are the words written in the same graphic form and representing the same real-world referent in both Mandarin and the dialects. Partly intelligible words are the words of which the main parts are written in the same form and represent the same real-world referents as in Mandarin, while other parts consist of dialect
characters which refer to different meanings from Mandarin. Feature words are the words that are only used by the speakers of the dialects. No matter the feature words have the same written forms as in Mandarin or not, the words do not refer to the same representation in meaning as in Mandarin. The percentage of each category in the dialect-written posts was computed and plotted.

Figure 6. The types of vocabulary in Cantonese written posts
As shown in Figure 6 and 7, the common words accounted for the largest proportion of the whole vocabulary no matter in the Cantonese written posts or in the Shanghainese written posts. The percentage of the common words was around 60% for the two dialects written posts. The feature words accounted for a certain amount of the total lexicon in the two dialectal posts, about 26.9% for Cantonese, 32.2% for Shanghainese. The partly intelligible words took the least proportion out of the entire lexicon pool in the present study.

As pointed out by many researchers (Li & Wu, 2001; Zhu, 2004; Wang, 1935; Xu & Tang, 1981; etc.), the lexicon of a dialect is notable because of the feature words. A close-up look at the feature words was made to the written dialectal pool. In terms of dialect writing,
the feature words identified by the data coders usually fell onto the categories of native
characters, phonetically borrowed characters and previously unattested forms of characters.
The semantically borrowed characters, which bear the same meaning and the same graphic
form in both Mandarin and in the dialects, were not considered as feature words but usually
categorized into common words. For example, to look was referred as 看 in some Cantonese
written postings. Although 看([tʰɐi]) in Cantonese is pronounced differently from its
Mandarin counterpart [kan], it represents the same meaning ‘to look’ and bears the same
written form 看 in both Mandarin and Cantonese. Therefore, it was considered no difference
conceptually and was categorized as a common word.

Feature words in the written dialectal posts not only included the words that are in
quite different graphic forms from the modern Chinese words, but also included the words
that are in the same graphic forms as in Mandarin and yet refer to different meanings. It was
easier to identify the feature words which are not included in modern written Chinese
characters and unique to the dialects, such as native characters and previously unattested
forms of characters. The Mandarin coders in the present study can easily identify words such
as 啊 (‘that’), 啊 (a suffix, indicating plural nouns), 撈 (‘look for’), and 疲 (‘tired’) as
feature words in Cantonese, or 葛咾 (‘therefore’), 哎 (a function word, indicating
questions), 你 (plural noun of ‘you’), and 囫 (‘hide’) as feature words in Shanghainese. In
addition, feature words also included the characters that share the same written forms as in
modern written Chinese characters but represent a different meaning or refer to a certain idea
which does not exist in Mandarin. For example, the wife of boss was referred as 事头婆
‘boss-wife’ in Cantonese postings, while in Mandarin the graphic form can be only interpreted as ‘thing-head-woman’ which does not represent anything. To understand the implied meaning of an utterance was referred as 接铃子 in Shanghainese postings, while the literal meaning of the three characters in Mandarin is ‘pick-ring-son’, which does not make any sense.

The sociolinguistic variables and the use of the dialects online

Social variables and the use of the dialects online. For the analysis of the correlation between the use of dialect online and the social variables, the frequency of using dialects online was counted by each social variable from the questionnaire data.

Figure 8. The use of dialects online by age

Among the participants who used the dialects online in the present study, people aging
from 21 to 30 years old accounted for the largest proportion, while people older than 30 years old only accounted for the least proportion. People who were younger than 20 years old also showed great interest in using the written dialects online. The results showed that people who were younger than thirty years old in the survey are more likely to write the dialects online.

As shown in Figure 9, among the participants who used the dialects online in the present study, females were far more than males. 62.5% of the participants who used the dialects online in the survey were females, while only 37.5% were males.
As shown in Figure 10, among the participants who use the dialects online in the present study, 77.5% have received or are receiving higher education, no matter in four-year universities or at three-year postsecondary colleges. Only 5% of the dialect speakers involved in the survey was at the education level of junior high school or below. The result showed that the respondents who received higher education were more likely to write the dialects online.
As shown in Figure 11, among the participants who use the dialects online, 90% were engaged in an intellectual type of occupations, while 10% of the dialect writers reported they were working non-intellectually. It indicated that the occupation type correlated strongly with the preference for writing the dialects online in the survey.
As shown in Figure 12, among the participants who used the dialects online, 67.5% were local residents in Guangzhou or Shanghai, while 32.5% were non-local residents but were able to speak the dialects.
As regards the living status in the local areas, 52.5% of the dialect online writers in the present study have been living in the local area for their whole life, while 47.5% have lived outside the region for a certain period of time. Among those who have lived in outer areas, 63.1% have stayed outside for over five years. The result indicated the use of the dialects online did not decrease as the period of time when the dialect speakers live outside increased; it showed an opposite trend.

*Discourse variables and the use of the dialects online.* For the analysis of the correlation between the use of dialects online and the discourse variables, such as topic types, website types and participant type, the frequency of the use of dialects online by each variable was computed from the survey data and plotted.
As shown in Figure 14, 67.5% of the respondents participating in the survey reported that they would choose to write the dialects on certain types of topics. Among the 67.5% respondents, about a half chose to use the written dialects for computer-mediated communication when they “talk” on topics regarding Shanghai or Guangzhou. About 30% of the topic-wise respondents chose to write in dialect writings when the topics of the conversations involved different themes, such as economics, politics, life, and entertainment. The results indicated that the respondents in the survey had a tendency to use their dialects online when they were involved in the topics regarding the local issues or a certain type of themes.
As shown in Figure 15, among all the participants in the survey, 82.5% reported they would choose to use the written dialects on a certain type of website or forum. Among the forum-wise respondents, 60% reported to use their dialects on a forum where most members are local residents, and 67.5% reported they preferred a regular website where a section is devoted to the local issues. The results showed that the participants in the survey had a strong tendency to use the dialects on a certain type of website which relates closely to regional issues.

The correlation between the use of dialects online and the website type was also examined through observations on the posts with written dialects in different types of forums.
A one-way variance analysis (ANOVA) in the SPSS was conducted to see the difference among the mean percentage of the dialect-written posts over every 100 most recent posts for 30 days in three different types of forum. For Shanghainese, the dependent variable was the percentage of the Shanghainese written posts over 100 most recent posts per day. The independent variable was the type of forum at three levels: a region forum in a regular website in China, a region forum in a regular website in oversea, and a local forum. The results was significant, $F(2, 87) = 163.09$, $p < .001$ ($M_{\text{regular, cn}} = 7.93$, $M_{\text{regular, oversea}} = 4.50$, $M_{\text{local}} = 23.33$). The strength of relationship between the percentage of Shanghainese written posts and the type of websites, as assessed by $r$, was strong, with the website type factor accounting for 68.2% of the variance of the dependent variable. Therefore, with 95% confidence, we can conclude that the percentages of the Shanghainese-written postings in different types of forums are different. There is a positive, significant correlation between the percentage of Shanghainese written posts and the type of websites.

Fisher’s least significant difference (LSD$^6$) post hoc test was conducted to see the pair-wise differences. The results showed there was a significant difference in the means between the percentage of Shanghainese written posts from a local forum and that from a region forum in a regular website in China. There was also a significant difference in the means between the percentage of Shanghainese written posts from a local forum and that from a region forum in a regular website oversea. In addition, there was a significant difference in the means between the percentage of Shanghainese written posts from a region forum of a regular website in China and from a region forum of a regular website oversea. A
visualized result is provided in Figure 16. In the boxplots, the whisks indicate the range of the data points in the three datasets, the boxes indicate the range from 25% to 75% of the dataset, and the middle lines in the boxes indicate the medians. In Figure 16, the boxplot of the local forum is way higher than that of the region forums in regular websites. In conclusion, we can tell with 95% confidence that the percentage of Shanghainese-written posts is higher in the local forum than in region forums of regular websites.

**Figure 16. The Percentage of Shanghainese written posts in three website types**

For Cantonese, the dependent variable was the percentage of the Cantonese written posts. The independent variable was the type of forum at three levels: a region forum in a regular website in China, a region forum in a regular website in oversea, and a local forum.
The results were significant, $F(2, 87) = 970.98$, $p < .001$ ($M_{\text{regular.cn}} = 7.93$, $M_{\text{regular.oversea}} = 12.37$, $M_{\text{local}} = 57.30$). The strength of relationship between the percentage of Cantonese written posts and the type of websites, as assessed by $r$, was strong, with the website type factor accounting for 88.4% of the variance of the dependent variable. A LSD post hoc test was conducted to see the pair-wise differences. The results show there was a significant difference in the means between the percentage of Cantonese written posts from a local forum and that from a region forum in a regular website in China. There was also a significant difference in the means between the percentage of Cantonese written posts from a local forum and that from a region forum in a regular website overseas. In addition, there was a significant difference in the means between the percentage of Cantonese written posts from a region forum of a regular website in China and from a region forum of a regular website overseas. A visualized result is provided in Figure 17. In Figure 17, the boxplot of the local forum is way higher than that of the region forums in regular websites. In conclusion, we can tell with 95% confidence that the percentage of Cantonese-written posts is higher in the local forum than in region forums of regular websites, and there is a positive, significant correlation between the percentage of Shanghainese written posts and the type of websites.
Both the reported use and the observed use of dialects online show that, among the three types of websites, the dialect speakers write their dialects more often in the local forums where most members are local residents than in the region forums of regular websites members of which are not restricted to Guangzhou or Shanghai. The observed statistics show that Cantonese speakers write Cantonese more in the region forum of the regular website oversea, while the Shanghainese speakers write Shanghainese more in the region forum of the regular website in China.
As shown in Figure 18, among the three discourse variables examined in the present study, the type of the interlocutors contributed most to the use of the dialects online; 92.5% of the respondents were reportedly inclined to use the dialects online with particular types of people. Among those interlocutor-wise respondents, 90% indicated they would use the dialects online with local people, while 25% reported the use of the dialects with outsiders who lived in Guangzhou or Shanghai for a long time as well. It was noted that 15% of the participants mentioned that they would “talk” in the dialects online with the interlocutors who can speak the dialects no matter they were from local places or not.

*The reasons and attitudes of the use of the written dialects online*
To address the third research question about the reasons and attitudes of the dialect speakers write dialect online, the frequency of the reasons and the attitude options each respondent of the questionnaire survey indicated was computed and plotted.

The reasons of using the written dialects online. As shown in Figure 19, 52.5% of the respondents thought they used the dialects online because everybody else on the forum used the dialects, and in this speech community they need to speak the vernacular to communicate
with others. 55% of the respondents attributed the reason of using the dialects online to the regional identity. They believe by presenting their linguistic identity — the dialects, they were not only telling the audience where they were from, but also further reinforced the connection to the regions. 50% of the respondents indicated that by using the dialects online they can quickly shorten or create the social distance with the interlocutors.

For the post analysis of the gender variable correlating with the reasons why people use the dialects online, the frequencies of each reason each participant indicated were computed in SPSS. As shown in Figure 20, for the reason that the use of the dialects is out of the need to enter into the speech community of Cantonese or Shanghainese, females and
males did not show any difference. But for the reason that the use of the dialects is a way to express the regional identity, female respondents in the survey were 75% more than males voting for this reason; and women three times as many as men chose the reason that the use of the dialects can establish solidarity or create social distance.
The attitudes towards the use of the dialects online. For the analysis of attitudes people hold to the use of dialects online in public media, the percentage of each option for the attitude survey was computed. For the idea of having Cantonese news report online, 70.73% of the respondents agreed with it, while only 7.32% were against it. For the idea of having Shanghainese news report online, 64.29% of the respondents agreed with it, while only 10.71% were against it. There were a slightly higher percentage of Cantonese respondents who voted for the dialectal news report online than Shanghainese respondents.
Figure 23. The impact of internet on the promotion of Cantonese
For the analysis of the attitudes towards the internet and the promotion of the dialects, the percentage of each option for the attitude survey was computed. Among the Cantonese respondents, 63.41% considered there was a positive effect of the internet on the promotion of Shanghainese, while 2.44% considered the internet would counteract with the spread of Shanghainese. Among the Shanghainese respondents, 67.86% think there was a positive effect of the internet on the promotion of Shanghainese, while no one considers the internet would counteract with the spread of Shanghainese. In conclusion, around two thirds of the survey respondents agreed on the contribution the internet makes to the promotion of the dialect.
Conclusion

To the large extent, the present study of the dialect used for computer-mediated communication is an exploratory research. Findings about the way people use for communicating online with their dialects are complicating yet intriguing. First, people involved in the present study wrote the dialects online in five types of writings: the common characters, the phonetically borrowed characters, the previously unattested forms of characters, the semantically borrowed characters and the native characters. Except the common characters, the dialect speakers were most likely to use the phonetically borrowed characters for online communication. Through analyzing the texts of the dialect written posts, I found that there were some patterns of writing the dialects. In terms of the dialect lexicon, the most part of the dialectal vocabulary were shared with the lingua franca (Mandarin), while the feature words that only exist in the dialectal region accounted for a certain part of the vocabulary. We can also see that the feature words were usually written in phonetic borrowed characters, previously unattested forms of characters or native characters.

Secondly, social variables, such as gender, age, and occupation, correlated with the use of the dialects online in the present study. For example, well educated young females with intellectual occupations were most likely to use dialects for CMC according to the results in the present study. In addition, the dialect speakers in the present study used the dialects more often on local forums with people who were able to write the dialects as well. Thirdly, people attributed the use of the written dialects online to the need of entering a speech community, while female respondents in the survey considered the use of the dialects online was a way to
express their regional identity and can establish solidarity or create social distance from others. For the survey of the attitudes towards the regional dialects online, most people in the survey advocated the use of dialects for public media online and believed the internet contributed to the promotion of the dialects. In the next chapter, the findings will be discussed and analyzed for implications.
Chapter four

Discussion, Conclusion and Implications

Discussion

Since computer-mediated communication entered the life of millions of internet users, sociolinguists have brought studies on social variables into this new frontier, such as the gender difference on the internet (Panyametheekul & Herring, 2003; Rossetti, 1998; Herring & Paolillo, 2006; Savicki, Kelley & Oesterreich, 1999), the net-speak of different age groups (Huffaker, & Calvert, 2005; Rodgers & Gauntlett, 2002), and the online gaming language (Driscoll, 2008; Herring, et.al. 2009). However, as an important characteristic of people’s social identity, use of regional dialects online has not been well studied. As an exploratory research, the present study touches upon the broad issues around regional dialects (Cantonese and Shanghainese) and the computer-mediated communication. Three research questions are addressed in the research. In what ways do speakers of Cantonese and Shanghainese write their dialects for computer-mediated communication? How do the dialect speakers use the written dialects online? Why do the dialect speakers choose to communicate in their dialects online? In the following three sections, the findings reported in the last chapter will be discussed correspondingly by the research questions, and limitations of the present study will also be addressed.
Dialect writing and lexicon. To address the first research question, the dialect characters and the dialect vocabulary were studied through observing the dialect written posts and surveying the dialect speakers online. The findings about the dialect writing shows that people write the dialects online by using five types of characters: the common characters, the phonetically borrowed characters, the previously unattested forms of characters, the semantically borrowed characters and the native characters.

The common characters proved to be used statistically most by the dialect speakers for online communication in the present study even though the dialect writers were not aware of it. This is probably due to the compulsory study of modern simplified Chinese in mainland China since 1949. As we have discussed in Chapter one, China has long history of implementing one written language policy for the purpose of unification. In the school systems in China, including Shanghai and Guangzhou, no education about the literacy of the dialects is available. To some extent, the modern written Chinese has become the only resource to transcribe the spoken dialects by the dialect speakers who were educated in modern written Chinese from an early age. However, traditional written Chinese is an important source for the study of written Cantonese in Guangzhou. People in Guangzhou are under the administration by the central government of mainland China and yet strongly influenced by Hong Kong public media. In Hong Kong public media, written Cantonese, a writing system based on the traditional Chinese characters, is widely used. Because of the easy access to the Hong Kong print and electronic media, such as magazines, newspaper and websites, people in Guangzhou are able to learn the Cantonese writing naturally outside
schools. Due to the constraints of time and research resources, the present study did not take into account the traditional written Chinese.

The speakers of Shanghainese did not show big differences from the speakers of Cantonese in terms of the use of common characters. The present study showed that both the dialect writing systems included over 70% common characters from modern written Chinese, and yet Cantonese writings showed slightly higher percentage than Shanghainese ones. This may be because the lexicon Cantonese shares with Mandarin are more than Shanghainese with Mandarin, according to the findings in the present study. The shared lexicon will be discussed later in this section.

Except for the common characters, the phonetically borrowed characters are the ones people most likely to use for online communication. The preference for this type of characters may be due to the function of writing systems. The main purpose of having a writing system is to transcribe the speech sound in a spoken language. Once the association between the sound and the written form is established and widely used by its speakers, this association will be applied by the informants of the writing system to the transcriptions of speech sound of any other languages. Evidence can be found in the transcription of English in modern written Chinese by Chinese learners of English. For the spoken dialects, the phonetically borrowed characters which have similar or the same pronunciations as in Mandarin reflect the representation of the speech sound and thus are more likely to be employed to transcribe their dialects.

The findings in the present study also showed that the use of phonetically borrowed
characters was nonsystematic and therefore variable. Variations in written forms for a single word were fairly common in the dialect written posts. This may be due to the unofficial status of the written dialects. As discussed in Chapter one, the written Shanghainese and written Cantonese have not been standardized among the linguistic scholarship or by the local or central administration. The dialect speakers are free to use whatever characters or written forms they can to represent the dialectal speech sound, and accordingly the use of the phonetically borrowed characters is not completely regular or standardized. However, a phonetically borrowed character in most cases cannot represent the unique phonetic features of the dialects, and dialectologists (Lin, 2003; Pawel, 2002; Snow, 2004; Xu, 2005; Qian, 2003) have for long argued against the overuse of phonetically borrowed characters. They proposed that the dialect speakers should learn to use native characters and previously unattested forms of characters so as to maintain the features of the dialects.

Among the five types of characters, the previously unattested forms of characters in the present study accounted for a certain percentage as well. Through the qualitative analysis, the previously unattested forms of characters are found to be constructed usually by attaching a radical which indicates meaning to a phonetic morpheme, rather than changing an existing character to represent the meaning. This may be due to the habitual ways of Chinese people to construct a character. In written Chinese, the characters which are created by combining semantic and phonetic morphemes are named “phono-semantic compounds”. The phono-semantic compounds accounts for about 90% of Chinese characters (Li, et.al., 1992). As a member of Chinese languages, Cantonese and Shanghainese employ the same way of
creating their own characters, of which phono-semantic compounds account for the most part. In the present study, no statistic computation has been conducted over the percentage of phono-semantic characters, but it is reasonable to infer the predominance of phono-semantic characters in the whole word pool.

Through the statistic analysis on the dialect written posts, 6.44% of the Cantonese writings were found to be previously unattested forms of characters, while in Shanghainese writings, there was only 1%. The imbalanced percentages of previously unattested forms of characters in Cantonese and Shanghainese are probably resulted from the developmental disparity. The development of a writing system for Cantonese, prompted by the Hong Kong administration, flourished since 1950s, while the development of a writing system for Shanghainese writings thrived just around the 1990s (Qian, 2003). Some phono-semantic compounds, such as 喺 (‘is’), 嘅 (‘of’), 喺 (suffix to the plural pronouns), which are categorized to the Cantonese previously unattested forms of characters, were widely used by Cantonese speakers, while Shanghainese speakers and the dialectologists have not achieved consistency over the previously unattested forms of characters yet. The developmental disparity can also explain the difference in the percentages of the semantically borrowed characters and the phonetically borrowed characters between Cantonese and Shanghainese.

No matter in Cantonese or Shanghainese written posts, the native characters accounted for the least proportion, although the dialect speakers reported the use of the native characters to some degree. This is probably because the dialect speakers mistook the previously unattested forms of characters for the native characters since some previously unattested
forms of characters have been used by the dialect speakers over years. According to Dong (2005), the previously unattested forms of characters cannot be categorized into the standard writing and are popular only among the uneducated readers and writers (p. 83); the native characters (characters that can be traced back to ancient sources), on the other hand, are considered orthodox in writing. However, for the computer-mediated communication, the writers are usually not educated for the dialect writings, and are not equipped with the knowledge as the dialectologists do to tell the differences between previously unattested forms of characters and native characters. What’s more, as we have discussed in Chapter One, the distinctions between previously unattested forms of characters and native characters are problematic and still under debates.

The findings of the percentage each type of the dialect writing accounts for in the present study is not in line with the findings from previous studies (Pawel, 2002; Zhao, 2007) in this regard. Phonetically borrowed characters, rather than previously unattested forms of characters or semantically borrowed characters, accounted for the largest part of the dialect writing except for the common characters. As I have pointed out in Chapter One, this is due to the different sources the studies base on. Previous studies limit their research to dictionaries and literary works, while the dialect written posts for online communication are employed in the present study.

In terms of dialect lexicon, the findings from the present study show that common words account for about 60% of the vocabulary pool in dialect written posts. This result is consistent with the previous study on Shanghainese, while higher than the study on
Cantonese (You & Yang, 1998). This inconsistent result may be due to the different methodologies used for computing the ratio. The study by You & Yang used tape recording as the main source for research. Dialect vocabulary with the same pronunciation and the same meaning would be considered common words, no matter what written form the words would take. So there is possibility that the semantically borrowed characters would be excluded from the category of common words because they are pronounced differently in the dialects from Mandarin. However, in the present study, the data coders and the researcher drew on the dialect written posts as sources, which focused on the written form and the meaning. As described in Chapter Three, the semantically borrowed characters were usually categorized into common words because they share the same meaning and the same written form as in written Chinese. Another explanation for the inconsistency between the present study and the previous study is that the computation methods are different. The present study only used the regular computing and types of words were simply added up no matter a word appeared in the written post once or repeatedly, while in the previous study, the authors employed the weighted average computing method which takes the frequency of the repeated words into account.

*The sociolinguistic variables and the use of the dialects online.* The social variables examined in the present study showed that the respondents who received education at college level, aged from 21 to 30, and were females with intellectual occupations were more likely to write the dialects online. The findings are consistent with the previous studies to some extent. In the report of the language use in China issued by the Department of Education (2006), the
speakers of the dialects at the college level speak the dialects more than those who are at lower education level. The results in the present study also showed that people who had four-year university background and aged from 21 to 30 are more likely to use the written dialects. This may lie in two reasons. First, well educated people aging from 21 to 30 are basically matured and their awareness of the regional identity has been awakened. To show their loyalty to the region (Guangzhou or Shanghai) where a large number of migrants moved in each year (Tang, 2006; Zhang, 2006), the use of the written dialects is a good way to distinguish themselves from those new residents. Second, compared with people with lower level of education, people with four-year university education are more inclined to learn the written dialects as a cool net-speak and more interested in perfecting this skill for online communication.

However, there is possibility that the results may be biased by the participants involved. According to the annual survey on internet users (CNNIC, 2008), about 50% of the internet users in China age from 21 to 30, and about 40% of the internet user in cities are at the education level of university. Therefore, the survey conducted in the present study may recruit proportionally larger ratio of people who are between 21 and 30 with university education background. This may have in turn increased the percentage of the particular groups of the dialect writers.

Unlike the findings from previous studies (Wang & Ladegaard, 2008) that females prefer the standard variety (Mandarin) to the vernaculars, the results in the present study indicated that there are more females than males using the written dialects online. This may
be because the written dialects online is not only a regional dialect, but also an internet language. Females are more sensitive and more open to “cool and fashion” languages than men (Li, 2007). However, there are also limitations in the study. The participants involved in the survey are not equal by gender due to the constraints of practical recruitment. Males who participated in the survey are three fifths as many as females. The imbalanced size of participants by gender may bias the results.

The results in the present study also showed that people with intellectual occupation are more likely to write dialects online. It is probably people who engage in intellectual professions may have more time spending on the internet and practicing the written dialects, while work requiring labor may not allow workers to access to the internet. According to the annual survey on internet users (CNNIC, 2008), only about 30% of the internet users are engaged in non-intellectual professions. However, there are limitations with the question design. First, the understanding upon “intellectual” may vary from person to person. One may consider an electrical engineer as an intellectual employee, while the other may consider him as a non-intellectual employee. Second, people may report more intellectual occupations than they actually have, because the society in China values the intellectual professions higher than non-intellectual work.

In the survey of the present study, most of the participants who wrote the dialects online were reported to be the local residents in Guangzhou or Shanghai, while about one third dialect writers were non-local residents but can speak the dialects. The result indicated the scale of the dialect writers was increasing on the internet where no boundary acts upon
communication. However, there is a possibility that the non-local residents in Guangzhou were able to write the dialect online only because Cantonese is a lingua franca in Guangdong and Guangxi province. The relevant findings also showed that the period of time when the participants stay in other regions did not correlate positively with their preference for the written dialects online. On the contrary, the ones who stayed outside the regions longer were more likely to use the dialects online. This may be due to the need of the online dialect users who tried to show their regional identity by using the regional dialects online. First, in the places where the outside dialect speakers live, they may not have opportunities to talk in their regional dialects, but on the internet they were free to use their dialects. Second, by using their dialects, the outside dialect speakers are able to reinforce the connections to the region and the cultures the region features.

The discourse variables examined in the present study indicate that the use of the written dialects online has strong correlation with the type of the communicating participants and the type of the forums where communications take place. In general, people are more likely to write the dialects online when they communicate with the local people or the people who are able to speak their dialects. However, the findings on the type of the forums show that people use the written dialects extensively in the local forums, which implies that no matter the communicating participants are able to speak the local dialects or not, the dialect speakers maintain writing the dialects to initiate or respond to a discussion.

Compared with the region forums in regular websites, the local forums have more local visitors and actually form the speech communities even in the non-physical internet. People
who posted messages or respond to discussions in the local forums assume their communicating participants are from the local areas and able to speak the dialects, even though some participants are not able to “write” the dialects. In contrast, the region forums in regular websites are more open to both local visitors and outsiders. Local people in the region forums may use the written dialects less so as to be able to communicate with outsiders. In conclusion, the local forums are more of a closed community of local residents while the region forums in the regular websites are more of an open community.

Compared with the type of forums and the type of participants, the type of topics did not prove to be a strong factor that contributes to the use of written dialects, even though the dialect speakers reported a tendency to write the dialects on a certain type of topics. However, when cross-examining the type of topics with the type of forums, it is noted that the local forums where people write the dialects more often involve more personal issues, such as complaining about the stress from work, asking for help with a relationship, or complaining about parents. But in the region forum from regular websites where less written dialects are used, people talk about the local issues in a more objective or distant manner, such as the dialects, the local jobs, and the featured restaurant in the neighborhoods. However, in the present study the observations made above are not verified with quantitative analysis. For future studies, the statistical cross-examination of the topics and the type of forums should be included.

The reasons and attitudes of the use of the written dialects online. To address the third research question, the present study explored the reasons for the use of the written dialects
online. The findings show there are three reasons, first, to enter into a speech community, second, to express regional identity, and third, to establish solidarity or create social distance from others. In a local forum or a region forum of regular websites, people feel a necessity to write the dialects because they assume other participants in the discussions on the forums would do so as well. The use of the written dialects can prompt the two-way communication to work out. In order to display the regional identity to other participants in the discussion, people also use the written dialects. Because the topics involved in the discussions cover both personal and impersonal issues, people choose to use the written dialects to establish solidarity with other local participants for the discussions of personal issues while create social distance from outsiders. The results show more females than males go for the latter two reasons. This might be because females are more aware of the social function of a language than males.

As for the attitudes towards having the dialect written news online, a slightly higher percentage of Cantonese respondents voted for the dialectal news report than Shanghainese respondents. This may be due to the relative higher prestige of Cantonese than Shanghainese. Cantonese is used by a larger number of people in more extensive regions than Shanghainese. However, the relatively large proportion of supporting having the dialect written news online may be biased by the type of the respondents recruited in the survey. All the respondents are the speakers of the dialects who have the ability to read and write the written dialects, while other dialect speakers may not totally agree with these respondents.

As regards the evaluation on the relation between the internet and the dialects, the
major respondents gave positive evaluations. As we examined the history of dialect writing in Cantonese and Shanghainese in Chapter One, the input systems, such as Jyutping for Cantonese and Qian’s input scheme for Shanghainese, were widely spread among the dialect speakers because of the internet.

**Conclusion and implications**

The present study examines the use of the written dialects for computer-mediated communication in three issues. First, the type of dialect characters and the type of dialect lexicon were studied. The results showed that as members of Chinese languages and as unofficial written forms in China, people use a large proportion of characters and words in written Chinese to transcribe their dialects. In addition, the dialect writings are quite inconsistent among the dialect speakers, especially when the dialect writers employ various phonetically borrowed characters to represent one meaning. For the convenience, the online dialect writers avoid using native characters although the linguists made great effort to encourage the use of native characters. Feature words that exist only in the dialects are maintained in the written forms through various ways, such as in native characters, with phonetically borrowed characters or previously unattested forms of characters. Second, the social and discourses variables correlating with the use of the written dialects online were examined. Consistent with the general description of internet users and previous studies on dialect speakers, gender, age, education level and the type of occupation proved to correlate with the use of the written dialects. In terms of the discourse variables, the type of websites and the type of communicating participants are strong factors that contribute to the use of
written dialects online. Thirdly, reasons for the use of written dialects online were also explored in the present study.

On the whole, the present study provides an overview of the current situation of the use of Cantonese and Shanghainese online. It provides a new avenue to look at regional dialects from the perspective of computer-mediated communication. No matter to control over the dialect use for certain purposes by the central government in China, or to promote the dialects by the dialectologists, the research can be used as a reference. However, the study only touches upon the linguistic and sociolinguistic issues in a broad way. In future studies, focus on either linguistic or sociolinguistic variables should be made. For example, the previously unattested forms of characters in Cantonese can be contrasted with those in Shanghainese, given the written Cantonese is much more developed than the written Shanghainese. The sentence structure of the written dialects in the discussion posts can be analyzed to see if the written dialects are strongly interrupted by written Chinese. In addition, more participants should be recruited in surveys to maintain statistical significance of future studies according to Cohen’s effect size table (1992), and to contrast with the dialect writer groups, control groups who are not able to write the dialects should be included in future studies of the sociolinguistic issues.
Notes

1 In a narrow sense, Cantonese refers only to Guangzhou-Cantonese because Cantonese is originated from Guangzhou. In a broad sense, Cantonese refers to all the sub-dialects developed on the basis of Guangzhou-Cantonese, such as Hong Kong Cantonese. In the thesis, Cantonese is used in a narrow way, only referring to the Guangzhou-Cantonese because the thesis focuses on the dialects in mainland China, where the developments of dialects are in the similar political context.

2 The transcription systems of pronunciations of the two dialects will be discussed in detail in the third section of this chapter.

3 [a] is a function word in Shanghainese, Cantonese and Mandarin, indicating the questions, or sometimes used as interjections.

4 [ŋə] is a function word in Shanghainese, indicating the present perfect tense.

5 “Sofa” means the first one to reply the beginning post. It is internet slang in Chinese.

6 LSD is a post hoc test which computes all pairwise t-tests and evaluates results using standard p values. It somewhat leads to inflated alpha levels in a study.
References


http://www.shtong.gov.cn/node2/node2245/node74288/node74301/node74521/node74531/userobject1ai90701.html


(CNKI:CDMD:2.2007.081776).


vernacular. Hong Kong: Hong Kong University Press.


Appendix I Questionnaire

1. How old are you?
________________

2. What’s your gender? (simple selection)
   A. male
   B. female

3. What’s your highest education level? (simple selection)
   A. 4-year undergraduate study or above
   B. 3-year postsecondary college
   C. senior high or postsecondary technical school
   D. junior high or below

4. What’s the type of your career? (simple selection)
   A. intellectual
   B. non-intellectual

5. Is your native dialect Cantonese? (simple selection)
   A. Yes
   B. No

6. (1) Are you local residents in Guangzhou? (simple selection)
   A. Yes (if you choose this one, please go straight to No.7)
   B. No (if you choose this one, please answer the second question below)

   (2) How long have you lived outside Guangzhou? (simple selection)
A. never
B. less than one year
C. one to five years
D. over five years

7. (1) Do you use Cantonese online? (simple selection)
A. Yes. (if you choose this one, please answer the second question below)
B. No. (if you choose this one, please go straight to No.22)

(2) Does the use of Cantonese online depend on different situations? (simple selection)
A. Yes. (if you choose this one, please go to No.8)
B. No. (if you choose this one, please go straight to No.12)

8. (1) Do you use Cantonese in some websites but not in the others? (simple selection)
A. Yes. (if you choose this one, please answer the second question below)
B. No. (if you choose this one, please go straight to No.9)

(2) In what type of websites do you usually use Cantonese? (multi selection if applicable)
A. most members of the website are from Guangzhou
B. in the section which is especially for local people in Guangzhou in an ordinary website
C. other types, for example, ______________________________

9. (1) Do you chat online in Cantonese with some people, but not with others? (simple selection)
A. Yes. (if you choose this one, please answer the second question below)
B. No. (if you choose this one, please go straight to No.10)
(2) With whom do you usually use Cantonese for online chatting or posting? (multi selection if applicable)

A. people from Guangzhou

B. people from other places but living in Guangzhou for a long while

C. It has nothing to do with where he/she is from.

D. Other types of people, for example, ____________________-

10. (1) Do you discuss about some topics online in Cantonese, but not on the others? (simple selection)

A. Yes. (if you choose this one, please answer the second question below)

B. No. (if you choose this one, please go straight to No.11)

(2) What topic would encourage you to use Cantonese for online discussion? (multi selection if applicable)

A. It depends on the region the topic is about. For example, I’ll use Cantonese when the topic is about Guangzhou.

B. It depends on the theme of the topic, such as politic issues, economic issues, life, or entertainment.

C. It has nothing to do with the region or the theme.

D. other topics, for example, _______________________________________________

11. Do you use Cantonese selectively under other circumstances? (simple selection)

A. Yes, for example, __________________________

B. No.
12. Why do you use Cantonese for online chatting and posting? (multi selection if applicable)

A. everybody use Cantonese; so do I

B. I’m from Guangzhou, so I use Cantonese

C. to establish solidarity with the person I’m talking to, or create social distance.

D. for other reasons, for example,__________________________________________________

13. If a stranger online initiates a conversation in Cantonese, you will (simple selection)

A. use Cantonese as he/she does

B. not always respond in Cantonese. It depends.

C. usually respond in Mandarin.

14. If you come across a stranger online, you will (simple selection)

A. initiate the conversation in Mandarin

B. initiate the conversation in Cantonese

C. initiate the conversation sometimes in Mandarin, sometimes in Cantonese

D. use other varieties other than Cantonese or Mandarin

15. Do you know how to write Cantonese? (simple selection)

A. proficiently

B. just so so.

C. I can only write a few frequently used words in Cantonese

D. I have no idea at all (if you choose this one, please go straight to No.19)

16. How do you “speak” Cantonese online? (multi selection if applicable)
A. use the Cantonese characters

B. use the homophone in Mandarin

C. use the synonym in Mandarin

D. create a character as long as others understand

E. use Pinyin

F. others, for example, __________________________

17. How do you learn Cantonese characters? (multi selection if applicable)

A. learn from peers online

B. see what comes from the Pinyin input system and pick one that I favor

C. learn it from the standard Cantonese online dictionary

D. use the Cantonese input system to write

E. learn from newspapers, magazines or Cantonese soap operas with subtitles

F. it is taught at school

F. others, for example, __________________________

18. If you happen to write a word that you don’t know how to do it with the input system, you will (multi selection if applicable)

A. replace it with the word in Mandarin which shares similar pronunciation.

B. check online for the standard character

C. use Pinyin instead

D. replace it with the word in Mandarin which shares similar meaning

E. create a character by myself
F. others, for example, ___________________________

19. How do you tell whether people who are involved in a discussion are using Cantonese?
(multi selection if applicable)
A. see if the vocabulary features Cantonese
B. see if the characters are Cantonese characters
C. see if the sentence fits the bill of Cantonese order
D. see if she is from Guangzhou and if she is familiar with topics related to Guangzhou
E. in other ways, for example, ____________________________

20. Do you use Mandarin for online chatting and posting in addition to Cantonese? (simple selection)
A. Yes
B. No

21. (1) when you are chatting or posting online in Cantonese, do you switch to Mandarin sometimes? (simple selection)
A. Yes. (if you choose this one, please answer the second question below)
B. Never. (if you choose this one, please go straight to No.23)
(2)Under what circumstance do you switch to Mandarin for online chatting and posting?
(multi selection if applicable)
A. when I find all of others use Mandarin
B. when I get to know the interlocutor is not from Guangzhou
C. when the topic involved has nothing to do with Guangzhou
D. when I am talking something serious or related to work

E. under other circumstances, for example, ___________________________________________

(If you finish Question 21, please go straight to Question 23.)

22. When you don’t use Cantonese online, it is because (multi selection if applicable)

A. everybody online uses Mandarin

B. I don’t know how to use the input of Cantonese characters

C. I get used to use Mandarin for work and life

D. for other reasons, like ______________________________

23. Do you support the idea of having Cantonese news report online? (simple selection)

A. yes, the reason is ______________________________

B. no, the reason is ______________________________

C. it depends on ______________________________

24. What do you think are the impacts of internet on the spread of Cantonese? (simple selection)

A. positive

B. no effects

C. negative

D. hard to tell. Maybe positive, maybe negative