8-2014

Linguistic Variation in Contact: the Use of erhua and rusheng in the Xianggang Community in China

Lei He
Syracuse University

Abstract

This study considers dialect variation by young and middle-aged speakers in the Xianggang community in China.[1] The Xianggang community is located at the southern part of China, but it used to be composed of northern people. The contact between northern and southern dialects has given birth to a dialect mixed situation. With the attempt to know the driving force of language change within this speech community, both social factors and linguistic factors have been taken into consideration. In this study, *Erhua* is chosen as the representation of northern dialects and *rusheng* is chosen on behalf of the southern dialects. Linguistically, *erhua* has two environments: post-vocalic and post-nasal. Against the background of Chinese education, social factors such as age, gender, and period of residence as well as the division of northern and southern characteristics have been studied.

This study suggests that there is dialect variation within the speech community, while more evidence is needed to support the assumption of dialect acquisition. Age and the division between northern and southern characteristics play significant roles on the use of *rusheng* and *erhua* within the speech community. Period of residence has effect on the use of *erhua* but not on *rusheng*. Influenced by Putonghua education, young speakers have less variation in their use of these two linguistic variables than middle-aged speakers do. Within the social context of a steel company, the use of *rusheng* is lexically conditioned because both northern and southern-characterized speakers choose to apply this tone to similar characters. Post-vocalic *erhua* is a significant linguistic environment for the use of *erhua* of the northerners. This is changing tendency towards the use of Putonghua within the speech community.

[1] This study focuses on varieties of Chinese. Although they are not 'mutually intelligible', they are treated as dialects in this study. According to linguistic classifications, they should be considered languages. However, in Chinese linguistics, they are often treated as dialects, because of their shared grammar and writing systems. In this study, I follow the classification of mainstream Chinese linguists.

LINGUISTIC VARIATION IN CONTACT: THE USE OF *RUSHENG* AND *ERHUA* IN THE XIANGGANG COMMUNITY IN CHINA

By

Lei He

B.A., Ludong University, 2012

THESIS
Submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Linguistics
In the Graduate School of Syracuse University

Syracuse University
August 2014

Acknowledgements

My deep appreciation goes to my thesis advisor Dr. Rania Habib. I feel very thankful to the inspiration from your course *language change and variation*. It is your insights and discussions that have encouraged me to make up my mind to conduct this work. I really enjoyed this experience, and with your guidance I finally understood how to organize my thought and make arguments. The process was hard, but luckily I have you as my guide who has spent countless hours to help me improve the work and has always been responsive to my ideas. Thanks for your critical thinking, and frank words, I couldn't do this without your devotion.

Many thanks are given to my academic advisor, Dr. Mary Louise Edwards, for your kindness and support throughout my graduate study. Your calmness left me the first expression on academics when I first walked into your office. Thank you for spending a whole summer with me to do the independent study. Besides the academic knowledge, I gained pleasure from reading and the ability to learn and solve problems. Thanks for all the conversation that happened between us, your words gave me power and supported me to finish this work.

Thanks to my committee member, Dr. Tej. Bhatia, and my mentor Dr. Amanda Brown for your encouragement and believing in me.

Special thanks go to all the participants of my study, and my friends in Syracuse. Without your support, this work couldn't be done.

Finally, I would like to say thank you to my parents, for your unconditional support and endless love. Thank you for giving me a happy life.

TABLE OF CONTENT

INDEX OF TABLES

INDEX OF FIGURES

CHAPTER ONE: INTRODUCTION

Much empirical linguistic work has been conducted in the context of speech community. As a concept that has been defined and implemented by many researchers (Gumperz, 1968; Labov, 1972; Kerswill, 1993; Milroy & Milroy, 1998), the study of speech community is at the intersection of many principal problems in sociolinguistic theory and methods. In the interest of knowing the driving force of linguistic change in a small speech community in which dialect contact (between the northern dialect and the southern dialect of China) takes place, this study explores dialect variation in the Xianggang community in the city of Xiangtan, Hunan province in China. Following the procedure of sociolinguistics surveys (Patrick, 2002), this study investigates the history and makeup of the community and examines if social variables such as age, gender, and period of residence as well as the division between northern and southern characteristics can be used as explanatory factors for the variation observed in the dialect and to predict the tendency of language change. The Xianggang community is located at the southern part of China but was initially composed of people who migrated from northern China to the South; linguistically, northern and southern dialects have become mixed within this community. To study the linguistic behavior of the members of the community, *Erhua* and *rusheng* have been chosen as representatives of northern and southern dialects, respectively. *Erhua* is an r-colored phonetic feature that widely exists in the speech of northern people (Zhang, 2005; Wang, 2007; Pankhurst, 2012) and has two phonological constraints: post-vocalic and post-nasal. *Rusheng*, however, as a tonal feature in Middle Chinese and does not exist in Modern Mandarin and other northern dialects but is preserved in many southern dialects in China (Chen, 1983; Yang, 2004).

Xiang dialect, a dialect spoken in the Xianggang community, is one of these southern dialects, which still preserves *rusheng*.

The Xianggang community is a small speech community inside Xiangtan city (Figure 1.2), Hunan province. Geographically, Hunan province is located in central south of China. As described in Figure 1.1, the Huai River–Qin Mountain line (red line in the map) divides China into the north and the south.

Figure 1.1: The division of North and South China and the location of Hunan province

http://upload.wikimedia.org/wikipedia/commons/d/db/%E7%A7%A6%E5%B2%AD-%E6%B7%AE%E6%B2%B3%E7%BA%BF.PNG

http://www.kingwu.com/china_map_provinces.gif

Figure 1.2: The location of Xiangtan city in Hunan province

Adopted from http://english.xiangtan.cn/

As recorded, people living to the north of the red line have a significant difference in their living style compared to people living to the south of the line. Moreover, there are some stereotypes about residents in the North and the South, which have been identical since 2000 years ago when Middle Chinese was used. Eberhard's (1965) work has confirmed these stereotypes by conducting a survey of speakers in China, including the northern and southern regions. He found that the northern people were considered "strong, frank and honest," while the southern people were "small, cunning and smart." Since communication is the main criterion that has an effect on the formation of stereotype according to Eberhard (1965), regional-related stereotypes have probably affected the selection of language used by northern and southern people.

Using quantitative methods, this study has focused on social factors such as age, gender, and period of residence as well as the division of northern/southern characteristics to see if any

of the factors has an influence on the use of *erhua* and *rusheng* within the speech community. Also, at the background of China's social context, language planning is an inevitable social factor, which is worth considering in this study. It is well known that in 1955, the Chinese government implemented a policy that set Putonghua (Mandarin) as the standard Chinese language. This policy has given Putonghua a prestigious status in the society, especially through education and social media. At the same time, however, many dialects are still used by local people in different regions of China, which makes dialect study valuable and necessary. As a native speaker of Mandarin who has grown up in the area of the Xiang dialect, the researcher feels there are some social functions that can be fulfilled only by using this dialect rather than Mandarin. In this study, Xianggang community members are the subjects of dialect variation and change. As a group that initially migrated from the northern part of China, the social and linguistic process underlying speech modification in different generations who have been living in the south for a long time would be at their most visible.

This chapter provides some background to the study. First, the research topic is introduced by clearly demonstrating the research questions. Second, based on the accomplishment of historical and contemporary studies on the Chinese language, a unified structure in Chinese phonology as well as the diverse representations of *erhua* and *rusheng* by people in different dialect areas is presented. Third, to examine the role of social factors on the possible language change, a historical overview of the Xianggang community is provided.

1.1 RESEARCH QUESTIONS

To raise awareness of dialect diversity in Mainland China and to explore the nature of dialect variation or acquisition in a dialect contact situation, this study makes its predictions by focusing on the subject of Xianggang community members and their linguistic behaviors.

Through the study of their use of *erhua* and *rusheng* within their social contexts, dialect variation within the community is investigated by answering the following questions:

(i) Is there variation in the use of *erhua* and *rusheng* in the speech of young and middle-aged speakers in the dialect contact situation in the Xianggang community?

(ii) Do linguistic factors and social factors, such as age, gender, period of residence and the division between northern and southern characteristics, play a role in this variation?

(iii) What is the pattern of linguistic variation in the community; is there a change toward Mandarin or the Xiang dialect?

1.2 BACKGROUND OF CHINESE LANGUAGE

Modern Chinese is to Archaic Chinese as Modern English is to Old English. Like many other languages in the world, Chinese language has experienced dramatic changes over 3000 years. These changes are embedded in the background of societal changes. Chen (1999) summarized the periodization of Chinese clearly in his book. This study is based on the Modern Chinese Language,[2] particularly the language spoken in Mainland China. Many related works have been done on Chinese spoken by people in Taiwan (Liao, 2010) and Singapore (Chen, 1983). These works inspired the focus on this topic, thereby filling a major gap in Chinese sociolinguistics with this current study.

In historical Chinese phonology, many studies have been done to reconstruct the sounds of the past, and their achievements are reflected in rhyme dictionaries such as *Qie Yun* (601 CE), *Guang Yun* (1008 CE), and *Zhong yuan yin yun*（1324 CE). These dictionaries provide a system

[2] In this paper, the Chinese language is a broad category of languages spoken in China, not confined to Mandarin and includes many other dialects mentioned in Figure 1.3.

of categorized rhymes, but the arrangement is not alphabetical as in Modern Chinese; rather, they use Chinese characters (see 1.2.2 "fanqie" method). Although many works in Chinese linguistics still use rhyme dictionary/tables as reference when trying to trace the history of dialect development, such as Yang's (2004) work on *rusheng*, the history of sounds is not included in this research. Rather, this study largely focuses on the northern and southern features *erhua* and *rusheng* and analyzes the occurrences of *erhua* and *rusheng* to explain the observed variation and change in the speech community under investigation.

1.2.1 Relationship between Putonghua and Chinese dialects

When talking about Chinese spoken in Mainland China, one cannot avoid Putonghua. Putonghua, sometimes translated as Mandarin, is the standard Chinese in Mainland China. It was formally defined in 1956, and translated as follows:

> "Putonghua is the standard form of Modern Chinese with the Beijing phonological system as its norm of pronunciation, and Northern dialects as its base dialect, and looking to exemplary modern work in Bai Hua 'vernacular literacy language' for its grammatical norms" (Chen, 1999).

Based on this definition, there is a division in dialects in Chinese. Putonghua is based on the northern dialects. Figure 1.3 shows that the Chinese language includes Mandarin and Southern Chinese. Within the category of Mandarin, there are three dialects: Northern, Eastern and Southwestern. Meanwhile, there are six dialects in Southern Chinese, including: Wu, Gan, Hakka, Xiang, Min, and Yue dialects. This research focuses on the variation in the contact between the northern dialect and the Xiang dialect, at the background of Putonghua promotion.

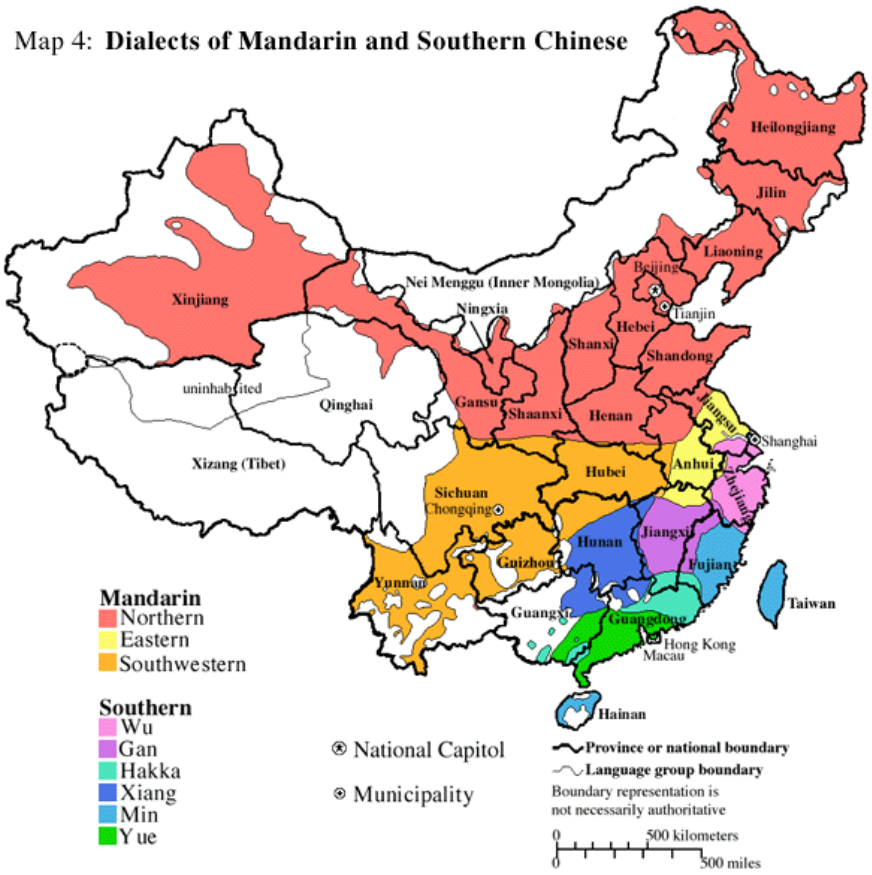

Figure 1.3: Dialects of Mandarin and Southern Chinese

Adopted from http://www.dartmouth.edu/~chinese/maps/maps.html

1.2.2 Phonological rules in Putonghua

Putonghua phonological system is not divided into three parts (onset, nuclei and coda). However, influenced by the ancient phonological study, Chinese linguists customarily separate the segment into two parts: initial syllable (onset) and final syllable (nuclei and coda). There are 21 initial syllables (see Table 1.2) in Mandarin Chinese, and all of them are consonants. Though important, the initial syllable is not mandatory in the Chinese phonological structure, but the final syllable is. There are 37 final syllables in Chinese and they include everything except for the initial syllable (see Table 1.3.) Chinese is quite different from other alphabetic languages because it has a logographic writing system whose symbols are meaning based rather than

pronunciation based. Chinese took on the alphabet form in the 1950s as one implementation of the policy on Chinese Romanization; previously, only characters were used for all Chinese linguistics studies. For example, as one part of phonological studies in Chinese, to identify the sound for a third character, traditional Chinese dictionaries set the model to use two characters, by assembling the initial syllable of A with the final syllable and tone of B. This method is called "Fanqie." For example, when a character is pronounced as "[kuŋ]," they would use two characters, pronounced as [ku] and [huŋ] separately. In this case, the third character received the [k] sound from the first character and the [uŋ] sound from the second character and was then assembled together to generate the sound of [kuŋ]. This background knowledge provides theoretical guidance to this study.

Table 1.1: Initials of Modern Standard Chinese (Chen, 1999)

Bilabial	[p]	[p']	[m]	[f]
Alveolar	[t]	[t']	[n]	[l]
Dental Sibilant	[ts]	[ts']	[s]	
Retroflex	[tʂ]	[tʂ']	[ʂ]	[ɹ]
Palatal	[tç]	[tç']	[ç]	
Velar	[k]	[k']	[x]	

In regards to the two linguistic variables, *Erhua* is a phonological process that is related to the components of the final syllable. According to Chao's study (1968), *erhua* is a process that involves a retroflex sound [ɻ] that is added to the syllable final and causes the final to become rhotacized. *Rusheng* is a suprasegmental feature of Chinese. As a tone, it existed in Middle Chinese and has been studied historically by many linguists in and out of China (e.g. Chen, 1983; Yang, 2004). *Rusheng* has different features from the other four tones in Mandarin Chinese. The pattern of the 'fifth tone' *rusheng* (tone 2-4) is similar to "Tone 6" in Stanford and Evans (2012)

and has a short duration and narrow range (Figure 1.4).[3]

Table 1.2: Finals of Modern Standard Chinese (Chen, 1999)

	[i]	[u]	[y]
[a]	[ia]	[ua]	
[o]	[uo]		
[ɤ]			
[ɛ]	[iɛ]	[yɛ]	
[ai]	[uai]		
[ei]	[uei]		
[au]	[iau]		
[ou]	[iou]		
[an]	[iɛn]	[uan]	[yɛn]
[ən]	[in]	[un]	[yn]
[aŋ]	[iaŋ]	[uaŋ]	
[əŋ]	[iŋ]	[uəŋ]	
[uŋ]	[yŋ]		
[ɚ]			

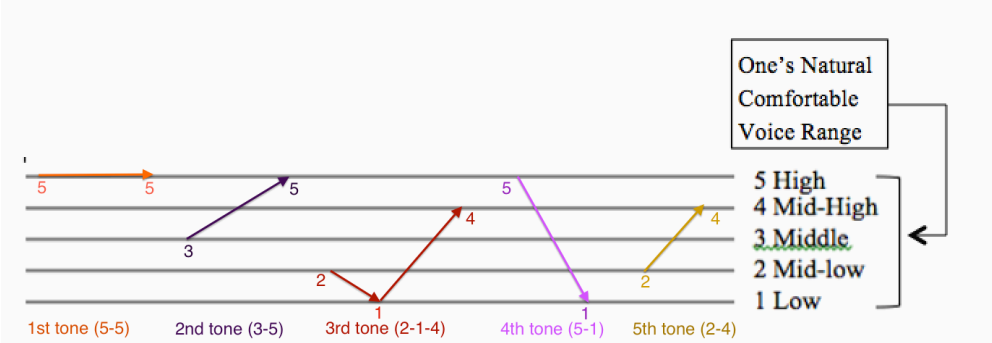

Figure 1.4: Pitch Contour of Chinese Tones, including the 5th tone (in Xiang Dialect)

Modified based upon http://web.mit.edu/~jinzhang/www/pinyin/tones/

[3] Chen (1983) named it the fifth tone and Stanford & Evans (2012) named it tone 6. Although they are using different names, they are referring to the same tone with the contour 2-4, which is the *rusheng* tone in this study.

1.3 SOCIAL AND HISTORICAL BACKGROUND OF THE XIANGGANG COMMUNITY

The Xianggang community is located in Xiangtan city, Hunan province. In this southern area, people speak the Xiang dialect. Xiangtan City, a medium-sized city in Hunan province, is situated 40 miles south of Changsha, the capital city of Hunan province. There are two districts within Xiangtan City: Yuhu and Yuetang. The Xianggang community is located in the Yuetang district, at the southwest part of Xiangtan City. It was formed along with the foundation of the country-owned steel company in Xiangtan City. The company was founded in 1958, accompanied with a flood of worker migrants from the northern part of China to the south (Xiangtan City). As an inner-city community, the Xianggang community has approximately 20 residential sub-districts. As is shown in the labor records statistics, the community has a population of 67,000, which includes 18,000 current workers, 18,000 retired workers, and around 30,000 workers' relatives. Because more than 50 years have passed since they initially immigrated to Xiangtan City, most of the community members have experienced changes over generations. As an inner-city community, the community members' linguistic features are not only different from other city members, but also from ordinary Mandarin, which has been promoted by the Chinese government since 1956.

> "We are considered Northern people when living in the Xiangtan city; however, when we travel to the North China, people in that area say: 'you are Southerners.' Our speech brings us to an awkward situation" (From two informants who originated from the North).

As one of the most typical organizational social structures in China during the 20th century, the Xianggang community represents the phenomena of "unit cultural circle" (Liu, 2009). Unit cultural circle is a specialized type of culture in China, which is a culture of state-owned companies. They formed a 'small society' within the city. Community members have their own

welfare system that includes school, hospital and restaurants, even grocery stores. Because of the relatively stabilized feature of their jobs, community members are proud of their company and also their community culture. Research on the "unit culture" becomes valuable as many country-owned companies in China are closed down because of their failure to compete in the market with other companies with the introduction of the market economy in 1997 (Liu, 2009). As described in Liu (2009), the author himself is a Xianggang community member. The reason he chose to study the "unit culture," by focusing on the Xianggang community, is because his accent was noticeable; therefore, he was labeled as a Xianggang community member when he spoke with members of Xiangtan city. Through the study of their language variation, this study contributed to the study of "unit culture" and the relationship between culture and language.

1.4 SUMMARY

Briefly speaking, in the midst of a dialect contact situation created by labor immigration, this study investigates sociolinguistic variation in Xianggang community in southern China. In addition to the social factor of Putonghua promotion in Mainland China, the influence of four social variables (age, gender, period of residence and the division of northern/southern characteristics) on Xianggang community members' use of two linguistics variables (*erhua* and *rusheng*) is studied. As a consequence of the effects from these social variables, the tendency of language change has been studied.

This paper is organized as follows: in Chapter 2, literature on the topic of dialect acquisition and that related to the observed linguistic variables are reviewed, situating this study on a theoretical basis. Chapter 3 presents the methodology of the research in detail, together with the rationales. Chapter 4 provides the organization and presentation of data collected for this study. Features of the northern and southern colored speech are analyzed separately within this

chapter. A discussion section is presented in Chapter 5, providing a deeper interpretation of the data following the lead of former literatures. Chapter 6 briefly summarizes the findings of this study as a conclusion. The last section outlines limitations of this study and indications for future researches.

CHAPTER TWO: LITERATURE REVIEW

In Siegel (2010), studies on second dialect acquisition have been divided into two broad contexts: naturalistic and educational. Studies on naturalistic context draw attention to the adaptation of new dialects without formal teaching, and studies on educational context often connect dialects and language with classroom education. Because both of these two contexts are related to this research, this study adopts his clarification. Accordingly, this chapter consists of two joint parts. First, literature on "dialect in contact" in the naturalistic context is reviewed. Based on the definition of this term given by Kerswill (1996:179), both linguistic variation and change within a language community is a matter of dialect contact that is not restricted to those involves mobility and migration. It can be continuous contact among speakers who has different backgrounds or use different varieties of language. Therefore, studies that have described the linguistic variation or that have given explanations to the dialect acquisition by speech community members are presented. Second, in the context of education, the study focuses on the studies regarding the language policy of promoting Putonghua in the P. R. China, which has great influence on the educational system and, in many cases, has affected people's speech patterns. The theory and linguistic models mentioned in this chapter are the foundation drawn upon for the whole study and are repeated in the following chapters.

2.1 DIALECT IN CONTACT IN THE NATURALISTIC CONTEXT

Trudgill's (1986) definition of "dialect in contact" gives a clear line between language and dialect, referring to "contact between varieties of language that are mutually intelligible at least to some degree." His definition fits in with my study because both the Xiang dialect and the

northern dialect are varieties of Chinese. Even though dialects in China are not always mutually intelligible, Chinese people and linguists perceive them as varieties of one language because of the shared writing system. The term "dialect in contact" is developed from "language in contact," a term that Weinreich proposed in 1953. To differentiate the study of these two terms, Trudgill (1986) stated that instead of studying the behavior of individual bilingualism that takes place when people who are fluent in more than one language converse with speakers and switch their language patterns very often, studies on dialect contact focus on linguistic changes within one language. These changes take place particularly during or at the sequence of contacts between closely related varieties within a language. For example, Chamber (2002) analyzed the use of language varieties and their relationship with social variables such as social class, age, and gender.

To better situate this study and make a connection with previous studies, studies on the relationship among dialect contact, dialect variation, and dialect acquisition in the naturalistic context are organized by giving answers to the following questions:

1) How can one language variety have influence on another?

2) What are the patterns of linguistic changes?

3) What are the social meanings associated with one language variety that may influence another?

2.1.1 How can one variety of language have influence on another?

The primary theory that must be dealt with when talking about the influence from one language variety to another is accommodation theory, proposed by Giles in 1973. In his paper, Giles (1973) divided accent change into two directions depending on the motivational tendencies of the sender: accent convergence and accent divergence. To make it clear, he specified "accent

convergence" as the situation where the sender adapts his accent patterns toward that of the receiver to gain the receiver's social approval. On the contrary, "accent divergence" takes place when the sender wants to dissociate himself from the receiver. Connected with the social meaning embedded within accents, he states that speakers adjust their speech up and down from high-prestige to low-prestige accents. Moreover, through a matched-guised test, Giles, Taylor, and Bourhis (1973) found that bilingual speakers gain more approval from another ethnic group when they accommodate toward the group. Switching from the approach of socio-psychology to that of sociolinguistics, Trudgill (1986) paid attention to dialects that are different regionally and explored the convergence and divergence patterns through linguistic analyses, which suggests some benefits for both psychological and linguistic studies. Evidence of sociolinguistics is illustrated in his long- and short-term accommodation theory. Short-term accommodation is often preferred by social psychologists. By using linguistic analysis, Trudgill found that Norwich people accommodate with each other in their use of the linguistic variable [t] but not in another linguistic variable [a:]. Then, adopting the two terms "marker" and "indicator" that were proposed by Labov (1972), he demonstrated the reason: because [t] is a linguistic marker, it is subject to both social stratification and stylistic differences, while indicators like [a:] only experience social class variation. Thus, Trudgill has proved that linguistic studies on short-term accommodation can bring some explanation to the adaptation of one variable over another, which cannot be reached by psychologists. As to the long-term accommodation, Trudgill (1986) used the example of American English and British English and pointed out that the accommodation by British English speakers to American English is mainly because of their residence in the United States. In some occasions such as telling jokes involving Americans and playing American roles in British drama or, most obviously, in British pop singers, British

English speakers choose the most salient feature in American English to imitate in their speech. He also noted that the social and geographic factors embedded within one particular linguistic variable project it as a marker of the community. However, there are also some limitations of solely using accommodation theory to explain linguistic variation and language changes: one of which is children need less time to fully accommodate to another dialect than adults do (Trudgill, 1986). Therefore, it is difficult to study dialect acquisition by looking into the frequency of usage of a particular language variety, especially for children within the critical period (younger than 11 years old).

To sum up, former studies (Trudgill, 1986; Labov, 1972) depicted acquisition processes in the facet of both long- and short-term accommodations. In addition to demonstrating the different roles linguistic convergence and divergence play in communication, their research illustrated how linguistic analyses play a role in determining influential factors. These works provide theoretical guidance to this study. First, linguistic analyses on *erhua* and *rusheng* have been conducted to test the frequency of the use of these two linguistic variables. Second, in the context of long-term accommodation, this study explores how the factors (age, gender, period of residence and the division between northern and southern characteristics) affect the choice of one language variety over another. Third, the research questions of this study are tailored to both workers and students.

2.1.2 What are the patterns of linguistic changes?

To explain the social and geographical spread of linguistic forms from one dialect to another, much research has focused on the diffusion of linguistic patterns. Diffusion is the trajectory of language changes used by dialectologists and sociolinguists to map out the geographically formed language variation. According to Britain (2002), by applying the

geographical notion of space, data could be presented on a map, and variation studies could be analyzed within the apparent-time model. Furthermore, Labov (2007) studied the spread of linguistic forms within the speech community. On one hand, he defined speech community as a concept with "well-defined limits, a common structural base and a unified set of sociolinguistics norms" (Labov, 2007, pg. 347). On the other hand, to explain the linguistic changes in and out of the speech community, he applied "transmission" as the primary source of diversity and "diffusion" as the secondary process. The transmission proceeds in the family tree model and it is often applied to children who preserve language as a faithful manner or dialect acquisition from their parents within the speech community. However, diffusion occurs across speech communities through language/dialect contact or what is referred to as adult language/dialect acquisition. Obviously, when considering the dialect/language acquisition within a speech community, there is a clear line drawn by Labov (2007) in his definition, which is the difference between children dialect acquisition and adult dialect acquisition.

To consider the relationship between the spread of linguistic forms and children dialect acquisition, much attention has been paid to the acquisition order and age. Chamber (1992) proposed eight principles based on evidence from his study on six Canadian children who migrated to the UK and some other related studies. In his research, children's age determines which rules are acquired earlier than others. For example, he proposed that lexical replacements are acquired faster than pronunciation and phonological variation (Principle 1). He compared children at the same age from Canada to the control group, who were native to southern England. The age-graded pattern of children's linguistic performance suggested the acquisition of phonological variants was lagging behind the acquisition of lexical ones. More directly, Kerswill (1996) took a precise account of the dialect acquisition, considering the factor of age and age-

related social contexts. On one hand, his study illustrates the acquisition order of different linguistic aspects of dialects in contact, that is, borrowings are the easiest to acquire and the lexical, unpredictable, phonological rules are the hardest. On the other hand, he extended the study of age to its related social context (interlocutor types) that individuals within a society would experience. With the consideration of both social and linguistic factors, he mentioned that adults have limited access to the adoption of language changes and some changes are only restricted to children. To find the origin of language change, he proposed a hierarchical structure of second dialect acquisition, which is related to different sociolinguistic conditions. The youngest children have the greatest capacity to acquire complex linguistic features. In addition to discussing the acquisition of linguistic rules, he applied the notion of sociolinguistic competence to explain the reason why adults have more homogeneous language use than children. It is a gradually developed ability for community members to recognize linguistic varieties and its evaluation within the community. In this sense, he compared the speech of children and their caregivers in different types of interlocutor combinations that are considered essential in different life stages. For example, for young children ages 0 to 6, caregivers' input and conversational style were regarded as the primary source of children's acquisition. However, other research suggested the linguistic development in children is not necessarily influenced by the adults' input. He explored this point by examining phonological rules, taking into consideration syntactic, morphological, and lexical conditions and found that older children in some cases are models for younger children. For pre-adolescents, ages 6 to 12, contact with peer groups seemed very important and called for some degree of conformity within the group. They moved away from the caregivers' norm to a new one that is related to their social network. Adolescents aged 12 to 17 were aware of adults' stylistic changes in different contexts and more

susceptible to peer-based norms. He found that adolescents are the most influential transmitters of linguistic change because of their wider contacts and willingness to modify their speech to project certain social identities. This is related to this study, as the informants are middle age speakers and younger age speakers above 13. Age is an essential social variable in this study.

To shift the topic from children subjects to adult ones, Stanford's (2008) study highlighted the role of language attitude. At the background of a diverse culture with exogamous customs, Sui people (children) switched from their mother's speech to their peers' speech. This is a transition from caregiver's influence to social influence. This point is quite important for adult dialect acquisition, because adults are more aware of the social norms of the language. More related to the context of migrated groups in this study, Hiramoto's (2010) research has provided an attitude-related example. Japanese immigrants from Tohoku to Hawaii experienced "dialect discrimination" when they first moved to the island. This is because these newcomers hold different accents from old migrants who have been living in Hawaii for a long time. Similar to the background of the current test base, these migrants first moved to Hawaii along with the labor output and then continued to live on the island. The data was collected from 1972 to 1975, over 50 years after their migrations. As stated in the study, it is the obvious dialect stigmatization that pushes new migrants to adopt the local dialect spoken by Japanese people. His finding is also in line with a previous study by Chamber (1992): lexically bounded features are acquired earlier than phonologically bounded ones. Following Kerswill's (1996) statement that sociolinguistic competence is learned over the life span, it gives inspiration to this work that speakers tend to switch their language use toward the community norm.

2.1.3 What are the social meanings associated with one language variety that may influence another?

In addition to considering the social norm that lead to the adoption of a particular linguistic structure, there is some research focusing on individual and community identity. With the consideration of language identity, dialect acquisition becomes a conscious procedure in language change and variation. Identity is not new in the study of language variation and, quite often, it becomes the reason researchers adopt to explain the uneven use of one language variety over another. Early in Labov's (1972) study of Martha's Vineyard, phonetic differences became stronger when the group tried to maintain their identity. It is the centralization of [ay] and [aw] that marks informants' attitudes toward Martha's Vineyard. He found that the centralized form is representative of the Martha's Vineyard identity. Those who had strong attachment to the island used the centralized form more to assert their local identity. Moreover, Hazen (2002) proposed another type of identity, cultural identity, and considered it a sociolinguistic factor concerned with how speakers conceive of their relationships with the local and regional communities. The attachment to the localized hip-hop culture explained the reason why two speakers spoke differently regardless of the similarity in age and ethnic identity. Habib's (2010) study in Hims, Syria demonstrated a clear linguistic shift from the voiceless uvular stop [q] to the glottal stop [ʔ]. She studied linguistic accommodation in relation to identity, and she linked the adoption of [ʔ] to its embedded urban identity, thus, the adoption of an urban identity. Also, based on her quantitative analysis, there is a clear shift in the speech of the younger generation towards the use of the urban form [ʔ] compared to mere variation or maintenance of the rural form in the speech of the older generation.

This study, however, explores community members' use of *erhua* and *rusheng* as the

consequence of dialect contact against the background of country-owned speech community in China. In detail, it considers how those social variables, such as age, gender and period of residence, as well as the division of northern/southern characteristics, affect the use of *erhua* and *rusheng* within the community.

2.2 STUDIES ON LANGUAGE POLICY IN CHINA

In addition to the naturalistic context in which dialect acquisition often takes place, the educational context can also influence dialect acquisition in China because the use of standard Mandarin penetrates the educational system. This study first reviews the literature on the language policy itself and then lists some related research on the correlation between the language policy and language use.

First, language policy is part of Chinese ideology. China's language policy is rooted in Chinese history. The new China (PRC) was founded in 1949; prior to that, China was a feudal society for over 5000 years. Looking back at history, a unified language worked pretty well to facilitate communication among different dialect speakers. From the view of rulers, it even helped them retain the unification of a nation. Historically, the unification of language use started in 221BC. The Qin government applied an extreme policy to unify the script. They burned books written in other Chinese scripts and executed many scholars who disobeyed the rule. Oppressive as this action was, it broke down many communicative barriers and served well to control people. According to Zhou and Sun (2004), the centralization of power was favored, and from that time on, language planning became part of the political tendency and consideration of the government. The Chinese government considered Putonghua as the standard Chinese in the Forum of Language Reform held in Beijing in 1954. Since 1955, the policy of promoting Putonghua as a common language has been applied all over the country. Basically, the policy

serves the area of education and social media. As a reaction to the prestige status of Putonghua, as is mentioned in Lee and Li (2013), some local dialects or minority languages have gradually given way to Modern Standard Mandarin, and therefore, the study of variation in a dialect contact situation becomes valuable.

Second, even though there is a lack of assessment of China's language policy (Zhou and Sun, 2004), there are generally two directions of research related to this topic. The first direction focuses on the multilingual situation in overseas Chinese migrated groups. In the situation of minority groups living in mainland China, according to Lee and Li (2013), some bilingual policies were applied and bilingual education became available. Meanwhile, migrated Chinese speakers become an important component of multilingual societies in other countries. Lee and Li (2013) refer to the complex situation of the Chinese language used by speakers living outside of China as Chinese language diaspora, and they have conducted research within these Chinese diasporic communities considering the background of each different country. In light of this, this paper tries to study the language variation within a speech community that has experienced labor migration from the north to the south of China.

The second direction is on the languages of minority ethnic groups and dialects of the majority group within China. These studies focus the influence of the language policy in mainland China. First, the division of majority and minority languages in Chinese is based on the amount of each ethnic group's population. Han people are the majority group, and therefore, their language—the "Han language"—is the majority language in China, which is the base of promoted Putonghua. At the same time, another 55 ethnic groups have their own language as well, leading scholars to raise awareness about preserving the language of minority groups (Zhou and Sun, 2004). Second, there are seven different dialects spoken by the Han people in mainland

China (Figure 1.3). There is a clear division between the north and south regarding the dialects spoken in China, and as mentioned previously, Putonghua is based on the northern dialect. Thus, it is difficult for southern people to fully master the variants of some linguistic features in Putonghua. The linguistic variables in this study, *erhua* and *rushing,* are good examples of those linguistic features. As to *erhua*, scholars in China (Qian, 1985; Zhou, 1980) have proposed that people from the southern part of China don't have *erhua* in their dialect and it is hard for people in those regions to master the variants of *erhua* in Mandarin. *Erhua,* also translated as rhotacization, has been studied by many linguists. Zhang (2005) has conducted a sociolinguistic work on rhotacization that raised awareness about the stereotype of rhotacization in the Beijing dialect. Subsequently, much work has been conducted on this topic. Pankhurst (2012), for example, conducted research on Henan speakers' rhotic lenition. Although the author does not totally agree with him that rhotacization is a stigmatized linguistic form in Putonghua, his analysis on the linguistic environment of *erhua* provides some guidance to the current study. His labeling of rhotacization as a stigmatized form in Putonghua is considered by the author to be an overgeneralization. There are some situations in the actual linguistic environment where rhotics must be adopted to differentiate the meanings of words (see Chen, 1999). Therefore, we can hardly define rhotacization as a non-welcomed linguistic pattern by simply referring to northern people's negative language attitude toward the use of it. This study is based on a speech community located in south China and therefore, regionally it has the potential to supplement previous studies on this area. As to *rusheng*, it is a unique suprasegmental linguistic feature that has been preserved in the city's dialect (Xiang dialect), where the speech community under investigation is located. To examine the variation of *rusheng* is also one part of the study on the Xiang dialect. According to Chen (1999), the population who can speak the Xiang dialect is

declining. The economic situation is not as good as in Guangzhou or Hong Kong, where the Yue dialect is spoken. Some excellent Hunan people would not consider coming back to their original area once they have had a chance to study or work in a more developed region. Also, people who migrated from other areas to Hunan province do not learn Xiang dialect, as people who work in the Yue dialect speaking areas do. By considering two linguistic variables that distinctively exist in northern and southern speech, this research hopes to fill the gap in the study of sociolinguistics to detect the variational pattern of a speech community where dialect contact takes place. The study also investigates the use of Putonghua by Xianggang community members in comparison to that of overseas Chinese language speakers.

All in all, observed dialect variation in both naturalistic and educational contexts and influences from family members, peers and the language policy are considered. More specifically, this study focuses on the use of *erhua* and *rusheng* and applying different social groupings, age, gender, period of residence, and the northern and southern division to explain the observed linguistic variation in the Xianggang community.

CHAPTER THREE: METHODOLOGY

This study focuses on the sociolinguistic variables *erhua* and *rusheng* used by speakers in the Xianggang community in China. In this chapter, the variants of each of the two linguistic phenomena and the reason they were chosen as study variables are discussed. Second, the methods of data collection, which includes information about the participants, the questionnaire design and the wordlists are presented. Third, the social factors of age, gender, and period of residence as well as the division of northern and southern characteristics are included. The methods and rationale for the data analysis has also been presented.

3.1 LINGUISTIC VARIABLES

Erhua 'rhotacization' is the first linguistic variable examined in this study. As mentioned previously, *erhua* is treated as a northern-colored form in this study. It is by adding a retroflex [ɻ] at the final syllable that the preceding vowels or consonants become *r*-colored sounds. For example:

1) 小石块 [ɕiɑu ʂɻ kʰuai] 'small stone' – without Erhua

2) 小石块儿 [ɕiɑu ʂɻ kʰuaɻ] 'small stone' – with Erhua

There are three reasons for choosing *erhua* as a variable in this study. First, much research on the dialects in northern China has shown it is typical for speakers in northern China to use *erhua*. For example, in Chen (1999), there are three situations where rhotacized forms are used in the Beijing dialect: semantically differentiated words, habitually rhotacized words, and stylistically informal words. In this study, data was collected by asking informants to read

designed wordlists. There are semantically differentiated words and some habitually rhotacized words. Second, there have been some sociolinguistic works that considered the social meaning of the use of *erhua*; for example, considering socioeconomic factors in linguistic analysis, Zhang's (2005) study on the speech of "a Chinese yuppie in Beijing" has given evidence to the fact that language variation has a constructive force on the categorical change in language use. Basically, she divided interviewed speakers into two general categories: foreign business employees and state-owned business employees. *Erhua* was one of the local linguistic variables she focused on, and the results suggested state-owned business employees use more rhotacized words than foreign business employees. The heavy use of *erhua* by Beijingers has generated a stereotypical description of them as smooth (smooth sounding and as a characteristics trait in a person). Other than the socioeconomic status of these two categorical employees, it is suspected the use of "local" and "contemporary" linguistic features are not only determined by the community identity of these employees, but also what their businesses are oriented to, that is, their contact groups. Third, *Erhua* is a localized linguistic form, and some scholars in China (Qian, 1985; Zhou, 1980) have proposed that people from the southern part of China do not have *erhua* in their dialect, and it is difficult for people in those regions to master the variants of *erhua* in Mandarin. Therefore, the nature of *erhua* provides an obvious opportunity to research language changes in contact. Pankhurst's (2012) findings are considered to be an overgeneralization to mark *erhua* as a stigmatized linguistic form in Modern Chinese, and this study tries to supplement his findings by adding a study on people from the southern area of China.

There are two types of rhotacization according to Pankhurst (2012): vowel and consonant rhotacization. Vowel rhotacization involves applying an *r*-colored sound to the vowel, and consonant rhotacization is much like a rhotic lenition according to Pankhurst (2012). He

considered consonant rhotacization a less salient feature than the former one. Pankhurst (2012, pg. 30) considered the use of consonant rhotacization as a marker of "rebellious, unrestrained response to a national standard." Again, my study tends to test the reliability of his results and to see how much of his researched results can also be applied to migrated group members. His research provides some guidance to my study and especially the grouping of data analysis.

Rusheng, 'entering tone', is the other linguistic variable focused on in this research. *Rusheng* is treated as a southern-colored speech in this study. First, it is a typical southern-colored speech pattern from the view of historical linguistics. As described previously, *rusheng* existed in Middle Chinese and has been excluded from the standard Mandarin. As one of the representative dialects that exist in the southern part of China, the Xiang dialect is spoken in Hunan province, and it preserved the tone contour of *rusheng*. Second, the existence of *rusheng* is distinguished in the Xiang dialect from the other four tones in Mandarin Chinese. Chen (1983, pg. 95) named the *rusheng* spoken by Min people as the fifth tone, and has the feature of "falling contour, glottal [ʔ] stop, and increasing tenseness." However, as stated by Lin and Wang (1995), the Middle Chinese tone *rusheng* has different tone contours in different dialects in China. The dialect in Changsha, the representative dialect of the Xiang dialect, has only preserved the feature of *rusheng* with the contour of [2-4] (Figure 1.4). The same contour of [2-4] has been confirmed in He's (2010) acoustic study on the pitch of the Xiang dialect spoken in Xiangtan city. Third, the use of *rusheng* has been less studied than other supra-segmental features in Chinese, especially in relation to its social meanings. As Chen (1983) found in his research, the frequency of using the fifth tone is related to the dialect background of speakers; for example, 89.4 percent of Min dialect speakers displayed this feature. As a supplement to what he found, the researcher hopes to find some social meaning of the use of *rusheng* by studying the variable

use of *rusheng* by people in the Xiang dialect area.

Generally, *rusheng* is a more regulated feature than *erhua* in Chinese. That is, during the promotion of Putonghua, the use of four tones for each Chinese character has been stipulated in the officially published dictionary, but the rules for the *erhua* are comparatively vaguely stated. Therefore, it is expected that a comparatively more detailed understanding of language use by Xianggang community members will be obtained by considering these two variables together.

3.2 DATA COLLECTION

The data used for this study was collected in the city of Xiangtan, Hunan Province, China. The participants were asked to complete a questionnaire and then read designed wordlists. It took each person 10-15 minutes to complete the task. The whole procedure of the interview was recorded by a SONY IC Recorder (ICD-TX50) in a quiet, indoor environment. For the convenience of participants, the location was chosen near their living areas, such as a quiet, empty room near their offices or a nearby café. The interviews were conducted in Mandarin Chinese; the whole session was run accompanied by one or several Xianggang community members. To obtain relatively natural speech and release the tension on individual speakers, the interviews were mostly conducted in groups. During the interview of one speaker, other speakers were observing. Because of the massive amount of data given in the wordlists, and according to my oral survey after the recordings, participants did not know exactly which linguistic form was being examined.

3.2.1 Participants

Participants were recruited through "friend of a friend" (Milroy, 1980). Introduced by a community member with the northern characteristics, I interviewed 20 people in the middle age group. Also, guided by a teacher in a school within the community, I recruited 19 people in the

younger age group. All of the participants are currently Xianggang community members. Fifty-eight speakers were interviewed; however, only 39 speakers were included in the data analysis because of the quality of the actual recordings. Most of the nineteen speakers excluded from the data analysis were older than 70 years. Because of health constraints, it was difficult for them to finish reading all the wordlists. To make the results generally comparable, this study eliminates that elderly age group and focuses on the middle and younger age groups. Consequently, the age of participants range between 14 and 59 years old. In this study, I divided the 39 informants based on their ages almost equally into two age groups: younger age group (14-25 years) and middle age group (38-59 years). The purpose is to compare the linguistic behavior of these two groups and to see if there is a linguistic difference between these two generations as well as if there is a linguistic change within the community. Table 3.1 presents the social distribution of all the participants.

Table 3.1 Study participants

Speaker#	Age	N/S	Gender	Period of Residence
11.	22	N	F	20-40y
14.	15	N	F	10-20y
17.	14	N	F	10-20y
51.	16	N	F	10-20y
58.	24	N	F	20-40y
12.	23	N	M	20-40y
25.	23	N	M	20-40y
32.	24	N	M	20-40y
33.	23	N	M	20-40y
34.	24	N	M	20-40y
36.	25	N	M	20-40y
57.	25	N	M	10-20y
27.	38	N	F	20-40y
22.	46	N	F	>40y
28.	50	N	F	>40y
29.	49	N	F	10-20y
30.	52	N	F	>40y
37.	51	N	F	>40y

43.	50	N	F	>40y
45.	45	N	F	20-40y
4.	54	N	M	>40y
35.	49	N	M	20-40y
40.	58	N	M	>40y
41.	43	N	M	>40y
42.	50	N	M	>40y
44.	43	N	M	20-40y
48.	52	N	M	>40y
13.	14	S	F	10-20y
18.	14	S	F	10-20y
19.	14	S	F	10-20y
16.	14	S	M	10-20y
20.	14	S	M	10-20y
26.	23	S	M	20-40y
27.	24	S	M	20-40y
1.	53	S	F	20-40y
2.	54	S	F	20-40y
23.	35	S	F	20-40y
24.	45	S	F	>40y
3.	59	S	M	>40y

3.2.2 Questionnaire design

There were two different questionnaires used in the interview. They were designed for workers and student separately (Appendices 1 and 2). Speakers who work in the Xianggang community used the workers' form and speakers still in school used the students' form. The purpose of using questionnaires in this study was to obtain some direct information from the participants. The difference between these two forms considers their current situations and how they position themselves; for example, students were asked about what their desirable jobs are, while workers were asked about their expectations for their children, if any.

Each questionnaire contains ten questions, and the target for each question is different. Every two of these ten questions expects an answer for one of the following five points.

First, each participant was asked to indicate his or her age and gender. The goal of this point is to obtain the data for the first two social factors (age and gender) and to help analyze if

there is any relationship between the use of "*erhua*/*rusheng*" and informants' age/gender.

The participants were then asked about their jobs and whether they identify themselves as "Xianggang people" or not. This question was designed to test their community identity and if community members have the same sense of belonging within the Xianggang community.

Third, they were asked about how long they had worked or lived within the Xianggang community. This point tended to determine if the length of dialect contact had an effect on the adoption of *erhua* and *rusheng* by Xianggang community members.

The fourth point was the one that differed between the workers' (Appendix 1) and the students' questionnaires (Appendix 2). Workers were asked about the current job of their offspring or their expectations for their offspring, i.e. would s/he work within this community or not. Students were asked about their plans to continue working within this community or not. This question aimed to give some thought on the relationship between future plans and speakers' language patterns.

Finally, all the informants were asked about their living circle—if it is convenient to live in their community and if they feel the need to go to the city for any reason. This was to estimate the living circle of the community members and how often these people have contact with other city members who are not living within the Xianggang community.

3.2.3 Wordlist design

There are two wordlists targeting the use of *erhua* and *rusheng* separately (see Appendices 3 and 4). Wordlist 1 adapted 112 phrases from Wang's (2007) research and focuses on the *erhua* of each word within each provided phrase. Because the number of words within one phrase is not equal to another and phrases are meaning-based, it was difficult for informants to determine which linguistic feature was being focused on, which enhances the reliability of the results.

Following the division given by Pankhurst (2012), there are 201 total words in which *erhua* could occur within the 112 phrases; in 87 of the words, *erhua* could occur after nasals and the remaining 114 *erhua* could occur after vowels. Wordlist 2 contains 43 single characters; it is adapted from Yang's (2004) research on *rusheng* in the Xiangtan dialect. According to her research, these 43 sample words displayed most of the evidence on *rusheng* in the Xiangtan Dialect. However, they are pronounced differently and classified within four tones in Mandarin Chinese. The reason these 43 single characters were chosen was because of the "tone sandhi" that exists within tone languages. Especially in Chinese, each character carries one tone and when two characters combine together, the tones affect one another. The researcher chose to obtain a direct result for the study by focusing on one-character words; however, it would be worth studying the two-character words in the future.

As mentioned previously, *erhua* and *rusheng* are regarded as the respective representations of northern- and southern-colored speech patterns separately in Chinese. This paper counts the frequency of the occurrence of *erhua* and *rusheng* in the speakers' collected data.

3.2.4 Recordings and coding

The entire interview session was audio recorded in a quiet room by a Sony IC recorder (ICD-TX50). The recordings started before being noticed by the informants and they contain approximately 353 minutes of interviews. As a Xiang dialect speaker (#4) who is from Hunan province, the author (#1) and two other graduate students (all native Chinese speakers) worked together to code the data and count the frequency of occurrence of *erhua* and *rusheng* (Table 3.2).

Table 3.2 Demography of four researchers

Number#	Origins	Major	Location
1	Xiangtan, Hunan	The Author-M.A. Linguistics	Syracuse University

2	Shijiazhuang, Hebei	M.A. Linguistics	Syracuse University
3	Wuhan, Hubei	M.S. Computer Science	Syracuse University
4	Xiangtan, Hunan	Diploma. Marketing	BICT. Canada

In detail, the graduate student (#2) is from Hebei province, a native speaker of the northern dialect, majoring in linguistics; student (#3) is from Hubei province, a native speaker of the southern dialect, majoring in computer science; the other Xiang dialect speaker (#4) is from Xiangtan city, a native speaker of the Xiang dialect, majoring in accounting. Students #2 and #3 helped code the *erhua* recordings. First, Student #2 helped in listening to the recordings and the results indicate that as a speaker of the northern dialect where *erhua* is one of her dialect features, Student #2 has a more keen sense on the use of *erhua* than the author (#1) does. To ensure accuracy, if at any point the two disagreed with each other on the pronunciation of a certain word, Student #3 was asked to make the judgment after listening to the different pronunciations, with and without *erhua*, pronounced by Student #2. Most of the time, Student #3 agreed with the judgment given by Student #2.

Regarding *rusheng*, because it is typically a distinct tone contour in the Xiang dialect, Student #4 – a native Xiang dialect speaker – worked together with the author (#1) to code the data by listening to the recordings and counting the frequency of occurrence. Because they are both originally from the same area and *rusheng* is one of their linguistic features, they mostly agreed with each other. If not sure at any point, the recordings were played a number of times until total agreement was reached.

3.3 DATA ANALYSIS

3.3.1 Grouping

Speakers can be classified into three groups according to their age, gender and their period of residence as well as the division between northern and southern characteristics within the Xianggang community. All of this information was extracted from the questionnaires the informants were asked to fill out at the beginning of the interview. Table 3.3 shows that the number of speakers from the north is more than that from the local area. In this case, local speakers are more like a control group or a comparison group.

Table 3.3 Participant Demographics

Social Grouping	Numbers of speakers from the North area (N=27)		Numbers of speakers from the local area (N=12)	
AGE	M	F	M	F
14-25	7	5	4	3
38-59	7	8	1	4
TOTAL	14	13	5	7
PERIOD OF RESIDENCE	M	F	M	F
10-20	1	4	2	3
20-40	8	4	2	3
>40	5	5	1	1
TOTAL	14	13	5	7

3.3.2 Descriptive analysis

To offer a broad picture of the linguistic performance of the informants, a descriptive analysis is provided in this paper. Although it has many shortcomings, it is still the most direct and understandable way to display what was discovered. The percentage of the occurrence of *rusheng* and *erhua* has been calculated as in the following section.

$$\text{The percentage of the occurrence} = \frac{\text{Total tokens of occurrence}}{\text{201(total events) * number of speakers}}$$

3.3.3 Statistical analysis

Statistical analysis using SPSS is applied in this paper to help interpret the data. Brace, Kemp, and Snelgar (2000) indicated that the researcher could not make any statement about causation by natural observations. One goal in the current project is to explore the effect of social factors on people's dialect variation in a dialect contact situation. As stated previously, this study considers two linguistic variables and four social variables. To determine how those factors interact with one another, two statistical tests are applied: a t-test and a regression test under the generalized linear model. The t-test aims to check if two means are reliably different from each other. By applying this test, it was possible to determine if the performance of speakers was affected by one variable and had a significant difference from the other variable. For example, by running the data regarding the use of *erhua* by northern and southern-characterized people separately, it could be determined if the variables of northern and southern characteristics generated significant differences on the speech of informants by looking at the p-value. The regression test, particularly the binary logistic regression test, was applied to test the interaction of the four social factors and the two linguistics variables. This gave a more comprehensive statistical report based on the limited data at hand. To simplify the model, each subject was given the same number of words (201) to test their use of *erhua*, and for each word, they had two possible outcomes: either they used *erhua* (Y) or not (N). Therefore, this fits well with the binary regression model with a fixed number of trials (201). In this research, it is possible to predict which social factors are the driving forces of dialect acquisition in a contact situation, as well as how the linguistic environment affects it.

CHAPTER FOUR: DATA ANALYSIS

This chapter consists of three parts. First, the results indicated in the designed questionnaires are presented. Second, the data in the recordings is analyzed, considering both the social and linguistic factors. As stated, social factors include age, gender and period of residence as well as the division between northern and southern characteristics; linguistic factors include the use of *erhua* and *rusheng*. Finally, the findings are summarized.

4.1 QUESTIONNAIRE

Among all 39 informants who attended the interviews, ten used the students' form and 29 used the workers' form. Judging from all the collected data, in addition to the divided categorical groups such as age and gender every individual speaker belongs to, results can be organized as an answer to each of the following four questions.

"Do you identify yourself as a Xianggang community member?" As one of the requirements for the participants' recruitment, all the informants are socially recognized as Xianggang community members. In Hazen (2002), cultural identity has a significant effect on language variation. Meanwhile, self-identification is an essential part of cultural identity; as such, it is worthwhile to count this factor in the study of language variation in the Xianggang community. The results suggest that 35 out of 39 informants identified themselves as Xianggang community members. Among the four people who do not consider themselves as community member, three are male and one is female; three are from the north and one comes from the locale (see Table 4.1).

Table 4.1: Participants who do not consider themselves community members

Subject	Age	Gender	N/S Characteristics	Period of residence
14	15	Female	Northern	10-20 years
25	23	Male	Northern	20-40 years
4	54	Male	Northern	>40 years
16	14	Male	Southern	10-20 years

"Do you work in the Xianggang community, and how long have you been working there?" This question is set for adults only. Different from the question that asks about the "period of residence," workplace is a place where contact happens at a higher frequency in comparison with the broad definition of "living within the community." As is shown in the results, among the 29 adult participants, 24 are currently working within the community; two have retired from the same company within the community. Seventeen informants have been working for a company other than the steel company within the Xianggang community, and all of them are from the middle age group. There are five informants who have been working within the community for less than five years and all are from the younger age group.

"What are your expectations for yourself (in the students' form) or your offspring (in the workers' form)?" Results here are demonstrated separately for each group. For those studying at school, all ten participants attend middle school within the Xianggang community. But as is shown in their responses, none want to come back and work within the same community in the future: two want to work in the capital city of Hunan province; seven want to either go abroad or work in Shanghai or in some provinces that are located beside the sea; one wants to work on her own farm.

For the adult participants, twenty have children and nineteen responded to the question regarding their aspirations for their children. As is suggested in the data, eleven of their children have already had a job, but only three are now working within the Xianggang community. For the eight children who are still in school, five attend school within the community and three do not. However, regarding the expectations these adults have for their children, only one interviewed subject wants her son to work within this community. Four interviewees directly stated they do not want their children to work within this community or in the local area; they hope their children go out to see a broader world than they themselves have.

Therefore, we can briefly conclude from the answers that the desired work for the majority of the interviewees is not within the community or locally. Students want to go out from the local city to more developed cities, which also fits with their parents' expectations.

"How do you feel about the convenience of living within the community?" The data indicates it is quite convenient for the majority of subjects to live within this community. The equipment and grocery stores meet their needs quite well, but still they often go to other places in the city for some recreational needs. Only three out of 39 interviewees mentioned they do not go to the city very often.

Generally, the questionnaire indicates that the Xianggang community is an open and convenient place to live in. The community itself does not have many geographical constraints for dialect contacts between northern and southern dialects. Workers and students have been living in different social conditions: for workers, their most frequently contacted group members are living within the Xianggang community at their work places; for students, their mostly contacted group members are at the school. Workers' self-identification plays a role in their language use, while students' future desires are important as well. Based on the results of the

questionnaires, social factors such as age, gender and period of residence as well as the division of northern/southern characteristics are examined as follows.

4.2 QUANTITATIVE RESEARCH ON THE USE OF *ERHUA* AND *RUSHENG*

After looking at the result from questionnaire, the use of *erhua* and *rusheng* has been studied with the consideration of four social variables: age, gender, and period of residence as well as the division between northern and southern characteristics. A quantitative research has been provided in this session by using SPSS. Before investigating the influence of each social factor, the distribution of the use of the two linguistic variables, *erhua* and *rusheng*, is presented in Table 4.2. This table provides the basic information for the following analyses, and it will be separated into two parts, the analysis of *erhua* and the analysis of *rusheng*.

Table 4.2: Distribution of *erhua* and *rusheng* for each speaker

Speaker#	N. tokens				%			
	erhua after vowel	*erhua after nasal*	*erhua*	*rusheng*	*erhua after vowel*	*erhua after nasal*	*erhua*	*rusheng*
11.	4/114	0/87	4/201	0/43	3.5%	0%	2%	0%
14.	7/114	0/87	7/201	0/43	6.1%	0%	3.5%	0%
17.	0/114	0/87	0/201	0/43	0%	0%	0%	0%
51.	2/114	0/87	2/201	0/43	1.8%	0%	1%	0%
58.	3/114	0/87	3/201	0/43	2.6%	0%	1.5%	0%
12.	3/114	0/87	3/201	0/43	2.6%	0%	1.5%	0%
25.	3/114	0/87	3/201	0/43	2.6%	0%	1.5%	0%
32.	2/114	0/87	2/201	0/43	1.8%	0%	1%	0%
33.	3/114	1/87	4/201	0/43	2.6%	1%	2%	0%
34.	6/114	3/87	9/201	0/43	5.3%	3%	4.5%	0%
36.	2/114	1/87	3/201	0/43	1.8%	1%	1.5%	0%
57.	6/114	0/87	6/201	0/43	5.3%	0%	3%	0%
27.	9/114	1/87	10/201	0/43	7.9%	1.1%	5%	0%
22.	9/114	0/87	9/201	2/43	7.9%	0%	4.5%	4.7%
28.	4/114	1/87	5/201	3/43	3.5%	1.1%	2.5%	7%
29.	0/114	0/87	0/201	1/43	0%	0%	0%	2.3%
30.	2/114	0/87	2/201	1/43	1.8%	0%	1%	2.3%
37.	2/114	0/87	2/201	4/43	1.8%	0%	1%	9.3%
43.	7/114	0/87	7/201	2/43	6.1%	0%	3.5%	4.7%
45.	8/114	7/87	15/201	2/43	7%	8%	7.5%	4.7%
4.	9/114	0/87	9/201	6/43	7.9%	0%	4.5%	14%
35.	8/114	9/87	17/201	1/43	7%	22%	8.5%	2.3%

40.	15/114	7/87	22/201	3/43	13.2%	14.6%	10.9%	7%
41.	4/114	0/87	4/201	0/43	5.6%	0%	2%	0%
42.	17/114	6/87	23/201	4/43	23.6%	14.6%	11.4%	9.3%
44.	3/114	2/87	5/201	4/43	4.2%	4.9%	2.5%	9.3%
48.	25/114	11/87	36/201	5/43	38.9%	29.3%	17.9%	11.6%
13.	2/114	0/87	2/201	0/43	1.8%	0%	1%	0%
18.	0/114	0/87	0/201	0/43	0%	0%	0%	0%
19.	0/114	0/87	0/201	0/43	0%	0%	0%	0%
16.	0/114	0/87	0/201	0/43	0%	0%	0%	0%
20.	1/114	0/87	1/201	0/43	0.9%	0%	0.5%	0%
26.	2/114	0/87	2/201	0/43	1.8%	0%	1%	0%
27.	4/114	0/87	4/201	0/43	3.5%	0%	2%	0%
1.	9/114	4/87	13/201	1/43	7.9%	4.6%	6.5%	2.3%
2.	2/114	0/87	2/201	2/43	1.8%	0%	1%	4.7%
23.	4/114	1/87	5/201	0/43	3.5%	1.1%	2.5%	0%
24.	1/114	0/87	1/201	1/43	0.9%	0%	0.5%	2.3%
3.	2/114	0/87	2/201	0/43	1.8%	0%	1%	0%
Total	190/4446	54/3393	244/7839	42/1677	4.3%	1.6%	3.1%	2.5%

4.2.1 Analysis of *erhua*

Erhua is a phonological process that takes place at the end of a syllable. As introduced before, the Chinese syllable consists of three parts: initial syllables, final syllables, and tone (Chen, 1999:34). In other words, the final syllable includes nucleus and coda; the initial syllable is the onset. A Chinese syllable can end with either consonants or vowels; therefore, there are two kinds of linguistic environment for *erhua*: post-vocalic and post-nasal. This section starts with the analysis of the influence of the four social factors on the use of *erhua*. Then, the effects of the two different linguistic environments are considered.

First, to determine whether the frequency of the use of *erhua* is different between the speech pattern of younger and middle age groups, the mean percentages of *erhua*, regardless of the gender differences, are analyzed, as shown in Figures 4.1-4.3.

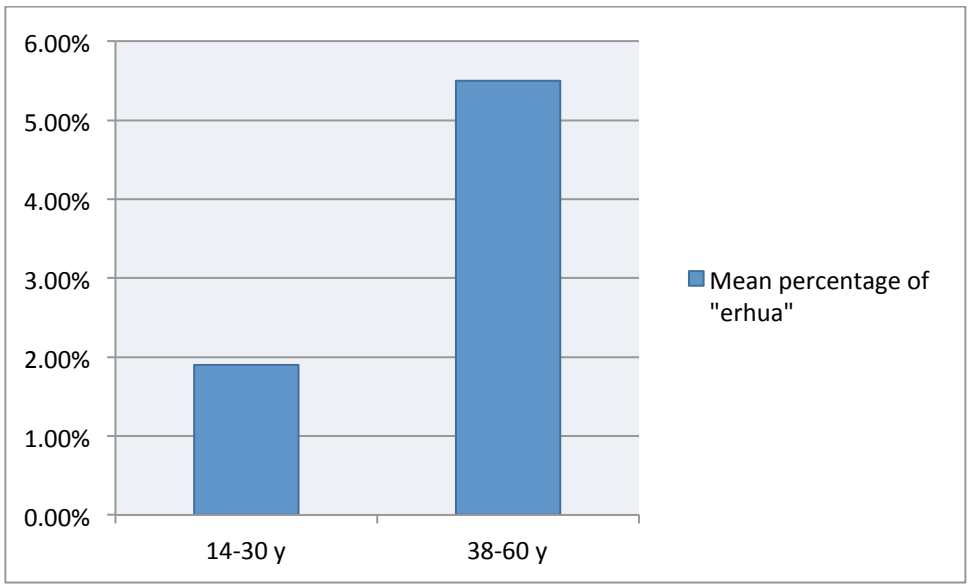

Figure 4.1: The percentage of the occurrence of *erhua* - people with the northern characteristics in the two age groups

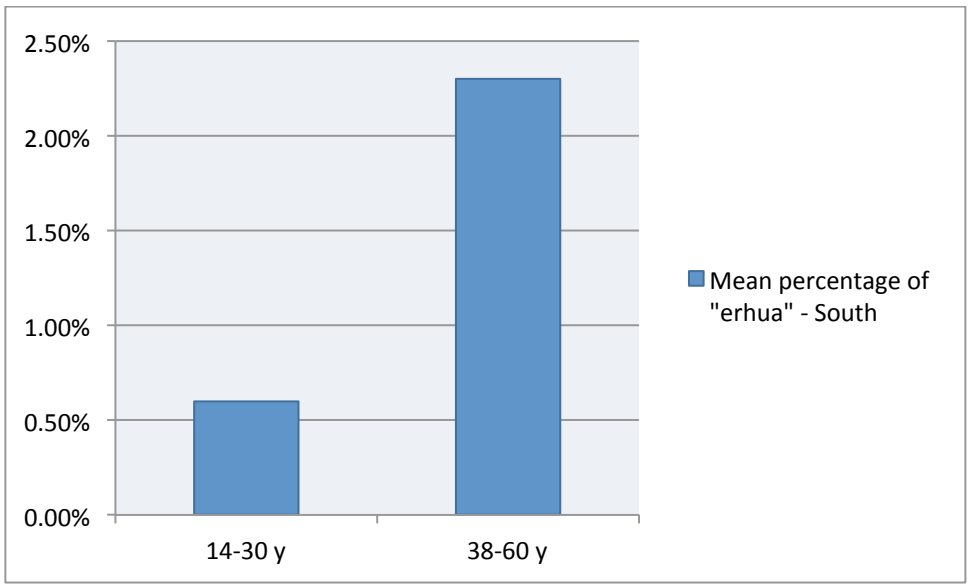

Figure 4.2: The percentage of the occurrence of *erhua* - people with the southern characteristics in the two age groups

As is shown in Figure 4.1, for people with northern characteristics, the middle age group has higher potential in their use of *erhua* than people in the younger age group. On average, the

use of *erhua* by the middle age group members is around three times more than that of the younger age group members.

Figure 4.2 suggests that the southern-characterized people present a similar trend of language use to the northern-characterized people. That is, the percentage of the use of *erhua* is higher in the middle age group than in the younger age group.

Regardless of the north and south division, t-tests were conducted to check if the different use of *erhua* is significantly different from one age group to another. The results confirm that for both southern and northern-characterized people, the middle age group *(M=9.45; SD=9.24)* report significantly more use of *erhua* than the younger people *(M=2.89; SD=2.45), t(21)=-3.059, P=.006.*

Second, to compare the use of *erhua* by northern and southern-characterized people, Figures 4.1 and 4.2 are combined together. Figure 4.3 suggests northern-characterized people use more *erhua* than southern-characterized people do on average. Both younger and older age groups with northern characteristics use more *erhua* than those from the same age groups with southern characteristics. This correlates with my initial/introductory statement that *erhua* is a northern-colored speech pattern.

Also, the result of the t-test confirms that northern-characterized people *(M=7.85; SD=8.29)* report significantly more use of *erhua* than southern-characterized people *(M=2.67; SD=3.60), t (37)=2.72, P=.01.*

Third, Figures 4.4 and 4.5 are presented to indicate if there is a difference between males and females in the northern and the southern-characterized speakers.

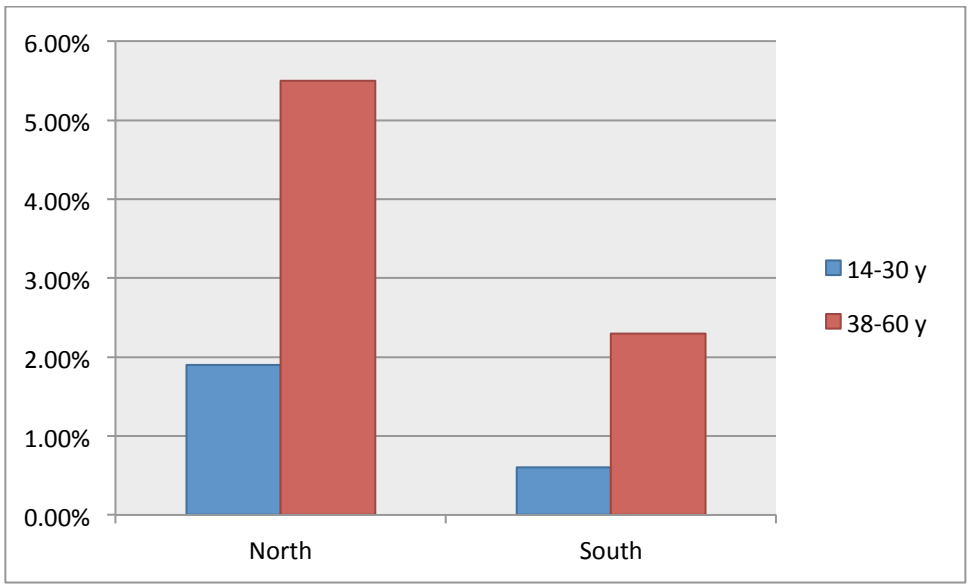

Figure 4.3: The percentage of the occurrence of *erhua* - all of the informants in the two age groups

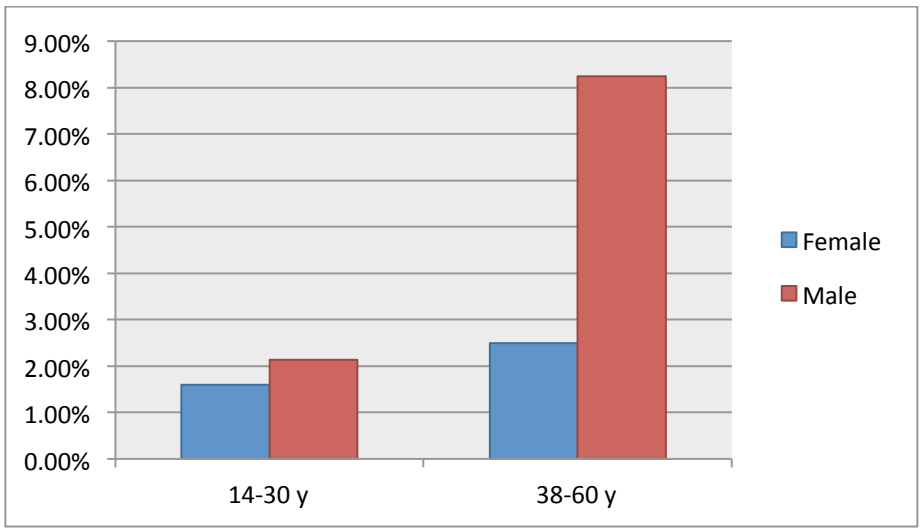

Figure 4.4: The percentage of the occurrence of *erhua* in different age and gender groups - northern characterized people

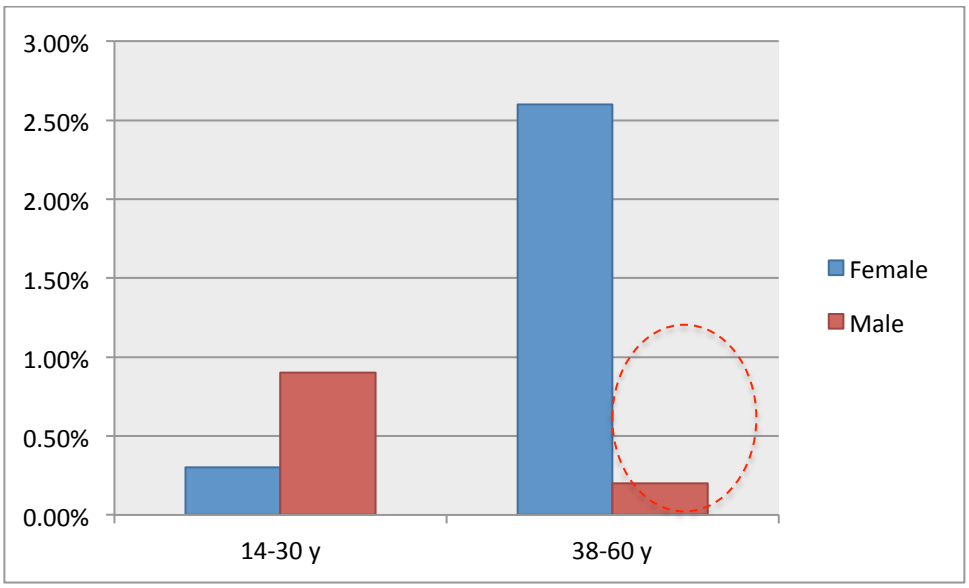

Figure 4.5: The percentage of the occurrence of *erhua* in different age and gender groups - southern characterized people

Figure 4.4 indicates that males from both the younger and middle age groups with northern characteristics display, on average, higher frequency of *erhua* than females do. However, Figure 4.5 shows that males in the middle age group with southern characteristics show less use of *erhua* than females. This seems to go against what was observed in Figure 4.4, that males have more use of *erhua* than females. To test the reliability of this information, the members of the southern group have been examined. As indicated in Table 3.1, three females and one male compose the middle-aged group with southern characteristics. Thus, the comparison may not be sufficient at this stage because of the limited data from males in this age group and the unbalanced number of males and females. The t-tests for these two groups also confirm there is no significant difference generated by gender (*P=.136*).

Fourth, to consider the effect from period of residence to community members' use of erhua and rusheng, together with three different periods, a general profile of all the speakers is presented in Table 4.3. As we can see from the table, the majority of informants have been living within this community for more than 20 years. Comparing the two groups, the total number of

people who carry northern characteristics is higher than the number of informants with southern characteristics.

Table 4.3: Period of residence

	10 – 20 years		20 - 40 years		More than 40 years		TOTAL
N/S C	M	F	M	F	M	F	
Northern	1	4	8	4	5	5	27
Southern	2	3	2	3	1	1	12
TOTAL	3	7	10	7	6	6	39

To consider the use of *erhua* by people in different periods of residence, and the division between northern and southern characteristics, the possibility of the occurrence of *erhua* in the speech of residents in the Xianggang community can be clarified as follows (Figure 4.5). The differences between northern and southern-characterized speakers are at their most obvious for residents who have lived in the Xianggang community for more than 40 years. For those living there for 20-40 years, the differences between the two are the least. Because the analyzed informants are the second and third generations of Xianggang community people, they themselves did not experience any migration as their elder generations did. The "period of residence" might have some overlap with "age." The regression test has shown significance in the interaction between "age" and "period of residence" (*P=.002*).

In addition to the three social factors, two linguistics environments were taken into consideration. To provide a clear and reliable interpretation of all these factors, typically how close these factors relate to the use of *erhua*, a binary regression test under the generalized linear

model was conducted in SPSS. Before starting the analysis, four social factors are categorically labeled (see Table 4.4).

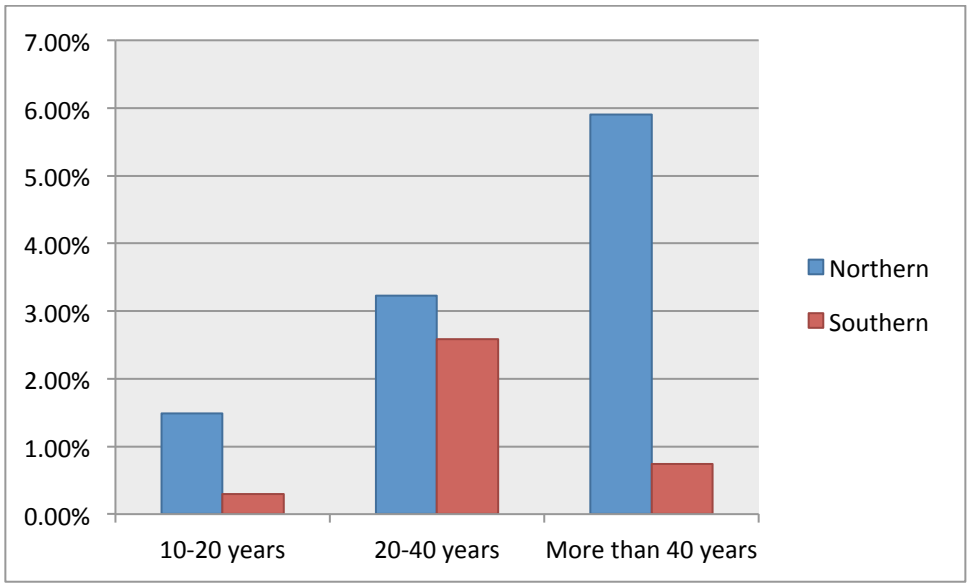

Figure 4.6: The percentage of the occurrence of *erhua* according to period of residence

Table 4.4: Categorical labels of social factors

	Labels		
Social factors	*Label "1"*	*Label "2"*	*Label "3"*
Age	10-20y	20-40y	>40y
Gender	Female	Male	
Period of Residence	10-20y	20-40y	>40y
N/S characteristics	North	South	

First, the use of *erhua* as the dependent variable was chosen and the number of trials was entered as 201. Then, under the column of "predictors," this study's four social factors were selected as "factors" because all of them are categorical factors. Meanwhile, the two linguistic environments, post-vocalic and post-nasal, were selected as covariates because they are

continuous numbers. Third, the "main effect" model was chosen. It aims to test the individual effects of these social factors and linguistic factors on the use of *erhua*. For estimation, the Pearson chi-square scale parameter method was applied as the scale parameter method to test goodness of fit of the chosen model. The result of the main effect model suggested three social factors, "age" *(P=.041),* "north-south division" *(P=.031)* and "period of residence" *(P=.007),* were significant predictors of community members' use of *erhua,* especially in the post-vocalic linguistic environment *(P=.000)*. Similar to the results of the t-test, gender emerged as being statistically insignificant.

Second, to test for interaction between the individually significant factors, the three social factors, and the one linguistic environment from the main effects model, three regression tests were implemented, using the "all two-way" model. The results revealed that the interaction between post-vocalic *erhua* and "period of residence" is statistically significant *(P=.000)*. According to the value of coefficient B,[4] and the exponential of the coefficient *Exp(B)*, the weight for each of these three environments could be measured. Thus, comparing among these three groups of data, residence between 10 and 20 years (*B=. 201, Exp(B)=1.223*), residence between 20 and 40 years (*B=.250, Exp(B)=1.283*), and residence for more than 40 years (*B=.143, Exp(B)=1.154*), the second group had the highest value of Exp(B*)*. As such, there is little difference in the use of post-vocalic *erhua* among these three residential groups. Also, the interaction between the two social factors the division between northern and southern characteristics and age is significant *(P=.000)*. The results of Exp(B) indicate that the effect on the use of post-vocalic *erhua* by people from two different groups is quite close to each other:

[4] The B coefficient is the estimates resulting from an analysis performed on variables that have been standardized so that they have the same variances (zero in this study). This usually helps to explain which of the independent variables has greater impact on the performance of the dependent variable.

B=.147 and *Exp(B)=1.1595* for the younger age group while *B=.132* and *Exp(B)=1.14* for the middle age group. Younger group speakers' use of post-vocalic *erhua* has slightly less influence than middle age group speakers' use. Moreover, there is a significant interaction between the division between northern and southern characteristics and post-vocalic *erhua (P=.000)*; southern-characterized speakers' use of post-vocalic *erhua (B=.449, Exp(B)=1.191)* has a slightly higher influence on community members' use of *erhua* than northern-characterized speakers' use *(B=.105, Exp(B)=1.141).*

4.2.2 Analysis of *rusheng*

Rusheng has its 2-4 tones preserved in the Xiang dialect. Because it is an obvious supra-segmental feature that does not exist in Mandarin Chinese, the use of *rusheng* can be explored just by counting its frequency of occurrence.

As in the case of *erhua*, regardless of the gender differences, the use of *rusheng* is examined against age differences. Figure 4.7 indicates the results.

Figure 4.7 indicates there is a large gap between these two age groups: younger age group informants have zero percentage of the use of *rusheng*, while on average the middle age group uses *rusheng* 2.1 percent of the time. The results of the t-test show that the older age group *(M=2.1, SD=1.78)* uses significantly more *rusheng* than the younger age group *(M=0, SD=0*, *t (19)=-5.3, P=.000).*

Second, when comparing the use of *rusheng* by females and males, males seem to use more *rusheng* than females do. However, the t-test suggests this difference is not statistically significant (*P=.63*) (see Figure 4.8).

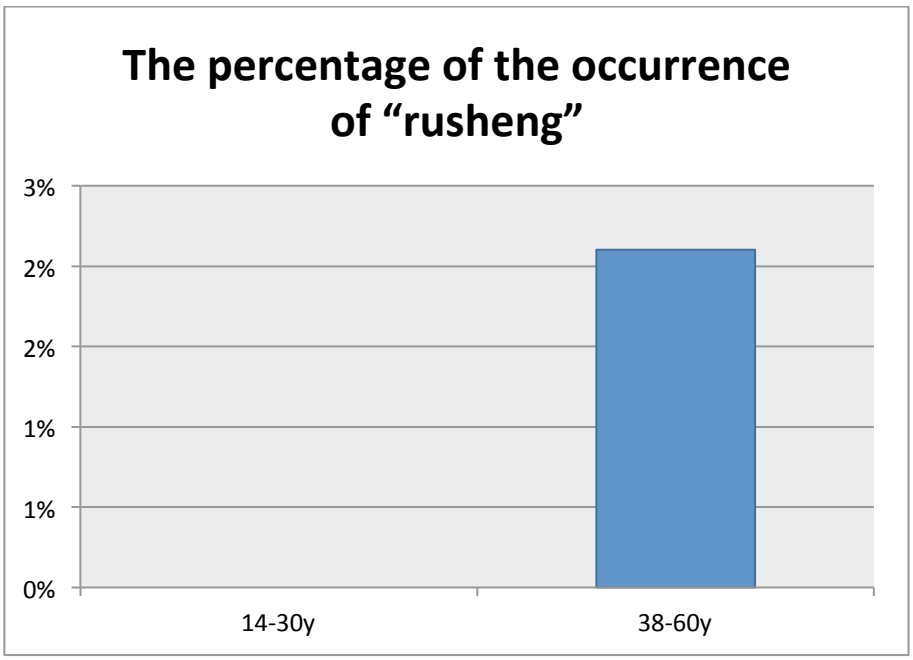

Figure 4.7: The percentage of the occurrence of *rusheng* in all speakers

Third, to examine how the use of *rusheng* relates to the factor of “period of residence,” the percentage of the occurrence of *rusheng* is displayed in Figure 4.9. Similar to what we have found in the use of *erhua,* residents living in the community for 20-40 years have the same percentage in their use of *rusheng*. Also, for residents living in the community for more than 40 years, the differences between northern and southern people are the highest. However, the regression tests suggest that the effectiveness of the “period of residence” only plays a significant role when solely considering its effect (*df=1.474; P=.000*).

Fourth, considering the factor of “north and south characterized divisions,” the results are presented in Figure 4.10.

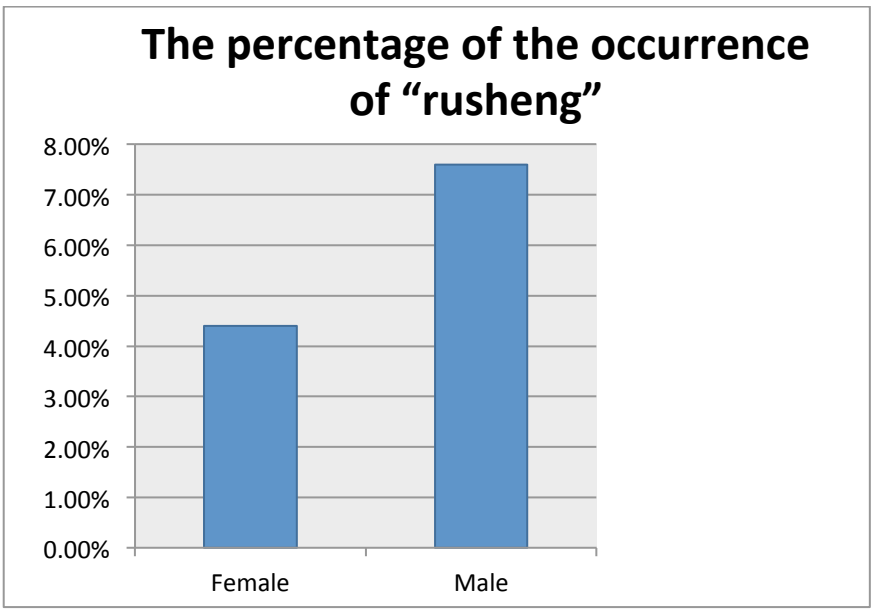

Figure 4.8: The percentage of the occurrence of *rusheng* by northern characterized speakers in both genders

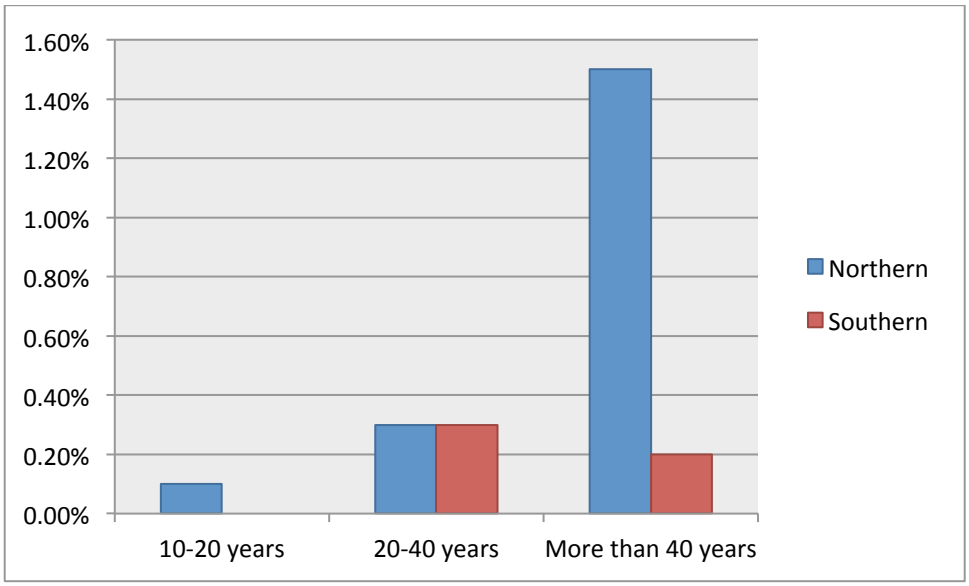

Figure 4.9: The percentage of the occurrence of *rusheng* according to the period of residence

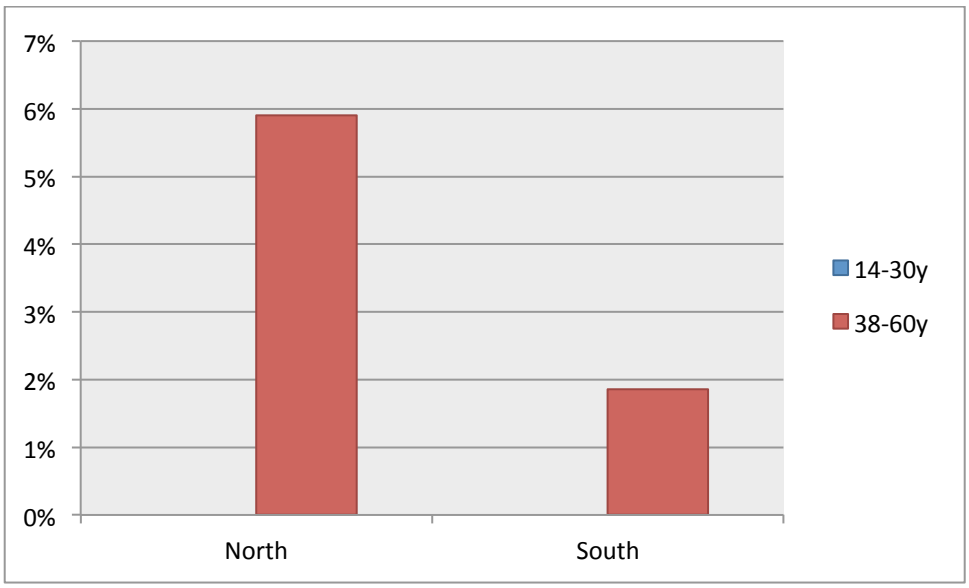

Figure 4.10: The percentage of the occurrence of *rusheng* by all informants in both age groups

Figure 4.10 indicates some differences among middle age people. People with northern characteristics have more use of *rusheng* than those who have southern characteristics. The t-test suggests the northern-characterized people (*M*=1.41, *SD*=1.85) use *rusheng* significantly more than southern-characterized people (*M*=.33, *SD*=.65, *t* (35)=2.67, *P*=.01).

To consider all of these four social factors together, a binary logistic regression test suggests that age groups *(P=.000)* and the division between north and south characteristics *(P=.008)* have significant effects on the use of *rusheng*. Gender *(P=.134)* and the period of residence *(P=.228)* are not statistically significant. However, when solely considering the factor of period of residence, the result suggests it is significant (*P=.000*). More specifically, the factor of period of residence is significant in the community members' use of *erhua (P=.000)* but not *rusheng (P=.228)*. Also, northern-characterized people *(B=1.030, Exp(B)=2.8)* use more *rusheng* than the southern-characterized people within the Xianggang community. However, this result can only be applied to the middle age group because younger age group didn't have any use of *rusheng* in this study.

According to the linguistics environment observations, there are three characters that are mostl often pronounced with *rusheng* (see Table 4.5). As recorded, the syllabic structure of *rusheng* is also different. Syllables always end with a stop consonant, such as [–p], [-t], [-k], or [-ʔ] (Chen, 1983; Lin and Wang, 1995). However, based on the meaning of these three most-frequently occurring characters, it can also be assumed that the use of *rusheng* within the Xianggang community is lexically related. These three characters (see Table 4.5) might be used frequently in the context of a steel company, where 'quality,' 'lost,' and 'fit' are commonly used in their working environments.

Table 4.5: Three characters with the most use of *rusheng*

	Word	IPA transcript in Mandarin	Tone in Mandarin	Glossary
18	湿	'tʂɿ'	4	'Quality'
19	失	'ʂɿ'	1	'Lost'
37	适	'ʂɿ'	4	'Fit'

To summarize what was found in the analysis of recordings, age and the division between northern and southern characteristics are significant social variables to the explanation of the use of both *erhua* and *rusheng*. Period of residence emerged as statistically significant only in the case of erhua. The middle age group has more use of these two linguistics variables than the younger group; also, the northern-characterized people use more of both *erhua* and *rusheng* than southern-characterized people do. Linguistically, on one hand, *erhua* is used at a higher frequency than *rusheng* in general; on the other hand, post-vocalic is the preferable linguistic environment for the use of *erhua* by speech community members. There is significant interaction between post-vocalic *erhua* with each of the three social factors in this study: age, north/south division, and period of residence. However, although *rusheng* is used less frequently compared to *erhua*, the use of it is mainly in three characters. Also, there is a big difference between the

performance of students and that of workers. Only workers use *rusheng* in their speech patterns, and the use of *rusheng* is reserved mostly in three lexical characters.

In general, to combine the information gathered from the questionnaires and that of the data analysis, Xianggang community members' use of *rusheng* is not divided along male/female lines but rather along the lines of age and northern/southern characteristics. The use of *erhua,* however, in addition to age and the division between northern and southern characteristics, is also divided along the line of period of residence. The difference between northern and southern characterized speakers in their use of *erhua* is at the most obvious for those who have been living within the speech community for more than 40 years. The older speakers tend to use more of both *erhua* and *rusheng* than the younger interviewees. Also, northern-characterized speakers use more *erhua* and *rusheng* than southern-characterized speakers do. It is understandable for the northern-characterized speakers to adopt more *erhua* because it is a north-colored speech pattern. But how can the higher use of *rusheng* by the middle-aged, northern-characterized speakers than the southern people be explained? One may also wonder about the young group: why did they refuse to use *rusheng* regardless of the north/south characteristics they have? To offer a better interpretation of these results, a deeper discussion is given in the next chapter.

CHAPTER FIVE: DISCUSSION

To interpret our findings in this study, this chapter will lead a discussion as a response to the three research questions proposed at the beginning of this paper.

First, there is "dialect variation" within the Xianggang community. Looking back to what has been learned from previous analysis, even the northern people use more *rusheng* than the southern people despite the fact that it is a southern feature. Hence, one cannot in this situation say acquisition, but variation only. If the adoption of *rusheng* were based on acquisition, we would see a similar pattern among the southern people. Rather, it seems that the northern people are adopting *rusheng* more than the southern people either to sound more southern or because they work in the factory, where three words with *rusheng* are used frequently. The fact that southern people are using less *rusheng* than *erhua* probably suggests their accommodation to the standard Putonghua. Also, young speakers have different language patterns to those of middle-age speakers. Specifically, young speakers have no use of *rusheng* and low use of *erhua*, while the middle age group shows variation and differences in the use of these two variables. The result strongly supports that there is dialect variation according to age and the division between northern and southern characteristics within the community. As to the acquisition process, because the author only collected data at one point in time, we can't conclude any from this study. As defined in Rys (2007), second dialect acquisition is the acquisition of varieties of one language chronologically later than the first language. This study is on varieties of Chinese, but it is hard to tell when speakers have their first exposure to the use of *erhua* (a feature of the northern dialect) and *rusheng* (a feature of the southern dialect), and which one is earlier than the

other. However, timeline has been considered essential to the study of dialect acquisition (Chamber, 1992; Straw & Patrick, 2007.) As such, to suggest there is an acquisition process, a continuous record on the use of *erhua* and *rusheng* by all age group speakers would be essential. However, we don't have all the data at this point. On one hand, we do not have data from younger speakers to see if they show the same pattern like students and workers or not; On the other hand, all the informants in this research are from the second or third generations of Xianggang community members, and they did not experience migration like the first generation did. Even though they have been classified as either northern or southern characteristics according to their family background (migrants or not), the fact that they were born locally in the Xiang dialect speaking area requires more evidence to support "acquisition" than could be provided in this study.

Age and the division between northern and southern characteristics emerged as significant social variables that have effects on community members' use of both *erhua* and *rusheng*. Period of residence, additionally, is a significant social variable that has an effect on the use of *erhua*. Linguistically, *erhua* has been used more frequently than *rusheng,* particularly more frequently in the post-vocalic environment than the post-nasal one. Pankhurst (2012) has a similar finding on Henan Mandarin speakers that use *erhua* after vowels as an overt marker of northeastern identity, while *erhua* after a nasal tone is less overt and not related to that identity. Since Henan is located in north China and northern-characterized speakers also favor post-vocalic *erhua* more than southern characteristics within the Xianggang community, this study to some extent supports Pankhurst's (2012) finding that post-vocalic *erhua* is an overt marker of northern people. However, instead of the social variable of identity, two other social factors in this study have been identified to have significant influence on community members' use of *rusheng,* and

three factors have suggested having impact on community members' use of *erhua*. A detailed explanation of these social factors will be presented below.

In order to interpret the relationship between age and the use of *erhua* and *rusheng* within the Xianggang community, scatter diagrams are presented (Figure 5.1-5.2). On the basis of the results of the statistical analysis, these scatter plots help us organize the evidence and then lead the discussion about the result by viewing the variation pattern of three age groups. In Figures 5.1-5.2, each symbol in the scatter diagrams represents one speaker, and different kinds of symbols in the diagrams represent speakers in different age groups. The x-axis stands for the time of the use of *erhua*, and the y-axis represents the time of the use of *rusheng*. For example, the symbol at (5, 4) means that this speaker has used *erhua* five times and *rusheng* 4 times. Generally, there is a changing tendency towards lower use of both variables *rusheng* and *erhua* from the middle to young age groups. Also, people in the middle age group display more variation than people in the young age group.

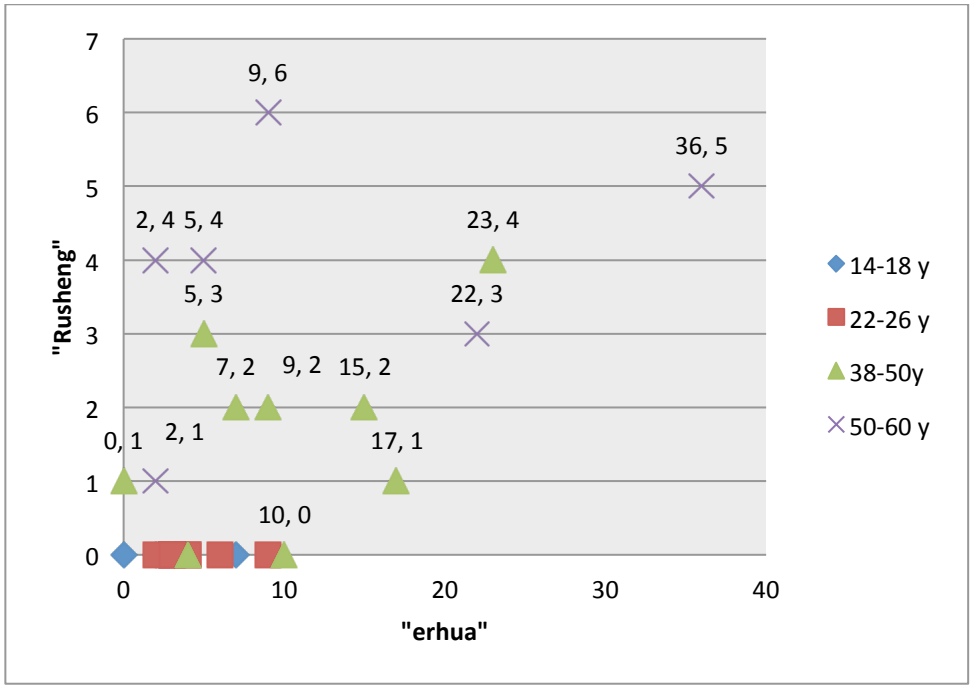

Figure 5.1: Scatter diagram of the two linguistic variables by Northern characterized speaker

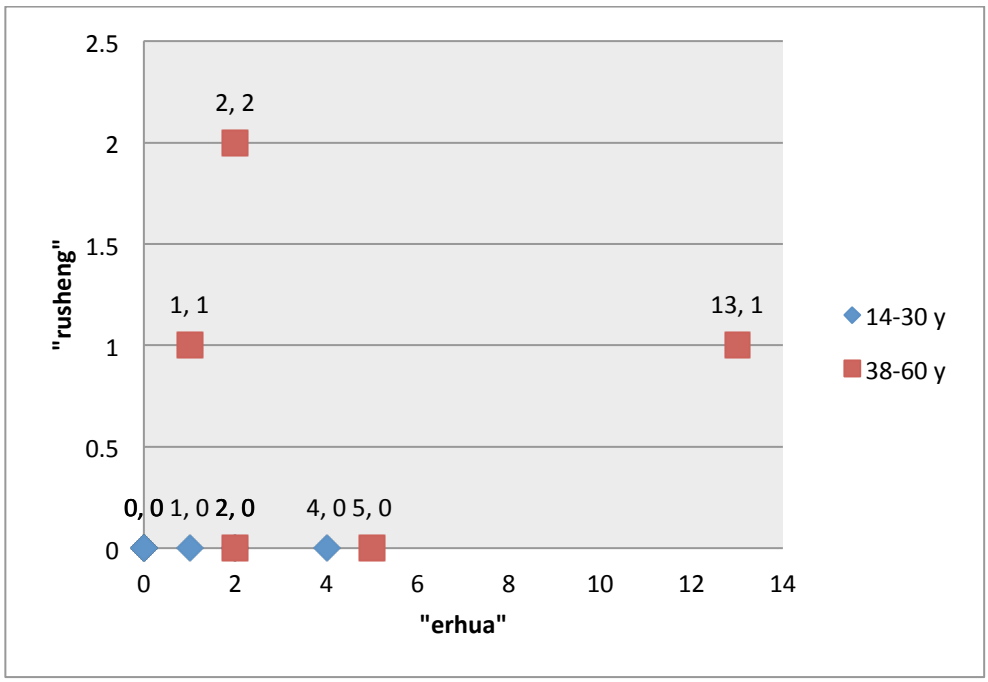

Figure 5.2: Scatter diagram of the two linguistic variables by southern characterized speakers

One of the differences between people in the middle and young age groups is that the team could hardly find any use of *rusheng* by speakers within the younger age groups (14-30). This suggests that the use of *rusheng* is only preserved in the middle age group, at least in the context of an interview. However, the use of *rusheng* by middle age people is mainly in three specific lexical characters that are frequently used in their social context, the steel company. This context-related linguistic pattern supports Kerswill's (1996) consideration of social contact as an important factor in the spread of linguistic changes. And it is speculated that these middle age group speakers acquire these three lexical words with the tone of *rusheng.* In addition, social contacts and social contexts for different age groups affect community members' use of either *erhua* or *rusheng*. Another difference is that young speakers use less *erhua* than middle age speakers and have less variation as well. One special situation among the young speakers is that

they have lately been experiencing Putonghua education.[5] In the context of Chinese society, it is time to consider the effect of government language planning on language education. Chen (2002) stated in his article that Putonghua education brings about much language variation in contemporary China. Variation occurs in Chinese dialects and minority languages as a response to the overt prestige of Putonghua. He mentioned that for dialects that have more tones than the standard Mandarin, the number of their tones has decreased (towards four tones). Based on his survey, *rusheng* has disappeared in the speech of adolescents in many dialect areas. The current study on a younger age group's linguistic variation also supports his findings. The same matter applies to *erhua*. Chen (1999, pg. 39) states that Modern Standard Mandarin (Putonghua) is "very selective in the admission of rhotacized '*erhua*' words and only those semantically distinct words or customarily preferred sounds have been admitted." In fact, one of the differences between Putonghua and northern dialects is that Putonghua has less *erhua* than northern dialects do. Thus, the decreased use of *erhua* in younger people compared to middle age speakers becomes understandable.

The effect from the division between northern and southern characteristics is significant in the dialect variation in the Xianggang community. Since the classification of northern and southern characteristics is determined by informants' family background, this suggests that there might be an impact due to the family's linguistic background. As presented previously in Figure 4.3, which displays perfect evidence regarding the influence of the family's linguistic background: northern-characterized people's use of *erhua* (the northern feature) is more than

[5] Putonghua education focuses on the Chinese pinyin, that is, it teaches the pronunciation of each letter. Phonological processes like "rhotacization" (*erhua*) are not included or are strictly regulated. Therefore, the use of *erhua* is eliminated but still preserved in some meaning-distinguished words.

southern-characterized people. However, this is not the whole picture. What is displayed in Figure 4.9 suggests a more complex picture than before: northern-characterized speakers use more *rusheng* (southern speech pattern) than southern-characterized people do. Since only middle age speakers use *rusheng*, in order to compare northern-characterized speakers' use of *rusheng* with that of southern-characterized speakers, middle age speakers' usage is presented in Tables 5.1 and 5.2.

Table 5.1: The distribution of *rusheng* by middle-aged northern speakers

#Speakers	Total Occurrence	%	Lexical items	
27	0/43	0%		
22	2/43	4.7%	适'fit', 质'quality'	
28	3/43	7%	质'quality',失'lost',戚'relative'	
29	1/43	2.3%	适'fit'	
30	1/43	2.3%	质'quality'	
37	4/43	9.3%	擦'wipe', 质'quality',失'lost', 适'fit'	
43	2/43	4.7%	失'lost', 戚'relative'	
45	2/43	4.7%	适'fit', 质'quality'	
4	6/43	14%	质 'quality', 失 'lost', 适 'fit', 戚 'relative', 莫'don't', 咋'how and why'	
35	1/43	2.3%	质'quality'	
40	3/43	7%	擦'wipe', 质'quality',失'lost'	
41	0/43	0%		
42	4/43	9.3%	擦'wipe', 质'quality',失'lost', 适'fit'	
44	4/43	9.3%	恰'just', 质'quality',失'lost', 适'fit'	
48	5/43	11.6%	失'lost',湿'wet', 质'quality', 适'fit', 踢'kick'	
TOTAL	38/645	6%	Lexical items	%
			质'quality'*11 失'lost'*8 适'fit'*8	29% 21% 21%
			戚'relative'*3 擦'wipe'*3 踢'kick'*1 莫'don't'*1 咋'how and why'*1 湿'wet'*1 恰'just'*1	8% 8% 2.6% 2.6% 2.6% 2.6% 2.6%

Table 5.2:The distribution of *rusheng* by middle-aged southern speakers

#Speakers	Total Occurrence	%	Lexical items	
1	1/43	2.3%	适'fit'	
2	2/43	4.7%	蜜'honey', 逆'opposite'	
23	0/43	0%		
24	1/43	2.3%	质'quality'	
3	0/43	0%		
TOTAL	4/215	2%	Lexical items	%
			适'fit'*1	25%
			蜜'honey'*1	25%
			逆'opposite'*1	25%
			质'quality'*1	25%

As can be seen in the two tables, northern-characterized people's use of *rusheng* has been focused on three lexical items. 15 northern-characterized people used *rusheng* 39 times in total, and 27 of those times were among these three lexical items (质'quality' 29%, 失'lost' 21%, 适'fit' 21%). Southern-characterized people (five people in total), however, used *rusheng* four times and evenly distributed into four different lexical items that did not include 失 'fit'. Since two out of four words used by southerners are similar to the three frequent words used by northerners, this suggests that northern-characterized speakers choose to use *rusheng* in the same characters as southerners do. Even though we don't have enough data to further prove whether the adoption of *rusheng* is systematic or not, we could conclude from the data that northern-characterized people's adoption of *rusheng* is lexically related. Since most of the speakers from the middle age group are working within the steel company, it can be assumed that social context, such as a factory environment, has affected Xianggang community members' linguistic behavior.

The social variable of period of residence has been indicated as significant in the community members' use of *erhua,* but not necessarily in *rusheng*. However, it sheds light on

the possible development trajectory of the community language. As is shown in Figures 4.6 and 4.9, among the three different groups of period of residence (10-20 years, 20-40 years, and more than 40 years), the difference between northerners and southerners is bigger in the more than 40 group than the other two groups in the use of *erhua* (5.16% more) and *rusheng* (1.3%). In contrast, the smallest gap of northerners and southerners exists in the 20-40 group (*erhua* 0.44%; *rusheng* 0%). The gap between the northern and southern groups runs in the middle for the 10-20 group (*erhua* 1.19%; *rusheng* 0.1%). Because almost all speakers in this group were born in the community, it can be speculated that the community language was their first language. As such, residents who have been living in the community for 20-40 years probably have adopted their use of *erhua* and *rusheng* from their peers. Also, that the community language has changed over time has likely affected the use of *erhua* and *rusheng* in the speech of community members. The community language for residents who have been living in the speech community for more than 40 years is not the same as that of those who have been living there for 20-40 years. For example, over 40 years ago, there might have been little mixture between northern and southern speech within the Xianggang community, and because of that, southerners and northerners had the least similarity in their use of *erhua* and *rushing*. The community language changed 20-40 years ago and southerners and northerners started adopting *erhua* and *rusheng* from each other. As a result, the most similar use of *erhua* and *rusheng* is among people who have resided in the community for 20-40 years, regardless of their northern or southern characteristics.

As is shown in the analysis of the questionnaire, there are four speakers who do not identify themselves as Xianggang community member. As can be seen from Table 5.1, subject 4 is a middle-aged, northern-characterized speaker who used *rusheng* six times and mostly used the tone in different lexical items compared to other people in his group. Although subject 4 has

been living in the community for over 40 years, he still does not identify himself as a Xianggang community member, and as is shown in the data, his different self-identification probably affected his adoption of *rusheng.*

There is a linguistic changing tendency towards the use of Putonghua in the Xianggang community. The most obvious evidence exists in the young group, which has been confirmed in the literature (Chen, 2002; Chen, 1999), that is, they have a lower use of *erhua* and avoidance of *rusheng.* The author noticed that during the interview, the younger aged speakers would correct each other if any word has been pronounced in a "wrong" way. For example:

- No. 13: Is this one ("ta") pronounced as [lai]?
- No. 14: It is [ta], the third tone, marked in the book.

Even though this correction is not directly related to the use of *erhua* or *rusheng,* it suggests that there is a discipline of 'right or wrong' in the students' mind. They are trying to bring the 'right' answer for the interview. In contrast, the Xiang dialect is still alive in the area particularly in some informal settings and has its unique function. In one of the interviews, an interviewee (middle age) who is in charge of the bureau of the Xianggang culture told me that: "We Xianggang community members have an advantage in the use of Putonghua speech because of our migration history. As recorded, a lot of prizes in Xiangtan city's speech contests have been given to our community members because we speak better Putonghua than people from other areas in the city; the Xiang dialect is not elegant at all; it is rough and sounds uneducated." Then, when asked, "Do you speak the Xiang dialect?" he smiled and replied, "Yes, when I go to the city and sometimes in contexts such as the farmers' market or on the bus."

Therefore, judging from the results, there is a changing tendency towards the use of Putonghua because of the education and prestige attached to it. At the same time, the dialect is still preserved in the middle age group, as a reaction to social contact.

CHAPTER SIX: CONCLUSION

This project investigated the use of *erhua* and *rusheng* by workers and students in the Xianggang community in China. Dialect use within the Xianggang community is quite complex, because some of the community members are initially from the northern part of China, while some others are locally from the south. The north-south dialects contact situation builds up a unique speech community within Xiangtan. Collecting data from both younger and middle age group people, community members' language patterns and the effect of social factors on their language use have been explored.

The language variables that have been focused on are the use of *erhua* and *rusheng*. *Erhua* is a northern-colored speech pattern, and *rusheng* is a southern-colored speech pattern. Putonghua is quite selective regarding the use of *erhua*, but excludes the use of *rusheng*. The purpose of testing community members' use of these two variables was to find the changing patterns of their language use towards the local dialect or Putonghua. Furthermore, by exploring the influence of four social variables on the use of these two linguistic variables, those that have impact on language change within the Xianggang community have been investigated. As such, a changing tendency of the community language has been predicted.

This study found that linguistically, post-vocalic *erhua* is the overt marker of northern people. Social variables such as age and the division of northern/southern characteristics have affected the use of both *erhua* and *rusheng* within the Xianggang community, while the period of residence has an impact on the use of *erhua* only. As discussed, language planning in China has the most influence on the younger speakers due to Putonghua education, which has been

reflected in their zero percentage use of *rusheng.* For middle-age speakers, the social context of their workplace has affected their lexical use, i.e., focused use of *rusheng* in three lexical items that are related to the social context of the steel company.

6.1 IMPLICATIONS FOR DIALECT VARIATION

First, the study confirmed what Kerswill (1996) stated, that the social factors of age and age-related social contexts play a significant role in the spread of language change. In this study, young speakers and middle-aged speakers are living in different social contexts. For the younger speakers, the social context that they are currently in is the school system. Their zero use of *rusheng* is a result of the language policy of the government that promotes the standard and prestigious form of Chinese language. Middle-aged speakers obviously show more variation in their language use than the younger ones. Southern-characterized speakers eliminated their use of *rusheng*, but northern-characterized speakers adopted some lexical items with the same tone.

Second, the results of this study suggest that in addition to the effect of interlocutors (Kerswill, 1996), the way individual speakers identify themselves and the social evaluation of language use have affected the spread of language change. The language policy that calls for conformity towards the use of Putonghua has a great effect on community members' speech. Even though there is no speaker whose native language is Putonghua, young speakers in this community tend to favor the Putonghua-related linguistic patterns such as *erhua* over the unrelated ones such as *rusheng*. This is because *erhua* is a feature in the northern dialect (Putonghua's base dialect) and has been partially preserved in Putonghua, while *rusheng* is a feature in southern dialect (nothing to do with *Putonghua*) and has been excluded in modern Chinese.

6.2 LIMITATIONS

There are some limitations to this study in terms of the experimental design and the participants' demography. First, the questionnaire could have elicited more information from the participants if the provided answers were not categorically "yes" or "no." It would have been more helpful to give them a scale for the degree of their agreement, such as "1 2 3 4 5." Instead of asking, "Do you identify yourself as a Xianggang community member?" the question could be "How do you feel if someone identifies you as a Xianggang community member? Please circle the number from 1 to 5." This issue was encountered when organizing the responses given by some students. They answered the "identification" question with "yes," but five out of ten students stated that they themselves were not familiar with "Xianggang." Second, limited by the time to conduct interviews, "wordlists" were chosen as the fastest way to get relevant speech patterns. However, if the participants had been given some questions to answer rather than giving them "wordlists" to read, the participants would have shown more linguistic features than they did by reading the "wordlists." Even though interviews were conducted in groups to make it casual, the person who read the wordlist and noticed that it was recorded may have behaved with some degree of formality. Since there is an overt prestige in Putonghua because of the policy and education, dialect study should be conducted in a less formal social situation, such as in natural conversations. Third, to collect data in groups might have some negative effects on the test results when speakers are asked to read designed wordlists. One informant's pronunciation might affect another's performance. As such, it may cause some sort of consistency among speakers within the same group. Fourth, as mentioned in the analysis chapter, only one middle-aged male with southern characteristics was interviewed. Even though the three middle aged females had already shown significantly more use of *erhua* than any of the rest groups, the result would have

been more reliable if the number of speakers in each group were equal. Fifth, the grouping of the younger (14 – 25 years old) and middle-age speakers (38 – 59 years old) is board in this study and may not give us a precise picture of what is going on linguistically in the community at large. Because I do not have any informants ages 26 to 38 years, or any informants ages 12 to 13 years, I divided the sample almost equally between a younger age group (14-25) and a middle age group (38-59).

6.3 FUTURE RESEARCH

Based on the results of this study, there might be two related areas that are worth exploring in future studies. First, there are many other studies (e.g., Zhang, 2005; Pankhurst, 2012) that focused on the social evaluation of *erhua* used by northern people in China (rhoticization, rhotic lenition). They found that the northern people might treat the use of *erhua* as a stigmatized form in their speech pattern. However, in this study, people from the south of China did not treat *erhua* as a stigmatized form. Rather, they tended to imitate or acquire some *erhua* to make them sound more like "pure Putonghua speakers." Therefore, there is a potential to study this phenomenon further using a "matched guised test",[6] and asking people from the south of China to evaluate such language use. The response given by people from the north of China could serve as a control group. This might suggest that language evaluation is also region related.

Second, much discussion has been conducted in this paper about the differences between variation and acquisition. As concluded, to study the process of second dialect acquisition, more data is needed to compare, for example, the speech of the first generation of Xianggang

[6] Matched guised test is a test designed to gauge unexpressed language attitudes by asking subjects to rate recorded speakers on a scale according to traits like social class, intelligence, and friendliness; however, subjects are actually listening to the same speaker or speakers several times, using different accents or speaking different languages (Cited by Herk, 2012)

community members with the speech of newcomers from the north. When data are ready, the same social factors considered for dialect variation could be considered for dialect acquisition.

Third, a comparison study on the speech community's language use between informal and formal settings would be interesting. This study has suggested that a collection of natural speech of the speech community members would be valuable for dialect study in China, but the data that have been analyzed so far are collected through formal interview (read wordlists). If I had collected some natural speech from the community members, I could have more evidence to support or challenge what I have observed in my current research.

Fourth, when there is a bigger linguistic sample for each age group within the Xianggang community, a study based on narrower categorization of speakers (young speakers, preadolescents, adolescents, early adulthood, adults, middle-aged, and elders) would be valuable, especially if it focuses on the social factor of age. We could see a clearer pattern of linguistic change in that situation.

REFERENCES

Brace, N., Kemp, R., & Snelgar, R. (2000). *SPSS for psychologists: a guide to data analysis using SPSS for windows.* London: Lawrence Erlbaum Associates, INC.

Britain, D. (2002). Space and Spatial Diffusion. *The Handbook of Language Variation and Change, ed.* by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 349-372. Malden, MA: Blackwell

Chao, Y. (1968). *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chambers, J. K. (2002). Studying language variation: An informal Epistemology. *The Handbook of Language Variation and Change, ed.* by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 349-372. Malden, MA: Blackwell.

Chamber, J. K. (1992). Dialect acquisition. *Language,* 68(4), 673-705.

Chen, C. (1983). A fifth tone in the mandarin spoken in Singapore. *Journal of Chinese Linguistics, 11(1)*, 92-121.

Chen, P. (1999). *Modern Chinese: History and sociolinguistics.* Cambridge & New York: Cambridge University Press.

Chen, Zhangtai. 2002. Lve lun wo guo xin shi qi de yu yan bian yi[On the language variation in the new period].Beijing: *Language Teaching and Linguistic Studies*, 6, 27-36.

Eberhard, W. (1965). Chinese regional stereotypes. *Asian Survey*, 5(12), 596-608.

Giles, H. (1973). A model and some data. *Anthropological linguistics*, 15(2), 87-105.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*,2(2), 177-192.

Gumperz, J.J. (1986). The speech community. In D.L. Sills (Ed.), *International encyclopedia of the social sciences,* 381-386. New York: Macmillan.

Habib, R. (2010). Rural migration and language variation in Hims, Syria. *SKY Journal of Linguistics,* 23, 61-99.

Hazen, K. (2002). Identity and language variation in a rural community. *Language* 78(2), 240-257

He, L. (2010). Acoustic study of the tones of Xiangtan dialect. Unpublished manuscript, Ludong University.

Herk, G. (2012). *What is sociolinguistics*. Chichester, West Sussex, UK: Wiley-Blackwell.

Hiramoto, M. (2010). Dialect contact and change of the northern Japanese plantation immigrants in *Hawai'i. Journal of Pidgin and Creole Languages*, 25(2), 229–262.

Kerswill, P. (1993). Rural dialect speakers in an urban speech community: the role of dialect contact in defining a sociolinguistic concept. *Internal Journal of Applied Linguistics*, 3(1), 33-45.

Kerswill, P. (1996). Children, adolescent, and language change. *Language Variation and Change* 8, 177-202.

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

Labov, W. (2007). Transmission and diffusion. *Language*, 83(2), 344-387.

Lee, S., & Li, D. C. S. (2013). Multilingualism in greater China and the Chinese language diaspora. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism (2 ed.)*. Blackwell Publishing Ltd.

Liao. S. (2010). Identity, ideology, and language variation: A sociolinguistic study of mandarin in central Taiwan. (Doctoral dissertation).

Liu, X. (2009). Under the influence of the collective memory of the "unit cultural circle" phenomenon and its significance in public. (Master's thesis, Xiangtan University, Xiangtan, Xiangtan, China).

Milroy, J. & L. Milroy. 1998. Mechanisms of change in urban dialects: The role of class, social network and gender. In P. Trudgill & J. Cheshire (Eds.). *The sociolinguistics reader, vol. I: Multilingualism and variation*, 179 -195. London: Arnold. (Originally published 1993 in International Journal of Applied Linguistics, 3(1).)

Milroy, Lesley (1980). *Language and Social Networks*. Baltimore, MD: University Park Press.

Pankhurst, M. M. (2012). Rhotic lenition as a marker of a dominant character type in northern mandarin Chinese. Working Papers of the Linguistics Circle of the University of Victoria, 22(1), 27-40.

Patrick, P. (2002). The Speech Community. *The Handbook of Language Variation and Change, ed.* by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 349-372. Malden, MA: Blackwell

Qian, X. (1985). shi lun pu tong hua de "qingsheng" zi he "erhua" ci [a study about zero-tone character and "erhua" words in Mandarin]. *Shenzhen University Journal*, 71–78.

Rys, Kathy. 2007. Dialect as second language: Linguistic and non-linguistic factors in secondary dialect acquisition by children and adolescents. PhD dissertation. Ghent University.

Siegel, J. (2010). *Second dialect acquisition*. United States of America: Cambridge University Press.

Stanford, James N. (2008). Child dialect acquisition: New perspectives on parent/peer influence. *Journal of Sociolinguistics* 12(5), 567–596.

Stanford, J. N., & Evans, J. P. (2012). The influence of mandarin Chinese on minority languages in rural southwest china: a sociolinguistic study of tones in contact. *International Journal of the Sociology of Language*, 215, 79-100.

Treffers-Daller, J., & Mougeon, R. (2005). The role of transfer in language variation and change: Evidence from contact varieties of French. *Bilingualism: Language and Cognition*, 8(2), 93-98

Trudgill, P. (1986). *Dialect in contact*. New York, NY: Basil Blackwell Inc.

Wang, Y. (2007). Han yu "erhua" yan yan jiu [A study on "rhotic" in Chinese]. (Doctoral dissertation, Jinan University, Guangzhou, Guangdong, Guangzhou, China).

Wang, L., & Lin, T. (1995). *Yu yin yue jiao cheng[chinese phonetics*]. Taipei: Wu-Nan Book Inc.

Weinreich, U. (1953) *Languages in Contact*. New York: Linguistic Circle.

Yang, S. (2004). Xiangtan fang yan ru sheng yan jiu [study on entering tones of Xiangtan dialect]. (Master's thesis, Shantou University, Shantou, Guangdong, Shantou, China).

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society*, (34), 431-466.

Zhou, M., & Sun, H. (2004). Language policy in the people's republic of china theory and practice since 1949. *Norwell: Kluwer Academic Publishers*.

Zhou, D. (1980). Yi shu yuyan fa sheng ji chu [the phonation of language]. Beijing: *China Social Sciences Press*.

APPENDICES

APPENDIX 1：WORKERS' QUESTIONNAIRE

Name	Age	Gender

1. What is your job?
   a. Current Workers in Xianggang
   b. Retired workers in Xianggang
   c. Other：_____

2. If you selected "other" in the former question, are you a relative of a Xianggang's worker?
   a. Yes b. No

3. Do you identify yourself as a "Xianggang Person"?
   a. Yes b. No

3. How many years have you been working in Xianggang?
   a. Less than 5 years
   b. 5 – 10 years
   c. 10 – 20 years
   d. More than 20 years, specifically: _____years
   e. N/A

4.Where are you originally from?
   a. Hebei Province
   b. Hunan Province
   c. Other:_____

5. How many years have you lived in Xianggang?
   a. More than 40 years
   b. 20 – 40 years
   c. 10 – 20 years
   d. Less than 10 years, specifically: _____years

5. Are your children working now?
   a. Yes b. Haven't yet

If he/she is working now, are they also working in Xianggang?

a. Yes b. No

6. If your child is still in school, is the school that she/he is attending located in the Xianggang community?

a. Yes b. No c. If no, where：_____

7. If your child is still studying at school, what are your expectations for him/her? Do you want your children to work in Xianggang, or outside it?

_____

8. Do you think living in your community is convenient enough for you? or Do you still feel the need to get stuff from the city?

_____

APPENDIX 2：STUDENTS' QUESTIONNAIRE

Name	Age	Gender

1. Are you a student?
   a. Yes b. No

2. Do you think that you are very familiar with "Xianggang"?
   a. Familiar b. Not Familiar

3. Do you identify yourself as a "Xianggang person"?
   a. Yes b. No

3. Is anyone of your parents working in "Xianggang" ? Identify all that apply.
   a. Mother
   b. Father
   c. Both of them
   d. Neither of them
   e. None of them

4. Do you have any idea about their jobs? Do you know what they often do at work?
   a. Yes, I know.
   b. No, I don't know.

5. Do you know where is your family originally from?
   a. Hebei Province
   b. Hunan Province
   c. Others：_____

6. Is the school you are now attending is located in the Xianggang community?
   a. Yes b. No c. Other

7. Are many of your classmates' families working in Xianggang?
   a. A lot
   b. Not much

8. How about your friends? Do they come from Xianggang?
   a. Yes b. No c. some of them

7. What are your hobbies? Where do you often go to have fun?

_____

8. Do you want to work in Xianggang when you finish school? Alternatively, where do you want to go, and what kind of job you would like to do?

_____

APPENDIX 3: THE LIST OF *ERHUA* WORDS

Table 1: The list of post-vocalic *erhua*

	Word	IPA Transcript	Glossary
1	雨	'y'	‘rain’
2	小雨	'ɕiɑu y'	‘light rain’
3	毛毛雨	'mɑu mɑu y'	‘drizzle’
4	雪	'ɕyɛ'	‘snow’
5	羊羔	'iɑŋ kɑu'	‘lamb’
6	篮子	'lan tsɻ'	‘basket’
7	小雪	'ɕiɑu ɕyɛ'	‘light snow’
8	猫	'mɑu'	‘cat’
9	被子	'pei tsɻ'	‘quilt’
10	岩石	'iɛn ʂɻ'	‘stone’
11	小猫	'ɕiɑu mɑu'	‘kitty’
12	小石块	'ɕiɑu ʂɻ kʰuai'	‘small stone’
13	鸟	'niɑu'	‘bird’
14 15	较小的 石块	'tɕiɑu ɕiɑu tɤ ʂɻ kʰuai'	‘the smaller stone’
16	手帕	'ʂou pʰa'	‘handkerchief’
17	斧子	'fu tsɻ'	‘axe’
18	猴子	'xou tsɻ'	‘monkey’
19	小猴子	'ɕiɑu xou tsɻ'	‘baby monkey’
20	米	'mi'	‘rice’
21	碎米	'suei mi'	‘rice powder’
22 23 24	颗粒 小的 糯米	kʰɤ li ɕiɑu tɤ' nuo mi	‘fine sticky rice’
25	棍子	'kun tsɻ'	‘stick’
26	小米	'ɕiɑu mi'	‘millet’
27	玉米	'y mi'	‘corn’
28	蚊子	'un tsɻ'	‘mosquito’
29	芋头	'y tʰou'	‘taro’
30	鱼	'y'	‘fish’
31	鱼苗	'y miɑu'	‘fry’
32	虾	'ɕia'	‘shrimp’
33 34	手扶 拖拉机	'ʂou fu tʰuo la tɕI'	walking tractor

35	辫子	'piɛn tsɿ'	‘pigtail’
36	大芥菜	'ta tɕiɛ tsʰai'	‘one kind of Brassica juncea’
37 38	极小的 虾	'tɕiɑu ɕiɑu tɤ ɕia'	‘tiny shrimp’
39	短辫子	'tuan pan tsɿ'	‘short pigtail’
40	小芥菜	'ɕiɑu tɕiɛ tsʰai'	‘the other kind of Brassica juncea’
41	小辣椒	'ɕiɑu la tɕiɑu'	‘little pepper’
42	鼻子	‘pi tsɿ’	‘nose’
43	石磨	'ʂɻ mo'	‘stone mill’
44 45	手推的 小石磨	‘ʂou tʰuei tɤ ɕiɑu ʂɻ mo'	‘manpowered small stone mill’
46	舌鳎	'ʂɤ tʰa'	‘Cynoglossus robustus’
47	口	'kʰou'	‘mouth’
48	马	'ma'	‘horse’
49 50	幼小的 舌鳎	'iou ɕiɑu tɤ ʂɤ tʰa'	“baby Cynoglossus robustus ”
51 52	短小的 梯子	'tuan ɕiɑu tɤ' tʰi tsɿ'	‘short stairs’
53 54 55 56	比喻 人的 小巧的 嘴巴	'pi y rən tɤ' ɕiɑu tɕʰiɑu tɤ tsuei pa'	‘use metaphor to describe human’s tiny mouth’
57	小马	'ɕiɑu ma'	‘pony’
58 59	个人儿大的 小草鱼	'kɤ rən ər ta tɤ ɕiɑu tsʰɑu y'	‘Grass carp that is very big
60	楼梯	'lou tʰI'	‘stairs’
61	齿	'tʂʰɻ'	‘teeth’
62	牛	'niou'	‘cattle’
63	草鱼	'tsʰɑu y'	‘Grass carp’
64	小牛	'ɕiɑu niou'	‘caff’
65	猪	'tʂu'	‘pig’
66	椅子	'i tsɿ'	‘chair’
67	拳头	'tɕʰyɛn tʰou'	‘fist’
68	小猪	'ɕiɑu tʂu'	‘piggy’

Table 2: The list of post-nasal *erhua*

	Word	IPA transcript	Glossary
1	鸡蛋	‘tɕi tan’	‘egg’
2	羊	'iɑŋ'	‘sheep’
3	毛巾	'mɑu tɕin'	‘towel’
4	砖	'tʂuan'	‘brick’
5	碎砖	'suei tʂuan'	‘rubble’
6	蚕	'tsʰan'	‘silkworm’
7	铃	'liŋ'	‘bell’
8	铃铛	'liŋ tɑŋ'	‘tinkle bell’
9	臭虫	'tʂʰou tʂʰuŋ'	‘bugs’
10	脸盆	'liɛn pʰən'	‘washbasin’
11	沙中小虫	'ʂa tʂuŋ ɕiɑu tʂʰuŋ'	‘small bugs living by sand’
12	箩筐	'luo kʰuɑŋ'	‘bamboo crate’
13	小箩筐	'ɕiɑu luo kʰuɑŋ'	‘small bamboo crate’
14 15	线的 一端	'ɕiɛn tɤ i tuan'	‘one side of the thread’
16	生姜	'ʂəŋ tɕiɑŋ'	‘fresh ginger’
17	稻草人	'tɑu tsʰɑu rən'	‘scarecrow’
18	嫩姜	'nən tɕiɑŋ'	‘baby ginger’
19 20 21 22 23	供 包装 商品 用的 木箱	'kuŋ pɑu tʂuɑŋ ʂɑŋ pʰin yŋ tɤ mu ɕiɑŋ'	‘wool cases for packaging goods’
24 25 26 27 28 29	供 商品 包装 用的 瓦楞 纸箱	'kuŋ ʂɑŋ pʰin pɑu tʂuɑŋ yŋ tɤ uɑ ləŋ tʂʅ ɕiɑŋ’	‘corrugated cases for packaging’
30	银鲳	'in tʂʰɑŋ'	‘pomfret’
31	河鳗	'xɤ man'	‘eel’
32	高低床	'kɑu ti tʂʰuɑŋ'	‘double-leveled bed’
33	木板	'mu pan'	‘board’
34	鼻尖	'pi tɕiɛn'	‘apex nasi (the tip of nose)’
35 36	狭长的 木板	'ɕia tʂʰɑŋ tɤ' mu pan'	‘long and narrrow board’

37	厢房	'ɕiɑŋ fɑŋ'	‘guest room’
38	房	'fɑŋ'	‘room’
39	厨房	'tʂʰu fɑŋ'	‘kitchen’

Table 3: The list of phrases that consist of post-vocalic *erhua* and post-nasal *erhua*

	Word	IPA transcript	Glossary
1 2 3	体型 较小的 鸟	'tʰi ɕiŋ tɕiɑu ɕiɑu tɤ niɑu'	‘Smaller shaped birds’
4 5 6 7 8 9	小号的 脸盆 多用来 贮存 食物 或汤汁	'ɕiɑu xɑu tɤ liɛn pʰən tuo yŋ lai tʂu tsʰun ʂʅ u xuo tʰɑŋ tʂʅ'	‘small sized washbasin for storing foods or soup’
2 3	低山 小丘	'ti ʂan' 'ɕiɑu tɕʰiou'	‘hill’
10 11 12	线头子 很短的 一段线	'ɕiɛn tʰou tsɿ, xən tuan tɤ i tuan ɕiɛn'	‘very short length of thread’
13 14	厢房旁的 小屋子	'ɕiɑŋ fɑŋ pʰɑŋ tɤ ɕiɑu u tsɿ'	‘a small room besides the guest room’
15 16 17 18	刮 胡子 用的 薄刀片	‘kuA xu tsɿ yŋ tɤ pɑu tɑu pʰiɛn'	‘razor blade’
19 20 21 22	母鸡 腹内的 未成形的 小蛋	'mu tɕi fu nei tɤ uei tʂʰəŋ ɕiŋ tɤ ɕiɑu tan'	‘an egg growing in a hen’s stomach’
23 24 25 26	老式的 摆 在桌子上的 钟	'lɑu ʂʅ tɤ pai tsai tʂuo tsɿ ʂɑŋ tɤ' tʂuŋ'	‘an old-styled table clock’
27 28 29	女人 用的 小皮包	'ny rən yŋ tɤ' 'ɕiɑu pʰi pɑu'	‘Women’s small handbag’
30 31	尼龙 编织袋	'ni luŋ' 'piɛn tʂʅ tai'	‘Nylon bag’
32	盛	'ʂəŋ	‘basket for holding

33 34	食物的 篮子	ʂʅ u tɤ' 'lan tsʅ'	foods’
35 36 37 38	织 网袋 用的 小梭子	'tʂʅ uɑŋ tai yŋ tɤ' 'ɕiɑu suo tsʅ'	‘the shuttle that used for making mesh bag’
39 40 41 42	专供 幼儿 盖的 小被子	‘tʂuan kuŋ iou ər kai tɤ ɕiɑu pei tsʅ'	‘the small quilt that used by babies’
43 44 45 46 47	用于 装 各种 常用工具的 小箱子	‘yuŋ y tʂuɑŋ kɤ tʂuŋ tʂʰɑŋ yŋ kuŋ tɕy tɤ ɕiɑu ɕiɑŋ tsʅ'	‘the small case that used to hold common tools’
48 49 50 51 52 53 54 55 56	一种 供 瓦工 斩断 砖头, 修削 砖瓦 用的 小斧子	‘i tʂuŋ kuŋ ua kuŋ tʂan tuan tʂuan tʰou', ɕiou ɕyɛ tʂuan ua yŋ' tɤ ɕiɑu fu tsʅ'	‘A small axe that tiler uses for cutting down bricks and mending tile’
57 58 59 60	锤头 为 铁质的 锤子	'tʂʰuei tʰou uei tʰiɛ tʂʅ tɤ tʂʰuei tsʅ'	‘A hammer that is made from iron’
61 62 63 64	鞋匠 专用的 一种 铁锤子	'ɕiɛ tɕiɑŋ tʂuan yŋ tɤ i tʂuŋ tʰiɛ tʂʰuei tsʅ'	‘A small axe that tiler uses for cutting down bricks and mending tile’
65 66 67 68	装 东西的 口袋 或包裹	'tʂuɑŋ tuŋ ɕI tɤ kʰou tai xuo bao guo'	‘pockets or package for carrying or storing stuff’
69 70 71	淘米 用的 石臼	'tɑu mi yŋ tɤ' ʂʅ tɕiou'	‘rice grinder’
72 73	小的 瓦楞	'ɕiɑu tɤ uɑ ləŋ	‘small corrugated cases’

74	纸箱	tʂʅ ɕiɑŋ'	
75 76 77 78 79 80 81	一般的 摆 在桌子上的 钟 较 老式座钟 小些	'i pan tɤ pai ' tsai' tʂuo tsʅ ʂɑŋ tɤ' tʂuŋ, tɕiɑu lɑu ʂʅ tɤ tsuo tʂuŋ ɕiɑu ɕiɛ'	‘an ordinary table clock, which is smaller than old-styled ones’
82 83 84 85	用 毛竹 制成的 扁担	'yŋ mɑu tʂu tʂʅ tʂʰəŋ' tɤ' piɛn tan'	‘the carrier that is made from bamboo’
86 87 88 89 90	用 木板 临时 搭成的 床铺	'yŋ mu pan lin ʂʅ ta tʂʰəŋ tɤ tʂʰuɑŋ pʰu'	‘bed that temporarily made from board’
91 92 93	儿童的 胖乎乎的 胳膊	'ər tʰuŋ tɤ pʰɑŋ xu xu tɤ kɤ puo'	‘child's chubby arms’

Table 4: The list of words that have neither post-vocalic nor post-nasal *erhua*

	word	IPA transcript	Glossary
1	泪花儿	'lei xua eɻ’	‘teardrop’
2	鳗苗儿	man miɑu əɻ’	‘baby eel’
3 4 5 6	比喻 牙龈 细密 洁白	'pi y ia tʂʰʅ ɕi mi ' tɕiɛ pai'	‘use metaphor to describe the whiteness of teeth’

APPENDIX 4: THE LIST OF *RUSHENG* WORDS

	Word	IPA transcript	Tone in Mandarin	Glosary
1	答	‘tʌ’	2	‘answer’
2	喝	‘xɤ’	1	‘drink’
3	塔	‘tʰʌ’	3	‘tower’
4	插	‘tʂʰʌ’	1	‘insert’
5	恰	‘tɕʰiʌ’	4	‘just’
6	立	‘li’	4	‘stand’
7	集	‘tɕI’	2	‘collect’
8	湿	‘ʂʅ’	1	‘wet’
9	习	‘ɕI’	2	‘habit’
10	及	‘tɕI’	2	‘and’
11	獭	‘tʰʌ’	3	‘otter’
12	达	‘tʌ’	2	‘reach’
13	擦	‘tsʰʌ’	1	‘wipe’
14	八	‘pʌ’	1	‘eight’
15	没	‘mei’	2	‘no’
16	毕	‘pi’	4	‘finish’
17	蜜	‘mi’	4	‘honey’
18	质	‘tʂʅ’	4	‘quality’
19	失	‘ʂʅ’	1	‘lost’
20	物	‘u’	4	‘object’
21	泊	‘pʰuo’	1	‘lake’
22	莫	‘muo’	4	‘don’t’
23	咋	‘tsʌ’	3	‘how and why’
24	恶	‘ɤ’	4	‘evil’
25	霍	‘xuo’	4	‘a sername’
26	觉	‘tɕyɛ’	2	‘think’
27	角	‘tɕiɑu’	3	‘angle’
28	确	‘tɕʰyɛ’	4	‘reliable’
29	岳	‘yɛ’	4	‘high mountain’
30	乐	‘lɤ’	4	‘happy’
31	北	‘pei’	3	‘north’
32	肋	‘lɤ’	4	‘rib’
33	识	‘ʂʅ’	2	‘know’
34	式	‘ʂʅ’	4	‘style’
35	窄	‘tʂai’	3	‘narrow’
36	逆	‘ni’	4	‘opposite’
37	适	‘ʂʅ’	4	‘fit’

38	踢	‘tʰI’	1	‘kick’
39	戚	‘tɕʰI’	1	‘relative’
40	族	‘tsu’	2	‘clan’
41	督	‘tu’	1	‘superintend’
42	毒	‘tu’	2	‘poison’
43	仆	‘pʰu’	2	‘servant’

(Yang, 2004)

APPENDIX 5: OUTPUT OF T-TESTS

1. *erhua*

1.1 All Female vs. All Male

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Rhotacization	female	20	4.4500	4.43046	.99068
	male	19	8.1579	9.55838	2.19284

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Rhotacization	Equal variances assumed	5.440	.025	-1.567	37	.126	-3.70789	2.36561	-8.50108	1.08529
	Equal variances not assumed			-1.541	25.107	.136	-3.70789	2.40624	-8.66258	1.24679

1.2 All Age – young vs. old

Group Statistics

	Age	N	Mean	Std. Deviation	Std. Error Mean
Rhotacization	younger group	19	2.8947	2.44710	.56140
	older group	20	9.4500	9.24790	2.06789

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Rhotacization	Equal variances assumed	13.958	.001	-2.990	37	.005	-6.55526	2.19233	-10.99735	-2.11318
	Equal variances not assumed			-3.059	21.779	.006	-6.55526	2.14275	-11.00166	-2.10887

1.3 North – age group

Group Statistics

	Age	N	Mean	Std. Deviation	Std. Error Mean
Rhotacization	younger	12	3.8333	2.44330	.70532
	older	15	11.0667	9.89565	2.55504

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Rhotacization	Equal variances assumed	11.296	.002	-2.464	25	.021	-7.23333	2.93592	-13.27996	-1.18670
	Equal variances not assumed			-2.729	16.096	.015	-7.23333	2.65061	-12.84965	-1.61702

1.4 North – gender

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Rhotacization	female	13	5.0769	4.38675	1.21667
	male	14	10.4286	10.23354	2.73503

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Rhotacization	Equal variances assumed	6.618	.016	-1.741	25	.094	-5.35165	3.07394	-11.68255	.97925
	Equal variances not assumed			-1.788	17.895	.091	-5.35165	2.99343	-11.64327	.93997

2 *rusheng*

2.1 age group

Group Statistics

	Grouped age	N	Mean	Std. Deviation	Std. Error Mean
Tone	1.00	18	.0000	.00000	.00000
	2.00	20	2.1000	1.77408	.39670

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Tone	Equal variances assumed	36.897	.000	-5.015	36	.000	-2.10000	.41874	-2.94924	-1.25076
	Equal variances not assumed			-5.294	19.000	.000	-2.10000	.39670	-2.93030	-1.26970

2.2 gender

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Tone	Female	19	1.0000	1.20185	.27572
	Male	19	1.2105	2.04339	.46879

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Tone	Equal variances assumed	5.984	.019	-.387	36	.701	-.21053	.54386	-1.31352	.89247
	Equal variances not assumed			-.387	29.123	.701	-.21053	.54386	-1.32264	.90159

2.3 North and South division

Group Statistics

	geographical division	N	Mean	Std. Deviation	Std. Error Mean
rusheng	north	27	1.4074	1.84514	.35510
	south	12	.3333	.65134	.18803

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rusheng	Equal variances assumed	13.390	.001	1.951	37	.059	1.07407	.55060	-.04154	2.18969
	Equal variances not assumed			2.673	35.945	.011	1.07407	.40181	.25913	1.88902

APPENDIX 6: OUTPUT OF REGRESSION TESTS

1. *erhua*

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	female and male	female	20	51.3%
		male	19	48.7%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 201

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	rhotic after vowel	39	.00	25.00	4.8718	5.12568
	rhotic after nasal	39	.00	11.00	1.3846	2.78733

Goodness of Fit[a]

	Value	df	Value/df
Deviance	37.340	31	1.205
Scaled Deviance	36.646	31	
Pearson Chi-Square	31.587	31	1.019
Scaled Pearson Chi-Square	31.000	31	
Log Likelihood[b,c]	-76.523		
Adjusted Log Likelihood[d]	-75.101		
Akaike's Information Criterion (AIC)	169.046		
Finite Sample Corrected AIC (AICC)	169.065		
Bayesian Information Criterion (BIC)	224.781		
Consistent AIC (CAIC)	232.781		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, rhotic2, north_south, age_group, gender, period

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
239.103	7	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, rhotic2, north_south, age_group, gender, period

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	7356.559	1	.000	7356.559	1	31	.000
rhotic1	26.019	1	.000	26.019	1	31	.000
rhotic2	.262	1	.609	.262	1	31	.613
north_south	5.121	1	.024	5.121	1	31	.031
age_group	4.557	1	.033	4.557	1	31	.041
gender	.116	1	.733	.116	1	31	.735
period	11.586	2	.003	5.793	2	31	.007

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, rhotic2, north_south, age_group, gender, period

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.736	.2979	-5.320	-4.152	252.683	1	.000	.009	.005	.016
rhotic1	.114	.0225	.070	.158	25.874	1	.000	1.121	1.073	1.172
rhotic2	.020	.0397	-.057	.098	.261	1	.609	1.020	.944	1.103
[north_south=1.00]	.457	.2093	.047	.868	4.775	1	.029	1.580	1.048	2.381
[north_south=2.00]	0[a]	.	.	.	.	.	.	1	.	.
[age_group=1.00]	-.533	.2480	-1.019	-.046	4.611	1	.032	.587	.361	.955
[age_group=2.00]	0[a]	.	.	.	.	.	.	1	.	.
[gender=1.00]	-.063	.1837	-.423	.297	.116	1	.733	.939	.655	1.346
[gender=2.00]	0[a]	.	.	.	.	.	.	1	.	.
[period=1.00]	.005	.3727	-.726	.735	.000	1	.990	1.005	.484	2.086
[period=2.00]	.604	.2345	.144	1.064	6.634	1	.010	1.830	1.155	2.897
[period=3.00]	0[a]	.	.	.	.	.	.	1	.	.
(Scale)	1.019[b]									

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, rhotic2, north_south, age_group, gender, period

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

1.1 Affected factors - output

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 201

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	rhotic after vowel	39	.00	25.00	4.8718	5.12568

Goodness of Fit[a]

	Value	df	Value/df
Deviance	37.947	33	1.150
Scaled Deviance	38.911	33	
Pearson Chi-Square	32.182	33	.975
Scaled Pearson Chi-Square	33.000	33	
Log Likelihood[b,c]	-76.827		
Adjusted Log Likelihood[d]	-78.779		
Akaike's Information Criterion (AIC)	165.654		
Finite Sample Corrected AIC (AICC)	165.665		
Bayesian Information Criterion (BIC)	207.455		
Consistent AIC (CAIC)	213.455		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, age_group, period, north_south

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
249.197	5	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, age_group, period, north_south

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	7710.759	1	.000	7710.759	1	33	.000
rhotic1	129.933	1	.000	129.933	1	33	.000
age_group	7.281	1	.007	7.281	1	33	.011
period	17.907	2	.000	8.953	2	33	.001
north_south	5.910	1	.015	5.910	1	33	.021

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, age_group, period, north_south

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.843	.2557	-5.344	-4.341	358.564	1	.000	.008	.005	.013
rhotic1	.127	.0113	.105	.149	126.298	1	.000	1.135	1.110	1.160
[age_group=1.00]	-.535	.1993	-.926	-.144	7.207	1	.007	.586	.396	.866
[age_group=2.00]	0[a]	.	.	.	.	.	.	1	.	.
[period=1.00]	.021	.3254	-.617	.659	.004	1	.949	1.021	.540	1.932
[period=2.00]	.654	.1851	.291	1.017	12.490	1	.000	1.923	1.338	2.765
[period=3.00]	0[a]	.	.	.	.	.	.	1	.	.
[north_south=1.00]	.477	.2035	.078	.875	5.488	1	.019	1.611	1.081	2.400
[north_south=2.00]	0[a]	.	.	.	.	.	.	1	.	.
(Scale)	.975[b]									

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), rhotic1, age_group, period, north_south

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

2. *rusheng*

2.1 social factors - output

Model Information

Events Variable	rusheng
Number of Trials	43
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rusheng	Events	42	2.5%
		Non-Events	1635	97.5%
		Total	1677	100.0%
Factor	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	female and male	female	20	51.3%
		male	19	48.7%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 43

Goodness of Fit[a]

	Value	df	Value/df
Deviance	23.684	33	.718
Scaled Deviance	42.589	33	
Pearson Chi-Square	18.352	33	.556
Scaled Pearson Chi-Square	33.000	33	
Log Likelihood[b,c]	-33.021		
Adjusted Log Likelihood[d]	-59.379		
Akaike's Information Criterion (AIC)	78.042		
Finite Sample Corrected AIC (AICC)	78.093		
Bayesian Information Criterion (BIC)	110.591		
Consistent AIC (CAIC)	116.591		

Events: rusheng

Trials: 43

Model: (Intercept), age_group, north_south, gender, period

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
121.796	5	.000

Events: rusheng

Trials: 43

Model: (Intercept), age_group, north_south, gender, period

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	3083.859	1	.000	3083.859	1	33	.000
age_group	45.536	1	.000	45.536	1	33	.000
north_south	8.032	1	.005	8.032	1	33	.008
gender	2.365	1	.124	2.365	1	33	.134
period	3.093	2	.213	1.547	2	33	.228

Events: rusheng

Trials: 43

Model: (Intercept), age_group, north_south, gender, period

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-3.489	.4294	-4.331	-2.648	66.041	1	.000	.031	.013	.071
[age_group=1.00]	-31.006[a]	.	.	.	.	.	.	3.422E-014	.000	.000
[age_group=2.00]	0[b]	.	.	.	.	.	.	1	.	.
[north_south=1.00]	1.030	.4078	.230	1.829	6.376	1	.012	2.800	1.259	6.227
[north_south=2.00]	0[b]	.	.	.	.	.	.	1	.	.
[gender=1.00]	-.376	.2455	-.857	.105	2.342	1	.126	.687	.424	1.111
[gender=2.00]	0[b]	.	.	.	.	.	.	1	.	.
[period=1.00]	-.902	.7817	-2.434	.630	1.332	1	.248	.406	.088	1.878
[period=2.00]	-.368	.2827	-.922	.186	1.698	1	.193	.692	.398	1.204
[period=3.00]	0[b]	.	.	.	.	.	.	1	.	.
(Scale)	.556[c]									

Events: rusheng

Trials: 43

Model: (Intercept), age_group, north_south, gender, period

a. Hessian matrix singularity is caused by this parameter. The parameter estimate at the last iteration is displayed.

b. Set to zero because this parameter is redundant.

c. Computed based on the Pearson chi-square.

3. Two-way regression test output
3.1 "*erhua* after vowel" & "period of residence"

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 201

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	rhotic after vowel	39	.00	25.00	4.8718	5.12568

Goodness of Fit[a]

	Value	df	Value/df
Deviance	41.056	35	1.173
Scaled Deviance	42.768	35	
Pearson Chi-Square	33.599	35	.960
Scaled Pearson Chi-Square	35.000	35	
Log Likelihood[b,c]	-78.382		
Adjusted Log Likelihood[d]	-81.649		
Akaike's Information Criterion (AIC)	164.763		
Finite Sample Corrected AIC (AICC)	164.768		
Bayesian Information Criterion (BIC)	192.631		
Consistent AIC (CAIC)	196.631		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * rhotic1

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
249.914	3	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * rhotic1

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	4957.350	1	.000	4957.350	1	35	.000
period * rhotic1	249.914	3	.000	83.305	3	35	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * rhotic1

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.753	.1399	-5.027	-4.479	1154.259	1	.000	.009	.007	.011
[period=1.00] * rhotic1	.201	.0482	.107	.295	17.424	1	.000	1.223	1.113	1.344
[period=2.00] * rhotic1	.250	.0243	.202	.297	105.526	1	.000	1.283	1.224	1.346
[period=3.00] * rhotic1	.143	.0088	.126	.160	263.264	1	.000	1.154	1.134	1.174
(Scale)	.960[a]									

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * rhotic1

a. Computed based on the Pearson chi-square.

3.2 “erhua after vowel” & “age group”

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 201

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	rhotic after vowel	39	.00	25.00	4.8718	5.12568

Goodness of Fit[a]

	Value	df	Value/df
Deviance	65.745	36	1.826
Scaled Deviance	39.694	36	
Pearson Chi-Square	59.627	36	1.656
Scaled Pearson Chi-Square	36.000	36	
Log Likelihood[b,c]	-90.726		
Adjusted Log Likelihood[d]	-54.776		
Akaike's Information Criterion (AIC)	187.452		
Finite Sample Corrected AIC (AICC)	187.455		
Bayesian Information Criterion (BIC)	208.353		
Consistent AIC (CAIC)	211.353		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), age_group * rhotic1

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
129.943	2	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), age_group * rhotic1

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	5353.978	1	.000	5353.978	1	36	.000
age_group * rhotic1	129.943	2	.000	64.971	2	36	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), age_group * rhotic1

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.395	.1586	-4.706	-4.084	767.463	1	.000	.012	.009	.017
[age_group=1.00] * rhotic1	.147	.0520	.045	.249	8.018	1	.005	1.159	1.046	1.283
[age_group=2.00] * rhotic1	.132	.0110	.110	.154	142.868	1	.000	1.141	1.117	1.166
(Scale)	1.656[a]									

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), age_group * rhotic1

a. Computed based on the Pearson chi-square.

3.3 “erhua after vowel” & “north-south division”

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	geographical division	north	27	69.2%
		south	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%
	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%

a. Number of trials = 201

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	rhotic after vowel	39	.00	25.00	4.8718	5.12568

Goodness of Fit[a]

	Value	df	Value/df
Deviance	64.419	36	1.789
Scaled Deviance	39.670	36	
Pearson Chi-Square	58.459	36	1.624
Scaled Pearson Chi-Square	36.000	36	
Log Likelihood[b,c]	-90.063		
Adjusted Log Likelihood[d]	-55.462		
Akaike's Information Criterion (AIC)	186.125		
Finite Sample Corrected AIC (AICC)	186.129		
Bayesian Information Criterion (BIC)	207.026		
Consistent AIC (CAIC)	210.026		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), north_south * rhotic1

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
133.355	2	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), north_south * rhotic1

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df1	df2	Sig.
(Intercept)	5461.819	1	.000	5461.819	1	36	.000
north_south * rhotic1	133.355	2	.000	66.677	2	36	.000

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), north_south * rhotic1

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.411	.1412	-4.687	-4.134	975.867	1	.000	.012	.009	.016
[north_south=1.00] * rhotic1	.132	.0105	.112	.153	160.017	1	.000	1.141	1.118	1.165
[north_south=2.00] * rhotic1	.174	.0449	.086	.262	15.082	1	.000	1.191	1.090	1.300
(Scale)	1.624[a]									

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), north_south * rhotic1

a. Computed based on the Pearson chi-square.

4. Output of factor interactions
4.1 “period of residence” & “age”

Model Information

Events Variable	rhotic=rhotic1+rhotic2
Number of Trials	201
Probability Distribution	Binomial
Link Function	Logit

Case Processing Summary

	N	Percent
Included	39	100.0%
Excluded	0	0.0%
Total	39	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable[a]	rhotic=rhotic1+rhotic2	Events	244	3.1%
		Non-Events	7595	96.9%
		Total	7839	100.0%
Factor	period of residence	10-20	10	25.6%
		20-40	17	43.6%
		>40	12	30.8%
		Total	39	100.0%
	younger and older	younger age group	19	48.7%
		older age group	20	51.3%
		Total	39	100.0%

a. Number of trials = 201

Goodness of Fit[a]

	Value	df	Value/df
Deviance	182.749	34	5.375
Scaled Deviance	31.244	34	
Pearson Chi-Square	198.870	34	5.849
Scaled Pearson Chi-Square	34.000	34	
Log Likelihood[b,c]	-149.228		
Adjusted Log Likelihood[d]	-25.513		
Akaike's Information Criterion (AIC)	308.456		
Finite Sample Corrected AIC (AICC)	308.464		
Bayesian Information Criterion (BIC)	343.290		
Consistent AIC (CAIC)	348.290		

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * age_group

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
16.792	4	.002

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * age_group

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III						
	Likelihood Ratio Chi-Square	df	Sig.	F	df 1	df 2	Sig.
(Intercept)	682.499	1	.000	682.499	1	34	.000
period * age_group	16.792	4	.002	4.198	4	34	.007

Events: rhotic=rhotic1+rhotic2

Trials: 201

Model: (Intercept), period * age_group

VITA

Lei He

E: norahe201231@gmail.com

EDUCATION

B.A., Chinese Linguistics — Jun. 2012
Ludong University, College of Chinese Language and Literature, Yantai, Shandong, China
THESIS: Xiang fang yan dui yingyu zhongjieyu zhuofuyin de yingxiang [The influence of Xiang Dialect on Voiced Consonants _m/n/ l/ŋ_in "Interlanguage" between English and Chinese]. Undergraduate Thesis. Ludong University, Yantai, Shandong Province, China

M.A., Linguistics — Aug. 2014
Syracuse University, College of Arts and Science, Syracuse, NY

CERTIFICATION

Certificate of Advanced Study TESOL & TLOTE — Dec 2013
Syracuse University, College of Arts and Sciences, Syracuse, NY

HONOR AND AWARD

Chancellor's Awards for Public Engagement and Scholarship (2014)
Phi Beta Delta Honor Society (2013-present)
Third Prize in the National University English Competition in China (2009)

CONFERENCE PRESENTATION

Language variation in contact: dialect acquisition in the Xianggang community in China. The 22nd Annual Conference of the IACL & the 26th North American Conference on Chinese Linguistics, Maryland, U.S.A. May 2 -4, 2014

TEACHING EXPERIENCE

ESL Teacher Assistant — Jul. 2014 – Aug. 2014
West Side Learning Center, Syracuse, NY

ESL Teacher — Sept. 2013 – Nov. 2013
Missio Church, Syracuse, NY

MEMBERSHIPS

Linguistics Society of America (LSA)
International Association of Chinese Linguistics (IACL)