

SubGlyphs: 通用拼合注音字系統

HyLu

前言

漢語族語言 (Sinitic languages) 雖種類繁多，然迄今唯有國語／普通話具備官方地位與成熟、普及的書面語體系。其他語言如台語、客語、粵語等，雖具有準官方地位，並各自發展出一套約定俗成的書寫標準，然而仍然面臨一些問題。

採用「純漢字書寫」時，遇到的主要問題是「有音無字」的處理困境：

- 造新字：粵語白話文傾向創造新字（如「嚟」），通常以口字旁加聲旁構形，雖然較能準確表音但字型可能複雜；
- 用華語音近字：台語白話文則常寫作字型簡單、音近的華語字（如「袂」），惟這一做法致使「不表音、不表意」的現象普遍存在；
- 用同音字：台語白話文也經常使用，如「魯」肉飯，但失去表意能力，例如曾經有外國人將其翻譯為「Shandong meat rice」（誤以為「魯」是山東的意思）；
- 用本字：看似是最佳方案，但首先台語、粵語乃至吳語、閩東語等都存在底層非漢語詞，其次即使有本字也可能生僻難解，如表野草莓之「蔗」字不認識者十之八九。
- 純拉丁書寫：雖然有一定社群，但此外，變調豐富的語言如台語、福州話或蘇州話等，拉丁字母無疑需要引入大量的連字符表示變調單位，

採用「純拉丁書寫」時，遇到的主要是方案設計困境：

- 語言差異：拉丁字母擅長表達的音節與漢語族差異太大，這導致羅馬化方案一般都存在複雜的二合、三合乃至多合字母。例如，表達「咀嚼脆的食物」的擬聲詞以台羅方案書寫是難以認讀的「kháunnh」，包括二合字母 kh、四合韻母 aunnh、入聲韻尾 h 等部分，而這個字往往以疊詞的形式出現。
- 連字符：連讀變調豐富的語言需要明確區分變調單位，這導致書寫時引入大量的連字符，可讀性與美觀大打折扣，還是以上一個例子，該擬聲詞往往是疊詞的形式「kháunnh-kháunnh」。
- 方案衝突：成熟的拼音方案往往有極大的影響力，迫使設計其他語言的方案時不得不配合已經成熟的方案，例如一些中國民間的拼音方案不得使用 bb 或 bh 表達「濁雙唇塞音」，因為 b 已經被漢語拼音用來表達「清不送氣雙唇塞音」了。

漢字與羅馬字混寫雖然看似綜合了上述兩者的好處，但它在美學上缺點極其顯著：

- 看起來極度不成熟，像是國小學生遇到不會寫的字，就寫成拼音代替那樣。
- 排版美感糟糕：拉丁文字與東亞文字各自發展出了成熟的排版美學，但這兩種美學本質上有衝突：拉丁文字讓每個字元有不同的寬度使得書面流暢靈動，東亞文字保證每個字元等寬使得書面整齊乾淨。混合排版時就會發生嚴重的不協調感，理工類書籍可以接受，但文學類書籍的排版美學也是其重要的組成部分。

既然混排看起來是有潛力的方向，我們可從已有成熟東亞文字混排歷史的日文與韓文的書寫策略獲得啟發：

- 日文的策略在於：能書寫則用漢字，無合適漢字則改用假名標音，既保留語意，也兼顧語音，實用而靈活。
- 韓文的策略則體現在：設計出諺文體系，將音節結構壓縮於一個「漢字大小」的方格中，實現書寫單元的簡潔與規律。

有鑑於此，本文嘗試提出一種結合兩者優點的書寫設計：以注音符號為基礎，並參考韓文的音節壓縮方式，將單一音節（含擴展符號）封裝為一個方塊單位，進而配合類似日文的漢字—假名混排機制，達成語音準確與排版美觀的平衡。

在技術層面上，為避免新增 Unicode 碼位，本文提出的擴展策略採用既有字符結合方式表現濁音與鼻化：

- 使用濁點（゛）與半濁點（゜）擴展符號表；
- 適當引入一些筆畫簡單的漢字或片假名字元；

目前，為兼容所有必要字符，本文使用支援完整注音符號集及片假名的字型 TW-98 楷體（及其增補字型）。展望未來，若本系統獲得應用與推廣，致使 Unicode 接納新注音符號字元，則期逐步汰換借用的日文假名與簡單漢字，構建更加純粹注音風格的注音合字系統。

第 1 章 聲母

調用聲母符號的宏為 `\bpmf{主碼, 附碼}`，定義於 `SGBasic.sty`。

- 主碼：填入下表所示的調音碼，先寫調音部位、再寫調音方式。例如，`\bpmf{pq}` 表示ㄑ。如果主碼解析失敗，會產生一個問號。
- 附碼：無附加符號可以省略或填入附碼 `o`，濁點的附碼為 `v`，半濁點的附碼為 `˘`。如果附碼解析失敗，則產生一個上標問號。務必注意，該宏只能解析「半形」逗號。

聲門鼻音本來在國際音標是不存在的符號，鑑於某些語言裡鼻化與鼻音韻尾有同樣的音系地位，我以不可能發音的聲門鼻音代表「地位相當於鼻音韻尾的鼻化」，以便形式一致。

Table 1: 聲母表

方式\部位	唇 p	齦 t	齦齶 s	齦腭 c	捲舌 r	軟腭 k	聲門 h	唇腭 w
全清 q	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ	ㄙ	ㄖ
次清 a	ㄒ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		ㄖ
全濁 q[v]	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		ㄖ
全濁 z	ㄑ		ㄓ	ㄔ		ㄖ		
次濁 n	ㄒ	ㄒ		ㄔ	ㄕ	ㄖ	(ㄙ)	
次濁 l		ㄒ						
清 f	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		
濁 f[v]	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		
濁 v	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		

部分語言需要把近音分析為聲母，不過注音符號沒有區分近音與元音，因此我暫且讓介音命令調用對應元音的注音符號。如果未來有拓展或修訂，可直接修改介音符號的命令。

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

第 2 章 韻母

第 2.1 節 單元音

調用元音符號的宏為**\bpmf{主碼, 附碼}**，定義於 SGBasic.sty。

- 主碼：開口度（兩個字母）、舌位（一個字母）、圓唇度（一個字母）。例如，「閉前展唇元音（一）」為**\bpmf{ccfu}**。如果四字母主碼不夠用，填入阿拉伯數字 1~9 可得到楷體蘇州碼子作為臨時符號。
- 附碼：與輔音一致。

方音符號裡表示鼻化韻的符號，在這裡拿來表示非鼻音的單元音。這是因為鼻化既有形式韻尾符號，又可以用附碼，無需單獨符號。鑑於羅馬字系統不區分普通輔音與輔音成音節的拼式，輔音成音節在本系統使用普通輔音注音符號。

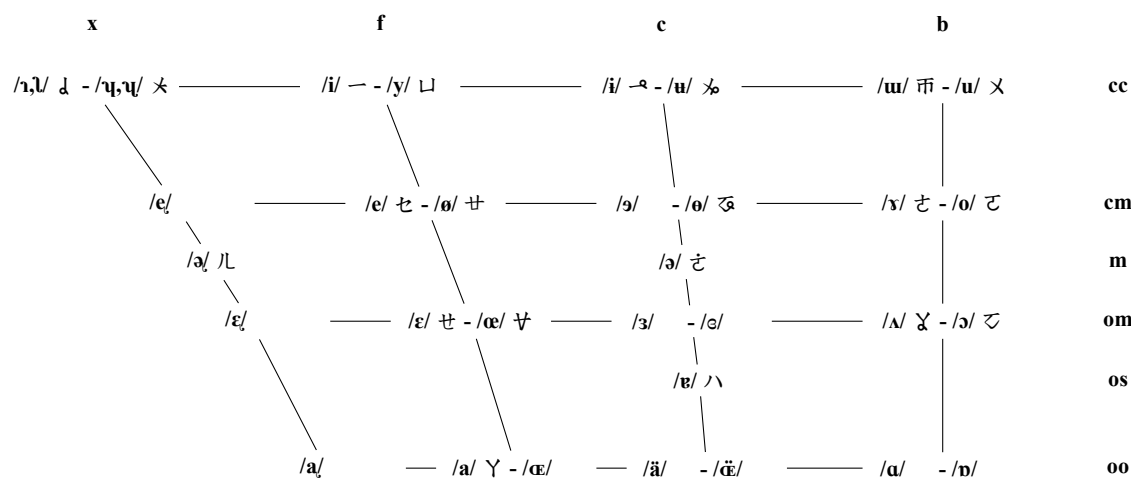


Figure 1: 單元音舌位圖

第 2.2 節 雙元音

原本的鼻化韻符號也改為非鼻化的其他預組元音。

Table 2: 預組元音

字串開頭	韻尾 韻腹	ㄚ a	ㄟ e	ㄛ o	ㄛ oo	一 i	ㄨ u
vc	一 i	ㄚ	ㄟ	ㄛ	ㄛ		ㄨ
vc	ㄨ u	ㄚ	ㄟ	ㄛ	ㄛ	ㄨ	
nc	ㄇ m	ㄚ		ㄛ			
nc	ㄋ n	ㄚ	ㄟ				
nc	ㄋ ng	ㄚ	ㄟ	ㄛ			

第 3 章 聲調

調用聲調符號的宏為\ToneCoda{調型碼}{調類碼}{韻尾碼}，定義於 SGBasic.sty 內。聲調標識遵循以下標準：

- 調型碼：平、上、去、入分別對應 p, s, q, r。
- 調類碼：陰、陽、中分別對應 i, a, z，中調只是表明其屬於擴展調型，並不表示其無法明定陰陽調型，例如粵語之「下陰入」調。
- 韻尾碼：m, n, ng 可以應用於平上去入四個調型，在遇到入聲時會自動轉為 p, t, k 形式；p, t, k 韻尾只能應用於入聲調，應用於其他調型會導致解析失敗；若有需求，可以使用對應輔音或元音的碼產生擴展標調，例如以下範例提供了填入 hn 產生鼻化音的四聲。

Table 3: 聲調韻尾組合表

韻尾\聲調	平	上	去	入
無	、	ㄥ	ㄛ	ㄜ
唇	ㄇ	ㄣ	ㄌ	ㄎ
舌尖	ㄣ	ㄣ	ㄣ	ㄣ
舌根	ㄣ	ㄣ	ㄣ	ㄣ
鼻化	ㄣ	ㄣ	ㄣ	ㄣ

第 4 章 拼合

第 4.1 節 基本層

定義在 SGBasic.sty 的底層拼合命令為：

`\SubGlyph{聲母}{介音}{韻腹}{韻尾一}{韻尾二}{調型}{調類}`。

用不到的部分直接在花括號中間留空，以下說明各個部件的格式。

- 聲母：同聲母部分
- 介音：直接寫 i、u、y 或留空。
- 韻腹：同韻母部分。
- 擴展韻尾：顯示為與韻母拼合的韻尾，一般是 offglide 形式的 i, u, y。
- 韻尾二：顯示為與聲調耦合的韻尾。
- 調型、調類：同聲調部分。

以下是示例表，以聲母ㄅ、韻母ㄩ為例。

Table 4: 拼合示例圖

韻末	陰				陽				中			
	平	上	去	入	平	上	去	入	平	上	去	入
無	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ
唇	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ
舌尖	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ
舌根	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ
鼻化	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ	ㄅ

第 4.2 節 映射層

底層命令過於繁瑣。例如，漢語拼音的「xióng」，需要歷經如下推導過程：

- x 對應聲母「ㄒ」，其主碼為 cf；
- iong 是一個不透明拼式，等效於「ü+eng」，應拆解為「ㄩ」與「ㄥ」，其主碼分別為 y 與 nceng；
- 華語的二聲對應陽平，其調型碼為 p，調類碼為 a；
- 組合以上資訊，調用代碼為 `\SubGlyph {cf}{y}{nceng}{}{}{p}{a}`，效果為ㄒㄩㄥ。

有鑑於此，package 內提供了六種語言的範例及接口 sty 檔案 SubGlyph.sty。只需要了解拼式的聲韻調分別是哪些，即可自動產生對應的通用拼合式注音，例如：`\ToSubGlyph {Hanpin} {x} {iong} {2}` 直接得到𪛗。

以下是六種語言的拼字示例。如有疏漏，敬請母語者不吝賜教。

Table 5: 數字讀音範例

語言	零	一	二	三	四	五	六	七	八	九	十
華語	𪛗	𪛘	𪛙	𪛚	𪛛	𪛜	𪛝	𪛞	𪛟	𪛠	𪛡
台語白讀	𪛢	𪛣	𪛤	𪛥	𪛦	𪛧	𪛨	𪛩	𪛪	𪛫	𪛬
台語文讀		𪛭		𪛮	𪛯	𪛰	𪛱		𪛲	𪛳	𪛴
四縣客語	𪛵	𪛶	𪛷	𪛸	𪛹	𪛺	𪛻	𪛼	𪛽	𪛾	𪛿
粵語	𪛢	𪛣	𪛤	𪛥	𪛦	𪛧	𪛨	𪛩	𪛪	𪛫	𪛬
蘇州話	𪛗	𪛘	𪛙	𪛚	𪛛	𪛜	𪛝	𪛞	𪛟	𪛠	𪛡
福州話	𪛢	𪛣	𪛤	𪛥	𪛦	𪛧	𪛨	𪛩	𪛪	𪛫	𪛬

作者也提供了 python 腳本，只需要整理該語言的羅馬字方案並參考寫好的檔案撰寫配置檔，即可自動產生對應的 sty 檔案。腳本會自動更新頂層的接口檔案 SubGlyph.sty，不需要手動新增`\usepackage {SGYoursystem.sty}`。為了防止撞名，建議命名方式為：

- 若該拼音方案標準化程度高，請直接使用該語言的簡稱加上該語言「拼」或「羅」的讀音，例如「Jyutping（粵拼）」、「Tailo（台羅）」、「Hanpin（漢拼）」。
- 若該拼音方案屬於一個大類的子方案，請在後面加上地域名稱在該方案中的拼式，例如「WuphinSoutseu（吳拼蘇州）」、「HagpinXiian（客拼四縣）」。華語音節組合少，羅馬化後極易重名，所以若您的語言羅馬字方案與漢語拼音有所差異，請直接使用該方案書寫地域名稱。

第 4.3 節 混合排版示例

- 華語：「和」讀音異常複雜，有𪛗、𪛘、𪛙、𪛚、𪛛等讀音。
- 台語：伊佇𪛢毋知企𪛣久矣。

第 5 章 結語

本系統比起純漢字、純拉丁或漢字拉丁混合排版有以下優勢：

- 排版東亞化：為一個漢字大小的純粹表音字符，既避免借用同音或諧音漢字的歧義，又保留了東亞文字排版一致性。
- 語音模組化：理論上可以依照各種漢語族語言的語音學特性設計出適合的拼合方式。
- 跨語言通用：系統保留了不少擴展空間，有潛力作為東亞聲調語言的共通表音文字。

不過，本系統僅為原型，尚存在以下問題：

- 字型無法遷移：由於 TikZ 與字型交互的特性，目前的最佳化都是使用 TW-Kai 調整的。改換字型後可能會發生意想不到的偏移錯位。
- 字型不統一：少數符號 TW-Kai 顯示為明體而非楷體，因此混用了一部分標楷體。字重與基線可能輕微不協調。
- 線條風格太細：這是 TikZ 縮放帶來的副作用，若想要克服這個問題，就需要設計專門的字型。作者暫時沒有這個時間、精力與能力。
- 即使僥倖推廣，Unicode 支援很可能等不到。以目前 Unicode 處理拼合文字諺文的策略來說，它傾向於枚舉。這個系統能寫出的拼合形式顯然遠超韓文。

總而言之，這是我的小小拋磚引玉，期望語言學者、文字設計專家與技術社群能發展出完整、優雅而功能強大的東亞聲調語言表音系統，並能達到類似日文那樣毫不違和地混合排版的效果，這應該是非常有利於語言保護、傳承與白話文學發展的工作。