

A PROPOSAL TO ENCODE TAIWANESE KANA (TÁI-OÂN-GÍ-KHÁ-NAH 豪語假名) IN THE UCS

Fredrick R. Brennan



ふるやときお
古家時雄を追悼して
もじきょう
文字鏡研究会に心から感謝します

ACKNOWLEDGEMENTS

in no particular order...

たに もと さち ひろ
Sachihiro Tanimoto (谷本玲大), Waseda University, Kokugakuin University

For his patient explanation of the history of Mojikyō, and his priceless help in getting Mojikyō Character Map working on my computer.

Deborah Anderson, University of California @ Berkeley Script Encoding Initiative

For her tireless review of script proposals by n00bs like me.

み うら だいすけ
Daisuke Miura (三浦大介), World Special-Characters Wiki (世界の特殊文字 ウィキ)

For his recommendation that I name the tone letters like MODIFIER LETTER KATAKANA instead of just KATAKANA as I originally planned; I did not know modifier letters could be non-Latin, but he knew of the precedent of U+10FC —MODIFIER LETTER GEORGIAN NAR⁽⁶⁾.

やま ぐち りゅう せい
Ryūsei Yamaguchi (山口 隆成)

For his experience with Mojikyō, Unicode, and all around good advice.

Wil Lee (Lí Kho-Lūn, 李イ/
科ヲ
潤ル!), Patreon

For kindly giving me a Taiwanese Hokkien name (also usable for Mandarin Chinese), for use in this proposal.

こ ばやしけん
Ken Lunde (小林 劍), Unicode Consortium

For his font development advice, and helpful advice regarding Unihan.

さい もん
Simon Cozens (カズンズン才文) & Caleb Maclennan, SIL^E

For their help with SIL^E, an experimental typesetting engine and alternative to L^AT_EX.

Table of Contents

Quick summary	4
Tone marks	4
Nasalized tone marks	5
Combining characters	6
Overview	7
Linguistic	7
Historic (Taiwanese...kana?)	7
Legacy encodings	10
Pseudo-Unicode encodings	10
Mojikyō	10
Unicode considerations	12
On tone six	12
On the aspiration mark	12
On glyph order	13
Tone letters	13
Combining characters	13
Vertical typesetting	13
Evidence of all requested characters	15
Tone letters	15
Combining characters	15
Miscellaneous pages	17
Bibliography	19

Chapter 1

Quick summary

I propose a new block, in the Supplementary Multilingual Plane, to be called Katakana Phonetic Extensions Extension A, or perhaps, if the Unicode Technical Committee prefers, Taiwanese Kana. I propose that the new block is to have range U+1B300 – 1B31F, and is to include:

1.1 Tone marks

/ MODIFIER LETTER KATAKANA HOKKIEN TONE TWO

Mojikyō 69622; tentative encoding U+1B302

- Also used for tone six¹

\ MODIFIER LETTER KATAKANA HOKKIEN TONE THREE

Mojikyō 69623; tentative encoding U+1B303

• MODIFIER LETTER KATAKANA HOKKIEN TONE FOUR

Mojikyō 69624; tentative encoding U+1B304

⟨ MODIFIER LETTER KATAKANA HOKKIEN TONE FIVE

Mojikyō 69625; tentative encoding U+1B305

| MODIFIER LETTER KATAKANA HOKKIEN TONE SEVEN

Mojikyō 69627; tentative encoding U+1B307

• MODIFIER LETTER KATAKANA HOKKIEN TONE EIGHT

Mojikyō 69628; tentative encoding U+1B308

1. For more discussion of this issue, and why I'm proposing we leave U+1B306 and U+1B316 open, please see § On tone six.

1.2 Nasalized tone marks

⠇ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE ONE
Mojikyō 69629; tentative encoding U+1B311

⠋ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE TWO
Mojikyō 69630; tentative encoding U+1B312
• Also used for tone six¹

⠌ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE THREE
Mojikyō 69631; tentative encoding U+1B313

⠋⠄ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE FOUR
Mojikyō 69632; tentative encoding U+1B314

⠋⠄⠄ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE FIVE
Mojikyō 69633; tentative encoding U+1B315

⠋⠄⠄⠄ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE SEVEN
Mojikyō 69635; tentative encoding U+1B317

⠋⠄⠄⠄⠄ MODIFIER LETTER KATAKANA HOKKIEN NASALIZED TONE EIGHT
Mojikyō 69636; tentative encoding U+1B318

1. For more discussion of this issue, and why I'm proposing we leave U+1B306 and U+1B316 open, please see § On tone six.

1.3 Combining characters

◎ COMBINING KATAKANA DOT BELOW

Mojikyō 069202; tentative encoding U+1B31E

- Indicates aspiration; native name 送氣符. For example, while タア is /ta/, タア is /t^ha/.
- At least one source renders it identically to the *nakaguro* (中黒);
→ • U+30FB KATAKANA MIDDLE DOT

◎ COMBINING KATAKANA OVERLINE

Mojikyō 069637–069642;² tentative encoding U+1B31F

- Makes the kana it combines with /t/ in the case of チ (/ti/) and ツ (/tu/); or adds a /t/ as in ツサ (/tsa/ or /tса/) and ツセ (/tse/ or /tсе/); or changes vowel sound as in ヴ (/i/) and ヴ (/ə/)
- Sometimes rendered identically to the *chōonpu* (長音符; U+30FC; KATAKANA-HIRAGANA PROLONGED SOUND MARK), so ツ in place of ヴ.

2. Mojikyō chose to encode the versions of the katakana with combining overline above separately, so this refers to the individual katakana they encoded. In Mojikyō order: ヂチツセツツ. They are missing ツ and ヴ.

Chapter 2

Overview

2.1 Linguistic

Taiwanese Hokkien (臺灣語) is a Sinitic Southern Min language spoken throughout the world; as of 2013, it was spoken by 48 million people.¹ It is spoken as far afield as the Philippines, where it is known as *Fookien*.

In Taiwan, Hokkien is a widely spoken language, and is even used in academic writing. The most popular way of writing Hokkien is known as *Hàn-lô* (漢羅), which to untrained eyes may appear to be *mojibake*...

“ M̄ 是講攏無台灣來 ê 學生來修。Tí 中級班，有一個讀電腦 ê 博士生；初級班有一個客家人，讀設計碩士班。”

However, it is not; what is happening is that 10–15% of Hokkien phonemes are not representable as a standardized Chinese character. So, reading Hokkien only in Chinese characters, called *Tāng-lâng-jī* (唐儂字), is quite difficult. Therefore, other systems, among them the Taiwanese government’s official *Tâi-lô* (臺閩) and the older *Peh-ōe-jī* (白話字) upon which it is based, along with, more rarely, *chù-im hû-hō* (注音符號), known elsewhere as *bopomofo*, are mixed in with the Chinese characters. Indeed, there’s no reason Taiwanese kana can’t be used for this purpose:

“ ム̄是講攏無台灣來工〈學生來修。矛|中級班，有一個讀電腦工〈博士生；初級班有一個客家人，讀設計碩士班。”

But perhaps we’re getting ahead of ourselves.

2.2 Historic (Taiwanese...kana?)

If I were given the task of naming Taiwanese kana, or, perhaps better said, translating their Hokkien/Japanese name, I would have named them “Hokkien kana”. Because,

1. Sun, Obed [孫暉皓] (2017). “台灣16年「母語教育」宣告失敗？我們需要「搶救台語」嗎？——我在新加坡的四個觀察”. Crossing [換日線]. Tiānxià Magazine [天下雜誌]. Quote: “全球的閩南語（包括潮汕話、海南話等）使用者有超過 4,800 萬人”.

2. Khîn-hōaⁿ, Lí [李勤岸] (2007). 台語飯桌. Sìn Bōng Ài [信望愛].

indeed, that's what they are; historically speaking, there's nothing Taiwanese about Taiwanese kana except that they can be used to write Taiwanese Hokkien. Indeed, they are not even exclusively *Taiwanese*, as they are also suitable for writing the Quanzhou dialect.

Taiwanese kana were invented by Japanese linguists on behalf of the imperial Japanese government, which at the time of their introduction, was pursuing a policy of *dōka* (同化), meaning, “assimilation”. The period we most remember them for today was the period of the 1930's, when they were spread primarily to teach Taiwanese people Japanese.³

The primary person involved in their creation was the linguist Naoyoshi Ogawa [小川尚義].⁴ The first of Ogawa's works that they make a notable appearance in is the *Japanese–Taiwanese Daijiten* (日台大辭典) of 1907.⁵ Ogawa continued to refine them until he published his greatest work, the two volume *Comprehensive Taiwanese–Japanese Dictionary* [臺日大辭典] (1931). Indeed, it is entirely on the back of this work that modern interest in them continues, with the dictionary continuing to receive updates long after the collapse of the Japanese empire; even famed Taiwanese linguist Âng Úi-jîn contributed substantially to an update of the dictionary in 1993, published as the *Minnan Classic Dictionary Collection* (閩南語經典辭書彙編).

Due to the high quality of this dictionary, which even informed the Taiwanese government's own *Dictionary of Frequently-Used Taiwan Minnan* [臺灣閩南語常用詞辭典] (1st ed. 2011),⁶ Japanese learners of Hokkien continue to use Taiwanese kana even today.⁷ But, of course, this is not the only reason to encode Taiwanese kana; they are interesting in and of themselves, as a historical writing system with a large body of work behind it.

Even in Ogawa's time, he was not the only author; other Japanese scholars recognized the quality of his work, and many other works exist which integrate Taiwanese kana, such as the *Taiwan Proverb Collection* [臺灣俚諺集覽] (1914) and the *State Monopoly Bureau's Taiwanese Dictionary* [專賣局台灣語典] (1923), which, unlike Ogawa's work, mixed *Peh-ōe-jī* with Taiwanese kana.

3. Chen, Chun-Hui [陳君慧] (June 2002); Lin, Ching-Hsiun [林慶勳], ed. “《訂正台灣十五音字母詳解》音系研究”. National Sun Yat-sen University Chinese Literature Department: Master's Thesis.

4. Ibid, p. 19.

5. Ibid, p. 19–20.

6. “Bibliography” [參考書目]. *臺灣閩南語常用詞辭典*. Retrieved August 1, 2020.

7. For two examples of Japanese learners of Hokkien who use Taiwanese kana, see the Twitter profiles of Soaⁿ-kiā (@suannkia) and Taipa (@taipalogy).

In our time, besides being used as a substantial basis for the *Dictionary of Frequently-Used Taiwan Minnan*, Âng Úi-jîn's version of the dictionary lives on due to a project of Sih Sîng-hông [薛丞宏] and Lîm Chùn-iok [林俊育] of the Academia Sinica: taigi.fhl.net/dict . This impressive work is a complete digitization of the 1993 dictionary, and when you search for terms, you are given links to individual pages of it. For example, searching 夢, meaning “dream”, brings up [page 555 of volume 2](#), which brings up quite a detailed entry, which differentiates between frightening dreams/nightmares (恐ろしい夢) and merely bad dreams (悪い夢); being told about a dream (夢の告を受ける) from dream interpretation (夢判断):

	バ
	ジ
	夢。
託す。	夢め。
出現。	完ヌ。(姓)
夢枕に立つ。	夢ぼう。
夢枕に立つ。	夢め。
夢枕に立つ。	夢め。
夢枕に立つ。	惡オ。
夢枕に立つ。	恐おぞ。
圓ヌ。	し。
夢判斷。	夢め。
	タバ
	一
	夢めはんたん。
	悪わる。

2.3 Legacy encodings

2.3.1 Pseudo-Unicode encodings

A “pseudo-Unicode” encoding is currently in use on Wikipedia and elsewhere. This “encoding”, to the extent that it is one, mixes SVG glyphs, HTML markup, and a few Unicode characters intended for other purposes.

- U+1B302–1B318 are represented with SVG’s;
- U+1B31E, the COMBINING KATAKANA DOT BELOW, is represented by the COMBINING DOT BELOW, U+0323;
- U+1B31F, the COMBINING KATAKANA OVERLINE, is represented variably by a CSS `border-top`, or else by a COMBINING MACRON (U+0304) or other script-inappropriate combining mark.

A second pseudo-Unicode encoding exists, which I call the “[Taipa encoding](#)”. It misuses characters like U+27E8, MATHEMATICAL LEFT ANGLE BRACKET, because of its similar appearance to tone 5, ጀ. It does not seem to have widely caught on outside of Taipa’s tweets; certainly this kind of misuse is a good reason to introduce real characters for the task.

2.3.2 Mojikyō

Mojikyō almost fully supports Taiwanese kana. Its only deficiency is that it is missing ㄉ and ㄉ, which means that the Quanzhou dialect is not reproducible in the Mojikyō encoding.

In Mojikyō, the Taiwanese kana are available in the font `Mojikm0D.TTF`; 20 codepoints are devoted to them. Mojikyō chooses to encode the COMBINING KATAKANA OVERLINE as precombined with the kana it combines with, and it encodes the COMBINING KATAKANA DOT BELOW as a spacing mark.



As mentioned in § Acknowledgements, Mr. Tanimoto and Mr. Yamaguchi helped me a lot with getting this program working. I was quite touched by Mr. Tanimoto's story, so I chose to memorialize his friend, Mr. Tokio Furuya, who is now deceased, on the front cover of this paper. Mr. Furuya was the main developer of Mojikyō.

Below are the Taiwanese kana glyphs as they appear in `Mojikm0D.TTF`; the top row are the tone letters, middle row nasalized tone letters, and bottom row overlined katakana:

ノ ャ ゝ ハ ワ
 レ ハ キ ゞ ブ ヲ
 サチツセソウ

Chapter 3

Unicode considerations

3.1 On tone six

In modern Hokkien, tone six is equal to tone two.¹ A symbol, therefore, for tone six, is not in frequent use. However, I found a single source with a symbol for tone six, Taigikho (2017), and recommend we leave a slot open for it should it become more popular in the future, or should more sources emerge, so we have a logical place to put it. Even in Taigikho (2017), however, the symbol is barely discernible, as shown below, and it seems to be that they are trying to discourage its use.

表 0 18		調序	第1聲	第2聲	第3聲	第4聲	第5聲	第6聲	第7聲	第8聲
◎ 聲調		調類	陰平	陰上	陰去	陰入	陽平	陽上	陽去	陽入
臺羅調符		○	ó	ò	oh	ô	ó	ó	oh	
臺語假名調符	常音		/	\	•	<			•	
	鼻音	ɓ	ɓ	ɳ	ɳ	ɸ	ɓ	ɳ	ɳ	

(舊化元音·類化輔音)

✓ 後表標注「○」的音節，臺語假名使用「鼻音」調符。

3.2 On the aspiration mark

The aspiration mark is quite unusual in that in some sources it acts as a combining character,² and in other sources, it appears identically to the *nakaguro* (中黑).³

- Chiung, Wi-vun Taiffalo (2003). “Tone Change in Taiwanese: Age and Geographic Factors”. University of Pennsylvania Working Papers in Linguistics. 8 (1).
- Ogawa (1932), Âng and Ogawa (1992), Taigikho (2017), among others.
- State Monopoly Bureau’s Taiwanese Dictionary* [專賣局台灣語典] (1923), among others.

As they are semantically the same, this issue should be left to higher-level protocols, such as fonts, to solve, via e.g. stylistic alternates. It is likely that in the sources where it appears as a *nakaguro*, this is only due to subpar printing technology, as it's clear it's meant to have been combining by Ogawa.

3.3 On glyph order

It makes the most sense to use the glyph order from Ång and Ogawa (1992). It is a simple order: basic Japanese kana order (*gojūon*, 五十音), with tone letters in numerical order and nasalized versions after normal versions. Overlined versions of characters come after regular versions, and aspirated versions come after those.

So:

3.3.1 Tone letters

/ < 6 < \ < 9 < 𢂀 < 𢂁 < 𢂂 < 𢂃 < 𢂄 < | < 𢂅 < 𢂆 < 𢂇

3.3.2 Combining characters

タ < タ⁴

サ < サ⁵ < サ̄ < ソ

3.4 Vertical typesetting

Taiwanese kana are most often typeset vertically historically. All horizontally typeset examples I was able to find are from this century.

For this paper, I am using a SIL^E package I wrote to do the vertical layout. Looking at the vertical layout, there are two ways to do it. In this paper, I consider it to be two lines, with the tone letter on the second line. It is also possible to do so in OpenType via contextual positioning, like this:

```
feature valt {
    lookup valt2 {
        # Make glyph 0-width and move it back 1.5 glyphs;
        # takes advantage of fact kana are monospace.
        # x     y     xadv yadv
        pos @toneletters.valt <-1300 500 0      0>;
    } valt2;
} valt;
```

4. Ång and Ogawa (1992), vol. 2, p. 2.

5. Ång and Ogawa (1992), vol. 1, pp. 610–611.

Vertical layout is really a problem to be handled by higher level protocols, and not a Unicode consideration, but is mentioned for completeness.

For a fully working OpenType implementation, see [FRB Taiwanese Kana](#), which is a font I made originally for this proposal but which can be used for Taiwanese kana in general. It works with HTML vertical layout, and has bolder glyphs for ruby:

臺 タイ < 語 ギイ /

タイ<ギイ /
臺 語

臺 タイ <
語 ギイ

No special actions need be taken, nor JavaScript be used, the HTML for the last example is just:

```
<p class="vert"><ruby>臺<rt>タイ</rt>語<rt>ギイ</rt></ruby></p>
```

Chapter 4

Evidence of all requested characters

4.1 Tone letters

Âng and Ogawa (1992), vol. 1, p. 3.

		(2)	(1)	二 符
鼻音	常音	八	出氣	號○
l	○	上聲	符	音符
6	/	上聲	號	號
9	\	上去		
o	,	上入		
d	<	下平		
b		下去		
o	,	下入		

4.2 Combining characters

Âng and Ogawa (1992), vol. 1, p. 5.

ヲ	バ	タ	カ
モ	ビ	チ	キ
ツ	ブ	ツ	ク
セ	ペ	テ	ケ
ソ	ボ	ト	コ

Âng and Ogawa (1992), vol. 1, p. 3. Proof of オ and ウ, missing from main table.

モ「ツエ」の促りたる音、即ち(あ)の音を表はす。

渣ツエ 達ツエ 祭ツエ 仄ツエ 坐ツエ 絶ツエ

ヲ「ツオ」の促りたる音、即ち(お)の音を表はす、此の母音に廣狭の二種あり、(オ)列の假名の條参照)。

租ツオ 祖ツオ 作ツア 宗ツシ 憎ツチ 左ツジ 紙ツシ

チ「ティ」の促りたる音、即ち(ヒ)の音を表はす、

知チシ 朝チホ 稲チヌ 等チド 陣チミ 中チヨ 直チツ

ヲ「トウ」の促りたる音、即ち(ム)の音を表はす。

蝶ツウ 除ツム 長ツメ 對ツバ 墓ツツ 唇ツツ 突ツツ

ヲ唇「イ」舌「ウ」の位置にて發する音を表はす。

余ツクニ 居ツクニ 暑ツクニ 鋸ツクニ 除ツクニ 慮ツクニ 魚ツクニ

オ唇「エ」舌「ヲ」(狭き「オ」)の位置にて發する音を表はす。

鍋ツクダ 過ツクダ 稅ツクダ 火ツクダ 飛ツクダ 袋ツクダ 尾ツクダ

他の五十音假名及び符號假名の發音は、大體に於て、日本語と同一なりと雖も亦多少異なるものなきにあらず、今特に注意を要するものにつき左に説明すべし。

一、「ウ」列の假名。

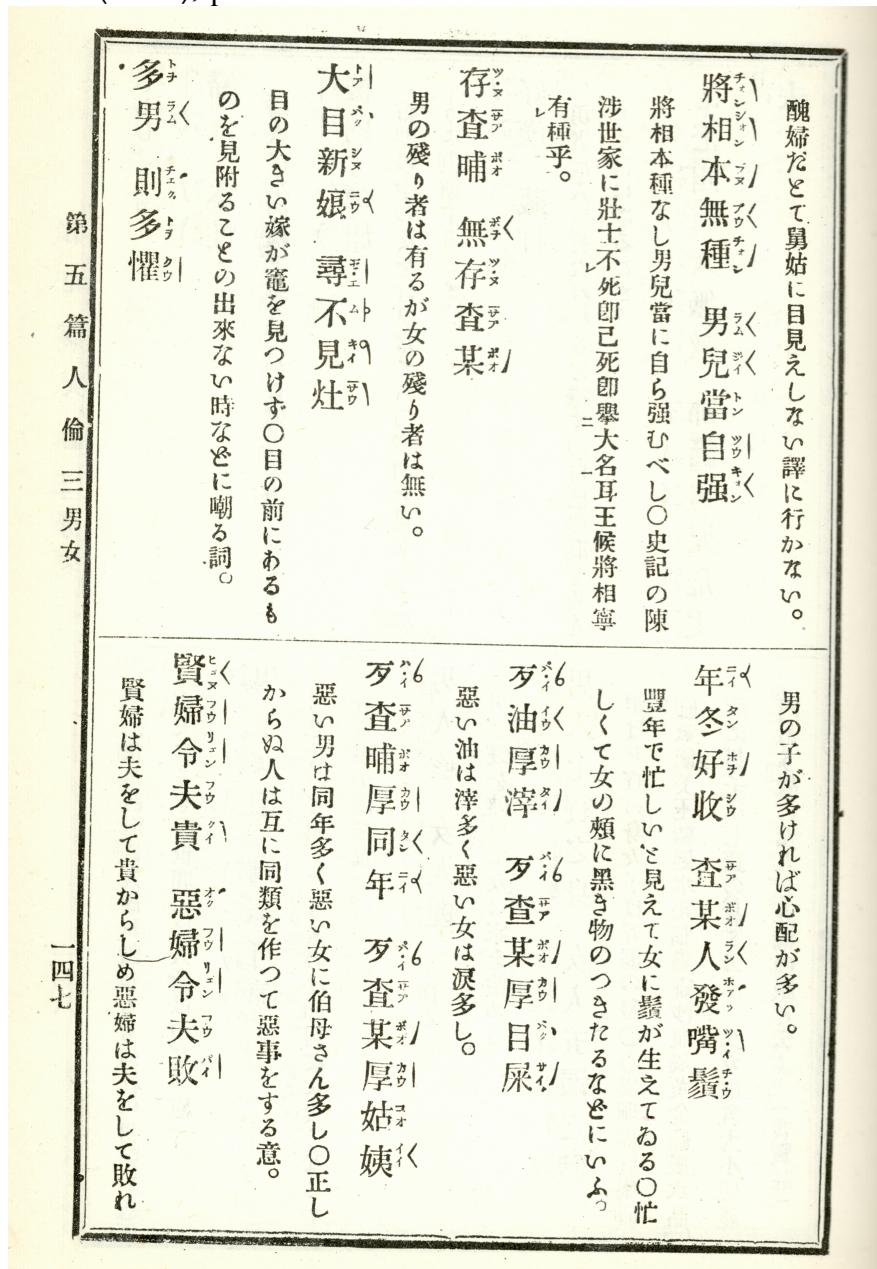
(1) 「ウ」列の假名にして「ウウ」「クウ」等の如き長音、又は「ウイ」

「クヌ」「クツ」の如く「イ」「ヌ」「ツ」の前に来るものは國語の「ウ」より狭く唇を突出して發音する「ウ」の母音を有す。

有ツクル 多ツクル 位ツクル 君ツクル 骨ツクル 憚ツクル 孫ツクル

4.3 Miscellaneous pages

Hirasawa (1914), p. 147.



Ogawa (1938). Unknown page, from Liang, et al (1999). “*The Japanese-Taiwanese Dictionary and The New Japanese-Taiwanese Dictionary* : a Comparison”. p. 16.

オチル(おつ)(自、ニ落。)	落葉(ロバ)。	落葉(ロバ)。	落葉(ロバ)。
起雷(キル)。	落葉(ロバ)。	落葉(ロバ)。	落葉(ロバ)。
鷦(ウマ)カラ。	日ガ山(ヒヤマ)ニ。	日落山(ヒヤマ)。	雷ガ。
跋落馬(ボロベイ)。	○「試驗ニ」	○「試驗ニ」	○「試驗ニ」
脱名(ルツメイ)。	落第(ロツテイ)。	○「垢ガ」	脱垢(ルツシユス)。
○「垢ガ」	○「垢ガ」	肉ガ。	肉ガ。
消去(シラキ)。	力ガ。	失望(シバツ)。	餒志(ロエチ)。
症脱去(チヨンル)。	色ガ。	退色(タクシヨン)。	瘧(オカリ)。
字ガ。	字落去(ジヨクキ)。	鼻(ハナ)。	乞食(キサム)。
○「字」	○「字」	四鼻(ヨクナ)	四鼻(ヨクナ)
字落去(ジヨクキ)。	脱字(ルツジ)。	○「鼻」	○「鼻」
○「字」	○「字」	○「鼻」	○「鼻」
聲(シテ)。	○「城」	對城(ライシヤ)逃走(トコトコ)。	○「計」
中計(チヨンケイ)。	○「城」	○「城」	○「計」
號較(カガ)輸(スウ)。	○「城」	○「城」	○「計」
較次(カガ)。	○「城」	○「城」	○「計」
能曉(オエヒア)。	○「城」	○「城」	○「計」
了解(リョウカイ)。	○「城」	○「城」	○「計」
此號(チホ)道(トド)理(リ)。	○「城」	○「城」	○「計」
此號(チホ)道(トド)理(リ)。	○「城」	○「城」	○「計」
受罪(シウブイ)。	○「城」	○「城」	○「計」
○「此」	○「此」	○「此」	○「此」

Chapter 5

Bibliography

- Hirasawa, Teiakira (1914). *Taiwan Proverb Collection* [臺灣俚諺集覽]. Governor-General of Taiwan, Imperial Government of Japan.
- Ogawa, Naoyoshi (1932). *Comprehensive Taiwanese-Japanese Dictionary* [臺日大辭典]. Governor-General of Taiwan, Imperial Government of Japan.
 - Volume 1 (1931)
 - Volume 2 (1932)
- Ogawa, Naoyoshi (1938). *New Japanese-Taiwanese Dictionary* [新訂日台大辭典]. (One volume only, second never completed.)
- Âng, Úi-jîn; Ogawa, Naoyoshi (1992). *Minnan Classic Dictionary Collection* (閩南語經典辭書彙編)
- *Tâi-lô-Taiwanese kana tables* (臺羅調符—臺語假名調符) (2017). Table 0/18. Taigikho (失控的台語課).