Language Policy, Dialect Writing and Linguistic Diversity[1]

Hongyuan Dong
George Washington University

This article studies the challenges encountered in the promotion of linguistic diversity in the context of Chinese dialects by examining the meta-data on Wikipedia sites written in major varieties of Chinese, with a focus on the type of writing systems used. The current language policy in China does not allow the explicit promotion of non-standard forms of Chinese in any official or national media. Therefore, online Wikipedia communities and sites of Chinese dialects have been flourishing. The choice of writing systems on these wiki sites to write Chinese dialects, including character-based and phonetic systems, is an important contributing factor to the success of these sites. I argue that the creation and practical use of an effective writing system conducive to literacy is a key issue in promoting dialects in the Chinese context.

1. Introduction

In this article, I study the effects of language policy and new collaborative technology on dialects from the perspective of the writing systems used by virtual linguistic communities. My focus here is on the different varieties of Chinese.[2]

In order to understand the current situation of linguistic diversity in terms of Chinese dialects and language policy making in China now, we need to take a historical perspective. The origins of modern language policy in China can be traced back to the year 1728 of the Qing Dynasty during the reign of Yongzheng Emperor, when an imperial edict was issued to order the establishments of local Mandarin schools in the Fujian and Guangdong areas (Dong 2014: 131; Wang 2014: 106). But this Mandarin Campaign was never met with any kind of enthusiasm from the local officials, and by 1775 during the reign of Qianlong Emperor the campaign was terminated (Deng 1994, Wu 2008, Dong 2015a). Consequently, the dialects in those areas were not affected at all.

Starting from the late 19th century until the founding of the People's Republic of China in 1949, another major wave of linguistic reform was implemented (Dong 2016,

---

[1] This paper benefitted from the discussions with the audience at NACCL-29, especially Miguel Cortiço dos Santos of The University of Tokyo.

[2] Here I will follow the traditional term "Chinese dialects" as a translation for "*Hànyǔ fāngyán*". Sometimes I refer to Chinese dialects as "varieties of Chinese". Many authors may prefer the term *topolects* or *Sinitic languages* (see e.g. Mair 1991).

Simons 2017). Although policies were made to promote Mandarin as the National Language, the implementations of these policies were not quite effective (Dong 2017). Thus, dialects were not affected much in this era either.

The new Chinese government after 1949 took a series of strong government measures to promote Putonghua as the national language (Zhou 2006, Zhou and Sun 2004). It is during this period up to the present time that usage of Chinese dialects has been gradually eroded. The situation resembles one of language loss. May (2006: 257–258) describes language decline and loss as occurring "most often in bilingual or multilingual contexts in which a majority language – that is, a language with greater political power, privilege, and social prestige – come to replace the range of functions of a minority language".

According to Baker and Jones (1998), and May (2006), there are three stages in the process of language shift. In terms of Chinese dialects, we may characterize these three stages as follows:

(1) Three Stages of Dialect Shift

° *Stage I*: increasing pressure on dialect speakers to speak the national language, particularly in formal language domains.
° *Stage II*: a decreasing number of fluent dialect speakers, especially among the younger generation.
° *Stage III*: replacement of dialects by the national language

Most varieties of Chinese, especially those in the south, are in the second stage of dialect shift as described above. This situation is directly related to the language laws in China. The most important one is the *Law of the People's Republic of China on the Standard Spoken and Written Chinese Language*, adopted at the 18th Meeting of Standing Committee of the Ninth National People's Congress on October 31, 2000. This law reflects various measures to promote Putonghua since 1949, and many of these measures are now officially codified to assume more power in its implementations. According to this law, "Putonghua and the standardized Chinese characters shall be used as the basic language in education and teaching in schools and other institutions of education, except where otherwise provided for in laws" (Article 10), "publications in Chinese shall be in conformity with the norms of the standard spoken and written Chinese language" (Article 11), and "Putonghua shall be used by the broadcasting and TV stations as the basic broadcasting language" (Article 12). Thus, dialects are restricted mostly to spoken forms in informal settings such as conversations at home.

Many scholars, dialect speakers, and dialect enthusiasts have started to try to preserve various dialects and, in some cases, oppose the promotion of Putonghua, e.g. resurgence of dialects in media (Liu 2013; Liu and Tao 2009, 2012), the campaign in Guangzhou to protect Cantonese from Putonghua erosion (Eng 2010), and etc. Much of

such efforts to preserve dialects started in online communities, and the organizers made good use of social media. This leads to my interest in studying the use of new technology to promote linguistic diversity in the Chinese context.

In this article, I use the metadata on Wikipedia sites written in Chinese dialects to study the promotion of dialects on the Internet (see also Dong 2015b). This can be considered a kind of "virtual linguistic landscape" (Ivkovic and Lotherington 2009). Linguistic landscape studies language displayed in public space (Shohamy and Gorter 2008: 1). To some extent, the web is the global public space where multilingualism can be displayed at its best with minimal restrictions imposed by national language policies. This article studies the linguistic landscape on Wikipedia in the Chinese context.

The remaining part of this article is structured as follows. In section 2, I summarize the metadata from Wikipedia, and point out issues highlighted by the numbers. In section 3, I give examples of all the Wikipedia sites written in Chinese dialects to illustrate how these websites are promoting their own version of dialects. In section 4, I connect the issues in section 2 with the writing systems used to write these dialects, and show that writing Chinese dialects is a key component to promoting linguistic diversity. In section 5, I make further remarks in conclusion.

2. Metadata on Wikipedia

The reason for using Wikipedia as a tool for promoting linguistic diversity in the Chinese context can be phrased as follows.

First, although there is content containing Chinese dialect elements on websites in China, such websites are nonetheless regulated by China's language laws, such as shown in the Introduction section. For example, the Chinese website *Bǎidù Bǎikē* 百度百科, which is the Chinese equivalent of Wikipedia, only allows content in the standard form of Chinese. There are no dialect versions of *Bǎidù Bǎikē*. Therefore, to fully promote dialects on the Internet, tools from outside China will be more effective because they are less subject to the laws within China.[3]

Second, Wikipedia has become the go-to site for information on any kind of topic. It is always listed on top of google search results. Therefore, by using Wikipedia, it can be guaranteed that the information will reach the widest audience and be used by the most readers, for purposes of gaining information, or simply learning a new language.

Third, the global reach of the Internet can make collaboration more easily achievable. The community of content contributors on Wikipedia consists of people from

---

[3] This is not to say that websites operated outside China are totally free from the influence of language policy in China. In effect, China's language policy has global reach in the linguistic standardizations adopted by international organizations and more recently in the establishments of language institutes around the globe. But indeed these websites are less restricted by language laws in China. For example, the Mandarin Wikipedia pages are often written with a mixture of simplified and traditional characters, likely due to the geographical regions of contributors. Such mixed use of Chinese characters is definitely not allowed by the linguistic laws in China.

different areas of expertise, not just linguists. Therefore, to my knowledge there is no other online tool or community that can compare to Wikipedia in its size and its power to pool resources globally to create content in a dialect.

Another important aspect about Wikipedia is that the content, including multi-media content, such as recordings and videos, creates a library, or a body of literature, of some sort in a language or a dialect. The existence of written documentation and other types of texts is the basis for the preservation and promotion of a language or a dialect.

Additionally, the official use of dialects is limited in China, but to create content on Wikipedia gives users and readers the practical opportunity to use the dialect. As shown in (1), one of the stages of language shift is the decreased use of dialects, and in this sense, to actually use dialects to do something is an important step towards preserving such dialects in the sense of increasing the use of such dialects.

Therefore, Wikipedia serves as the best model, so far, for bringing people in an online linguistic community to create a presence, or rather the virtual linguistic landscape, in order to preserve and promote linguistic diversity. Thus, studying these Wikipedia sites can tell us a great deal about how such efforts are faring and what challenges they encounter, so that we may better understand the promotion of linguistic diversity in terms of Chinese dialects. On a related note, the multi-language list for the same topic on Wikipedia can help us compare different languages or dialects easily. This is another advantage of using such data to study Chinese dialects on the web systematically.

Before discussing the meta-wiki data, let me introduce the major varieties of Chinese. According to the traditional classification of Chinese dialects, e.g. Yuan et al. (1960), there are seven major dialects of Chinese: Mandarin, Wu, Xiang, Gan, Min, Hakka, and Cantonese[4]. But the internal differences in each of these groups are still quite considerable, especially in the Min dialect, within which mutual intelligibility is the lowest of these seven groups. According to the *Language Atlas of China* (Wurm et al. 1987), the Min dialect can be further distinguished among the following subgroups in (2).

(2) Subgroups of the Min dialect

- ° Northern Min or Min Bei (Nanping Prefecture)
- ° Shaojiang Min (Shaowu, Jiangle, etc.)
- ° Eastern Min or Min Dong (Fuzhou, etc.)
- ° Central Min (Sanming Prefecture)
- ° Pu-Xian Min (Putian and Xianyou)
- ° Southern Min or Min Nan (Xiamen, Taiwan, etc.)
- ° Leizhou Min (Leizhou City)
- ° Hainan Min (Wenchang)

---

[4] The more accurate term here is the Yue dialect, instead of Cantonese.

The subgroups in (2) are arranged roughly from north to south. The place names in the parentheses are the representative versions of each subgroup.

A more recently recognized new group is the Jin dialect[5] spoken in Shanxi and the surrounding areas such as Hebei, Inner Mongolia, Henan and Shaanxi. It was included in the Mandarin group in the traditional classification. But in many newer classification systems such as in the *Language Atlas of China* (Wurm et al. 1987), the Jin dialect is a separate primary group on par with Mandarin.

Table 1 shows the relative proportion of each dialect among speakers of the major varieties of Chinese.

TABLE 1. Size of Chinese Dialects[6]

Chinese varieties	% of L1 Speakers
Mandarin	66.2%
Jin	5.2%
Min (all subgroups)	6.2%
Wu	6.1%
Cantonese	4.9%
Gan	4.0%
Hakka	3.5%
Xiang	3.0%
Other	0.9%

The percentage is the proportion of first-language speakers. The largest group in Table 1 is Mandarin at 66.2%. If we combine Jin and Mandarin it is almost ¾ of all speakers (71.4%). The second largest group is Min (6.2%), as one group including all the varieties in (2). The Wu dialect has more or less the same number of speakers (6.1%) as the Min dialect. Cantonese (4.9%) follows Wu. Then the next groups are Gan (4.0%), Hakka (3.5%) and Xiang (3.0%). The "Other" category includes smaller dialects such as Pinghua and Huizhou. Since there are no Wikipedia sites written in Pinghua, Huizhou and other lesser-known dialects, I will not discuss these dialects in the "Other" category in this current article.

Now let's see the data regarding the Wikipedia sites written in Chinese dialects. In my research, data were collected over two years. I look at two snapshots of Chinese dialect Wikipedia sites. Table 2 shows the data recorded on March 9, 2015. Table 3 shows the data recorded on May 18, 2017.

---

[5] Jìn Yǔ 晋语.

[6] Source: http://en.wikipedia.org/wiki/Varieties_of_Chinese [Retrieved on November 20, 2017], where the data are taken from the 2nd edition of *Language Atlas of Chinese* (Chinese version), edited by the Chinese Academy of Social Sciences, published by the Commercial Press in 2012.

TABLE 2. Meta-wiki data of sites in Chinese dialects as of March 9, 2015

Rank	Dialect	Articles	Admins	Users	Active Users
15	Mandarin	814322	80	2007603	7949
79	Cantonese	35317	8	100829	167
119	Min Nan	12798	6	21324	38
143	Gan	6305	2	21862	24
161	Hakka	4512	0	13473	16
175	Wu	3536	3	31800	22
195	Min Dong	2518	1	8907	11

TABLE 3. Meta-wiki data of sites in Chinese dialects as of May 18, 2017

Rank	Dialect	Articles	Admins	Users	Active Users
15	Mandarin	941817	81	2375687	7363
39	Min Nan	208033	5	28898	66
76	Cantonese	53986	10	136487	239
147	Hakka	7423	0	18904	22
153	Min Dong	6432	3	11532	19
154	Gan	6388	2	26784	17
159	Wu	5812	3	49594	19

The data here were downloaded from the meta wiki webpage that can be easily retrieved from the follow address https://meta.wikimedia.org/wiki/List_of_Wikipedias. The different columns represent the overall ranking of the website among all Wikipedia websites in terms of total number of articles, the dialect used on the website, the total number of articles on that website, the total number of administrators in that specific wiki community, the total number of users, and the active users among them. According to the meta-wiki page, "Active Users" are defined as those that have registered and "have made at least one edit in the last thirty days" as of the date of the data collection. Thus "users" are those that have registered, being part of the relevant virtual linguistic community. The number of users is an indicator of the size of the virtual linguistic community, and the number of articles is an indicator of how well each site is doing generally.

Now let's examine the numbers in Table 2 in detail first. The relative rankings of all Wikipedia websites of a variety of Chinese in terms of the total number of articles are Mandarin, Cantonese, Min Nan, Gan, Hakka, Wu and Min Dong. The Xiang, Min Bei and Pu-Xian versions of Wikipedia were being incubated at the time of data collection in

Table 2. Mandarin as the largest group of dialects (Table 1) has the largest Wikipedia site in terms of the number of articles, administrators, users and active users.[7]

Cantonese ranks second in both the number of users and the total number of articles, although in terms of speakers, Cantonese is behind Min and Wu. Some explanations for this relatively higher ranking of Cantonese can be found in the high internal homogeneity among all varieties of Cantonese, and the existence of a regional *lingua franca* based on the Guangzhou version of Cantonese. In this sense, the Cantonese linguistic community can pool the resources together more easily. Another reason might be due to the large number of overseas Cantonese speakers, e.g. in Europe and North America. In terms of Min, if we add the numbers of articles of Min Nan and Min Dong, their combined ranking is still third, right after Cantonese. Note that the size of Min in Table 1 is based on all varieties of Min. Thus the actual number of speakers of Min Nan an Min Dong should be much smaller, which can partially explain the ranking of Min Nan Wikipedia after Cantonese. The total number of users in the Min Nan and Min Dong virtual linguistic community ranks after Cantonese and Wu, but it is quite close to Wu.

The Gan and Hakka rankings on meta-wiki are more or less comparable to their real linguistic communities (Table 1). Xiang is the smallest among these major groups, and it is not surprising that its Wikipedia site was being incubated.

The only surprising fact from Table 2 is the low ranking of Wu in terms of total number of articles. But in terms of the total number of users, the virtual linguistic community of Wu ranks third, right after Cantonese. This is more in line with the size of the linguistic community in Table 1. This suggests that there are more people who are interested in the project of Wu Wikipedia than those who are actually contributing to the content creation.

To summarize the data in Table 2. The relative rankings of Wikipedia sites in major Chinese dialects are more or less comparable to their linguistic community sizes (Table 1). This shows that most of these linguistic communities are actively using Wikipedia as a way to promote their own dialects.

Now let's compare the data from May 18, 2017 as shown in Table 3, with the data in Table 2 to see the growth of these Wikipedia sites. One trend is that most of these sites have higher rankings in Table 3 in terms of both the number of articles and number of users than their own rankings in Table 2, thus showing growth and maintenance of these sites over time. The Mandarin site has grown but maintains its ranking at 15. One

---

[7] As a comparison, English ranks No. 1 of all Wikipedia sites. As a global language, it is easy to see why English ranks No. 1 on Wikipedia. However, with the largest number of speakers, Mandarin's ranking of No. 15 seems a little too low. There may be several reasons for this. For example, censorship within China intermittently blocks access to Wikipedia. Also there are Chinese equivalents of Wikipedia, such as *Bǎidù Bǎikē* 百度百科 and *Hùdòng Bǎikē* 互动百科, thus diluting the resources that users devote to one particular website. But since my focus is on Chinese dialects, instead of Mandarin in comparison to other major world languages, I will not go into any details here.

exception is the Gan Wikipedia, which dropped in its ranking from 143 to 154, although the number of articles and the number of users both increased. This shows a lack of momentum in the development of the Gan Wikipedia project. Those that were incubated in 2015 were still not up and running as of May 18, 2017, thus showing lack of growth.

The site that shows the most growth is Min Nan, which jumped from 119 in 2015 to 39 in 2017. Min Dong has also increased its ranking considerably as well. Although the Wu Wikipedia has also increased its ranking from 175 to 159, it is ranked last now among all these sites in terms of the total number of articles, although the number of users on the Wu Wikipedia is still third right after Mandarin and Cantonese. On the other hand, Cantonese has improved slightly in its ranking, and it seems that the Cantonese site is becoming quite stable and shows the highest number of administrators, users and active users after Mandarin.

To sum up the data in Table 3, we still see that the relative sizes of these Wikipedia sites are more or less proportional to those of their linguistic communities (Table 1), except in the case of Wu. Most of these sites have improved their overall rankings within the two years. Min Nan shows the largest growth, while Cantonese is stabilizing and becoming a more mature website.

By examining and comparing the data from Table 1, Table 2 and Table 3, we may give the following factors as contributing to the growth of a Wikipedia site written in a Chinese dialect.

First the internal homogeneity is a very important factor. Although officially speaking, Wu ranks higher than Cantonese in terms of the total number of speakers, the internal homogeneity of Cantonese is much higher than that of Wu. Some southern Wu dialects are actually not mutually intelligible with the northern Wu dialects. Even among the northern Wu dialects, Shanghainese as the prestigious variety can be understood by many speakers of Wu but they may not be able to contribute to creating content in Shanghainese.

The second major factor is the existence of overseas diaspora communities. In terms of both Cantonese and Min Nan, there are large linguistic communities in Europe, North America and Southeast Asia. These communities can help to bypass the restrictions on Internet access set forth within China. In this aspect, Wu dialect has much smaller overseas communities compared to Cantonese and Min.

Third, political factors also play a major role. For example, the growth of Min Nan Wikipedia is likely supported by the linguistic movements in Taiwan. The stabilization of Cantonese Wikipedia is likely supported by the fact that the majority language in Hong Kong is Cantonese, not Mandarin or English. The Taiwan government and the Hong Kong government, together with the local linguistic communities, have also taken measures to standardize aspects of Min Nan, Cantonese and Hakka.

Another factor is writing systems. This will be the main focus of this article. In the next two sections, I will show examples of the type of writing systems in each of the

Wikipedia sites in Chinese dialects, and then I will compare these writing systems to how the Wikipedia sites in these writing systems are faring.

3. Writing Chinese Dialects

A Chinese dialect can be written in either a character-based system or a phonetic writing system. The Wikipedia sites that are written in a character-based system include Mandarin, Cantonese, Wu and Gan. Let's take a look at a snapshot of these websites by using the article on the city of Shanghai as an example, as shown in Figures 1, 2 and 3. I omit Mandarin because the writing system is standardized and well-known.

Figure 1 shows the article from the Cantonese Wikipedia site.

FIGURE 1. Wikipedia page about Shanghai written in Cantonese

Cantonese is the only Chinese dialect that has developed a stable popular writing system which has been standardized to a greater extent than other dialects. According to Snow (2004: 6), written Cantonese can be traced back to the late Ming Dynasty (1368-1644), when books of verse were printed. Cantonese opera scripts were written down in characters in the early 20th century. Nowadays, although written Cantonese in many cases may contain elements from standard Chinese and Classical Chinese, the writing system is nonetheless capable of writing down spoken Cantonese (Snow 2004: 60).

Figure 2 shows the article from the Wu Wikipedia.

FIGURE 2. Wikipedia page about Shanghai written in Shanghainese

Traditionally the representative version of Wu is that of Suzhou. Vernacular writing based on the Suzhou dialect can be traced as far back as early Qing Dynasty (1644-1912). There are texts of fiction and opera written in mixed Classical Chinese and Suzhou dialect by using characters. In the formation of the Shanghai dialect, one important contribution is Suzhou dialect. Therefore even though the contemporary representative version of the Wu dialect is that of Shanghai, the tradition of writing Wu dialects has been present in Shanghai as well. According to the texts cited by Qian (2003: 357–394) from the mid-19th and early 20th centuries, colloquial Shanghainese could be written down with characters. The degree of popularity and standardization of written vernacular Shanghainese is to a much lesser degree compared to Cantonese.

Figure 3 shows the article from the Gan Wikipedia.

上海

上海（上海話，[zɑ̃'hɛ]），簡稱滬，又叫申，係中華人民共和國一隻城市，係四隻直轄市之一。佢坐落到中國東部嗰長江三角洲，東臨東中國海。佢係眼下中國發展最快嘅城市，更加係經濟

FIGURE 3. Wikipedia page about Shanghai written in Gan

The representative version of the Gan dialect is that of Nanchang. The internal homogeneity of the Gan dialect is relatively high. Although the Gan dialect can be written with a character-based writing system, e.g. as in the dictionary by Xiong (1995), there has not been a tradition of a popular vernacular writing in the Gan dialect.

All of the other Chinese dialect Wikipedia sites are currently written in a phonetic writing system. Figure 4 is the Min Nan page about Shanghai.

Siōng-hái-chhī

Siōng-hái-chhī (上海市), kán-chheng Hō͘ (滬), piat-chheng Sin (申), Tiong-kok siōng tōa ê siâⁿ-chhī, tiàm-tī Tiong-kok tang-pō͘ iân-hái, Tiông-kang chhut-hái-kháu ê lâm-sì. Siōng-hái sī Tiong-

FIGURE 4. Wikipedia page about Shanghai written in Southern Min

As with all of the other southern Chinese dialects, Southern Min can be written with characters. The earliest known written vernacular Southern Min is an opera script titled *The Tale of the Lychee Mirror* [Lì Jìng Jì 荔镜记] dated 1566 in the Ming Dynasty. According to Lin (1999), the development of written Taiwanese using a character-based system has not been up to the degree of Cantonese, and there are more issues with standardization as well, although speakers of Taiwanese nowadays do use the character-

based writing system, especially in popular culture, e.g. song lyrics, film subtitles, etc. The Taiwan government has taken measures to standardize the character set used for Taiwanese Southern Min since 2007.

On the other hand, Southern Min has a long tradition of phonetic writing, such as those designed by early missionaries. Some of these systems were once quite popular and had a basis of literacy among speakers who might not know how to write Chinese characters. One system is the POJ system (Pėh-ōe-jī 白话字), or Church Romanization, designed by the Presbyterian Church in the 19th century. It has a sizable literature as well. Apart from political reasons that might disfavor using a character-based system, the practical usefulness of the phonetic writing system does seem to show the choice is reasonable. However, as shown in Figure 7, on the discussion page the contributors also use the character-based system almost exclusively.

FIGURE 5. The discussion page in Southern Min

Figure 6 shows the article about Shanghai writing in Min Dong based on Fuzhou.

FIGURE 6. Wikipedia page about Shanghai written in Min Dong

The character-based writing of Fuzhou can be traced back to the 16th century. The early records include the rime book *Qī Lín Bāyīn* [戚林八音 The Book of Eight Tones],

473

and the fiction writing *Mǐn Dū Bié Jì* [闽都别记 Alternative Records of the Capital of the Min] from the mid-Qing Dynasty. However the writing tradition in characters in Eastern Min has not been as popular as in Southern Min. Consequently practice of writing Eastern Min in characters is confined to a limited group of people. The once popular form is the BUC system (Bàng-uâ-cê 平话字) designed by missionaries in the 19th century.

Figure 7 shows the article on Shanghai written in Hakka. Note there is one line of characters after the title, which gives a link to edit the article. But the article itself is written in a phonetic writing system.

FIGURE 7. Wikipedia page about Shanghai written in Hakka

Hakka can be written in Chinese characters, although there has not been much study on this topic. In terms of the phonetic systems, there have been systems designed by missionaries, e.g. Phàk-fa-sṳ (白話字) created by the Presbyterian church in the 19th century. The Taiwanese Hakka linguistic community and the Taiwan government also adopted the Taiwanese Hakka Romanization System in 2012.

Although the Wikipedia sites in Xiang, Min Bei and Pu-Xian Min are still being incubated, some pages exist nonetheless. The Xiang Wikipedia uses a character-based system, but has two side-by-side versions, one for Old Xiang, and one for New Xiang, which is due to the significant differences between these two versions of Xiang. In this sense, the Wu Wikipedia could also have multiple versions. The Min Bei and Pu-Xian Min Wikipedia sites use a phonetic system similar to earlier systems designed by missionaries in the 19th century.

The data here are summarized in Table 4. The dialects in parentheses are those Wikipedia sites still being incubated. Although in theory and in practice (to varying degrees) all Chinese dialects can be written with a character-based writing system, writing tradition and practical needs vary and therefore on these Wikipedia sites, different writing systems are used, among other reasons. Character-based systems are used on the Wikipedia sites of Mandarin, Cantonese, Wu, and Gan, and also on the preliminary pages of Xiang. In the Min dialects (i.e. the four Min Wikipedia sites), and in Hakka, a phonetic

writing system is used, which mostly can be traced back to earlier systems designed by missionaries in the 19th century.

TABLE 4. Writing Chinese Dialects on Wikipedia

Character-Based	Letter-Based
Mandarin	Southern Min
Cantonese	Hakka
Gan	Min Dong
Wu	(Min Bei)
(Xiang)	(Pu-Xian Min)

In the next section, I look at the choice of writing system in connection with the development and growth of the Wikipedia sites.

4. Writing system and linguistic diversity

Systematic research on the writing systems used in Chinese dialects is quite rare. The practice of writing Chinese dialects has also been equally sparse for the most part of the history of the Chinese language. This can be explained by the following factors.

First, the *Law of the People's Republic of China on the Standard Spoken and Written Chinese Language* recognizes the use of languages of different ethnic groups within China. The minority languages, e.g. Mongolian, Zhuang etc., have the legal rights to use their own languages alongside Putonghua. For the minority languages that did not have a writing system, or in the case of the Zhuang language which has a character-based writing system[8], new phonetic writing systems were created to standardize the use of these languages by the Chinese government since 1949 (Zhou 2003). Despite the various issues with the language policy towards minority languages in China, the legal status of minority languages at least draws attention to the use and standardization of these languages both in the spoken form and in the written form. However, the various Chinese dialects are not recognized as such. Therefore, the standardization and the creation of a writing system for Chinese dialects were never formally considered. Even in Taiwan, the standardization of the writing systems for Taiwanese and Hakka is still quite recent, and these measures have limited effects outside Taiwan in the Southern Min and Hakka linguistic communities.

Second, the language laws in China also do not allow the explicit use of dialects in all official media. Although there have always been gaps between language laws and the implementation of such laws in language practices, in most cases dialect writings are not possible. Especially in primary education, no explicit teaching in writing dialects is

---

[8] Gǔ Zhuàngzì 古壮字 in Chinese, or Sawndip 倳字 ("saw + ndip": writing raw) in Zhuang. It is a similar system to the Chữ Nôm 字喃 used in Vietnam.

allowed, although some areas, e.g. Shanghai, have introduced classes of dialects outside the normal curriculum in elementary schools. More importantly, the language laws command economic incentives. Learning Mandarin means more economic and employment opportunities, and the use of writing in dialects is practically quite limited.

Third, traditionally the use of Chinese dialects mostly is confined to the spoken form, and this is true of most dialects even nowadays. Thus when people write, they tend to write standard Chinese. The need to write dialects is not strong enough to call for a full writing system for most dialects.

Fourth, all Chinese dialects share a core vocabulary to different extents (Wang 1994: 1448; Wang 1998: 530), and therefore writing Chinese dialects have always been possible with Chinese characters, with additional dialect characters[9] added. The need to create a dialect writing system has not been urgent for most dialects, because they can all be written somehow and to some degree for practical purposes. In cases of words for which the etymologically correct characters[10] cannot be determined, or are too specialist for the average speaker to use, homophonous characters can be used to write those words.

For all these reasons, the research and practice in writing dialects in the Chinese context have been quite rare. Now with the emergence of new technology and media such as Wikipedia, which gives Chinese dialects a channel to become fully functional in both the spoken form and the written forms, the lack of systematic research and practice in writing definitely is a major obstacle to the growth of these dialect Wikipedia sites.

But all dialects are not equal. As I have discussed in section 3, Cantonese has created and standardized the writing system to the most degree among all Chinese dialects. Writing Cantonese is not really an issue. This can be shown in the relative high ranking of the Cantonese Wikipedia as shown in Table 2 and Table 3. The Cantonese Wikipedia is relatively stable and has the largest user base after Mandarin Wikipedia.

In contrast, the Wu dialect has a large linguistic community but ranks last in Table 3 in terms of the number of articles, although the total number of users ranks right after Cantonese. Among the factors mentioned before, e.g. the actual speakers of Shanghainese being much smaller than all Wu dialect speakers, the lack of a standardized writing system and the lack of basic literacy education might also be factors.

Although the Gan Wikipedia is written in a character-based system, it is to an even lesser degree in terms of standardization and basic literacy education. Thus Gan Wikipedia is actually losing its momentum, as shown in the data in Table 2 and Table 3. Within the two years, there was little increase of the total number of articles and the ranking of the Gan Wikipedia dropped from 143 to 154. Similarly, in the Xiang Wikipedia, the same issues exist, in addition to the fact that the two versions of Xiang, i.e. Old Xiang and New Xiang, are so different that they call for two versions of the Xiang Wikipedia.

---

[9] Fāngyán zì 方言字
[10] Fāngyán běnzì 方言本字

476

Regarding Min Nan, people have been using characters to write in recent decades, especially in Taiwanese popular culture. However Min Nan Wikipedia uses a phonetic writing system. This might be due to three factors. First, the need for a unique identity as a political factor can lead some speakers to favor a phonetic system, since it looks radically different from Mandarin Chinese writing. Second, the Southern Min dialect is probably the most advanced among all Chinese dialects in terms of the phonetic writing system. Although phonetic writing systems were created by missionaries in the 19th century for many varieties of Chinese, the POJ system was the most successful in producing a large body of literature and in its literacy education. Third, the standardization that took place in Taiwan only has limited effects on Southern Min spoken outside Taiwan. Therefore to reach a larger readership, a phonetic writing system does seem to have its advantage given the high internal homogeneity among the major Southern Min speaker communities. As can be seen from Table 2 and Table 3, the growth of Min Nan Wikipedia within the two years was phenomenal! Although this has to be ascribed to the enthusiasm of a smaller number of contributors, as can be seen from the increase of the total number of articles from 12,798 to 208,033, a 15-time increase, while the total number of users only increased from 21,324 to 28,898. But there is no doubt the phonetic writing system facilitates the creation of articles.

Hakka has a similar situation in terms of its writing system compared to Min Nan, although the practice of writing Hakka in characters has not been to the same extent as in Min Nan. The Hakka Wikipedia grew tremendously, as can be seen by the 65% increase of total number of articles, and 40% increase in total number of users. The ease of the phonetic writing system is likely a contributing factor.

For the other two Min dialect Wikipedia sites, i.e. Min Bei and Pu-Xian, their choice of using a phonetic writing system is based on a lack of character-based writing. But the phonetic writing system is equally less popular in practical use. Therefore there is no actual momentum in bringing these sites out of the incubator. We see here the lack of a practical popular writing system does seem to be an obstacle to the growth of these sites.

In summary, I argue that a practical popular writing system is an important factor in the growth and maintenance of Chinese dialect Wikipedia sites. By "popular" I mean the actual use of the writing by the average speakers. For the most successful ones, i.e. Cantonese and Min Nan, both enjoy a popular writing system that has a large user base, and their virtual linguistic communities can build upon such a user base to promote these dialects. For the less successful ones, e.g. Xiang, Wu, Min Bei, Pu-Xian, and Gan, the lack of a practical popular writing system impedes the growth and maintenance of these sites, hence hampering efforts to promote these dialects. Compared to these two groups, the Hakka Wikipedia seems to be doing quite well, maybe more or less in the middle.

5. Conclusions

This article is part of my larger project to explore the creation of the standard form of modern Chinese, i.e. Putonghua, and its relation to nation-building. Here I have

shown that Wikipedia is an important tool to promote linguistic diversity. A practical popular writing system is needed to guarantee the success of such sites. In connection to what writing systems to use, there are various other issues.

One issue is related to the classification of Chinese dialects. Although there are seven major groups, the actual mutually-unintelligible forms of Chinese can be much greater than seven. Even among the Mandarin group, speakers from different areas do not necessarily understand each other. Moreover, the Jin dialect has been recognized by many scholars as a separate group. Therefore there is the issue of how many Wikipedia sites of Chinese dialects should be recognized. As Ensslin (2011) points out, "Wikipedia defines itself as 'the biggest multilingual free-content encyclopedia on the internet', thus featuring an explicit language policy in its mission statement". Thus to be recognized as a language by Wikipedia is not an automatic process.

Another issue is internal homogeneity. Among many dialect groups, there are local speech forms that are not mutually-intelligible. For example, the distinction between Northern Wu and Southern Wu, and that between Old Xiang and New Xiang. Even among groups or subgroups that have greater internal homogeneity, which version should be regarded as the representative is a major issue, such as in the case of Wu. These two issues need to be sorted out before standardization on the form and writing of dialects can be carried out. Then after standardization, literacy education and content or literature creation need to be addressed.

Furthermore for the majority of Chinese dialects, there has never been a writing system, either character-based or phonetic. If one is to create a writing system, which way is to go? In terms of the advantages and disadvantages of these two types of writing, the character-based system is considered more authentically Chinese, and can be partially understood by speakers of other dialects. But for the uniquely local vocabulary, it is more difficult to write with characters. Moreover, the etymologically correct characters might be very rare characters that can be difficult to input. The unique dialect characters may also be difficult to input. The phonetic system can be considered less authentically Chinese, and the diacritics for tones and vowels can be overwhelming both typographically and in terms of readability. However a phonetic system is much easier to create and to learn for everyone, including people who do not know Chinese characters. Therefore a phonetic writing system is more efficient if one is to create a writing system for a dialect that has never been systematically written. Such systems can be very instrumental in promoting linguistic diversity, especially by using Wikipedia sites.

This paper has drawn attention to the importance of writing systems for Chinese dialects in the process of promoting linguistic diversity, especially with new technological tools and channels such as Wikipedia, given the context where language policy restricts the maintenance of dialects. It is my hope that more research will be conducted in this respect in the future to solve both the theoretical and practical issues.

REFERENCES

BAKER, COLIN, and SYLVIA PRYS JONES (eds). 1998. *Encylcopedia of bilingualism and bilingual education.* Multilingual Matters.

DENG, HONGBO (邓洪波). 1994. Zhèngyīn shūyuàn yǔ Qīngdài de guānhuà yùndòng 正音书院与清代的官话运动 [Mandarin Academies and the Mandarin Campaign in Qing Dynasty]. Journal of East China Normal University 3: 79-86. Shanghai, China.

DONG, HONGYUAN. 2014. *A history of the Chinese language*. New York, NY, USA and Abingdon, UK: Routledge.

DONG, HONGYUAN. 2015a. Mandatory Mandarin: An archival study of Qing dynasty language policy. Manuscript, George Washington University.

DONG, HONGYUAN. 2015b. New media as channels for linguistic diversity: A case study of Chinese. Manuscript, George Washington University.

DONG, HONGYUAN. 2016. An archival study on linguistic reforms in pre-modern East Asia. Manuscript, George Washington University.

DONG, HONGYUAN. 2017. An historical comparative view on contemporary language policy in China. Manuscript, George Washington University.

ENG, ROBERT Y. 2010. Is Cantonese in danger of extinction? The politics and culture of language policy in China. Blog post August 20, 2010 on *China Notes: Superfluous Musings of a Chinese Historian*. Retrieved on November 20, 2017 from http://chinamusictech.blogspot.com/2010/08/is-cantonese-in-danger-of-extinction.html

ENSSLIN, ASTRID. 2011. What an un-wiki way of doing things: Wikipedia's multilingual policy and metalinguistic practice. *Journal of Language and Politics* 10(4): 535-561.

IVKOVIC, DEJAN., and HEATHER LOTHERINGTON. 2009. Multilingualism in cyberspace: Conceptualising the virtual linguistic landscape. *International Journal of Multilingualism*, 6(1): 17-36.

LIN, ALVIN. 1999. Writing Taiwanese: The development of Modern Written Taiwanese. *Sino-Platonic Papers* 89, Dept. of Oriental Studies, University of Pennsylvania.

LIU, JIN. 2013. *Signifying the local: Media productions rendered in local languages in Mainland China in the new millennium*. Leiden: Brill

LIU, JIN, and HONGYIN TAO. 2009. Negotiating linguistic identities under globalization: language use in contemporary China. *Harvard Asia Pacific Review*, 10(1): 7-10.

LIU, JIN, and HONGYIN TAO. 2012. *Chinese under globalization: Emerging trends in language use in China*. World Scientific.

MAIR, VICTOR H. 1991. What is a Chinese "dialect/dialect"? Reflections on some key Sino-English linguistic terms. *Sino-Platonic Papers, 29*. Department of Oriental Studies, University of Pennsylvania.

MAY, STEPHEN. 2006. Language policy and minority rights. In Thomas Ricento (ed.) *An introduction to language policy: Theory and method.* 255-272. Blackwell.

QIAN, NAIRONG (钱乃荣). 2003. *Shànghǎi yǔyán fāzhǎn shǐ* 上海语言发展史 [A history of the development of the language of Shanghai]. Shanghai, China: Shànghǎi Rénmín Chūbǎnshè 上海人民出版社 [Shanghai People's Publishing House].

SHOHAMY, ELANA, and DURK GORTER (eds). 2008. *Linguistic landscape: Expanding the scenery*. Routledge.

SIMMONS, RICHARD VANNESS. 2017. Whence came Mandarin? Qīng guānhuà, the Běijīng dialect, and the national language standard in early Republican China. *Journal of American Oriental Society* 137, 1: 63-88.

SNOW, DON. 2004. *Cantonese as written language: The growth of a written Chinese vernacular*. Hong Kong, China: Hong Kong University Press.

WANG, HUI. 2014. *China from empire to nation-state*, translated by Michael Gibbs Hill. Cambridge, MA: Harvard University Press.

WANG, WILLIAM S.-Y. 1994. Glottochronology, lexicostatistics and other numerical methods. In the *Encyclopedia of language and linguistics*, edited by R. E. Asher and J. M. Y. Simpson. 1445-1450. Oxford and New York: Pergamon Press.

WANG, WILLIAM S.-Y. 1998. Three windows on the past. In *The Bronze Age and early Iron Age peoples of eastern Central Asia*, edited by Victor H. Mair. University of Pennsylvania Museum Publications. 508-534.

WU, YONGBIN (吴永斌). 2008. Shì xī Yōng-Qián nián jiān de guānhuà yùndòng 试析雍乾年间的官话运动 [An analysis of the Mandarin Campaign in the Yongzheng-Qianlong era]. *Mínzú Jiàoyù Yánjiū* 民族教育研究 [Journal of Research on Education in Ethnic Minorities]. 2:113-116. Beijing, China.

WURM, STEPHEN ADOLPHE; RONG LI; THEO BAUMANN; and MEI W. LEE (eds). 1987. *Language Atlas of China*. Hong Kong: Longman.

Xiong, Zhenghui (熊正辉). 1995. *Nánchāng fāngyán cídiǎn* 南昌方言词典 [A dictionary of the Nanchang dialect]. Nanjing, China: Jiāngsū Jiàoyù Chūbǎnshè 江苏教育出版社 [Jiangsu Education Publishing House].

YUAN, JIAHUA (袁家骅) et al. 1960. *Hànyǔ fāngyán gàiyào* 汉语方言概要 [An outline of Chinese dialects]. Beijing, China: Wénzì Gǎigé Chūbǎnshè 文字改革出版社 [Writing System Reform Publishing House].

ZHOU, MINGLANG. 2003. *Multilingualism in China: The politics of writing reforms for minority languages 1949-2002*. Walter de Gruyter.

ZHOU, MINGLANG. 2006. Theorizing language contact, spread, and variation in status planning: A case study of Modern Standard Chinese. *Journal of Asian Pacific Communication,* 16, 2: 159-174.

ZHOU, MINGLANG, and HONGKAI SUN (eds). 2004. *Language policy in the People's Republic of China: Theory and practice since 1949*. Springer.

Path of Vowel Raising in the Chengdu Dialect of Mandarin[1]

Hai Hu and Yiwen Zhang
Indiana University Bloomington

He and Rao (2013) reported a raising phenomenon of /a/ in /Xan/ (X being a consonant or a vowel) in Chengdu dialect of Mandarin, i.e. /a/ is realized as [ɛ] for young speakers but [æ] for older speakers, but they offered no acoustic analysis. We designed an acoustic study that examined the realization of /Xan/ in speakers of different age (old vs. young) and gender (male vs. female) groups, where X is one of the three conditions: 1) *unaspirated consonants* C ([p], [t], [k]), 2) *aspirated consonants* C^h ([p^h], [t^h], [k^h]), and 3) *high vowels* V ([i], [y], [u]). 17 native speakers were asked to read /Xan/ characters and the F1 values were extracted for comparison. Our results confirmed the raising effect in He and Rao (2013), i.e., young speakers realize /a/ as [ɛ] in /an/, whereas older speakers in the most part realize it as [æ]. Also, female speakers raise more than male speakers within the same age group. Interestingly, within the /Van/ condition, older speakers do raise /a/ in /ian/ and /yan/. We interpret this as /a/ first assimilates to its preceding front high vowels /i/ and /y/ for older speakers, which then becomes phonologized in younger speakers in all conditions, including /C^han/ and /Can/. This shows a possible trajectory of the ongoing sound change in the Chengdu dialect.

0. Introduction

He and Rao (2013) report a raising phenomenon in Chengdu dialect of Mandarin. Specifically, native speakers born in the 1950s pronounce the phoneme /a/ as [æ] in nasal environment /an/, whereas younger generations (born after the 1980s, roughly) raise [æ] to [ɛ] in most instances. The degree of raising from [æ] to [ɛ] seems to be also related to the environment for /an/. It is more likely to occur in environments where there are adjacent high vowels [i], [u] or [y]. He and Rao (2013) also report that the raising effect originates from conditions where the consonant preceding /an/ is aspirated.

This acoustic study tries to examine such raising effect by comparing the vowel height of /an/ in different phonological environments of two age groups, the young and old. We will also look at the role of gender, and phonological environment in the raising phenomenon.

The results show that in general the raising of /an/ in young speakers is confirmed, but the aspirated-unaspirated contrast is not born out. We also discuss the implications of the study and briefly discuss the possible path for raising in the Chengdu dialect.

1. Literature review

In this section, we review previous literature on vowels in Chengdu dialect with a focus on the rhyme /an/, and how others have measured vowel raising phenomena in other languages.

Year of Analysis	1941	1956	1956		c.a. 1982	c.a. 1983	c.a. 2006
Published in	Yang (1984)	Zhen (1958)	Zhen, and (1960)	Hao Chen	Liang (1982)	Zhen (1983)	He and Rao (2013)
/iai/	a	ε	ε		ε	NA	ε
/ian/	e	æ	ε		ε	ε̃	æ / ε
/yan/	e	æ	ε		ε	ε̃	æ / ε
/Can/	a	A	æ		NA	ã	æ / ε
/uan/	a	A	æ		NA	ã	NA

Table 1: /an/ in previous studies

1.1. Literature on vowels in Chengdu dialect

Chengdu is the capital city of Sichuan Province in southwest China; Chengdu dialect is usually categorized as Southwest Mandarin, which is similar to Standard Mandarin. In Chengdu dialect, one syllable usually corresponds to one morpheme, allowing us to use mono-syllabic tokens in the recording. There are 6 possible syllable patterns: CV, CVN, CVV, CVVN, V and VN. /an/ can appear in CVN, CVVN and VN.

Many previous studies have documented the realization of /an/ (see Table 1). From Table 1, we see almost a three-way distinction:

/iai/ – /ian/ /yan/ – /Can/ /uan/

It can be drawn that historically /iai/ is almost always realized as [ε]. On the other hand, /ian/ and /yan/ are gradually raised whereas /Can/ and /uan/ seem to be raised only very recently, if raised at all. He and Rao (2013) is the first and possibly the only literature that documents the raising of /Can/, despite the fact that such raising effect has been noticed by many new comers to Chengdu who immediately notice that Chengdu people pronounce /pan/ or /fan/ very differently from people in nearby cities. He and Rao (2013) report that female speakers born after 1980s exhibit a strong raising of /a/ in /an/,

and male speakers born after 1990s generally raise /a/ in /an/. Thus female speakers seem to lead the vowel change.

In addition, He and Rao (2013) also reports that the raising is more prominent when /an/ is preceded by an aspirated consonant, but no reason is provided. We will test this in our study by comparing the vowel height of /Can/ and /C^{h}an/.

However, they do not mention if their study is an acoustic analysis. Apart from that, very few studies have investigated the age difference and phonological environment of the Chengdu vowel raising. Therefore, we feel the need to document this possible vowel change.

1.2. Vowel raising measurement
1.2.1. /æ/ raising in GA Northern Cities Chain Shift

Many studies have looked at vowel raising in different languages/varieties. Clopper, Pisoni and de Jong (2005) collected data from 6 dialectal regions of the US and plotted vowel charts of these 6 regions. From their vowel chart, we can clearly see the raised /æ/ in Northerners which is part of the Northern Cities Chain Shift. Statistical analysis such as post-hoc Tukey is also used to confirm the raising of /æ/. In addition, they used Labonov normalization for all speakers (Labonov, 1971).

1.2.2. New Zealand vowel raising

Watson, Maclagan and Harrington (2000) compare recordings of 1948 to 1995 to see if the vowels of New Zealand English (NZE) have changed, particularly whether /ɛ/ has been raised or not.

Two methods are used to determine the raising of /ɛ/. First, they use /i/ and /æ/ as reference vowels and discover that in Old NZE /ɛ/ is almost in the mid point of /i/ and /æ/, whereas in Modern NZE, /ɛ/ is very close to /i/, which indicates the raising of /ɛ/. Second, they use t-test on the F1 and F2 values of the vowels they are investigating. The significance difference found between /ɛ/ in Old and Modern NZE confirms the raising. Labonov normalization is also used in their study.

In sum, two methods are commonly used in determining vowel raising: 1) plotting vowel chart and eyeballing, 2) statistical test, be it t-test or post-hoc Tukey tests in ANOVA.

2. Method
2.1. Research questions

Based on previous literature, we ask the following research questions in this study.

1. **Age**: Does the young age group raise /a/ more than the old age group when /a/ is followed by a nasal coda /n/? That is, are there differences in the height of /a/ in the experimental conditions of /an/ between the young and older age groups?
2. **Gender**: Within each age group, do female subjects exhibit more raising?
3. **Phonological environment**: Is the height of /a/ different in different phonological environments?

2.2. Subjects

Altogether 21 native speakers of Chengdu dialect were recruited originally, of which three are older female speakers, four are older male speakers, eight are young female speakers and six are young male speakers. Four female subjects are excluded for poor quality of recording.

Background questionnaires are collected from the young subjects (unfortunately we were unable to collect background questionnaires from the old age group, but it is ensured that they are all native speakers of Chengdu dialect). All of them are native speakers of Chengdu dialect. The parents of all young female subjects are also born in Chengdu, whereas only two of the young male subjects' parents are born in Chengdu. The daily communication in homes of all subjects are Chengdu dialect. All young subjects have spent most of their lives (mostly more than 18 years) in Chengdu (see Table 2).

	No.	Age	Gender	ParentsChengdu?	YrsInChengdu
Young Age Group	ym1	27	M	n	19
	ym2	27	M	n	18
	ym3	27	M	n	20
	ym4	25	M	n	6
	ym5	27	M	y	22
	ym6	27	M	n	18
	yf1	26	F	y	18
	yf2	26	F	y	18
	yf3	26	F	y	18
	yf4	28	F	y	22
Old Age Group	om1	59	M	NA	59
	om2	57	M	NA	38
	om3	57	M	NA	57
	om4	56	M	NA	56
	of1	49	F	NA	49
	of2	49	F	NA	49
	of3	43	F	NA	43

Table 2: Language background of subjects

2.3. Material
2.3.1. Reference vowels

From the literature review (see Table 1), /iai/ seems to be a good choice for reference vowel. However, two reasons exclude /iai/ as the possible reference vowel in this study. First, /iai/ can only appear after two consonants /ɕ/ and /t͡ɕ/ in the dialect,

whereas the other four rhymes have a much more varied environment, thus making the comparison very limited. Second, to our knowledge, young speakers of the dialect seldom pronounce the rhyme /iai/ now, because of the influence of standard Mandarin. /ɕiai/ is now more often pronounced as /ɕiɛ/ by young speakers, most likely to be influenced by its pronunciation in standard Mandarin (see Zhou, 2001).

For these reasons, we choose /i/ and /a/ as the reference vowel. [ta] will be the lowest vowel in the vowel space, whereas [ti] will mark the highest vowel height. We will then compare the height of /a/ in /an/ with the two reference vowels to see if it is raised in the experimental conditions. Details will be introduced in section 2.4.

2.3.2. Experimental tokens

In order to see if there are differences among phonological environments, we designed three environments. The first is /Can/ and /Cʰan/, which include both unaspirated and aspirated consonants, where /an/ is preceded by a consonant. This contrast is to test what He and Rao (2013) has reported. Namely the /an/ raising is more prominent when preceded by an aspirated consonant. The second is /Van/ which we termed "Diphthong". This is to test whether there is a difference in /an/ realization when it is preceded by a vowel, which corresponds to /ian/, /yan/ and /uan/ in Table 1. The last environment is the reference vowels. Apart from the highest point in the vowel chart /i/ and the lowest /a/, we also include /e/, /o/ and /u/ in order to plot the vowel chart of each participant.

The experimental tokens are presented in Table 3. We use three different characters for each syllable, indicated by the number after the syllable.

Condition	Environment					
			an	an	an	
	Unaspirated	p	pan1	pan2	pan3	
		t	tan1	tan2	tan3	
		k	kan1	kan2	kan3	
	Aspirated	pʰ	pʰan1	pʰan2	pʰan3	
		tʰ	tʰan1	tʰan2	tʰan3	
		kʰ	kʰan1	kʰan2	kʰan3	
EXPERIMENTAL						
			ian	uan	yan	
		0	ian1	uan1	yan1	
	Diphthong	0	ian2	uan2	yan2	
		0	ian3	uan3	yan3	
			a	i	u	e	o

REFERENCE	Reference	t	ta1	ti1	tu1	te1	to1
		t	ta2	ti2	tu2	te2	to2

Table 3: Experimental tokens in IPA

Altogether 37 experimental tokens are used in this study. In addition, in order to avoid interference between the experimental tokens, we added fillers ending with mid vowels /o/ or /oŋ/ between experiment tokens, since they are mid vowels which we expect will not interfere with the height of the other vowels we are examining. Altogether 70 tokens are recorded by each subject.

Figure 1: An example of annotation for /ian/.

2.4. Procedures
Now we describe the recording and analysis procedures.

2.4.1. Recording
Recordings are all done by the subjects reading a list of characters (in Chinese one character corresponds to one syllable and usually one morpheme). Recording apps on smart phones are used; these apps are capable of recording .wav format. All 17 recordings used in this study are in .wav format (i.e. mp3 recordings were deleted).

2.4.2. Labeling
Next, we use Praat (Boersma, 2002) to manually label the data. For "Unaspirated", "Aspirated" and "Reference" conditions the most stable part of the F1 is labeled. For the "Diphthong" condition (/ian/, /yan/ and /uan/), we labeled the highest point of F1 in the syllable. The reason is that /a/ is the lowest vowel in "Diphthong" condition (compared to /i/, /y/ and /u/) and should therefore have the highest F1 (see Figure 1).

Then we use a script to extract the F1 and F2 of the labeled points in all recordings.

2.4.3. Analyzing

Vowel charts are plotted using R (R Core Team, 2014).[1] They provide a direct visualization of the vowel spaces of the subjects, through which we can eyeball whether there is raising.

In addition, post-hoc Tukey tests on F1 are performed to determine whether there exists statistically significant differences between different age groups, genders and phonological environments. Then we will use the *diff* value to calculate the 'Height' of /an/ with reference to two reference vowels /i/ and /a/. That is, the F1 value of /i/ and /a/ will be the lower and upper bound of the F1 range; the F1 of /an/ of different groups will be plotted on to the scale and then normalized. For example, on a F1 scale of 250 Hz (/i/) - 900 Hz (/a/), if the F1 of /an/ is 700 Hz, the 'Height' of /an/ will be:

$$Height_{/an/} = \frac{F1_{/a/} - F1_{/an/}}{F1_{/a/} - F1_{/i/}} \times 100\% = \frac{900 - 750}{900 - 250} \times 100\% = 23.1\% \qquad (\star)$$

Thus we can compare the different 'Height' of /an/ in different groups.

2.5. Hypothesis

Based on the research questions and literature review, we have the following hypotheses.

H1: Degree of raising is more significant for young group than for old group:
- The realization of /a/ is lower for old group than for young group.
- The degree of raising of /a/ from reference vowel to [an], [uan], [ian], [yan] is larger for young group than for old group.

H2: Raising is more significant for female group than for male group:
- Within each age group, the degree of raising is larger in female speakers.
- Within each gender, young speakers still raise /an/ more than the old speakers.
- The difference in the degree of raising between young and old female participants is larger than that between two age groups of male speakers.

H3: Raising is influenced by phonological environment:
- The height of /a/ is higher in environments where there is an immediately preceding high vowel (i.e. Diphthong [uan], [ian] and [yan]).
- The height of /a/ is higher in aspirated environments (i.e. [pʰ], [tʰ], [kʰ]) than in unaspirated environments (i.e. [p], [t], [k]).
- The height of /a/ is different among [uan], [ian] and [yan]. Specifically, /a/ in [ian] and [yan] is higher than /a/ in [uan].

3. Results

In this section, we report both the statistical results and the descriptive analysis of our data.

---

[1] We would like to thank Phillip Weirich for the help with R code.

3.1. Statistical results
3.1.1.   Between age groups
     First we test whether it is true that the young age group exhibit more raising in all experimental environments. That is, "Unaspirated", "Aspirated" and "Diphthong" will together be collapsed under the *experimental condition* (i.e. "_an"), and it will be compared to the *reference vowel* /a/ (see Figure 2).
     A two way ANOVA with condition (experimental "_an", reference "a") and age (young, old) reveals a main effect of condition on the value of F1, F = 478.98, p<.001, and a main effect of age on the value of F1, F = 73.18, p<.001. It also reveals an interaction between condition and age, F = 4.48, p=.01. Post-hoc analysis using Tukey's HSD shows the following results.
     Between young and old group, the value of F1 of [a] in experimental condition for young age group is statistically lower than [a] of old group (diff =-101.31, p <.001). This indicates that the vowel height of /an/ in the younger age group is higher than that of the old age group, thus suggesting the young speakers are raising /an/ compared to their parents' generation, which is captured by the red box for condition "_an" condition in the middle of Figure 2.
     Within both the young and old age group, the value of F1 of [a] in experimental condition is significantly lower than the F1 of reference vowel [a] (diff =-104.56, p =.03). This suggests that in both age groups the height of /an/ is significantly higher than reference vowel /a/.
     Finally, there is no significant difference in the F1 value of Reference vowel [a] between two age groups (p =.98), thus indicating that the young and old speakers have a similar vowel height in the reference condition. This is shown in the similar length of the green boxes of both age groups in Figure 2.
     Using the equation (✶) and the mean value from Table 4, we can calculate the mean 'Height' of /an/ in both the young and old age group. The result is that the Height$_{/an/}$ for young age group is

$$\frac{926.2534 - 706.0808}{926.2543 - 302.7332} \times 100\% = 35.3\%$$

whereas the Height$_{/an/}$ for old age group is

$$\frac{911.9501 - 807.3867}{911.9501 - 344.5206} \times 100\% = 18.4\%$$

     This shows that the experimental /an/ in young age group is at about a third (35.3%) in height on the scale established by their two reference vowels /i/ and /a/. However, the experimental /an/ in old age group is at a position of 18.4% on the

reference scale between /i/ and /a/. This lends further support for our argument that the younger age group raises /an/ more than the old age group.
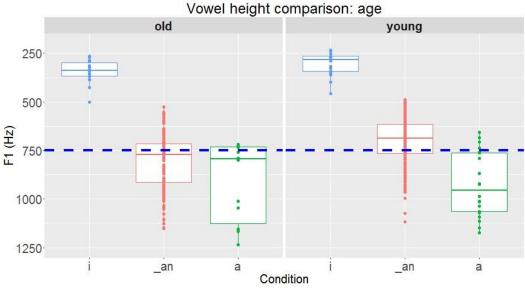

Figure 2: Old vs. young age group.

Condition	Age group	F1(Hz)
an	old	807.3867
an	young	706.0808
a	old	911.9501
a	young	926.2534
i	old	344.5206
i	young	302.7332

Table 4: Mean F1 of Experimental /an/, Reference /a/ and /i/

3.1.2. Between two genders

A three way ANOVA with condition (Experimental /_an/, Reference /a/) age (young, old) and gender (male, female) reveals a main effect of condition, age and a main effect of gender on the value of F1 of [a], $F = 414.27$, p<.001. It also reveals an interaction between condition and gender, $F = 15.99$, $p<.001$; and interaction between age and gender, $F = 24.19$, $p<.001$; but no interaction between condition, age and gender, $F = 0.67$, $p=.57$. These are shown in Figure 3. Post-hoc analysis using Tukey's HSD shows the following results.

Within each age group, the female speakers raise /an/ more than the male speakers. That is, in the **young** age group, the difference in F1 value of Experimental /an/ and Reference /a/ for female speakers is 295.00 (p=.000); the difference for the male speakers

is 75.97 (p=.001). This shows clearly that the female speakers in young age group raise much more than the male speakers.

In the **older** age group, the difference in F1 value of Experimental /an/ and Reference /a/ for female speakers is 206.34 (p=.000); the difference for the male speakers is 28.23 (p=.999). This shows that the female speakers also raise much more than male speakers in the old age group.

If we compare within each gender group, the young speakers still raise more than the old speakers. That is, within the female group, the degree of raising is 206.34 for old speakers but 295.00 for young speakers. Within the male group, the degree of raising is 28.34 for old speakers but 75.97 for young speakers.

3.1.3. Between phonological environments

A three-way ANOVA with gender (male, female), age (young, old) and environment (Diphthong, Aspirated, Unaspirated, Reference /a/) reveals a main effect of gender, a main effect of age and a main effect of environment on the value of F1 of [a], $F = 365.22$, $p<.001$. It also reveals an interaction between age and environment, $F = 3.802$, $p =.002$; an interaction effect between gender and environment, $F =10.51$, $p <.001$; but no interaction effect among age, gender and environment, $F =.49$, $p =.78$. Post-hoc analysis using Tukey's HSD indicates that: Within each gender and age group, the value of F1 of [a] in Diphthong, Aspirated and Unaspirated are not statistically different from each other ($p >.05$). These are also demonstrated in Figure 3.

Crucially, the fact that no significant difference between the aspirated and the unaspirated environments (the yellow and pink bars in Figure 3) is discovered does not support the analysis in He and Rao (2013) where aspirated environments are more prone to raising.

A one way ANOVA with high vowel-preceding environments (i.e. [uan], [yan], [ian]) and a post-hoc analysis using Tukey's HSD indicates that the F1 value of /a/ in [yan] is statistically significantly lower than [uan] ($F =-92.28$, $p <.02$). There is no statistically significantly difference between the F1 value of /a/ in [uan] and [ian] as well as in [yan] and [ian]. This is shown in Figure 4, and provides partial support for previous literature (see Table 1) to group [yan] and [ian] together, but exclude [uan].

3.2. Descriptive summary

Two representative vowel charts are shown in Figure 5. We see that in the representative old male speaker the reference /a/ is of the same height to experimental tokens of different phonological conditions. Sometimes it is even higher (e.g. compared with /uan/). However, the representative young female speaker demonstrates neatly the raising phenomenon, as her /_an/ is much closer to /e/ than to /a/. The remaining question is that whether it is possible that the old male speaker has raised /a/.
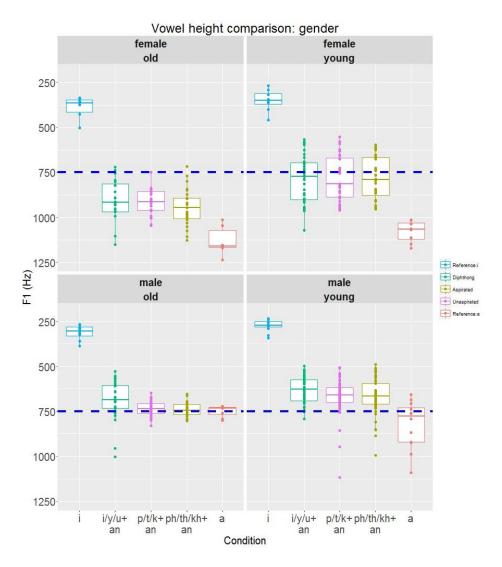
Figure 3: Comparison across age, gender and phonological environment.

When we look at Figure 6 where the vowel chart of the mean of two age groups are plotted one on top of another, it is more straight-forward that the young age group (green) raises the experimental /_an/ whereas the old age group (red) does not. The two reference vowels /i/ and /a/ are almost in the same position on the chart, suggesting that they are stable reference points for the study.

Another very intriguing point in Figure 6 is that for the old age group, their /ian/ and /yan/ are almost as high as green cluster of /an/ for young speakers, whereas their /uan/ has the lowest height. This seems to suggest that /ian/ and /yan/ are the first to raise and /uan/ is the last one to join the raising phenomenon. What Figure 6 shows is

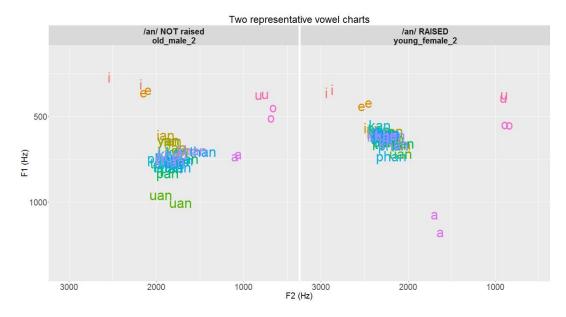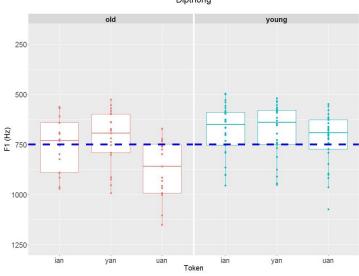particularly interesting as it demonstrates the possible order of raising for different conditions.

Figure 5: Two representative vowel chart.

Figure 4: Comparison between old and young age groups in i/y/u+an.

To sum up, our data first shows that raising is more significant for the young age group than the old group, supporting our first hypothesis H1. Specifically, the height of the vowel in experimental condition, i.e. /_an/ in the young age group is higher than that of the old age group. In addition, the degree of raising of from the reference vowel /a/ to experimental condition /_an/ is greater for the young age group.

Second, raising is more significant for the female group than for the male group within each age group, supporting H2. Specifically, the degree of raising of /a/ from reference condition to experimental condition is larger for female participants than for male participants in both young and old age groups. Also, the contrast between young and old age groups is still attested within each gender group. That is, the young female speakers still have more raising than the old female speakers. The same is true for male speakers.
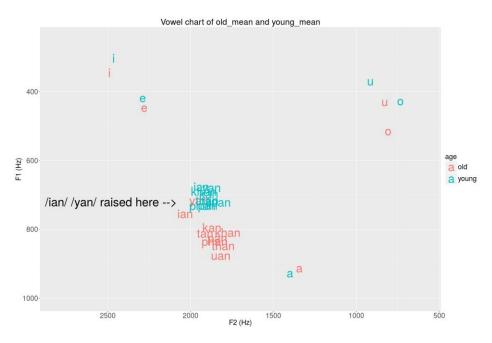

Figure 6: Combined vowel chart of the mean of two age groups.

Finally, raising is influenced by some, but not all phonological environment, supporting only part of our third hypothesis. There is no significant difference between the three environments: Diphthong, Aspirated and Unaspirated. Specifically, Aspirated environment does not have higher vowel height, providing no evidence for the aspirated unaspirated distinction in He and Rao (2013). In the Diphthong environment, the height of /a/ in /yan/ is significantly higher than /uan/, lending support to categorize /uan/ differently from /yan/ and /ian/.

4. Discussion

Our results confirmed the observation in previous literature (He & Rao, 2013) that young speakers tend to raise the vowel /a/ in /an/ and young *female* speakers are leading the change. This is in accordance with Labov's statement that women are usually the innovators in unconscious sound change (Labov, 1990, pp. 215-218). However, there are several interesting points that deserve our attention.

4.1. Individual variances in raising

First, old female speaker No.1, who is younger than her male counterparts in the old age group (49 yrs vs. ca 57 yrs), seems to show unexpected raising (see Figure 7). This indicates that the vowel change may not be an absolute phenomenon, though it is manifested mostly in young speakers. It may also be found in some speakers from older generation.
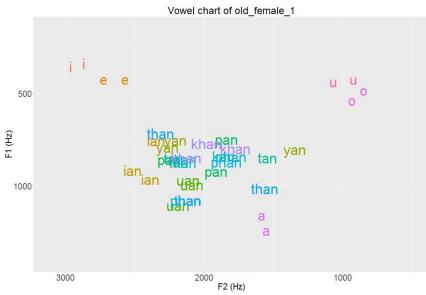

Figure 7: Some raising in old female speaker No.1.

The analysis of He and Rao (2013) shows that female speakers born after the 1980s have almost all raised /an/, whereas for male speakers only those born after the 1990s demonstrate the same raising effect. Our young male speakers are mostly born around 1990 so they can be said to be somewhat in between the "raising" and the "non-raising". Their data (as shown in Figure 3) is still significantly different from male speakers in the old age group, suggesting that most of them have raised /an/. The next step is to collect data from both middle-aged speakers and teenage speakers to determine where exactly the cut-off year may be.

At the same time, not all young speakers exhibit raising. Young female speaker No.1 for instance does not seem to raise /an/ at all (see Figure 8). This is much unexpected since young female speakers as innovators are more likely to show raising. It

494

is more surprising considering the fact that this speaker has lived in near-downtown Chengdu before the age of 19 where standard Chengdu dialect is spoken and where He and Rao (2013) did their study. Her parents are also born and raised in Chengdu. So she should be very representative of the vowel raising effect. But our data suggests otherwise. One possible reason might be that the speaker went to college in Beijing and lived there for altogether 7 years.

4.2. Possible path of /an/ raising

As shown in Figure 4, the vowel height in /uan/ is different from /ian/ and /yan/. That is, although /i/, /y/ and /u/ are all high vowels, the [front] feature actually influences the height of /a/ considerably (t-test shows that in the old age group, the difference in F1 between /yan/ and /uan/ is significant).
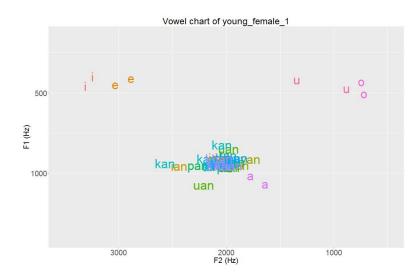

Figure 8: Unexpected: no raising for young female speaker No.1.

This result explains why in the previous literature summarized in Table 1, /uan/ never patterns with /ian/ or /yan/. In fact, Table 1 shows that /uan/ has long been categorized with /Can/. This is exactly the case for the old male speakers in this study, as shown in a more detailed vowel height plot of all the tokens (Figure 9).

Interestingly, Figure 6 suggests within the /Van/ condition, older speakers do raise /a/ in /ian/ and /yan/ (**ian**, **yan** is closer to **Xan**). We interpret this as /a/ first assimilates to its preceding front high vowels /i/ and /y/ for older speakers, which then becomes phonologized in younger speakers in all conditions, including /Cʰan/ and /Can/. This shows a possible trajectory of the ongoing sound change in Chengdu dialect.
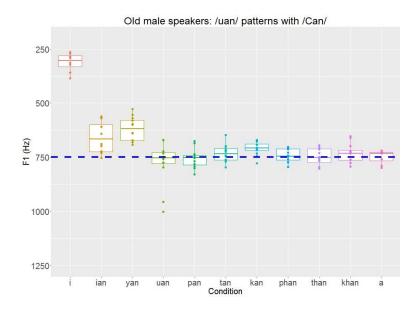
Figure 9: All tokens for old male speakers: /uan/ patterns with /Can/.

4.3. No distinction between aspirated & unaspirated

As suggested in He and Rao (2013), the aspirated environment is the first to be raised, according to their data obtained from speakers of 5 age groups. But the data in our study does not show any difference between aspirated and unaspirated environments, as shown in Figure 3 where the unaspirated (pink) and aspirated (yellow) are very close. If we zoom in and only look at these two environments (Figure 10 and 11), it is clearer that there is no systematic pattern between aspirated and unaspirated environments. That is, the F1 of both aspirated and unaspirated environments for old age group is right below the 750Hz line, whereas the F1 of both aspirated and unaspirated in young age group is just above the 750Hz line.

The only noticeable pattern seems to be the co-articulation effect of labial /p/, alveolar /t/ and velar /k/. That is, the height of /a/ seems to have the following order: /pan/ < /tan/ < /kan/ where < indicates 'lower than'. We still need to further explore the reasons for this phenomena.
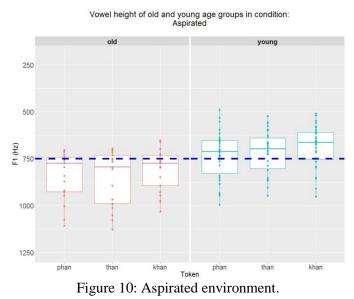
5. Conclusion and future work

To conclude, the raising of /a/ in young speakers of Chengdu dialect in /Van/, /Can/ and /Cʰan/ are all attested. It is also clear from our data that female speakers lead the change. However, there is no difference between aspirated and unaspirated conditions.

Another finding is that the /a/ in /ian/ and /yan/ possible undergo raising first because of vowel height assimilation, which then spreads to other conditions such as /Can/. This is a potential path for the vowel raising phenomenon in Chengdu dialect.

In the future, we first need to obtain better quality recordings using more professional recording devices. Second, to understand the effect of age on vowel raising in greater detail, we need subjects from a more diverse range of ages. Third, to test whether there is a path for vowel raising, we need to analyze data from more age groups and see if there is a pattern that shows a clear path for raising.
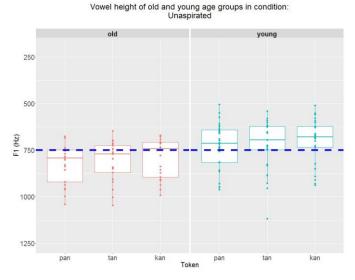

Figure 10: Aspirated environment.

Figure 11: Unaspirated environment.

REFERENCES

BOERSMA, PAUL. 2002. Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

CLOPPER, CYNTHIA G.; DAVID B. PISONI; and KENNETH DE JONG. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676.

DAILEY-O'CAIN, JENNIFER. 1997. Canadian raising in a midwestern US city. *Language Variation and Change*, *9*(01), 107–120.

HE, WAN AND RAO DONGMEI. 2013. *Sichuan chengduhua yinxi cihui diaocha yanjiu (A survey of Sichuan Chengdu dialect)*. Sichuan University Press.

LABONOV, BORIS M. 1971. Classification of russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608.

LABOV, WILLIAM. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change*, *2*(02), 205–254.

LABOV, WILLIAM. 1994. *Principles of language change: internal factors*. Oxford: Blackwell.

LABOV, WILLIAM. 1998. The three dialects of English. In M. D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39–81). Academic Press, San Diego.

LIANG, DEMAN. 1982. *Sichuan fangyan yu putonghua (Sichuan dialect and Mandarin)*. Sichuan People's Press.

R CORE TEAM. 2014. R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

WATSON, CATHERINE I.; MARGARET MACLAGAN; and JONATHAN HARRINGTON. 2000. Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change*, *12*(01), 51–68.

YANG, SHIFENG. 1984. *Sichuan fangyan diaocha baogao (A report on Sichuan Dialect)*. Institute of Linguistics Academia Sinica.

ZHEN, SHANGLING. 1958. Chengdu yuyin de chubu yanjiu (A preliminary phonological analysis of Chengdu dialect). *Journal of Sichuan University*, (1), 5–34.

ZHEN, SHANGLING. 1983. Sichuan fangyan de biweiyun (Nasal codas in Sichuan dialect). *Fangyan(Dialect)*, (4), 241–243.

ZHEN, SHANGLING; HAO, XIJIONG; and CHEN, SHAOLING. 1960. Sichuan fangyan yinxi (Phonlogy of Sichuan dialect). *Journal of Sichuan University*, *3*.

ZHOU, JIXU. 2001. Study of phonetic changes in Chengdu dialect in the 20th century– patterns under the influence of standard Mandarin. *Journal of Sichuan Normal University*, *28*(4), 47–59.

臺灣桃園市觀音區客家話的微觀空間分布

黃菊芳
國立中央大學

本研究運用 GIS 等技術建置語言空間分布的資料庫，主要關注不同的客家方言在臺灣的分布及所產生的相關變體。本研究選擇桃園市的觀音區做為空間分布研究的對象，主要原因在於桃園市是北臺灣客語使用人口最多的區域，也是四縣客家話和海陸客家話接觸最頻繁的區域。本研究採用語言微觀研究的家戶調查方式，該研究方法運用 GIS、GPS、衛星航空照片等先進科技，紀錄並繪製精確的語言分布地圖。觀音區靠海，是閩南語與客語接觸的鄉鎮，本研究透過家戶語言調查完成觀音區的語言空間分布，紀錄並繪製地圖分析閩南語與客語的空間分布界限，詳細紀錄客家話的地域變體地理分布。

1. 前言
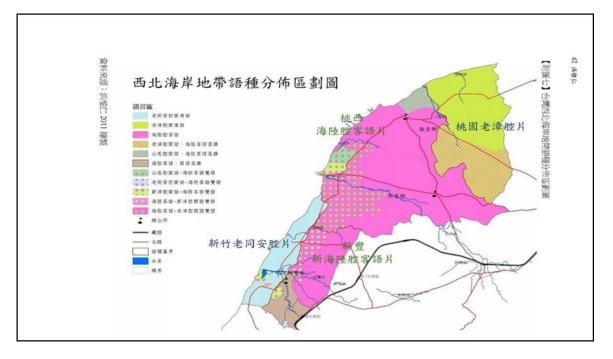
　　客家話在台灣的分布及所產生的相關變體是我們關注的焦點。本研究選擇桃園市的觀音區做為空間分布研究的對象，主要原因在於桃園市是全台灣客語人口最多的區域，另一方面則是因為我們過去已經進行苗栗縣的後龍鎮、南庄鄉、新竹縣的新豐鄉、新埔鎮及桃園縣新屋鄉的家戶語言分布調查。其中苗栗縣後龍鎮及南庄鄉的調查結果顯示，靠海的後龍鎮閩客雜居，而靠山的南庄鄉原客雜居，這兩個鄉鎮都分別有四縣客語及海陸客語的使用家庭。

　　以後龍鎮而言，海陸客語的使用人口居絕對弱勢，我們分析仍熟稔海陸客語的中老年人口，發現聲調仍保持海陸客語的特色，聲母及韻母則趨近四縣客語，以聲母而言，相對於新竹的海陸客語，少了[ʧ, ʧʰ, ʃ, ʒ]。有趣的是，南庄鄉的海陸客語使用人口居相對弱勢，我們分析仍熟稔海陸客語的中老年人口，發現聲調仍保持海陸客語的特色，聲母及韻母則趨近四縣客語，以聲母而言，相對於新竹的海陸客語，少了[ʒ]。

黃: 客家話

　　這個現象告訴我們進行中的語音變異似乎可以找到規律，不過必須與語言的空間分布進行詳細的論證，才能說明接觸的環境如何影響語音的變異，而變異的方向有沒有其規律。我們試圖延續這個議題，探討更多的客語方言接觸的鄉鎮，希望能夠利用語言微觀分布的角度解釋方言接觸後語音的演變機制。在這些基礎資料的建構與分析完成後，更希望進一步討論空間與時間的運作如何促使語音演變。

　　我們的研究從共時的角度出發，並且集中在客語，因為台灣是非常適合觀察方言接觸的地域，往往在一個小鄉鎮裡，便存在不同的客語方言。這些密切接觸的客語方言，經由姻親的密切往來，彼此溝通發展，語音的變異可以說隨處可見。如何將進行中的語音變異記錄並加以研究，是我們關心的議題。

　　關於桃園觀音的語言分布，較詳細的是 2011 年洪惟仁發表的〈臺灣西北岸閩客交界地帶的語言分佈〉一文，此文描寫新竹的新豐鄉、桃園的新屋鄉及觀音鄉的語言分布情形。該文對觀音的語言分布研究指出：「東北邊的草漯區是漳州腔閩南語區，西邊的觀音區是海陸客語區，東南部的新坡區除廣福村是純閩南語區以外都是閩客混雜區。」（2011：37）所繪語言分布如【圖 1】。

資料來源：洪惟仁 2011 繪製
【圖 1】 西北海岸地帶語種分佈區劃圖

　　該圖根據地方傳統分區將觀音區劃為三大部分：觀音區、草漯區及新坡區，並認為這三區的劃分和族群語言分布有相當程度的契合。（2011：37）該論文指出：「白玉村以黃姓為主；溝尾以許姓為主，源自廣東陸豐，都說海陸漳腔閩南語，和大牛欄方言屬於一個系統。本村原本以閩南語為優勢，但因海陸客語的強勢影響及客家人的遷入，現在已經變成閩客混雜區，所有人都是閩客雙語，甚至有部分閩南人反而變成「客鶴佬」（不會說閩南語只會說客語的閩南人）。」（2011：37）文中的「大牛欄方言」就是本研究的「大牛欄閩南語」，是聚居在新屋鄉西北的閩客混合語，原本鄰近的觀音區也有不少的黃姓居民使用大牛欄閩南語，只是「建立大潭電廠以後，黃姓大量遷出，只有少數人留下。大潭村黃姓居民多改說海陸客語，只有八九十歲的黃姓以大牛欄方言為主要語言。」（陳淑娟 2004：11）研讀前人的研究，桃園觀音的客家話似乎只有海陸腔，然而觀音區其實有不少的四縣客家話及饒平客家話，因此本研究將聚焦於客家話的分布，希望能進一步補充不同腔調客家話的空間分布。

　　本研究採用之研究方法是由中央研究院鄭錦全院士於 2004 年首先提出之語言微觀研究的家戶調查方式，該研究方法運用 GIS、GPS、衛星航空照片等先進科技，紀錄並繪製精確的語言分布地圖。本研究認為該研究方法可以有效的紀錄及描繪語言的空間分布，而且可以透過語言的空間分布進而探討語言接觸及變異的相關議題。本研究將持續建立最微觀的語言分布電子地圖，觀音區與我們過去調查的新屋區毗鄰，本研究將透過家戶語言調查完成觀音鄉的語言空間分布，紀錄並分析比較觀音鄉與新屋強勢語言及弱勢語言的語音異同。
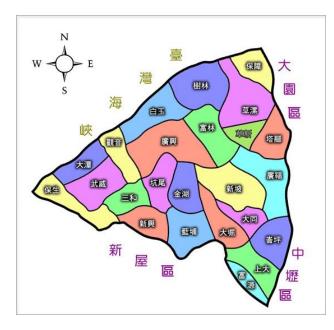
2. 桃園市觀音區的行政區劃與在地家族分布

　　桃園縣在中華民國 103 年 12 月 25 日正式改制升格為直轄市，改稱「桃園市」。觀音區是桃園市的沿海鄉鎮，東北接大園區、東南鄰中壢區，南邊是新屋區，住民以閩、客居多。觀音區的地理位置請參考【圖 2】。觀音區行政劃分為觀音里、白玉里、廣興里、大潭里、保生里、武威里、三和里、新興里、坑尾里、金湖里、藍埔里、大同里、大堀里、崙坪里、富源里、上大里、新坡里、廣福里、塔腳里、保障里、草漯里、樹林里、富林里、草新里（103 年 1 月 1 日起行政區域重新劃分新增），共計 24 個里。行政分區圖請參考【圖 3】。

黃: 客家話

【圖 2】 桃園縣市行政區劃圖

【圖 3】 桃園市觀音區行政分區示意圖

　　觀音區是北部最大工業區，而桃園科技園區及觀塘工業區也都正在開發，因此外來人口頗多，勢必在不久的將來對本地的語言產生巨大的影響。本區面積 87.79 平方公里，人口數有 6 萬餘人。（觀音區公所網頁資料）除了是工業重鎮，本地的農業及觀光業也很發達，每年的蓮花季都湧入不少的觀光人潮，而觀音區的埤塘數全桃園市最多，孕育出特有的埤塘文化。在地家族有黃姓、廖姓、李姓、徐姓、陳姓、楊姓、許姓、林姓、彭姓、吳姓、卓姓、謝姓、莊姓、古姓、歐姓、向姓、江姓、周姓、梁姓等。根據《觀音鄉志》的文字描述，這些在地家族的來台祖、祖籍地、遷台時間、子孫分布地等資料整理如【表 1】。（2014：402-421）

【表 1】觀音在地家族表

在地家族	來台祖	祖籍地	遷台時間	公元	子孫分布地	備註
江夏堂黃姓	黃鼎坤	廣東省惠州府陸豐縣寬塘回瑤	清雍正 6 年	1728	大潭村小飯壢、保生村、三和村	
江夏堂黃姓	黃鼎圳	廣東省惠州府陸豐縣寬塘回瑤			保生村	
江夏堂黃姓	黃鼎交	廣東省惠州府陸豐縣寬塘回瑤			新屋鄉永安村下庄子	
江夏堂黃姓	黃英飄	廣東省惠州府陸豐縣房長鄉	清乾隆年間		觀音莊白沙屯、坡寮后湖（今金湖村）	
江夏堂黃姓	黃英照	廣東省惠州府陸豐縣大安邨方塘下社	清雍正年間或乾隆初年		觀音莊白玉村白沙墩	
江夏堂黃姓	黃如泗	廣東省嘉應州鎮平縣白馬鄉新坊村寨背	清乾隆 21 年	1756	觀音莊苦練腳莊（富源村）、新坡村、廣福村	
武威堂廖姓	廖世崇	廣東省惠州府吉康都陸豐縣閭羅崆長崗背	清乾隆 19 年	1754	大潭庄塘背（今武威村）、觀音村、三和村、保生村	武威村 90%均為廖世崇公派下子孫
隴西堂李姓	李必正	漳州府詔安縣青龍山接應寺人			新坡張厝、富林村、過溪仔、草漯	
隴西堂李姓	李裕寧	福建省同安縣坑口鎮	清康熙、雍正年間		草漯村、保障村	

黃: 客家話

在地家族	來台祖	祖籍地	遷台時間	公元	子孫分布地	備註
東海堂徐姓	徐子貴	廣東省惠州府陸豐縣三溪鄉四旗田	清乾隆 17 年	1752	觀音村新坡、橫圳頂（現在三和村）	
東海堂徐姓	徐子堅	廣東省惠州府陸豐縣三溪鄉四旗田	清乾隆 17 年	1752	保生村對面屋、三和村	
東海堂徐姓	徐仕愷		渡台時間不詳		觀音村下埔頂	
東海堂徐姓	徐子鼎		渡台時間不詳		觀音村新坡下大路墘	
東海堂徐姓	徐永廷		渡台時間不詳		坑尾村藍埔段	
東海堂徐姓	徐常政	廣東省蕉嶺谷倉下	渡台時間不詳		白玉白沙屯、藍埔段	
東海堂徐姓	徐星桂	廣東省蕉嶺谷倉下	渡台時間不詳		白玉村 9 鄰下庄仔 19 號	
東海堂徐姓	徐相庚	廣東省蕉嶺北角鋪	清乾隆 40 年	1755	廣興村水溝尾莊、廣興村 1 鄰溝尾 31 號	
東海堂徐姓	徐錦文	廣東省蕉嶺北角鋪	清乾隆 40 年	1755	桃園龍潭、白玉村 12 鄰下庄仔 49 號	
東海堂徐姓	徐俊美	廣東省蕉嶺谷倉下	渡台時間不詳		廣興村 8 鄰溝尾 21、23 號	
東海堂徐姓	徐啟德	廣東省惠州府陸豐縣南阿甲龍盤石下	渡台時間不詳		三和村三座屋 29 號	
東海堂徐姓	徐振連	廣東省惠州府陸豐縣南阿甲龍盤石下	渡台時間不詳		廣興村、白玉村、坑尾村	
穎川堂陳姓	陳樵	福建省漳州府南靖縣習賢里碇總三平堡永豐湯兜社山邊厝小地號睡眠地牛穴	清 乾 隆 20、21 年	1755、1756	蘆洲、大同村、大堀一帶（大崙、山上、山下、崙坪、新坡、青埔子、上大堀、下大堀等地）	

黃: 客家話

在地家族	來台祖	祖籍地	遷台時間	公元	子孫分布地	備註
弘農堂楊姓	楊寬柔、楊敦朴、楊秉正	福建省漳州府漳浦縣浮南橋人	清雍正、乾隆年間		樹林村一帶	
高陽堂許姓	許忠良	福建省漳州府南靖縣習賢里碇總黃坑深區社	清乾隆年間		廣福村一帶	
高陽堂許姓	許勝光	廣東省陸豐縣大安鄉官塘壠	清嘉慶年間		石觀音坡寮水尾子、金湖村一帶	
高陽堂許姓	許勝元	廣東省陸豐縣大安鄉官塘壠許厝鄉	清嘉慶年間		金湖村	
高陽堂許姓	許錦哲	廣東省陸豐縣大安鄉官塘壠許厝鄉	清乾隆年間	1736-1795	廣興村 6 鄰溝尾 16 號	
西河堂林姓	林君海	廣東省陸豐縣黃塘油車街烏石腳	清乾隆 21 年	1756	坡寮後湖	
西河堂林姓	林萬成		清雍正 9 年	1731	廣興村坑背、溝尾	
隴西堂彭姓	彭茂松	廣東省惠州府陸豐縣大安墟吉康郡五雲洞歐坑鄉	清雍正 8 年	1730	大潭村塘尾、新屋鄉社子村	彭氏族人由於居住海邊，除農耕外，也擅長岸上漁產捕撈。
草漯吳姓	吳振利				草漯	觀音地區大地主
西河堂卓姓	卓乃文	廣東省潮州府海陽縣汶路鄉東鳳鎮文路卓村	清乾隆 11 年	1746	上大村	

黃: 客家話

在地家族	來台祖	祖籍地	遷台時間	公元	子孫分布地	備註
寶村堂（陳留堂）謝姓	謝永錦	廣東省嘉應州府長樂縣（現五華縣）石船牛角墟	清乾隆 8 年	1743	保生村	謝永錦本人並未渡台，其夫人陳氏帶 3 子來台。
下大堀莊姓	不詳	廣東省惠州府陸豐縣			下大堀、大堀後庄子	
坑尾古家	古培坤	廣東省嘉應州長樂縣貯河石下（今廣東省五華縣）			新竹州楊梅庄、坑尾村 7 鄰 42 號	
千乘堂倪姓	倪朴直	福建省漳州府狗尾山	不詳		富林村	
平陽、河口堂歐姓	歐子深、歐元悅	廣東省惠州府陸豐縣歐田鄉（舊惠州府吉康郡）	清雍正 4 年	1726	白沙墩（現在的白玉村歐屋），即白玉村（平陽堂）、新興村（河口堂）、新鄉永安	
河南堂向姓	向振國	廣東省惠州府陸豐縣方閣都大安墩還珠寨	清康熙 36 年	1697	消飯阬莊（即小飯壢坑，今大潭村 9 至 13 鄰一帶）	
濟陽堂江姓	江琪臻	福建省汀洲府永定縣芹菜白石后	清雍正 5 年	1722	新北市新莊十八份、新竹芎林、金湖、新坡、廣福、崙坪、富源	
汝南堂周氏	周仕衛	廣東省惠州府海豐縣赦寮庄大德山中田墩			保生村對面厝 6 鄰一帶、新屋鄉笨港村、楊梅鎮富岡里、新竹縣峨眉鄉石井村	
渤海堂吳氏	吳上崑	廣東省惠州府陸豐縣方角都葫蘆崒（即陸豐縣八萬鎮上葫管區）			保生村、樹林村、新竹縣湖口鄉	
觀音梁姓	梁瑟璋	廣東省惠州府陸豐縣吉康都竹坑鄉	清康熙 50 年	1711	草漯	

資料來源：作者整理

黃: 客家話

　　根據上表，整理觀音區在地家族祖籍地統計圖如【圖 4】，來自廣東省惠州府的陸豐縣的人口最多，其次是廣東省嘉應州府蕉嶺縣、長樂縣，福建省以漳州府南靖縣、詔安縣、漳浦縣為主，泉州府則以同安縣為主，來台時間約在清朝雍正、乾隆、嘉慶年間。

觀音區在地家族祖籍統計圖

【圖 4】 觀音區在地家族祖籍統計圖

3. 桃園市觀音區的語言分布

　　本研究透過田野調查，進行觀音區的語言分布地理資訊系統建置，調查對象為世居本地的居民，運用 GIS 詳細描繪觀音區的家戶語言空間分布，本研究排除工業區成立後新搬來的住戶，共計調查 14,841 戶的家戶語言使用情形，其中閩南語使用戶數有 8,248 戶，大牛欄閩南語 5 戶，海陸客語使用戶數有 5,018 戶，四縣客語 928 戶，饒平客語 95 戶，豐順客語 16 戶，詔安客語 2 戶，其他 530 戶（主要是指隨國民政府來台講國語的族群）。【表 2】是本研究整理的觀音區分里語言統計表。
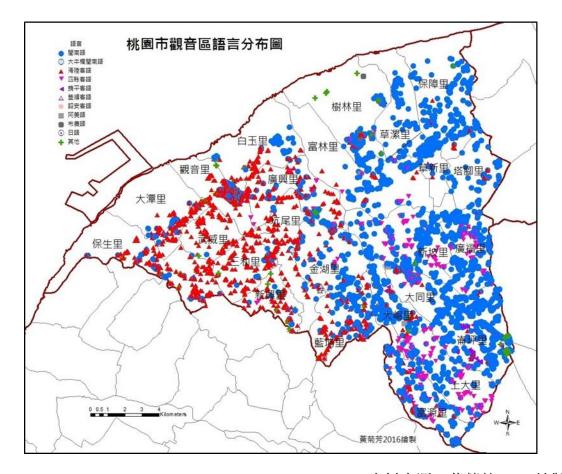
黃: 客家話

【表2】 觀音區分里語言統計表

單位：戶

語言 里名	海陸 客語	四縣 客語	饒平 客語	豐順 客語	詔安 客語	閩南語	大牛欄 閩南語	其他
新興里	213	10				20		17
武威里	339					2		3
金湖里	144	4				175		14
三和里	296					15		7
保生里	684					102	1	11
坑尾里	389	5	6			11	4	
藍埔里	393			16		485		5
觀音里	975	5				87		24
大潭里	586					67		6
新坡里		202				862		5
大崛里	65		7			412		14
大同里	65	115				271		28
廣興里	270	26				86		28
白玉里	242					195		26
富源里	145	198				832		1
富林里	16	1				135		
草新里	7	1				94		
廣福里	31	108				685		2
上大里	0	227				213		2
草漯里	11	18				569		
保障里	24					731		5
塔腳里	40					347		
樹林里	54	7				463		52
崙坪里	29	1	82		2	1389		279
總計	5018	928	95	16	2	8248	5	529

資料來源：作者整理

　　雖然本地外來人口較多，無法確認每一戶的語言使用，不過居民對外的日常生活語言以國語為主，本地的優勢語言有閩南語及海陸客語，因此大部分的住戶都是多語使用者。四縣客語集中在新坡里、大同里、富源里、廣福里、上大里，其中

上大里和新坡里完全沒有海陸客語的住戶。饒平客語集中分布在崙坪里，根據調查過程的瞭解，許多家中原本使用饒平客語的住戶已不講饒平客語，轉而使用當地的優勢語言。豐順客語集中於藍埔里的高姓，有 16 戶。

　　根據本研究建置的資料庫，繪製觀音區語言分布地圖如【圖 5】。圖中以深藍色圓形代表閩南語，淺藍色黑邊圓形代表大牛欄閩南語，紅色正三角形代表海陸客語，桃紅色倒三角形代表四縣客語，紫色 90 度翻轉三角形代表饒平客語，淺紫黑邊正三角形代表豐順客語，淺膚色五邊形代表詔安客語，淺灰正方形代表阿美語，深灰圓邊正方形代表布農語，綠色十字代表其他。這張圖顯示，觀音區可以從中切半，東片優勢語言為閩南語，夾雜四縣客語及饒平客語，西片的優勢語言為海陸客語。本地的海陸客語與新屋區的海陸客語聲韻調都相同，小稱詞尾也和新屋區的海陸客語相同，主要以變調處理，不同於新竹海陸的[ə]。四縣客語有[ʧ, ʧʰ, ʃ, ʒ]，但與新竹新埔及桃園新屋的四縣客語後接的韻母不盡相同，觀察不到規律。
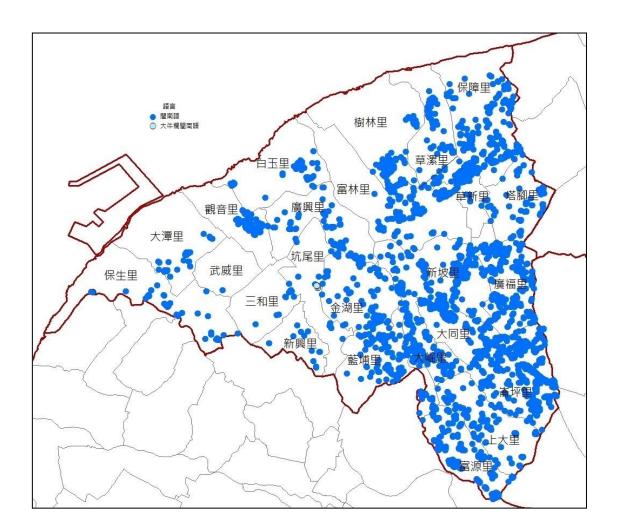

資料來源：黃菊芳 2016 繪製

【圖 5】 觀音區語言分布地圖

黃: 客家話

4. 桃園市觀音區的閩南語空間分布

　　觀音區的閩南語有偏漳腔閩南語，也就是一般的通行腔閩南語，本文以閩南語稱之，較集中分布在藍埔里、新坡里、大崛里、大同里、白玉里、富源里、草新里、廣福里、上大里、草漯里、保障里、塔腳里、樹林里、崙坪里。此外還有大牛欄閩南語，分布在坑尾里和保生里，戶數很少。【圖 6】是桃園市觀音區閩南語的分布簡圖。

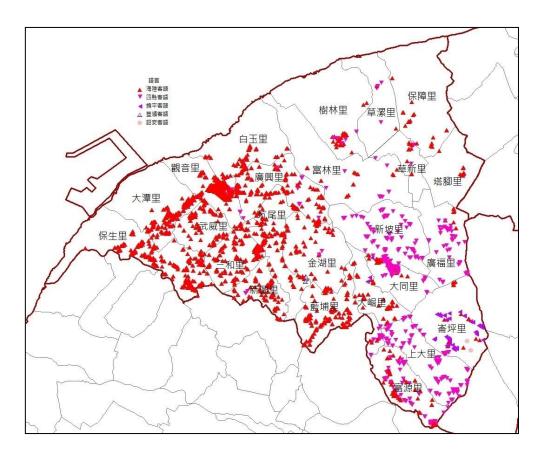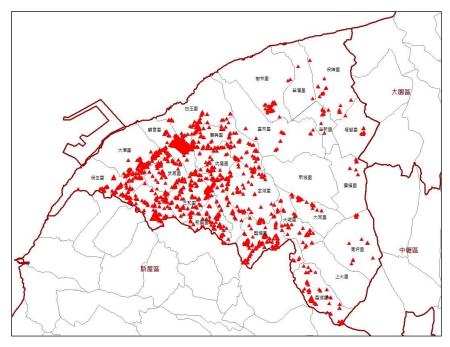【圖 6】　觀音區閩南語分布簡圖

黃: 客家話

5. 桃園市觀音區的客家話空間分布

　　觀音區的客家話有海陸腔、四縣腔、饒平腔、豐順腔及詔安腔，【圖 7】是觀音區客家話的分布簡圖。海陸客家話是本地的優勢客家話，分布在全觀音區，只有上大里和新坡里沒有，【圖 8】是桃園市觀音區海陸客家話的分布簡圖。四縣客家話集中分布在上大里和新坡里，此外還分布在大同里、富源里、廣福里等，【圖 9】是桃園市觀音區四縣客家話的分布簡圖。饒平客家話分布在崙坪里、藍埔里、大崛里，豐順客家話集中在藍埔里，詔安客語僅兩戶，位於崙坪里，【圖 10】是桃園市觀音區饒平、豐順、詔安客家話的分布簡圖。
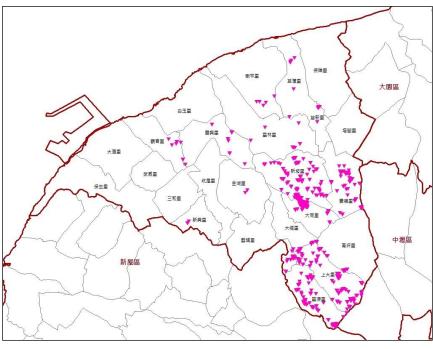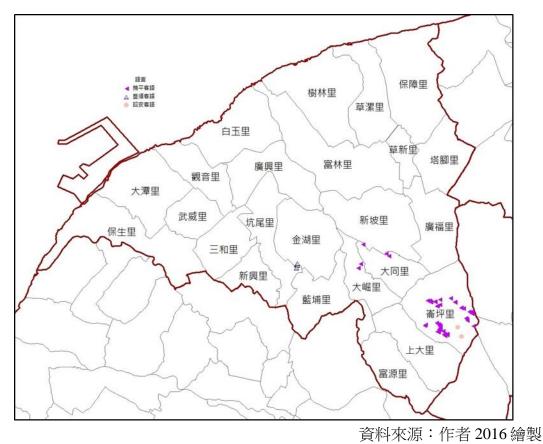

資料來源：作者 2016 繪製

【圖 7】 觀音區客家話分布簡圖

黃: 客家話

【圖 8】 觀音區海陸客家話分布簡圖

【圖 9】 觀音區四縣客家話分布簡圖

【圖 10】 觀音區饒平、豐順、詔安客家話分布簡圖

6. 觀音區的四縣客家話

　　本地的海陸客家話與新屋區的海陸客家話聲韻調都相同，名詞後綴也和新屋區的海陸客語相同，主要以變調處理，不同於新竹海陸的[ə]和苗栗四縣的[e]。四縣客家話處於以閩南語及海陸客家話為優勢的環境下，雖然聚居在一起，仍不免受優勢語言的影響而產生變化，因此四縣客家話變異的情形較為複雜。簡單比較目前已調查的新竹新埔、桃園新屋以及本文的桃園觀音這三個鄉鎮區居弱勢的四縣客家話，可以從聲母、韻母、聲調及名詞後綴進行討論。

　　首先是聲母，如果把新埔、新屋和觀音這三個地方的四縣客家話拿來與苗栗主流的四縣客家話比較，最大的特色是舌葉音[ʧ, ʧʰ, ʃ, ʒ]的增生，如表 3 所示。不過增生的詞彙比例不同，新埔多於新屋，新屋又多於觀音。

【表 3】 四縣客家話聲母比較表

鄉鎮＼聲母	p	pʰ	m	f	v	ts	tsʰ	s	t	tʰ	n	l	tʃ	tʃʰ	ʃ	ʒ	ȵ	k	kʰ	ŋ	ø	h
新埔四縣	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
新屋四縣	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
觀音四縣	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
苗栗四縣	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
新竹海陸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

資料來源：作者整理

除了舌葉音的增生之外，各地客家話[ts, tsʰ, s]前接介音[i]都有顎化為[tɕ, tɕʰ, ɕ] 的現象，例如「酒」[tsiu³¹]> [tɕiu³¹]，不過新埔四縣客家話沒有出現這個現象。表 4 是[ts, tsʰ, s]顎化分布表。

【表 4】 四縣客家話[ts, tsʰ, s]顎化分布表

聲母＼變體	新埔四縣	新屋四縣	觀音四縣	苗栗四縣	新竹海陸	備註
ts, tsʰ, s	＋	＋	＋	＋	＋	
tɕ, tɕʰ, ɕ	（－）	＋	＋	＋	－（＋）	與ts, tsʰ, s互補
tʃ, tʃʰ, ʃ	＋	＋	＋	－	＋	

資料來源：作者整理

新埔、新屋及觀音的四縣客家話[ts, tsʰ, s]改讀海陸客家話[tʃ, tʃ, ʃ]發音全同的 字例如表 5，此表顯示，[u]韻和[oŋ]韻最容易改讀[tʃ, tʃ, ʃ]。

韻母＼聲母	ɿ	a	u	ui		
tʃ	紙	遮	豬煮珠祖阻主蛀注註			
tʃʰ		車	住鼠處助訴初			
ʃ	屎時市	蛇	樹書疏薯蜍素守數手	水		
韻母＼聲母	in	on	un	aŋ	oŋ	uŋ
tʃ	珍	轉鑽	準		張裝章樟彰	中鍾終
tʃʰ		賺串		鄭	昌丈滄床長腸廠創常	蟲重
ʃ			唇損順		商傷上賞	
韻母＼聲母	ɿp	ɿt	ak	uk		
tʃ		職	燭祝竹			
tʃʰ						
ʃ	濕十	實失	石	叔熟		

【表 5】 四縣客家話[ts tsʰ, s]改讀海陸客家話[tʃ, tʃʰ, ʃ]發音全同字例表

資料來源：作者整理

新埔、新屋、觀音雖然都增生[ʒ]聲母，不過增生的詞彙數有異，新埔四縣大部分的韻母都增生[ʒ]聲母，新屋四縣及觀音四縣增生[ʒ]聲母的韻母數較少，表6是四縣客家話[ʒ]聲母增生詞彙比較表，此表顯示，新埔[ʒ]聲母增生的詞彙數多於新屋，而新屋又多於觀音。本文認為這個現象與海陸客家話的密切接觸相關，新埔的海陸客家話使用人口居絕對優勢，新屋的海陸客家話使用人口數也高居七成以上，觀音的海陸客家話使用人口則相對較少，本地的閩南語使用人口較高。

【表 6】 四縣客家話[ʒ]聲母增生詞彙比較表

變體＼例字	醫	員	央	容	野	襜	烟	翼	越	邀
新埔四縣	ʒi	ʒien	ʒioŋ	ʒiuŋ	ʒia	ʒiam	ʒian	ʒit	ʒiet	iau
新屋四縣	ʒi	ʒien	ʒioŋ	iuŋ	ia	iam	ian	it	iet	iau
觀音四縣	ʒi	ien	ioŋ	iuŋ	ia	iam	ian	it	iet	iau
苗栗四縣	i	ien	ioŋ	iuŋ	ia	iam	ian	it	iet	iau

黃: 客家話

　　韻母的部分，新竹新埔及桃園新屋、觀音的四縣客家話韻母有些微差異。新埔四縣的韻母與新竹海陸趨同，新屋和觀音四縣的韻母保留較多苗栗四縣的發音習慣。表 7 是四縣客家話韻母比較表，其中與唇音結合的[i]已經改讀海陸客家話的[ui]。

【表 7】 四縣客家話[ʒ]聲母增生詞彙比較表

例字 \ 變體	新埔四縣	新屋四縣	觀音四縣	新竹海陸	苗栗四縣
試始	i	ɿ	ɿ	i	ɿ
針深審	im	ɿm	im	im	ɿm
汁	ip	ɿp	ɿp	ip	ɿp
直值食	it	ɿt	ɿt	it	ɿt
街雞	ai	e	e	ai	e
趙照少	au	eu	eu	au	eu
廟笑	iau	eu	eu	iau	eu
齊蕭小	iau	iau	eu	iau	eu
書臭	iu	u	u	iu	u
真陳稱	in	ɿn	ɿn	in	ɿn
味會買	ui	ui	ui	ui	i

　　四縣客家話弱勢變體的聲調目前保持在相對穩定的狀態，表 8 是四縣客家話的聲調比較表。

【表 8】 四縣客家話聲調比較表

例字 \ 調類 \ 變體	陰平 包鮮	陽平 銅籃	上聲 飽火	陰去 報線	陽去 大換	陰入 百節	陽入 白合
新埔四縣調值	24	11	31	55		32	55
新屋四縣調值	24	11	31	55		32	55
觀音四縣調值	24	11	31	55		32	55
苗栗四縣調值	24	11	31	55		32	55
新竹海陸調值	53	55	24	11	33	55	32

資料來源：作者整理

桃園新屋及觀音的四縣客家話是當地的弱勢語言，但是在名詞後綴的表現上，仍全部與苗栗四縣客家話相同，都使用[e]，例如：「車子」讀[$ʧʰa^{24}\ e^{31}$]。只有新竹新埔的四縣客家話用[ə]，例如：「車子」讀[$ʧʰa^{24}\ ə^{31}$]，但其他也都讀[e]，例如「蚊子」讀[$mun^{24}\ e^{31}$]、「扇子」讀[$san^{55}\ e^{31}$]、「鳥」讀[$tiau^{24}\ e^{31}$]等。

7. 結論

桃園市觀音區是工業區，外來人口頗多，本研究主要調查世居本地的住戶，以戶為單位進行語言的空間分布調查。調查結果顯示，閩南語是本地最多人口使用的地方語言，海陸客家話居次，四縣客家話再其次。有意思的是，本區東片以閩南語為優勢語言，西片以海陸客家話為優勢語言，而四縣客家話集中在新坡里及上大里，這兩個里都沒有海陸客家話的住戶，饒平客家話集中在以閩南語為優勢語的崙坪里，會說豐順客家話的是藍埔里的高姓家族共有 16 戶，四縣、饒平和豐順客家話都有群聚現象，尚能保持語言的活力。

本研究著眼於共時分布的描寫，經由時間與空間座標將調查的資料進行定位，除了提供精確的第一手資料於研究及教學，更可以進一步將共時的調查資料拿來做歷時的語言變化推論根據。當然值得研究的是，這些源於移民帶來的不同客語在臺灣生根發芽，彼此接觸，所產生的變異與發展如何。目前的研究成果顯示，在語音方面，四縣與海陸的接觸影響與當地優勢腔呈現正相關，單就語音而言，變異出現在聲母及韻母，聲調則保持在相對穩定的狀態。本地的海陸客語與新屋區的海陸客語聲韻調都相同，名詞後綴也和新屋區的海陸客語相同，主要以變調處理，不同於新竹海陸的[ə]。四縣客語有[ʧ, ʧʰ, ʃ, ʒ]，但與新竹新埔及桃園新屋的四縣客語後接的韻母不盡相同，詞彙的增生數以新竹新埔最多，桃園新屋其次，桃園觀音最少。

參考文獻

岩田禮編，2009，《漢語方言解釋地圖》。[日本]東京都：白帝社。

洪惟仁，2011，〈台灣西北海岸的語言分佈與閩客互動〉，《臺灣語文研究》第 6 卷第 2 期：29-64。台北：萬卷樓圖書股份有限公司。

洪惟仁，2003，〈桃園大牛欄方言的形成與發展：發祥地的追溯與語言層次、共時演變的分析〉，《臺灣語文研究》第 1 卷第 1 期：25-67。台北：萬卷樓圖書股份有限公司。

洪惟仁，1992，《台灣方言之旅》。台北：前衛出版社。

曹志耘主編，2008，《漢語方言地圖集》。北京：商務印書館。

黃菊芳、陳秀琪、江敏華、鄭錦全，2013B，〈台灣南庄海陸客家話舌葉音的變異與消失〉，《客家研究》第 6 卷第 2 期：129-66，桃園：國立中央大學。

黃菊芳、江敏華、鄭錦全，2013A，〈台灣新埔四縣客語舌葉音的產生〉，《語言學論叢》第 48 輯：140-166，北京：北京大學中國語言學研究中心。

黃菊芳、江敏華、鄭錦全，2012B，〈後龍海陸客家話的語音變異〉，《臺灣語文研究》第 7 卷第 1 期：129-150。台北：萬卷樓圖書股份有限公司。

黃菊芳、江敏華、鄭錦全，2012A，〈台灣桃園縣新屋鄉客語的地理分布微觀〉，THE 24TH NORTH AMERICAN CONFERENCE ON CHINESE LINGUISTICS--CELEBRATING THE 120TH ANNIVERSARY OF THE BIRTH OF YUEN REN CHAO ENGLISH（第 24 屆北美漢語語言學會議–紀念趙元任先生誕辰 120 周年）。UNIVERSITY OF SAN FRANCISCO（舊金山：舊金山大學）。

黃菊芳、蔡素娟、鄭錦全，2012，〈台灣雲林縣崙背鄉客家話分布微觀〉，《語言暨語言學》專刊《語言時空變異微觀》，95-123，2012-1-049-005-000332-1。台北：中央研究院語言學研究所。

鄭錦全，2004，〈語言與資訊：釐清台灣地名厝屋〉，羅鳳珠編《語言文學與資訊》1-24。新竹：國立清華大學出版社。

尹章義總編纂，2014，《觀音鄉志》（下冊）。桃園縣：桃園縣觀音鄉公所。

陳淑娟，2004，《桃園大牛欄方言的語音變化與語言轉移》。臺北市：臺灣大學。

「桃園郵局」，<HTTPS://WWW.POST.GOV.TW/POST/INTERNET/Q_LOCALPOST/INDEX.JSP?ID=12130201>，查詢日期：2016 年 11 月 22 日。

「桃園市觀音區公所」，<HTTP://WWW.GUANYIN.TYCG.GOV.TW/HOME.JSP?ID=101&PARENTPATH=0,11,60>，查詢日期：2016 年 11 月 22 日。

ESRI.（ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE）2010. ARCGIS 10 SOFTWARE.

WANG, WILLIAM S-Y. 1968. Vowel features, paired variables, and the English vowel shift. *Language* 44.4:695-708.

Strong and Weak Personal Pronouns in Tunxi Hui Chinese

LU Wen

The University of Hong Kong

Tunxi Hui, a very little studied variety of the Hui Chinese, demonstrates a dichotomy of strong and weak personal pronouns. The strong pronouns are formed by suffixing –*le* to the corresponding weak forms. The weak pronouns, furthermore, can be divided into uncliticized ones as proclitics and cliticized alternatives as endoclitics. Phonologically, the strong forms are higher in pitch, longer in duration and clearer in timbre. Syntactically, only the strong forms can occur in isolation or clause final positions, and they are preferred in information discourse focus positions like answers, contrast or cleft sentences. The uncliticized and cliticized weak forms differ from each other in that the former are proclitics while the latter are endoclitics. In terms of distribution, therefore, the uncliticized weak forms are commonly seen in subject positions whereas the cliticized weak forms typically appear in pivotal constructions or double object constructions.

1. Background of Tunxi Hui

Tunxi Hui is a very little-studied Hui variety of Chinese, one of the ten 'dialect' groups of Sinitic languages, which has only around 70,000 speakers. Furthermore, the Hui group of Chinese can be divided into five subgroups, and Tunxi Hui belongs to the Xiu-yi subgroup, according to the second edition of the *Linguistic Atlas of Chinese Dialects* (2008). As an under-studied group of Sinitic languages, Hui Chinese is the latest 'dialect' group to be established among all Sinitic languages. Hui Chinese also demonstrate exceptionally substantial internal diversity, as has been recorded in the Ming Dynasty text *Huizhoufu zhi* 徽州府志 'A history of the Huizhou Provincial Capital' that

六邑之語不能相通[1]. This high level of internal unintelligibility has also been confirmed by the late renowned linguists Luo Changpei and Chao Yuen-ren during their pioneering field work on Hui Chinese in 1940s and 1960s.

In terms of typology, Tunxi Hui is a 'transitional' type of Sinitic languages, which showcases intermediate features between the Northern Chinese like Mandarin and the Southern Sinitic languages such as Min dialects and Yue dialects, in line with Hashimoto (1976, 1986), and later Norman (1988) and Chappell (2016)'s distinction of the North-South divide of Sinitic languages. This North-South division is based on a set of phonological, morphological and syntactic features of representative languages from northern and southern groups of Chinese languages, including number of tones, morpheme per word ratio, syllable structure, the inventory of classifiers, relative linear order of modifier and modifiee, the relative order of adverbs and predicate, sequence of comparative constructions, source of markers of passive constructions, relative linear order of direct and indirect objects in double object constructions, etc.

As a transitional Sinitic language, Tunxi Hui has six tones, more than the four tones in Mandarin Chinese but fewer than the nine tones in Cantonese. Besides, Tunxi Hui possesses a merged final stop *ʔ,* simpler than the full set of final stops *p*, *t*, *k* as preserved in Cantonese but more complex than Mandarin which has no final stops.

Morphologically, the ration of monosyllabic words in Tunxi Hui is also higher than Mandarin Chinese. For animal names, Tunxi Hui exhibits head-initial tendency like many Southern Sinitic varieties, with animal names preceding their gender.

Regarding syntax, the source of pretransitive markers in Tunxi Hui varies from either Mandarin or Cantonese, which is grammaticalized from the HELP/GIVE verbs. Topicalization in Tunxi Hui also stands out as the highest in frequency among all types of Sinitic languages. Other syntactic constructions such as passive constructions, comparative constructions and adverbial constructions in Tunxi Hui fluctuate between the Northern Mandarin and the Southern Sintic languages in that it resembles Mandarin sometimes, and will take after Southern Sinitic languages such as Cantonese at other times.

---

[1] In English it means 'The languages of the six counties under the Huizhou Provincial Capital are mutually intelligible.'

To sum up, the major typological features of Tunxi Hui as a transitional Sinitic language are shown below:

Type of Chinese	**Northern Type**	**Transitional type**	**Southern Type**
Example	Mandarin	Tunxi Hui	Hong Kong Yue
Tones	4 tones	6 tones	9 tones
Phonology	no final stops	merged final stop ʔ	full set of final stops p, t, k
Morphology	more disyllabic words, agglutinating tendency	more monosyllabic words, isolating tendency	more monosyllabicwords, isolating tendency
Inventory of classifiers	relatively small	relatively big	relatively big
Animal names	head final	head initial	head initial
Reduplicative pattern	head final	head final	head initial
Pretransitive construction	pretransitive marker grammaticalized from HOLD/TAKE verb	Pretransitive marker grammaticalized from GIVE/HELP verbs	SVCs with TAKE verbs
Passive construction	the SUFFER type	the GIVE type	the GIVE type
Comparative construction	head final	head final	head initial
Adverbial Construction	head final tendency	head initial tendency	head initial tendency
Topicalization	frequent	very frequent	less frequent

Diagram 1 Tunxi Hui as a Transitional Sinitic Language

After a brief overview of the typology, we will focus on the strong and weak personal pronouns in Tunxi Hui.

2. The strong and weak personal pronouns in Tunxi Hui

In the literature, the majority of Sinitic languages are understood to have only one simple paradigm of personal pronouns, except for a few Wu varieties such as Fuyang Wu, which possesses both 'simple singular pronouns' and 'complex singular pronouns' as stated in Li (2015: 226-247). In this regard, Tunxi Hui is exceptionally distinctive in that it has two paradigms of personal pronouns, with both strong and weak forms of singular personal pronouns, as many Indo-European languages do. They are shown below:

Number	Singular			Plural
Form	Weak Form		Strong Form	
	cliticized	uncliticized		
1st person	a^{23}	a^{24}	a^{24}-le	a^{24}-ian^{44}
2nd person	$\underset{.}{n}^{0}$	$\underset{.}{n}^{44}$	$\underset{.}{n}^{44}$-le	$\underset{.}{n}^{44}$-ian^{44}
3rd person	$k^{h}\partial^{0}$	$k^{h}\partial^{44}$	$k^{h}\partial^{44}$-le	$k^{h}\partial^{44}$-ian^{44}

Diagram 2 Two Paradigms of Pronominal System in Tunxi Hui: Weak and Strong

As can be seen from the above table, the plural personal pronouns are formed by adding a plural suffix *-ian*^{44} 'person' 人[2]. Besides, a strong and weak distinction is made between the singular personal pronouns. Specifically, the strong forms are formed by suffixing *–le*[3] to the uncliticized weak form of the same singular personal pronouns. It can be reasonably suggested that the plural personal pronouns do not need strong counterparts as the singular personal pronouns do, because they are disyllabic, and hence already 'strong' forms themselves prosodically.

2.1 The strong forms of singular personal pronouns

The forms *a*^{24}*-le*, *ṇ*^{44}*-le* and *kʰə*^{44}*-le* are believed to be the strong forms because they are prosodically prominent, in terms of not only volume and length, but also and pitch and timbre. In other words, personal pronouns *a*^{24}*-le* '1.SG', *ṇ*^{44}*-le* '2.SG' and *kʰə*^{44}*-le* '3.SG' display a greater volume, longer duration, higher pitch and clearer timbre

---

[2] *ian*^{44} can stand alone as a free morpheme meaning 'person' 人 in Tunxi Hui, which is at the same time the source of its plural marker for personal pronouns. Apart from Hui, some Min varieties of Chinese also make use of the plural marker grammaticalizaed from the noun meaning 'person' 人.

[3] The form (-)*le* is a multi-functional morpheme in Tunxi Hui. As a suffix, apart from its function as forming a strong counterpart of the weak form of singular personal pronouns, it can otherwise function as a nominal affix to nominalize a verb, or to address a particular person with affection.

compared with their weaker counterparts, including both uncliticized and cliticized weak forms, i.e. *a²⁴* '1.SG'/*a²³* '1.SG', *ṇ⁴⁴* '2.SG'/*ṇ⁰* '2.SG' and *kʰə⁴⁴* '3.SG'/*kʰə⁰* '3.SG'. It is therefore no wonder that the strong form will be always paired with those situations where prosodically prominent forms are required, for example, in isolation, the clause final position and the focus position in the information structure. The scenarios where strong forms are obligatorily required or preferred are elaborated below:

(i)　　In isolation, for instance:

(1) A:　　*la³¹⁻⁵-ka⁴²*　　*a?*
　　　　　which-CL　　Q
　　　　'Who is it?'
　　B:　　*a²⁴e/*a²⁴/*a²³.*
　　　　　1.SG
　　　　'me'

　　　　In example (1), since the answer contains only one morpheme, i.e. standing alone, only the strong form ***a²⁴e*** is allowed.

(ii)　　In clausal final position, for example:

First person singular forms *a²⁴e/*a²⁴/*a²³*

(2) *pau¹¹　mo³¹　　ka⁴²　　mə¹¹⁻²¹sʔ¹¹　　tɔ¹¹　　ti⁴²　　a²⁴e/*a²⁴/*a²³.*
　　OM　that　　CL　　thing　　　carry　　COV　　1.SG
　　'Please pass that thing to me.'

(3) *ṇ⁴⁴　　ɕi¹¹⁻²¹-ŋiau¹¹　kau³¹　　a²⁴e/*a²⁴/*a²³?*
　　2.SG　how-way　　talk　　1.SG
　　'Why are you scolding me?'

Second person singular forms *ṇ⁴⁴e/*ṇ⁴⁴/*ṇ⁰*

(4) *a²⁴　　kau³¹　　pu¹¹　　ko⁴²　　ṇ⁴⁴e/*ṇ⁴⁴/*ṇ⁰.*
　　1.SG　talk　　NEG　　pass　　2.SG
　　'I cannot talk you into it.'

(5) $k^h\vartheta^{44}$ $\textit{tɕiau}^{24}\textit{miɛ}^{11}$ $ma\textit{ʔ}^{5}$ **ņ ^{44}le/*ņ ^{44}/*ņ ^{0}?** $t^han^{42}pu^{11}$ t^han^{42} $a\textit{ʔ}$

  3.SG   just now   squeeze   2.SG   hurt   NEG   hurt   Q

  'He/she just squeezed you? Does it hurt?'

Third person singular forms $k^h\vartheta^{44}$le/*$k^h\vartheta^{44}$/*$k^h\vartheta^{0}$

(6) a^{24} $\textit{tɕ}^hio^{11}xa\textit{ʔ}^{5}$ **$k^h\vartheta^{44}$le/*$k^h\vartheta^{44}$/*$k^h\vartheta^{0}$.**

  1.SG   fear   3.SG

  'I fear him/her.'

(7) pau^{11} mo^{31} $p\varepsilon^{31}$ cy^{11} $t\mathfrak{o}^{11}$ ti^{42} **$k^h\vartheta^{44}$le/*$k^h\vartheta^{44}$/* $k^h\vartheta^{0}$.**

  OM   that   CL   book   carry   COV   3.SG

  'Please bring that book to him/her.'

In the above examples, all the clausal final positions require the strong forms with the suffix –*le*.

In addition to occurring in isolation and the clause final position, there is a third scenario where the strong forms are usually required. Nevertheless, the uncliticized weak form, but not the cliticized weak form, can sometimes be employed in the following as well:

(iii) In the focus position, including both information focus and discourse focus position, whereby three situations are often included:

(a) Default focus position, like answers:

(8) A: $la\textit{ʔ}^{24-5}\textit{-}ka^{42}$ $k\vartheta\textit{ʔ}^{5}$ $\textit{ņ}^{44}$ kau^{31} $ka\textit{ʔ}$

     Which-one   COV   2.SG   speak   SFP

     'Who told you?'

   B: $k^h\vartheta^{44}$le/$k^h\vartheta^{44}$/*$k^h\vartheta^{0}$ $k\vartheta\textit{ʔ}^{5}$ $\textit{ņ}^{44}$ kau^{31} $ka.$

     3.SG   COV   2.SG   speak   SFP

     'He/she told me.'

Answers are the default information focus in a clause. Therefore, both the strong

form and the uncliticized weak forms are acceptable in example (8).

(b) In contrast

(9) *a²^{4}le/a²^{4}/*a²³*   *tɕʰi?^{5}*   *ɕy³¹,*   *kʰə^{44}le/kʰə^{44}/*kʰə^{0}*   *tɕʰi?^{5}*   *tsɔ^{44}.*
　　1.SG　　　　　eat　　water　　3.SG　　　　　　　　eat　　　tea
　　'I will have water, while he/she will have tea'

Example (9) represents a contrastive construction. Again, the strong form and the uncliticized weak form are grammatical, while the unstressed weak forms are ungrammatical.

(c) With intensification

(10) *ɕi²^{4}*   *a²^{4}le/a²^{4}/*a²³*   *tɕio^{42}*   *kʰə^{44}le/kʰə^{44}/*kʰə^{0}*   *kʰə^{42}*   *ka.*
　　COP　　1.SG　　　　　ask　　3.SG　　　　　　　　　　go　　SFP
　　'It is me who ask him/her to go.'

Sentence (10) is a cleft sentence, with the pronoun *a²^{4}le* '1.SG' immediately following the copular verb *ɕi²^{4}* 'be' as the discourse focus. Hence, only the disyllabic strong form *a²^{4}le* '1.SG' or the monosyllabic uncliticized weak form *a²^{4}* '1.SG' can fill this position.

2.2　The weak forms of singular personal pronouns

The weak forms exhibit both an uncliticized and cliticized alternative. Compared with their strong counterparts, the weak forms cannot stand in isolation or appear in the sentence final position, which means they are clitics which always require something to 'lean on', typically verbs. Furthermore, the uncliticized and cliticized weak forms differ in their hosts. While the uncliticized weak forms require a morpheme immediately following themselves to lean on, the cliticized weak forms have to be surrounded by morphemes both before and after themselves, with at least one morpheme being a verb. In other words, the uncliticized weak forms are proclitics which need to join with the following morphemes, whereas the cliticized weak forms are endoclitics which need to be surrounded and lean both forward and backward onto.

In this section, we will start with the distinctions between the strong and weak forms, followed by distinctions between the two weak forms.

2.2.1 Weak forms versus strong forms

As mentioned above, the distinction between the strong and the weak forms of singular personal pronouns is more straightforward: the weak forms cannot appear in isolation or the sentence final positions as their strong counterparts do, for example:

(11) A:　*la³¹⁻⁵-ka⁴²*　　*ɕi²⁴*　　　*mo³¹le*　*a*?

　　　which-CL　　COP　　　there　　Q

　　　'Who is there?'

　　B:　*a²⁴le/*a²⁴/*a²³*.

　　　1.SG

　　　'me'

In sentence (11), neither the uncliticized weak form *a²⁴* '1.SG' nor the cliticized weak form *a²³* '1.SG' is possible because weak forms are not allowed to stand alone.

Similarly, sentence final position also prohibits both the uncliticized and the cliticized weak forms of personal pronouns, for example:

(12) *kʰə⁴⁴*　　*pu¹¹*　　*xu:ə¹¹ɕi³¹*　　*ŋ̍⁴⁴le/*ŋ̍⁴⁴/*ŋ̍⁰*.

　　3.SG　　NEG　　like　　　2.SG

　　'He/she doesn't like you.'

In example (12), the second personal singular form appears in the sentence-final position. As has been explained in the previous section, sentence-final positions only license the occurrence of the strong forms, in this instance *ŋ̍⁴⁴le* '2.SG', not the uncliticized weak form *ŋ̍⁴⁴* '2.SG' or the cliticized weak form *ŋ̍⁰* '2.SG'.

However, the differences between the uncliticized and cliticized weak forms are more subtle. We will discuss them in the next section.

2.2.2 Uncliticized weak forms versus cliticized weak forms

Compared with the strong personal pronouns, the weak personal pronouns are more like clitics which needs to 'lean on' something, usually a verb. The uncliticized weak forms and the cliticized weak forms mainly differ from each other in their hosts. The uncliticized weak forms are proclitics which need to lean on the following morphemes. The cliticized weak forms, however, are more of the 'clingy' type: they need to 'lean on' hosts both before and after themselves, with at least one host being a verb. Here is an example:

(13) a.   **ṇ^{44}/ʔṇ^{0}**   *tɕʰiʔ^{5}*   *pan^{11}*   *tɕʰiʔ^{5}*   *fu:ə^{11}?*

           2.SG   eat   not yet   eat   rice

           'Have you had your meal?'

   b.   *kʰə^{44}*   *mɛ^{11}*   **ṇ^{44}/ṇ^{0}**   *fu:ə^{11}*   *tɕʰiʔ^{5}*   *pan^{11}*   *tɕʰiʔ^{5}?*

           3.SG   ask   2.SG   rice   eat   not yet   eat

           'He/she asked whether you had had your meal.'

In example (13a), although the cliticized weak form *ṇ^{0}* '2.SG' has a host to lean on, it does not have a host to lean forward onto, therefore the uncliticized weak form *ṇ^{44}* '2.SG' is preferred , which is not as bound or 'clingy' as the cliticized counterpart. In example (13b), the cliticized weak form *ṇ^{0}* '2.SG' appears in a pivotal construction, both as the object of the verb *mɛ^{11}* 'ask' in the main clause and the subject in the subordinate clause. Surrounded by two morphemes with at least one being the verb, the existence of the cliticized weak form *ṇ^{0}* '2.SG' is hence justified.

Apart from pivotal construction, the uncliticized weak forms are also seen in double object constructions where they can lean on the morphemes before and after themselves, as in Sentence (14b):

(14)   a.   *ṇ^{44}*   *tɔ^{44}*   *ka^{31} ka^{42}*   *mə^{11}*   *ti^{42}*   **kʰə^{44}/*kʰə^{0}!*

            2.SG   bring   this CL   thing   COV   3.SG

            Intended meaning: 'Please bring this to him/her!'

   b.   *a^{24}*   *tsʰo$^{11-21}$ɲie^{11}*   *ti^{42}*   *kʰə^{44}/kʰə^{0}*   *liau^{24}*   *pʰɛ^{44}*   *tsiu^{31}*.

            2.SG   yesterday   give   3.SG   two   CL   wine

            'I gave him/her two bottles of wine yesterday.'

In the serial verb construction (14a), neither the uncliticized weak form $k^h\vartheta^{44}$ '3.SG' nor the cliticized weak form $k^h\vartheta$ '3.SG' is allowed since they occupy a sentence final position where only the strong form is acceptable. Nevertheless, in the double object construction (14b), both the uncliticized and the cliticized weak form $\boldsymbol{k^h\vartheta^{44}/k^h\vartheta^0}$ '3.SG' are grammatical, because they have the verb host ti^{42} 'give' and/or the direct object 'two bottles of wine' to lean on.

To sum up, the strong forms of personal pronouns is the most free among all forms of singular personal pronouns, followed by the uncliticized weak personal pronouns which can generally appear in positions where they can lean backward on. The cliticized weak personal pronouns are the most 'clingy' of all, which can only be seen when surrounded by other morphemes, with at least one host being the verb. The hierarchy of freedom of the three forms of singular personal pronouns is schematized below:

Strong forms > Uncliticized weak forms > Cliticized weak forms

Free ← →Bound

One final note about the strong personal pronouns is that, although in theory the strong forms can appear in any position, in reality the uncliticized weak form is sometimes preferred in positions such as the subject, as long as no special emphasis is placed on it. This phenomenon may be explained in terms of the principle of economy.

2.3 Distribution of the strong and weak personal pronouns

Following the above criteria on the strong forms, the uncliticized and cliticized weak forms, we can reasonably predict the distribution, and hence conclude the nature of these forms as follows:

Singular Personal Pronouns	Criteria	Nature	Distribution
Strong Forms	Free	Free	In isolation; In sentence final positions; In discourse focus positions
Uncliticized	Lean backward	Proclitics	Most often seen in the

Weak Forms	on		subject positions
Cliticized Weak Forms	Surrounded by other morphemes	Endoclitics	In pivotal constructions; In double object constructions

Diagram 3 Distribution of the Three Forms of Singular Personal Pronouns in Tunxi Hui

The strong forms of personal pronouns, namely *a² ⁴le* '1.SG', *ņ ⁴⁴le* '2.SG' and *kʰə ⁴⁴le* '3.SG', are more free in nature. Technically they can appear in all occasions, especially in isolation, in the sentence final position, and in the discourse focus position. For instance:

The Strong Forms (*a² ⁴le* '1.SG', *ņ ⁴⁴le* '2.SG' and *kʰə ⁴⁴le* '3.SG')

(15) A: *laʔ² ⁴⁻⁵-ka ⁴²   kəʔ⁵   ņ ⁴⁴le   kau³¹   ka?*
       Which-one   COV   2.SG   talk   Q
       'Who told you?'

   B: ***kʰə ⁴⁴le.***
       3.SG
       'him/her.'

(16) A: *kʰə ⁴⁴   tɕi³¹   tiɛ³¹   tsan¹¹   tə ⁴²   kɔ¹¹   a?*
       3.SG   how many   point   clock   reach   home   Q
       'When will he/she arrive at home?'

   B: *a² ⁴   pu¹¹   ɕi² ⁴   tiʔ⁵   ņ ⁴⁴da³¹ ti ⁴²   **kʰə ⁴⁴le!***
       1.SG   NEG know   RVC   2.SG   call   COV   3.SG
       'I don't know. Please call him/her!'

(17) *a² ⁴   kau³¹   pu¹¹   ko ⁴²   **kʰə ⁴⁴le.***
     1.SG   talk   NEG   pass   3.SG
     'I cannot talk him/her into it.'

(18) *ņ ⁴⁴   mau¹¹tɕi³¹   tɕia?*
     2.SG   forget   PERF.SFP
     'You have forgotten about it?'

ɕiau¹¹ tsʰŋ⁴² ɕi²⁴ **ṇ⁴⁴le** io⁴² tɕʰiʔ, pu¹¹ ɕi²⁴ **a²⁴le** io⁴² tɕʰiʔ⁵.
last ime COP 2.SG want eat NEG COP 1.SG want eat
'It is you who wanted to eat it last time, not me.'

On the other hand, the uncliticized weak forms of singular personal pronouns, i.e. a^{24} '1.SG', $ṇ^{44}$ '2.SG' and $kʰə^{44}$ '3.SG', are not as free as their strong counterparts, and they are proclitics which need something to 'lean backward' on, usually verbs. The subject positions are one of the positions where they are often observed, since they can most conveniently 'lean on' the following verbs as their hosts. For example:

The Uncliticized Weak Forms (a^{24} '1.SG', $ṇ^{44}$ '2.SG' and $kʰə^{44}$ '3.SG')
(19) **kʰə⁴⁴** tɕio⁴² ṇ⁴⁴le to¹¹ tɕʰiʔ⁵ tin²⁴.
3.SG ask 2.SG more eat a bit
'He/she asks you to eat more.'

(20) **a²⁴** pu¹¹ ɕi²⁴ tiʔ⁵ kʰə⁴⁴le.
3.SG NEG know RVC 3.SG
'I don't know about him/her! (You need to check with him/her yourself.)'

(21) A: **ṇ⁴⁴** ɕi²⁴ laʔ³¹⁻⁵-ka⁴² a?
2.SG COP which-CL Q
'Who are you?'
   B: **a²⁴** ɕi²⁴ kʰə⁴⁴ ka len²⁴.
1.SG COP 3.SG GEN daughter
'I am her/his daughter.'

Last but not least, the cliticized weak forms of singular personal pronouns, namely a^{23} '1.SG', $ṇ^{0}$ '2.SG' and $kʰə^{0}$ '3.SG', are the most bound of all forms. They need to lean on morphemes both before and after themselves, with at least one morpheme as the verb. Hence, they are more like endoclitics in nature and they are the most restricted in distribution. The double object constructions and the pivotal constructions are the most common structures to accommodate the cliticized weak forms, where they can both 'lean

forward' and 'lean backward' onto, with at least one morpheme being verb, as their host. For example:

The Cliticized Weak Forms (a^{23}'1.SG', $ŋ^{0}$'2.SG' and $k^{h}ə^{0}$'3.SG')

In a pivotal construction

(22) $ŋ^{44}$ $ŋiau^{11}$ $\boldsymbol{k^{h}ə^{0}}$ $k^{h}ə^{42}$ $uɛ!$
     2.SG   let   3.SG   go   SFP

     'Please let him/her go!'

In a double object construction

(23) a^{24} ti^{42} $\boldsymbol{k^{h}ə^{0}}$ $tɕ^{h}iu^{11}$-$ɕi^{24}$ $liau^{24}$ $pu:ə^{11}$-$tɕiau^{31}$.
     1.SG   give 3.SG   just   two   hand

     'I slapped him/her twice on the face.'

3. Conclusions

       This study has focused on the strong and weak personal pronouns in an under-studied Hui variety of Chinese, namely Tunxi Hui. While the bi-morphemic strong personal pronouns in Tunxi Hui can stand alone, the mono-morphemic uncliticized and cliticized personal pronouns are proclitics and endoclitics in nature, which must lean on other morphemes, typically verbs. To be specific, the uncliticized personal pronouns need to lean on the following morphemes, whereas the more 'clingy' cliticized personal pronouns require the existence of morphemes surrounding themselves.

       The distinct natures of personal pronouns result in different distributions. The strong forms are the most free type, which can appear in most situations, especially in isolation, clause final positions and discourse focus positions. The uncliticized forms prefer the subject positions where they can lean on the following verbs. Finally, the cliticized forms need to be surrounded by other morphemes, and are hence most seen in pivotal constructions and double object constructions.

REFERENCES

CAO, ZHIYUN (eds.) 2008. *Hanyu Fangyan Dituji* 漢語方言地圖集 (Linguistic Atlas of Chinese Dialects). 2nd Edition. Beijing: Commercial Press.

CHAPPELL, HILARY M. 2016. 'Linguistic Areas in China for Differential Object Marking, Passive, and Comparative Constructions'. In Chappel, Hilary M. (ed). Diversity in Sinitic Languages: 13-52. Oxford: Oxford University Press.

HASHIMOTO, M. J. 1976. Language diffusion on the Asian continent. *Computational Analyses of Asian and African Languages* vol. 3: 49-63.

HASHIMOTO, M. J. 1986. The Altaicization of Northern Chinese. In J. McCoy& T. Light (eds.), *Contributions to Sino-Tibetan Studies*. Leiden: E.J. Brill, 76-97.

LI, XUPING. 2015. 'Complex Pronouns in Wu Chinese: Focalization and Topicalization'. In Hiliary. M. Chappell (ed.) *Diversity in Sinitic Languages*. Oxford: Oxford University Press: 226-247.

NORMAN, JERRY. 1988. *Chinese*. Cambridge: Cambridge University Press.