

Yuè & the Sinitic Languages of the Pearl River Region — Unit 1.3

13th EACL Summer School, Universität Stuttgart

Hilário de Sousa 蘇沙

EHESS–CRLAO

membre associé

Monday 27 July 2026

丙午六月十四

hilario@bambooradical.com

Models of syllabification

框 ‘frame’

[k^h ɿ a ŋ ʔ] (Taiwan Mandarin; [ŋ] in Beijing Mandarin)

[k^h ɿ a ŋ ʔ] (Cantonese)

Initial–Final model: (C) | (G)V(X)^T

the traditional model used for Mandarin and most Sinitic languages

Initial	Final		
	medial glide	rime	
		nucleus	coda
C	G	V	X
k ^h	ɿ	a	ŋ

e.g. Taiwanese Braille
⠫⠨⠭⠎⠏⠒⠐⠶⠑⠔⠂

Models of syllabification

框 ‘frame’

[k^h ɿ̯ a ŋ ɿ̯] (TW Mandarin; [ŋ] in BJ Mandarin)

[k^h ɿ̯ a ŋ ɿ̯] (Cantonese)

often applied to many Yuè and S Pínghuà dialects

Onset–Rime model: (C)(G) | V(X)^T

(a model more familiar to SE Asian / Western linguistics)

Onset		Rime	
		nucleus	coda
C	G	V	X
k ^h	ɿ̯	a	ŋ

Models of syllabification

Onset–Rime model:

Onset		Rime	
		nucleus	coda
C	G	V	X
k ^h	ᵘ	a	ŋ

Most Yuè and S Pínghuà dialects are poor in medial glides
Standard Cantonese onsets with glides: [kᵘ- k^hᵘ- ᵘ- ɿ-]

Models of syllabification

(Instead of (C)(G)V(X)^T)

Standard Cantonese usually analysed as (C)V(X)^T

/k^w- k^{wh}- w- j- /

e.g. /k^{wh} aŋ ɿ/

/k^{wh}- / onset, /aŋ/ rime

Cantonese Braille

⠠⠠⠠ kw-aang(-1)

(vs. Taiwanese Braille

⠠⠠⠠ k-uang-1)

Many people do not distinguish the terminologies of “Initial vs. Final” and “Onset vs. Rime”

both are called 聲母-韻母 in Chinese

Models of syllabification

The changes from CGVX to CVX:

Merger of GV, e.g.

香 MC $hjaŋ^A$ > Canto. ($*hjaŋ >$) $/hœŋ\downarrow/$
(cf. HK Hakka $/hjaŋ\downarrow/$)

Deletion of G, e.g.

菊 MC $kju^w k^D$ > Canto. $/kuk\downarrow/$
(cf. HK Hakka $/k^hjuk\downarrow/$)

Deletion of V, e.g.

偏 MC p^hjen^A > Canto. $/p^hin\downarrow/$
(cf. HK Hakka $/p^hjen\downarrow/$)

Merger of GG, deletion of V, e.g.

勸 MC $k^h wjən^C$ > Canto. $/hyn\downarrow/$
(Nanning Pinghua $/k^{wh}in\downarrow/$)
(HK Hakka $/k^hjen\downarrow/$)

Romanisation

Cantonese transcription rarely taught in the education system



Romanisation



<https://www.threads.com/@tommykm.eth/post/DYWKMLskc9>

香港輸入法鄙視鏈: 倉頡 > 速成 > 筆畫 > 手寫 > 粵拼
Hong Kong hierarchy of disdain for [漢字] input methods:

- 倉頡 Cong1 Kit3 / Cāngjié (e.g. 倉 = OIAR 人戈日口)
- 速成 Cuk1 Sing4 (e.g. 倉 = OR 人口)
- 筆畫 Bat1 Waak6 (e.g. 倉 = 3445113251 丿、丶、冫一一 丨冫一)
- 手寫 Sau2 Se2 handwriting recognition
- 粵拼 Jyut6 Ping3 (e.g. 倉 = CONG)

Romanisation

There exist many Cantonese Romanisation schemes

Most commonly used nowadays amongst linguists:

Jyut6 Ping3 粵拼

(Linguistic Society of Hong Kong Cantonese Romanization Scheme

香港語言學學會粵語拼音方案) <https://lshk.org/jyutping-scheme/>

See, e.g.,

https://en.wikipedia.org/wiki/Comparison_of_Cantonese_transcription_systems

for comparison of Cantonese transcription systems

Beware: people may use “Jyutping” 「粵拼」 informally to mean other types of Cantonese romanisation schemes

Onsets

Onsets in Standard Cantonese

b /p/	p /p ^h /	m /m/	f /f/
d /t/	t /t ^h /	n /n/	l /l/
z /ts/	c /ts ^h /		s /s/
g /k/	k /k ^h /	ng /ŋ/	h /h/
gw /k ^w /	kw /k ^{wh} /		w /w/

- zero onset is represented by [zero/nothing]
e.g. 屋 /ukɿ/ *uk1* 'house'
- *b-* *p-*, *d-* *t-*, *z-* *c-*, *g-* *k-*, *gw-* *kw-* are voiceless (unaspirated vs. aspirated)
- some speakers have merged *n-* into *l-*, e.g. 你 *nei5* / *lei5* 'you (sg)' (vs. 李 *lei5* surname)

Onsets

Onsets in Standard Cantonese

b /p/	p /p ^h /	m /m/	f /f/
d /t/	t /t ^h /	n /n/	l /l/
z /ts/	c /ts ^h /	s /s/	j /j/
g /k/	k /k ^h /	ng /ŋ/	h /h/
gw /k ^w /	kw /k ^{wh} /		w /w/

- z- c- s- =
[ts- ts^h- s-] ~ [tʃ- tʃ^h- ʃ-]

Onsets

Onsets in Standard Cantonese

b /p/	p /p ^h /	m /m/	f /f/
d /t/	t /t ^h /	n /n/	l /l/
z /ts/	c /ts ^h /		s /s/
			j /j/
g /k/	k /k ^h /	ng /ŋ/	h /h/
gw /k ^w /	kw /k ^{wh} /		w /w/

ng- and 0-:

- older speakers often turn 0- into *ng-* (e.g. 亞 *aa3* > *ngaa3*) (common in singing)
- many younger speakers do not have *ng-* (e.g. 我 *ngo5* > *o5*)
- 0- and *ng-* are nearly in complementary distribution: 0- with high tone (1 2 3; e.g. 矮 *ai2* ‘short stature’); *ng-* with low tone (4 5 6; e.g. 蟻 *ngai5* ‘ant’)

Onsets

Onsets in Standard Cantonese

b	p	m	f
/p/	/p ^h /	/m/	/f/
d	t	n	l
/t/	/t ^h /	/n/	/l/
z	c	s	j
/ts/	/ts ^h /	/s/	/j/
g	k	ng	h
/k/	/k ^h /	/ŋ/	/h/
gw	kw	w	
/k ^w /	/k ^{wh} /	/w/	

- exception 1: 勾鈎鉤 **gau1*
> *ngau1* (taboo avoidance; 勾鈎鉤 *ngau1* contrasting with 歐區 *au1*)
- exception 2: non-sinitic words like 𠵼 *ngaam1* ‘correct, just, exactly’ (cf. Zhuang *ngamq* ‘exactly’), Thai งาม /ŋa:m/ ‘pretty’; Viet *ngám* ‘to fit’)

Onsets

Onsets in Standard Cantonese

b /p/	p /p ^h /	m /m/	f /f/
d /t/	t /t ^h /	n /n/	l /l/
z /ts/	c /ts ^h /		s /s/
j /j/			
g /k/	k /k ^h /	ng /ŋ/	h /h/
gw /k ^w /	kw /k ^{wh} /		w /w/

- *h-* is glottal [h], e.g. 哈 [ha˥] (not velar [x] like Beijing Mandarin, 哈 [xa˥])
- some people merge *gwo kwo* into *go ko*
e.g. 廣 *gwong2* > *gong2*
(= 港 *gong2*)
- /j-/ is j-
- /j-/ is usually rounded in front of rounded vowel,
e.g. 完 'end' **jyun4** /jyn˥/ [ɥyn˥] ~ [jyn˥]

Codas

-p -t -k -m -n -ŋ -i -u

/-p -t -k -m -n -ŋ -j -w/ (and 0)

- /-p -t -k/: unreleased, e.g. [ap̚ at̚ ak̚]
- stigmatised trend: total or partial merger of /-k -ŋ/ into /-t -n/
e.g. 香港 [hœŋ̚ kœŋ̚] > [hœn̚ kœn̚], 北角 [pœk̚ kœk̚] > [pœt̚ kœt̚]
- newer trend: “indeterminate” merger of /-n -ŋ/ as something like $\tilde{V}$, and /-t -k/ as something like ?
e.g. 公安 [kœŋ̚ œn̚] > [kũ ɔ̃], 失約 [sœt̚ jœk̚] > [sœ? jœ?]

Codas

Stereotypical non-native accents, e.g.:

- Canto. /-p -m/ > /-t -n/:
e.g. Mandarin, Wu, Xiang
e.g. 深圳 [səmɿ tsənɿ] > [sənɿ tsənɿ]
- Canto. /-t -n/ > /-k -ŋ/:
e.g. 潮州 Teochew or 雷州 Luichew
e.g. 深圳 [səmɿ tsənɿ] > [səmɿ tsəŋɿ]

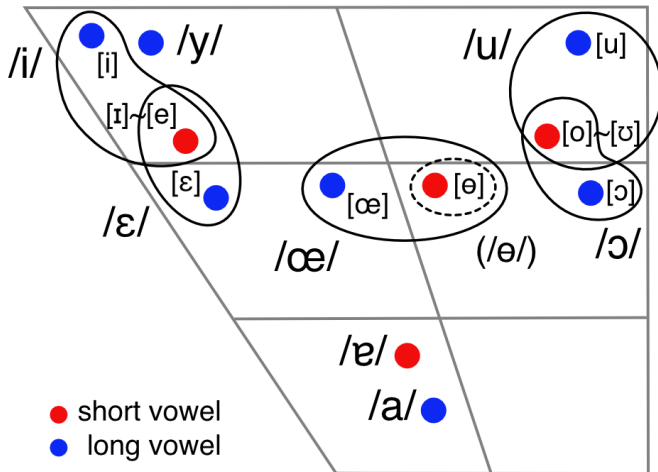
<https://www.tiktok.com/@leederu>

<https://www.douyin.com/video/7483381330437262626>

https://www.instagram.com/p/DLAr6_iNkY_/



Vowels



Based on Zee, Eric. 1999. Chinese (Hong Kong Cantonese). In The International Phonetic Association (ed.), *Handbook of the International Phonetic Association*, 58–60. Cambridge: Cambridge University Press.
Modified from https://upload.wikimedia.org/wikipedia/commons/c/c6/Cantonese_vowel_chart.svg

Rimes

	-0	/-p/	/-m/	/-t/	/-n/	/-k/	/-ŋ/	/-j/	/-w/
/i/	i [i]	ip [ip]	im [im]	it [it]	in [in]	ik [ek]	ing [eŋ]		iu [iu]
/u/	u [u]			ut [ut]	un [un]	uk [ok]	ung [oŋ]	ui [ui]	
/y/	yu [y]			yut [yt]	yun [yn]				
/ɛ/	e [ɛ]	ep [ɛp]	em [ɛm]	(et) [[ɛt]]	(en) [[ɛn]]	ek [ɛk]	eng [ɛŋ]	ei [ei]	eu [ɛu]
/ɔ/	o [ɔ]			ot [ɔt]	on [ɔn]	ok [ɔk]	ong [ɔŋ]	oi [ɔi]	ou [ou]
/œ/	oe [œ]			(oet) [[œt]]		oek [œk]	oeng [œŋ]		
				eot [ət]	eon [ən]			eoī [øy]	
/ɐ/	(a) [[ɐ]]	ap [ɐp]	am [ɐm]	at [ɐt]	an [ɐn]	ak [ɛk]	ang [ɛŋ]	ai [ɛi]	au [ɛu]
/a/	aa [a]	aap [ap]	aam [am]	aat [at]	aan [an]	aak [ak]	aang [aŋ]	aai [ai]	aaū [au]
0			m [m]				ng [ŋ]		

Rimes

粵拼 Jyut6 Ping3:

- /y/ is rendered **yu** (/j-/ is **j-**), e.g. 雨 ‘rain’ **jyu5** /jyɿ/ [ɥyɿ]
- only contrastive **long**–**short** vowel pairs are distinguished:
 - **aa** /a/ vs. **a** /ɐ/
 - **oe** /œ/ vs. **eo** /ə/ (*very marginally contrastive*)

(otherwise the vowel symbols are not distinguished,
e.g. **it** [it] vs. **ik** [ek])

- (unlike Dutch spelling) **long** /a/ in open syllables is still spelt **aa**, e.g. **jaa6** 廿 ‘twenty’

More generally:

- vowels in open syllables have to be **long**
(see below for exceptional **short a** [ɐ] in open syllables)
like e.g. most Germanic and Kra-Dai languages

Rimes

Cantonese “vowel length” contrast: involves both length and quality differences

Many perceptual studies on **long /a/** vs. **short /e/** (e.g. Zhang 2011, Zhang et al. 2011); opposing conclusions on whether duration cues or spectral cues are primary
(people probably normally rely on both)

Zhang, Ling. 2011. Vowel length perception in Cantonese. *Online Proceedings of the ICPHS XVII 2011*. International Phonetics Association. (<https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Zhang,%20Ling/Zhang,%20Ling.pdf>).

Zhang, Caicai & Peng, Gang & Wang, William S.-Y. 2011. Effect of language experience on the categorical perception of Cantonese vowel duration. *Interspeech 2011*, 169–172. ISCA. (<https://doi.org/10.21437/Interspeech.2011-74>).

Rimes

In Cantopop at least in the 80s/90s, when syllables are lengthened:

long vowels are nearly always sung

“long” (vowel lengthened, coda short)

short vowels can be sung long or short, usually

“short” (vowel short, coda lengthened)

Exception: a short vowel with a plosive coda is usually sung “long”

e.g. 合 hap6 [hɛ::p-]

Reminder: “vowel length” also involves vowel quality differences

Rimes

〈上海灘〉 Soeng6 Hoi2 Taan1 (1980)

轉千灣 轉千灘

zyun2 cin1 waan1 zyun2 cin2 taan1

everyone:

[tsy:n ts^hi:n wa:::n tsy:n ts^hi:n t^ha:::n]

Rimes

〈上海灘〉 Soeng6 Hoi2 Taan1 (1980)

亦未平 復此中爭 鬥

jik6 mei6 ping4 fuk6 ci2 zung1 zang1 dau3

葉麗儀 Jip6 Lai6 Ji4 Fancis Yip (1947 Hong Kong –)

[jek mej peŋ::: fok ts^{hi}: tson tsəŋ::: tɛ:w]

https://www.youtube.com/watch?v=PpVNdL9Wz_M&t=81s

甄妮 Jan1 Nei4 Jenny Tseng (1953 Macau –)

[jek mej peŋ::: fok ts^{hi}: tson tsəŋ::: tɛu::]

<https://www.youtube.com/watch?v=QZ9IC8MB30Q&t=74s>

Maria Cordero (肥媽 Fei4 Maa1) (1954 Macau –)

[jek mej pe::ŋ fok ts^{hi}: tson tseũre tɛu::]

<https://www.youtube.com/watch?v=by1-rYoGDWQ&t=87s>

Rimes

〈風的季節〉 Fung1 dik1 Gwai3 Zit3 (1981)

卻 也不需再 驚懼 風雨侵

koek3 jaa5 bat1 seoi1 zoi3 ging1 geoi6 fung1 jyu5 cam1

徐小鳳 Ceoi4 Siu2 Fung6*2 Paula Tsui (1949 Wuhan –)

[k^hœ::k ja: pet səɥ tsɔ::j keŋ køy:: foŋ jy: ts^hem::]

<https://www.youtube.com/watch?v=by1-rYoGDWQ&t=87s>

Soler (Giulio & Dino Acconci) (1972^{Macau} –)

[k^hœ::k ja: pet səɥ tsɔ::j keŋ kø:y: foŋ jy: ts^hem::]

<https://www.youtube.com/watch?v=YFs32GMc7u8&t=55s>

Rimes

Other comments on the segments:

- /i/ /u/ /y/ (except /u/[o]) cannot be preceded by [ʔ-]
TW Mand.^{initial-final} 雨 /yʌ/ [ɥyʌ] ~ [ʔyʌ]
Cantonese^{onset-rime} 雨 /(j)yʌ/ [ɥyʌ] * [ʔyʌ]

Jyutping spells the glide onset when it is present phonetically

雨 **jyu5** /(j)yʌ/ [ɥyʌ], 月 **jyut6** /(j)ytʌ/ [ɥytʌ]

熱 **jit6** /(j)itʌ/ [jitʌ], 翼 **jik6** /(j)ikʌ/ [jekʌ]

烏 **wu1** /(w)uʌ/ [wuʌ], 活 **wut6** /(w)utʌ/ [wutʌ]

(but 屋 **uk1** /ukʌ/ [(ʔ)okʌ], 甕 **ung3** /uŋʌ/ [(ʔ)onʌ])

Cantonese Braille spells these with zero onset,

except /jik/ [jek] /jin/ [jen]

雨 **y5** ⠨ ⠆, 月 **yt9** ⠨ ⠆.

熱 **it9** ⠨ ⠆, (but 翼 **jik9** ⠨ ⠆⠨ ⠆.)

烏 **u** ⠨ ⠆, 活 **ut9** ⠨ ⠆.

屋 **uk** ⠨ ⠆, 甕 **un3** ⠨ ⠆ ⠨ ⠆.

Rimes

- syllabic nasal **m** [m̩] **ng** [ŋ̩]:
cannot take an onset (except interjection 哼 [hm̩]~[hŋ̩])
唔 **m4** NEG
吳蜈吾梧𩚑 **ng4** 五伍午忤 **ng5** 誤悟晤梧梧寤忤𩚑 **ng6**
(from *ŋu, cf. 五官 Canto *ng5 gun1* Sino-Viet *ngũ quan*)
- Many people have merged syllabic **ng** into syllabic **m**
- codas /-j/ /-w/: Jyutping uses **-i -u**
phonological/phonetic studies may use vowel or
consonant symbols, e.g. **aa** /au/~/aw/, **eo** /œi/~/œj/
[əy]~[əɥ]

Rimes

- /uj/[ui] and /œj/[øy] are nearly in complementary distribution:

/uj/ [ui]: /p- p^h- m- f- w-/ (i.e. [LABIAL])

/œj/ [øy]: /t- t^h- n- l- ts- ts^h- s- j- h-/

contrast:

/kuiɿ/ [kuiɿ] **gui6** 𢵇/瘡 ‘tired’

/kœjɿ/ [kœjɿ] **geoi6** 𡇗 ‘huge’

/k^huiɿ/ [k^huiɿ] **kui2** 繪 ‘draw’

/k^hœjɿ/ [k^hœjɿ] (**kam4**) **keoi4*2** 蟾蜍 ‘toad’

/k^hœjɿ/ [k^hœjɿ] **keoi5** 𡇗 3SG)

(alternative model: [kui] /k^wui/, [kœj] /kui/)

Rimes

- **(et)**: only in ideophones (e.g. nam4 bet6 bet6 ‘(squishy)’) and *perhaps* non-native words (e.g. pet1 pet1 ‘bum’, get1 ‘get, comprehend’)
- **(en)**: only in non-native words (e.g. yen1 ‘(Japanese) Yen’)
- **eot** [ət] vs. **(oet)** ([œt]):
eot [ət] is the normal allorime,
e.g. 出 **ceot1** [ts^hətʰ] ‘exit’
(oet) ([œt]) is only found in ideophones,
e.g. 𠂇𠂇𠂇𠂇轉 cit1 **coet1** gam2 zyun3 ‘pace back and forth’, **oet1** ~ **oet6** [œtʰ] ~ [œtʰ] ‘belch, burp’

Rimes

- open rime **(a)** (**short [ɐ]**): only occurs in “unstressed” situations where **aa** [a] can be phonetically reduced
e.g. mou5 **laa3** can be lexical 冇罅 ‘there is no gap’
or grammatical 冇罅 ‘there is now none’ (sentence-final particle)
mou5 **la3** is grammatical 冇罅 ‘there is now none’