

SubGlyphs: 通用拼合注音字系统

HyLu

前言

汉语族语言（Sinitic languages）虽种类繁多，然迄今唯有普通话具备官方地位与成熟、普及的书面语体系。其他语言如台语、客语、粤语等，虽具有准官方地位，并各自发展出一套约定俗成的书写标准，然而仍然面临一些问题。

采用“纯汉字书写”时，遇到的主要问题是“有音无字”的处理困境：

- 造新字：粤语白话文倾向创造新字（如“嚟”），通常以口字旁加声旁构形，虽然较能准确表音但字形可能复杂；
- 用普通话音近字：台语白话文则常写作字形简单、音近的普通话汉字（如“袂”），唯这一做法致使“不表音、不表意”的现象普遍存在；
- 用同音字：台语白话文也经常使用，如「鲁」肉饭，但失去表意能力，例如曾经有外国人将其翻译为“Shandong meat rice”（误以为“鲁”是山东的意思）；
- 用本字：看似是最佳方案，但首先台语、粤语乃至吴语、闽东语等都存在底层非汉语词，其次即使有本字也可能生僻难解，如表野草莓之“蔗”字不认识者十之八九。

采用“纯拉丁书写”时，遇到的主要是方案设计困境：

- 语言差异：拉丁字母擅长表达的音节与汉语族差异太大，这导致拉丁化方案一般都存在复杂的二合、三合乃至多合字母。例如，表达「咀嚼脆的食物」的拟声词以台罗方案书写是难以认读的“kháunnh”，包括二合字母 kh、四合韵母 aunh、入声韵尾 h 等部分，而这个字往往以叠词的形式出现。
- 连字符：连读变调丰富的语言需要明确区分变调单位，这导致书写时引入大量的连字符，可读性与美观大打折扣。还是以上一个例子，该拟声词往往是叠词的形式“kháunnh-kháunnh”。
- 方案冲突：成熟的拼音方案往往有极大的影响力，迫使设计其他语言的方案时不得不配合已经成熟的方案，例如一些中国民间的拼音方案不得使用 bb 或 bh 表达“浊双唇塞音”，因为 b 已经被汉语拼音用来表达“清不送气双唇塞音”了。

汉字与罗马字混写虽然看似综合了上述两者的好处，但它在美学上缺点极其显著：

- 看起来极度不成熟，像是小学生遇到不会写的字，就写成拼音代替那样。

- 排版美感糟糕：拉丁文字与东亚文字各自发展出了成熟的排版美学，但这两种美学本质上有冲突：拉丁文字让每个字符有不同的宽度使得书面流畅灵动，东亚文字保证每个字符等宽使得书面整齐干净。混合排版时就会发生严重的不协调感。理工科书籍或许可以接受，但文学类书籍的排版美学是其重要的组成部分，不成熟的排版美学会限制其书面文学的发展。

既然混排看起来是有潜力的方向，我们可从已有成熟东亚文字混排历史的日文与韩文的书写策略获得启发：

- 日文的策略在于：能书写则用汉字，无合适汉字则改用假名标音，既保留语意，也兼顾语音，实用而灵活。
- 韩文的策略则体现在：设计出谚文体系，将音节结构压缩于“一个汉字大小”的方格中，实现书写单元的简洁与规律。

有鉴于此，本文尝试提出一种结合两者优点的书写设计：以注音符号为基础，并参考韩文的音节压缩方式，将单一音节（含扩展符号）封装为一个方块单位，进而配合类似日文的汉字—假名混排机制，达成语音准确与排版美观的平衡。

在技术层面上，为避免新增 Unicode 码位，本文提出的扩展策略采用既有字符结合方式表现浊音与鼻化：

- 使用浊点（゛）与半浊点（゜）扩展符号表；
- 适当引入一些笔画简单的汉字或片假名字符；

目前，为兼容所有必要字符，本文使用支援完整注音符号集及片假名的字型 TW-98 楷体（及其增补字体）与标楷体。展望未来，若本系统获得应用与推广，致使 Unicode 接纳新注音符号字符，则期逐步汰换借用的日文假名与简单汉字，构建风格更加纯粹的注音合字系统。

第 1 章 聲母

调用声母符号的宏为 `\bpmf{主码, 附码}`，定义于 `SGBasic.sty`。

- 主码：填入下表所示的调音码，先写调音部位、再写调音方式。例如，`\bpmf{pq}` 表示ㄑ。如果主码解析失败，会产生一个问号。
- 附码：无附加符号可以省略或填入附码 `o`，浊点的附码为 `v`，半浊点的附码为 `˘`。如果附码解析失败，则产生一个上标问号。务必注意，该宏只能解析「半角」逗号。

声门鼻音本来在国际音标是不存在的符号，鉴于某些语言里鼻化与鼻音韵尾有同样的音系地位，我以不可能发音的声门鼻音代表「地位相当于鼻音韵尾的鼻化」，以便形式一致。

Table 1: 声母表

方式\部位	唇 p	龈 t	龈腭 s	龈腭 c	卷舌 r	软腭 k	声门 h	唇腭 w
全清 q	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ	ㄙ	ㄖ
次清 a	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ	ㄙ		ㄖ
全浊 q,v	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		ㄖ
全浊 z	ㄑ		ㄓ	ㄔ		ㄖ		
次浊 n	ㄑ	ㄒ		ㄕ	ㄖ	ㄙ	(ㄙ)	
次浊 l		ㄒ						
清 f	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		
浊 f,v	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		
浊 v	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ		

不分语言需要把近音分析为声母，不过注音符号没有区分近音与元音，因此我暂且让介音命令调用对应元音的注音符号。如果未来有拓展或修订，可直接修改介音符号的命令。

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

第 2 章 韵母

第 2.1 節 单元音

调用元音符号的宏为**\bpmf{主码, 附码}**，定义于 SGBasic.sty。

- 主码：开口度（两个字母）、舌位（一个字母）、圆唇度（一个字母）。例如，「闭前不圆唇元音（一）」为**\bpmf{ccfu}**。如果四字母主码不够用，填入阿拉伯数字 1~9 可得到楷体苏州码子作为临时符号。
- 附码：与辅音一致。

方音符号里表示鼻化韵的符号，在这里拿来表示非鼻音的单元音。这是因为鼻化既有形式韵尾符号，又可以用附码，无需单独符号。鉴于拉丁化方案不区分普通辅音与辅音成音节的拼写，辅音成音节在本系统使用普通辅音注音符号。

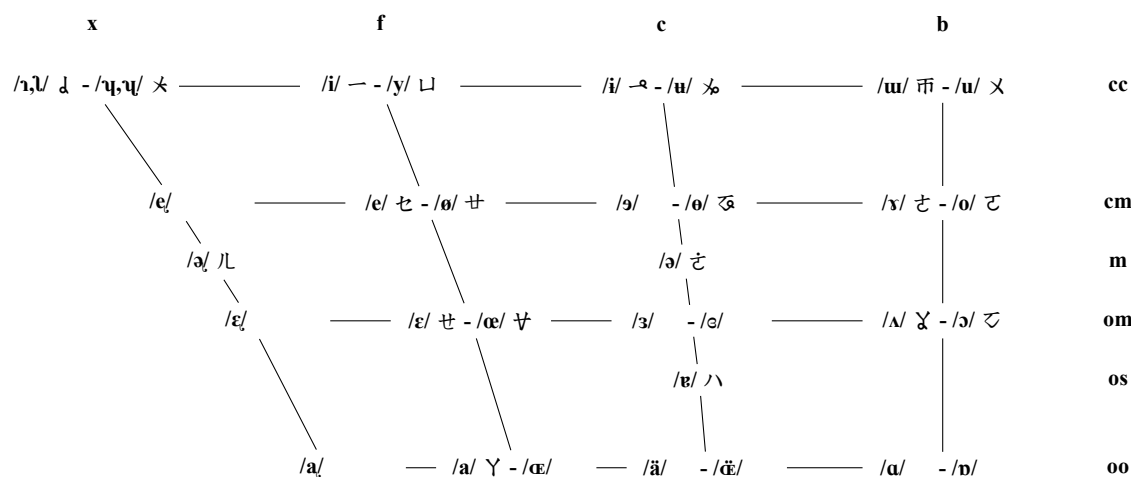


Figure 1: 单元音舌位圖

第 2.2 節 雙元音

原本的鼻化韵符号也改为非鼻化的其他预组元音。

Table 2: 预组元音

字符串开头	韵尾 韵腹	Y a	せ e	ㄛ o	ㄛ oo	一 i	ㄨ u
vc	一 i	ㄚ	ㄟ	ㄛ	ㄛ		ㄨ
vc	ㄨ u	ㄜ	ㄞ	ㄨ	ㄨ	ㄟ	
nc	ㄇ m	ㄇ		ㄇ			
nc	ㄋ n	ㄋ	ㄋ				
nc	ㄋ ng	ㄋ	ㄋ	ㄋ			

第 3 章 聲調

调用声调符号的宏为\ToneCoda{调型码}{调类码}{韵尾码}，定义于 SGBasic.sty 内。声调标识遵循以下标准：

- 调型码：平、上、去、入分别对应 p, s, q, r。
- 调类码：陰、陽、中分别对应 i, a, z，中调只是表明其属于扩展调型，并不表示其无法明定阴阳调型，例如粤语之「下阴入」调。
- 韵尾码：m, n, ng 可以应用于平上去入四个调型，在遇到入声时会自动转为 p, t, k 形式；p, t, k 韵尾只能应用于入声调，应用于其他调型会导致解析失败；若有需求，可以使用对应辅音或元音的码产生扩展标调，例如以下范例提供了填入 hn 产生鼻化音的四声。

Table 3: 声调韵尾组合表

韵尾\声调	平	上	去	入
无	、	ㄥ	ㄣ	ㄛ
唇	ㄇ	ㄇ	マ	ㄣ
舌尖	ㄣ	ㄣ	ノ	ㄣ
舌根	ㄣ	ㄣ	元	ㄣ
鼻化	ン	ㄣ	ㄣ	ㄣ

第 4 章 拼合

第 4.1 節 基本层

定义在 SGBasic.sty 的底层拼合命令为：

`\SubGlyph{声母}{介音}{韵腹}{韵尾一}{韵尾二}{调型}{调类}`。

用不到的部分直接在花括号中间留空，以下说明各个部件的格式。

- 声母：同声母部分
- 介音：直接写 i、u、y 或留空。
- 韵腹：同韵母部分。
- 扩展韵尾：显示为与韵母拼合的韵尾，一般是 offglide 形式的 i, u, y。
- 韵尾二：显示为与声调耦合的韵尾。
- 调型、调类：同声调部分。

以下是示例表，以声母ㄅ、韵母ㄩ为例。

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 映射层

底层命令过于繁琐。例如，汉语拼音的「xióng」，需要历经如下推导过程：

- x 对应声母「ㄒ」，其主码为 cf；
- iong 是一个不透明拼式，等效于「ü+eng」，应拆解为「ㄩ」与「ㄥ」，其主码分别为 y 与 nceng；
- 普通话的二声对应阳平，其调型码为 p，调类码为 a；
- 根据以上信息，调用代码为 `\SubGlyph{cf}{y}{nceng}{}{}{p}{a}`，效果为ㄒㄩㄥ。

有鉴于此，package 内提供了六种语言的范例及接口 sty 档案 SubGlyph.sty。只需要了解拼式的声韵调分别是哪些，即可自动产生对应的通用拼合式注音，例如：`\ToSubGlyph {Hanpin}{x}{iong}{2}` 直接得到𪛗。

以下是六种语言的拼字示例。如有疏漏，敬请母语者不吝赐教。

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 脚本，只需要整理该语言的拉丁化方案并参考写好的档案撰写配置档，即可自动产生对应的 sty 档案。脚本会自动更新顶层的接口档案 SubGlyph.sty，不需要手动新增`\usepackage {SGYoursystem.sty}`。为了防止撞名，建议命名方式为：

- 若该拼音方案标准化程度高，请直接使用该语言的简称加上该语言「拼」或「罗」的读音，例如「Jyutping（粤拼）」、「Tailo（台罗）」、「Hanpin（汉拼）」。
- 若该拼音方案属于一个大类的子方案，请在后面加上地域名称在该方案中的拼式，例如「WuphinSoutseu（吴拼苏州）」、「HagpinXiian（客拼四县）」。普通话音节组合少，拉丁化后极易重名，所以若您的语言拉丁化方案与汉语拼音有所差异，请直接使用该方案书写地域名称。

第 4.3 節 混合排版示例

- 普通话：「和」读音异常复杂，有𪛗、𪛘、𪛙、𪛚、𪛛等读音。
- 台语：伊佇𪛢毋知企偌久矣。

第 5 章 结语

本系统比起纯汉字、纯拉丁或汉字拉丁混合排版有以下优势：

- 排版东亚化：为一个汉字大小的纯粹表音字符，既避免借用同音或谐音汉字的歧义，又保留了东亚文字排版一致性。
- 语音模组化：理论上可以依照各种汉语族语言的语音学特性设计出适合的拼合方式。
- 跨语言通用：系统保留了不少扩展空间，有潜力作为东亚声调语言的共通表音文字。

不过，本系统仅为原型，尚存在以下问题：

- 字体无法迁移：由于 TikZ 与字体交互的特性，目前的优化都是使用 TW-Kai 字体调整的。改换字体后可能会发生意想不到的偏移错位。
- 字体不统一：少数符号 TW-Kai 显示为明体而非楷体，因此混用了一部分标楷体。字重与基线可能轻微不协调。
- 线条风格太细：这是 TikZ 缩放带来的副作用，若想要克服这个问题，就需要设计专门的字体。作者暂时没有这个时间、精力与能力。
- 即使侥幸推广，Unicode 支援很可能等不到。以目前 Unicode 处理拼合文字谚文的策略来说，它倾向于穷举。这个系统能写出的拼合形式显然远超韩文。

总而言之，这是我的小小抛砖引玉，期望语言学者、文字设计专家与技术社群能发展出完整、优雅而功能强大的东亚声调语言表音系统，并能达到类似日文那样毫不违和地混合排版的效果，这应该是非常有利于语言保护、传承与白话文学发展的工作。