

SubGlyphs: 通用拼合注音字系統

HyLu

Introduction

There are many East asian tonal languages using Han characters as their writing system, either is in or influenced strongly by Sinitic languages. Despite, only Mandarin, as an official language, has stable, mature and common writing system. Some quasi official languages like Taiwanese, Hakka and Cantonese, have developed a convention writing standard. But they are still facing some problems.

When adopting pure Han character writing system, the main problem goes to those words without conventional characters:

- New characters: Cantonese writing system tends to use this method, like '嚟'. Usually they are created with a '口' and a phonetic component. Though the phone is mostly correct, the character is sometimes complicated, and new words may not be supported by Unicode.
- Character with similar Mandarin phone: Taiwanese writing system tends to use simple characters with this feature, like '袂'. Despite, these characters are neither correct in pronunciation nor related in meaning.
- Character with the same pronunciation in Taiwanese: also a common method in Taiwanese writing system, like '魯肉飯' (Minced meat rice), but may leads to misunderstanding. For instance this was once translated as 'Shandong meat rice', since he thought '魯' stands for Shandong province.
- Original character: It seems to be the best method. Nevertheless, some languages have words either without Sinitic origin or the original character is difficult to read and write. For instance, more than 80% people does not know character '薦' (wild berries), which is the original character of Taiwanese 'pho' in word 'tsì-pho'.

When adopting pure latin letter writing system, the main problem is the system design problem.

- Differences of phonology: Latin letters are not the best system to write East asian tonal languages, which leads to complicated digraph, trigraph or multigraph. For example, the word 'sound of biting crispy food' will be written as a complicated syllable 'khàunnh', including digraph onset kh, quadragraph rime aunnn, coda -h.

- Hyphen: it is significant to represent tone sandhi unit for languages with highly developed tone sandhi. Thus, the pure latin writing system will be full of hyphens, causing low readability and typesetting beauty. The word of 'biting crispy food', to exemplify, appears mostly in the form 'kháunnh-kháunnh'.
- System clash: The developed system will be too influencing and forces other systems to avoid clashing with it. For example, some Chinese dialect systems have to avoid use 'b' for 'voiced bilabial plosive' even that it seems to be a natural choice, since letter 'b' stands for 'tenius bilabial plosive' in Hanyu Pinyin. They have to use 'bb' or 'bh' instead.

The Han character - lating letter mixed writing system seems to be a good solution. Despite, it is a nightmare for typesetting.

- It looks immature, like a primary school student use latin letters instead when he/she encounters a complicated character.
- Chaos visual performance: latin letters and east asian symbols have developed highly mature typeset art in history. But they clash to each other when mixed: different latin letters has different width to make typeset fluent and relaxing, while east asian symbols make each symbol sharing the same width to make typeset unified and clean. The mixed typeset will introduce serious sense of incongruity. For literal works, it is unacceptable.

Now that mixed typesetting is a potential way, I was inspired by Japanese and Korean writing systems, both with long history of mixed typesetting:

- Japanese: characters and kana mixed writing system. It keeps the ideography and pronunciation simultaneously, flexible and useful.
- Korean writing system: Hangul system compresses each syllable into a Han character size. It supports either mixed typeset or pure Hangul typeset.

Therefore, I tried to proposed a system combining the advantages: an expandable phonogram for East Asian tonal languages based on Bopomofo and referring to Hangul compression. So that it may be mixed with Han characters like Japanese kana.

To avoid new unicode, I used some extra symbols to expand current Bopomofo system:

- Japanese (ゝ) and (゜) .
- Some strokes, simple Han characters and Japanese Katakana.

At present, in order to display all symbols, I used font TW-98-Kai (and its additional font) and Biau-kai. In the future, if this system may be used and widely spread so that unicode accepts new Bopomofo symbols, the Japanese katakana and simple Han characters will be replaced to unify the style.

第 1 章 Onsets

The macro for onset glyphs is `\bpmf{Main-code,Sub-code}`, defined in `SGBasic.sty`.

- Main-code: fill the symbol codes shown below, Place of articulation first and then manner. For instance, `\bpmf{pq}` represents ㄑ. A Question mark will be displayed if it fails to parse the main-code.
- Sub-code: Please omit it or use o if no diacritical symbols is needed. Dakuten is v and hann-dakuten is e. The case of failing to parse is a superscript question mark. Only the 'half-width' comma is accepted.

Glottal nasal is an impossible phoneme in IPA. Since in some languages the nasalization of vowel shares the same function as normal nasal consonant, I utilize this grid to represent 'Nasalization as the function of coda'.

Table 1: Table of Onset

Manner\Place	bilabial p	alveolar t	alveolar s	post- alveolar c	retroflex r	velar k	glottal h	labio- velar w
Tenius q	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ	ㄗ	ㄘ
Aspirated a	ㄑˊ	ㄒˊ	ㄓˊ	ㄔˊ	ㄕˊ	ㄖˊ		ㄘˊ
Voiced q,v	ㄑˊ	ㄒˊ	ㄓˊ	ㄔˊ	ㄕˊ	ㄖˊ		ㄘˊ
Voiced z	ㄑˊ		ㄓˊ	ㄔˊ		ㄖˊ		ㄘˊ
Nasal n	ㄑ	ㄒ		ㄔ	ㄕ	ㄖ	(ㄗ)	
Lateral l		ㄒ						
Voiceless f	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		
Voiced f,v	ㄑˊ	ㄒˊ	ㄓˊ	ㄔˊ	ㄕˊ	ㄖˊ		
Voiced v	ㄑ	ㄒ		ㄔ	ㄕ	ㄖ		

In some languages, half-vowel are viewed as onsets. Despite, Bopomofo does not differ half-vowel with its vowel counterpart. Therefore, I use the symbols of the vowels for glides. It might be modified if there is any expansion of change in the future.

- `\bpmf{i}` = ㄟ ;
- `\bpmf{u}` = ㄨ ;
- `\bpmf{y}` = ㄩ °

第 2 章 Rime

第 2.1 節 Single vowel

The macro for rimes is `\bpmf{main-code,sub-code}`. It is defined in `SGBasic.sty`.

- Main-code: Openess (Two letters), Backness(One letter), Roundedness (One letter). For example, 'Close Front Unrounded Vowel (一)' is `\bpmf{ccfu}`. If they are not enough, please fill main-code with arabic 1 9 and use Suzhou numeral as temporary symbols.
- Sub-code: Same with Onset part.

The symbols for nasalized vowels in Bopomofo Extended are used as other normal vowels. It is because one may either use nasalization-coda or use sub-code to decorate normal vowels to represent its nasalized counterpart. This system use normal consonant symbols for syllabic consonants, as most Romanization systems do.

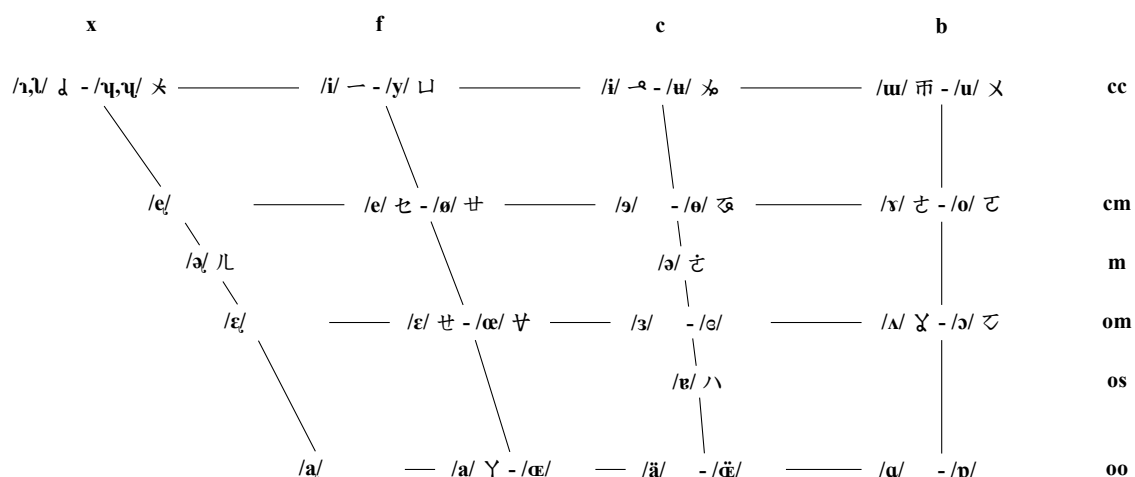


Figure 1: 單元音舌位圖

第 2.2 節 diphthongs

The nasalized rimes are also used for other pre-combined rimes.

Table 2: 預組元音

Start of main-code	coda nucleus	ㄚ a	ㄝ e	ㄛ o	ㄜ oo	ㄧ i	ㄩ u
vc	ㄧ i	ㄚ	ㄝ	ㄛ	ㄜ		ㄩ
vc	ㄩ u	ㄘ	ㄝ	ㄛ	ㄜ	ㄩ	
nc	ㄇ m	ㄇ		ㄇ			
nc	ㄋ n	ㄋ	ㄣ				
nc	ㄋ ng	ㄋ	ㄣ	ㄣ			

第 3 章 Tone and Coda

The macro for tones is `\ToneCoda{ Tone Type } { Tone Class } { Coda }`. It is defined in `SGBasic.sty`. The symbols follow the rules below:

- Tone type: Leveling, rising, falling and entering are p, s, q and r, respectively.
- Tone class: yin, yang and neutral are i, a and z respectively. Neutral means it is an expanded tone from other tones, and does not indicate that the 'yin' or 'yang' tone class is unspecified. For instance, 'Down Yin entering' in Cantonese will be recorded as 'Neutral Entering Tone', or 'p, z'.
- Coda: m, n and ng can be used in all four tone types and will be transformed to p, t, k when they encounters entering tone. Meanwhile, p, t, k codas are used in Entering tone only, and use them to match any other tones will fail to parse. If required, please use the symbol code in this and there will be expanded tone-coda form, like the example of 'nasalization as special coda' provided in the table below.

Table 3: Combination of Tone and Coda

Coda\Tone	Leveling	Rising	Falling	Entering
Empty	、	ㄥ	ㄣ	ㄛ
Labial	ㄇ	ㄇ	ㄇ	ㄇ
Coronal	ㄋ	ㄋ	ㄋ	ㄋ
Dorsal	ㄌ	ㄌ	ㄌ	ㄌ
Nasalized	ㄣ	ㄣ	ㄣ	ㄣ

第 4 章 Stacking

第 4.1 節 Basic Layer

Basic stacking macro defined in SGBasic.sty is as below:

`\SubGlyph{Onset}{Glide}{Nucleus}{Coda extended}{coda}{Tone type}{Tone class}`。

Please leave empty brackets for non-used part. The parts are introduced below:

- Onset: onset symbol code
- Glide: write i, u or y, or leave it empty
- Nucleus: rime symbol code
- Coda extended: off-glide i, u, y if there is another coda
- Coda: coda coupled with tone
- Tone type and tone class: tone symbol

Below is the example of onset ㄅ, rime ㄩ.

Table 4: Stacking Example

Coda	Yin				Yang				Neutral			
	Leveling	Rising	Falling	Entering	Leveling	Rising	Falling	Entering	Leveling	Rising	Falling	Entering
Empty	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ
Labial	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ
Coronal	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ
Dorsal	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ
Nasalized	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ	ㄅ	ㄆ	ㄇ	ㄏ

第 4.2 節 Mapping Layer

The basic layer macro is too complicated. For instance, one has to derive SuGlyphs of 'xióng' in Hanyu Pinyin as below:

- x for onset 'ㄒ', main-code of which is cf.
- iong is intransparent, equivalent to 'ü+eng', which should be separated to 'ㄩ' and 'ㄥ' with main-code 'y' and 'nceng'.
- Tone 2 is Yang leveling, with tone type 'p' and tone class 'a'.
- From above, it is `\SubGlyph {cf}{y}{nceng}{ }{ }{p}{a}`, which will be displayed as ㄒㄩㄥ².

Thus, the package provides six examples of languages and an interface SubGlyph.sty. If understanding the onset, rime and tone of the romanization system, one may directly produce the SubGlyph with interface macro. For example: `\ToSubGlyph {Hanpin} {x} {iong} {2}` will be displayed as ㄒㄩㄥ^2 as well.

Below are examples of the six language provided. Please contact me if there is any mistake.

Table 5: Example: Pronunciation of Numbers

Language	Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten
Mandarin	ㄓ	一	二	三	四	五	六	七	八	九	十
Taiwanese Colloquial	ㄓ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ
Taiwanese Literary		ㄟ		ㄟ	ㄟ	ㄟ	ㄟ		ㄟ	ㄟ	ㄟ
Hakka (Xiian)	ㄓ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ
Cantonese	ㄓ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ
Suzhounese	ㄓ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ
Fuzhounese	ㄓ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ	ㄟ

I provided python script as well. Once you refer to the .py files to write your own romanization system, the .sty file may be generated automatically. The script also updates the interface SubGlyph.sty, and you do not have to write `\usepackage {SGYoursystem.sty}`. To avoid name clash, I recommend you to name your system as:

- If the system is highly standardized, please use the abbreviated name of the language and the pronunciation of '拼' or '羅' in your language, like 'Jyutping (粵拼)', 'Tailo (台羅)', 'Hanpin (漢拼)'.
- If the system is a subsystem of a large system, please attach the region name of the subsystem. For example, 'WuphinSoutseu (吳拼蘇州)', 'HagpinXiian (客拼四縣)'. Please do not use Hanyu Pinyin region name if the system has different spelling to further avoid name clash.

第 4.3 節 Example of Mixed Typetting

- Mandarin: 「和」讀音異常複雜，有ㄏ、ㄉ、ㄆ、ㄌ、ㄆ等讀音。(The pronunciation of '和' is extremely complicated. There are ...)
- Taiwanese: 伊佇ㄟ毋知企佢久矣。(Nobody knows how long he has been standing there.)

第 5 章 Conclusion

The proposed system has advantages comparing to pure Han character, pure latin letter or Han-Latin mixed writing systems:

- East Asian typeset style: Phonograms with the size of one Han character, unified typeset style while avoiding the misunderstanding.
- Modularized phones: in theory, one may design a matched system for each East Asian tonal languages.
- Extension potential: might be used as an unified phonogram system for East Asian languages.

Despite, this is a prototype. It still has problems below:

- Font unchangeable: The current optimization is based on the display of TW-Kai-98 with TikZ. If the font changes, misplacement may happen.
- Font ununified: some symbols are not displayed correctly in TW-Kai as Kai, for which some characters are assigned as Biau-Kai. Baselines and weights may not fit perfectly.
- Line style: A result of TikZ scaling. The only way to overcome it is to design a new font, which is beyond my capability.
- May not be supported by Unicode: Unicode pre-combined 11172 Hanguls. But this system has even more possible combinations than Hangul.

In all, this is my prototype. I hope linguistic experts, symbol designers and tech groups may develop complete, elegant and powerful system for East Asian tonal languages and achieve the performance of Japanese mixed typesetting. This will highly benefit the language protection, inheritance and literal development.