

Literary Disyllabism vs. Colloquial Trisyllabification in Taiwanese*

1. Introduction

The argument in this paper is simply that Taiwanese Hokkien (hereafter Taiwanese) probably has already entered the stage of trisyllabism and that the process of trisyllabification is being hindered by the disyllabism of borrowed forms from literary Chinese and literary Japanese, most of which are written in two characters.

The first point that Taiwanese has become trisyllabic will be presented from three respects. First, Taiwanese phonemic inventory and syllable inventory are compared with those in other languages. The comparison shows that Taiwanese has a similar number of contrastive sounds as many polysyllabic languages do and that Taiwanese has much less syllables than polysyllabic English. Furthermore, calculations of the syllable types and monosyllabic lexical items in Taiwanese and some other languages indicate that languages leave excessive number of redundant gaps open. Instead of utilizing the empty slots for new monosyllabic lexical items, speakers form disyllabic and longer lexical items by reusing occupied slots. It seems that there is an anticipation that new lexical items could occur someday and occupy these empty slots. In order to avoid the occurrence of homophones in these slots in the future, it would be better to keep the slots empty now. It happened that even occupied slots were emptied out, as illustrated in §2.2. The anticipation also partially explains the occurrences of many “unnecessary” trisyllabic lexical items in Taiwanese. In other words, many Taiwanese disyllabic lexical items are trisyllabified so that the constituent disyllabic phonemic word slots are left open.¹ Finally, data show that a large number of older place names in Taiwan were trisyllabic and that there are many trisyllabic common lexical items pairing with synonymous disyllabic lexical items, such as *m̄ng-kháu* and *m̄ng-kha-kháu*, both meaning ‘doorway’ (cf. Figure 6).

To show that written Chinese and written Japanese has been hindering Taiwanese trisyllabification, three sets of examples are given, which are the place names and the doublets mentioned above, and competing synonyms. Most of the trisyllabic place names were either shortened or changed to two characters each in the written forms by the Chinese and the Japanese administrators in conformity with their literary traditions. The chronology of the trisyllabic and disyllabic doublets of common words is mostly unknown.² Either they were trisyllabic in the beginning and shortened to become disyllabic under the pressure of the literary word pattern, or they are lengthened from disyllabic to trisyllabic so as to create more disyllabic gaps, but their original disyllabic counterparts are unfortunately sustained by the literary pattern. In either case, the impact of literary disyllabism is felt. The last set of examples are the competing synonyms between older trisyllabic lexical items and newer disyllabic ones, the latter being primarily from written Chinese (cf. Figure 7).

The hindrance of Taiwanese from freely becoming trisyllabic entails the accumulation of disyllabic lexical items. The accumulation, in turn, entails more and more disyllabic

*An earlier version of this paper was presented at the 27th International Conference on Sino-Tibetan Languages and Linguistics, 12–14 October 1994, Paris. Subsequent revisions were made after giving two lectures on the same subject in Tâi-pak and Tâi-tiong to concerned groups in Taiwanese language and culture. [編按: 本文原載於 *National Taiwan University Working Papers in Linguistics*, Dec 1998, 1:1–19.]

¹The practice of slot reservation is not always desirable, for it often makes other slots more crowded, as can be seen in §2.

²The doublets whose chronology is known are not cases of lengthening or shortening. See §3.3 for this category of lexical items.

homophones. In languages using characters or used to be using characters as their written media, such as Taiwanese, Chinese, Vietnamese, Japanese, and Korean, the great majority of homophones are literary forms (cf. Figure 10). The revival of Taiwanese language and writing in recent years witnesses the siphoning in of written Chinese words at a high speed. More and more disyllabic homophones are expected as a result of the influx.

2. Trisyllabification in Taiwanese

A calculation of the distribution of Taiwanese lexical items according to word length like Figure 1 below indicates that Taiwanese lexical items are predominantly disyllabic.³

(1) Distribution of Taiwanese Lexical Items According to Syllable Number

Monosyllabic	20.42%
Disyllabic	68.84%
Trisyllabic	10.12%
Quadrisyllabic	0.55%
Quintisyllabic	0.07%

The comparatively low percentage of the trisyllabic lexical items is due to the dictionary structure of the source (Murakami 1983), which ignores many common trisyllabic compounds with a “unit”.⁴ It is expected that as the lexicon expands, the percentage of monosyllabics and disyllabics will decrease, and that of trisyllabics will increase. To explore the possibility that Taiwanese trisyllabification is in progress, two more kinds of calculations are made and presented in §2.1 and §2.2. These are a calculation of contrastive sounds and a calculation of possible phonemic words and the actual occurrences of lexical items. In addition, §2.3 gives examples of two categories of trisyllabic lexical items, place names and doublets, to illustrate the claim in this section. The place names are the most convincing evidence.

2.1 Phonemic Inventory and Syllable Types

Since the number of phonemic inventory in a language can serve as an indicator of the canonic lexical length, Tiuⁿ J. (1995a) provides a list of phonemic inventories of the first 100 languages in Ruhlen (1976:153–173). It is repeated as follows.⁵

³Almost all the percentages in this paper are estimations. It is not easy to count syllables that actually occur, for different speakers have very different inventories – at least this is the case in Taiwanese, and different dictionaries collect different words. The following calculation is based on Murakami (1983). The ratio between disyllabic words and monosyllabic words in our current calculation is 1:3.37, but it is 1:2.36 in Ñg (1988).

⁴For instance, *kàu-ibk-pō* ‘ministry of education’ is a compound of the base *kàu-ibk* ‘education’ and a “unit” *pō*; meaning ‘ministry’ in the current context. *Kàu-ibk-pō*; incidentally, is not entered in Murakami (1983).

⁵The manuscript of this paper was written before that of Tiuⁿ J. (1995a), but the latter was published first. As a result, there is also a reference to the conference version of this paper concerning the following table.

(2) Phonemic Inventory in 100 Languages

Phonemes	Languages	Counter	Phonemes	Languages	Counter
70	1	001	37	3	044
69	1	002	36	3	047
65	1	003	35	2	049
64	1	004	34	2	051
62	1	005	33	1	052
61	1	006	32	3	055
60	1	007	31	5	060
59	2	009	30	3	063
58	2	011	29	3	066
57	2	013	28	3	069
55	2	015	27	4	073
54	2	017	26	3	076
50	2	019	25	1	077
47	1	020	24	2	079
46	2	022	22	5	084
43	1	023	21	4	088
42	7	030	20	3	091
41	1	031	19	2	093
40	2	033	18	5	098
39	3	036	17	1	099
38	5	041	16	1	100

The purpose of the list is to compare Taiwanese with other languages so as to decide whether Taiwanese ought to be considered monosyllabic, disyllabic, or polysyllabic according to its phonemic inventory. Taiwanese phonological common core has 31 contrastive sounds, and its overall pattern has 40. The majority of the dialects have 35 phonemes. If the 42 ranks of phoneme numbers in (2) are evenly divided into 6 groups, with 7 ranks in each group, Taiwanese falls into the fourth group. In the same group where Taiwanese belong, there are Swahili (31 phonemes), Malay (32 phonemes) and English (37 phonemes). Modern English is polysyllabic. Malay is commonly believed to be a disyllabic language, but Malay verbs often have to be affixed in order to function in a sentence, and a large portion of Malay nouns are polysyllabic.⁶ Swahili is basically a trisyllabic language, as shown by the following calculation, which is based on Morino (1975) and is another reproduction from Tiuⁿ J. (1995a).

(3) Syllable Numbers of Swahili Common Words

	Disyllabic	Trisyllabic	Quadrisyllabic	Quintisyllabic
Words	63	101	16	3

The comparison of Taiwanese with these languages may not be convincing that Taiwanese is a trisyllabic language. However, it shows that Taiwanese, in terms of its phonemic inventory, falls in the same group where trisyllabic languages and languages in the process of trisyllabification belong.

Another indicator that can be used is the number of syllable types that have actual monosyllabic lexical items (i.e., excluding accidental gaps). For instance, English has at least 3,765 monosyllabic words by type (i.e., phonemic words), excluding plural forms and forms in the past tense, whereas Taiwanese has about 2,000. If a language like English that is rich in actual syllable types (still not counting accidental gaps) has to become polysyllabic, for a

⁶If one wants to argue that Malay is disyllabic on the ground that its word bases are predominantly disyllabic, one has to argue on the same ground that Sinitic languages are still monosyllabic, which is not true.

language like Taiwanese that is not as rich in the same respect to become polysyllabic is very natural. The following figure is a comparison of monosyllabic words by type and monosyllabic words by token (i.e., lexical items) in English and Taiwanese.

(4) Monosyllabic Lexical Items in English and Taiwanese⁷

	Word Types	Word Tokens	Homophone Types	Homophone Tokens
English	3,765	4,104	945	1,284
Taiwanese	1,754	2,650	645	1,541

According to this figure, English has 25.10% monosyllabic homophones by type and 31.29% by token, whereas Taiwanese has 36.77% by type and 58.15% by token. If homophones are a motivation to polysyllabify, Taiwanese has more reasons than English to be considered a polysyllabic language.

2.2 Accidental Gaps

In my variety of Taiwanese, there are 4152 possible syllables, excluding syllables with non-historical tones, which are ignored by all lexicographers to date. This figure is similar to that in most varieties of Taiwanese, i.e., around 4,000. Of these 4,000 or so monosyllabic phonemic words, only less than 2,000 of them have actual lexical items (cf. Figure 4, which shows 1,754 syllables). About 2,700 monosyllabic items (i.e., actual words), excluding proper names, are jammed in the 2,000⁻ slots.⁸ This means that Taiwanese has a serious problem of monosyllabic homophones. In Figure 4, one sees that only 1/3 (36.77%) of the monosyllabic phonemic words (by type) are favored and that more than 1/2 of the monosyllabic lexical items (by token) are homophonous. The concentration of monosyllabic colloquial lexical items in a favored monosyllabic phonemic word can be as dense as 8 items. The situation is like that of cities and the countryside. Many cities are overcrowded, but the countryside is scarcely inhabited, even empty.

Language speakers apparently tend to avoid rare syllables in word-making and cling to familiar syllables. Supposing Taiwanese has 4,000 possible syllable types and supposing the rate of tonal neutralization in combination is flatly 3/9 (33.33%) which makes 2,667 possible beginning syllable types, the language can have as many as 10,668,000 disyllabic phonemic words (i.e., types). There is, therefore, no need for trisyllabic words. Nevertheless, language is not mathematics. Most of the 10 million phonemic words are “barrens”, and only a very small portion of the possible combinations is “arable”, and yet an even smaller portion is “exploited”.

In Ogawa (1931, 1932), only about 100,000 of the 10 million strong possible Taiwanese disyllabic phonemic words are recorded as occupied by lexical items. That is, more than 1% of the disyllabic phonemic word slots are filled. When there is an increase of disyllabic lexical items, the new lexical items are more likely to dwell in the more populated “cities”. The consequence is that there will be an increase of disyllabic homophones. This is the situation that will be discussed in §4.

On the other hand, language speakers also tend to prepare excessive room for clearer communication. As stated above, more than a half of the possible Taiwanese syllable types are “empty rimes” (in historical Chinese terminology) currently not occupied by monosyllabic lexical items. For instance, there is the phonemic word /gín/, but it only occurs in a single disyllabic lexeme GÍN-Á ‘child’.⁹ There is no monosyllabic lexeme *GÍN. Furthermore, originally

⁷Based on Steinmets and Baraham (1993) and Murakami (1983). The calculation of English is a collective work of Misses 廖慧玲, 陳慧娟, and 林美燕 in my 1993–1994 Introduction to English Linguistics class.

⁸Murakami (1983) is a small dictionary and, therefore, contains less monosyllabic words (2,650 items, as shown in Figure 4).

⁹The definition of “lexeme” follows Matthews (1993).

occupied slots were even emptied out. For instance, there is the phonemic word /hau/, but it now only occurs in a disyllabic lexeme HAU-SIĀU ‘to lie’ (considered vulgar by some speakers, for *siâu* literally means ‘semen’), for the monosyllabic lexeme *HAU ‘to lie’ is lost in modern Taiwanese except in extant old idioms. By the same token, there are sequences of syllables that cease to realize or do not realize as disyllabic lexical items but only form a part of polysyllabic words. For instances, *ông-lók-á(-sian)* ‘swindler’ and *ko-put-chiang* ‘cannot help but...’. The base of *ông-lók-á(-sian)* is *ông-lók*, which is now lost. Again, there is *ko-put-chiang*, but there is no **ko-put* or **put-chiang*. Many more lexical items simply died out without lengthening.

The anticipation or avoidance of homophones probably is the main motivation of trisyllabification, as well as disyllabification from monosyllabics. Envisaging invisible disyllabic homophones, together with other reasons, if any, the native speakers would expand existing disyllabic items to trisyllabic or coin new trisyllabic lexical items.

2.3 Trisyllabic Words

There are three major categories of trisyllabic lexical items in Taiwanese. The trisyllabics are not categorized according to their structure. Rather, they are classified in a way convenient to illustrate the status of Taiwanese trisyllabism. The first category is place names. Although old disyllabic Taiwanese place names were abundant, they were not as many as trisyllabic names. Many of the trisyllabic place names are phonemic loans from Formosan languages, and all of the trisyllabic place names that were given by early immigrants are descriptions. Many of the descriptions cannot be rendered in two syllables orally. They have to be in three syllables. These are “necessary” trisyllabic words. There are also descriptive place names with an optional deminutive suffix *-á*. These are “unnecessary” trisyllabic words.

(5) Some Trisyllabic Taiwanese Place Names

Trisyllabic Names		Disyllabic Doublets	
Chhân-tiong-ng	田中央	Tiân-tiong	田中
Chhiū-nâ-á	樹林仔	Chhiū-nâ	樹林
Chiòh-kng-á	石岡仔	Chiòh-kng	石岡
Chiòh-pâi-á	石牌仔	Chhiòh-pâi	石牌
Chúi-lí-khe ⁿ	水裏坑	Chúi-lí	水裏
Eng-ko-chiòh	鶯歌石	Eng-ko	鶯歌
Î ⁿ -nâ-á	員林仔	Oân-lîm	員林
Î ⁿ -soa ⁿ -á	圓山仔	Î ⁿ -soa ⁿ	圓山
Kiô-á-thâu	橋仔頭	Kiô-thâu	橋頭
Lêng-bák-ché ⁿ	龍目井	Liông-ché ⁿ	龍井
Pat-lí-hun	八里坌	Pat-lí	八里
Poát-á-nâ	拔仔林	Poát-lîm	拔林
Sa ⁿ -têng-po	三重埔	Sam-tiông	三重
Sin-siā-á	新社仔	Sin-siā	新社
Tang-koe-soa ⁿ	冬瓜山	Tang-soa ⁿ	冬山
Tang-sì-kak	東勢角	Tang-sì	東勢
Thô-á-hûg	桃仔園	Thô-hûg	桃園
Tiong-chiu-á	中洲仔	Tiong-chiu	中洲
Khong-á-liâu	槓仔寮	Kòng-liâu	貢寮
Phoh-á-kha	朴仔腳	Phoh-chú	朴子
Soa ⁿ -á-kha	山仔腳	San-ka	山佳
Thâm-á-kí ⁿ	潭仔墘	Thâm-chú	潭子
A-tà-bū	阿罩霧	Bū-hong	霧峰
Âng-thâu-sū	紅頭嶼	Lân-sū	蘭嶼
Cháp-lák-hūn	十六份	Sèng-hin	勝興
Chhia-láng-po	車籠埔	Kong-liông	光隆
Chúi-túg-kha	水返腳	Sít-chí	汐止
Gû-mā-thâu	牛罵頭	Chheng-chúi	清水
Hô-lô-tun	葫蘆墩	Hong-goân	豐原
Hóe-sio-tó	火燒島	Lék-tó	綠島
Kap-má-lân	噶瑪難	Gî-lân	宜蘭
Lîm-kí-po	林圯埔	Tek-san	竹山
Sa ⁿ -kak-éng	三角湧	Sam-kiap	三峽
Ta-pa-ní	礁吧嘸	Giók-ché ⁿ	玉井

The second category consists of trisyllabic words in the pattern ABC, where each letter represents a syllable, and having one or more synonymous words in the pattern AB, AC, or BC. The following is a selective list of 20 examples.

(6) Taiwanese Synonyms of Different Lengths

Trisyllabic Words	Disyllabic Synonyms	Gloss
ah-bó-nng	ah-nng	duck egg
ām-á-kún	ām-kún	neck
bīn-thâu-chêng	bīn-chêng/thâu-chêng	(in) front
chhin-bák-chiu	chhin-bák	with one's own eyes
chhiú-hó-kháu/ hó-kháu-chhùi	hó-kháu	the space between the thumb and the forefinger
chhùi-ē-táu	ē-táu	the chin
chóng-háp-kiōng	chóng-kiōng	total
èng-tong-jiân	tong-jiân	of course
goân-khí-thâu	khí-thâu	in the beginning
hiân-chú-sī	hiân-sī	now
keng-kah-thâu	keng-thâu	shoulder
kha-āu-te ⁿ	āu-te ⁿ	the heel
lòh-bóe-chhiú	bóe-chhiú/lòh-bóe	later; finally
mī-chioh-phōe	mī-phōe	cotton quilt
mng-kha-kháu	mng-kháu	doorway
po-lê-kià ⁿ	po-lê	glass
put-jī-kò	put-kò	but
sín-bêng-toh	sín-toh	table serving as an altar
thang-á-mng	thang-á	window
tng-á-tō	tng-tō	intestines

The third category are all other trisyllabic words, some of which that have corresponding new disyllabic equivalents are listed in Figure 7, under §3.3.

The two lists of trisyllabic words in (5) and (6) will be discussed again in §§3.1–3.2. Their purpose in this section is to illustrate that trisyllabic words are numerous in Taiwanese, 10.12% according to Murakami (1983), as shown in Figure 1 (see also its corresponding note), but actually the percentage should be higher. Most original place names in (5) are “necessary” trisyllables, but all except the last item in (6) are “unnecessary” trisyllables.¹⁰ If a word is “unnecessarily” trisyllabic, why does it remain trisyllabic, or why did it become trisyllabic? The only answer is in §2.2, i.e., the anticipation and avoidance of homophones.

3. Literary Disyllabism

In spite of the internal urge for Taiwanese to become trisyllabic, external interferences force it to be confined in disyllabism. On the one hand, monosyllables, whether ambiguous or not, are disyllabified. On the other hand, longer words are shortened, mostly to become disyllabic.

Besides the “anticipation” discussed earlier, many instances of the unnecessary disyllabification of monosyllables are due to the impact of Chinese literary lexical items. Written Chinese and written Japanese tend to form new words in two characters. Single-character words will often be expanded to having two characters in written Chinese. Nevertheless, although there are some expanded disyllables of this category in Taiwanese partly as a result of Chinese influence, such as *khàu* → *thi-khàu* ‘weep’ and *hoe* → *hoe-lúi* ‘flower’, their total number is not large at all.

The stronger Chinese and Japanese literary influences manifest in the shortening of longer words. In both written Chinese and written Japanese, longer lexical items are compressed to two characters in writing language whenever possible. Thus, Chinese 全國人民代表大會 (*Quánguó Rénmín Dàibiǎo Dàhuì*) ‘people’s congress’ is abbreviated as 人大 (*Rén-Dà*). Absorbing Chinese and Japanese lexical items and imitating Chinese and Japanese literary conventions, Taiwanese

¹⁰ *Tng-tō* also means ‘length’. Besides, some other doublets have become contrastive pairs, e.g., *kún-thâu* ‘martial art’ versus *kún-thâu-bó* ‘fist’.

lexical items are often trimmed to stay within the two-character limit. The disyllabic literary norm indeed has become a casting mold that often limits the elasticity of Taiwanese lexical length.

3.1 Place Names

The observance of the written Chinese and written Japanese lexicological traditions is most obviously shown in Taiwanese place names. Many place names were polysyllabic, but the Chinese and the Japanese administrators shortened the majority of them or even changed some of them to two characters in the written forms.

Figure 5 is only a small portion of trisyllabic place names that are shortened. Before the Japanese government in Taiwan launched the major change of place names in 1920, there were more trisyllabic place names than disyllabic ones in Taiwan. Take the area enclosed in what is modern Taipei Prefecture, including Taipei City and Keelung City, for example. Excluding mounds, waters, “markets” (where merchants concentrated), and “streets” (where merchants and residents concentrated) the names of which were almost all polysyllabic, there are 64 disyllabic place names, 71 trisyllabic place names, and 4 quadrisyllabic place names recorded in Abe (1937). Nowadays, on the contrary, all of the official names of districts, towns, and cities in Taipei Prefecture are disyllabic, though some of the old trisyllabic names still occur orally, such as that of the site of the controversial 4th Nuclear Power Plant, *Khong-á-liâu*.

3.2 Synonyms of Different Lengths

In §2.3 it is mentioned that (6) can be looked at from two different sides of the shield. From one side, one would reckon that a portion of the disyllabic forms are expanded to trisyllabic words. From the other side, one would reckon that another portion of the trisyllabic forms are abbreviated to become disyllabic. If one looks at it from the latter side, it is easy to attribute the abbreviation to the pressure of the Chinese and Japanese literary lexical pattern. If one looks at it from the former side, one has to explain the coexistence of the disyllabic forms along with the trisyllabic ones. The explanation is then that the disyllabic forms are sustained by the literary lexical pattern. No matter which way the change goes, literary disyllabism is there to control and to direct.

3.3 New Words

There are also polysyllabic words that are competing with or already lost to more recent disyllabic loans from written Japanese and, especially, written Chinese. Polysyllabification, especially trisyllabification, is indeed a strong trend in Taiwanese word formation. Nevertheless, the trend is arrested as a result of language contact. Many things and matters, especially modern ones, are expressed in different ways in (spoken) Taiwanese, in (written) Chinese, and in Japanese. For instance, ‘sound film’ was rendered as *hoat-siaⁿ oáh-tāng ê tiān-hì* (發聲活動的電戲) in Taiwanese (Barclay 1923:86b), but as 有聲電影 (*yǒushēng diànyǐng*) in Chinese. When pre-Japanese Taiwanese came into contact with classical Chinese before 1895, when Taiwanese came into contact with written Japanese since 1895, and especially when post-Japanese Taiwanese came into contact with written Chinese in later years, many Chinese and Japanese equivalents, ancient and modern, were borrowed and rendered according to Taiwanese pronunciations of the characters. 有聲電影 therefore becomes *iú-siaⁿ tiān-iáⁿ* in literary Taiwanese. These new words from written Chinese, predominantly disyllabic or having disyllabic stems, started to compete with their older indigenous equivalents and borrowed equivalents from Japanese and earlier Chinese.¹¹ Since Chinese has had an overwhelming power

¹¹One counterexample is ‘potato’. The newer literary word for it, *bé-lêng-chî/má-lêng-chú* (馬鈴薯), is trisyllabic, whereas the older colloquial word for it is disyllabic, i.e., *iúⁿ-chî/-chú*. However, longer literary

over Taiwanese in the past five decades, many new Taiwanese disyllabic words from written Chinese have replaced their older trisyllabic equivalents.¹²

In Figure 7, a majority (23 out of 30) of the lexical items listed are words for early modern things new to the Taiwanese. Only 5 of the 23 modern polysyllabic words are from written Japanese (in pairs 1, 4, 12, 13 and 30). The rest are indigenous colloquial coinage or phonemic loans from Japanese. The trisyllabic forms of pairs 1, 3–5, 7–9, 11, 13, 16, 17, 22, 24, 26, and 28–30 now either have been or are in the process of being completely replaced in common Taiwanese. They are marked with an asterisk.

(7) Older Polysyllabic Words and Corresponding New Disyllabic Words¹³

Polysyllabic Words	Disyllabic Synonyms	Gloss
01. bān-liān-pit*	kàng-pit 鋼筆	fountain pen
02. chāi-sek-lú	chhú-lú 處女	virgin
03. chhân-thâu-ke*	tē-chú 地主	farm landlord
04. chū-tōng-chhia*	khì-chhia 汽車	motor vehicle
05. gū-tō-kin-á*	mô-kin 毛巾	towel
06. hoan-á-hóe	hóe-chhâ 火柴	the match
07. hoan-á-iú*	chiòh-iú 石油	petroleum
08. hóe-ian-chhia*	hóe-chhia 火車	[railway] train
09. hóe-ian-chhûn*	khì-chhûn 汽船	steamboat
10. hông-tè-niú	hông-hiō 皇后	empress
11. hông-tè-tiān*	hông-kióng 皇宮	imperial palace
12. hui-hêng-ki	hui-ki 飛機	airplane
13. hui-hoat-iú*	khì-iú 汽油	gasoline
14. ke-kui-á	khì-kiú 氣球	balloon
15. kiát-thâu-chhài	kam-nâ 甘藍	kohlrabi
16. lâ-giâ-chhia*	chúi-chhia 水車	waterwheel
17. liān-khò-tò*	hong-i 風衣	raincoat
18. ne-khú-tái	niá-tò 領帶	necktie
19. o-thâu-á-chhia	kiau-chhia 轎車	sedan (car)
20. o-tó-bái	ki-chhia 機車	motorcycle
21. peng-á-sa ⁿ	kun-hók 軍服	military uniform
22. se-bí-loh*	se-chong 西裝	lounge suit
23. sé-seng-khu-keng	ék-sit 浴室	bath[room]
24. tâng-sò ⁿ -phe*	tiān-pò 電報	telegraph
25. the-ní-suh	bāng-kiú 網球	tennis
26. thih-ki-á-lō*	thih-lō 鐵路	rail[road]
27. tho-lák-khuh	khah-chhia 卡車	truck
28. tiān-iá ⁿ -hi*	tiān-iá ⁿ 電影	motion picture
29. tōa-thâu-chhang*	iú ⁿ -chhang 洋蔥	onion
30. ūn-choán-chhiú*	su-ki 司機	driver

words are exceptions.

¹²Competetion and replacement also occur between polysyllabics of course. ‘Sound film’ is an example. *La-jí-oh* ‘radio’ competing with and losing to *síu-im-ki* (收音機) is another. Nevertheless, our issue here is the impact of literary disyllabism.

¹³The disyllabic form in pair 19 is a case of erroneous reading of Chinese characters. Cf. Tiuⁿ J. (1995b).

4. Homophones

A language without a written tradition or one with a weak written tradition, such as Taiwanese, is strong in purging “actual homophones” (as contrasting to “virtual homophones”).¹⁴ However, the purging function can be handicapped when the language is under great influence from another language. Vietnamese borrowed from French *ga* ‘accelerator (Fr. **gas**)’ and *ga* ‘station (Fr. **gare**)’, which are homophonous with the indigenous *ga* ‘gauge’. Taiwanese borrowed from Japanese *bát-tah* (a variant of *bá-tah*) ‘butter (Jap. **batā**)’, *bát-tah* ‘[baseball] batter (Jap. **battā**)’, and *bát-tah* ‘[baseball] bat (Jap. **batto**)’, all becoming homophones. Examples multiply, but, with regard to homophones, the impact of phonemic loans on the borrowing language is weak compared with borrowing through written forms.

It has been a tradition of lexical borrowing for thousands of years that written words and phrases in Chinese, Japanese, etc., are rendered in different languages and dialects according to their respective local pronunciations of the written characters. Therefore, for instance, the Japanese translated ‘chemistry’ as 化學 (pronounced as *kagaku*) which becomes *hùaxiúe* in Chinese, *hóa-học* in Vietnamese, and *hò-a-hàk* in Taiwanese. Again, in the former Malaya, either the Hokkiens or the Teochews borrowed English ‘orchid’ and rendered it as *o-kiit* and then wrote it as 胡姬, which, then, was borrowed into Singaporean Mandarin and pronounced as *hújī*.

The “corruption” of sounds does not concern us here. The point is that during the process of borrowing from written words, sound is ignored. Homophones are created or added to the lexicon without the awareness of the borrowers because of the ignoring, and there is therefore no purging of homophones.

The long history of the borrowing tradition, in addition to the simplification of phonological systems, causes languages using Chinese characters to be haunted by monosyllabic homophones. Even polysyllabic homophones can be bothersome in communication. Figures 8–10 show the percentage of homophones calculated in different ways. In Figure 8, all the languages except English and Malay are using or used to be using characters in writing. They borrowed written words from one another and accumulated a great deal of homophones in their lexicons. There is therefore a conspicuous difference between these languages on the one hand and English and Malay on the other in terms of the percentage of monosyllabic homophones.¹⁵

Lexical Homophones

(8) *All Items with the Same Number of syllable(s)*

	Monosyllabic	Polysyllabic
Taiwanese	(小川) 86.28% (村上) 58.15%	(小川) 3.25% (村上) 1.38%
Chinese	76.92%	6.67%
Japanese	86.57%	12.88%
Vietnamese	53.81%	0.47%
English	31.29%	–
Malay	15.05%	1.98%

Taking the whole lexicons into consideration, the difference between Malay and the languages in question is even greater, as shown in Figure 9, which considers lexical items of all lengths.¹⁶

¹⁴For the terminology see T’sou (1976:68,71f).

¹⁵小川 (Ogawa 1931, 1932) is a full-sized dictionary with about 100,000 entries. The rest of the dictionaries consulted for this table are small ones. In the table, the calculation of English polysyllabic homophones is not available, and the low percentage of Vietnamese polysyllabic homophones probably is due to the high number of syllable types and the change of the script from characters to letters.

¹⁶Malay homophones include seeming derivations. The calculation of Korean lexical homophones is a collective

(9) $\frac{\textit{Lexical Homophones}}{\textit{All Lexical Items}}$

Taiwanese	(小川) 10.70% (村上) 12.70%
Chinese	14.38%
Japanese	13.69%
Vietnamese	15.74%
Korean	21.40%
Malay	2.06%

Polysyllabic homophones are not yet a threat to any of the languages in (9). However, since the majority of Sinitic, Japanese, Vietnamese, and Korean polysyllabic homophones came from written sources, and since these languages apparently are unable to purge literary homophones effectively, the long-term result will be a higher and higher percentage of polysyllabic homophones in their respective lexicons. As a matter of fact, Japanese is already rich in quadrisyllabic homophones from written sources. Figure 10 is the result of a calculation of disyllabics.

(10) $\frac{\textit{Disyllabic Literary Homophones}}{\textit{All Disyllabic Homophones}}$

Taiwanese	(小川) 83.99% (村上) 81.25%
Chinese	90.39%
Vietnamese	80.88%

With respect to the increase of homophones as a result of borrowing from written materials, Taiwanese is not very different from any other languages that are associated with Chinese characters. However, facing the influx of two-character literary words, the urge to trisyllabify colloquial words has been suppressed. This curbing of trisyllabification probably is peculiar to all Sinitic languages, and Taiwanese is but an example.

5. Conclusion

As reiterated above, there is a need for spoken Taiwanese to become more trisyllabic. In fact, there are traces in Taiwanese lexicon showing that Taiwanese probably must have been in the process of trisyllabification. The most obvious trace is the old place names, more than a half of which were trisyllabic. Next comes the existence of homonymous polysyllabic doublets, one having two syllables and the other having one or more extra syllables. Although it is not known whether certain disyllabic common lexical items (i.e., not proper names) are the abbreviations of their respective trisyllabic counterpart or whether the trisyllabic forms are the expansion of their respective disyllabic counterpart, the doublets indicate that there are two forces that are operating within Taiwanese lexicon. On the one hand, oral communication demands longer words for clearness. On the other hand, the disyllabism that is conditioned by the two-character convention of written Chinese and written Japanese circumvents the trisyllabification. That is, the preference for two written characters per lexical item in visual communication becomes a bondage limiting the length of most new words to two syllables each and hinders the development of the other force. The circumvention yields a very undesirable result in oral communication, that is, the increase of homophones. Currently, Taiwanese still functions orally in almost all

work of Misses 袁玉玲 and 李美和 in my 1993–1994 Introduction to English Linguistics class. They failed to calculate the percentage of literary homophones for Figure 10, however.

occasions and retains a very high power of purging homophones. As such, the homophone problems as seen in Taipei Chinese and in Japanese are not yet noticeable in Taiwanese.¹⁷ However, since more and more Taiwanese publications are appearing, it is predictable that continuous and heavier influence from written Chinese on written Taiwanese will gradually impair spoken Taiwanese's natural ability to control the number of homophones.

¹⁷“Standard” Taipei Chinese is losing the contrast between *r*- and *l*- initials in a portion of the lexicon, and retroflex initials in another portion have become non-retroflexes. The suffix *-r* and the neutral tones rarely occur. And *-in* has merged with *-ing*. As such, a much higher percentage of homophones than Pekinese is expected in this Mandarin dialect.