

LAYERS OF CHINESE LOANWORDS IN PROTO- SOUTHWESTERN TAI AS EVIDENCE FOR THE DATING OF THE SPREAD OF SOUTHWESTERN TAI

Pittayawat Pittayaporn¹

Abstract

The current ethno-linguistic landscape of mainland Southeast Asia is a result of the spread of Tai speakers from southern China. This study examines Chinese loanwords in Proto-Southwestern Tai, the hypothetical ancestor of all modern Southwestern Tai varieties and proposes a dating of the spread of Southwestern Tai languages. By comparing the reconstructed Proto-Southwestern Tai forms with corresponding Chinese forms, four layers of Chinese loanwords existed in Proto-Southwestern Tai, namely Pre-Later Han, Later Han Chinese, Early Middle Chinese, and Late Middle Chinese layers. These layers indicate that Proto-Southwestern Tai was in contact with Chinese at least until the Tang era. In collaboration with non-linguistic evidence, this paper therefore proposes that Southwestern Tai languages began to spread southward sometime during the eighth and the tenth centuries CE.

Introduction

Inspiration is perhaps the most precious gift a teacher can offer her students. Back in the year 2000, one assignment in my

first course in historical linguistics was to read a paper discussing the different layers of Chinese loanwords in Thai. Fascinated, I stood in front of the classroom and reported on what I had read. At the end of the presentation, Professor Pranee Kullavanijaya plainly asked me a short but penetrating question: “How do we know which layers each of the loanwords belong to?” That was one of the defining moments in my life, one which led me onto this journey as a historical linguist. This humble paper is an attestation of Professor Pranee’s dedication to inspiring and illuminating her students.

The ethno-linguistic make-up of mainland Southeast Asia observed today is a product of successive migrations, ethno-linguistic shifts, and hybridization that started a few millennia ago. It might not be amiss to say that the most critical period of the transformation that gave rise to the current ethno-linguistic landscape was when the Tai, the Burmese, and the Vietnamese came to dominate the peninsula through an ethnic and political succession rooted in a southward spread of agricultural practices (O’Connor 1995). However, it is unclear when the transformation occurred. O’Connor places the Tai migration into Southeast Asia in the first millennium CE. Diller (2000) more specifically suggests that the southwestward migration of Tai speakers started in the 10th century. In contrast, Saraya (2002:24) believes that, due to population growth, Tai speakers began to spread south in the 12th century at the latest. Similarly, Diskul (1996) suggests that the migration probably started before Kublai Khan’s southern campaign in the 13th century. The earliest date proposed is perhaps that by Wongthes (1994:22-24; 2005:180-184) who speculates that Tai speakers started their movement around the first century BCE. Because

¹ Lecturer, Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

historical records only take us so far back in time, a picture of when the ethno-linguistic expansion occurred must come from other types of evidence.

From a linguistic point of view, the ethno-linguistic transformation resulted in a vast Tai-speaking territory stretching from Northern Vietnam to Northeast India, and from Southern China down to Malaysia. Among the branches of the Tai language family, according to Li (1960, 1977), an overwhelming majority of Tai varieties now spoken in mainland Southeast Asia all belong to one branch, namely Southwestern Tai (SWT). Better-known SWT languages include Thai², Lao, Yuan or Northern Thai, Lue, Shan, Black Tai, and White Tai. Therefore, dating the spread of SWT languages from southern China in present-day Guangxi or eastern Yunnan may facilitate a more precise identification of when the ethno-linguistic transformation took place. To this end, Chinese loanwords may be used as powerful evidence for situating the expansion of SWT languages in prehistorical chronology. This paper thus examines layers of Chinese loanwords in Proto-Southwestern Tai (PSWT), the hypothetical ancestor of all modern SWT varieties, as evidence for the spread of SWT into mainland Southeast Asia.

Loanwords as evidence for Sino-Tai contact

Loanwords are words adopted into one language from other languages. For example, English has borrowed a large number of words from Spanish, including *cargo*, *mosquito*, *plaza*, *salsa*, *ranch*,

rodeo, etc. These loanwords are clear attestations of the linguistic contact between English and Spanish, which in turn testifies to historical contact between speakers of the two languages. Not only can they provide a picture of how the contact occurred, but they can also, in many cases, reveal when it took place. For example, Dahl (1951) identifies Malagasy etyma of Sanskrit origin and proposes that its speakers left Borneo after the arrival of Indian influence in the Indonesian archipelago in the fifth century CE. Likewise, Chinese loanwords in PSWT can be used as evidence for Sino-Tai contact, particularly in speculating as to when SWT languages began to spread from their Sinospheric homeland into mainland Southeast Asia.

Although Tai and Chinese are nowadays classified in two distinct language families, their historical connection has long been a topic of great interest. While Tai is a branch of the Kra-Dai language family, also known as Tai-Kadai, Chinese belongs to the Sino-Tibetan family. Nonetheless, the fact that they share a number of typological characteristics including tonality and monosyllabicity, as well as a large set of common vocabulary, led many to view them as genetically related (Conrady 1896; Grierson 1903; Li 1976; Luo 1997; Manomaivibool 1976b; Nishida 1975; Schmidt 1926; Wulff 1934). According to this view, the shared etyma include cognates as well as a large set of loanwords from different periods. However, an alternative view that the similarities between Tai and Chinese are results of extensive and prolonged language contact (Benedict 1942, 1975, 1997; Haudricourt 1954; Ostapirat 2005; Sagart 2004, 2005; Wulff 1942) has been gaining acceptance in the past few decades. If the two language groups are

² While 'Thai' refers specifically to the Tai language of Thailand, 'Tai' is used for the entire language family.

indeed genetically unrelated, all the shared Sino-Tai vocabulary must be considered loanwords. Without denying the possibility of the Sino-Tai hypothesis, this paper, in accordance with the latter view, assumes that all etyma shared by Tai and Chinese are due to borrowing.

With respect to Chinese vocabulary in Tai languages, a number of studies (Li 1976; Nishida 1975; Wang 1966) have clearly shown that regular phonological correspondences exist among the Tai and Chinese forms. Extremely important is the work by Wulff (1934) who identifies a great number of shared Sino-Tai etyma. Most crucially, he establishes tonal correspondences between Chinese and Tai. Specifically, etyma that had *A in Proto-Tai (PT) regularly showed Even tone (平聲) in Middle Chinese. Similarly, etyma that show *B and *C in PT had Departing tone (去聲) and Rising tone (上聲) in Middle Chinese, respectively. Lastly, etyma that had *D in PT had Entering tone (入聲) in Middle Chinese. The tonal correspondences are illustrated in Table 1. Note that, on the Chinese side, Departing and Rising tones are annotated as *C and *B, respectively. In contrast, Even and Entering tones are left unmarked.

Table 1 Correspondences between Proto-Tai and Middle Chinese tones

PT		Middle Chinese	
*A	*so:ŋ ^A ‘two’ *bwi: ^A ‘fat’	雙 ʃaiwŋ 肥 buj	Even 平
*B	*ha:n ^B ‘goose’ *ʏe:ŋ ^B ‘shin’	雁 ɲain ^C 脛 ʏejŋ ^{B/C}	Departing 去
*C	*ha: ^C ‘five’ *ma: ^C ‘horse’	五 ɲo ^B 馬 mai ^B	Rising 上
*D	*pe:t ‘eight’ *ŋuək ‘crocodile’	八 pɛ:t 鱷 ŋak	Entering 入

Most relevant to the issue at hand is the investigation of Sino-Thai lexical correspondences by Manomanivibool (1975, 1976a). Not only does this excellent study propose Chinese etymological sources for many words in Thai, but also shows that the language contains at least four different layers of Chinese-related vocabulary including Pre-Middle Chinese, Middle Chinese, and Post-Middle Chinese³. It uses phonological changes as criteria for placing the etyma in one of the layers. For example, Thai t^hâ: ‘wharf’ from 渡 dù must have been borrowed before Middle Chinese because it does not reflect the rounding of *a* to *ɔ*⁴. Similarly, Thai fūn ‘dust’ from 粉 fěn is analyzed as a Late Middle Chinese loan because the initial f- indicates that it was borrowed after the process of labiodentalization had taken place in Chinese. While this study successfully uncovered the multiple layers of Chinese loanwords, it cannot be used directly to infer the date of the spread of SWT. This is because it does not aim to determine which etyma were borrowed separately by Thai or which ones already existed in Proto-Southwestern Tai (PSWT), the reconstructed ancestor of SWT.

In summary, a sizeable body of literature demonstrates clearly that Chinese loanwords abound in Tai. These loanwords are clear attestations of Sino-Tai contact, for which historical documents only provide a very blurred picture. However, the present knowledge of Chinese loanwords in Tai still cannot provide a specific date of the spread of SWT into mainland Southeast

³ Monomaivibool (1976a) calls this layer “loanwords from certain Chinese dialects.”

⁴ Manomaivibool (1976a) adopts Li’s (1971) Old Chinese reconstruction and describes this change as a change from Old Chinese *ag to Middle Chinese uo.

Asia. Therefore, this paper specifically addresses layers of Chinese etyma in PSWT.

Method and Data

The dating of Chinese loanwords in SWT involves three major steps. The first one is positing etyma of Chinese origins that are reconstructible for PSWT. Lexical data from 28 languages from Gedney's Comparative Tai Source Book (Hudak 2008) as well as the author's field notes (Pittayaporn 2005) and other published wordlists and dictionaries (e.g. Harris 1975; Hudak 1994, 1996, 2001; Kullavanijaya 2001; Luo 1999; Xing 2000) were systematically compared. Using the Comparative Method, only those that show regular correspondences established in Li (1977), Jonsson (1991), Sarawit (1973), and Pittayaporn (2009b) are considered to have existed in PSWT. The PSWT forms are based on the phonological reconstruction proposed by Pittayaporn (2009b). From among the 1159 etyma in Gedney's wordlist, 91 were identified with high level of confidence as PSWT etyma of Chinese origin. Most etymologies are based on Manomaivibool (1975) as well as earlier works, especially those by Haudricourt (1954) and Wulff (1934, 1942).

The next step involves classifying the loanwords reconstructed in the first step according to the stages of Chinese during which they were borrowed. Following Manomaivibool (1976a), phonological changes in Chinese are used as criteria in placing PSWT etyma of different origins in their respective layers. In the current study, five stages of Chinese are used as reference points: Old Chinese (OC), Late Han Chinese

(LH), Early Middle Chinese (EMC)⁵, Late Middle Chinese (LMC), and Early Mandarin (EM). However, it adopts different phonological reconstructions from Manomaivibool, who bases her analysis on Li's (1971) Old Chinese and Karlgren's (1957) Middle Chinese. Table 2 summarizes the sources of Chinese reconstructed forms used in this paper.

The last step is to infer, based on the loanword data, the date of the spread of SWT languages into mainland Southeast Asia. Crucially, the latest layer of Chinese loanwords is taken as evidence for the dating. As a proto-language is by definition the language ancestral to modern varieties, etyma that are not reconstructible to PSWT must have been innovations introduced after the proto-language diversified into daughter languages. If the diversification is a consequence of language spread, the date of the diversification and the date of the SWT expansion should be quite close. Therefore, the date of the spread of SWT into mainland Southeast Asia can be estimated on the basis of the date given to the stage of Chinese from which the latest layer of loanwords came.

⁵ For ease of comparison, EMC Rising and Departing tones, transcribed in Pulleyblank (1991) as ' and ^h, are re-transcribed here as ^B and ^C respectively.

Table 2 Reconstructions adopted for different stages of Chinese⁶

Stages	Time periods	Sources
Old Chinese (OC)	7 th -11 th centuries BCE	Baxter and Sagart (n.d.) ⁷
Later Han Chinese (LH)	1 st -2 nd centuries CE	Schuessler (2007) and Schuessler (2009)
Early Middle Chinese (EMC)	6 th -7 th centuries CE	Pulleyblank (1991)
Late Middle Chinese (LMC)	7 th -11 th centuries CE	Pulleyblank (1991)
Early Mandarin (EM)	13 th -14 th centuries CE	Pulleyblank (1991)

Table 3 Pre-LH loanwords with initial clusters

Gloss	PSWT	OC	LH	EMC	LMC	EM	
shelf	*k ^h ra: ^B		ka ^C	kai ^C / ke: ^C	kja: ^C	kjà	架 jià
rice seedling	*kla: ^C	*k ^h ra-s	ka ^C	kai ^C / ke: ^C	kja: ^C	kjà	稼 jià
conical hat	*klup	*k.rəp	lip	lip	lip	li	笠 lì
fish scale	*klet ^D	*k ^h rep-s	kəs	kəi ^C / ke: ^C	kja:j ^C	kjāj	介 jiè
lazy	*gra:n ^C	*N-kə.r ^h an?		lan ^B	lan ^B	lăn	懶 lǎn
indigo	*gra:m ^A	*N-k.r ^h am	lam	lam	lam	lám	藍 lán

⁶ Retroflex and palatal consonants in Pulleyblank (1991), Schuessler (2007), and Schuessler (2009) are re-transcribed using IPA symbols.

⁷ Note the following notations for OC reconstructed forms: () the segment may or may not have been there; [] either that sound/string of sounds, or another sound/string of sounds that gives the same result in Middle Chinese; < > infix

Pre-Later Han layer

The oldest layer of Chinese loanwords in PSWT goes back earlier than the LH period. A number of Chinese loanwords in PSWT show phonological characteristics that had been lost by the first century CE. These etyma are identifiable by their conservative onsets intermediate between LH and OC, which is the language of the earliest written documents in Chinese. The first group includes those etyma that show onset clusters. According to Schuessler (2009: 29), LH did not have initial consonant clusters. Chinese loanwords in Table 3 are reconstructed with **-l-* and **-r-* in PSWT and must have been borrowed before the LH period. Note that PSWT **kla*^C ‘rice seedling’ has tone *C rather than the expected tone *B.

The second group consists of Chinese loanwords that begin with voiceless sonorants in PSWT. According to Baxter (1992: 188-220), OC voiceless sonorants had turned into other sounds such as *t^h-* and *l-* by the time of LH. These etyma are given in Table 4. Note that ‘gill’, ‘central root’, and ‘soot’ are not found in the list of reconstructed OC etyma but show initial voiceless sonorants, which suggests pre-LH origin. Also included in this set are PSWT etyma with initial **h-* corresponding to Chinese **ŋ-*. These must have been borrowed relatively early because they show voiceless onset pointing to the archaic initial voiceless sonorants.

The third group comprises a few etyma with initial uvular **q-* in PSWT. Baxter and Sagart (2007) propose that OC had uvular stop **q-* and **g-* but neither survived into LH as reconstructed by Schuessler (2007, 2009). Chinese loanwords in PSWT beginning with uvular

stops in Table 5 must have been borrowed quite early on. However, it is puzzling that two out of the three etyma starting with **q-* in PSWT are actually reconstructed with **k-* in OC.

The fourth group consists of a number of etyma with very conservative initial consonants, as shown in Table 6. The PSWT forms **ruə*^B ‘to leak’, **rak*^D ‘armpit’ and **rom*^B all show initial **r-*, reflecting medial **-r-* in OC. These must have been borrowed before LH because OC **-r-* had changed to *l-* by LH time (Baxter 1992; Schuessler 2009)⁸. Note that the reconstruction of medial *-r-* in the OC forms of ‘armpit’ and ‘shadow’ are not certain, but the liquid is attested in Vietnamese *râm* ‘shade’, most likely an ancient Chinese loan⁹. In addition, PSWT **liəŋ*^C ‘to nourish’ suggests OC **g(r)əŋ*?, which Baxter (1992) reconstructs as **(l)jaŋ*?. The initial **l-* in PSWT indicates that the etymon was borrowed into Tai before the LH period. Also included in this set are **tuə*^A and **xwa*^A. The Chinese sources of these two etyma had changed their onsets to *d-* and *w-* by the LH period. Note that the PSWT form of ‘right’ has tone *A rather than the expected *C tone.

⁸ Manomaivibool (1976a) claims that the etyma were borrowed from a non-standard dialect during Han times.

⁹ Chinese has two similar etyma differing in tone: 蔭 *yìn* ‘shade’ from OC **q(r)əm-s* and 陰 *yīn* ‘dark’ from **OC *q(r)um*. Vietnamese *râm* is more likely from the former but with an unexpected tone.

Table 4 Pre-LH loanwords with initial voiceless sonorants

Gloss	PSWT	OC	LH	EMC	LMC	EM	
fog	* ^h mɔ:k ^D	*kə.m(r)[o]k-s	muo ^C	muə ^C	vjyǎ ^C vuǎ ^C	vù	霧 wù
six	* ^h rok	*k.ruk	liuk	luwk	liwk	liw	六 liù
thread	* ^h maj ^A	*məj	hui	xuj	xyj	xuj	徽 huī
gill	* ^h ɲwək ^D		ɲak				鰓 è
central root	* ^h ɲaw ^C			ɲəw ^B	ɲəw ^B	ǎw	藕 ǒu
soot	* ^h mi: ^C		mə	məj	muaj	múj	煤 méi
shaman	* ^h mɔ: ^A	*C.m(r)[o]	mua	muə	vjyǎ/ vuǎ	vú	巫 wū
pus	* ^h nɔ:ɲ ^A	*C.n ^ʰ uɲ	nouɲ	nawɲ	nəwɲ	núnɲ	膿 nóng
iron	* ^h lek ^D	* ^ʰ ik	t ^h et	t ^h et	t ^h iat	t ^h jě	鐵 tiě
crossbow	* ^h na: ^C	*C.n ^ʰ aʔ	na ^B	nɔ ^B	nuə ^B	nǔ	弩 nǔ
five	*ha: ^C	*C.ɲ ^ʰ aʔ	ɲa ^B	ɲɔ ^B	ɲuə ^B	ǔ	五 wǔ
goose	*ha:n ^B	*C.[ɲ] ^ʰ rar-s	ɲan ^C	ɲain ^C ɲe:n ^C	ɲja:n ^C	jàn	雁 yàn

Table 5 Pre-LH with uvular onsets

Gloss	PSWT	OC	LH	EMC	LMC	EM	
needle	*qem ^A	*t.[k]əm	kim	tɛim	tɕim	tɕim	鍼 zhēn
arm	*qe:n ^A	*[k] ^ʰ e[n]	ken	kɛn	kjian	kjen	肩 jiān
soul	*qwan ^A	*[m].q ^w ə[n]	ɣuən	ɣwən	x ^ʰ un	xún	魂 hún
narrow	*ge:p ^D		gep	ɣəip/ ɣe:p	x ^ʰ ja:p	xjá	狹 xiá

Table 6 Pre-LH loanwords with conservative onsets

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to leak	*ruə ^B	*[Nə-r] ^ʰ ok-s	lo ^C	ləw ^C	ləw ^C	lòw	漏 lòu
armpit	*rak ^D	*m-q(r)ak	jak	jiajk	jiajk	jì	亦 yì
shadow	*rom ^B	*q(r)[ə]m-s	ʔim ^C	ʔim ^C	ʔim ^C	jìm	蔭 yìn
to nourish	*liəɲ ^C	*[ɕ](r)əɲʔ	jaɲ ^B	jiaɲ ^B	jiaɲ ^B	jǎɲ	養 yǎng
to exchange	*lɛ:k ^D	*lek	jek	jiajk	jiaik	jì	易 yì
classified for animals	*tuə ^A	*m-t ^ʰ o	do	dəw	t ^ʰ əw	t ^ʰ ów	頭 tóu
right	*xwa: ^A	*m-q ^w əʔ-s	wuə ^{B/C}	wuw ^B	iw ^B	jìw	右 yòu

One expected but very important generalization about these pre-LH loanwords is that they are all reconstructible to Proto-Tai (Li 1977; Pittayaporn 2009a). This means that they were borrowed during the PT period and then passed down to PSWT. Moreover, the reconstructed PSWT forms seem to be more evolved than OC but still recognizably older than LH. The only exceptions are ‘classifier for animal’ and ‘right’ for which modern Tai languages disagree on what the PT onsets might have been (Gedney 1989a; Pittayaporn 2009a; Thurgood 2002). This observation suggests that they were borrowed sometime between OC and LH. If OC was spoken in the first half of the first century BCE and LH started around the first century CE (Schuessler 2009: 29), then these pre-LH loans were possibly borrowed by Tai in the later half of the first millennium BCE.

Note that Li (1976) and Nishida (1975) consider a number of the etyma analyzed here as Pre-LH loans, e.g. ‘fog’, ‘to leak’, ‘six’, ‘lung’ etc., to be Sino-Tai cognates, evidence for a genetic relationship between Tai languages and Chinese. However, comparison between Tai and Chinese alone is not sufficient to ascertain the status of these items. Data from other branches of Kra-Dai must also be systematically compared to know whether they are reconstructible to a deeper level than PT.

Later Han Chinese layer

The second oldest layer corresponds to the LH period of Chinese, which was spoken around the first and second centuries. The reconstruction of LH is based on data from modern dialects including Min as well as

transcription of Sanskrit Buddhist texts from the Han period. Although LH is about 500 years earlier and clearly more conservative than EMC (Schuessler 2009: 29), loanwords borrowed during these two periods are extremely difficult to tell apart. The most reliable clues for distinguishing LH from EMC loanwords are in the rimes. Many Chinese loanwords in PSWT, given in Table 7, show conservative rimes that had been modified before the time of EMC.

Table 7 LH loanwords with conservative rimes

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to smear	*da: ^A	*l ^ʰ a	da	də	t ^h uə	t ^h ú	塗 tú
wharf	*da: ^B		da ^C	də ^C	t ^h uə ^C	tù	渡 dù
ivory	*ŋa: ^A	*m-ŋ ^ʰ r>a	ŋa	ŋai ŋe:	ŋja:	ja:	牙 yā
sand	*za:j ^A	*[s] ^ʰ raj	ʂai	ʂai ʂe:	ʂa:	ʂa	沙 shā
left	*za:j ^C	*ts ^ʰ a[j]?	tsai ^B	tsa ^B	tsa ^B	tsǎ	左 zuǒ
lady	*na:ŋ ^A			ŋian	ŋian	ŋján	娘 niáng
to peel	*pə:k ^D	*p ^ʰ rok	pək	paiwk/ pə:wk	pa:wk	pǎw	剥 bō
two	*sə:ŋ ^A	*[s] ^ʰ roŋ	ʂoŋ	ʂaiwŋ/ sə:wŋ	swa:wŋ	ʂwan	雙 shuang
mustard green	*ka:t ^D	*k ^ʰ r[e][t]-s	kes	kaij ^C / ke:j ^C	kja:j ^C	kjàj	芥 jiè
lung	*pə:t ^D	*p ^h o[t]-s	p ^h uas	p ^h uaj ^C	fjyàj fji	fi	肺 fèi
name	*ju: ^B	*mə-dzə(?) -s	dziə ^C	dzi ^C dzi ^C	ts ^h z ^C	tsz̥	字 zì

Table 8 LH loanwords with conservative onsets

Gloss	PSWT	OC	LH	EMC	LMC	EM	
price	ga: ^B	*C.q ^ʰ r>aʔ-s	ka ^C	kai ^C ke: ^C	kja: ^C	kjà	價 jià
salty	gem ^A	*Cə.[g] ^ʰ r[o]m	gem	ɣəim/ ɣe:m	x ^h ja:m	xjám	鹹 xián
to pinch	gi:p ^D	*m-k ^ʰ ep	gep	ɣep	x ^h jiap	xjé	挾 xié
seven	cet ^D	*[ts ^h]i[t]	ts ^h it	ts ^h it	ts ^h it	ts ^h ĩ	七 qī
duck	pet ^D		p ^h it				鴨 pī

The PSWT etyma **da.^A* ‘to smear’ and **da.^B* ‘landing’ are clearly LH loans as they have **a:* pointing to LH rime *a* rather than EMC *ɔ*. Similarly, the rimes of **ŋa.^A* ‘ivory’, *za:j^A* ‘sand’, and **za:j^C* ‘left’ resemble LH *a* and *ai* more than EMC *ai/ɛ:*. Note that the initial consonants in ‘sand’ and ‘left’ in PSWT were unexpectedly voiced. For **na:ŋ^A* ‘lady’, its Chinese correspondence was not attested before EMC but its vowel **a:* suggests a pre-EMC source¹⁰. In addition, the vowel of PSWT **pɔ:k^D* ‘to peel’ and **sɔ:ŋ^A* ‘two’ points to LH *ɔ* rather than EMC *aiw*. Most revealing is PSWT **ka:t^D* ‘mustard green’ whose back vowel *a:* resembles EMC *ai* but whose final coronal stop suggests LH final *-s*. Similarly, PSWT **pɔ:t^D* ‘lung’ also shows a final consonant **-t* which points to *-s* and must have been an LH loan. Although the PSWT form of this etymon is similar to both the OC and LH forms, the PT reconstruction for this etymon is **pwrɔt^D*, which favors viewing it as a LH loan. The most likely scenario is that it was borrowed before *-s* developed into the Departing tone but after the vowel had lost its front quality. Furthermore, the monophthongal rime in PSWT **ju.^B* ‘name’ from PT **jɜ.^B* points to OC **mǝ-dzǝ(ʔ)-s*.

In addition to rimes, a few etyma given in Table 8 can also be identified as LH loans from their onset. More specifically, OC **g-* had become *ɣ-* by the time of EMC. A few loanwords still retain the velar stop, which reveals their LH origin. Note that the PSWT **ga.^C* ‘to trade’ unexpectedly

shows the voiced **g-* instead of the expected voiceless **k-* as in LH. Also included are **cet^D* ‘seven’ and **pet^D* ‘duck’ whose onsets and rimes are quite stable from OC, LH, EMC, and even LMC. However, Pittayaporn (2009a: 100-101) argues that these two etyma and the word for ‘lung’ discussed above were borrowed extremely early as they predated the emergence of contrastive aspiration in Thai. If this is true, ‘seven’ and ‘duck’ might have been borrowed around the same time as ‘lung’ in the LH period.

Like their pre-LH counterparts, LH loanwords are also reconstructible at the PT level, indicating that they were borrowed in the PT period. The only possible exceptions are ‘salty’, ‘sand’, and ‘left’, which show dubious vowel reflexes in modern languages (Pittayaporn 2009a). Because the first century CE is thought to be the onset of the LH period (Schuessler 2009: 29), the borrowing must have taken place sometime in the first half of the first millennium CE. By the PSWT period, they were fully integrated as part of the native lexicon. In addition to the etyma discussed in this section, a number of others might have also been borrowed in this period. However, they are unfortunately indistinguishable from EMC loanwords.

Early Middle Chinese layer

A number of Chinese loanwords in PSWT belong to the third layer, which corresponds to the EMC period around the sixth to seventh centuries. The reconstruction of EMC is based mainly on the rhyming dictionary *Qièyùn* (切韻) compiled in 601 CE during the Sui dynasty (Baxter 1992:35-41; Pulleyblank 1970; 1984:2-3; 1991:1-3). Because EMC is very similar to LH, except for a few innovations, it is difficult to distinguish

¹⁰ Vietnamese has borrowed this Chinese etymon as *nàng* ‘princess, lady’, pointing to the same sources as PSWT **na:ŋ^A*. In contrast, the Sino-Vietnamese reading for the character 娘 is *nường*, showing closer resemblance to the EMC and LMC forms.

loans from the two periods. However, a few etyma, given in Table 9, display innovative initials or rimes characteristic of EMC.

The PSWT forms **ʃɛːŋ^B* ‘shin’ and **ʃɔː^A* ‘throat’ reflect the change from **g-* to EMC **ʃ-* that occurred sometime after LH, indicating that they were borrowed from EMC. For **kɛːŋ^A* ‘soup’, **kɛːw^A* ‘Vietnamese’, and **kʰɛːk^D* ‘guest’, the vowel **ɛː* clearly goes back to EMC *ɛː* or *ɛːj* rather than LH *a¹¹*. Similarly, the monophthongal **aː* in **pʰaː^B* ‘to split’ indicates that it was borrowed after LH *ai* changed to EMC *a*.

In addition to etyma that are clearly EMC loans, there are items that can be placed in either layer with equal plausibility. However, these loanwords are clearly older than LMC as they still preserve the EMC voiced stops that had been lost by the LMC period. Although the etyma in this group, provided in Table 10, are tentatively grouped with EMC loans, they could have been borrowed from LH as well. Note that a few items in this group also show conservative rimes that predate LMC. For example, PSWT **bɛː^A* ‘raft’ points to LH *ɛ* or EMC *ɛːj* rather than LMC *aːj*.

Moreover, there are a few etyma, provided in Table 11, that could be from either LH or EMC, but do not go back to voiced obstruents in Chinese. It is difficult to say whether they were borrowed from LH or EMC, but their rimes indicate clearly that they must have been borrowed before LMC. The front vowel **ɛː* in PSWT **kɛː^C* ‘to untie’, **ʔɛːk^D* ‘yoke’, **ʔɛːn^B* ‘swallow’, **lɛːw^C* ‘to finish’, and **pɛːt^D* ‘eight’ points to LH or EMC *ɛːj*, *ɛ*, and *ɛː* rather than

their LMC counterparts. Lastly, the palatal nasal in PSWT **ɲiː^B* ‘two’ and **ɲɔːm^C* ‘to dye’ points to the palatal nasal *ɲ-* in EMC rather than the liquid *r-* in LMC. Note, though, that the vowel of ‘to dye’ resembles more closely *nhuộm*, the corresponding Chinese loan in Vietnamese.

Unlike the LH layer, a sizeable number of EMC etyma are not old enough to have been part of PT. Although all of them are commonly found among modern Tai languages, a few are possibly not reconstructible back to the PT level, e.g. ‘to split’, ‘level’, ‘copper’, ‘soup’ and ‘ox’¹². This observation suggests that PT began to diverge sometime during the EMC period so that only etyma borrowed before the diversification became part of PT, while those that were not incorporated early enough can only be reconstructed to intermediate daughter languages.

¹¹ These forms very closely resemble Vietnamese *canh* ‘soup’ and *khách* ‘guest’.

¹² The status of **ɲuə^A* ‘ox’ is unclear because it is found only in SWT varieties (Gedney 1989b). However, it is possible that it is a PT etymon that was lost elsewhere outside of SWT.

Table 9 EMC loanwords with innovative initials or rimes

Gloss	PSWT	OC	LH	EMC	LMC	EM	
shin	*ʏɛ:ŋ ^B	*m-k ^h ɛŋ-s	geŋ ^C	ʏɛŋ ^{B/C}	x ^h jiəŋ ^B	xìŋ	脛 xìng
throat	*ʏɔ: ^A	*[g] ^ʕ (r)o	go	ʏəw	x ^h əw	xəw	喉 hóu
soup	*kɛ:ŋ ^A	*k ^ʕ raŋ	kaŋ	kəiŋ/ kɛ:ŋ	kja:ŋ	kiŋ	羹 gēng
Vietnamese	*kɛ:w ^A	*[k] ^ʕ raw	kau	kaiw/ kɛ:w	kja:w	kjaw	交 jiāo
guest	*k ^h ɛ:k ^D	*k ^h ɛrak	k ^h ak	k ^h aijk/ k ^h ɛ:jk	k ^h ja:jk	k ^h ǎj/ k ^h ǎj	客 kè
to split	*p ^h a: ^B	*p ^h aj-s	p ^h ai ^C	p ^h a ^C	p ^h ua ^C	p ^h ò	破 pò

Table 10 Loanwords with initial voiced obstruents either from LH or EMC

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to defeat	*ba:j ^B	*N-p ^ʕ ra[t]-s	bas	baij ^C / bɛ:j ^C	p ^h a:j ^C	pàj	敗 bài
raft	*bɛ: ^A		bɛ	baij/ bɛ:j	p ^h a:j	p ^h áj	簰 pài
level	*biəŋ ^A	*breŋ	bieŋ	biaŋ	p ^h iaŋ	p ^h ín	平 píng
fat	*bi: ^A	*[b][ə]r	bui	buɣ	f ^h jiy/ f ^h ji	fǐ	肥 féi
ground	*di: ^B	*[l] ^ʕ ej-s	di ^C	di ^C	t ^h i ^C	tì	地 dì
path	*da:ŋ ^A	*[N-]r ^ʕ aŋ	daŋ	daŋ	t ^h aŋ	t ^h án	唐 táng
lump, piece	*dɔ:n ^B	*N-t ^ʕ o[n]ʔ-s	duan ^C	dwan ^C	t ^h uan ^C	tòn	斷 duàn
copper	*dɔ:ŋ ^A		doŋ	dəwŋ	t ^h əwŋ	t ^h úŋ	銅 tóng
plier	*gi:m ^A	*C.[g]r ^ʕ em	giam	giam	k ^h iam	k ^h jém	鉗 qián
artisan	*ja:ŋ ^B	*s.[b]aŋ-s	dziaŋ ^C	dziaŋ ^C	ts ^h iaŋ ^C	tsjàŋ	匠 jiàng
elephant	*ja:ŋ ^C	*s-[d]aŋ?	ziaŋ ^B	ziaŋ ^B	s ^h iaŋ ^B	sjàŋ	象 xiàng

Table 11 Loanwords with conservative rimes either from LH or EMC

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to untie	*kɛ: ^C	*kʰreʔ	kɛ ^B	kai:j ^B / kɛ:j ^B	kja:j ^B	kjǎj	解 jiě
yoke	*ʔɛ:k ^D	*qʰ<ɾ>[i]k	ʔɛk	ʔəijk/ ʔɛ:jk	ʔja:jk	jàj	軛 è
swallow	*ʔɛ:n ^B	*ʔe[n]-s	ʔen ^C	ʔen ^C	ʔjian ^C	jèn	燕 yàn
to finish	*lɛ:w ^C			lɛw ^B	liaw ^B	ljěw	了 liǎo
eight	*pɛ:t ^D	*pʰret	pɛt	pəit/ pɛ:t	pa:t	pǎ	八 bā
ox	*ɲuə ^A	*[ŋ]wə	ɲu < ɲwuə	ɲuw	ɲiw	jíw	牛 niú
two	*ni: ^B	*ni[j]-s	nis/ ni ^C	ni ^C	ri ^C	rɿ	二 èr
to dye	*ɲɔ:m ^C	*C.n[a]mʔ	ɲam ^{B/C}	ɲiam ^B	riam ^C	rěm	染 rǎn

Late Middle Chinese layer

The last layer of Chinese loanwords in SWT consists of forms from the LMC period. LMC is thought to have been spoken in the latter half of the first millennium from the end of the seventh century. The reconstruction of LMC phonology is based primarily on the *Yunjing* (韻鏡), the earliest attested rhyme table, dating from 1161 CE in the Northern Song dynasty. It is, however, believed to reflect the standard language of the late Tang period (Baxter 1992: 41-43; Pulleyblank 1970; 1984: 3-4; 1991: 3)

(Pulleyblank 1970; 1984: 67-68), EMC voiced obstruents had become partially devoiced consonants with accompanying murmur by the time of the *Yunjing*. LMC loanwords beginning with these devoiced stops were thus borrowed into Tai as voiceless aspirated rather than voiced stops. Note that the Chinese character 騎 has two readings *qí* ‘to ride’ from LMC *kʰi* and *jì* ‘horsemen’ from LMC *kʰi^C*. While the meaning of PSWT **kʰwi:^B* agree with the former, its form is more similar to the latter.

The first set of etyma, given in Table 12, comprises those that show voiceless aspirated stops corresponding to EMC voiced stops. According to Pulleyblank

Table 12 Loanwords with PSWT aspirated stops corresponding to LMC breathy voiced stops

Gloss	PSWT	OC	LH	EMC	LMC	EM	
eggplant	*k ^h uə ^A			gia	k ^h ia	k ^h jé	茄 qié
to ride	*k ^h wi ^B	*[C.g](r)aj	gi <i>ai</i>	gi/ giǎ	k ^h i	k ^h í	騎 qí
young male animal	*t ^h ɿk ^D	*[d]ʰək	dək	dək	t ^h əǎk	táj	特 tè
bean	*t ^h uə ^B	*N.t ^h o-s	do ^C	dəw ^C	t ^h əw ^C	təw	荳 dòu
bowl	*t ^h uəj ^C	*[d]ʰok-s	do ^C / dôh				豆 dòu

Table 13 Loanwords with PSWT voiceless onsets corresponding to LMC breathy voiced fricatives

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to close	*hap ^D		gəp	ɣap	x ^h ap	xó	闔 hé
turbid	*k ^h un ^B		guən ^B	ɣwən ^B	x ^h un ^C	xùn	混 huǐn
horizontal	*xwa:ŋ ^A	*C.g ^w raŋ	ɣuaŋ	ɣwaijŋ/ ɣwe:ŋ	x ^h wa:jŋ	xwəŋ/ xúŋ	橫 héng
ten	*sip ^D	*t.[g]əp	dzip < gip	dzip	ʂ ^h ip	ʂí	十 shí
cooked, ripe	*suk ^D	*[d]uk	džuk	dzuwk	ʂ ^h iwk	ʂíw/ ʂý	熟 shú
enemy	*sɿk ^D	*k.dzʰək	dzək	dzək	ts ^h əǎk	tsáj	賊 zéi

Table 14 LMC loanwords showing traces of labiodentalization

Gloss	PSWT	OC	LH	EMC	LMC	EM	
person	*p ^h u: ^C	*p(r)aʔ	pua ^B	puǎ ^B	fjyǎ ^B / fuǎ ^B	fǔ	甫 fǔ
husband	*p ^h uə ^A	*p(r)a	pua	puǎ	fjyǎ/ fuǎ	fu	夫 fū
bee, beewax	*p ^h ɿŋ ^C	*p ^h (r)oŋ	p ^h uoŋ	p ^h uawŋ	fjyawŋ/ fəwŋ	fuŋ	蜂 fēng
to float	*vu: ^A	*m.b(r)u	bu	buw	f ^h jy ^w / f ^h uw	fú	浮 fú

Similarly, a number of Chinese etyma in PSWT, given in Table 13, show voiceless onsets corresponding to EMC voiced fricatives. This correspondence also reflects the LMC devoicing that affected all EMC voiced obstruents. Curiously, the LMC initial x^h is reflected differently in PSWT $*hap^D$ ‘to close’, $*k^hun^B$ ‘turbid’, and $*xwa:\eta^A$ ‘horizontal’. However, this variability in adaptation is understandable as PSWT $*h$ -, $*k^h$ - and $*x$ - are all equally similar to x^h . Also note that the PSWT form for $*k^hun^B$ ‘turbid’ has *B tone, which also points clearly to LMC.

The next group of Chinese loanwords in PSWT, given in Table 14, consists of a few etyma with onsets that suggest LMC labial fricatives, which developed from EMC labial stops followed by palatal and labial glides (Baxter 1992; Karlgren 1957; Li 1971; Pulleyblank 1984). The PSWT forms $*p^hu:C$ ‘person’ and $*p^hu\partial^A$ ‘husband’ had initial p^h - which suggests that an intermediate stage between EMC p - and LMC f -. As for PSWT $*p^h\eta^C$ ‘bee, beeswax’, the aspirated onset $*p^h$ - itself could have come directly from EMC p^h - but the rime $*\eta$ points to LMC $\partial w\eta$ rather than EMC $uaw\eta$. More interesting is the voiced fricative $*v$ - in PSWT $*vu^A$ ‘to float’. The form must have been borrowed after the labiodentalization but before the obstruent devoicing.

Also included in this layer are a few etyma, provided in Table 15, that could have been from either EMC or LMC with equal probability. Because neither their onsets nor their rimes are useful in precisely placing them in one of the two layers, they are tentatively put together with the clear LMC etyma.

Unlike loanwords in earlier layers, an overwhelming majority of PSWT forms of LMC origin cannot be reconstructed to PT. Those that are possibly reconstructible are ‘insect’, ‘first (month)’, and ‘to transcend’, but they are all etyma that could be placed in either the EMC, or the LMC layers. This observation suggests that PSWT was contemporaneous to LMC.

Other common SWT etyma

In addition to etyma belonging to the four layers, there are a number of forms, given in Table 16, whose periods of borrowing are unclear. Their rimes and consonants resemble forms from multiple stages of Chinese, including LMC and even EM. The most plausible hypothesis is that these words were borrowed at different periods. For example, the low numbers $*sa:m^A$ ‘three’ and $*si^B$ ‘four’ are certainly reconstructible all the way to PT, and might thus have been borrowed in the Pre-LH or LH period. In contrast, the high number $*kaw^C$ ‘nine’ shows semi-regular sound correspondences among its modern reflexes, and might thus have been borrowed in the LMC or EM period. It is also interesting to note that the low numerals were borrowed in earlier periods than higher numerals. This is contradictory to the cross-linguistic observation that lower numerals are more resistant to borrowing (Matras 2007).

Table 15 Ambiguous loanwords either from EMC or LMC

Gloss	PSWT	OC	LH	EMC	LMC	EM	
first (month)	*ciəŋ ^A	*C.teŋ	tseŋ	teiaŋ	tɕiaŋ	tɕiŋ	正 zhēng
to dig	*k ^h ut ^D	*k ^h ut	k ^h uət	k ^h wət	k ^h ut	k ^h ũ	掘 kū
to transcend	*kwa. ^B	*k ^w aj-s	kuai ^C	kwa ^C	kua ^C	kwò	過 guò
insect	*mlɛ:ŋ ^A		meŋ-leŋ	meŋ-leŋ	mjaŋ-ljaŋ	mín-lín	螟蛉 míng-líng
sound	*siəŋ ^A	*[l]eŋ	ɕeŋ	ɕiaŋ	ɕiaŋ	ɕiŋ	聲 shēng
to remove	*t ^h ɔ:t ^D	*l ^ʰ ot	t ^h uat/ duat	t ^h wat	t ^h uat	thǒ	脫 tuō

Table 16 Loanwords that cannot be dated with certainty

Gloss	PSWT	OC	LH	EMC	LMC	EM	
to mend	*fu:ŋ ^A	*(mə-)[b](r)on	buon	buawŋ	f ^h jyawn/ f ^h əwŋ	fún	縫 féng
dust	*fun ^B	*mə.pən?	pun ^B	pun ^B	fjyn ^B / fun ^B	fün	粉 fěn
stem	*ka:n ^C		kan ^C	kan ^C	kan ^C	kàn	幹 gàn
to open	*k ^h aj ^A	*[k] ^h əj	k ^h əi	k ^h əj	k ^h aj	k ^h aj	開 kāi
saw	*ku: ^C	*k(r)a-s	kiə ^C	kiə ^C	kiə ^C kyə ^C	kỳ	鋸 jù
wide	*kwa:ŋ ^C	*k ^w əŋ?	kuəŋ ^B	kwaŋ ^B	kuəŋ ^B	kwǎŋ	廣 guǎng
melon	*kwa: ^A	*k ^w əra	kua	kwai/ kwɛ:	kwa:	kwa	瓜 guā
to embroider	*sɛ:w ^B	*[s]iw(k)-s	siu ^C	suw ^C	siw ^C	siw	繡 xiù
charcoal	*t ^h a:n ^B	*[t ^h] ^ʰ a[n]-s	t ^h an ^C	t ^h an ^C	t ^h an ^C	t ^h àn	炭 tàn
waist	*ʔɛ:w ^A	*ʔew	ʔiaw	ʔjiaw	ʔjiaw	jew	腰 yāo
nine	*kau ^C	*[k]u?	ku ^B	kuw ^B	kiw ^B	kĩw	九 jiǔ
three	*sa:m ^A	*sr[u]m	sam	sam	sam	sam	三 sān
four	*si: ^B	*s.li[j]-s	si ^C	si ^C	sɿ ^C	sɿ	四 sì
to deliver	*son ^B	*[s] ^ʰ on-s	son ^C	səwŋ ^C	səwŋ ^C	sùn	送 sòng
book	*su: ^A	*s-ta	ɛa	ɛǎ	ɕiǎ/ syǎ	ɕy	書 shū
stool	*taŋ ^B			təŋ ^C	təǎŋ ^C	tàn	凳 dèng
saddle	*ʔa:n ^A	*[ʔ] ^ʰ an	ʔan	ʔan	ʔan	an	鞍 ān
basin	*ʔa:ŋ ^B		ʔaŋ ^C	ʔaŋ ^C	ʔaŋ ^C	àn	盥 àn

Table 17 Common etyma of Chinese origin not reconstructible to PSWT

Gloss	BT	OC	LH	EMC	LMC	EM	
silver	ɲɿn ^{A2}	*ɲrə[n]	ɲin	ɲin	ɲin	jín	銀 yín
ink	muuk ^{DS2}	*C.mək	mək	mək	muək	mùj	墨 mò
work	viəʔ ^{DL2}	*C ^w ek	wek	jwajk	jyajk	jì	役 yì
to fry	cən ^{A1}		tsian	tsian	tsian	tsjen	煎 jiān

Last but not least, a few etyma of Chinese origin, given in Table 17, are commonly found in modern SWT languages but they cannot be reconstructed for PSWT. This indicates that they were borrowed into daughter languages after the diversification of PT. All of these forms appear to have Mandarin-like features. Since these post-PSWT loans do not have reconstructed PSWT forms, Black Tai (BT) forms are given as representatives. Interestingly, they resemble very closely their corresponding forms in Vietnamese, i.e. *ngân* ‘silver’, *mực* ‘ink’, *việc* ‘work’, and *chiên* ‘to fry’. It is very likely that they are in fact not direct loans from Chinese but were borrowed through Vietnamese.

Dating SWT migration

From the chronological classification of Chinese loanwords in PSWT, a clearer picture of the contact history between Chinese and Tai has emerged. The multiple layers of Chinese loanwords in PSWT clearly demonstrate that Tai was in close contact with Chinese for at least one thousand years before the spread of its SWT branch. As the earliest layer displays archaic features intermediate between OC and LH, the Sino-Tai contact must have started at least in the Western Han time toward the end of the first millennium BCE. This contact persisted through the Later Han, Sui, and Tang periods as testified by the LH, EMC, and LMC layers

respectively. However, the fact that the latest layer of Chinese etyma in PSWT corresponds to LMC suggests that the dating of LMC itself is key to dating the spread of SWT.

According to Pulleyblank (1970; 1984: 3-4; 1991: 3), LMC was the standard language of the Tang Empire. Its reconstruction is based mainly on the *Yùnjìng* compiled sometime between 742 and 756 CE. Although this rhyme dictionary is not attested in its original form, it served as the basis for the sound glosses in *Yíqiè Jīngyīngyì* (一切經音義) compiled by *Huì Lín* (慧琳) who died in 810 CE. However, the earliest evidence for LMC phonological features is from sound glosses in *Yán Shīgǔ*’s (顏師古) commentary to the *Hànshū* (漢書) as well as Tang Buddhist transcriptions from the latter half of the seventh century. Therefore, the LMC layer indicates that PSWT, the ancestor of modern SWT languages, was still in contact with Chinese in the late seventh century. If one assumes that the diversification of PSWT into its daughter languages is due to the movement of its speakers out of the homeland, the spread of SWT into mainland Southeast Asia must have started after the LMC period. This suggests the eighth century as the upper bound of the period of SWT expansion.

To estimate the lower bound is harder and thus needs collaboration from non-linguistic evidence. First considering the linguistic evidence, it is significant that none of the Chinese loanwords in PSWT can be analyzed with confidence as coming from EM. This means that PSWT had developed into daughter languages before the time EM supplanted LMC as the lingua franca of the Chinese empire. Because the first attestation of EM is the Yuan rhyme dictionary *Méngǔ Zìyùn* (蒙古字韻) compiled in 1308 CE (Pulleyblank 1984: 3-4; 1991: 3-4), PSWT must have developed into daughter languages by the end of the thirteenth century. However, the history of SWT-speaking kingdoms indicates that this date is probably too late as the early Tai-speaking kingdoms had already emerged by the time (Gosling 1998: 11-36; Hall 1968: 170-172; Sai Aung Tun 2009: 14-20; Wyatt 1984: 5-6, 33-49). Moreover, the word *syām*, an ethnonym for various SWT-speaking groups, is first attested in Cham and Pagan inscriptions from the eleventh and twelfth centuries. (Ferus 2006; Luce 1958, 1959; Phumisak 1976). Therefore, the tenth century is the most plausible lower bound for the period of SWT expansion.

In summary, the LMC layer of Chinese loanwords in PSWT indicates that SWT began to spread into mainland Southeast Asia after the seventh century. On the other hand, non-linguistic evidence suggests that the linguistic expansion must have started before the eleventh century. Therefore, this paper proposes that SWT speakers began to spread southwestward from Guangxi sometime between the eighth and the tenth centuries.

Conclusion

The current ethno-linguistic landscape of mainland Southeast Asia is due in large part to the spread of SWT speakers from southern China. A lack of historical record makes the dating of the SWT expansion a difficult task. As this study has shown, linguistic evidence can help shed light on this important prehistorical event. More specifically, Chinese loanwords in PSWT, the hypothetical ancestor of all SWT languages, are important evidence for the dating of the spread of SWT languages. Altogether four layers of Chinese loanwords existed in PSWT, namely Pre-Later Han Chinese (Pre-LH), Later Han Chinese (LH), Early Middle Chinese (EMC), and Late Middle Chinese (LMC). Crucially, the LMC layer indicates that the SWT spread must have started after the EMC period. In collaboration with non-linguistic evidence, the loanword evidence suggests that SWT began to spread southwestward in the last quarter of the first millennium CE.

Acknowledgment

I would like to thank Laurent Sagart, John Whitman, John Phan, Jakrabhop Iamdanush, the audience at conference on “Rice and Language across Asia: Crops, Movement, and Social Change”, and two anonymous reviewers for their contributions at various stages of the paper. All shortcomings this article may have are naturally mine. This research is supported by the Ratchadaphiseksomphot Endowment Fund of Chulalongkorn University (RES560530179-HS).

References

- Baxter, W. H. 1992. *A handbook of Old Chinese phonology*. Berlin; New York: Mouton de Gruyter.
- Baxter, W. H., & Sagart, L. 2007. *Reconstructing Old Chinese uvulars in the Baxter-Sagart system (ver. 0.97)*. Paper presented at the 40th International Conference on Sino-Tibetan Languages and Linguistics, Harbin, China.
- . n.d. *Baxter-Sagart Old Chinese reconstruction (Version 1.00)*. Online at <http://crlao.ehess.fr/document.php?id=1217>.
- Benedict, P. K. 1942. Thai, Kadai and Indonesian: a new alignment in Southeastern Asia. *American Anthropologist*, 44: 576-601.
- . 1975. *Austro-Thai language and culture, with a glossary of roots*. New Haven: HRAF Press.
- . 1997. The Tai languages: key to Austro-Tai. In *Southeast Asian Linguistic Studies in Honor of Vichin Panupong*, edited by A. S. Abamson. Bangkok: Chulalongkorn University.
- Conrady, A. 1896. *Eine indochinesische Causative-Denominativ-Bildung und ihr Zusammenhang mit den Tonaccenten*. Leipzig: Otto Harrassowitz.
- Dahl, O. C. 1951. *Malgache et Maanjan: une comparaison linguistique*. Oslo: Egede Instituttet.
- Diller, A. 2000. The Tai language family and the Comparative Method. In S. Burusphat (Ed.), *Proceedings : the International Conference on Tai Studies, July 29-31, 1998*, pp. 1-32. Nakhon Pathom: Institute of Language and Culture for Rural Development, Mahidol University.
- Diskul, S. 1996. Foreword. In *Ancient capitals of Thailand*, edited by E. Moore, P. Stott & S. Sukhavasti. Bangkok: River Books.
- Ferlus, M. 2006. Sur l'origine de la dénomination "siam". *Aséanie* 18: 115-117.
- Gedney, W. J. 1989a. Evidence for another series of voiced initials in Proto-Tai. In R. J. Bickner, J. Hartmann, T. J. Hudak & P. Peyasantiwong (Eds.), *Selected Papers on Comparative Tai Studies: William J. Gedney*. pp. 229-270. Ann Arbor: Center for South and Southeast Asian Studies: University of Michigan.
- . 1989b. Tai names for the ox. In *South-East Asian Linguistics in Honour of Eugenie J. A. Henderson*, edited by J. H. C. S. Davidson, pp. 89-110. London: School of Oriental and African Studies.
- Gosling, B. 1998. *A chronology of religious architecture at Sukhothai, late thirteenth to early fifteenth century*. Chiang Mai, Thailand: Silkworm Books.
- Grierson, G. A. 1903. *Linguistic survey of India*. Calcutta: Office of the Superintendent of Government Printing House.
- Hall, D. G. E. 1968. *A history of Southeast Asia*. London: Macmillan.

- Harris, J. G. 1975. A comparative word list of three Tai Nüa dialects. In *Studies in Tai linguistics in honor of William J. Gedney*, edited by J. G. Harris & J. R. Chamberlain, pp. 202-230. Bangkok: Central Institute of the English Language and Office of State Universities.
- Haudricourt, A.-G. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242:69-82.
- Hudak, T. J. (Ed.). 1994. *William J. Gedney's Southwestern Tai dialects : glossary, texts, and translations*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- (Ed.). 1996. *William J. Gedney's the Lue language : glossary, texts, and translations*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- (Ed.). 2001. *Minot's White Tai dictionary: a phonetic version*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- (Ed.). 2008. *William J. Gedney's Comparative Tai source book*. Honolulu: University of Hawaii Press.
- Jonsson, N. L. 1991. *Proto-Southwestern Tai*. Ph.D. dissertation, State University of New York at Albany, Albany, NY.
- Karlgren, B. 1957. *Grammata serica recensa*. Stockholm: Museum of Far Eastern Antiquities.
- Kullavanijaya, P. 2001. *Phasa Thai Nuea [The Tai Nüa language]*. Bangkok: Faculty of Arts, Chulalongkorn University.
- Li, F.-K. 1960. A tentative classification of Tai dialects. In *Culture in history: essays in honor of Paul Radin*, edited by S. Diamond, pp. 951-959. New York: Columbia University Press.
- . 1971. Shangguyin yanjiu [Studies on Archaic Chinese phonology]. *Bulletin of the Institute of History and Philology*, 9.1-2, 1-61.
- . 1976. Sino-Tai. *Computational Analysis of Asian and African Languages*, 3:39-48.
- . 1977. *Handbook of Comparative Tai*. Honolulu: University of Hawai'i Press.
- Luce, G. H. 1958. The early Syam in Burma's history. *Journal of the Siam Society* 46: 123-214.
- . 1959. Note on the peoples of Burma in the 12th-13th century A.D. *Journal of the Burma Research Society* 42: 52-74.
- Luo, Y. 1997. *The subgroup structure of the Tai languages: A historical-comparative study*. Berkeley: Project on Linguistic Analysis, University of California, Berkeley.
- . 1999. *A dictionary of Dehong, Southwest China*. Canberra: Pacific Linguistics.
- Manomaivibool, P. 1975. *Study of Sino-Thai lexical correspondences*. Ph.D. dissertation, University of Washington, Seattle.
- . 1976a. Layers of Chinese loanwords in Thai. In *Tai linguistics in honor of Fang-Kuei Li*, edited by T. W. Gething, J. G. Harris & P. Kullavanijaya, pp. 179-184. Bangkok: Central Institute of

- English Language, Office of State Universities.
- . 1976b. Thai and Chinese- are they genetically related? *Computational Analyses of Asian and African Language* 6: 11-31.
- Matras, Y. 2007. The borrowability of structural categories. In *Grammatical borrowing in cross-linguistic perspective*, edited by Y. Matras & J. Sakel, pp. 31-74. Berlin: Walter de Gruyter.
- Nishida, T. 1975. Common Tai and Archaic Chinese. *Studia Phonologica* 9.2-12).
- O'Connor, R. 1995. Agricultural change and ethnic succession in Southeast Asian states: A case for regional anthropology. *Journal of Asian Studies* 54.4: 968-996.
- Ostapirat, W. 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In *The peopling of East Asia: putting together archaeology, linguistics and genetics*, edited by L. Sagart, R. Blench and A. Sanchez-Mazas, pp. 107-131. London; New York: RoutledgeCurzon.
- Phumisak, C. 1976. *Khwampenma khong kham Sayam, Thai, Lao lae Khom, lae laksana thang sangkhom khong chue chonchat [History of the words Siam, Thai, Lao, and Khom, and Social aspects of ethnonyms]*. Bangkok: The Foundation for the Promotion of Social Sciences and Humanities Textbooks Project.
- Pittayaporn, P. 2005. *Phu Thai Kapong*. field note. Cornell University.
- . 2009a. *Phonology of Proto-Tai*. Ph.D. dissertation, Cornell University, Ithaca, NY.
- . 2009b. Proto-Southwestern-Tai revised: A new reconstruction. *Journal of the Southeast Asian Linguistics Society* 2: 121-144.
- Pulleyblank, E. G. 1970. Late Middle Chinese. *Asia Major* 15.2: 197-239.
- . 1984. *Middle Chinese : a study in historical phonology*. Vancouver: University of British Columbia Press.
- . 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: University of British Columbia Press.
- Sagart, L. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43.2: 411-444.
- . 2005. Tai-Kadai as a subgroup of Austronesian. In *The peopling of East Asia: putting together archaeology, linguistics and genetics*, edited by L. Sagart, R. Blench and A. Sanchez-Mazas, pp. 177-181. London; New York: RoutledgeCurzon.
- Sai Aung Tun, U. 2009. *History of the Shan state: From its origins to 1962*. Chiang Mai: Silkworm Books.
- Sarawit, M. E. 1973. *The Proto-Tai vowel system*. Ph.D. dissertation, University of Michigan, Ann Arbor.
- Saraya, D. 2002. *Kwa cha pen khon Thai [Becoming Thai]*. Bangkok: The Institute of Asian Studies, Chulalongkorn University.

- Schmidt, W. 1926. *Die Sprachfamilien und Sprachenkreise der Erde* (Vol. Carl Winter). Heidelberg.
- Schuessler, A. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawai'i Press.
- . 2009. *Minimal Old Chinese and Later Han Chinese: A companion to Grammatica serica recenset*. Honolulu: University of Hawai'i Press.
- Thurgood, G. 2002. A comment on Gedney's proposal for another series of voiced initials in Proto-Tai. In *Studies in Southeast Asian Languages*, edited by R. B. Bauer, pp. 169-183. Canberra: Pacific Linguistics.
- Wang, S. S. 1966. *Phonology of Chinese loanwords in a northern Tai dialect*. Ph.D. dissertation, University of Washington, Seattle.
- Wongthes, S. 1994. *Khon Thai yu thini thi Usakhane [The Thai were here in Southeast Asia]*. Bangkok: Silpakorn University and Art & Culture.
- . 2005. *Khon Thai ma chak nai? [Where did Thais come from?]*. Bangkok: Matichon.
- Wulff, K. 1934. *Chinesisch und Tai : sprachvergleichende Untersuchungen*. København [Copenhagen]: Levin und Munksgaard, Ejnar Munksgaard.
- . 1942. *Über das Verhältnis des Malayo-Polynesischen zum Indochinesischen*. København [Copenhagen]: Levin und Munksgaard.
- Wyatt, D. K. 1984. *Thailand: a short history*. Bangkok: Silkworm Books.
- Xing, G. 2000. *Honghe shangyou Daiyu [Tai Ya of the Upper Reach of the Red River]*. Beijing: Language and Literature Publishing.