Advancing explanatory and tonal dialectometry

Sung, H.W.M.

CHAPTER 2

Traditional Classification of Yue-Pinghua Dialects

2.1 Introduction

Yue and Pinghua are two sub-branches of Sinitic from the Sino-Tibetan language family, which are spoken in Southern China and the diaspora communities in North America, the UK, Malaysia etc. (Wu 2012).

The existing classifications of Yue and Pinghua dialects have largely been proposed based on some phonetic features. More specifically, these features mainly consist of segments (i.e. consonants and vowels, opposed to supra-segments, or tones).

Early classifications provide only rough groupings of the dialects, of which many do not include classification maps. This is because before the 1980s, dialect data were scarce in the Yue-speaking area. This was especially the case for Pinghua dialects, as more extensive surveys were only published during the 2000s. We can see the impact from the lack of data in Yue-Hashimoto's (1988) classification, which utilises more than 100 features in her study, but due to the lack of data, many dialects were not taken into account.

Most of the classifications in Yue can be described as incomplete. Other than the lack of data for the earlier classifications to take into account, scholars also tended to focus on a particular province in which Yue is spoken. For example, in Zhan's (2002) introduction to Yue di-

alects, only dialects spoken in the Guangdong province were described. On the other hand, Xie (2007) focused on the dialects spoken in Guangxi. Although some dialects from Guangdong are mentioned by Xie for comparative purposes, the focus is still on Guangxi mainly.

We can speak of a more complete classification in the *Language Atlas of China* (*LAC*, Chinese Academy of Social Sciences (CASS) 2012), as well as Carlyle's (2020) study. LAC's classification divides the whole Yue-speaking area into seven groups, and Pingua, as a separate Sinitic language, which consists of two subgroups. Carlyle (2020) on the other hand uses dialectometric techniques to detect groupings within the Yue and Southern Pinghua dialect area. Both studies are not without problems, though, which are discussed in the Section 2.2 below.

In general, most of the existing classifications are often unclear in the choice and motivation of features used in the classification. It is possible that extralinguistic factors played a role in the scholars' judgements when drafting a classification scheme. This chapter offers an overview of the state-of-the-arts of Yue and Pinghua dialect classifications, which provides the background knowledge for the data-driven approaches to the classification problem of Yue and Pinghua in Chapter 5.

2.2 Traditional classification of Yue and Pinghua dialects

The classification of the dialects in Chinese dialectology is motivated by dialect descriptions and finding historical affiliation (Yue-Hashimoto 1988:7-9). In order to indicate how one group of dialects is different from another, a dialect classification serves the "convenience of accurately summarizing the characteristics of each dialect group" (Yue-Hashimoto 1988:7). Furthermore, it has been found that the historical background in the formation of these dialects can also be uncovered through classification, which "reveal[s] the interrelationship among dialects" (Yue-Hashimoto 1988:9).

2.2.1 Previous classifications of Yue and Pinghua

Zhan (1988) has discussed the methodology for the classification of Yue varieties. Zhan has acknowledged that it is difficult to have a clear-cut separation between dialect groups in the Yue-speaking area (in Guang-

dong). The main reason is that dialect groups are not homogeneous.[1] Zhan proposed using distinctive, characteristic features to separate each dialect group as the criteria for the groupings.

Yue-Hashimoto (1988:8-10) raised that classification of dialects relies heavily on geographical divisions (due to the scarcity of data), even though linguistic relationship and geography do not always coincide. Furthermore, the subjectivity problem was also identified. Depending on the choice of features, different scholars would produce different groupings of the same dialects.

Although the principles and pitfalls in dialect classification have been discussed, classifications in Yue seem not to follow Zhan's principle and Yue-Hashimoto's reflections closely. Before diving deeper into the problems different Yue classification schemes have, the following subsections will first give an overview of the major dialect classifications of Yue in the literature.

Yuan's (1960) and Zhan's (1981) classifications

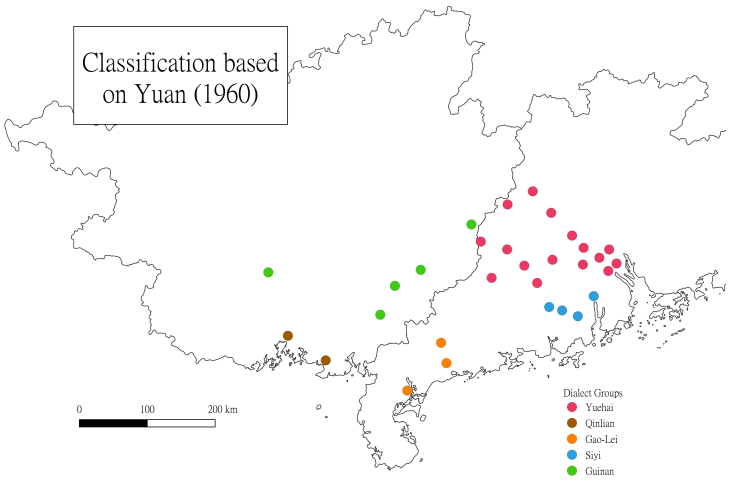

Figure 2.1: Map of Yuan's (1960) classification

Yuan (1960) divided dialects of Yue into five groups, namely 1) Yue-

[1]The exact wording in Zhan (1988:92-93) is "同中有異", which literally means 'there are differences within the likes' (self translation).

hai (粵海), 2) Qinlian (欽廉), 3) Gaolei (高雷), 4) Siyi (四邑) and 5) Guinan (桂南) dialects. Yuehai dialects include dialects spoken in the majority of the area in the Pearl River Delta (珠江三角洲) and along the Xijiang river (西江); Qinlian dialects and Gaolei dialects are found near the Qinzhou (欽州), Lianzhou (廉州) and Gaozhou (高州), Leizhou (雷州) areas respectively. Siyi dialects are found in Taishan (台山), Xinhui (新會), Kaiping (開平) areas and lastly, Guinan dialects are found in Southern Guangxi, including Wuzhou (梧州), Rongxian (容縣), Yulin (玉林), Bobai (博白). Yuan did not provide any criteria for his classification, and due to the lack of data, he also did not mention dialects spoken in the central-western part of Guangxi (Yu 2016: 99). This classification is mostly based on geographical areas, also due to the lack of data (Yue-Hashimoto 1991: 165).

Zhan's (1981) classification is basically the same as Yuan's, except he also listed some characteristic features for each group. Yue-Hashimoto (1991:166) believes Zhan's classification is not a new proposal, rather, it is providing more evidence for Yuan's classification.

Zhan's (2002) classification

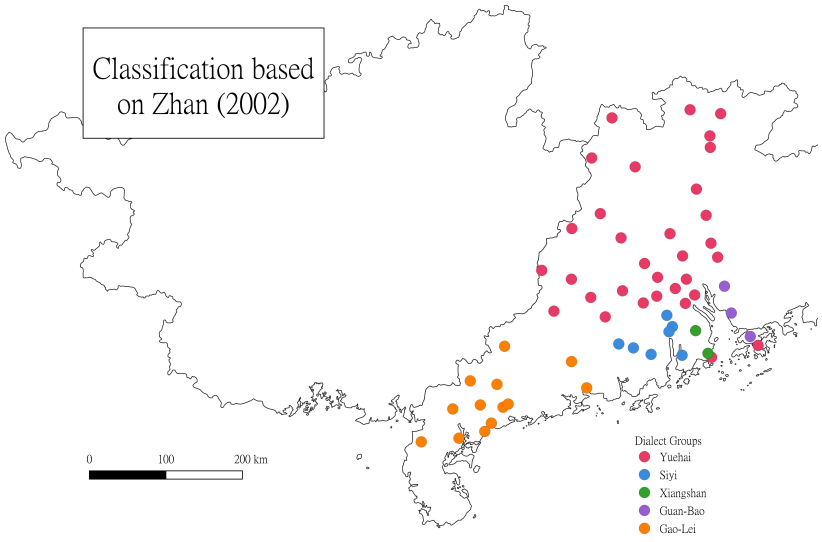

Figure 2.2: Map of Zhan's (2002) classification

After the publication of the major dialect surveys in Guangdong, Zhan (2002) proposed a Guangdong-specific dialect classification. It consists of five groups: 1) Yuehai/ Guangfu (廣府) dialects, 2) Siyi dialects, 3) Xiangshan (香山) dialects (found in Zhongshan (中山) and Zhuhai (珠海)), 4) Guan-Bao (莞寶) dialects (found in Dongguan (東莞), Shenzhen (深圳) and the New Territories in Hong Kong) and 5) Gaolei dialects (found in the Zhan-Mao (湛茂) area as well as Yangjiang (陽江), Yangchun (陽春)).

Zhan (2002) also provides a description of these dialect groups, some with more details, but there is no explicit explanation on how these groups were classified.

Li's (1994) classification

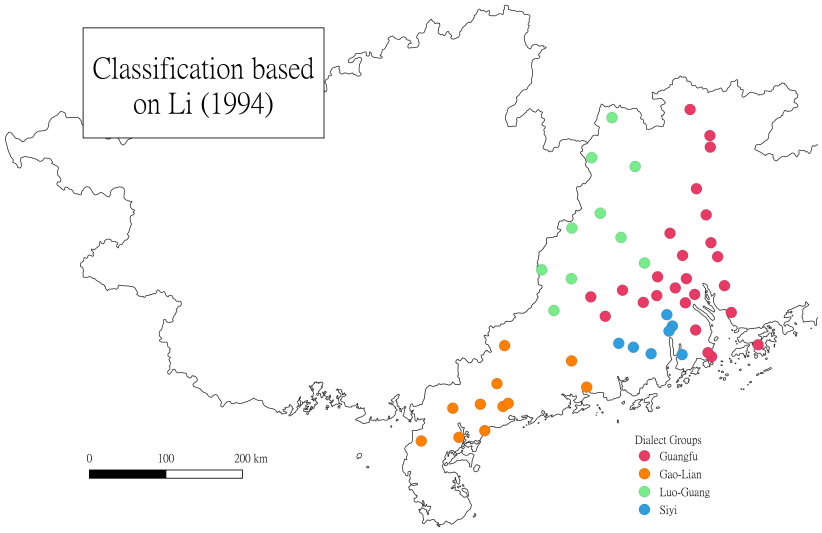

Figure 2.3: Map of Li's (1994) classification

Another classification similar to Zhan's can be found in Li (1994), which is also specific to Guangdong. He proposed that there can be a few groups of Yue dialects, including 1) Guangfu (largely the same as other classifications), 2) Gao-Lian (高廉, found in South-Western Guangdong), 3) Luo-Guang (羅廣, found in Central- and North-western Guangdong) and 4) Siyi dialects (largely the same as other classifica-

tions). Li also acknowledged that there are also Yue dialects spoken in Guangxi, though the authors did not specify whether they are part of any of the groups in Guangdong or not. Like the other classifications mentioned above, no criteria were specified for how these groupings were made.

Classification in the *Language Atlas of China (LAC, 1st and 2nd Edition)*

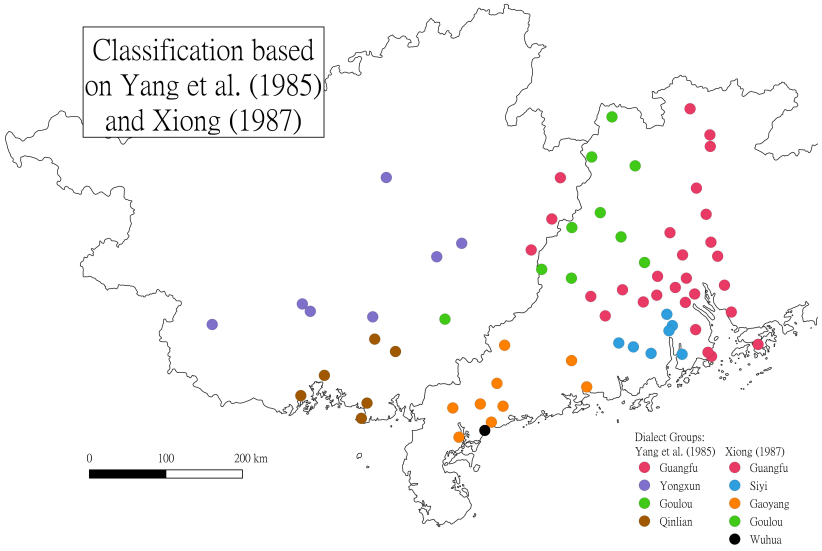

Figure 2.4: Map of Yang et al. (1985) and Xiong's (1987) classification

The classification of Yue dialects in Guangxi and Guangdong was elaborated in two separate articles, Yang et al. (1985) and Xiong (1987) respectively. Each article only discusses the classification for one province. These classifications have been adapted in the *Language Atlas of China* (1st Edition, Li et al. 1987).

Yang et al.'s (1985) classification divided the Yue dialects in Guangxi into four groups, namely 1) Guangfu (includes Urban Wuzhou (梧州), Cangwu (蒼梧), Hexian (賀縣)), 2) Yongxun (邕潯, includes Nanning (南寧), Liuzhou (柳州), Yongning (邕寧), Chongzuo (崇左), Hengxian (橫縣), Guiping (桂平), Pingnan (平南)), 3) Goulou (勾漏, includes Yulin and rural area of Wuzhou) and 4) Qinlian (includes Qinzhou,

Lianzhou, Pubei (浦北), Lingshan (靈山), Fancheng (防城) and Beihai (北海)). Yang et al.'s classification has a brief description of each dialect group by listing some features, but they did not explicitly define their criteria for grouping these dialects in this specific way.

Xiong's (1987) classification of Yue dialects in Guangdong, on the other hand, has a clear description of the classification procedures. In Guangdong, Yue dialects are divided into five groups, namely 1) Guangfu (mostly dialects in the Pearl River Delta, but also Northern areas such as Qujiang (曲江), Lechang (樂昌), Western areas like Yunfu (雲浮) and Dianbai (電白)), 2) Siyi (spoken in Heshan (鶴山), Xinhui (新會), Jiangmen (江門), Doumen (斗門), Enping (恩平), Kaiping (開平) and Taishan (台山)), 3) Gaoyang (found in the Zhan-Mao area, including Zhanjiang (湛江), Maoming (茂名), Gaozhou (高州), Xinyi (信宜) and the Liangyang (兩陽) areas, which include Yangjiang (陽江) and Yangchun (陽春)) and, 4) Goulou (spoken in the central-/ north-western part of Guangdong, including Sihui (四會), Guangning (廣寧), Deqing (德慶), Huaji (懷集), Yangshan (陽山)) and 5) Wuhua (吳化, distributed in Wuchuan (吳川), Huazhou (化州)). Wu (2007) later expanded on Xiong's classification, and added two features in order to account for the Yue dialects in Guangxi, yielding 7 dialect groups in total in both provinces. This classification is adapted in the 2nd edition of the LAC.[2] The methodology is identical with Xiong's, which is explained below.

In the LAC (both Xiong and Wu), dialects are classified through a process of elimination. Dialects are first classified into a certain group if they have a certain feature, and the remaining dialects which do not have the feature will then be assessed by the next feature. Once a group of dialects has been assigned to a group, it will not be assessed again by any following features, even if these features are present. This process iterates until all the criteria have been assessed. The presence and absence of the features for each dialect group are visualised in Figure 2.5, and the illustration of the elimination method is shown as a tree in Figure 2.6 (the presence of features are represented by the numbers on the branches). The first column in Figure 2.5 contains the five features which the LAC used for the classification, and the plus and minus symbols

[2]Wu (2007) is an article which is included and adapted in the LAC (2nd Ed.), featured as Wu (2012). Hence Wu (2007) and the LAC (Chinese Academy of Social Sciences (CASS) 2012) are used interchangeably when referring to Wu's classification.

indicate whether the dialect group has a certain feature or not (the ± in the first feature indicates partial possession of the feature, depending on the phonological context). The '+/-'symbol means that within the dialect group, some dialects have the feature, and some do not. Shaded cells indicate that the features are no longer considered.

	Wuhua	Goulou	Siyi	Yongxun	Qinlian	Gaoyang	Guangfu
1. MC* [+voi, -cont] > [-voi, +s.g.]	+	-	±	±	±	±	±
2. MC *t^h- > h-	-	-	+	-	-	-	-
3. Have contrastive [i] and [ɿ]	±	-	+/-	+	-	-	-
4. No Yin-Yang distinction for Shang/ Qu	-	-	-	-	+	-	-
5. MC *s- > [ɬ] or [θ]	+	+/-	+/-	+/-	+	+	-

Figure 2.5: LAC Classification Criteria (modified from Wu 2007)

As shown in Figures 2.5 and 2.6, the first feature used in the classification concerns the reflexes of Middle Chinese voiced stops and affricates (obstruents).[3] From Middle Chinese to present-day Yue dialects, the voiced obstruents went through devoicing, but the presence of aspiration is not homogenous among these dialects. Goulou dialects have been grouped together because their reflexes are unaspirated across all tone categories; the Wuhua dialects were separated from the rest of the dialects because their obstruent reflexes are aspirated across all tone categories. The remaining dialects only have aspirated obstruents if the word belongs to the Middle Chinese Ping 'level' and Shang 'rising' tone categories, otherwise, the obstruents are unaspirated. These remaining dialects are then assessed with a second criterion, namely based on the reflex of Middle Chinese *t^h-.[4] Dialects with a reflex of [h-] would be eliminated and grouped as Siyi dialects, and the remaining dialects are then assessed for: whether they have a contrast between [i] and [ɿ][5] (Feature 3), whether there is a Yin-Yang distinction for the Shang and Qu tones[6] (Feature 4) and lastly, whether they have a lateral or interdental fricative as a reflex of Middle Chinese *s-.[7] The presence of Feature 3 will make the dialects part of the Yongxun dialect group; a dialect with

[3]Original text: "全濁聲母清化後今送氣" (Wu 2007:170).
[4]Original text: "透母讀 [h-]".
[5]Original text: "止攝開口韻精莊、知章兩分".
[6]Original text: "上聲、去聲分別只有一類".
[7]Original text: "心母讀邊擦音或齒間擦音".

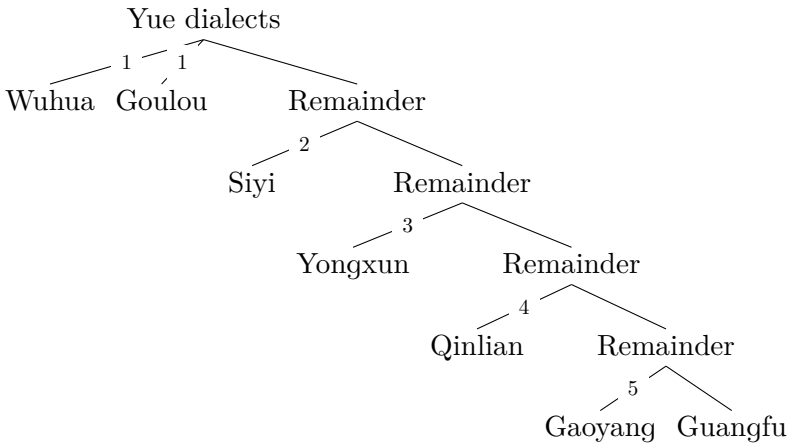

Figure 2.6: LAC Classification of Yue dialects (Tree representation), based on Wu (2007)

Feature 4 will be considered as a Qinlian dialect and lastly, if a dialect has Feature 5, then it is a Gaoyang dialect. Otherwise, if the dialect does not have Feature 2-4, it is a Guangfu dialect.

Yue-Hashimoto's (2006) classification

Yue-Hashimoto (1988, 1991) have taken a different approach to the dialect classification of Yue from the previous proposals. Instead of focusing on a handful of features, she proposed to use "as many features as possible... to get a general picture of what features are shared by which dialects"(Yue-Hashimoto 1988:14). In this classification, features from the phonological and lexical levels were considered. Morphological and syntactic features were excluded due to the paucity of information in the studied areas. Another major difference is the use of isoglosses. Yue-Hashimoto argued that the isogloss method and the calculation of degree of similarity should yield the same results; a preference on the isogloss method is supported by its visual representation. Even though isoglosses are often interweaving with each other, they represent the real complex situation which linguists should "face and not escape from" (Yue-Hashimoto 1988:14). However, Yue-Hashimoto's classification suffered from the lack of data, and could not cover the entire Yue-speaking area.

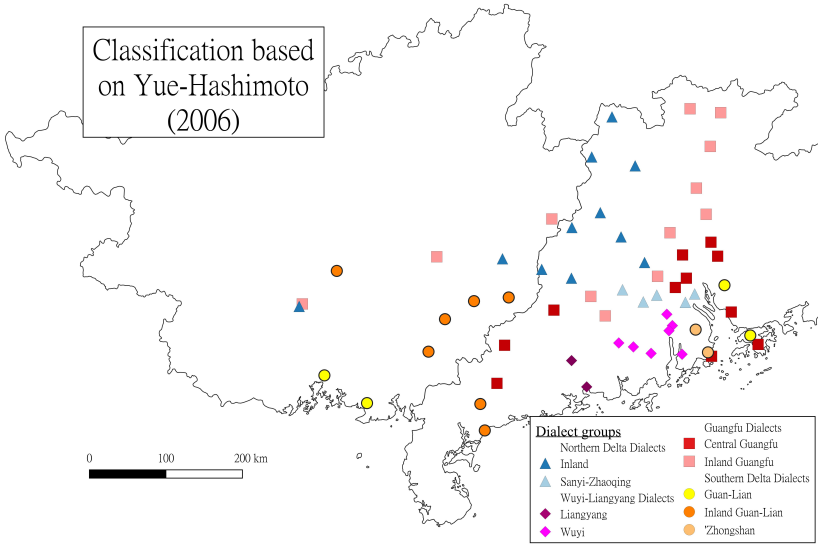

Figure 2.7: Map of Yue-Hashimoto's (2006) classification

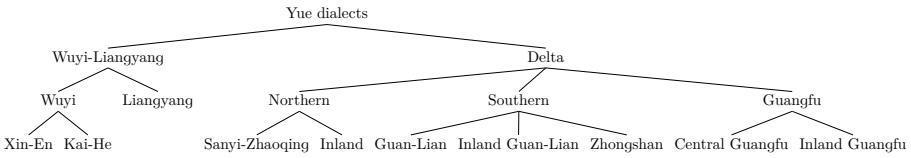

Figure 2.8: Yue-Hashimoto's (2006) Classification (Tree representation)

Yue-Hashimoto's classification has since been updated, as there were several dialect surveys published after her previous classifications. A later, revised version of her classification can be found in Yue-Hashimoto (2006). The dialect data come from the later published dialect surveys, as well as numerous unpublished materials, including Chao Yuan-Ren's fieldwork data in Guangdong and Guangxi between 1928 and 1929. This classification has a wider geographical coverage than her previous classification. It includes northern Guangdong, for instance, which was largely missing before. It is, however, unclear what Yue-Hashimoto's revisions were based on, as only the updated classification was given, and nothing like the previous isogloss analysis was provided.

Yue-Hashimoto's classification divides Yue into two main dialect ar-

eas, namely Wuyi-Liangyang and Delta dialects, and each group has their sub-dialect groups. Wuyi-Liangyang dialects are further divided into Wuyi (五邑) dialects (Siyi dialects in the LAC) and Liangyang dialects (dialects spoken in Yangjiang and Yangchun). The Delta dialects consist of the Northern Delta dialects, spoken in the Sanyi (三邑) area in Guangdong as well as the border region between Guangdong and Guangxi (and Nanning Tingzi (南寧亭子)). Southern Delta dialects consists of mainly dialects which Zhan et al. (2002) classifies as Guan-Bao, Xiangshan and Gao-Lei dialects, together with Rongxian (容縣), Cenxi (岑溪), Bobai (博白), Yulin (鬱林), Shinan (石南), Binyang (賓陽) in Guangxi. Lastly, the third sub-group within the Delta dialects are the Guangfu dialects, which are dialects that share a large degree of similarity with the Guangzhou dialect, including some (LAC) Yongxun dialects. Yue-Hashimoto's dialect classification can be found in Figure 2.7 and 2.8.

Xie's (2007) classification

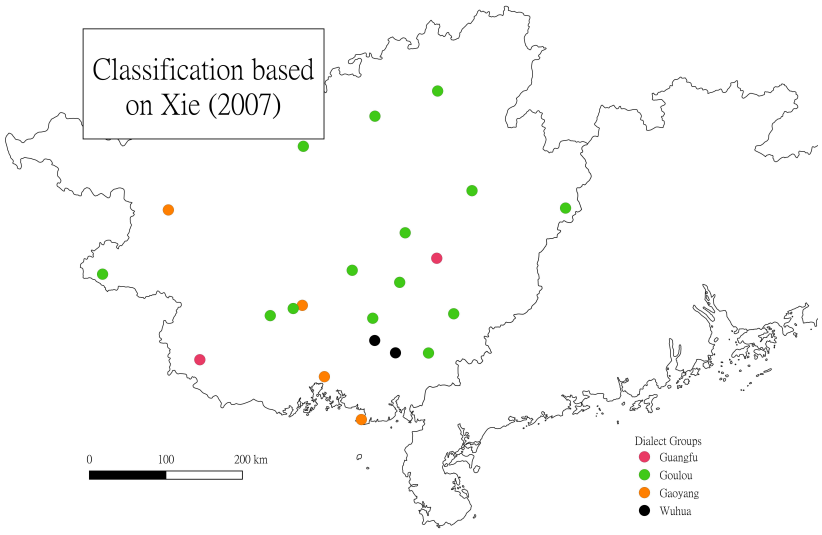

Figure 2.9: Map of Xie's (2007) classification

Xie (2007) has assessed both similarities and differences between Yue varieties in Guangdong and Guangxi, as well as differences between Yue

varieties in Guangxi. Xie (2007:174) noticed the separation in the studies of Yue dialects in the two provinces and argued for a unified classification of the whole Yue-speaking area, unlike previous classifications.

When applying Xiong's criteria to the Guangxi dialect data, Xie did not agree with the classification they yielded, because mutually unintelligible dialects were grouped together (e.g. Mengshan and Bose dialects were grouped as Goulou) and mutually intelligible dialects were assigned to different groups (e.g. Urban Nanning and Ningming dialects were assigned to Gaoyang and Guangfu dialect groups respectively). In order to make the classification of Guangxi dialects more coherent with the dialects in Guangdong, Xie added two criteria which help to identify misclassified Goulou dialects according to Xiong's criteria.

Xie's classification procedure starts with Xiong's list of features, and he applies the same elimination method as Xiong (1987) and Wu (2007). In addition to Xiong's criteria, Xie added two additional criteria, i) whether there is a stop coda [-t] for the word 鼻 'nose' and ii) having two tone reflexes for the Middle Chinese *Yang Ru* tone category.[8] The additional criteria does not yield new dialect groups, but they put certain dialects into the Goulou dialect group.

Xie (2007) has kept the same 5-group division of Yue as Xiong (1987) in his classification (see Figure 2.9). Since these groups were originally only applied on dialects in Guangdong, the additional criteria better reflect Xie's understanding of the relationship between the dialects in Guangxi.

Sung's (2020) classification

Sung (2020) focuses on one feature, the reflexes of Middle Chinese *ɣu- in this classification. Using dialectometric approaches (Relative Distance Value and multidimensional scaling, see Section 3.3), Sung discovered three geographical patterns regarding their reflexes of *ɣu-, namely 1) loss of *ɣ- in all phonological environments, 2) loss of *ɣ- in all environments except words with *-u# (which have the [f-] reflex) and 3) loss of *ɣ- in all environments, and only some words with *-u# retained the [f-] reflex. These three types of reflex patterns are found in the Inland, Coastal and Siyi dialect groups respectively in Figure 2.10. Sung further explored the origin of the geographical patterns of the reflexes, and discovered a very complicated interaction between two competing sound

[8] Original text: "以陽入讀兩個調值為特殊條件" (Xie 2007:188).

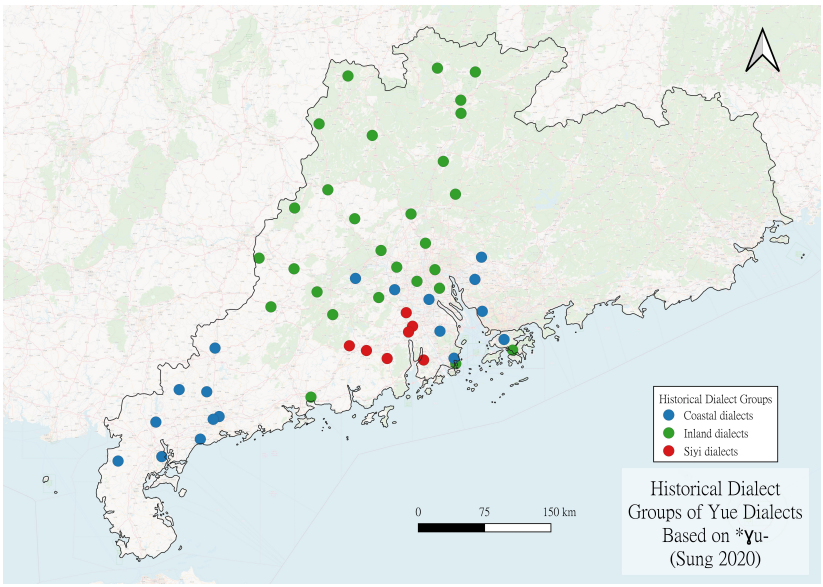

Figure 2.10: Map of Sung's (2020) classification of historical Yue dialects based on the reflexes of *ɣu-.

changes, namely ɣ-deletion and o-raising. ɣ-deletion was found before high vowels (like *u), and the difference between Coastal dialects and Inland dialects differ in when they had o-raising. Inland dialects already had o-raising (> -u#) and that allowed ɣ-deletion to apply to all the words with historical *-o. However, with the Coastal dialects, o-raising did not reach the area when ɣ-deletion was in operation, which then led to the preservation of *ɣ- and later became an [f-] reflex (<*h-), as a result of labio-dentalisation. Based on the historical changes described above, Sung reconstructed the following: historically, there were two dialect areas, Inland vs. Coastal, and they differ in the consequences of the timing of o-raising. This explains why two coastal regions (near Dongguan and Zhongshan in the East and the Zhan-Mao area in the West) share the same reflexes. The Siyi region is believed to be part of the Coastal dialect region in the past too, as it still has traces of the [f-] reflexes. Sung hypothesised that there could be a population inflex from the Inland dialect area, which caused a disruption of the Coastal dialect pattern (i.e. separating the coastal dialect group into two areas and showing a mixed pattern for the [f-] reflex).

Carlyle's (2020) classification

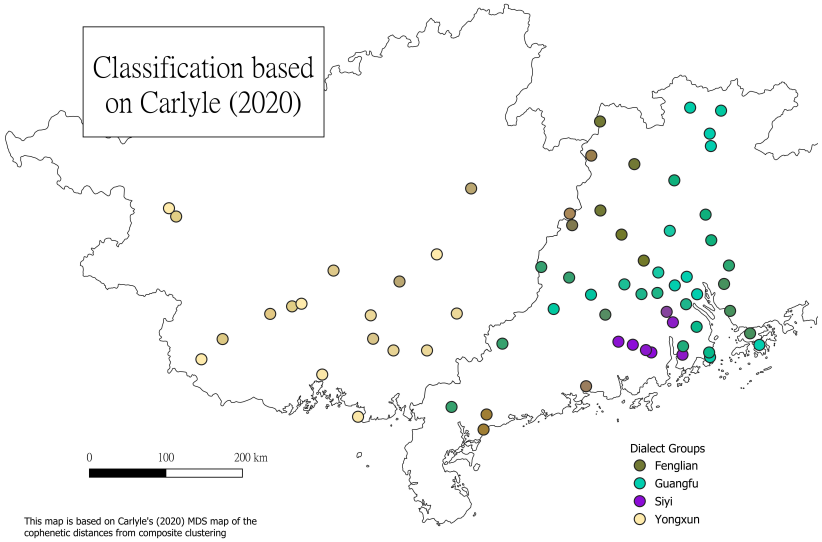

Figure 2.11: Map of Carlyle's (2020) classification (replicated)

Carlyle (2020) offers a dialectometric analysis of Yue and Pinghua dialects. Before the analysis, Carlyle first eliminated 'non-Yue' dialects by using 9 proposed features which are considered as "diagnostic criteria that are generally thought of as unique characteristics of Yue dialects" (Carlyle 2020:32). Whenever a dialect possesses fewer than 5/9 of these features, it is not considered to be a dialect of Yue. The purpose of this procedure is to filter dialects which are not considered to be part of the dialect family. As a result, a number of traditional Pinghua and Tuhua varieties (as well as some other Sinitic varieties) were eliminated from the main analysis.

For the classification procedure, Carlyle calculated the *lexical distance* (Nerbonne and Kleiweg 2003) between the dialects by comparing a total of 77 words over 70 dialects. 'Lexical distance' here refers to a lexico-phonetic distance, where Levenshtein distances (see Section 3.3.1) were calculated from both cognates and non-cognates.[9] Instead of using monosyllables as recorded in most dialect surveys, Carlyle used both

[9]Kessler (1995) refers to this distance as *all-word distance*, in contrast to *same-word distances*, where only cognates are considered.

mono- and poly-syllabic words. Furthermore, this analysis also include tones. In the data, tones are transcribed in Chao's (1930) tone letters. These tones were treated as strings just like the segments in the analysis.

Carlyle's (2020) classification consists of two large regions, Eastern and Western, and within the Eastern region, there are three dialect groups, namely 1) Fenglian (封連), 2) Guangfu and 3) Siyi dialects, and within the Western group, the Yongxun dialects. There are also five unclassified dialects in the analysis, due to the sparse documentation in some regions. These dialects include, Wuchuan, Huazhou, Yangjiang, Xindu, Mengshan. Carlyle's map can be found in Figure 2.11.

The status of Pinghua

The status of Pinghua as a major branch of Sinitic languages has been controversial since its first proposal. In the 1980s, Pinghua was classified as a separate Sinitic language in the *Language Atlas of China* (*LAC*, Li et al. 1987). Since then, scholars have been debating whether Pinghua belongs to Yue or not. A group of scholars (represented by Zhang (1982), Liang and Zhang (1999), Wei (1996), Li (2000), cited in Tan (2012)) focused more on the differences of features between Pinghua and Yue, and they argued that Pinghua should be viewed as independent from Yue. Their opponents (represented by Wu (2001), Tan (2000) and Liang (1997), cited in Tan (2012)) examined several features and they mostly argued that a sub-branch of Pinghua, Guinan (Southern) Pinghua, is actually very similar to Yue, and therefore should be classified under Yue. Liang (1997) made an even more radical claim that the entire Pinghua branch should merge with Yue, based on the examination of seven features. The controversy of the relationship between Pinghua and Yue remains.

In Tan's (2012) introduction to Pinghua varieties (published in the LAC (2nd edition, Chinese Academy of Social Sciences (CASS) 2012)), Pinghua is split into two sub-groups, namely Guinan (Southern) and Guibei (Northern) Pinghua. Three features have been proposed to distinguish Guinan Pinghua from other varieties. Feature 1 is the possession of unaspirated obstruents across all Middle Chinese tone categories as the reflexes of the Middle Chinese voiced obstruents. This feature is used to distinguish Guinan Pinghua from 'Baihua' (roughly Yongxun and Guangfu) dialects. Feature 2 is the possession of only one series of

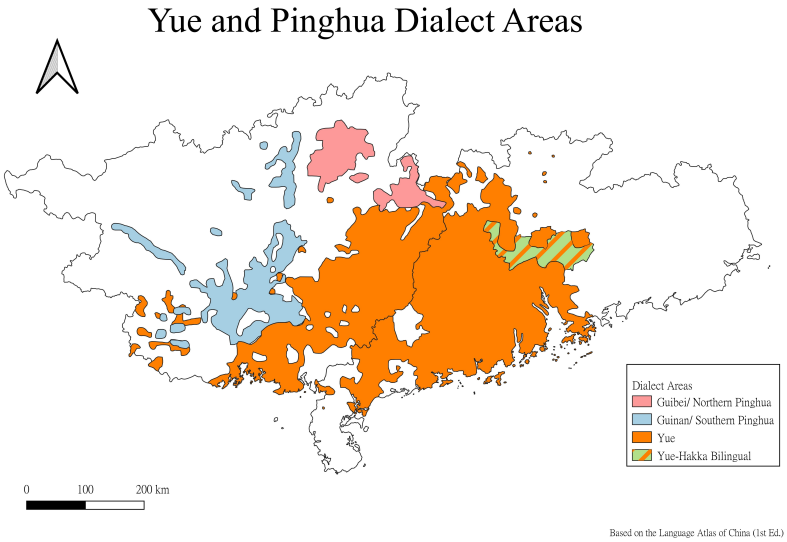

Figure 2.12: Geographical distribution of Yue, Guinan and Guibei Pinghua dialects (based on LAC 1st Edition)

affricates.[10] This criterion was set to distinguish Guinan Pinghua from neighbouring Goulou dialects. The last feature is the possession of [-p, -t, -k] codas. This feature can separate Guinan and Guibei Pinghua varieties.

The geographical distribution of Yue, Guinan Pinghua and Guibei Pinghua can be found in Figure 2.12.

2.2.2 Limitations to the existing classifications

Early classifications of Yue dialects did not provide explicit criteria and motivations for the groupings. Classifications from e.g. Yuan (1960) and Zhan (2002) may not be pure linguistic classification of Yue dialects, but simply divisions based on geographical areas (Yuan 1960) or some personal judgements (Zhan 2002). Yue-Hashimoto (1988) use more than 100 features in her classification. However, these proposals suffered from the lack of data in certain regions, which yielded incomplete classifications. Her later 2006 classification had a wider coverage, but the linguistic jus-

[10] Original text: "精、清、從、邪與知、莊、章合併為一套塞擦音".

tifications of the classification were not provided. Lastly, a number of classifications only focused on one province, e.g. Zhan (2002). Dialects in Guangxi are often missing in these classifications.

Yue-Hashimoto's (2006) classification

Yue-Hashimoto's revised classification has filled in the geographical areas which were not accounted for in her earlier 1988 classification. However, her revision did not provide any justification for revisions she made (Yue-Hashimoto (2006) was an article for reconstruction), and the author did not publish any further articles discussing the matter. The lack of transparency puts Yue-Hashimoto's classification into the same group of classifications as Yuan (1960) and Zhan (2002).

Xie's (2007) classification

Xie's (2007) classification attempted to account for the Yue dialects spoken in both provinces. However, Xie kept the 5-group division from Xiong (which was for Guangdong Yue dialects only), and then added two additional criteria so that some 'misclassified' Guangxi dialects according to Xie can be reassigned to the 'correct' dialect group, namely the Goulou dialect group. In this classification, Xie did not propose a new classification from Xiong's by proposing a revised set of features as criteria. Instead, the purpose of these additional criteria is to make what he considered as 'misclassified' dialects (under Xiong's criteria) to be reclassified as members of the 'correct' dialect group (based on Xie's (2007:189) own perception). Xie's classification hence adds further subjectivity to Xiong's classification.

The LAC (2nd Edition) classification

The method used in the LAC is problematic, because this method of dialect classification is not based on exclusive features, unlike what Zhan (1988) proposed. This can be seen through the lack of exclusive features for the Guangfu dialect group in Figure 2.5. In addition, the elimination method is dubious. As mentioned in the previous section, once the dialects have been assigned to a group, the later criteria will not matter anymore (indicated by the gray cells in Figure 2.5), even though some of these features are shared with other dialect groups (see Feature 3 and 5). Hence, the methodology appears to be stipulative. There is no

explanation to why dialects have to be assessed by a certain order of criteria, and not others. This also means that if one changes the order of the criteria, the result would be drastically different This is illustrated through the example below. The presence and absence of features are shown in Figure 2.13 (and the tree diagram in Figure 2.14 to illustrate the elimination method). Here, the original fifth feature is now the second feature in the classification, and the result is already very different from the original LAC classification. We can see that the Siyi dialects (in green) and the Yongxun dialects (in orange) are now split into two groups, since only some of the dialects possess the lateral fricatives in these two dialect groups. This feature also groups Qinlian and Gaoyang dialects (in addition to Siyi and Yongxun dialects) as one group. Furthermore, Feature 4 is now redundant in the classification. Based on one change, the membership of dialect groups and the total number of dialect groups have already changed drastically. This shows how unstable and potentially unreliable this classification method is, especially when there is no explanation why dialects should be classified this way. In Wu (2012), the author provided a revision of the classification found in Wu (2007). Both classifications return the same divisions of dialects, even though the order of features in the assessment is different (neither the ones shown in Figure 2.5 and 2.13). Despite the change, the LAC or Wu (2012) somehow gives the same classification as Wu (2007), even though if you follow the order of elimination, one sould yield a different classification as the one stated in the atlas.

	Wuhua	Goulou	Siyi (Taishan, Kaiping)	Yongxun (Nanning, Ningming)	Qinlian	Gaoyang	Siyi (Enping, Jiangmen, Doumen)	Yongxun (Pingnan)	Guangfu
1. MC* [+voi, -cont] > [-voi, +s.g.]	+	-	±	±	±	±	±	±	±
5. MC *s- > [ɬ] or [θ]	+	+/-	+	+	+	+	-	-	-
2. MC *t^h- > h-	-	-	+	-	-	-	+	-	-
3. Have contrastive [i] and [ɿ]	±	-	+/-	+	-	-	+/-	+	-
4. No Yin-Yang distinction for Shang/ Qu	-	-	-	-	+	-	-	-	-

Figure 2.13: LAC Classification Criteria with a Different Criteria Order

Despite the problems listed above, Wu's (2007, 2012) classification is still the most widely used classification of Yue in Chinese dialectology. In addition, the maps also indicate where each dialect group is roughly

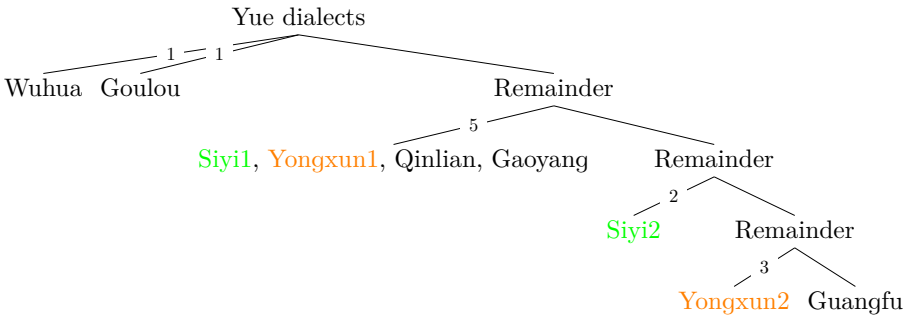

Figure 2.14: LAC Classification of Yue dialects with a different feature order (Tree representation)

distributed. This element is lacking in Xie's (2007) classification.

Carlyle's (2020) classification

While Carlyle (2020) utilises computational methods to seek partitions of Yue dialects, this procedure did not cover the dialects that were filtered at the start by the 9 features. The choice of these features was unexplained, and this is contradictory with the computational methods used later on, since the use of computational methods is meant to provide a more objective view in the relationship of dialects. Filtering dialects with these manually chosen features does not help in addressing the controversy over Pinghua and Yue, which means it remains unresolved. The controversy will be revisited in Chapter 5.

Furthermore, Carlyle's analysis combines tones and segments together. While it is useful to look at both tones and segments in one single analysis, Carlyle's classification has one problem: choices on the dialectometrical methods of tones have not been tested and compared systematically on many closely-related dialects. It is unclear whether Carlyle's application of Levenshtein distance on Chao's (1930) tone letters yield a linguistic-coherent distance. These problems are addressed in Chapter 7.

Other than the validity of the tone distance metric, since Carlyle combined both tones and segments together, the nature of tonal variation across space and its correlation with segments remain unexplored (though it was not the goal of his dissertation). Chapter 8 explores this aspect of tonal variation in greater depth.

The status of Pinghua

One of the reasons why the status of Pinghua is controversial is that each scholar has focused on a different set of features, some on the differences (e.g. the reflex of Middle Chinese *k^h- is often a [k^h-] in Pinghua, but [h-] in Yue, according to Li (2000:36)), while others on the similarities (e.g. the retention of codas -p, -t, -k, according to Wu (2001:137)). Furthermore, the criteria proposed by Tan (2012) on Guinan Pinghua has a strange rationale. Instead of finding exclusive features, he proposed features which are useful in separating the group from 1) Guangfu or Yongxun dialects, 2) Goulou dialects and lastly 3) Guibei Pinghua. These features tell us nothing that is unique to the Guinan Pinghua, as Feature 1 is shared with Goulou dialects, Feature 2 is shared with Guangfu dialects and Feature 3 is also shared with many Guangfu dialects. These criteria give an impression that the author simply wants to separate this group of varieties (perhaps based on the author's judgements), but no concrete evidence is given (e.g. using distinctive features, cf. Zhan (1988)).

Reference classification of Yue throughout the dissertation

In the following chapters, the classification of Yue will be re-assessed using dialectometric methods. The quantitative classification will be compared to a traditional classification.

The LAC classification has the following properties: 1) the LAC is widely used and is often cited in the literature; 2) it provides maps which indicate the approximate distribution of the dialect areas; 3) it covers the entire Yue-speaking area; 4) while the features are possibly subjectively selected, they are all clearly stated (useful for assessing the criteria used) and 5) unlike Xie's classification, there are no further manipulations of the groupings after the elimination procedures.

For the reasons above, the LAC classification is preferred as a reference classification from traditional Yue dialectology throughout the dissertation. Carlyle's (2020) classification is not a traditional classification, hence, it is not considered as a reference classification.