Early Sino Vietnamese sub-layers and a constraint against final *-s in Proto-Vietic

Thanh Viet Cao (高成越)

Graduate Institute of Linguistics
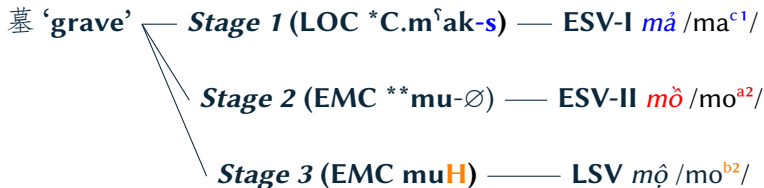National Tsing Hua University

Outline

Figure 1: *Nhật Tảo* bell, 948 CE

Setting the stage

- Some *Qusheng* (< Old Chinese ***-s**) Sinitic etyma are reflected in Vietnamese with A-tone (*ngang/huyền*) marked forms, in addition to C-tone (*ngã/nặng*) and B-tone (*sắc/nặng*) unmarked forms in *Early Sino-Vietnamese* (**ESV**) and *Literary Sino-Vietnamese* (**LSV**) layers, respectively.
- Unlike Wang Li (1948) who attributes these forms to the so-called *Vietnamized Sino-Vietnamese* (VSV) without any transparent phonological grounds, Alves (2018) assumes these to be early remnants of Early Middle Chinese (**EMC**) borrowed into atonal Vietic as toneless forms. This is the three-stage hypothesis on relativized tonogenesis in Chinese and Vietic.

Table 1: two examples of marked ESV bearing A-tone

Gloss	OC	MC	ESV	LSV
'grave'	*C.mˤak-**s**	muH	mả/mồ	mộ
'to forge'	*[r]ˤen-**s**	lenH	rèn	luyện

墓 **'grave'** —— *Stage 1* **(LOC *C.mˤak-s)** —— **ESV-I** *mǎ* /ma^{c1}/

Stage 2 **(EMC **mu-∅)** —— **ESV-II** *mò* /mo^{a2}/

Stage 3 **(EMC muH)** —— **LSV** *mộ* /mo^{b2}/

Figure 2: A simplified illustration for the three-stage hypothesis

- Given Alves' (2018) three-stage hypothesis, A-tone reflexes for *Qusheng* etyma (e.g. *mò* 'grave') can be regarded as items of an ESV sub-layer (let term it **ESV (II)**) mediating between the canonical ESV (i.e. the **ESV(I)**) and the LSV strata.
- **Stratifying problem**: Can this diagnosis be generalized to all cases? Specifically, on which phonological grounds can we identify items like *rèn* as **ESV(II)** but not **ESV(I)** (or just plainly **ESV**)?

Proposal in a nutshell

(i) Items like *rèn* (i.e. A-tone ESV ended with a nasal + don't have other ESV cognate) seem unable to be stratified on the tonal ground.

(ii) By appealing to segmental correspondences, a portion of them (e.g. *rèn* 'to forge', *tên* 'arrow', etc.) can be identified as earlier loans, i.e. **ESV(I)**, whereas others can be later ESV loans or cannot be specified in terms of timing.

Other than donor-related factors, all of these may have something to do with a receiver-related factor, i.e. a constraint against final ***-s** in Proto-Vietic.

(1) **The Constraint**
*The final sequence with a **nasal** followed by an **-s** is illicit in Proto-Vietic.*

Languages	Main sources
Vietnamese	Alves (2018), author's data (in Quốc Ngữ)
Vietic	Ferlus (2007), SEAlang database
Austroasiatic	SEAlang database
Old Chinese	Baxter & Sagart (2014), 古音小鏡 Database
Middle Chinese	Baxter & Sagart (2014), Kroll (2017)
Chinese varieties	MCC Database
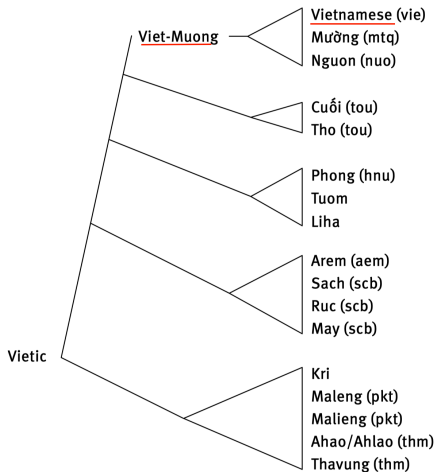
Figure 3: Vietic subgrouping following Sidwell & Alves (2021:183)
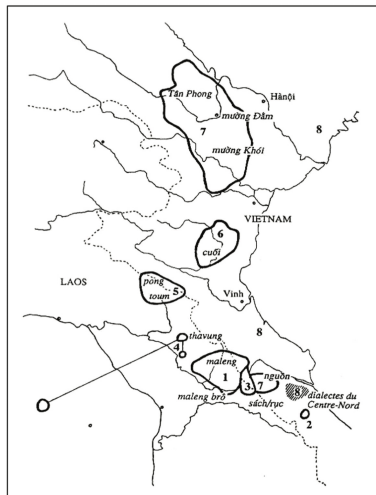
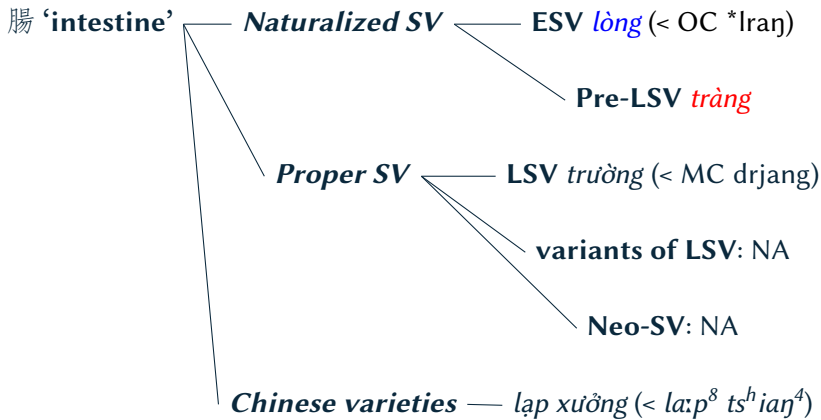**Figure 3:** Vietic languages distribution (Ferlus 1998:27)

The three-stage hypothesis

Sinitic loans in Vietnamese: A stratification

- Sinitic is the richest and most important non-AA layer in the lexicon of Vietnamese (cf. Wang Li 1948; Nguyễn Tài Cẩn 2000; Alves 2009, 2016, 2018, 2022; Vũ Đức Nghiệu 2011; among others)

- Presumably, there are 3 essential layers of Sinitic loans in Vietnamese:
  1. **Naturalized Sino-Vietnamese** (*before the 10th cent.*)
     - *Early Sino-Vietnamese* (i.e. ESV(I) in this study)
     - *Pre-Literary Sino-Vietnamese* (Pre-LSV, or ESV(II) in this study)

  2. **Proper Sino-Vietnamese** (*10th–12th cent.*)
     - *Literary Sino-Vietnamese* (i.e. Sino-Vietnamese Proper, LSV)
     - *Variants of LSV* (i.e. Post-Song variants of LSV)
     - *Neo-Sino-Vietnamese graphic loans* (i.e. Sino-neologisms)

  3. **Loans from Chinese varieties**: loans from Cantonese, Min languages, etc. in recent times (*since the 17th cent.*).

- **Early Sino-Vietnamese** is the earliest Sinitic layer, with some even borrowed from Late Old Chinese.

腸 '**intestine**' —— ***Naturalized SV*** —— **ESV** *lòng* (< OC *lraŋ)

Pre-LSV *tràng*

Proper SV —— **LSV** *trường* (< MC drjang)

variants of LSV: NA

Neo-SV: NA

Chinese varieties —— *lạp xưởng* (< *laːp⁸ tsʰiaŋ⁴*)

- Under Haudricourt's (1954) account, the tonal system developed in Vietnamese was a compensation for the loss of two codas in Proto-Vietic following an areal tendency (also happened in Sinitic, Kra-Dai, and Hmong-Mien languages by the same mechanism).

ta	ta**ʔ**	ta**h**
da	da**ʔ**	da**h**

⟶

ta	**tone**	ta	**tone**	ta	**tone**
da	**A**	da	**B**	da	**C**

The development from atonal to 3 tones (assumed in the 6th century) in <u>unchecked syllables</u> (i.e. without stop codas /-p, -t, -k/)

ta	**tone**	ta	**tone**	ta	**tone**
da	**A**	da	**B**	da	**C**

⟶

ta	**A1**	ta	**B1**	ta	**C1**
ta	**A2**	ta	**B2**	ta	**C2**

The development from 3 to 6 tones in un-checked syllables[1]

---

[1]Recent insights from register studies in Vietic suggest that this registrogenesis (i.e. the historical development of register) may have happened before the tonogenesis (cf. Brunelle & Ta 2021, Ta 2023, etc.)

Table 2: Alves' (2018:11-12) three-stage hypothesis (revised version)

Donor's stage	Evolution of tones		Relative chronology	Example: 墓 'grave (n.)'
	Chinese	*Vietic*		
Stage 1 (*with* *-s*)	**atonal**	**atonal**	*pre-Han* or Han eras	ESV *mả* (< LOC *C.mˤak-s)
Stage 2 (*with* **H**)	**tonal**	**atonal**	*Jin-Tang* era	ESV *mồ* (< EMC **mu)
Stage 3 (*with* **H**)	**tonal**	**tonal**	*post-Tang* era	LSV *mộ* (< LMC muH)

- **Implication** #1: The timing of Vietic tonogenesis might be later...
- **Implication** #2: LOC without *-s or EMC without *Qusheng* H
- **Implication** #3: All reflexes of *Qusheng* LOC/EMC with A-tone in Vietnamese are **ESV(II)** items (in parallel with mồ 'grave').

Table 3: Sino-Vietnamese triplets (2 ESV + 1 LSV)

Gloss	OC	MC	ESV(I)	ESV(II)	LSV
墓 'grave'	*C.mˤak-**s**	muH	mả	mồ	mộ
共 'together'	*N-k(r)oŋ?-**s**	gjowngH	cũng	cùng	cộng
帶 'belt'	*C.tˤa[t]-**s**	tajH	dải	đai	đái
利 'interest'	*C.ri[t]-**s**	lijH	lãi	lời	lợi
二 'two'	*ni[j]-**s**	nyijH	nhĩ	nhì	nhị
念 'sentiment'	*nˤim-**s**	nemH	ngẫm	niềm	niệm
萬 '10 thousand'	*C.ma[n]-**s**	mjonH	muôn	vàn	vạn

- **Implication** #2: LOC without *-**s** or EMC without *Qusheng* H?
- **Answer**: EMC without *Qusheng* H forms seem to be the donor for ESV(II) more frequently, given the convergence between ESV(II) and LSV reflexes in most of the cases.

• All reflexes of *Qusheng* LOC/EMC with A-tone in Vietnamese should be regarded as **ESV(II)** items (in parallel with mồ 'grave').

Table 4: Sino-Vietnamese doublets (1 ESV + 1 LSV)

#	Gloss	OC	MC	ESV(I)	ESV(II)
1	眾 'crowded'	*tuŋ-s	tsyuwngH	−	đông
2	劍 'sword'	*s.kr[a]m-s	kjaemH	−	gươm
3	放 'to release'	*paŋ-s	pjangH	−	buông
4	慣 'familiar'	*kˤro[n]-s	kwaenH	−	quen
5	箭 'arrow'	*[ts]en-s	tsjenH	−	tên
6	信 'believe'	*s-ni[ŋ]-s	sinH	−	tin
7	分 'part'	*[m]-pə[n]-s	pjunH	−	phần
8	算 'calculate'	*[s]ˤor?-s	swanH	−	toan
9	韻 'rime'	*[m-qʷ]<r>i[n]-s	hjunH	−	vần

• BUT, there's a missing piece in this picture...

A missing piece

Brunelle's remark

• However, there is a missing point in our up-till-now analysis concerning the phonotactics of Proto-Vietic.

> *As predicted by Haudricourt's reconstruction, tone C (qù, hỏi-ngã) does NOT occur on nasal-final syllables of Austroasiatic stock, where it would imply earlier sequences of* **a nasal followed by /h/**. *Tone C on nasal-final syllables is only found in borrowings from Chinese, and in words of expressive origin (Ferlus 2004: 299, citing Maspero).[2]*
>
> *Brunelle's remark in his translation of Haudricourt (1954)*

• There's evidence suggesting that **early** borrowings from Chinese also pattern with Austroasiatic natives with respect to this constraint.

---

[2]On the other hand, there exist numerous nasal-final syllables carrying tone B, suggesting that sequences of a nasal followed by /-ʔ/ are permitted.

• Unlike Old Chinese, Proto-Vietic lacks patterns ending with a nasal followed by a fricative, i.e. [CV-N-s] or [CvCV-N-s] .

• Given this lack, we can propose the constraint against the final *-s (and *-h) in Proto-Vietic as follows.

(1)   **The constraint**
      *The final sequence with a **nasal** followed by a fricative is illicit in Proto-Vietic.*

• In other words, there's a rule of deleting the final *-s (or *-h) after the penultimate nasals (viz. /m,n,ŋ/) for the [majority] of OC loans.[3]

(2)   **Final fricative dropping rule in Proto-Vietic**
      -s/-h > ∅ / [+nasal] _____#

---

[3]I set aside the conceptual distinction between a rule vs. a constraint.

The proposal

• Given the internal constraint against the final ***-s** in Proto-Vietic as observed:

(i) Items like *rèn* (i.e. A-tone ESV ended with a nasal + don't have other ESV cognate) seem unable to be stratified on the tonal ground.

(ii) By appealing to (segmental grounds), a portion of these items (e.g. *rèn* 'to forge', *tên* 'arrow', etc.) can be identified as earlier loans, i.e. **ESV(I)**, whereas others as later **ESV(II)** loans or cannot be specified in a more detailed chronology.

Table 5: a sample of data-to-be-substratified

#	Gloss	OC	MC	ESV(I)	ESV(II)
1	眾 'crowded'	*tuŋ-s	tsyuwngH	đông?	đông?
2	劍 'sword'	*s.kr[a]m-s	kjaemH	gươm?	gươm?
3	放 'to release'	*paŋ-s	pjangH	buông?	buông?
4	慣 'familiar'	*kˤro[n]-s	kwaenH	quen?	quen?
5	箭 'arrow'	*[ts]en-s	tsjenH	tên?	tên?
6	信 'believe'	*s-ni[ŋ]-s	sinH	tin?	tin?
7	分 'part'	*[m]-pə[n]-s	pjunH	phần?	phần?
8	算 'calculate'	*[s]ˤor?-s	swanH	toan?	toan?
9	韻 'rime'	*[m-qʷ]<r>i[n]-s	hjunH	vần?	vần?

• To be **ESV(I)** or **ESV(II)**? That's the question!

The analysis
& Concluding remarks

• Set tone aside, on what grounds can we employ to distinguish ESV(I) from ESV(II) items?

- **Initial-related indicators for ESV(I)**:
    - ▶ OC initial retentions: *k-, *r-, *b, *pʰ (> Middle Vietnamese *pʰ > f-)
    - ▶ OC initial alternations: **voicing** (*t- > ɗ-, *p- > ɓ-), **fortition** (*ts- > t-), **spirantization** (*C.k- > ɣ-)

- **Vowel-related indicators for ESV(I)**:
    - ▶ OC vowel retentions: *a, *e, *i, *o
    - ▶ OC vowel alternations: **diphthongization** (*a > uo/ɨə, *e > ie)[4], *u > o, *o > e

• Other than that, ⌈semantic change⌉ can play some role in the diagnosis (to be extended later)

---

[4]Both seemingly are conditioned by the presence of ***-r-**

Table 6: ESV(I) bearing A-tone

#	Gloss	OC	MC	ESV(I)	LSV
1	眾 'crowded'	*tuŋ-s	tsyuwngH	đông	chúng
2	劍 'sword'	*s.kr[a]m-s	kjaemH	gươm	kiếm
3	放 'to release'	*paŋ-s	pjangH	buông	phóng
4	慣 'familiar'	*kˤro[n]-s	kwaenH	quen	quán
5	箭 'arrow'	*[ts]en-s	tsjenH	tên	tiễn
6	鍊 'to forge'	*[r]ˤen-s	lenH	rèn	luyện
7	樣 'manner'	NA	yangH	dường	dạng
8	敬 'to respect'	*kreŋ(?)-s	kjaengH	kiêng	kính
9	印 'to seal'	*[?]iŋ-s	jinH	in	ấn
10	片 'flat piece'	*pʰˤe[n]-s	phenH	phên	phiến
11	便 'then, handy'	*ben-s	bjienH	bèn	tiện
12	弄 'play'	*[r]ˤoŋ-s	luwngH	rong/rông	lộng
13	陣 'period'	*lri[n]-s	drinH	lần	trận

Typical ESV(II) cases

• Signature traits: ESV(II) items have (identical segments) with LSV cognates.

Table 7: Timing-underspecified ESV

#	Gloss	OC	MC	ESV(II)	LSV
1	勸 'advise'	*C.qʷʰar-s	khjwonH	khuyên	khuyến
2	願 'wish'	*[ŋ]o[n]-s	ngjwonH	nguyền	nguyện
3	讓 'to yield'	*naŋ-s	nyangH	nhường	nhượng
4	嘆 'to lament'	*n̥ˤar-s	thanH	than	thán
5	信 'believe'	*s-ni[ŋ]-s	sinH	tin	tín
6	分 'part'	*[m]-pə[n]-s	pjunH	phần	phận
7	算 'calculate'	*[s]ˤor?-s	swanH	toan	toán
8	韻 'rime'	*[m-qʷ]<r>i[n]-s	hjunH	vần	vận

- The third possibility: underspecified cases

Table 8: Timing-underspecified ESV

#	Gloss	OC	MC	ESV	LSV
1	幔 'curtain'	*mˤa[n]-s	manH	màn	mạn

Problematic cases

• Given the constraint at issue, how do we account for C-tone (*ngã/nặng*) ESV items ending with nasal?

Table 9: Two problematic cases

Gloss	OC	MC	ESV(I)	ESV(II)	LSV
共 'together'	*N-k(r)oŋ?-**s**	gjowngH	cũng	cùng	cộng
念 'sentiment'	*nˤim-**s**	nemH	ngẫm	niềm	niệm

• **Suggestion**: a hidden factor related to phonation feature?

Remaining issues

(i) How to account for [retaining-*s] cases like ***cũng***/*cùng* 'also', ***ngẫm***/*niềm* 'sentiment' in ESV(I)?

(ii) Finer-grained donor forms: LOC without ***-s** or EMC without *Qusheng* H?

(iii) Can we extend the analysis to the final glottal *-ʔ in LOC?

(iv) The cross-linguistic perspective: can we find comparable data in other Sinospheric languages?

• There is evidence suggesting that there also was a medial, transitional stage of losing *-ʔ (leading to Shangsheng X) during LOC-EMC periods before the tonogenesis in Vietic.

• In other words, along the lines with the three-stage hypothesis on Chinese and Vietic tonal developments, it is very likely that these Sinitic *Shangsheng* etyma were borrowed into Vietic as toneless forms surfacing as A-tone reflexes.

Table 9: Evidence supporting a losing *-ʔ stage

# Gloss	Character	OC	MC	ESV	Pre-LSV	LSV
1 patch, make up for	補 bǔ	*[Cə]-pˤaʔ	puX	vá	bù	bổ
2 tub, bucket	桶 tǒng	*lˤoŋʔ	thuwngX	-	thùng	dũng
3 clf for book	部 bù	*[b]ˤ(r)oʔ	buwX	-	pho	bộ
4 summer	夏 xià	*[ɢ]ˤraʔ	haeH	-	hè	hạ
5 lazy	懶 lǎn	*[N-kəə.]rˤanʔ	lanX	-	lười	lãn

• Comparable data can also be found in Tai Yo, a Kra-Dai language spoken in North-Central Vietnam.

Table 10: Comparable data from Tai Yo [Kra-Dai]

Gloss	OC	MC	Tai Yo[5]	LSV
獻 'offer sacrifices'	*ŋ̊ar-**s**	xjonH	xển /seːn A1/	hiến
禁 'prohibit, taboo'	*kr[əə]m-**s**	kimH	kằm /kam A1/	cấm
晻 'dark'	*qˤ[o]m(ʔ)-**s**	ʼomH	ằm /ʔam A1/	ám

---

[5]Data informed by Sầm C.D. (p.c.).

- Unlike Old Chinese, Proto-Vietic lacks patterns ending with a nasal followed by a fricative, i.e. $\boxed{\textbf{CVN-s}}$ or $\boxed{\textbf{CvCVN-s}}$.
- In other words, a constraint against the final *-**s** (and *-**h**) in Proto-Vietic can be proposed as follows (cf. (1)).
- Given this constraint, there's a big portion of early Qusheng Sinitic loans (at least) in Vietnamese are expected to stay indifferent to the final *-**s** of the donor forms, resulting in A-tone reflexes.
- In light of this observation, two sub-layers of ESV can be identified, i.e. **ESV(I)** and **ESV(II)** (besides a few underspecified cases).
- This constraint, along with Alves' (2018) three-stage hypothesis, provides a finer-grained stratification for both nasal-final and non-nasal-final cases of A-tone reflexes of early Qusheng Sinitic loans in Vietnamese.

THANK YOU!
Cám ơn quý vị đã chú ý lắng nghe!

- Alves, Mark J. (2009). Loanwords in Vietnamese. *Loanwords in the World's Languages: A Comparative Handbook*, 617-637.
- Alves, Mark J. (2018). Early Sino-Vietnamese Lexical Data and the Relative Chronology of Tonogenesis in Chinese and Vietnamese. *Bulletin of Chinese Linguistics*, 11(1-2), 3-33.
- Alves, Mark J. (2022). Lexical evidence of the Vietic household before and after language contact with Sinitic. *Vietnamese Linguistics: A State of the Field*, 15-58.
- Baxter, William H. & Sagart, Laurent. (2014). *Old Chinese: A new reconstruction*. Oxford: Oxford University Press.
- Ferlus, M. (1998). Les systèmes de tons dans les langues viet-muong. *Diachronica*, 15(1), 1-27.
- Haudricourt, A.G. (1954). De l'origine des tons en Vietnamien. *Imprimerie Nationale*.

Selected References (2)

- Nguyen, Tài Cẩn (1995). *A Course in Vietnamese Historical Phonology*. Hà Nội: Education Publishing House.
- Nguyễn, Tài Cẩn. (2000). *Nguồn gốc và quá trình hình thành cách đọc tiếng Hán Việt* [The Origin and Development of Sino-Vietnamese readings]. Hà Nội: VNU.
- Sidwell, P. & Alves, M. (2021). The Vietic Languages: A Phylogenetic Analysis. *Journal of Language Relationship*, 166-194.
- Sidwell, P. & Jenny, M. (Eds.) (2021). *The Languages and Linguistics of Mainland Southeast Asia: A Comprehensive Guide* (Vol. 8). Walter de Gruyter.
- Vũ, Đức Nghiệu. (2011).*Lược khảo lịch sử từ vựng tiếng Việt* [An inquiry into the history of Vietnamese lexicon]. Vietnam Education Publishing House.
- Wáng, Lì (王力). (1948). HànYuèyǔ yánjiū (漢越語研究) [A Study on Sino-Vietnamese Language]. *Língnán Xuébào* (岭南學報) 9(1), 1-96.