An introduction to Taiwan Hakka
Focusing on its sounds, morph-syntax and social background

Huei-ling Lai

Introduction

The word 客家 kejiā ‘Hakka’, literally denoting ‘guests’, symbolizes the social status of the Hakka people. Two views are held in the literature as to the formation of Hakka people. One view holds that the Hakka people originated from the Central Plains of China, and because of foreign invasions, civil wars, and other historical reasons, they moved southwards to mainly the southern areas of China. The word 客家 ‘Hakka’ is hence used as a category to refer to those migrant northerners as a distinction of the local inhabitants. Subsequently, after a couple of major moves, around the middle of the 19th century, some Hakka people finally migrated to Taiwan and scattered around the island (cf. Hashimoto, 1973; Luo, 1998). An alternative, and perhaps linguistically more reasonable, view holds that the so-called 客家 ‘Hakka’ refers to a hybrid group emerging from the southern Gan during the Song Dynasty (10th~13th centuries). The Hakka language is a mixture of Chinese and non-Chinese languages, bearing certain linguistic features similar to non-Chinese languages such as She and Yao around Gan areas. With subsequent migrations, they moved down to southeastern Jiangxi and western Fujian, and further south to northeastern Guangdong. Afterwards, some Hakka people migrated southwards to Taiwan in the early Qing Dynasty (cf. Chappell and Lamarre, 2005). After several generations, the Hakka language featured dialectal differences because of geographical and historical matters. Taiwan Hakka (台湾客语) used in this paper refers to the various varieties spoken by the Hakka ethnic groups in Taiwan. These Hakka dialects used in Taiwan may have fostered variations in sounds and usages different from the Hakka language still used in the Mainland. This discussion is based on Taiwan Hakka.

Note that Taiwan Mandarin here is what is called Taiwan Standard Chinese by Lin (2007: 267). Lin (2007) deems that Taiwan Mandarin (Taiwan Guoyu) is ambiguous as it can “refer to either the local norm of Standard Chinese in Taiwan or the non-standard accents heavily influenced by Taiwanese or both” (266). Instead, she uses Taiwan Standard Chinese to refer to the former and Taiwanese-accented Standard Chinese to refer to the latter. The typical norm of Standard Chinese as pointed out by Lin (2007) is the prescribed average or the textbook standard. In Taiwan, most Mandarin speakers speak the local norm of Standard Chinese. To illustrate, the neutral tone, which is commonly found in Standard Mandarin dialect, is much less
commonly used by speakers of Taiwan Mandarin. The lack of post-alveolars is also pervasive in Taiwan Mandarin (Lin, 2007: 267). For the sake of simplification, Taiwan Mandarin and Taiwan Hakka are used to refer to the variants spoken in Taiwan. Examples of Taiwan Mandarin will be illustrated as a comparison and contrast to Taiwan Hakka examples from time to time.

The content of this chapter is composed as follows. For the first part, a general linguistic introduction of the Taiwan Hakka language (TH) will be given, including the sounds, and the morpho-syntactic features. Next, the current state of TH will be presented in the social-cultural context of Taiwan. Specific attention will focus on the decline of Taiwan Hakka together with the transformation of language policy of the current regime. Then, a Taiwan Hakka spoken corpus will be introduced regarding its construction, its content and its significance for the documentation and preservation of the Taiwan Hakka language. Finally, a brief analysis of the semantic and discourse functions of degree words based on the data drawn from the spoken corpus will be discussed.

An introduction of Taiwan Hakka

According to the demographic survey report by Hakka Affairs Council (客委會HAC, 2017) in Taiwan, the Hakka population comprises 19.3% of the total population, approximating 4.5 million people, and is the second largest ethnic group. Hakka language refers to the Hakka dialects spoken in Taiwan. In general, six dialect varieties can be identified: Sixian (Northern and Southern varieties), Hailu, Daipu, Raoping, and Zhoan. According to a survey by HAC (2013), the distribution of Hakka speakers of each variety can be shown as follows: 北四县 Northern Sixian Hakka (56.1%, mainly spoken in Miaoli), 南四县Southern Sixian Hakka (4.8%, mainly spoken in Kaohsiung and Pingtung), 海陆Hailu Hakka (41.5%, mainly spoken in Hsinchu and Taoyuan), 大埔Dapu Hakka (4.2%, mainly spoken in Tungshih and Hsinser), 饶平Raoping Hakka (1.6%, sporadically spoken in Changhua, Yuanlin and Jiayi) and 诏安Zhaon Hakka (1.3%, mainly spoken in Lunbei and Erlun in Yunlin). The survey also shows that among the Hakka speakers, 85.9% can speak one dialect, 22.4% can speak two dialects, 1.2% can speak three dialects, and another 14.1% of speakers can speak Hakka, but are not sure which kind of dialect.

The sound systems of TH can be shown by the consonants, vowels and tones (Taiwan Ministry of Education, 2008; Xu, 2009). There are 21 consonants in TH as given in Table 39.1. Notice that Chung (2004) gives 24 consonants, including the three post-alveolar variants [tɕ, tɕʰ, ɕ] of palato-alveolars [ʃ, ʒ, ɲ]. As stated in Chung (2004: 64), these three variants are found only in Raoping dialect. On the other hand, Luo (1996) gives 17 consonants, stating that the

```
<table>
<thead>
<tr>
<th>manner</th>
<th>stop</th>
<th>affricate</th>
<th>fricative</th>
<th>nasal</th>
<th>lateral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>p</td>
<td>pʰ</td>
<td></td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Labiodental</td>
<td></td>
<td></td>
<td>f</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>Dental</td>
<td>t</td>
<td>tʰ</td>
<td>ʦ</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>palato-alveolar</td>
<td>ʧ</td>
<td>ʧʰ</td>
<td>ʃ</td>
<td>ʒ</td>
<td>n</td>
</tr>
<tr>
<td>Velar</td>
<td>k</td>
<td>kʰ</td>
<td></td>
<td>ŋ</td>
<td></td>
</tr>
<tr>
<td>Glottal</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* For the two symbols of stops and affricates, the left one is voiceless unaspirated and the right one is voiceless aspirated. The shaded ones are voiced.
```
four consonants [ʧ, ʧʰ, ʃ, ʒ] are used in Hailu dialect. To represent most of the surface forms used in TH, 21 consonants are identified. As for vowels, consensus is high among scholars, and six vowels are given in Table 39.2.

While Hailu Hakka has all 21 of them, Sixian Hakka has only 16 consonant phonemes with the four palato-alveolars [ʧ, ʧʰ, ʃ, ʒ] merged into dentals, [ts], [tsʰ], [s]. The three alveolo-palatal sounds [tɛ], [tɛʰ], and [ɛ] occur in front of high front vowels as in 井 [tɛIan3] ‘shallow’, and 想 [gioN2] ‘think’. Like Taiwan Mandarin (TM), TH has six oral stops, three of which are voiceless unaspirated and three voiceless aspirated. The places or articulation, bilabial, dental, and velar, also feature three nasals. In addition, like TM, TH has four affricates, but they differ in place of articulation. Unlike TM, which has four fricatives, TH has six fricatives, with two extra ones not found in TM: voiced labiodental fricative [v], and voiced palato-alveolar [ʒ].

Examples of TH will be shown in Sixian Hakka (SH) forms. Forms from other TH dialects will be specified when necessary. Tone numbers are used to represent the pitch values in the examples: 1, rising; 2, falling; 3, high level; 4, short low; 5, low level; 8, short high. Several phonological features can further distinguish TH from TM. The [v] sound is unique in TH, as shown in the contrast of 湾 ‘bay’ [uan1] in TM and [van1] in TH; and the contrast of 黄 ‘yellow’ [huang2] in TM, and [von3] in TH. Second, the sound [ʒ], which does not occur in TM, occurs only in Hailu Hakka as in 云 ‘cloud’ [ʒun5]. In addition, the velar nasal [ŋ] can occur as an onset in TH, while it cannot in TM. For example, the word 乐 ‘music’ is pronounced as [ŋok8] in TH. Like TM, TH allows onsetless syllable structures when the rime starts with the two vowels [o], and [a], as in 暗 ‘dark’ [am3], and 愛 ‘love’ [oi3]. Furthermore, TH maintains the paradigm of three nasals [m, n, ŋ] in a word-final position whereas TM has only two nasals [n, ŋ]. For example, the word 反 shows a contrast [kan1] in TM and [kam1] in TH. Moreover, TH maintains the paradigm of three stops [p, t, k] in a word-final position whereas TM does not allow any stops as a coda. For example, the word 罚 ‘to punish’ shows a contrast [ta4] in TM and [tap8] in TH; 罰 ‘to punish’ shows a contrast [fa2] in TM and [fat8] in TH; and 醫 ‘to support’ shows a contrast [tʰuo] in TM and [tʰok4] in TH, in addition to the difference in the vowels.

Like TM, TH has five vowel phonemes: three unrounded and two rounded. In TH, the two rounded vowels are back vowels as in 步 ‘step’ [pʰu3], and 锁 ‘key’ [so2]. However, in TM, mid back rounded [o] is not realized. On the other hand, TM has a high front rounded [y], as in 綠 ‘green’ [ly4]. This sound, however, is always realized as high front unrounded [i] in TH as contrasted by 雨 ‘to rain’ [y3] in TM and [i1] in TH. In addition, the high central unrounded [i], although often identified as one of the main vowels in TH, is an allophone of [i]; that is, [i] occurs only after dental fricatives and affricates [s, tsʰ, ts], as in 吃 ‘to eat’ [sIt8], 齒 ‘tooth’ [tsʰi2], and 紙 ‘paper’ [tsi2].

Only four tones exist in TM – high level, high rising, low falling-rising, and high falling. On the contrary, most varieties of TH have six tones except for Hailu, which has seven tones. Table 39.3 gives the tone systems of Sixian Hakka (SH) and Hailu Hakka (HH), which show

### Table 39.2 Vowels in TH*

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>ð</td>
<td>u</td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td></td>
<td>o</td>
</tr>
<tr>
<td>Low</td>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The shaded ones are rounded vowels.
The difference also shows in morpho-syntactic formation of combining words. For instance, the three daily meals in TM are 早餐 [zao3can1] ‘breakfast’, 午餐 [wo3can1] ‘lunch’, and 晚餐 [wan3can1] ‘dinner’. But in TH, the corresponding usages are 食朝 [siid8 zu1] ‘to eat breakfast’, 食昼 [siid8 zu3] ‘to eat lunch’ and 食夜 [siid8 ia3] ‘to eat dinner’, in which the verb 食 [siit8] ‘to eat’ needs to collocate with the earlier monosyllabic word forms for ‘morning’, ‘noon’ and ‘evening’ in TH to metonymically refer to the meal at that particular time. In TM, on the other hand, the verb 吃 [chi1] ‘to eat’ can be freely combined with the three meals to form a larger three-syllabic chunk as in 吃早餐 [chi1zao3can1] ‘to eat breakfast’.

Moreover, when some disyllabic compounds are found in both TM and TH, they show reverse word order combinations. Contrast the following examples in Table 39.4:

<table>
<thead>
<tr>
<th>Gloss</th>
<th>TM</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘beautiful’</td>
<td>美丽 [mei3li4]</td>
<td>靓 [jiang1]</td>
</tr>
<tr>
<td>‘happy’</td>
<td>高兴 [gao11xing4]</td>
<td>畅 [tiong3]</td>
</tr>
<tr>
<td>‘to fix’</td>
<td>修理 [xiu1li3]</td>
<td>整 [zang2]</td>
</tr>
<tr>
<td>‘to fear’</td>
<td>害怕 [hai4pa4]</td>
<td>惊 [giang1]</td>
</tr>
<tr>
<td>‘tunnel’</td>
<td>隧道 [sui4dao4]</td>
<td>窿 [lung5]</td>
</tr>
</tbody>
</table>

The difference also shows in morpho-syntactic formation of combining words. For instance, the three daily meals in TM are 早餐 [zao3can1] ‘breakfast’, 午餐 [wo3can1] ‘lunch’, and 晚餐 [wan3can1] ‘dinner’. But in TH, the corresponding usages are 食朝 [siid8 zu1] ‘to eat breakfast’, 食昼 [siid8 zu3] ‘to eat lunch’ and 食夜 [siid8 ia3] ‘to eat dinner’, in which the verb 食 [siit8] ‘to eat’ needs to collocate with the earlier monosyllabic word forms for ‘morning’, ‘noon’ and ‘evening’ in TH to metonymically refer to the meal at that particular time. In TM, on the other hand, the verb 吃 [chi1] ‘to eat’ can be freely combined with the three meals to form a larger three-syllabic chunk as in 吃早餐 [chi1zao3can1] ‘to eat breakfast’.

Moreover, when some disyllabic compounds are found in both TM and TH, they show reverse word order combinations. Contrast the following examples in Table 39.5.
In addition, while both TM and TH employ various types of compounding to give rise to a rich set of vocabulary, different roots are sometimes found among the same morpho-syntactic types and the same concepts. Contrast the following examples in Table 39.6:

Table 39.5 Examples of different word formation orders in TM and TH

<table>
<thead>
<tr>
<th>Gloss</th>
<th>TM</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘rooster’</td>
<td>公鸡 [gong1ji1]</td>
<td>鸡公 [jie1 gung1]</td>
</tr>
<tr>
<td>‘guest’</td>
<td>客人 [ke4ren2]</td>
<td>人客 [ngin5 ha4]</td>
</tr>
<tr>
<td>‘dust’</td>
<td>灰尘 [hui1chen2]</td>
<td>尘灰 [ciin5 fo1]</td>
</tr>
<tr>
<td>‘bustling’</td>
<td>热闹 [re4nao4]</td>
<td>闹热 [nau3 ngied8]</td>
</tr>
<tr>
<td>‘in front of’</td>
<td>前头 [qian2tou2]</td>
<td>头前 [teu5 qien5]</td>
</tr>
</tbody>
</table>

Table 39.6 Examples of different compounds in TM and TH

<table>
<thead>
<tr>
<th>morpho-syntactic type</th>
<th>Gloss</th>
<th>TM</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>coordinate</td>
<td>‘eyes’</td>
<td>眼睛 [yan3jing1]</td>
<td>目珠 [mug4 zu1]</td>
</tr>
<tr>
<td>modifier-head</td>
<td>‘noon’</td>
<td>中午 [zhong1wu3]</td>
<td>当昼 [dong1zu3]</td>
</tr>
<tr>
<td>verb-resultative</td>
<td>‘finish-eating’</td>
<td>吃完 [chi1wan2]</td>
<td>食忒 [siid8 ted4]</td>
</tr>
<tr>
<td>subject-predicate</td>
<td>‘earthquake’</td>
<td>地震 [di4zhen4]</td>
<td>地动 [di3tung1]</td>
</tr>
<tr>
<td>verb-object</td>
<td>‘angry’</td>
<td>生气 [shen1qi4]</td>
<td>发谴 [fad4 kien2]</td>
</tr>
</tbody>
</table>

Table 39.6 indicates that the five morpho-syntactic and semantic relations of categorizing compounds can occur in both TM and TH. Nevertheless, different roots are chosen for compounding. Furthermore, different semantic relations among the roots of compounds can cause various degrees of lexicalization, hence leading to differences in idiomatic meaning. To illustrate, the form 出手 occurs both in TM [chu1shou3] and in TH [cud4 su2]. Several idiomatic meanings can be detected with this case in TH, which correspond to different morpho-syntactic realizations in TM. Contrast the following examples in Table 39.7:

Table 39.7 Contrast of 出手 in TM and TH

<table>
<thead>
<tr>
<th>Gloss</th>
<th>TM</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘to sell goods when the price is right’</td>
<td>放手(卖掉) [fang4shou3mai4diao4]</td>
<td>出手(放忒) [cud4su2piong3ted4]</td>
</tr>
<tr>
<td>‘to help cope with troubles’</td>
<td>出面帮忙</td>
<td>出手 [cud4su2]</td>
</tr>
<tr>
<td>‘very generous’</td>
<td>用钱很大方</td>
<td>大出手 [qin3tai3cud4su2]</td>
</tr>
</tbody>
</table>

As illustrated by this particular example, different lexicalization patterns can be found in various kinds of compounding constructions (cf. Packard, 2000). It is also expected that TM and TH will show diversified representations of lexicalization patterns together with diversified shades of metaphorical meanings. Take another example for illustration. The word 习惯 [xi2guan4] in TM denotes ‘habit’, with quite a neutral connotation as in 饭前洗手是好习惯 [fan4qian2xi 3shou3shi4hao3xi2guan4] ‘Washing hands before eating is a good habit’. In TH, there are two
corresponding words. One is the exact word 习惯 [xíb à guān3] as in 饭前洗手系好习惯 [fàn-3qian5xe2su2he3xiō2xiō2xíb àguān3] ‘Washing hands before eating is a good habit’. The other highly synonymous word is 惯势 [guān3 se3]. Not only does the word have a reverse word order combination and a different root, but its meaning also carries a negative connotation – depreciating a person as being immature with improper character or behaviors acquired through accumulated habits and not knowing how to behave properly in front of others, as in the usage 老弟分人纵惯势 [lò2tái1bùn1yīng5 jiǜng2guān3se3] ‘The younger brother has been spoiled’.

Analyzing lexicalization and grammaticalization patterns together with their metaphorical extensions is an intriguing issue worthy of investigation. Yeh (2017) is an endeavor along this line of research in which three modal disjuncts in TH are analyzed regarding their emerging epistemic functions in TH corpora. The three targeted examples are 惊怕 [giàng1pa3], 敢怕 [gān2pa3], and 好得 [hō2ded4]. What is interesting is that the three compounds are composed of canonically separate verbs, 惊 [giàng1] ‘fear, be afraid’, 敢 [gān2] ‘dare, bold’, 怕 [pa3] ‘fear, be afraid’, 好 [hō2] ‘good’, and 得 [ded4] ‘obtain’, respectively. Examples include 惊舖娘 [giàng1bu1ngiōng5] ‘be afraid of (his) wife’, 女怕嫁错郎 [nǚ pa3 jià cuò láng] ‘women fear marrying wrong husbands’, 敢食苦瓜 [gān2sī dì fū2guá1] ‘dare eat bitter melons’, 好朋友 [hō2pén5yīu1] ‘good friends’, and 得金牌 [ded4 gíān1pái5] ‘obtained a gold medal’. The three compounds however developed various meanings and functions at different stages of lexicalization and grammaticalization, as have been extensively investigated by Yeh (2017). The final stages of their developments as observed in TH corpora have shown interesting pragmatic aspects that are lacking in TM. Specifically, both 惊怕 [giàng1pa3] and 敢怕 [gān2pa3] have emerged into down-toners that are used by speakers “to mitigate the face-affronting power or the force of a proposition for the sake of social harmony” (Yeh, 2017: 224). On the other hand, 好得 [hō2ded4] has developed into a pragmatic or discourse marker denoting speaker’s subjective evaluation toward the state of affairs. The following examples are adopted from Yeh (2017) for illustration.

Notice that the data used in Yeh (2017) are based on Hailu Hakka. For the sake of consistence of presentation throughout the paper, some of the lexical items are changed into Sixian Hakka, and The English free translations are revised when necessary.

1. 遮仔带等去，惊怕会落水。
   (Yeh, 2017: 224, 267a)
   
   I suggest you bring an umbrella with you. I’m afraid that it will probably rain.

2. 这碗橱仔做到真幼路，敢怕异贵哦？
   (Yeh, 2017: 224, 268a)
   
   This handcrafted cupboard is really sophisticated and elegant. It must be very expensive, right?

3. 嗅肉仰愁韧，好得佢切幼幼仔，2正咬得落。
   (Yeh, 2017: 274, 347)
   
   The pork is unexpectedly tough. Fortunately, he minced the pork into small pieces, and only then I was able to chew.
The first two cases show the pragmatic functions of 惊怕 [giang1pa3] and 敢怕 [gam2pa3] being down-toners. Both express value judgment of the speaker so as to characterize his attitude toward the state of affairs being denoted, and at the same time represent his assessment of the possibility of what he is saying. 惊怕 [giang1pa3] differs from 敢怕 [gam2pa3] in that the former is used in a declarative manner while that latter can be used in an interrogative manner, showing that 敢怕 [gam2pa3] leaves more room for the addressee, hence lessening the face-threatening act. In other words, 敢怕 [gam2pa3] used in an interrogative contexts is even softer in terms of illocutionary force than 惊怕 [giang1pa3]. On the other hand, the third example indicates 好得 [ho2ded4] is used as a pragmatic marker that expresses the speaker’s subjective perspective toward the state of affairs. As maintained by Yeh (2017: 274f), the first statement provides a background statement regarding the speaker’s evaluation of the state of the pork meat. Two possible situations can arise: either the speaker is able to chew the meat or the speaker is not able to chew the meat. Nevertheless, the statement denoted by the 好得 [ho2ded4] clause factors out the negative situation, and the positive situation is expressed by the final clause as the discourse expands. In other words, 好得 [ho2ded4], has developed into a pragmatic or discourse marker indicating speaker-oriented epistemic modality.

The decline of TH and measures for revival

Given the many interesting features at various linguistic levels, not many linguistic studies have paid attention to TH until recent two decades. While efforts still need to be exerted for more investigation of the language, TH has to face the music of losing speakers as time passes. The Hakka ethnic groups, owing to complicated historical, political, economic and social factors that led to the characteristics of their diaspora, have long been considered as invisible in a multilingual and multi-ethnic Taiwan society (cf. Hsu, 1994; Tsao, 1999; Chiou, 2006; Wang, 2007; Lai, 2017). Their invisible and diasporic nature has also led to the decline of their language and culture since it is not the dominant language prevalently used in public domains or in education.

In retrospect, TH has suffered under the language policy of the Mandarin Only Movement (推行国语运动) by the KMT government back in 1956 when Mandarin was proclaimed as the national language of Taiwan. It was extensively promoted in schools and public domains while at the same time other languages and/or dialects were forbidden to be used. Loss of the varieties of other Chinese dialects such as TH has resulted in the weakening of its public recognition in Taiwan society. The situation got worse when TH became relatively little spoken not only on public domains but also on a day-to-day basis within or across ethnic groups, due to the emigration of Hakka population out of Hakka villages to cities for economic needs. Since speaking Mandarin was convenient for communication, some native Hakka speakers stopped speaking their own mother tongues because they felt knowing Hakka was not very beneficial for survival in society. Most of them even hid their Hakka identities because Hakka people were stereotypically considered as stingy and impoverished.

Subsequently, after about 30 years, with the declaration of the end of martial law in Taiwan in 1987, a series of advocacy movements arose, challenging the Mandarin-only linguistic-cultural situation. Along with the change of political and societal environment, the Hakka ethnic groups launched their first social movement in 1988 – the Give My Mother Tongue Back Movement (还我母语运动). This event was a milestone as it raised the Hakka people’s awareness of the crisis of the serious loss of their culture, language and tradition. Twelve years later, the DPP won the presidential election in year 2000 for the very first time in Taiwan. The Hakka ethnic groups then urged the new government to put effort into launching Hakka policies including Hakka culture promotion, language preservation and development and media exposure. To echo their request, The Council for Hakka Affairs was established in 2001 at the central government
level and various Hakka-related administrative units were established in the cities and counties. Hakka cultural parks and local-level public buildings have been constructed to include Hakka-style village motifs. Annual Hakka traditional festivals, activities, and events have appeared in significant numbers ever since. With Hakka culturally related activities enthusiastically held and hardware constructions massively built up, the Hakka ethnic groups seem to be more visible in Taiwan society. In the year 2001, the government also launched a language policy regarding incorporating native languages into elementary school curriculum. Since 2001, students from each grade have been required to take a one-hour native-language course choosing from one of the native languages – Taiwan Hakka, Taiwan Southern Min or Aboriginal languages. Nevertheless, these efforts still cannot stop the trend in the decline of their language. A report from the HAC (2013) indicates that all the varieties of Taiwan Hakka language are still suffering a serious decline due to the loss of speakers among the younger generations. The report shows that almost all of the speakers of TH who are fluent in the language are over 60 years old; and only 15% of Hakka people in the age group of 13–18 can speak TH.

Taiwan Hakka spoken corpus

The decline of TH in Taiwan society has aroused concerns and awareness of the importance of preserving TH. Implementation plans such as incorporating TH into language teaching programs in primary schools have been adopted to promote TH. For the documentation and preservation of TH, methods and principles of corpus linguistics can be considered as one of the effective measures. The construction of the NCCU Corpus of Spoken Hakka, a sub-corpus of the NCCU Spoken Corpus of Chinese constructed in 2007, is such an effort (Chui and Lai, 2008; Chui et al., 2017). The Corpus of Spoken Hakka includes three varieties of spoken data: face-to-face conversations, TV talk programs, and oral narratives based on Frog, Where are You? by Mercer Mayer (1969). This book is a wordless children book with 24 illustrations. It is about a boy whose frog is missing and tries to look for it with his dog. The book is famous to and has been used by linguists around the world for elicitation of narratives among children for dozens of languages. It has also been used for collecting narratives of second language learners of different languages. Although a sociolinguistic stratification of sampling measure is adopted to take gender and age into consideration, almost all of the conversation participants were in the range of 46–60 or that of >60 years old. Since younger speakers were not confident to converse with TH for 60 minutes, they were hence recruited for frog storytelling for two purposes. For one, a balance of the age stratification is a bit more maintained; for the other, various aspects of language use from the young speakers can be documented, providing data for language changes. Up until 2017 summer, 96 recordings have been collected including 39 (40.63%) face-to-face daily conversations, 1 (1.04%) TV program and 56 (58.33%) oral narratives. All the transcriptions are under two-round editing to maintain coding consistence. Due to lack of web maintenance technicians, the Hakka sub-corpus of NCCU is currently closed. However, 44 out of the 96 recordings are now accessible at the CABank of TalkBank, including both audio sounds and transcripts that are browsable and downloadable (http://talkbank.org/access/CABank/TaiwanHakka.html).

A total of 158 participants, 95 females and 63 males, participated. They are speakers of different TH varieties, contributing different sub-dialects of TH to the corpus – 84 (53%) Sixian speakers, 57 (36%) Hailu speakers, 11 (7%) Raoping speakers, 3 (2%) Dapu speakers, and 3 (2%) Sihai (a hybrid of Sixian and Hailu) speakers. These speakers’ occupations are quite diversified ranging from homemakers and teachers to employees in business and workers of industries. In terms of their gender and age, Table 39.8 gives the distribution of conversations (CA) and narratives (NA) across gender and age.
Table 39.8 reflects the concerns mentioned above: young speakers in the age range of 18–30 years old were more willing to participate in storytelling; in contrast, face-to-face conversations were mostly contributed by speakers over 46 years old. The observation implies the divergent levels of command of TH between younger speakers and older speakers. Older speakers that use more TH on a day-to-day basis have better command of TH and are more confident to do a one-hour face-to-face conversation. On the contrary, younger speakers do not use TH on a daily basis, but use it only occasionally for the purpose of communicating with their family elder members such as grandparents, and hence do not have as good TH language abilities as older speakers.

### An analysis of degree adverbs in the CA data

The Corpus of Spoken Hakka, encompassing different varieties of TH, has accumulated a considerable amount of valuable spoken data that can be utilized for teaching or researching. One of the most frequently used syntactic markers found in the CA data is degree intensifying adverbs, partly because intensification is usually employed for emotional or expressive manifestation in speech, and partly because TH has more monosyllabic adjectives and hence degree adverbs are often used to form a disyllabic compound. There are six degree intensifying adverbs that can be detected among the varieties of TH – 恒 [an2] (e.g., 恒大 [an2tai3] ‘very big’), 当 [dong1] (e.g., 当靓 [dang1jiang1] ‘very pretty’), 盖 [goi3] (e.g., 盖闹 [goi3nau3] ‘very pretty’), 尽 [qin3] (e.g., 尽坏 [qin3fai3] ‘very bad’), 异 [i3] (e.g., 异大 [i3tai3] ‘very big’, and 还 [han5] (e.g., 还臭 [hai5 cu2] ‘very smelly’). They all indicate a grammatical meaning of proportionally high degree roughly similar to that of 非 [hen3] in TM, or that of very in English. Based on 32 CA recordings, 1,466 tokens are found. The distribution of the six degree intensifying adverbs in an ascending order is: 还 [han5], 41 (2.8%), 异 [i3], 56 (3.8%), 盖 [goi3], 63 (4.3%), 尽 [qin3], 176 (12.0%), 当 [dong1], 540 (36.8%), and 恒 [an2], 590 (40.2%). Notice that the word 盖 [goi3] occurs only in the Southern Sixian Hakka. The other five degree adverbs can occur in the other five sub-dialects. In Hu (2017), detailed investigation of what can be modified by a degree intensifying adverb has been conducted. A degree intensifying adverb construction ([Degree Intensifying Adverb + X]) is examined regarding what X can be in terms of its syntactic categories and its grammatical functions. Degree intensifying adverbs can generally modify adjectives, adverbs or verbal predicates. One observation is that a host-expansion process seems to be undergoing whereby X is observed to expand from the most prototypical relative adjectives and stative verbs and psychological verbs to optative verbs and to common nouns. Examples (4) and (5) show the most prototypical usage of the six degree intensifiers modifying a relative adjective; the construction serves a predicative function in (4), but an attributive function in (5). Example (6) and (7) show expanded usages in which a verb and a common noun are being modified. In both cases, the degree construction serves a predicative function.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>CA</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18–30</td>
<td>31–45</td>
<td>46–60</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>20</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 39.8 The distribution of CA and NA across gender and age (N = 158)
(4) 这房间恁当/尽/盖/异还阔哦!
Ia2 fong5gien1 an2/dong1/qin3/goi3/i3/han5 fad4 o1!
This room very spacious PART
‘This room looks very spacious’.

(5) 这恁当/尽/盖/异还阔个房间无人歇。
Ia2 an2/dong1/qin3/goi3/i3/han5 fad4 NOM fong5gien1 NEG person live
‘Nobody lives in this spacious room’.

(6) 阿婆恁当/尽/盖/异/还惜大细。
A1po5 an2/dong1/qin3/goi3/i3/han5 xiag4 tai3se3
Grandmother very love family
‘Grandmother loves the family very much’.

(7) 阿婆头摆歇恁当/尽/盖/异/还乡下。
A1po3 teu5bai2 hed8 an2/dong1/qin3/goi3/i3/han5 hiong1ha3
Grandmother past live very countryside
‘Grandmother used to live in a very rural area’.

Although predicative and attributive functions together still receive the lion’s share, other grammatical functions such as subject (恁多荬仔拿来分猪仔食‘So many of the cabbages are given to feed pigs’), object (阿婆看着当多毋煞猛个后生人‘Grandmother has seen many young men that are not hard-working’), and complement (阿民读书尽庆‘Amin is very good at studying’) are also observed.

The observed linguistic behaviors are not uncommon as cross-linguistic studies have shown how adverbs can develop into more grammatical markers. However, two interesting issues arise if the observations are delved into more deeply. First, a correlation between vaster host class expansion and a higher distribution of the degree word in the corpus is revealed. Specifically, the two adverbs that enjoy the highest distribution – namely 当 [dong1] and 恁 [an2] (77% together) – perform better in terms of host class expansion, modifying a diversifying array of different syntactic constructions. Table 39.9 lists examples of 当 [dong1] constructions with its collocates ranging from the most typical monosyllabic, disyllabic adjectives and stative verbs to the least typical idiomatic verbal predicates and even nouns. Constructions of 当 [dong1] modifying monosyllabic and disyllabic adjectives are quite productive with their meanings compositionally derived; those with 当 [dong1] collocating with verbal predicates and nouns are less productive often carrying idiomatic meanings and expressing subjective opinions. Each example is given with a word-for-word translation and a free translation.

Second, a correlation between a higher degree of lexicalization and a higher distribution of both the degree adverb and the modified elements is also revealed. The two adverbs likewise display the most lexicalized patterns with fixed and idiomatic meanings among the six. Adding up the tokens of the modified adjectives of the two adverbs, namely an2 +AP and dong1 + AP, from Hu (2017: 47, Table 39.9) gives to 78.15% (948/1213) out of the six adverbs. The more frequently the two adverbs are used in daily discourse, the more likely the usages will become routinized, and hence gaining a higher degree of lexicalization. Examples such as 恁好[an2ho2]/当好[dong1ho2] ‘pretty good; quite good’, 恁大[an2tai3]/当大[dong1tai3] ‘pretty big; quite big’, 恁早[an2zo2]/当早[dong1zo2] ‘pretty early; quite early’, and 恁久[an2giu2]/当久[dong1giu2] ‘it’s been a while’ have lexicalized into quite fixed and idiomatic forms frequently used in daily life conversation. Take 恁早[an2zo2] for illustration.
On top of the literal meaning denoting ‘very early’, it has become a routinized expression for greeting in the morning similar to ‘good morning’ in English. Another usage of this expression is an extended metaphorical euphemism indicating ‘someone passed away earlier than expected’. Consider the following excerpted utterance adopted from Hu (2017: 95, (83)).

(8) F1: 无想 … 无想着佢恁早 (过身)
mo5 xiong2 mo5 xiong2do2 gi5 an2zo2 go3shinl
NEG think NEG think-DO he very early (pass away)
F1: ‘I am quite surprised that he passed away so early’.

In a similar vein, the expression 悠久 [an2giu2], in addition to the literal meaning denoting ‘very long’, has developed into indicating an idiomatic meaning ‘it’s been a while’. Consider the following excerpted utterance adopted from Hu (2017: 95, (82)).

(9) F2: 悠久 又无看着佢了
an2giu2 nga5 iu3 mo5 kon3do2 gi5 le1
very long I again NEG see-DO him PRT
F2: ‘It’s been a while since I saw him last time’.

The development of these expressions in conversational discourse accords with what have been shown in the extant literature. Studies have proved that frequency and repetition in language
use evidences in language change processes such as grammaticalization and lexicalization (Bybee and Hopper, 2001; Bybee, 2010), leading to fixed forms that are less analyzable. A linguistic chunk often emerges as an established pattern through routinization and entrenchment. Frequent word combinations can bring about the emergence of multiword constructions that in turn can grammaticalize with repetition (Bybee, 2010: 203). Furthermore, a consequent effect that follows from semantic and pragmatic idiomaticization is for the holistic form to express a subjective opinion in a discourse. Examine the following usages of the two expressions 当大 [dong1tai3] and 恁大 [an2tai3] shown in the excerpted dialogue adopted from Hu (2017: 99, (85)).

(10) F2: 旧年个风搓乜当大啊
 F1: 系..(8)该摆又当大呢
 F2: 有当大哦 该摆有[当大]哦
 F2: [这摆]乜恁大

F2: kiu3ngien5 ge3 fung1cai me3 dong1tai3 a3
F2: last year NOM typhoon also very huge PART

F1: he3 ge3 bai2 iu3 iu2 dong1tai3 hol
F1: COP that time again have very huge PART
F2: ia2 bai2
F2: this time

F1: ia2 dong1tai3 o2 ge3 bai2 iu3 iu2 dong1tai3 o2
F1: have very huge PART that time again have very huge PART
F2: ia2 bai2 me3 an2tai3
F2: this time also very huge

F2: ‘The typhoon was also very huge and intense last year!’
F1: ‘Was it? It (the typhoon) was huge that year?’
F2: ‘This year . . . ’
F1: ‘Very large and intense! It was very large and intense last year’. F2: ‘It was also a huge and intense typhoon this year’.

This excerpt has shown several repetitions of 当大 [dong1tai3] /恁大 [an2tai3] ‘very huge’ by both speakers. They are discussing the size and intensity of two typhoons happening in the current year and in the past years. They both express subjective opinions regarding how huge and intense the typhoon is this current year or was the previous year. The two degree words are used not only to intensify the largeness and intensity of the two typhoons but also to display the two speakers’ subjective stance as their discussion has been based on their impressions lodging in the memory. An interesting observation is the switching by speaker F2 between the two usages – first 当大 [dong1tai3] and second 恁大 [an2tai3]. The usages reveal a nuancing difference especially associated with 恁 [an2]. In addition to indicating intensification, 恁 [an2], which has the highest frequency, further develops into an indexical function in certain discourse contexts, as shown by the second usage of speaker F2, who particularly refers to the typhoon this current year. This subtle difference also manifests in the developments of certain
frozen and fixed lexicalized expressions of恁[an2] as in恁形[an2hin3],恁呢[an2ne5], and恁样形[an2ngion5hin5] ‘kind of; as such’ to indicate speaker’s attitude. These expressions cannot be replaced by 当[dong1].

Concluding remarks

This chapter first gives an overview of Taiwan Hakka regarding its varieties and distributions, the sound and morpho-syntactic features, and certain lexicalized forms that carry pragmatic and discourse functions uniquely found in TH. The declining status of TH and the revival measures in Taiwan have also been presented. Then, the construction of Taiwan Hakka Spoken Corpus is introduced, followed by a semantic and pragmatic analysis of degree words used in conversational discourse. Although linguistic studies of TH have increased in the recent decades, most of them focus on the sound systems, some on the syntactic and semantic structures, few on the pragmatic and discourse perspectives. With more and more TH corpus data collected, it is hoped that more studies can pay attention to the linguistic analyses of TH based on authentic data.

Moreover, Taiwan is a multilingual and multicultural society, in which language contact and language changes are actively happening. The various dialects of TH have exhibited various linguistic features influenced by TM or Taiwan Southern Min. Language use is not only contingent upon discourse participants but also is deeply rooted in the social-cultural contexts. Further investigation into the interaction of identity, ideology and society based on analysis of various layers of discourses in Taiwan is an endeavor worthy of carrying out.

Acknowledgments

This chapter is partly based on research projects granted by Ministry of Science and Technology (MOST 104–2010–H–004–002–MY2; 105–2410–H004–179–MY3). I thank all the research assistants who have collected and transcribed the Hakka spoken corpus.

References

English references

An introduction to Taiwan Hakka


**Chinese references**


Xu, Zhao-Quan 徐兆泉 (2009) 台湾客家话辞典 (*Hakka Dictionary of Taiwan*). Taipei: Nantian.