

Chapter 16

Convergence and divergence of tone paradigms across Tai dialects in the 21st century

Chingduang Yurayong^{a,b}, Saknarin Pimvunkum^b & Yuttaporn Naksuk^b

^aUniversity of Helsinki ^bMahidol University

The present study examines tone paradigms of Tai dialects spoken in Laos, Malaysia, Myanmar, Thailand, and Vietnam. The focus is on the homophones between the paradigmatic tone D in checked syllables and tones A, B, and C in smooth syllables. The method comprises a conventional historical-comparative approach to dialectal classification and language contact among the Tai languages, complemented by a computer-aided quantitative approach which makes treatment of data collected from 315 different dialect speakers of 17 Tai languages possible. The data illustration generated through a Neighbor-Net algorithm identifies shift of the tone paradigm in some diaspora dialects which have converged towards models from sociolinguistically dominant dialects in their speaking locations.

1 Introduction

The Tai languages form a second level subbranch under the Tai-Kadai language family, one of the indigenous language groups in Mainland Southeast Asia. According to the mainstream view, the Proto-Tai-Kadai language is believed to have been spoken ca. 5000 years ago in South China (Ostapirat 2005: 126) before its dispersal into several first level subbranches: 1) Kra, 2) Hlai, and 3) Kam-Tai. The taxonomical structure, diversification of the Tai-Kadai language family and estimated dates of dispersals at different stages of intermediate protolanguages are given in Figure 1.

Chingduang Yurayong, Saknarin Pimvunkum & Yuttaporn Naksuk. 2025. Convergence and divergence of tone paradigms across Tai dialects in the 21st century. In Susanne Wagner & Ulrike Stange-Hundsdörfer (eds.), *(Dia)lects in the 21st century: Selected papers from Methods in Dialectology XVII*, 403–434. Berlin: Language Science Press. DOI: 10.5281/zenodo.15006623 



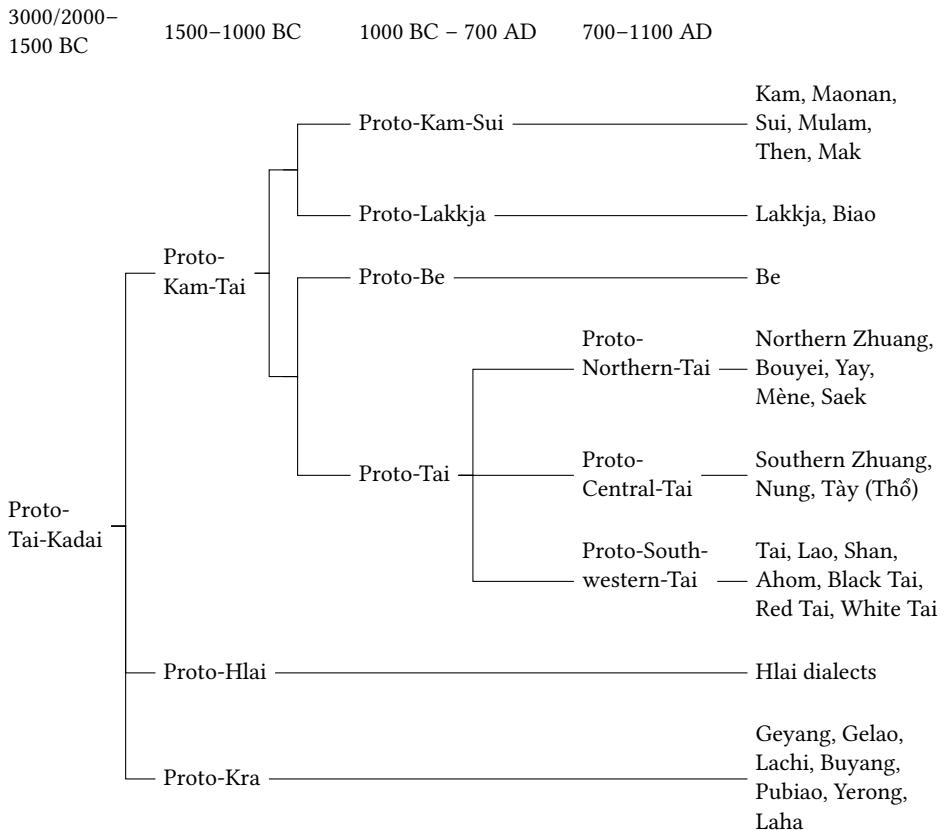


Figure 1: Diversification of Tai-Kadai languages (based on Mitani 1977, Edmondson & Solnit 1997, Pittayaporn 2014).

As is indicated in Figure 1, the Tai branch stems off from the Kam-Tai intermediate protolanguage, which makes the Kam-Sui languages linguistically closer to the Tai languages, compared to the more distantly related Kra and Hlai languages. This close genealogical tie is evident from a higher proportion of shared lexical items between Kam-Sui and Tai (see Luo 1988, Edmondson & Solnit 1997: 4).

In terms of geographical distribution, the Tai branch and its linguistic populations have spread out from their hypothetical Urheimat in South China towards South Asia and Mainland Southeast Asia. Figure 2 illustrates the speaking areas of modern Tai-Kadai languages. Today, there is a total of ca. 102 million Tai-Kadai speaking people (Liao 2023a: 199), of which ca. 21 million are living in China. As can be seen from Figure 2, the Tai branch has extended its speaking area to the

south and west, significantly farther than its sister languages, Kam-Sui, Kra, and Hlai.

Focusing on the Tai branch, previous studies have proposed a taxonomical structure with the following three major groups distributed across Mainland Southeast Asia (Chamberlain 1975, Li 1977; see also Luo 1997 for an alternative classification with four groups):

1. Northern Tai – the majority spoken in China, and the Saek language spoken in Laos, Thailand, and Vietnam
2. Central Tai – spoken in China and Vietnam
3. Southwestern Tai – spoken in China, India, Laos, Malaysia, Myanmar, Thailand, and Vietnam.

In any case, such a subgrouping remains disputable when certain isoglosses concerning sound changes occur across subbranches. By only applying a conventional comparative method of historical linguistics, common patterns and clustering of phonological innovations namely do not give a straightforward dialectal classification. This is likely due to later language contact among Tai dialects, resulting from migration and relocation of the linguistic populations and causing convergence among Tai dialects spoken in adjacent areas (Pittayaporn 2009, Liao 2023b). A method which carefully deals with potential contact-induced changes and incorporates them into the classification yields a result illustrated in Figure 3.

For instance, innovation Q concerns a sound change involving the merger $*kr-$ = $*h_r-$, which is the case in Shan, Thai (Siamese), Black Tai and Lue, but not in Sapa (Pittayaporn 2009: 301).

In terms of population history, it has been the case especially in Laos and Thailand where people were compelled to migrate to new locations in the region, often as captives during the multiple war time periods between 1750 and 1850 (Piyabhan 1998). The series of such historical events has given rise to new diaspora communities dispersing around Thailand in particular. The resettlement of Tai dialect speakers primarily concerned community members from the first generation, whereas the younger second and third generations of dialect speakers already exhibit signs of shifting towards a national language or the dominant regional dialect of their current location (Akharawatthanakun 2003, Bunyasathit et al. 2016).

For many decades, the evolution and convergence among Tai dialects have been popular research topics in Thai and Tai-Kadai linguistics among researchers and students, particularly in Thai universities. Among hundreds of individual

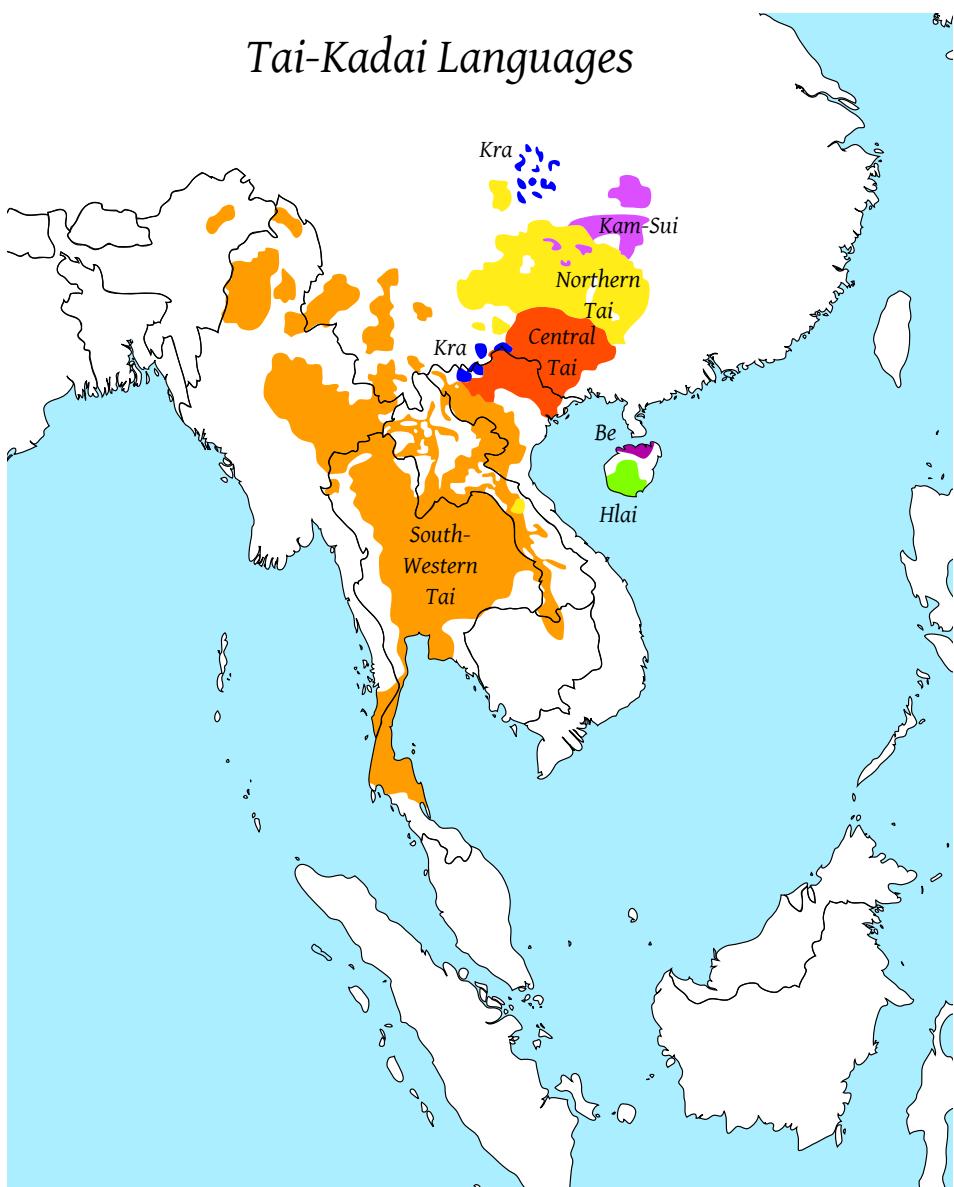


Figure 2: Geographical distribution of modern Tai-Kadai languages (public domain, <https://upload.wikimedia.org/wikipedia/commons/7/71/Taikadai-en.svg>).

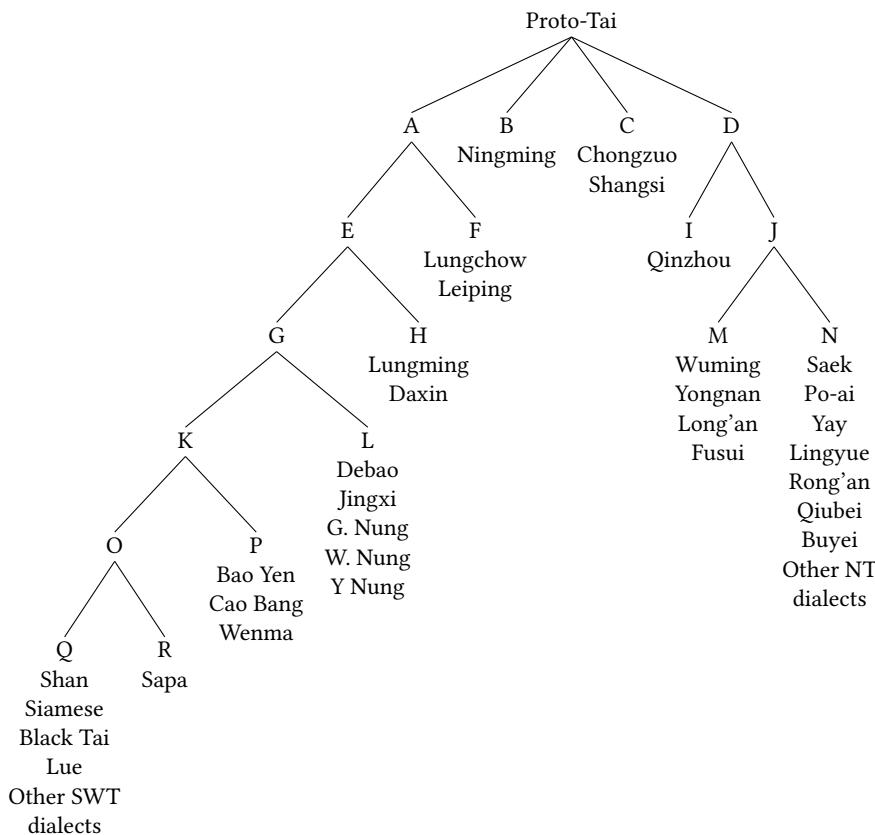


Figure 3: A classification of Tai dialects with reference to contact-induced changes (Pittayaporn 2009: 298).

studies and theses published after the comprehensive handbook of Tai dialects by Li (1977), extensive works on Tai dialects by Gedney (edited by Hudak 2008) and a comparative work on Tai tone paradigms by Brown (1985), there have been several large-scale comparative studies of tone paradigms. To name just few, the comparative approach has been applied for the investigation of tonal variation in Yo dialects (Koowatthanasiri 1981), Phu Thai dialects (Srithonrat 1983), Lao dialects (Akharawatthanakun 2003), Thai dialects in Malaysia (Damanhuri 2004), Central Thai dialects (Canilao 2010), Black Tai dialects (Burusphat 2012), and Tai Lue dialects (Akharawatthanakun 2020). Consequently, the consensus is continuously evolving due to newly collected data from individual dialects which can contradict and deviate from earlier proposals in the genealogical classification of the Tai languages.

Participating in the ongoing discussion in dialectological studies of Tai-Kadai languages, our first goal in the present study is to identify change in progress. We choose homophones in tone paradigms of different Tai dialects as our object of study. The patterns of these homophones may have either resisted changes or diverged from their protosystems as we have arrived in the 21st century. Concretely, our attention is directed towards (i) paradigmatic tones DL and DS, which represent checked syllables with final stops [-p/-t/-k/-ʔ], and (ii) their alignment with tones A, B and C, which represent smooth syllables with final vowels or sonorants (see Section 2 for the detailed definition of tone paradigm).

The nature of the current study involves historical-comparative linguistics in conjunction with a quantitative approach, utilising computer-aided tools to investigate data from more than 300 different dialect speakers under investigation and aiming to trace significant signals of convergence among their tone paradigms. With this method, our second goal is to demonstrate how to address language changes occurring in the 21st century and their potential impact on how we read and interpret the taxonomical structure of the Tai languages.

We divide our study into five sections. After introducing the research questions, hypotheses, and historical background in Section 1, in Section 2 we present a diachronic framework for the studies of tones, which has been previously employed for Tai as well as other Southeast Asian languages. Building upon the state of the art, in Section 3 we provide a description of how to apply a computer-aided quantitative method to investigate and analyse tone paradigm data from different Tai dialect speakers. Based on the data analysis, in Section 4 we discuss the scenarios of convergence and divergence of tone paradigms in general and in individual cases, which can be explained by sociolinguistic factors. Finally, in Section 5 we give some conclusional thoughts on the functionality of the applied quantitative method and highlight some areas which deserve attention in future studies and data collection, in order to conduct diachronic studies in an even more efficient manner.

2 Tone system in Tai and other Southeast Asian languages

Tones in Tai as well as in other languages in Southeast Asia, such as the cognate Kam-Sui, Hlai and Kra languages, along with neighbouring Chinese, Karenic, Hmong-Mien, and Vietic languages, are organised within a paradigm. A general principle dictates that the distribution of tones in paradigm is controlled by two factors: (i) the type of initial consonant, and (ii) the type of syllable. Both factors offer clues which can be traced back to phonological shapes at the protolanguage

stage. Therefore, the tone paradigm serves as an important piece of evidence for reconstructing the lexicon of the protolanguage because sound changes, such as merger or loss of specific consonant types, may have left their traces on tones. This mechanism is called tonogenesis, which gave rise to tones as compensation for loss of certain phonetic features, most notably a voice distinction in the domain of consonants (see Wulff 1934, Haudricourt 1954, Ferlus 2004, Kingston 2011, Michaud & Sands 2020). Naturally, this approach implies that the protolanguages did not originally possess tones, as has been discussed, most notably, for Chinese (Handel 2014: 593, Hill 2019: 182).

In the case of Tai languages, a classification of four types of initial consonants has been associated with different pitch patterns as given below (see e.g. Maddieson 1984 and Ratliff 2015 for the effect of voicing in pitch). One of the common practices to annotate pitches and contours in the studies of East Asian linguistics uses numbers to indicate pitch height: 1 = lowest pitch vs. 5 = highest pitch.

1. Aspirated – high pitch
2. Voiceless – mid pitch
3. Implosive – mid pitch
4. Voiced – low pitch

Note that the relation between manner of articulation and pitch presented here should not be taken as an absolute linguistic universal. Some Tai-Kadai languages, namely, exhibit no significant correlation between aspiration and high pitch (see Zhang 1980: 38, Edmondson 1990: 188, Liao 2016: 170–171, Zhu et al. 2016: 18). Furthermore, the status of the aspirated initials remains a subject of dispute, with questions surrounding whether they can be reconstructed as far as to Proto-Tai or only to the Proto-Southwestern-Tai intermediate stage (see arguments in Liang & Zhang 1996, Pittayaporn 2009, with a summary provided in Liao 2023a). Another dimension to consider is the syllable type, which leads to a division of four tone classes, a.k.a. tonal splits: A, B, C (smooth syllables ending with vowels or sonorants), and DL, DS (checked syllables ending with stops [-p/-t/-k/-?]).

The two factors discussed above serve as the foundation for the tool known as “Gedney’s tone box” (see Table 1), which was initially proposed by William J. Gedney (1972) and is recognised as one of the main comparative tools in the Tai-Kadai linguistics scholarship. Based on the original tone box by Gedney, we have systematised and relabelled the tone slots for the comparative purpose of

the present study (see Table 2). According to Pittayaporn (2009: 271), contour characteristics of each tone are reconstructed for Proto-Tai as follows: A = mid-level, B = low-rising, C = high-falling, and D = low-rising.

Table 1: The original Gedney's (1972) Tai tone box for Proto-Tai

		Proto-Tai tones					
		A	B	C	D-short	D-long	
Initials at time of tonal splits	Voiceless	Friction	5th tone	2nd tone	3rd tone	2nd tone	2nd tone
			1st tone				
	Voiced			3rd tone	4th tone	4th tone	3rd tone
Smooth syllables				Checked syllables			

Table 2: The modified Tai tone box (based on Gedney 1972).

Initial consonant class at time of tonal splits	Initial consonant class				
	A	B	C	DL(ong)	DS(hort)
Aspirated	A1	B1	C1	DL1	DS1
Plain	A2	B2	C2	DL2	DS2
Implosive	A3	B3	C3	DL3	DS3
Voiced	A4	B4	C4	DL4	DS4

In any case, it is important to note that Gedney's tone box has been primarily developed within the context of the Southwestern branch of Tai languages. Consequently, it often has issues when applied to explain the tone paradigms in other subbranches, i.e. Central and Northern Tai, not to mention the more distantly related Tai-Kadai languages like Kam-Sui, Kra and Hlai (see the criticisms on cross-family validity of Gedney's tone box tool in Liao 2023a).

A similar type of tone box is also employed in the studies of tones in Chinese, Karenic, Hmongic and Vietic languages. However, some languages may distinguish only a contrast between voiceless and voiced initial stops, while the aspirated and implosive classes are grouped within the same category as voiceless

stops, making only the voice distinction between the initial consonant classes. Similarly, the vowel length distinction is not considered a meaningful category in some languages. See, for instance, a tone box with only two initial consonant classes, based on the reconstructed Middle Chinese tone paradigm and used to describe tone paradigms across modern Chinese dialects in Table 3. Interestingly, certain Middle Chinese tone names appear to record the contour characteristics pronounced during the Middle Chinese period (cf. the reconstruction of Proto-Tai tone contours in Pittayaporn 2009: 271).

Table 3: The Chinese tone box

Middle Chinese name of tone				
	平 <i>píng</i> 'level'	上 <i>shǎng</i> 'rising'	去 <i>qù</i> 'falling'	入 <i>rù</i> 'entering'
Tone label	A	B	C	D
Voiceless initials	A1	B1	C1	D1
Voiced initials	A2	B2	C2	D2

From the paradigms in modern Tai languages, we know that tone D can be traced back to a checked syllable structure ending with stops [-p/-t/-k] in Proto-Tai. However, there is unfortunately no internal evidence available for syllable structures of tones A, B and C prior to the emergence of tones in Proto-Tai-Kadai. This is in contrast to some other neighbouring languages, for example, Chinese where sufficient evidence for syllable structures during the pre-tonal Old Chinese stage is attested in earlier written sources and phonological adaptation of Sanskrit loanwords in Old Chinese, as well as Old Chinese loanwords in Korean (Handel 2014: 594, Hill 2019: 184–185). A reconstruction of the Old Chinese syllable structures is given as follows:

1. A – a syllable ending with a vowel or sonorant
2. B – a syllable ending with a glottal stop [-?]
3. C – a syllable ending with a sibilant [-s]
4. D – a syllable ending with a stop [-p/-t/-k]

Similar patterns and mechanisms have also been applied together with evidence from Chinese loanwords to explain tonogenesis in Hmong-Mien (Ratliff

2010: 183–184) and Vietic (Thurgood 2002), both of which have undergone early contact with Chinese. Interestingly, in the field of Tai-Kadai linguistics, there has been no serious attempt to reconstruct distinct original syllable types for tones A, B and C in Proto-Tai-Kadai and their corresponding reflexes in subsequent intermediate protolanguages, although there are ample early Chinese loanwords available for similar analysis (cf. Pittayaporn 2014), as has been done for Hmong-Mien and Vietic.

Relevant to the current study which focuses on Tai tone paradigms is the observation that tones DL and DS always align paradigmatically with some of the twelve phonemic tone slots in columns A, B, and C (cf. the similar low-rising contour reconstructed for tones B and D as given in Pittayaporn 2009: 271). However, in practice, the maximum number of phonemic tone distinctions observed in Tai-Kadai languages is nine, often due to the merger of tones between the plain (tones A2/B2/C2/D2) and implosive (tones A3/B3/C3/D3) classes of initial consonants in most languages. Ultimately, only a few Tai-Kadai languages possess nine distinct phonemic tones in a symmetrical fashion, a characteristic primarily found in certain dialects of the Kam language within the Kam-Sui branch. Table 4 illustrates pairs of examples, where the former represents a smooth syllable (tone A, B, or C) and the latter represents a checked syllable (tone DL or DS), except the last three tones (53, 453, and 33) which only occur in smooth syllable. Despite the symmetrical case of three consonant classes combined with three syllable types in Kam, the average number of tones across the Tai-Kadai family typically falls within the range of five to seven tones (see e.g. Brown 1985, Hudak 2008).

Table 4: Tones of common Kam (modified from Yang & Edmondson 2008: 514).

Tone value (contour)	Example	Gloss	Tone value (contour)	Example	Gloss
55	<i>ma</i> ⁵⁵	‘vegetable’	13	<i>no</i> ¹³	‘rat’
	<i>jak</i> ⁵⁵	‘wet’		<i>p^hat</i> ¹³	‘blood’
35	<i>ma</i> ³⁵	‘to come’	31	<i>ma</i> ³¹	‘horse’
	<i>mat</i> ³⁵	‘flea’		<i>mak</i> ³¹	‘ink’
11	<i>ma</i> ¹¹	‘tongue’	53	<i>ja</i> ⁵³	‘paddy, wet field’
	<i>sak</i> ¹¹	‘thief’		<i>ma</i> ⁴⁵³	‘to soak (rice)’
323	<i>ma</i> ³²³	‘cloud, soft’	33	<i>pa</i> ³³	‘rice husk’
	<i>mak</i> ³²³	‘big’			

The subgrouping of individual Tai languages can be discerned through both consonant and vowel inventories, as well as their tone paradigm. Concerning the latter, the distinctive homophone patterns across tone slots exhibit characteristics specific to individual language groups. This aspect of tone paradigms has significantly received scholarly attention within the field of Tai dialectology (see e.g. Brown 1985, Dockum 2019). Below, we present several characteristics unique to individual dialects, which are not commonly observed in other dialectal groups. These examples are drawn from dialectal data in Brown (1985) and Hartmann (2008) and specifically focus on the fundamental paradigmatic tones A, B, and C. Further explanations regarding individual dialectal groups are provided below following the paradigmatic summary in Table 5.

Table 5: Some characteristics of the tone paradigms in the selected Tai dialectal groups.

Tai Lue & White Tai			Tai Yuan / Northern Thai			Lao		
A1	B1	C1	A1	B1	C1	A1	B1	C1
A2	B2	C2	A2	B2	C2	A2	B2	C2
A3	B3	C3	A3	B3	C3	A3	B3	C3
A4	B4	C4	A4	B4	C4	A4	B4	C4

Shan & Central Thai			Southern Thai		
A1	B1	C1	A1	B1	C1
A2	B2	C2	A2	B2	C2
A3	B3	C3	A3	B3	C3
A4	B4	C4	A4	B4	C4

1. Tai Lue and White Tai dialects – a symmetrical split of all the paradigmatic tones A, B and C for smooth syllables between the original unvoiced and voiced initials
 - A1, A2, A3 ≠ A4
 - B1, B2, B3 ≠ B4
 - C1, C2, C3 ≠ C4

2. Tai Yuan/Northern Thai dialects – a symmetrical split as observed in Tai Lue dialects, but with a distinction between the original aspirated/plain and implosive/voiced initials within the paradigmatic tone A
 - A1, A2 ≠ A3, A4
 - B1, B2, B3 ≠ B4
 - C1, C2, C3 ≠ C4
3. Lao dialects – a merger of all tone slots within the paradigmatic tone B
 - B1 = B2 = B3 = B4
4. Shan and Central Thai dialects – a diagonal alignment of the original aspirated/plain/implosive and voiced initials within paradigmatic tones B and C
 - C1, C2, C3 = B4
5. Southern Thai dialects – a symmetrical split of the paradigmatic tones within tone A with distinctions for the original aspirated, plain/implosive and voiced initials, as well as a symmetrical split of the paradigmatic tones A, B, and C with the original voiced initials
 - A1 ≠ A2, A3 ≠ A4
 - A4 ≠ B4 ≠ C4

Next, we shift our focus to the primary data of the present study and the computer-aided methods and processes employed therein.

3 Data and methods

3.1 Data coverage

In the present study, we conduct a comprehensive investigation using 315 datapoints obtained from data of speakers of Tai dialects, primarily from regions outside China. Our focus is on analysing patterns observed within the tone paradigm. The data is collected from a range of grammatical and phonological descriptions of Tai dialects in Mainland Southeast Asia. These descriptions have been published during the late-20th and early-21st centuries. Most, if not all, of the sources provide empirical data of individual dialect speakers, presented in the form of word lists covering all the tone slots outlined in Table 2. Alignment

patterns observed among different tone slots are also often discussed in a tone box similar to that of Table 5. The majority of these sources usually have their goal in describing how the observed tone paradigms correspond to or deviate from the common pattern of the specific Tai language. Some works, most notably Akharawatthanakun (2003), further delve into sociolinguistic factors which may have significantly influenced the evolution of the tone paradigm when such data is available and feasible from speech communities (see Section 4 for a similar attempt in the present study).

Out of the 315 datapoints, the data covers dialects from a total of 17 different Tai languages as listed below.¹

- | | |
|----------------------------------|-----------------|
| 1. Tai Lue | 10. Saek |
| 2. Tai Yuan/Northern Thai | 11. Shan |
| 3. Phu Thai | 12. White Tai |
| 4. Northern Lao (Luang Phrabang) | 13. Black Tai |
| 5. Central Lao (Vientiane) | 14. Yo |
| 6. Southern Lao (Champasak) | 15. Yoy |
| 7. Western Lao/Northeastern Thai | 16. Phuan |
| 8. Central Thai | 17. Khorat Thai |
| 9. Southern Thai | |

In the 21st century, dialects of various languages are spoken in diaspora areas, particularly among Lao dialect speakers who have relocated from the dialectal core regions in Luang Phrabang, Vientiane, or Champasak to various areas of Thailand. The geographic locations of the speakers included in the dataset are cartographically illustrated in Figure 4, using ArcGIS tools. As can be seen from the map in Figure 4, a noteworthy proportion of dialect speakers are not located within the core area of their respective languages, particularly the Lao diaspora dialects due to their migration history discussed by Piyabhan (1998, mentioned in Section 1). This phenomenon raises an important issue potentially contributing to the restructuring of tone paradigms among certain dialect speakers, a matter which will be discussed further in Section 4.

¹The full details and sources for each datapoint can be found in the supplementary material: https://researchportal.helsinki.fi/files/244396698/Convergence_and_divergence_of_tone_paradigms_across_Tai_dialects_in_the_21st_century_supplementum_.xlsx.

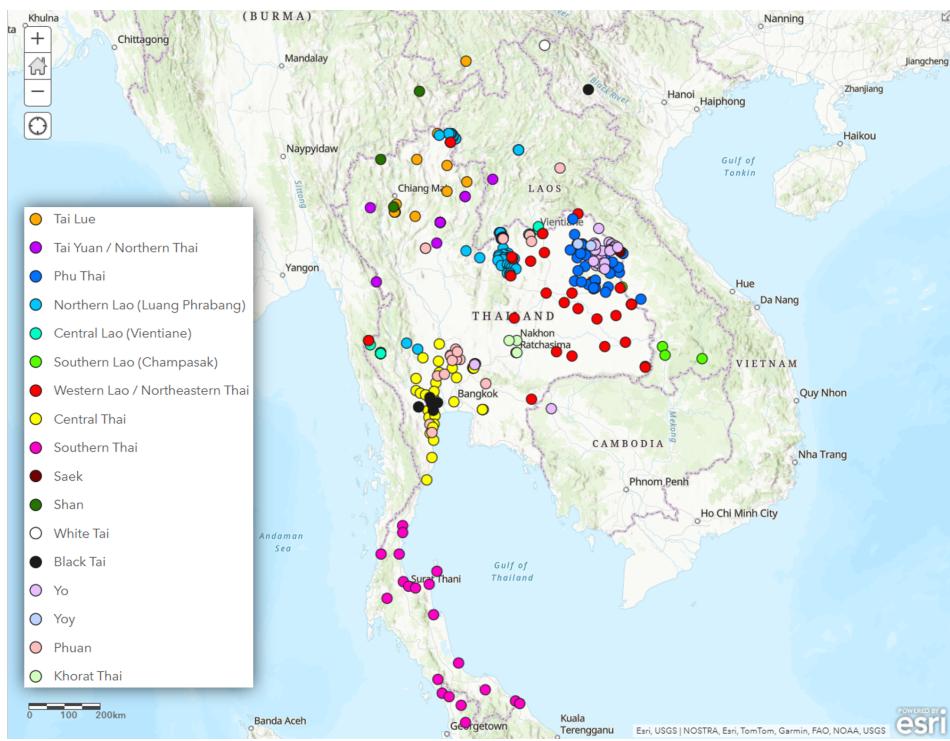


Figure 4: Datapoints of the Tai dialect speakers under investigation.

3.2 Methods for data illustration and analysis

As a general principle for organising the data, the information collected from various grammatical and phonological descriptions of Tai dialects is arranged based on whether tones DL and DS align with tones A, B, and/or C within individual dialects. This compiled information is then processed by a Neighbor-Net algorithm (Bryant & Moulton 2004). The outcome of this processing is a network diagram visualising the clustering and distances among typological profiles of each dialect under investigation. Importantly, this diagram is generated without exhibiting any bias towards the genealogical relationship or distances across dialects. This approach has been effectively applied to studies of several language families across Eurasia (e.g. Grünthal & Nichols 2016, Szeto et al. 2018, Nichols 2020, Yurayong & Szeto 2020, Szeto & Yurayong 2021). The data processing involves four different steps, which are elaborated upon in the following points below.

3.2.1 Step 1: From language description to tone paradigm

It is a common practice in Tai-Kadai linguistics for the phonological descriptions of Tai dialects to include details about tone contours (1 = lowest pitch vs. 5 = highest pitch) and paradigms presented in the form of tone box (as previously illustrated in Table 2 within Section 2). Utilising these tone boxes, the homophones between paradigmatic tones DL and DS, juxtaposed with tones A, B, and C, can be identified, as exemplified in Table 6. The tone paradigm in Table 6 reveals that the speaker of the Central Lao dialect spoken in Suan Pan, Nakhon Pathom province of Thailand, has six distinct phonemic tones. Among these, four alignment patterns of tones DL and DS are identified: (i) DL123 = C1, (ii) DL4 = C234, (iii) DS123 = A1, and (iv) DS4 = B1234. The alignment patterns, identified from the tone paradigms of individual dialect speakers, serve as the focal points of comparison in the quantitative analysis conducted in the present study.

Table 6: The tone paradigm of the Central Lao dialect speaker (LC24) in Suan Pan, Nakhon Pathom province, Thailand.

Initial consonant class at time of tonal splits	A	B	C	DL	DS	→	Alignment pattern
Aspirated	24			21			DL123 = C1
Plain		33			24		DL4 = C234
Implosive							DS123 = A1
Voiced	353		41		33		DS4 = B1234

3.2.2 Step 2: From tone paradigm to binary data

The homophones aligned between tones DL and DS and tones A, B, and C are then converted into binary values: 0 = no alignment vs. 1 = alignment present, as displayed in Table 7. This binary value representation characterises the tonological profile of each dialect speaker, drawing an analogy to chromosomes and their sequencing within a species. This is a concept commonly employed in evolutionary studies, in which the original development of this method finds its root (see general principles for the application of Neighbor-Net diagrams in evolutionary studies in Maddison et al. 1997, Huson & Bryant 2006). Subsequently, this dataset of tonological profiles will serve as a foundational component for data illustration model generated by the Neighbor-Net algorithm in Step 4 described in Section 3.2.4.

Table 7: The tonological profile of the Central Lao dialect speaker (LC24) in Suan Pan, Nakhon Pathom province, Thailand.

DL1												DL2											
A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
DL3												DL4											
A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
DS1												DS2											
A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
DS3												DS4											
A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

3.2.3 Step 3: From binary data to nexus format

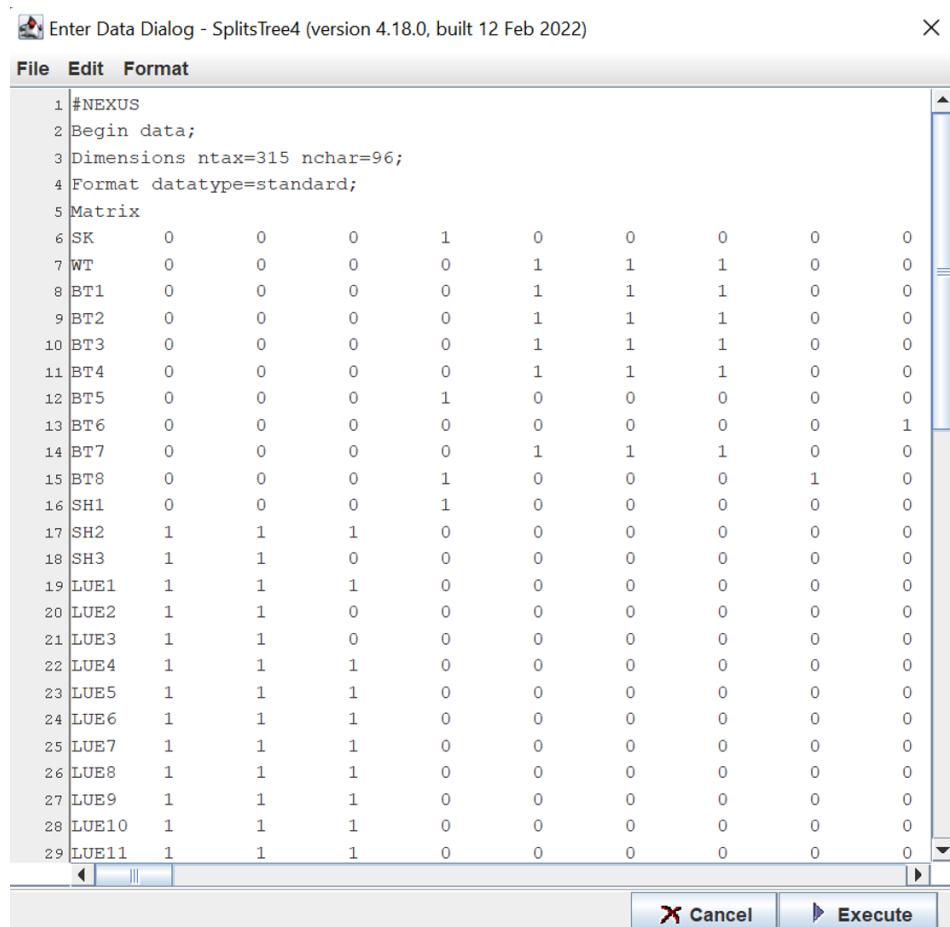
The binary data is converted into the nexus format (Maddison et al. 1997) which is a kind of structured and systematic data format for facilitating computational modelling tasks, such as the implementation of the Neighbor-Net algorithm, as demonstrated in Figure 5. The instructions for conducting the computational modelling process indicate several key parameters: (i) the total number of data-points (`ntax=315`) from 315 dialect speakers included, (ii) the total number of tone slots (`nchar=96`) from 12 tone slots within tones A, B and C \times 8 tone slots within tones DL and DS, and (iii) a standard type of data analysis (`datatype=standard`) which is predicated on binary values (0 = no alignment vs. 1 = alignment present). These parameters collectively guide the computational modelling process, enabling the extraction of meaningful insights from the dataset.

3.2.4 Step 4: From nexus format to Neighbor-Net diagram

The final step of the data illustration process involves feeding the organised data, previously formatted according to the nexus format, into the software `Splitstree4` (version 4.18.0, built on 12 February 2022). This software operates using the Neighbor-Net algorithm, and by executing the imported data (as exemplified in Figure 5), the software performs calculations and generates visual representations of the distances across tonological profiles of each dialect. The calculations

also generate clustering of datapoints with similar profiles beneath the same nodes creating a clear illustration of the relationships, as illustrated in Figure 6.

At this point, the computer-aided tools have fulfilled their role on the quantitative front, presenting the data in a visually interpretable fashion. The subsequent work phase involves a qualitative analysis based on the Tai-Kadai linguistics scholarship. This qualitative account seeks to identify significant signals which may be indicative of language changes.



The screenshot shows the 'Enter Data Dialog - SplitsTree4 (version 4.18.0, built 12 Feb 2022)' window. The data is in NEXUS format, starting with '#NEXUS' and 'Begin data;'. The 'Dimensions ntax=315 nchar=96;' line is present. The 'Format datatype=standard;' line follows. The 'Matrix' section contains 29 rows of data, each representing a taxon (SK, WT, BT1, BT2, BT3, BT4, BT5, BT6, BT7, BT8, SH1, SH2, SH3, LUE1, LUE2, LUE3, LUE4, LUE5, LUE6, LUE7, LUE8, LUE9, LUE10, LUE11) and 10 columns of binary data (0 or 1). The data shows various patterns of 1s and 0s across the taxa. The window has a standard Windows-style interface with a menu bar (File, Edit, Format), scroll bars, and buttons for 'Cancel' and 'Execute'.

```

1 #NEXUS
2 Begin data;
3 Dimensions ntax=315 nchar=96;
4 Format datatype=standard;
5 Matrix
6 SK 0 0 0 1 0 0 0 0 0
7 WT 0 0 0 0 1 1 1 0 0
8 BT1 0 0 0 0 1 1 1 0 0
9 BT2 0 0 0 0 1 1 1 0 0
10 BT3 0 0 0 0 1 1 1 0 0
11 BT4 0 0 0 0 1 1 1 0 0
12 BT5 0 0 0 1 0 0 0 0 0
13 BT6 0 0 0 0 0 0 0 0 1
14 BT7 0 0 0 0 1 1 1 0 0
15 BT8 0 0 0 1 0 0 0 0 0
16 SH1 0 0 0 1 0 0 0 0 0
17 SH2 1 1 1 0 0 0 0 0 0
18 SH3 1 1 0 0 0 0 0 0 0
19 LUE1 1 1 1 0 0 0 0 0 0
20 LUE2 1 1 0 0 0 0 0 0 0
21 LUE3 1 1 0 0 0 0 0 0 0
22 LUE4 1 1 1 0 0 0 0 0 0
23 LUE5 1 1 1 0 0 0 0 0 0
24 LUE6 1 1 1 0 0 0 0 0 0
25 LUE7 1 1 1 0 0 0 0 0 0
26 LUE8 1 1 1 0 0 0 0 0 0
27 LUE9 1 1 1 0 0 0 0 0 0
28 LUE10 1 1 1 0 0 0 0 0 0
29 LUE11 1 1 1 0 0 0 0 0 0

```

Figure 5: The nexus-formatted data for computational modelling.

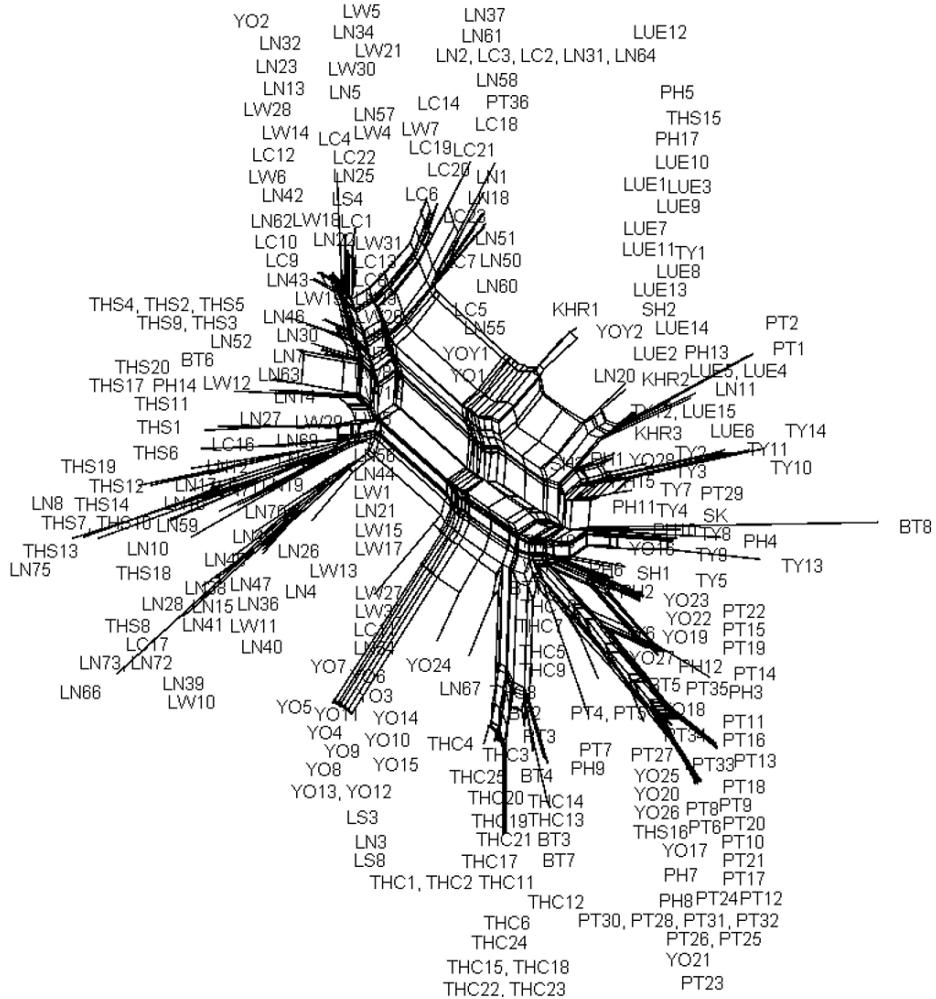


Figure 6: Neighbor-Net diagram for the tonological profiles of Tai dialects (n=315): LUE = Tai Lue; TY = Tai Yuan/Northern Thai; PT = Phu Thai; LN = Northern Lao; LC = Central Lao; LS = Southern Lao; LW = Western Lao/Northeastern Thai; THC = Central Thai; HIS = Southern Thai; SK = Saek; SH = Shan; WT = White Tai; BT = Black Tai; YO = Yo; YOY = Yoy; PH = Phuan; KHR = Khorat Thai.

4 Data interpretation and discussion

Through the Neighbor-Net diagram generated by the SplitsTree software (see Figure 6), several distinct dialect clusters can be identified. These clusters are discernible by the presence of dominant dialects indicated in the list below, and as annotated with different colours and arranged in an anti-clockwise direction within Figure 7.

- | | |
|---------------------------|----------------------|
| 1. Lao proper: LC, LS, LW | 5. Central Thai: THC |
| 2. Southern Thai: THS | 6. Phu Thai: PT |
| 3. Northern Lao: LN | 7. Tai Yuan: TY |
| 4. Yo: YO | 8. Tai Lue: LUE |

In general, the results largely agree with the genealogical classification proposed in the previous studies, as discussed in Section 1. By identifying dialects which do not belong to their respective genealogical cluster based on the histori-cal-comparative basis, we look further into their current speaking areas on the map and migration history of their respective speech communities. In this context, two specific cases will be discussed as illustrative examples.

First, the Central Lao dialect speaker LN11 is situated within a convergence of the Tai Yuan and Lue clusters, as noted in Figure 8. As highlighted on the map using a red circle, the Northern Lao dialect LN11 is presently spoken within the major Tai Yuan and Lue speaking areas.

Second, the Central Lao dialect speaker LN67 aligns with the Central Thai cluster, as indicated by a red circle in Figure 9. By examining the map, the location of the Northern Lao dialect LN67 is currently situated adjacent to the major speaking areas of Central Thai dialects.

Both of the aforementioned cases highlight a scenario where the examined dialects have undergone or are undergoing a shift in their homophonous tone pattern of DL and DS. This shift reflects a convergence towards the tonal profile of a regional dialect within their newly settled area, in line with their sociocultural assimilation. This phenomenon can be effectively demonstrated by comparing their tone paradigms with our reconstructed common Lao tone system in Table 8, serving as a baseline.

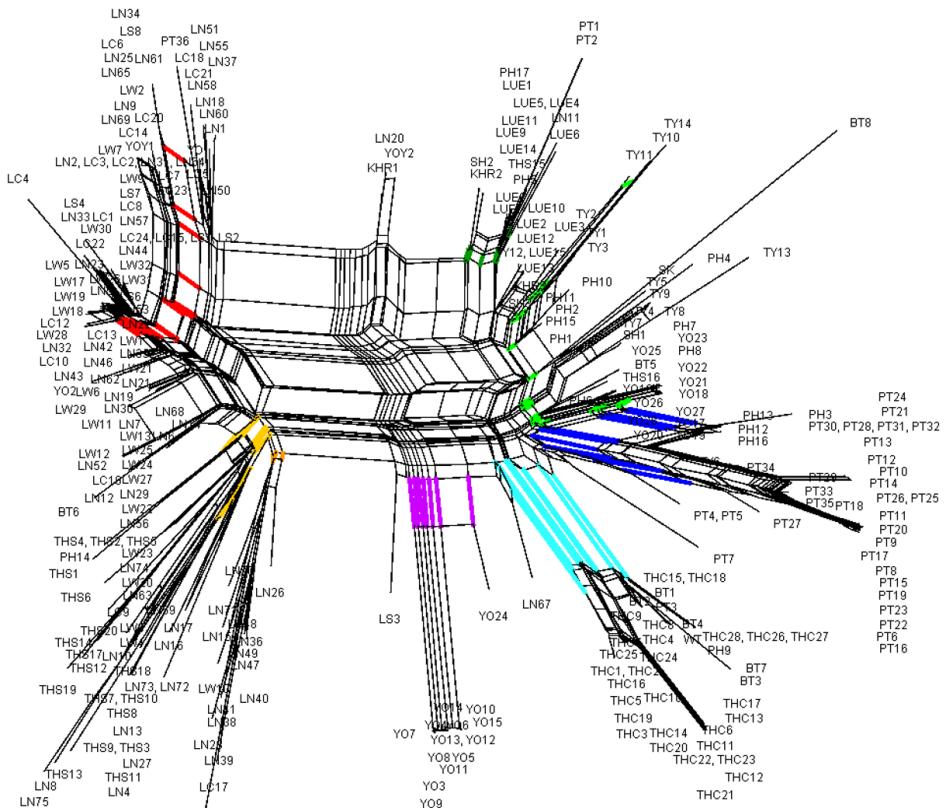


Figure 7: A Neighbor-Net diagram for the tonological profiles of Tai dialects (n=315) with identified clusters: LUE = Tai Lue; TY = Tai Yuan/Northern Thai; PT = Phu Thai; LN = Northern Lao; LC = Central Lao; LS = Southern Lao; LW = Western Lao/Northeastern Thai; THC = Central Thai; THS = Southern Thai; SK = Saek; SH = Shan; WT = White Tai; BT = Black Tai; YO = Yo; YOY = Yoy; PH = Phuan; KHR = Khorat Thai.

Table 8: A reconstructed tone paradigm of common Lao.

Initial consonant class at time of tonal splits	A	B	C	DL	DS
Aspirated	1		5		
Plain		4			1
Implosive	2				
Voiced	3		6		4

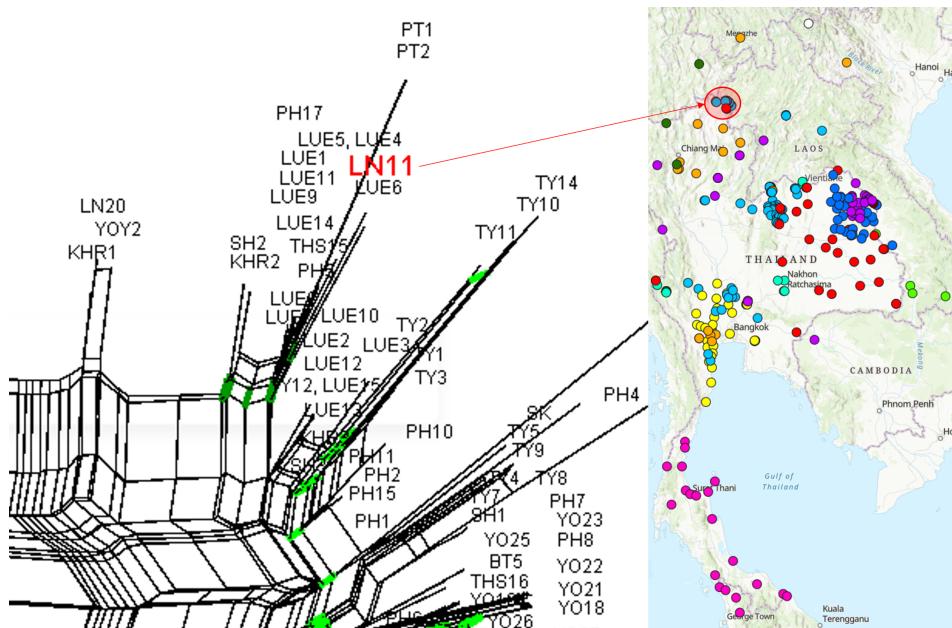


Figure 8: The Northern Lao dialect speaker (LN11) in Chiang Khong, Chiang Rai province, Thailand.

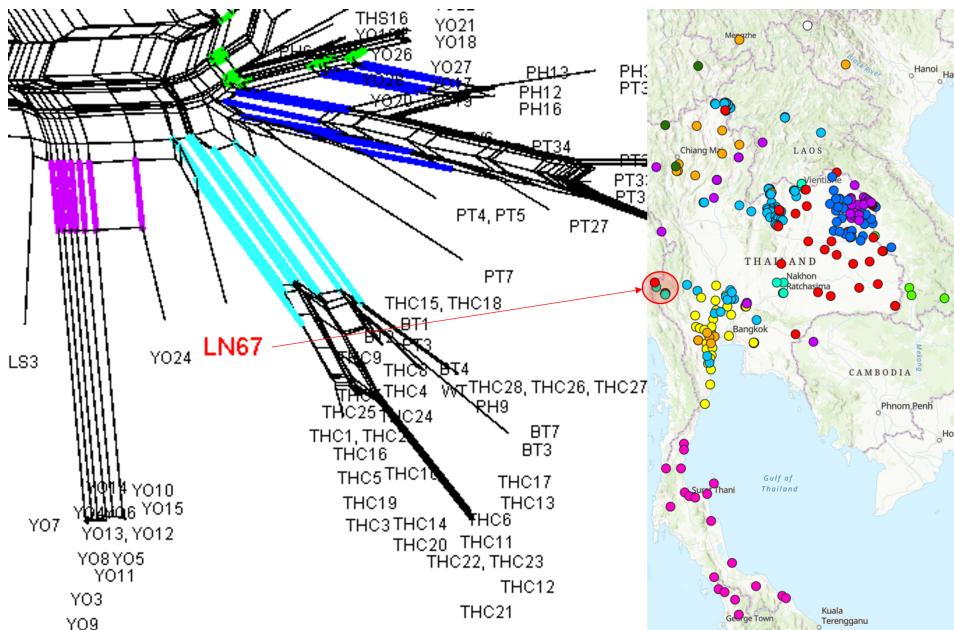


Figure 9: The Northern Lao dialect speaker (LN67) in Sangkhlaburi, Kanchanaburi province, Thailand.

There are two major characteristics of Lao tone paradigms in general and across dialects which have been, respectively, retained and undergone changes in the two dialects under investigation. The first feature is found within tone B where there is no distinction between the original initial consonant classes (as discussed in Section 2, Table 5). The second hallmark of the Lao tone paradigm is what is labelled in Tai-Kadai linguistics as the Lao ladders, which concerns the paradigmatic distribution of tones within tones C and DL, as marked in bold in Table 8.

On the one hand, the primary attribute of tone B remains largely intact across the majority of contemporary Lao dialects, including the two diaspora dialects under investigation (see Tables 9 and 10), irrespective of the geographical locations of their current speaking areas. Accordingly, Akharawatthanakun (2003: 337) observes that the split of tone B4 from tones B123 in any Lao dialects can be considered a strong indication of deviation from the common Lao pattern, as is evident in the case of Khorat Thai dialects regarded as hybrid dialects of Western Lao and Central Thai.

On the other hand, the second homophonous pattern between tones C and DL is gradually fading away within the two Northern Lao dialects under investigation. These dialects are spoken outside their core regions around Luang Phrabang in Northern Laos, as can be observed from the tone paradigms given in Tables 9 and 10.

Table 9: Tone paradigm of the Northern Lao dialect speaker (LN11) in Chiang Khong, Chiang Rai province, Thailand.

Initial consonant class at time of tonal splits	A	B	C	DL	DS
Aspirated			4		
Plain	1	3	5	3	1
Implosive					
Voiced	2		4	6	3

Our hypothesis is that the occurrence of a paradigmatic shift is ongoing, indicating a trend converging towards the tonal profiles prevalent in other Tai dialects, which are predominantly spoken by the majority populations in the areas. In order to assess the validity of our hypothesis, a comparative analysis is conducted using two additional baselines: Tai Yuan (Table 11) and Central Thai (Table 12).

Table 10: Tone paradigm of the Northern Lao dialect speaker (LN67) in Sangkhlaburi, Kanchanaburi province, Thailand.

Initial consonant class at time of tonal splits	A	B	C	DL	DS
Aspirated	1				
Plain			4	3	6
Implosive	2	3			5
Voiced			5	4	7

Table 11: Tone paradigm of common Tai Yuan.

Initial consonant class at time of tonal splits	A	B	C	DL	DS
Aspirated	1				
Plain		3	5	3	1
Implosive	2				
Voiced		4	6	4	5

Table 12: Tone paradigm of common Central Thai.

Initial consonant class at time of tonal splits	A	B	C	DL	DS
Aspirated	5				
Plain	1	2	3		2
Implosive					
Voiced	(6)	3	4	3	4

Upon comparing the paradigms of common Tai Yuan and common Central Thai with those present in the two Northern Lao dialects, it is obvious that the Lao ladders have been decomposed within the two diaspora dialects under investigation. Another observation pertains to tone A, wherein the Northern Lao dialect speaker in Sangkhlaburi (LN67) seems to have adopted a splitting pattern akin to that of Central Thai. This adaptation results in a differentiation of tone A1 (originating from aspirated initials) from the other tones, A234, within the paradigm.

Furthermore, we assess the significance of language contact by considering the sociolinguistic context of the two Northern Lao dialect speakers within the diaspora: LN11 and LN67. Although Akharawatthanakun (2003) does not explicitly report the extent of contact intensity and the specific domains of usage for both dialects, it is highly likely that bilingualism has resulted in the reorganisation and convergence of homophonous patterns within the Northern Lao tone paradigm towards the dominant regional dialects in their respective locations. In the case of LN11, this particular Northern Lao dialect speaker has primarily resided in the border region between northwestern Laos and northern Thailand on the Thai side, where the dominant regional language is Tai Yuan (Akharawatthanakun 2003: 150, 450). Similarly, the living environment of the Northern Lao dialect speaker LN67 involves contact with Central Thai, the dominant regional language in western Thailand, as well as with other diasporic Lao dialects (Akharawatthanakun 2003: 150, 448).

Based on our tonological profile data (as shown in Tables 9 and 10), exposure to Tai Yuan and Central Thai speaking environments, respectively, emerges as a highly plausible factor contributing to the decomposition of the Lao ladders within tones C and DL. Additionally, the Central Thai tone paradigm also provides a model for the reorganisation of tone slots within tone A for the Northern Lao dialect speaker LN67 (as discussed above). At the same time, it appears that homogeneity within the common Lao tones B1234 (as highlighted in Tables 5 and 8) remains stable and resistant to contact-induced changes in the tone paradigm of these two speakers.

As a supplementary remark, we may also consider a language-internal perspective and posit a hypothesis that tones DL and DS paradigmatically and cognitively lie at a deeper level of prominence and speaker's awareness. This heightened prominence might stem from their vague status within the tone paradigm and the variations observed among individual dialect speakers, due to their probabilistic status as a low-frequency syllable type and a low-probable tone type in the language system (see e.g. the case of Chinese dialects in Wiener & Ito 2015). Consequently, these distinctive characteristics may make them more prone to

alternation and change when compared to the salient tones A, B, and C, which are often emphasised in the process of acquisition, more systematically taught in school and acquired by children from their elementary education. This scenario is usually observed in the teaching of standard languages, Lao and Thai, as well as Vietnamese, Cantonese and Mandarin (see e.g. Bar-Lev 1991). However, the method used in the current study is not specifically designed to test this particular claim. Nevertheless, this aspect can offer an interesting direction for future research which could bridge the domains of dialectology and cognitive sciences. Such exploration could provide valuable insights into the dynamics of tone paradigms and their evolution.

5 Conclusions

In the present study, we have examined and discussed instances of language shift occurring in various regions where Tai dialect speakers are shifting their language whose tone paradigm structure subsequently also converges with a dominant regional dialect. From the perspective of tone paradigm, the convergence has sometimes resulted in the emergence of transitional dialect systems, in which a protosystem has undergone restructuring, aligning itself more closely with the model provided by the dominant regional dialect. These transitional dialect systems of tone paradigm stand out when examined through a quantitative approach in combination with the conventional comparative method. It enables identification of the protostructure of tone paradigm from which the transitional dialect systems have diverged. In any case, our results do not post significant challenges to the genealogical classification proposed in the previous studies, as the signals of change observed in the present study primarily pertain to contact-induced changes occurring subsequent to the dispersal stages of individual Tai subbranches.

At a methodological level, we have also demonstrated the utility of Neighbor-Net algorithm as an effective tool for identifying such instances from a big pool of data. Our method employed to gather, organise and analyse the data can potentially offer a preliminary model for scholars engaged in the studies of Tai dialects as well as for those researching dialects of other Mainland Southeast Asian languages with tones. As more data, particularly relating to tone paradigms, have been continuously collected from field during the recent decades, this methodological model should also facilitate scholars and dialectological studies with a focus on tonal aspects in embracing an emerging trend within digital humanities and big data studies.

Lastly, our vital message to scholars engaged in language documentation and description is that the collection of sociolinguistic data stands on equal footing with the description of language features. From the present study, we see that insufficient description of sociolinguistic context regarding informants may pose a challenge when attempting to establish a contact-based explanation for a language change. With the inclusion of such comprehensive sociolinguistic information about the speakers, their speech communities and their linguistic repertoires, numerous finely-tuned analyses centred around cross-factor correlations can yield significantly more insightful understanding of the language situation. Such analyses will contribute to the discussion of language change and the diversification of dialects, a dynamic process which continues to evolve as we advance into the 21st century.

Abbreviations

Tones

- A Tone A, smooth syllable
- B Tone B, smooth syllable
- C Tone C, smooth syllable
- DL Tone D, long vowel, checked syllable
- DS Tone D, short vowel, checked syllable

Languages

BT	Black Tai	SH	Shan
KHR	Khorat Thai	SK	Saek
LC	Central Lao	THC	Central Thai
LN	Northern Lao	THS	Southern Thai
LS	Southern Lao	TY	Tai Yuan/Northern Thai
LUE	Tai Lue	WT	White Tai
LW	Western Lao/Northeastern Thai	YO	Yo
PH	Phuan	YOY	Yoy
PT	Phu Thai		

Acknowledgements

This research was supported by the Grant for Graduate Student 2021 from the King Prajadhipok and Queen Rambhai Barni Memorial Foundation, awarded to

Saknarin Pimvunkum for the research project “Classification of Vientiane Lao dialects in Western Thailand by tone paradigm” and was part of the project “North-east and Southeast Asian Studies Network in Finland and Thailand (NSEANET)” funded by the Finnish National Agency for Education (EDUFI) during 2020–2023. This study was exempted from ethical review by the Committee for Research Ethics (Social Sciences), Faculty of Social Sciences and Humanities, Mahidol University (2022/001.1701, 17 January 2022).

References

- Akharawatthanakun, Phinnarat. 2003. *Tone change: A case study of the Lao language*. Chulalongkorn University. (Doctoral dissertation).
- Akharawatthanakun, Phinnarat. 2020. Tonal diversity and tone Sandhi in Lue. *Journal of Liberal Arts, Thammasat University* 20(2). 513–548. DOI: 10.14456/lartstu.2020.31.
- Bar-Lev, Zev. 1991. Two innovations for teaching tones. *Journal of the Chinese Language Teachers Association* 26(3). 1–24. <https://eric.ed.gov/?id=EJ447341>.
- Brown, J. Marvin. 1985. *From ancient Thai to modern dialects*. Bangkok: White Lotus.
- Bryant, David & Vincent Moulton. 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265. DOI: 10.1093/molbev/msh018.
- Bunyasathit, Wannaporn, Chantas Pientam & Theptida Silapabanleng. 2016. The historical development of LaoWiang people in U Thong district, Suphanburi province. *Journal of Nakhonratchasima College* 11(1). 26–38. http://journal.nmc.ac.th/th/admin/Journal/2560Vol11No1_702.pdf.
- Burusphat, Somsonge. 2012. Tones of Thai Song varieties. *Journal of the Southeast Asian Linguistics Society* 5. 32–48. <http://hdl.handle.net/1885/9118>.
- Canilao, Kritsana. 2010. *Tonal geography of the provinces of Central Thailand*. Mahidol University. (Doctoral dissertation). <http://mulinet11.li.mahidol.ac.th/e-thesis/2553/cd446.1/4836363.pdf>.
- Chamberlain, James R. 1975. A new look at the history and classification of the Tai languages. In Jimmy G. Harris & James R. Chamberlain (eds.), *Studies in Tai linguistics in honor of William J. Gedney*, 49–66. Bangkok: Central Institute of English Language. <http://sealang.net/sala/archives/pdf8/chamberlain1975new.pdf>.

- Damanhuri, Umayah. 2004. The classification of some Thai dialects spoken in Kedah. In Somsonge Burusphat (ed.), *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*, 167–182. Tempe, AZ: Arizona State University Programme for Southeast Asian Studies Monograph Series Press. <http://sealang.net/sala/archives/pdf4/damanhuri2004classification.pdf>.
- Dockum, Rikker. 2019. *The Tonal Comparative Method: Tai tone in historical perspective*. Yale University. (Doctoral dissertation).
- Edmondson, Jerold A. 1990. Kam tone splits and the variation of breathiness. In Jerold A. Edmondson, Crawford Feagin & Peter Mühlhäusler (eds.), *Development and diversity: Language variation across time and space (a Festschrift for Charles-James N. Bailey)*, 187–202. Arlington, TX: Summer Institute of Linguistics. <https://www.sil.org/resources/archives/8425>.
- Edmondson, Jerold A. & David B. Solnit. 1997. Introduction. In Jerold A. Edmondson & David B. Solnit (eds.), *Comparative Kadai: The Tai branch*, 1–32. Arlington, TX: Summer Institute of Linguistics.
- Ferlus, Michel. 2004. The origin of tones in Viet-Muong. In Somsonge Burusphat (ed.), *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*, 297–313. Tempe, AZ: Arizona State University Programme for Southeast Asian Studies Monograph Series Press. <http://sealang.net/sala/archives/pdf8/ferlus2004origin.pdf>.
- Gedney, William J. 1972. A checklist for determining tones in Tai dialects. In M. Estellie Smith (ed.), *Studies in linguistics in honor of George L. Trager*, 423–437. The Hague: de Gruyter Mouton.
- Grünthal, Riho & Johanna Nichols. 2016. Transitivizing-detransitivizing typology and language family history. *Lingua Posnaniensis* 58(2). 11–31. DOI: 10.1515/linpo-2016-0008.
- Handel, Zev. 2014. Historical phonology of Chinese. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 576–598. Oxford: John Wiley & Sons. DOI: 10.1002/9781118584552.ch22.
- Hartmann, John F. 2008. The Lue language. In Anthony V. N. Diller, Jerold A. Edmondson & Yongxian Luo (eds.), *The Tai-Kadai languages*, 254–297. London: Routledge.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242. 69–82. https://lacito.hypotheses.org/files/2015/12/Haudricourt_1954_Origine-Tons-Vietnamien_scan.pdf.
- Hill, Nathan. 2019. *The historical phonology of Tibetan, Burmese, and Chinese*. Cambridge: Cambridge University Press.

- Hudak, Thomas John. 2008. *William J. Gedney's comparative Tai source book* (Oceanic Linguistics Special Publications 34). Honolulu, HI: University of Hawai'i Press. <https://www.jstor.org/stable/20532978>.
- Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267. DOI: 10.1093/molbev/msj030.
- Kingston, John. 2011. Tonogenesis. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2304–2333. Oxford: Wiley-Blackwell. DOI: 10.1002/9781444335262.wbctp0097.
- Koowatthanasiri, Kanjana. 1981. *The tones of Nyo*. Chulalongkorn University. (MA thesis). <http://cuir.car.chula.ac.th/handle/123456789/35801>.
- Li, Fang-Kuei. 1977. *A handbook of comparative Tai* (Oceanic Linguistics Special Publications 15). Manoa, HI: University Press of Hawai'i. <https://www.jstor.org/stable/20006684>.
- Liang, Min & Junru Zhang. 1996. *Dòngtái yǔ Gailùn [An introduction to the Kam-Tai languages]*. Beijing: China Social Sciences Publishing House.
- Liao, Hanbo. 2016. *Tonal development of Tai languages*. Payap University. (MA thesis). https://www.academia.edu/download/49907598/Final-TONAL_DEVELOPMENT_OF_TAI_LANGUAGES-0714.pdf.
- Liao, Hanbo. 2023a. An integrated tone box scheme for determining tones in Tai varieties beyond Southwestern Tai: Diachronic and synchronic concerns. *Folia Linguistica* 57(s44–s1). 199–244. DOI: 10.1515/flin-2022-2048.
- Liao, Hanbo. 2023b. *The formation of Lingnan linguistic traits: Typological structures and diachronic issues*. The University of Hong Kong. (Doctoral dissertation).
- Luo, Meizhen. 1988. Dǎi Tàï cíhuì bìjiào [A comparison of Dai and Thai vocabularies]. *Mínzú Yǔwén [Minority Languages of China]* (2). 26–34.
- Luo, Yongxian. 1997. *The subgroup structure of the Tai languages: A historical-comparative study* (Journal of Chinese Linguistics Monograph Series 12). Hong Kong: The Chinese University of Hong Kong Press. <https://www.jstor.org/stable/23887080>.
- Maddieson, Ian. 1984. The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics* 12(1). 9–15. DOI: 10.1016/S0095-4470(19)30845-9.
- Maddison, David R., David L. Swofford & Wayne P. Maddison. 1997. Nexus: An extensible file format for systematic information. *Systematic Biology* 46(4). 590–621. DOI: 10.1093/sysbio/46.4.590.

- Michaud, Alexis & Bonny Sands. 2020. Tonogenesis. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.748.
- Mitani, Yasuyuki. 1977. Tai-Kadai shogo no gengonendaigaku-teki kōsatsu [Linguistic chronology of Tai-Kadai languages]. *Tōnan'ajia kenkyū [Southeast Asian Studies]* 15(3). 421–429. <https://kyoto-seas.org/pdf/15/3/150309.pdf>.
- Nichols, Johanna. 2020. Dispersal patterns shape areal typology. In Mily Crevels & Pieter Muysken (eds.), *Language dispersal, diversification, and contact: A global perspective*, 25–43. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198723813.003.0002.
- Ostapirat, Weera. 2005. Kra-dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In Roger Blench, Laurent Sagart & Alicia Sanchez-Mazas (eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, 107–131. London: Routledge. DOI: 10.4324/9780203343685.
- Pittayaporn, Pittayawat. 2009. *The phonology of Proto-Tai*. Cornell University. (Doctoral dissertation). <https://ecommons.cornell.edu/bitstream/handle/1813/13855/Pittayaporn,%20Pittayawat.pdf>.
- Pittayaporn, Pittayawat. 2014. Layer of Chinese loanwords in Proto-Southwestern Tai as evidence for the dating of the spread of Southwestern Tai. *Manusya: Journal of Humanities, Special Issue No 20* 17(3). 47–68. DOI: 10.1163/26659077-01703004.
- Piyabhan, Bung-on. 1998. *Laaw nai krung rāttanákoosin* [Lao people in the Rattanakosin kingdom]. Bangkok: The Thailand Research Fund.
- Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics. DOI: 10.15144/PL-613.
- Ratliff, Martha. 2015. Tonoexodus, tonogenesis, and tone change. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 245–261. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199232819.013.021.
- Srithonrat, Potjanee. 1983. *A tonal comparison of Phuthai dialect in three provinces*. Mahidol University. (MA thesis).
- Szeto, Pui Yiu, Umberto Ansaldi & Stephen Matthews. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22(2). 233–275. DOI: 10.1515/lingty-2018-0009.
- Szeto, Pui Yiu & Chingduang Yurayong. 2021. Sinitic as a typological sandwich: Revisiting the notions of altaicization and taicization. *Linguistic Typology* 25(3). 551–599. DOI: 10.1515/lingty-2021-2074.

- Thurgood, Graham. 2002. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica* 19(2). 333–363. DOI: 10.1075/dia.19.2.04thu.
- Wiener, Seth & Kiwako Ito. 2015. Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience* 30(9). 1048–1060. DOI: 10.3798/2014.946934.
- Wulff, Kurt. 1934. *Chinesisch und Tai: Sprachvergleichende Untersuchungen*. Copenhagen: Levin & Munksgaard.
- Yang, Tongyin & Jerold A. Edmondson. 2008. Kam. In Anthony V. N. Diller, Jerold A. Edmondson & Yongxian Luo (eds.), *The Tai-Kadai languages*, 509–584. London: Routledge.
- Yurayong, Chingduang & Pui Yiu Szeto. 2020. Altaicization and de-altaicization of Japonic and Koreanic. *International Journal of Eurasian Linguistics* 2(1). 108–148. DOI: 10.8833-12340026.
- Zhang, Junru. 1980. Yuánshǐ tái yǔ shēngmǔ lèibíé tànsuǒ [An exploration of the classification of initial consonants of proto-Tai]. *Mínzú Yǔwén [Minority Languages of China]* (2). 31–40.
- Zhu, Xiaonong, Mingying Wei & Junfang Wang. 2016. Shíwǔ tiáohé qìtiáo: Dòngyǔ róngjiāngxiàn kǒuzhài fāngyán ànlì [Fifteen tones and breathy tones: A case of the Kouzhai Dong dialect of Rongjiang county]. *Mínzú Yǔwén [Minority Languages of China]* 5. 12–24.