JOURNAL of the SOUTHEAST ASIAN LINGUISTICS SOCIETY

Volume 4.2
December 2011

JSEALS
Journal of the Southeast Asian Linguistics Society

JSEALS is the peer-reviewed journal of the Southeast Asian Linguistics Society, and is devoted to publishing research on the languages of mainland and insular Southeast Asia. It is an electronic journal, distributed freely by Pacific Linguistics (www.pacling.com) and the JSEALS website (jseals.org).

JSEALS was formally established by decision of the SEALS 17 meeting, held at the University of Maryland in September 2007. It supersedes the Conference Proceedings, previously published by Arizona State University and later by Pacific Linguistics.

JSEALS welcomes articles that are topical, focused on linguistic (as opposed to cultural or anthropological) issues, and which further the lively debate that characterizes the annual SEALS conferences. Although we expect in practice that most JSEALS articles will have been presented and discussed at the SEALS conference, submission is open to all regardless of their participation in SEALS meetings. Papers are expected to be written in English.

Each paper is reviewed by at least two scholars, usually a member of the Advisory Board and one or more independent readers. Reviewers are volunteers, and we are grateful for their assistance in ensuring the quality of this publication. As an additional service we also admit data papers, reports and notes, subject to an internal review process.

JSEALS is published twice a year. Papers can be submitted to the Managing Editor, electronically (paul.sidwell@anu.edu.au or paulsidwell@yahoo.com) at any time.

Contents

SEALS News

The 21st Southeast Asian Linguistics Society annual meeting was successfully conducted at Kasetsart University, Bangkok, May 11-13. Many thanks to Aj. Kitima Indrambarya and her staff for doing a great job. One important development was that the Business Meeting resolved to begin the process of formalizing the structure and membership of the Society, which until now has been merely a name attached to the annual meetings. An Executive, consisting of Aj. Kitima Indrambarya, Mathias Jenny and Paul Sidwell, was elected and charged with drafting a Statement of Purpose, which individual scholars will be invited to endorse as a condition of being listed in the new membership registry. This marks an important first step in creating an independent professional body that can represent the community of scholars concerned with Southeast Asian Languages and Linguistics.

The **Statement of Purpose** (also now online at jseals.org) is as follows:

1) The Southeast Asian Linguistics Society (SEALS, also the Society) is formed for the purpose of facilitating and promoting contact and communication among scholars and students of Southeast Asian Linguistics, and for the dissemination of their scholarly works.

2) The members of the Society advocate the documentation, study, analysis, teaching and maintenance of Southeast Asian Languages.

3) To these ends, the Society undertakes:
   a) to hold international meetings, normally annually, and in a manner that provides reasonable opportunity for scholars and students from Southeast Asia to participate
   b) to publish a journal, and such other publications and communications as deemed appropriate, in order to provide opportunity for the presentation of scholarly research and documentation on Southeast Asian Languages
   c) to maintain a website as a point of contact and information
   d) to maintain appropriate academic standards in meetings and publications, principally by means of peer review of papers and abstracts

4) The Society accepts as members those persons who, in good faith, make a declaration of endorsement of this statement of purpose.

In order to endorse the Statement of Purpose, please send an email to me <paulsidwell@yahoo.com> and your details will be entered into the registry.

Next SEALS Meeting

The business meeting also accepted the proposal to hold the 22nd meeting at Agay in France in 2012. Deth Thach (INaLCO, Paris) presented a detailed proposal for the meeting to be held in this beautiful seaside resort town between May 30 and June 2. Supporting institutions include:

- Institut des Sciences Humaines et Sociales
- Structure et Dynamique des Langues
- Institut Nationale des Langues et Civilisations Orientales
- Langues et Civilisation à Tradition Orale
- Centre Asie du Sud-Est

- Institut de Recherche sur le Sud-Est Asiatique/Maison Asie-Pacifique - Université de Provence
- Institut de Recherche pour le Développement

The deadline for abstracts if February 1st 2012, please check the website (jseals.org) for more information.

SEALS Archives

The business meeting also resolved to create an online archive of documentation and handouts from previous SEALS meetings, and the enthusiastic Elisabeth Ginsburg was elected to the position of SEALS Archivist. Elisabeth has already been busily attending to her new role, and extensive documentation is now available under **Conference Archives** at jseals.org.

Indexing of abstracts

Another item that will be keenly welcomed by contributors to this journal, important strides have been made towards improving the profile and discoverability of JEALS papers. Both *Linguistics & Language Behavior Abstracts* (LLBA) and *EBSCOhost™* databases will now be indexing JSEALS, with the latter providing full text access to subscribing institutions. This will begin in 2012, and will include all issues since we began as a journal 4 years ago. This is a great step forward for discovery and access for JSEALS that will benefit everyone.

SEALS book reviews

JSEALS has begun receiving unsolicited books for review. Initially I have taken the approach of advertising these on the SEALS facebook page, and this has successfully resulted in drumming up offers to review. We will trial this for a reasonable time and discuss it at subsequent business meetings.

I would expect reviewers to be notional members of SEALS, and for reviews to be submitted within 3 months, at minimum 1000 words for regular books, 2000 words for large books, no upper limit. Reviews are be expected to be more than long book notices, but offer some critical insight. Unsolicited ferviews are also welcome, as long as the books relate to our area of interest.

JSEALS Forum

Finally, this issue experiments with a new section called *JSEALS Forum*, in which scholars are given a chance to freely discuss and challenge each other over a particular topic. The idea arose after Anne Daladier of CNRS (Paris) privately challenged my presentation concerning the history Khasian languages given at this year's SEALS meeting in Bangkok. After negotiation, we agreed to draft papers, then exchange drafts and add or incorporate critiques of each other's papers. These are not externally peer reviewed, only internally reviewed, the purpose being to facilitate timely robust debate and discussion with minimal restrictions. We hope that this will stimulate others to take advantage of the opportunity to do something similar in future, and discussions on any topic within the scope of JSEALS are invited.

Paul Sidwell
Managing Editor
December 2011

SOUTHERN SUI: A FOURTH SUI DIALECT[1]

Andy Castro
Guizhou University and SIL International
<andy_castro@sil.org>

0 Abstract

Previous research claims that Sui can be divided into three dialect areas and that intelligibility both between and within these dialects is high (SDB 1956; Zhang 1980; Zeng 2004; Stanford 2007). This paper presents a historical comparison of previously published Sui dialect data with new data collected by the author, using Thurgood's (1988) Proto-Kam-Sui (PKS), Zeng's (1994) reconstruction of Proto-Sui initials, and data from Kam varieties to track phonemic innovations. Shared phonological innovations, in addition to lexical similarity counts, indicate that part of the area formerly classified as "Sandong dialect" should be considered a distinct, fourth dialect area, which the author calls "Southern Sui". Anecdotal low intelligibility between Southern Sui and the Sandong dialect spoken further north supports this proposal. Interestingly, and surely not coincidentally, the geographical area covered by Southern Sui largely corresponds to the homeland of a subset of Sui people who celebrate the "*Maox*" festival instead of the "*Dwac*" festival celebrated by almost all other Sui.

Key words: Classification, dialectology, Kam-Sui

1 Introduction

The Sui live in southeastern Guizhou province, China, concentrated in Sandu Sui Autonomous County and its locale. Sui is typically classified as belonging to the Kam-Sui branch of the Tai-Kadai language family (Diller 2008; Lewis 2009). Its closest relative is Maonan. Other languages in the Kam-Sui branch include Kam, Mulam, Mak and Then. Sui occupies an important position in Tai-Kadai comparative research due to its relatively conservative nature, particularly in terms of its rich inventory of sounds, many of which have been lost in other Kai-Tadai languages (Wei & Edmondson 2008).

Zhang (1980) and others (for example SDB 1956, Zeng 2004, Wei & Edmondson 2008, Lewis 2009) divide Sui into three dialects (sometimes referred to as "vernaculars"), as shown in Figure 1. This division is primarily based upon phonological and lexical similarity.

Figure 1: *Traditional grouping of Sui dialects dialects based on dialect locations given in Zhang 1980, showing Sandu and Libo county towns.*
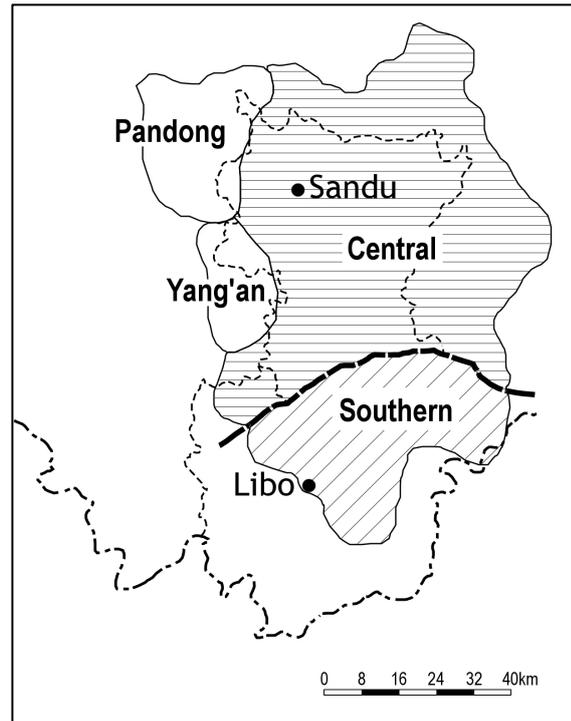
Figure 2: *Proposed grouping of Sui dialects based on the present study.*

Sandong dialect has the most number of speakers and is considered the "standard dialect". It has a rich inventory of over 60 initials (including bilabial, alveolar[2], palatal, velar and uvular consonants and preglottalised, palatalised and labialised obstruents) and over 50 finals (including nasal and -p, -t, -k codas). The other two dialects, Pandong and Yang'an, have smaller phonemic inventories. They lack preglottalised initials, have fewer voiceless nasals (Pandong has none at all) and fewer prenasalised stops. Sui has six contrastive tones on open syllables and two on closed syllables with different pitch contours depending on the length of the vowel nucleus. Typical tone values for the Sandong dialect are given in Table 1.

In his most recent Sui dialect study, Stanford examines phonetic features in the speech of 33 Sui speakers from 17 locations across the Sui region. He concludes by saying that his results "confirm a stable three-way distinction of major Sui dialects, as found in prior literature" (Stanford in prep.). Of course there are also numerous small phonetic differences in pronunciation across the Sui area, some regional, some clan-based and many individual-based (Stanford 2007, 2008, 2009). Until now, most scholars have considered differences between and within the Sui dialects to be minor and claim that all three dialects

---

[2] The "alveolar" series referred to in this paper is equivalent to Li Fang-kuei's (1965) "dental" series. Edmondson et. al. (2004:51) analysed these sounds as "denti-alveolars" because he found that there is a wide post-dental contact area. The author has noticed that for some Sui speakers there is no dental contact at all. Some speakers recorded for this study even articulate prenasalised "alveolar" stops with the tongue slightly retroflexed, touching the back of the alveolar ridge.

are mutually intelligible (Zhang 1980:75; Weng 2001:563; Zeng 2004:42; Edmondson et al. 2004:48; Stanford 2007:19).

Table 1: *Typical Sui tone values (Sandong dialect, Wei & Edmondson 2003:xxviii).*[3]

Syllable type	Tone number	Tone value
open (unchecked) (vowel and nasal codas)	1	11
	2	31
	3	33
	4	51
	5	35
	6	24
closed (checked) (-p, -t, -k codas)	7	35 (long), 55 (short)
	8	42 (long), 31 (short)

However, anecdotal reports of low intelligibility between some speakers within the traditional Sandong dialect area challenge this picture. Zhang (2008:52) writes of the Sui language spoken in Shuiyao, Libo county (Southern Sui area), that "during our survey we found the situation to be a little different [from that described in previous literature]. The Sui in Shuiyao consider their own language to be clearly different from Sandong Sui, of which they say they can only understand 60%-70%. They often cite their language differences as evidence to show that they are a separate community from Sandong Sui." Zhang then quotes a retired first school teacher from Shuiyao who said, "Sui people from Shuiyao who have never been to Sandong before find Sandong Sui extremely difficult to understand."

An examination of diachronic sound changes among Sandong speech varieties and a comparison of lexical similarity percentages show that varieties spoken in the south of the Sandong dialect area (mainly in Libo county but also crossing over into southeastern Sandu county) form a distinct dialect cluster of their own which the author labels "Southern Sui".[4] Sandong speech varieties to the north of this area, spoken by the majority of Sandong dialect speakers, are collectively referred to as "Central Sui". This new grouping is shown in Figure 2. Southern Sui appears to have undergone a unique set of shared phonological innovations and displays high internal phonetic and phonemic consistency. The bulk of this paper is devoted to tracking these innovations, using Thurgood's (1988) Proto-Kam-Sui (PKS), Zeng's (1994) reconstruction of Proto-Sui (PS) initials and data from other Kam-Sui varieties as references.

---

[3] Tone values are transcribed using Chao's (1930) pitch scale of 1 to 5, with 5 indicating highest pitch and 1 indicating lowest pitch. The same pitch number repeated indicates a level tone and different pitch numbers in succession indicate rising or falling contour.

[4] SDB (1956) includes data from three locations in the Southern Sui region: Jiuqian, Jiarong and Shuiqing. Some of these data support the notion of a southern region. Maps 5 and 6, showing isoglosses for a x-/kʰ- alternation in the word 'diligent' and a ʔd-/l- alternation in the word 'boat', hint at a southern dialect. However, the authors did not take the step of proposing a fourth dialect, perhaps because their data were limited.

Lack of data has thus far held back reconstructions in the Kam-Sui branch. Thurgood's (1988) and Peiros' (1998)[5] reconstructions of Proto-Kam-Sui refer to data from only two Sui dialects and one Kam dialect, despite the internal diversity of both of these languages. Zeng (1994, 2004) has reconstructed Proto-Sui initials based on data from five locations; one Central Sui, two Southern Sui, one Pandong and one Yang'an dialect areas. The data presented in this paper highlight the limitations and occasional inconsistencies of both Thurgood's and Zeng's reconstructions. A new reconstruction of Proto-Kam-Sui should perhaps only be attempted once thorough treatments of Proto-Sui and Proto-Kam have been made. Due to the limitations of Zeng's reconstructed PS initials (which will become apparent as this paper progresses), Thurgood's PKS is used as the primary reference point for demonstrating the phonological divergence of Central and Southern Sui.

2 Cultural distinctiveness of the Southern Sui area

Interestingly, the geographical area covered by Southern Sui corresponds to the homeland of a subset of the Sui people who celebrate the *Maox*[6] festival instead of the *Dwac* festival that is celebrated by almost all other Sui. *Maox* festival takes place during the fifth and sixth months of the lunar calendar (June and July in the solar calendar). During *Maox*, young, unmarried Sui gather on hillsides known as *ggumx Maox*, or "*Maox* slopes", and sing traditional songs to one another in an act of courtship. The *Dwac* festival takes places during the eighth and ninth months of the lunar calendar (September and October in the solar calendar) just after harvesting the rice. Horse racing is one of the main activities, taking place on *ggumx Dwac*, or "*Dwac* slopes" (Sandu County Education Bureau et al. 2007).

Dwac is celebrated in the Central, Pandong and Yang'an dialect areas by most Sui, although there are a number of Sui villages which celebrate Chinese New Year or other festivals in lieu of *Dwac*. *Maox* is celebrated exclusively in the Southern Sui area (Zhang 2008:282)[7]. The correlation between festival locations and the Central and Southern dialect areas is shown in Figure 3. *Ggumx Dwac* are indicated by numerals showing the order in which the festivals are celebrated (1 is first, at the beginning of September; 7 is last, towards the end of October). *Ggumx Maox* are indicated by letters, again showing the order of celebration (A is first, D is last).

---

[5] Peiros (1998:31) admits that his reconstruction is "remarkably similar" to Thurgood's; this is not surprising given that most of their sources appear to be identical.

[6] Sui words are written in italics. The author employs a Latin-based orthography originally created in the 1950s and recently revised by the Sui Research Institute, Sandu county. Tone is denoted by a word-final consonant: *-l* = tone 1; *-z* = tone 2; *-c* = tone 3; *-x* = tone 4; *-s* = tone 5, or tone 7 on checked syllables; and *-h* = tone 6. Tone 8 on checked syllables is unmarked.

[7] There are some villages in the Southern Sui area which do not celebrate *Maox*. For example, the home village of the Jiuqian informant celebrates Chinese New Year in lieu of *Maox*. Such villages are rare.
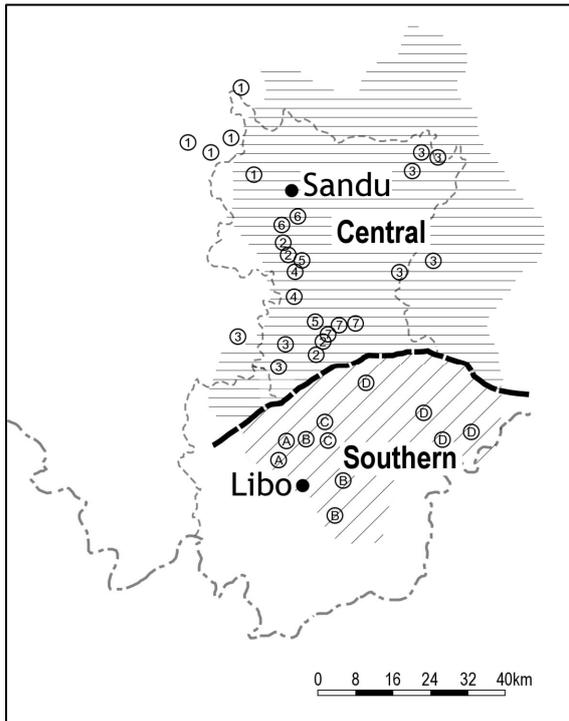
Figure 3: *Festival locations (source of locations: Weng 2001:613-4).*
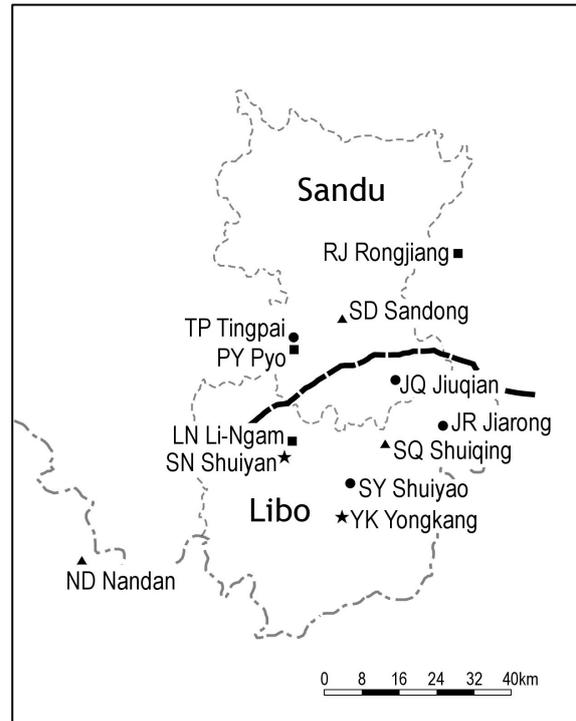
Figure 4. *Locations of data consulted for this study: ● = author's own data, ■ = Li (1965), ★ = Zeng (2004), ▲ = other sources, see Table 2.*

3 Data sources

This study makes use of a variety of data from old and new sources. These data are supplemented by the author's own field notes from three hitherto undocumented Southern Sui locations and one Central Sui location. Data are referred to by the abbreviations listed in Table 2. Further information about the data sources is given in Appendix A.

The Rongjiang (RJ) data proved almost identical to the data from Sandong (SD). Similarly, data from Pyo (PY) largely agreed with the Tingpai (TP) data collected by the author. RJ and PY are therefore rarely cited in the data tables. The Li-Ngam (LN) data is only referred to when different from Zeng's (2004) more recent Shuiyan (SN) data. The fact that Sui speech appears to have changed so little over such a long period of time in these places[8] supports Stanford's (2009) observations of the maintenance of phonetic speech features over time as a significant act of clan loyalty in Sui culture and his 50-year real-time comparison indicating stable preservation of Sui dialectal features over a period of more than 50 years (Stanford, in prep.).

---

[8] Li's (1965) data were collected in the early 1940s whereas the SD, TP and SN data were all collected much more recently. See Appendix A for more details. Li (1965) does not specify the exact source location of his Rongjiang data. Rongjiang is included as part of the Sandong dialect area by Zhang (1980), and the author's personal observations of a Sui speaker from the eastern fringe of the Sui area in Rongjiang indicates that their speech is close to "standard" Sandong Sui.

Table 2: *Sources of data used for this study.*

Ref	Location[9]	Sui toponym/s[10]	Sui dialect area	Source of data
LN	Shuili and Shuiyan, Shuili district, Libo county ("Li-Ngam")	li^6 & ŋɐm^2	Southern	Li (1965)
PY	Tingpai district, Sandu county ("Pyo")	pjo^2	Central	Li (1965)
RJ	Rongjiang county ("Jung-chiang")	ɬʰɛːk^7	Central	Li (1965)
SN	Shuiyan, Shuili district, Libo county	ŋɐm^2	Southern	Zeng (2004)
YK	Yongkang district, Libo county	la^{6}qɐŋ^4	Southern	Zeng (2004)
SD	Sandong district, Sandu county	tɔːŋ^6	Central	Zhang (1980)
SQ	Shuiqing, Maolan district, Libo county	la^{6}kʰən^1	Southern	ILCRD et al. (1996)
ND	Liuzhai district, Nandan county	(unknown)	Central	GZARMLC (2008)
JQ	Jiuqian district, Sandu county	muːi^6	Southern	author's field notes
SY	Shuiyao district, Libo county	la^{6}ʔjaːŋ^1	Southern	author's field notes
JR	Jiarong district, Libo county	ljɐŋ^1	Southern	author's field notes
TP	Tingpai district, Sandu county	pjo^2	Central	author's field notes

Various Kam data are also cited throughout the paper. All Kam words given in the data tables are taken from GZARMLC (2008). They were collected from Sanjiang county, which is at the southernmost end of the Southern Kam language area, about 160km east of Sandu county seat. All other Kam data referred to were collected between 2002 and 2004 by researchers from Guizhou University (led by Professor Shi Lin 石林) who visited 17 data points in Guizhou, Hunan and Guangxi provinces, covering most of the Kam area. These data are unpublished to date.

Early Mandarin (EM), Late Middle Chinese (LMC) and Early Middle Chinese (EMC) forms are from Pulleyblank (1991). Middle Chinese (MC) forms are from Baxter & Sagart (2011). Old Chinese (OC) forms are from SGYCX (2011) unless stated otherwise. This source gives OC forms proposed by Baxter (白一平), Karlgren (高本汉), Zhengzhang Shangfang (郑张尚芳) and Pan Wuyun (潘悟云).[11]

4 Phonological divergence

The vast majority of phonological divergence between Central and Southern Sui occurs in the onsets. In many cases, different developments of PKS sounds have resulted in different phonemic mergers within the two dialect clusters. In total, no fewer than eight divergent phonemic mergers are observed. In one instance, a phoneme in PKS has been lost in Central Sui but retained in Southern Sui, albeit in a different form (the velar fricative, x-). It is no wonder, then, that there are reports of comprehension difficulties between Central and Southern Sui speakers.

In this section, the diachronic innovations which most characterise Southern Sui are examined and described. Further innovations have taken place within both Central and

---

[9] Li's (1965) spelling of place names are given in quotation marks. In Sui orthography, "Li-Ngam" would be written *Lih-Ngamz* and "Pyo" would be written *Byoz*. "Jung-chiang" is an older romanised form of Chinese pinyin *Róngjiāng* (榕江).

[10] "Sui toponym" refers to the region in which the data point is located rather than to the specific village. A superscript zero refers to a "neutral" tone on a reduced syllable.

[11] This data is also provided at Thesaurus Linguae Sericae (Harbsmeier ed., 2011).

Southern Sui, often just in one or two locations. Some of these are described in Appendix B. A detailed description of all the phonemic innovations in individual lects within Sui would require a much more comprehensive set of data than is currently available and would make an entirely separate study.

4.1 PKS labialised velars

PKS *gw-, *kw- are generally realised as p- in Central Sui and q- in Southern Sui. Correspondences are given in Table 3. One variety of Northern Kam (Baojing district, Zhenyuan county) exhibits a q- reflex in common with Southern Sui. Other varieties of Kam (specifically on the northern edge of Northern Kam in Zhenyuan and Xinhuang counties, and to the very south of Southern Kam in Rongshui county, Guangxi) realise this sound as a k-, as does Mak (Thurgood 1988:194).

Table 3: *PKS *gw-, *kw- correspondences, words with initial q- shaded in grey.*

Gloss	PKS	PS	Central Sui			Southern Sui				Kam
			ND	SD	TP	JR	JQ	SY	SN	
'to sharpen'*	*gwan²	-	pan²	pan²	-	pɐn²	pɐn²	pɐn²	pan²	pan²
'dove'	*gwau²	*pq-	-	pau²	pʌu²	qʌu²	qʌu²	qɔu²	qau²	pau²
'horn'	*m-kwa:u¹	*pq-	pa:u¹	pa:u¹	pa:u¹	qa:u¹	qa:u¹	qa:u¹	qa:u¹	pa:u¹
'leg'	*kwa¹	*pq-	pa¹	pa¹	pa¹	qa¹	qa¹	qa¹	qa¹	pa¹
'to sell'	*kwe¹	*pq-	pe¹	pe¹	pɛ¹	qɛ¹	qe¹	qe¹	qe¹	pe¹

* The Mulam for 'to sharpen', kwan², leads Thurgood to his PKS reconstruction. Data for Southern Sui indicates a different Proto-Sui onset.

Li Fang-kwei (1965:156) suggests a series of labialised post-velars in Proto-Kam-Sui to explain this p- and q- alternation. Zeng (1994:26-28) proposes *pq- or *pk- for Proto-Sui (or possibly a pre-syllable *p-, thus *p-k- or *p-q-), citing ancient Chinese rhyming books which indicate that a similar sound could have existed in Old Chinese, for example鬢 'hair on the temples', 綆 'well-rope', 鵑 'hoopoe', 駁 'piebald horse' and皀 'kernel'. Indeed, Zhengzhang Shangfeng 郑张尚芳 and Pan Wuyun潘悟云 reconstruct Old Chinese *p-q- and *pk- respectively for the latter two characters.[12]

In a later publication, Zeng (2004:53) revises her Proto-Sui reconstruction for these words to *q-, claiming that there is little evidence for an earlier *p- pre-syllable because cognates for these words exhibit a single initial k- in almost all Tai languages. She says that the k-/p- alternation in Kam and Sui dialects is the result of a sporadic sound change which occasionally occurred in Sino-Tibetan languages when there was lip rounding on the rhyme (thus offering a similar solution to Thurgood's PKS *kw-). In disagreement with Zeng (2004), Pittayaporn (2009) proposes the pre-syllable *p- in his recent reconstruction of Proto-Tai, for example Proto-Tai *p.qa:^A 'leg' and *p.qa:j^D 'to sell'. The p-/q- correspondence between Central and Southern Sui appears to support Zeng's (1994) original reconstruction of *p-q-.

---

[12] Zhengzhang reconstructs Old Chinese for駁 'piebald horse' as *p-qreewg, Pan as *pkreewg. Zhengzhang reconstructs皀 'kernel' as *p-qrɯɡ, Pan as *pkrɯɡ (SGYCX 2011).

The distribution of these reflexes in Sui shows a clear division between Southern and Central varieties. These reflexes also constitute different mergers: in Southern Sui, PKS *gw-, *kw- merges with PKS *p-; in Central Sui with PKS *k-, illustrated in Table 4.

Table 4: *Divergent mergers involving PKS *k-, *kw-, *p- in Central and Southern Sui (mergers shown in grey).*

PKS[13]	PS	Central Sui		Southern Sui	
*k-	*q-, *ɢ-	q-	qa¹ 'crow'	q-	qa¹ 'crow'
*gw-, *kw-, *m-kw-	*pq-	p-	pa¹ 'leg'	q-	qa¹ 'leg'
*p-	*p-, *b-	p-	pa³ 'aunt'	p-	pa³ 'aunt'

The labialised velar fricative PKS *xw- has usually become f- in Central Sui and w- in Southern Sui through processes of fusion and lenition respectively. Its aspirated stop equivalent PKS *khw- appears to have become f- through a similar process in all Sui varieties, merging with *xw- in Central Sui, although the data are too few to draw any definitive deductions.[14] Correspondences are given in Table 5.

Thurgood's PKS *khw-, *xw-, *ɣw- and *hw- are all reconstructed as *pj- or *bj- in Proto-Sui by Zeng (1994:18-20). The data presented in Table 5 indicate PS *f- for words such as 'sweet', 'wide' and 'cotton'. Zeng's only evidence for not proposing PS *f- seems to be the fact that most cases of initial f- in Mandarin have developed from bilabial stop initials and complex finals in Middle and Old Chinese.[15]

Our data show that Southern Sui generally retains a distinction between initial f- and w- in these words which has been lost in Central Sui. In Central Sui, the development of PKS *xw- constitutes a merger with PKS *hw- and *ɣw-, all becoming f-. In Southern Sui (with the exception of Jiarong), PKS *xw- has usually merged with PKS *pw-, becoming w-.[16] This pattern is illustrated in Table 6.

---

[13] Thurgood does not reconstruct voiced equivalents for *kr-, *k- or *p- (all of which would bear even-numbered tones after the voiced onset tone split), with the exception of *gr- in the word 'to kneel' PKS *gruk⁸ > t̥ok⁸ (SD), cok⁸ (LN). However, Zeng (1994) cites several examples in modern Sui for each of PS *ɢ-, *g- and *b-, none of which appear in Thurgood's data. Oddly, she does not suggest a voiced equivalent of PS *pq- to explain the even-numbered tone on qau² 'dove'.

[14] The words 'kerchief', 'wide' and 'cotton' seem to be good candidates for PKS *khw-, especially given their consistent realisations in Southern Sui as f- and in Kam and other Kam-Sui languages as khw-, f- or v- (CNU 1985).

[15] For example, 发 'to send out' fa¹ (Mandarin) < Middle Chinese (MC) *pjot^D and 凡 'all, ordinary' fan² (Mandarin) < MC *bjom^A. Pulleyblank (1991) proposes *pu- and *bu- in Early Middle Chinese for words such as these. For Late Middle Chinese, Pulleyblank (1991) reconstructs *f- for many of these words.

[16] If Zeng's reconstruction for Proto-Sui is correct, the initial w- in Southern Sui in these words would constitute a partial merger of PS *pj- and *ʔp-, but the conditioning environment which led to this merger is unclear.

Table 5: *PKS *khw- and *xw- correspondences, words bearing voiced fricatives or approximants shaded in grey.*

Gloss	PKS	PS	Central Sui				Southern Sui				Kam[17]
			ND	SD	TP	PY	JR	JQ	SY	SN	
'delayed'	*khwe¹	-	-	fe¹	ve¹	we¹	(tjɐŋ¹)[18]	fe¹	fɛ¹	fe¹	we¹'
'sweet'	*khwaːn¹	*pj-	faːn¹	faːn¹	faːn¹	fan¹	faːn¹	faːn¹	faːn¹	hwaːn¹	kʰwaːn¹
'tired'*	*khwe³	-	-	fe³	fe³	-	ʋe³	we³	wɛ³	we³	-
'cloud'	*m-xwa³	*pj-	fa³	fa³	fa³	fa³	fa³	wa³	wa³	wa³	ma³
'rain'	*xwin¹	*pj-	fən¹	fən¹	fən¹	fən¹	fən¹	wən¹	wən¹	wən¹	pjən¹
'bamboo'	*xwan¹	*pj-	fan¹	fan¹	fɐn¹	-	fɐn¹	wɐn¹	(kwi¹)	wan¹	pan¹
'pod'†	*pwak⁷	-	-	fak⁷	-	fɐk⁷	fɐk⁷	wɐk⁷	wɐk⁷	wɐk⁷	-
'elder brother'	*ɣwaːi⁴	*bj-	faːi⁴	faːi⁴	faːi⁴	fai⁴	faːi⁴	faːi⁴	faːi⁴	faːi⁴	ȶaːi⁴
'right'	*hwa¹	*pj-	fa¹	fa¹	fa¹	fa¹	hua¹	fa¹	hwa¹	-	wa¹'
'kerchief'	-	*pj-	-	fja³	-	-	fa³	fja³	fa³	-	-
'wide'	-	*pj-	faːŋ³	faːŋ³	-	-	-	-	-	faːŋ³	khwaːŋ³
'cotton'	-	*pj-	faːi³	faːi⁵	-	fai⁵	-	faːi⁵	-	faːi⁵	-

* Thurgood's reconstruction is based solely on Central Sui and Mulam data, both of which pronounce this word [fe³]. The data presented here suggests a reconstruction of PKS *xwe³.

† Thurgood's reconstruction is based solely on Southern Sui [wak□] and Mak [vak□] data. The data presented here suggests PKS *xwak□.

Table 6: *Divergent mergers involving PKS *hw-, *xw- and *pw- in Central and Southern Sui (mergers shown in grey).*

PKS	PS	Central Sui		Southern Sui	
*ɣw-, *hw-, *khw-	*pj-, *bj-	f-	fa¹ 'right'	f-	fa¹ 'right'
*m-xw-, *xw-	*pj-	f-	fa³ 'cloud'	w-	wa³ 'cloud'
*pw-	*ʔp-	v-	va⁵ 'wing'	w-	wa⁵ 'wing'

4.2 PKS bilabial-velar clusters

PKS bilabial-velar clusters *pw- and *phw- have usually become voiced labio-dental fricatives in Sandong Sui through a process of lenition, often further weakening to w- in Southern Sui, although not always. Correspondences are given in Table 7. In Jiuqian, these initials are more often than not pronounced as [ʋ], sometimes becoming [v] for emphasis, sometimes weakening to [w] in quick speech. The author suspects that the exact realisation

---

[17] In many Kam varieties, tones 1, 3, 5 and 7 have undergone a subsequent tone split. The resulting tones are marked in this paper with a prime mark after the tone number (1', 3', 5', 7'), following the convention of Kam linguists (Edmondson & Solnit 1990:8-9; Shi 1997:156-166).

[18] [tjɐŋ¹] means 'a long time' in most Sui dialects. In Jiarong it is also used to mean 'late, delayed' instead of the form [we¹] or [fe¹] which is more common in other dialects.

of /v/ probably varies greatly from speaker to speaker, so this alternation does not provide strong evidence for proposing a Central/Southern Sui dialect division.

These initials tend to have lost the labio-velar release and become p- or ph- in most varieties of Kam, although the Northern Kam variety spoken in Xinhuang county consistently realises them as w- (deletion of initial bilabial stop), like Southern Sui.

Table 7: *PKS *pw- and *phw- correspondences, words with labio-dental fricatives shaded in grey.*

Gloss	PKS	PS	Central Sui				Southern Sui				Kam
			ND	SD	TP	PY	JR	JQ	SY	SN	
'fire'	*pwai¹	*ʔp-	vi¹	vi¹	vi¹	vui¹	ȵi¹	ȵi¹	vi¹	wi¹	wi¹
'wing'	*pwa⁵	-	va⁵	va⁵	va⁵	va⁵	wa⁵	ʋa⁵	wa⁵	wa⁵	pa⁵
'seed'	*pwan¹	*ʔp-	van¹	van¹	vɐn¹	wan¹	wan¹	ʋɐn¹	wɐn¹	wan¹	-
'leaf'	*pwa⁵	*ʔp-	va⁵	va⁵	va⁵	va⁵	wa⁵	ʋa⁵	wa⁵	wa⁵	pa⁵
'dream'	*pwjan¹	*ʔp-	jan¹	vjan¹	-	-	vjen¹	vjen¹	vjen¹	wjaːn¹	pjan¹
'to fly'*	-	*ʔp-	vjən³	vjən³	vjɛn³	vjen³	vjin³	vjin³	vjɪn³	win³	pʰən³
'palm (of hand)'	*phwa³	-	fa³	-	-	-	(faːn³)¹⁹	ʋa³	wa³	wa³	pa⁵
'ashes'	*phwuːk⁷	-	vuk⁷	vuk⁷	vʊk⁷	-	ɣʊk⁷	ʋʊk⁷	ɣʊk⁷	wuk⁷	pʰuk⁷
'tall, high'	-	*ʔp-	vjaːŋ¹	vaːŋ¹	vɰɐŋ¹	wuəŋ¹	waːŋ¹	ʋaːŋ¹	waːŋ¹	waːŋ¹	pʰaːŋ¹'
'day'	*hŋwan¹	*ʔp-	van¹	van¹	vɐn¹	wan¹	wɐn¹	ʋɐn¹	wɐn¹	wan¹	man¹

* The Sui and Kam reflexes suggest PKS initial *phwj-.

The palatalisation on initial labio-dental onsets in 'dream' and 'to fly' (PKS *pwj-, *phwj-) seems to have "prevented" v- from weakening to w- (for which a loss of palatalisation would be articulatorily almost unavoidable). The velar fricative in [ɣʊk⁷] 'ashes' (JR, SY) is probably due to conditioning by the high, back vowel which follows it.

Incidentally, the author has observed that almost all speakers under the age of 40 in Shuiyao consistently realise prenasalised bilabial stops (from PKS *mp-) as v- (sometimes weakening to ʋ- or w-). This would compound any difficulties they have in comprehending Central Sui, whose v- is a reflex of PKS *pw-, *phw-. For example, [ᵐbaːn¹] 'male' (from PKS *mpaːn¹), is pronounced as [vaːn¹] by young people in Shuiyao, which sounds very similar to Sandong Sui [vaːŋ¹] 'tall, high'. Similarly [ᵐbi³] 'leech' (from PKS *mpliŋ¹) is pronounced [vi³], differing from Sandong Sui [vi¹] 'fire' only in tone.

4.3 PKS alveolar stop-lateral clusters
PKS unaspirated alveolar-lateral clusters have lost their alveolar stop onset in Central Sui whereas they have lost their lateral release in Southern Sui. In all known Kam varieties, preglottalised stops are lost completely, and *tl- and *ʔdl- become l-. Correspondences are shown in Table 8.

---

¹⁹ faːn³ is possibly non-cognate. Wei & Edmondson (2003) record both faːn³ and fa³ for "palm" or "sole".

Table 8: *PKS *tl- and *ʔdl- correspondences, words bearing ʔd- shaded in grey.*

Gloss	PKS	PS[20]	Central Sui			Southern Sui				Kam
			ND	SD	TP	JR	JQ	SY	SN	
'bone'	*tla:k^7	*ʔdl-	la:k^7	la:k^7	la:k^7	ʔda:k^7	ʔda:k^7	ʔda:k^7	ʔdak^7	la:k^9
'lightning'	*ʔdla:p^7	-	la:p^8	la:p^7	-	ʔda:p^7	ʔda:p^7	ʔda:p^7	-	la:p^9
'boat'	-	-	lwə^1	lwa^1	lua^1	(çyən^2)	ʔda^1	ʔda^1	la^1	lo^1
'to turn inside out'		*ʔdl-	lin^3	-	lɪn^3	ʔdɪn^3	ʔdɪn^3	ʔdin^3	ʔdin^3	ljin^3
'wild boar'	*ʔdla:i^5	*ʔdl-	la:i^5	-	-	ʔda:i^5	ʔda:i^5	ʔda:i^5	ʔda:i^5	la:i^5
'hornet'	*ʔdlu^1	-	-	lu^1	lu^1	ʔdu^1	ʔdu^1	ʔdu^1	-	la:u^1
'to awaken'	-	-	lju^1	lju^1	lju^1	ʔdjo^1	ʔdju^1	ʔdjɔ^1	lju^1	lhjo$^{1'}$
'fingernail'	*ʔdlyap^7	*ʔdl-	ljap^7	ljap^7	ljɐp^7	ʔdjɐp^7	ʔdjɐp^7	ʔnjɐp^7	ʔdjap^7	nəp^7
'to pull'*	*ʔdla:k^7	-	ʔda:k^7	ʔda:k^7	ʔdak^7	qa:k^7	qa:k^7	qa:k^7	qa:k^7	kwa:k^9

* Thurgood's reconstruction is based solely upon the Maonan pronounciation [ʔda:k^7]. Based on Sui and Kam reflexes, PKS **kra:k^7 seems more likely (see 4.4 below).

In Central Sui, the loss of the alveolar stop onset has resulted in a merger with PKS *hl- and *l-, the former becoming voiced after the pan-Tai-Kadai voiced onset induced tone split.[21] In Southern Sui, the loss of the lateral release resulted in a merger with PKS *ɗ-, which has also become a voiced preglottalised alveolar stop. These different mergers are illustrated in Table 9.

Table 9: *Divergent mergers involving PKS *ɗ-, *tl- and *hl- in Central and Southern Sui (mergers shown in grey).*

PKS	PS	Central Sui		Southern Sui	
*ɗ-	*ʔd-	ʔd-	ʔda:i^1 'good'	ʔd-	ʔda:i^1 'good'
*tl-, *ʔdl-	*ʔdl-	l-	la:i^5 'wild boar'	ʔd-	ʔda:i^5 'wild boar'
*hl-, *l-	*l-	l-	la:i^1 'back(-bone)'	l-	la:i^1 'back(-bone)'

These two contrasting developments of PKS lateralised stops are particularly significant in some Southern Sui areas in terms of affecting comprehension between the dialects. The author has observed that in both Shuiyao and Jiarong, Sui is currently undergoing a sound change by which all prenasalised alveolar stops [ⁿd] are becoming

---

[20] In Zeng's (2004) modification of Proto-Sui, she eliminates *ʔdl- from her reconstruction. Alluding to the facts that in modern-day Thai these words all exhibit an initial *ʔd-, and that a ʔd-/l- alternation in Sui is common (and presumably not uniform), she proposes PS ʔd- for these words. Pittayaporn's (2009) reconstruction of Proto-Tai contradicts Zeng's modification. He posits PT initials with lateral elements for 'lightning' PT *m.le:p^D, 'to turn inside out' PT *pli:n^C, 'awaken' PT *plok^D and 'fingernail' PT *Ç.lep^D. We thus retain Zeng's original reconstruction *ʔdl- here.

[21] Tai-Kadai languages originally had four tones, each of which split into two depending on the voicing of the proto-initial, in common with Hmong-Mien and Sinitic languages (Edmondson & Solnit 1990:8; Shi 1997:161-2). Thus PKS words with voiceless initials always bear odd-numbered tones and PKS words with voiced initials always bear even-numbered tones.

alveolar laterals [l]. Comprehension difficulties between young speakers from these locations and Central Sui are therefore likely to be compounded. For example, the Central Sui word for 'wild boar', [laːi⁵], sounds identical to some Southern Sui speakers' pronunciation of the word 'dry field', [laːi⁵] ([ndaːi⁵] in most other Sui areas). Furthermore, l-, ⁿd- and ʔd- are all common initials in Sandong Sui.

Note that Thurgood's proposed aspirated PKS equivalent, *thl-, has become a prenasalised alveolar stop ⁿd- in all varieties of Sui, forming a merger with *thr- in most places, for example PKS *thla¹ 'eye' > ⁿda¹ (all dialects), PKS *thram⁵ 'low' > ⁿdɐm⁵ (most dialects). This is given more attention in Appendix B and correspondences are shown in Table 25.

4.4 PKS velar-lateral and velar-rhotic clusters

In general, Thurgood's proposed PKS *kl- and **kr-[22] have respectively become simple velar or uvular (> velar before a high vowel) stops in Southern Sui (and, it may be added, in many varieties of Kam). This k-/q- alternation is not accounted for in Zeng's PS reconstruction. In Central Sui, both of these PKS forms are realised as preglottalised alveolar stops. Correspondences are given in Table 10.

Table 10: *PKS *kl-, *kr- and **kr- correspondences, words bearing k-/q- shaded in grey.*

Gloss	PKS	PS	Central Sui			Southern Sui				Kam
			ND	SD	TP	JR	JQ	SY	SN	
'far'	*klaːi³	*kl-	ʔdi¹	ʔdi¹	-	(haːn⁴)²³	(χaːn⁴)	qaːi¹	qaːi¹, ʔdi¹	ljaːi¹
'left over'	*kla¹	-	ʔdje⁵	ʔdja¹	-	ka¹	ʈa¹	ka¹	ka¹	ka¹
'seedling'	*kla³	*kl-	ʔdje³	-	ʔdiːɛ³	ka³	ʈa³	ka³	ka³	ka³
'salt'	-	*kl-	ʔdwə¹	ʔdwa¹	ʔdua¹	kwa¹	kwa¹	kwa¹	kwa¹	-
'hard'	**kra³	*kl-	ʔda³	ʔda³	-	qa³	qa³	qa³	ka³	kwa³
'(fish) scales'*	**krin⁵	*ʔd-	-	ʔdjən⁵	-	kən⁵	ʈin⁵	kən⁵	kin⁵	kwan⁵
'bright'	**kraːŋ¹	-	ʔdaːŋ¹	ʔdaːŋ¹	-	qaːŋ¹	qaːŋ¹	qaːŋ¹	qaːŋ¹	kwaːŋ¹
'to pull'†	*ʔdlaːk⁷	*ʔd-	ʔdaːk⁷	ʔdaːk⁷	ʔdak⁷	qaːk⁷	qaːk⁷	qaːk⁷	qaːk⁷	kwaːk⁹
'clothes'	-	*kl-	ʔduk⁷	ʔduk⁷	ʔdʊk⁷	kuk⁷	qok⁷	kʊk⁷	kuk⁷	quk⁹
'to wait'	*kra³	*k-	ka³	ka³	ka³	ka³	ka³	ka³	ka³	-
'to laugh'	*kru¹	*k-	ku¹	ku¹	ku¹	kɔ¹	ku¹	kɔ¹	ku¹	ko¹
'egg'	*krai⁵	*k-	kai⁵	kai⁵	kɛi⁵	kɛi⁵	kɛi⁵	kɛi⁵	kai⁵	kai⁵

* Zeng (1994:222) indicates that this is an old Chinese loan word, from OC *ljin.

† Zeng (1994:214) claims that this is related to Chinese 擢 'to draw up, to pull up', from OC *drakw. Both Zeng's and Thurgood's data lacked the Southern Sui reflexes for this word. PKS **kraːk · seems more likely for this word.

---

[22] Thurgood uses a double asterisk to denote reconstruction of a Chinese loan word, but he does not identify source words.

[23] In Shuiqing they say [qaːi¹], as in Shuiyao.

A sound change similar to that seen in Central Sui has also been observed in White Hmong ("*Hmoob Daw*"): Proto-Miao *ql- > Proto-Farwestern Miao *tl- > White Hmong ʔd- (Castro & Gu 2010:19; Johnson 1998:22). As far as the author is aware, this sound change is unique to Central Sui within the Kam-Sui branch. This ʔd-, k- alternation only occurs on words bearing odd-numbered tones, thus there are no voiced counterparts reconstructed for Proto-Sui or Proto-Kam-Sui.

The ɬ- variant for k- seen in Jiuqian (and Shuiqing, e.g. ɬa³ 'seedling') is common in many Chinese dialects including Mandarin. For example, LMC *kja¹ 家 'family' > ka¹ (Cantonese), tɕa¹ (Mandarin); LMC *kjaːj¹ 街 'street' > kaːi¹ (Cantonese, southwestern Mandarin), tɕe¹ (Mandarin); LMC *kiajŋ³ 景 'scenery' > kiŋ² (Cantonese), tɕiŋ³ (Mandarin); and LMC *kɦiaŋ⁴ 強 'strong' > kʰœŋ⁴ (Cantonese), tɕʰiaŋ² (Mandarin). These k-, ɬ- alternations indicate a palatalised onset in Proto-Sui.[24]

In general, PKS *kl- and **kr- have merged with PKS *ɗ- (becoming ʔd-) in Central Sui, whereas they have merged with PKS *kr- (becoming k-) or PKS *k- (becoming q-) in Southern Sui. This is illustrated in Table 11.

Table 11: *Divergent mergers involving PKS *ɗ-, *kl-, *kr- and *k- in Central and Southern Sui (mergers shown in grey).*

PKS	PS	Central Sui		Southern Sui	
*ɗ-	*ʔd-	ʔd-	ʔdaːi¹ 'good'	ʔd-	ʔdaːi¹ 'good'
*kl-, **kr-	*kl-	ʔd-	ʔdje³ 'seedling', ʔda³ 'hard'	q-, k-	ka³ 'seedling', qa³ 'hard'
*kr-	*k-	k-	ka³ 'to wait'	k-	ka³ 'to wait'
*k-	*q-	q-	qa¹ 'crow'	q-	qa¹ 'crow'

Let's look briefly at the aspirated counterparts. In all varieties of Sandong Sui, PKS *khr- has consistently lost its rhotic release, either becoming kʰ- or qʰ-. The conditions under which its place of articulation moves back to the uvular position are unclear due to a lack of data. Zeng reconstructs both *kʰ- and *qʰ- in Proto-Sui. PKS *khl- is also realised as either kʰ- or qʰ- in Southern Sui, whereas it has become h- in Central Sui. Zeng reconstructs these initials as *χ- in Proto-Sui.[25] Correspondences are given in Table 12.

Considered in isolation, Sui data suggests that PKS *khl- > Proto-Sui *qʰ-, which then weakened in Central Sui to χ-[26] (before PKS *khr- > PS *qʰ-), but remained a stop in Southern Sui, becoming velar when followed by a short vowel in checked syllables (in the case of Shuiyan, becoming velar in all environments).

---

[24] Further evidence for palatalised onsets in Proto-Sui (which were not reconstructed by Zeng) is given in Appendix B.

[25] Pittayaporn (2009) proposes *ʰl- for cognates of 'iron', 'liquor', 'to fear' and 'grandchild' in Proto-Tai. Zeng later revised her reconstruction for this initial to *khl-, due primarily to Li Fang-kuei's suggested *hl- for Proto-Tai (Zeng 2004:53).

[26] χ- is often the actual pronunciation of what is generally transcribed as h- in this paper.

Table 12: *PKS *khl-, *khr- correspondences, words bearing kʰ- or qʰ- shaded in grey.*

Gloss	PKS	PS	Central Sui			Southern Sui				Kam
			ND	SD	TP	JR	JQ	SY	SN	
'iron'*	*khlit⁷	-	hjət⁷	ɕət⁷	ɕet⁷	kʰət⁷	ʈʰət⁷	kʰət⁷	kʰət⁷	kʰwət⁷ˈ
'liquor'	*khlaːu³	*χ-	haːu³	haːu³	haːu³	qʰaːu³	qʰaːu³	qʰaːu³	kʰaːu³	kʰwaːu³
'to fear'	-	*χ-	ho¹	ho¹	hɔ¹	qʰɔ¹	qʰo¹	qʰɔ¹	kʰo¹	-
'lazy'	*khlut⁷	*χ-	hat⁷	hət⁷	hət⁷	kʰət⁷	qʰət⁷	kʰət⁷	kʰət⁷	kʰut⁷ˈ
'earth'	*khlum⁵	-	hum⁵	hum⁵	hom⁵	kʰum⁵	qʰom⁵	kʰəm⁵	kʰum⁵	-
'grandchild'	*khlaːn¹	*χ-	haːn¹	haːn¹	-	qʰaːn¹	qʰaːn¹	qʰaːn¹	kʰaːn¹	kʰwaːn¹ˈ
'river snail'	*khruːi¹	-	qʰui¹	qʰui¹	qʰuːi¹	kʰui¹	qʰoːi¹	kʰui¹	kʰui¹	-
'ear'	*khra¹	*qʰ-	qʰa¹	qʰa¹	qʰa¹	qʰa¹	qʰa¹	qʰa¹	qʰa¹	kʰa¹ˈ
'tasty'	-	*qʰ-	-	qʰan¹		qʰɐn¹	qʰɐn¹	qʰɐn¹	qʰan¹	-
'to bark'	*khrau⁵	*kʰ-	kʰau⁵	kʰau⁵	kʰʌu⁵	kʰʌu⁵	kʰʌu⁵	kʰɔu⁵	kʰau⁵	kʰau⁵
'centipede'	*khryap⁷	*kʰ-	kʰʊp⁷	kʰʊp⁷	kʊp⁷	kʰɐp⁷	kʰʊp⁷	kəp⁷	kʰap⁷	kʰəp⁷ˈ

* Sui reflexes indicate a different PKS onset. See 4.5 below.

In Central Sui, PKS *khl- > h- constitutes a merger with PKS *khj-, whereas in Southern Sui PKS *khl- merges with Proto-Sui *qʰ- (from PKS *khr-). This is illustrated below.

Table 13: *Divergent mergers involving PKS *khr-, *khl- and *khj- in Central and Southern Sui (mergers shown in grey).*

PKS	PS	Central Sui		Southern Sui	
*khr-	*qʰ-, *kʰ-	qʰ-, kʰ-	qʰa¹ 'ear'	qʰ-, kʰ-	qʰa¹ 'ear'
*khl-	*χ-	h-	haːn¹ 'grandchild'	qʰ-,kʰ-	qʰaːn¹ 'grandchild'
*khj-	*h-	h-	haːn³ 'red'	h-	haːn³ 'red'

4.5 Pre-palatal and velar/uvular alternations

There is some inconsistent alternation between ɕ-, ʈ-, ʈʰ- and kʰ-, qʰ- in some words, shown in Table 14. Central Sui tends to exhibit pre-palatals whereas Southern Sui tends to exhibit velars or uvulars, although not uniformly. This alternation suggests a palatalised Proto-Sui initial such as *khj-. At least two of the words are almost certainly loans from Early Chinese: 'iron' (铁 EMC *thɛt^D) and 'congee' (羹 EMC *kəijŋ^A).

Table 14: *Correspondences showing ɕ-, ȶ-, ȶʰ- and kʰ-, qʰ- alternations, the latter shaded in grey.*

Gloss	PKS	PS	Central Sui			Southern Sui				Kam
			ND	SD	TP	JQ	SQ	JR	SN	
'iron'	*khlit⁷	-	hjət⁷	ɕət⁷	ɕet⁷	ȶʰət⁷	kʰət⁷	kʰət⁷	kʰət⁷	kʰwət⁷'
'pointed, sharp'*	-	*ɕ-	hje¹	ɕa¹	ɕiːɛ¹	ȶʰa¹	ɕa¹	kʰa¹	kʰa¹	ȶa¹'
'congee'	-	-	tsiŋ⁵	qeŋ¹	ȶiŋ¹	qɛːŋ¹	qeŋ¹	qɛːŋ¹	qɛːŋ¹	qeŋ¹
'arrow'	-	*qʰ-	-	ɕam³	ɕam³	qʰɐm³	qʰam³	qʰɐm³	qʰɐm³	-
'Rongjiang (toponym)'	-	-	-	ȶʰɛːk⁷	-	qʰɛːk⁷	-	-	-	-

* Kam word given here is from Jianhe and Tianzhu counties of northern Kam. Other Kam varieties use a non-cognate word.

4.6 Velar fricative /x/ in Southern Sui

The velar fricative /x/ is not identified as a separate phoneme in most published descriptions of Central Sui (Zhang 1980:8-9; Xia 1992:282; Zeng & Yao 1996:259; GZARMLC 2008:780-782)[27]. Wei & Edmondson (2008) include /x/ in their phonemic inventory but fail to cite any examples of it. However, the four published phonological sketches of Sui spoken in the Southern Sui dialect area *do* identify /x/ as an individual phoneme and cite examples (Li 1977:84; CNU 1985:10; Xia 1989:268-9; ILCRD 1996:56). Stanford (in prep.) does not provide a phonology of Sui as such, but the data which he presents indicates that [x] does not occur in Sandong Sui at all, apart from in the word [xʲɐk⁷][28] 'diligent' in a speaker from Jiuqian, which lies within the Southern Sui area.

Data collected by the author agrees with previous literature, indicating that /x/ is a distinct phoneme which occurs throughout the Southern Sui dialect area but not in Central Sui. Thus Southern Sui possesses a complete set of pre-palatal, velar and uvular obstruents, illustrated here (following Jiuqian pronunciation):

Pre-palatal, velar and uvular fricatives		*Pre-palatal, velar and uvular stops*	
ɕaːŋ¹	'heat of the sun'	ȶʰaːŋ¹	'handsome'
xaːŋ¹	'to roast (sticks of meat over a fire)'	kʰaːŋ¹	'to fry (lightly in oil)'
χaːŋ¹	'root (of a tree)'	qʰaːŋ¹	'rafter'

Thurgood did not deal with this correspondence pattern because he did not have sufficient Southern Sui data. Zeng proposes PS *khl- for this onset.[29] PS *khl- has

---

[27] SDB (1956:68) documents one instance of [x] in a Central Sui speaker, in the word [xak⁷] 'diligent'. This particular speaker was from Hengfeng, situated in the far south-west of the Central Sui dialect area. The Hengfeng speaker recorded by Stanford (in prep.) pronounced this word [kʰak⁷]. It is possible that the speaker transcribed in SDB (1956) was influenced by the pronunciation of Southern Sui speakers nearby in Shuili district (Li-Ngam).

[28] The palatalisation on the initial [x] has not been observed by the author anywhere else in the Southern Sui area.

[29] Zeng (2004:53) later revised her reconstructions for both this initial and her original PS *kʰ- to PS *kr-. In this case, /x/ in Southern Sui would constitute a split from *kʰ-, but the conditioning environment causing such a split is not evident.

generally become kʰ- in Central Sui (through cluster reduction, resulting in a merger with PKS *khr- or PS *kʰ-), x- in Southern Sui (through lenition), and j- in most varieties of Kam (through palatalisation and lenition)[30]. These correspondences are illustrated in Table 15.[31]

Table 15: *kʰ-, x- alternation in Central and Southern Sui. Words bearing x- are shaded in grey.*

Gloss	PS	Central Sui			Southern Sui				Kam
		ND	SD	TP	JR	JQ	SY	SN	
'crispy'*	-	kʰim⁵	-	kʰim¹	xɪm¹	xim¹	xim¹	kʰim¹	jim¹'
'net (for catching fish)'†	*khl-	-	-	kʰɛ¹	xe¹	xe¹	xe¹	kʰe¹	je¹'
'rib'	*khl-	-	kʰət⁷	-	xət⁷	xət⁷	xət⁷	xət⁷	-
'diligent'	*khl-	ʔɣak⁷	kʰak⁷	-	xɐk⁷	xɐk⁷	xɐk⁷	xak⁷	jak⁷'
'maple'	-	(fu¹)	kʰaːu¹	-	xo¹	(fu¹) [32]	xo¹	(fu¹)	jaːu¹'
'to roast (meat sticks over a fire)'	-	-	-	kʰaːŋ¹	(sʌu³) [33]	xaːŋ¹	xaːŋ¹	-	-

* Li (2008) has [xim¹] for LN.
† Li (1965, 2008) has [xe¹] for LN. Zeng (2004) has [kʰe¹] for SD.

Other instances of x- in Southern Sui appear to be reflexes of PS *h- and *ɦ- (from PKS *khj-, *dz-[34]) conditioned by a central or mid-centralised vowel, evidenced in Table 16. These data also show that in Shuiqing, Shuiyan and sometimes Jiarong, hw- or hu- > f-. The author has observed the same phenomenon among younger speakers in Shuiyao, except that [u] is lost entirely, for example: [hui⁶] 'to sit' > [fi⁶] (also observed in Jiarong); [hui²] 'snake' > [fi²]. In view of the widespread occurrence of a f-, hu- alternation (often in free variation) in southern Chinese dialects (see section 4.1 PKS *xw- above), this should not be viewed as compelling evidence for a Central/Southern Sui divide.

---

[30] There are two Kam varieties from Liping county (northeastern Southern Kam area) and one Kam variety from Congjiang county (which borders the Southern Sui area in Libo county) which also have an x- onset corresponding to Southern Sui x-.

[31] Incidentally, Zeng suggests that three of these words are possibly old loans from Chinese: 'diligent' [kʰak⁷] (SD) <仂 'diligent' OC *ljək (Zeng 1994:136); 'rib' [kʰət⁷] (SD) < 骨 'bone' OC *kwət (Zeng 1994:156); and 'net' [kʰe¹] (SD) < 罟 'fish-net' OC *kag (Zeng 1994:197).

[32] According to a Sui speaker from Jiuqian, [xo¹] refers to the sap of the maple tree, which is used to make medicine, rather than to the tree itself.

[33] Not to be confused with [saːu³], which is the general word for 'to cook (dishes)'.

[34] The PKS *dz- initial, which seems slightly surprising when considering Sui and Kam data alone, is proposed by Thurgood largely on the basis of regular z- and t- reflexes in Maonan and Mulam respectively.

Table 16: *PKS *khj-, *dz- correspondences, words with x- onset shaded in grey, words with f- onset framed in double lines.*

Gloss	PKS	PS	Central Sui			Southern Sui					Kam
			ND	SD	TP	JR	SQ	JQ	SY	SN	
'tail'	*khjut⁷	-	hət⁷	hət⁸	hət⁸	xət⁸	hət⁸	xət⁸	xət⁸	zət⁸*	sət⁷
'sour'	*khjum³	*h-	hum³	hum³	hom³	xʊm³	xəm³	χom³	xəm³	fum⁴	səm³
'early'	*khjam¹	*h-	ham⁵	ham¹	hɐm¹	hɐm¹	-	χɐm¹	hɐm¹	ham¹	sam¹
'intestines'	*khja:i³	*ɦ-	ha:i⁴	ha:i⁴	ha:i⁴	ha:i⁴	ha:i⁴	χa:i⁴	ha:i⁴	hai⁴	sa:i³
'to give'	*khja:i¹	-	ha:i¹	ha:i¹	ha:i¹	ha:i¹	ha:i¹	χa:i¹	ha:i¹	ha:i¹	sa:i¹
'place'†	-	-	-	hən²	-	xən²	χən²	xən²	xən²	hən²	-
'to blow'	*dzup⁸	-	hup⁸	hup⁸	hop⁸	fʊp⁸	xəp⁸	χop⁸	xəp⁸	hup⁸‡	səp⁸
'worm'	*dzan⁴	*ɦ-	han⁴	-	hɐn⁴	hɐn⁴	han⁴	χɐn⁴	hɐn⁴	-	san⁴
'snake'	*dzu:i²	*ɦ-	hui²	hu:i²	hui²	hui²	fui²	χu:i²	hui²	hui²‡	sui²
'to sit'	*dzu:i⁶	-	hui⁶	hui⁶	hu:i⁶	hui⁶	fui⁶	χui⁶	hui⁶	hui⁶‡	sui⁵
'to pound'	*kja:k⁷	*h-	ha:k⁷	ʁa:k⁷	ha:k⁷	fʊk⁷	fuk⁷	χok⁷	hʊk⁷	fuk⁷	sa:k⁹

* For LN. This word does not appear in Zeng's SN data.
† Zeng gives [hen²] for SD, YK and SN.
‡ Li (1965) gives initial f- for all these three words.

4.7 Vowels

In certain words, Central Sui has labio-velar onglides (seemingly retentions from Proto-Kam-Sui labialised initials) where Southern Sui does not. Correspondences are shown in Table 17. These onglides appear to have monophthongised in Southern Sui when preceded by alveolar consonants (thus *ɓlwa¹ 'navel' > [ʔda¹] in Southern Sui).

Table 17: *PKS *-wa correspondences, words with labio-velar onglides shaded in grey.*

Gloss	PKS	Central Sui				Southern Sui				Kam
		ND	SD	TP	RJ	JR	JQ	SY	SN	
'navel'	*ɓlwa¹	ʔdwə¹	ʔdwa¹	-	dwa¹	ʔda¹	ʔda¹	ʔda¹	ʔda¹	po⁵
'boat'	-	lwə¹	lwa¹	lu:ɐ¹	-	ʔda¹	ʔda¹	ʔda¹	la¹	lo¹
'to rest'	-	-	lwa²	lu:ɐ⁵	-	la⁵	la⁵	la⁵	la⁵	sa⁵
'*Dwac* festival'	-	-	twa³	tu:ɐ³	-	ta³	ta³	ta³	-	-
'salt'*	-	ʔdwə¹	ʔdwa¹	ʔdu:ɐ¹	-	kwa¹	kwa¹	kwa¹	kwa¹	-
'cloud'	*m-xwa³	fa³	fa³	fa³	fa³	fa³	ʋa³	wa³	wa³	ma³
'wing'	*pwa⁵	va⁵	va⁵	va⁵	va⁵	wa⁵	ʋa⁵	wa⁵	wa⁵	pa⁵

* The onset reflexes suggest PKS *klwa¹ (see section 4.4).

In Nandan and other Central varieties discussed in Stanford (in prep.), the [a] in these glides have raised and centralised to [ə]. Stanford found that the vowel raising in [ua]

onglides is a feature common to western Sui varieties which occurs concurrently with a similar vowel raising in [ia] onglides. The latter seems to be due to a breaking and raising (and sometimes centralising) of PKS *-ja which often occurred with a loss of palatalisation on the initial. Some examples are given in Table 18. There is no evidence of this sound change in Southern Sui.

Table 18: *PKS *-ja correspondences, words showing raising or centralising of final vowel shaded in grey.*

Gloss	PKS	Central Sui				Southern Sui				Kam
		ND	SD	TP	PY	JR	JQ	SY	SN	
'river'	*ʔnja¹	-	ʔnja¹	ʔniːɛ¹	ʔniə¹	ʔnja¹	ʔnja¹	ʔnja¹	ʔnja¹	ȵa¹
'bored'	*ɓja⁵	ʔbe⁵	-	ʔbiːɛ⁵	ʔbiə⁵	ʔbja⁵	ʔbja⁵	ʔbja⁵	ʔbja⁵	mja⁵
'hand'	*k-mja¹	mje¹	mja¹	miːɛ¹	miə¹	mja¹	mja¹	mja¹	mja¹	mja²
'tea'	-	tsje²	tsja²	tsiːɛ²	tsiə²	tsja²	tsja²	tsja²	tsja²	ɕe²

4.8 Tones

On words with PKS initial *hr-, the tone is odd-numbered in Southern Sui (as expected with a voiceless PKS initial) whereas it is even-numbered in Central Sui and all varieties of Kam. Correspondences are given in Table 19. It suggests that PKS *hr- acquired voicing *before* the general voiced onset induced tone split in Kam and Central Sui, whereas it became voiced *after* the tone split in Southern Sui. Thurgood (1988:191) draws attention to the same phenomenon, noting that in Mulam, Kam and Then, these particular tones are normally associated with initial voicing, whereas in Sui (his data includes Sandong Sui and Li-Ngam Sui) and Mak, they are normally associated with initial voicelessness. He does not, however, point out the discrepancy between Sandong Sui (a Central Sui lect) and Li-Ngam Sui (Southern Sui), perhaps because his data only contained two correspondences exhibiting this tone change.

Zeng reconstructed two Proto-Sui initials for PKS *hr-: PS *x- (which she later revised to *ⁿk-, Zeng 2004:52) for words with odd-numbered tones; and PS *ɣ- for words with even numbered tones. Inconsistencies in her data (as seen in, for example, the SN data in Table 19) seemingly obscured any regular pattern of tones among the dialects and led her to this hypothesis. The regularity of the tone differences within Central and Southern Sui on these words suggest that PS retained PKS *hr- and that the tone split occurred at a later stage.

A similar tone alternation occurs on PKS *mpr- in Kam, Nandan Sui and Tingpai Sui, evidenced in Table 20. A lack of data makes it difficult to draw concrete conclusions. It seems that, as with tones on words with PKS *hr- onset, Southern Sui always retains the original tone on PKS *mpr- words. But not all Central Sui lects have, like Kam, seen a tone shift on the same words.

Table 19: *PKS *hr- correspondences. Words with an even-numbered tone are shaded in grey.*

Gloss	PKS	PS	Central Sui				Southern Sui					Kam
			ND	SD	TP	RJ	JR	JQ	SY	LN	SN	
'home'	*hraːn¹	*x-	ɣaːn²	ɣaːn²	ɣaːn²	ɣan²	ɣaːn¹	ɣaːn¹	ɣaːn¹	ɣaːn¹	ɣaːn²	jaːn²
'two'	*hra¹	*ɣ-	ɣa²	ɣa²	ɣa²	ɣa²	ɣa¹	ɣa¹	ɣa¹	ɣa¹	ɣa²	-
'pear'	-	*ʔɣ-	ɣai²	ɣai²	ɣɛi²	-	ɣɛi¹	ɣɛi¹	ɣɛi¹	-	ɣai¹	jai²
'to drink'	-	*ɣ-	-	ɣum⁴	ɣom⁴	-	ɣʊm³	ɣʊm³	ɣʊm³	ɣum³	ɣum⁴	hum⁴
'footprint'	*hruːi¹	-	ŋui²	-	ɣui²	ɣui²	ɣi¹	ɣi¹	ɣi¹	ɣi¹	-	-
'to swim'	-	-	-	-	luːi²	-	ɣi¹	ɣi¹	ɣi¹	-	-	-
'sharp'	*hraːi⁵	-	-	-	-	-	-	-	-	-	-	jaːi⁶
'to cough'*	-	*ɣ-	-	ɣuk⁸	-		-	wok⁷	-	-	ɣuk⁸	-
'to know (a person)'	-	*ɣ-	-	ɣo⁴	-	ɣo⁴	ɣɔ³	ɣo³	ɣo³	ɣo³	ɣo⁴	jo⁴

* YK, SQ (both Southern Sui) =ɣuk□.

Table 20: *PKS *mpr- correspondences. Words with an even-numbered tone are shaded in grey.*

Gloss	PKS	PS[35]	Central Sui				Southern Sui				Kam
			ND	SD	TP	PY	JR	JQ	SY	SN	
'ear of grain'	*mpraːŋ¹	*mb-	ʔbjaːŋ²	ᵐbjaːŋ¹	mjaŋ²	mjaŋ¹	ᵐbjaːŋ¹	ᵐbjaːŋ¹	ᵐbjaːŋ¹	ᵐbjaŋ¹	mjeŋ²
'to plant (a seedling)'	*mpra¹	*mb-	mje²	ᵐbja¹	miɛ²	-	ᵐbja¹	ᵐbja¹	ᵐbja¹	-	mja²

There is some regional variation in the pitch values of Tones 1 and 6, Tone 6 being the most salient. Stanford (in prep.) describes the variation in Tone 6 across the Sui area. It is realised as a high level 55 tone in the Pandong dialect area and in a small area around Sandong township, whereas it is realised as a low rising 24 tone in the rest of the Central Sui area and in Yang'an. Data collected for this study shows that Tone 6 is consistently realised as a high level 55 tone in the Southern Sui area. Phonetic tone values are not discussed further here because they does not constitute phonological divergence.

4.9 Phonological innovations, summary
In summary, Southern Sui lects have undergone a series of phonemic mergers different from those in Central Sui varieties. These are summarised in Table 21. The fact that these mergers have happened so consistently across the Southern Sui region indicates that Southern Sui constitutes a separate dialect cluster within Sui.

---

[35] Zeng also reconstructs PS *mb- for words such as 'male' [ᵐbaːn¹] (all Sui dialects), 'to be near to, beside' [ᵐba³] (all dialects) and 'expensive' [ᵐbiŋ¹] (all dialects). Thurgood proposes PKS *mp- for these other words (the odd-numbered tone in all dialects indicates a voiceless onset). Zeng does not explain the development of palatalisation on 'ear of grain' and 'to plant (a seedling)'.

Table 21: *Summary of phonemic mergers in Central and Southern Sui.*[36]

PKS	Central Sui		Southern Sui	
	merges with (PKS)	reflex	merges with (PKS)	reflex
*xw-	*hw-, *ɣw-	f-	*pw-, *phw-	w-
*pw-, *phw-	(*hŋw-)[37]	v-	*xw-	w-
*gw-, *kw-	*p-	p-	*k-	q-
*tl-, *ʔdl-	*hl-, *l-	l-	*ɗ-	ʔd-
*kl-, **kr-	*ɗ-	ʔd-	*kr-, *k-	k-, q-
*khl-	*khj-	h-	*khr-	kʰ-, qʰ-
(PS *khl-)	*khr-	kʰ-	*khj-, *dz- (partial)	x-
*-wa / *alveolar C_____	-	-ua or -wə	*-a	-a

Of course, there are other minor phonemic and phonetic variations that have taken place among both Central and Southern Sui varieties, some of which are described in Appendix B. Further data from both the Central and Southern Sui areas are sure to bring to light other divergent diachronic innovations across the region.

5 Lexical similarity

In this section, I show that lexical similarity percentages confirm a two-way division between Central and Southern Sui lects. Heggarty (2010:307) affirms that lexicostatistical-type cognate counts can give us, in his words, "measures ... of divergence between given languages", based on two suppositions: 1. that the languages (or dialects) being compared are all descended from one proto language; and 2. that the more two languages (or dialects) have diverged, the more cognates inherited from their common ancestor language will have been lost. The first supposition is reasonable in the case of Sui, whose speakers all have a common autonym which is non-cognate with those of other ethnic groups living nearby.

Perceived "loss" of lexical cognates could be a result of either the replacement of older words by loan-words, or the meanings of cognate words diverging (through semantic broadening, narrowing or shift) such that they are no longer elicited for the same meaning slots. Therefore it is crucial that lexical items with precisely the same meanings are compared. Fortunately, both the Shuiqing and the Nandan data have tight meaning slots for each word. The Sandong glosses are vague, so they were cross-checked with the more precise entries in Zeng & Yao's (1996) *Chinese-Sui dictionary* (which is also based on Sui spoken in Sandong district) in order to reduce the chances of semantic mismatches. The author's own wordlists were elicited carefully to ensure semantic equivalence. Neither

---

[36] For the sake of clarity, only Thurgood's Proto-Kam-Sui forms are shown here. Readers should refer to previous sections to see the correspondences with Zeng's Proto-Sui.

[37] PKS *hŋw- is only attested in one word, *hŋwan¹ 'day'.

Zeng's (2004) nor Li Fang-kuei's (1965) data were included in the similarity counts because of their semantic ambiguities.[38]

For the lexical similarity count, then, seven Sui lects were compared, four in the Southern Sui cluster and three in the Central Sui cluster. All lexical items which appear in all seven sets of data, amounting to a total of 308 words, were included in the count. These included 90 words from the Swadesh 100 wordlist. Words were only considered "lexically similar" if they were proven historical cognates. Words which were likely cognates but which appeared to have undergone a sound change unattested elsewhere, such as the word for 'ox' (consistently pronounced [po^4] in Central Sui and [mo^4] in Southern Sui), were counted as "dissimilar".

The results, given in Table 22, reveal high lexical similarity (over 90%) among all four Southern Sui lects (Jiuqian, Jiarong, Shuiqing, Shuiyao) and a clear division between the Central and Southern Sui lects. Sandong and Jiuqian have relatively high average lexical similarity among all the lects (89.8% and 90.8% respectively) with relatively low standard deviations (3.00 and 3.54 respectively)[39], suggesting that they are lexically most "representative" of all seven lects.

Table 22: *Lexical similarity percentages among Sui lects. Percentages over 90% are shaded in grey.*

Nandan						
92.1	Tingpai					
91.7	93.5	Sandong				
85.9	89.0	90.6	Jiuqian			
83.8	86.3	89.2	96.7	Jiarong		
84.2	85.4	88.9	91.6	92.5	Shuiqing	
80.9	82.1	84.7	91.2	93.5	92.9	Shuiyao

Recently developed "Gabmap" software[40] was used to carry out clustering analysis (based on an average group method) and plot the results as a dendogram, shown here in Figure 5. This confirms a clear Central/Southern Sui division, although it disguises the overall relatively high similarity of both Jiuqian and Sandong to all the other dialects.

---

[38] All data sources are described in more detail in Appendix A.

[39] Mean lexical similarity percentages range from 86.4% (Nandan) to 90.8% (Jiuqian) and standard deviations range from 3.00 (Sandong) to 5.64 (Shuiyao).

[40] Gabmap is a dialectometry web application developed at the Centre for Language and Cognition at the University of Groningen. It is based on RuG/L04 software which was developed to perform Levenshtein distance calculations, clustering and multidimensional scaling. See Nerbonne et. al. (2011). Gabmap is accessible at http://www.gabmap.nl/
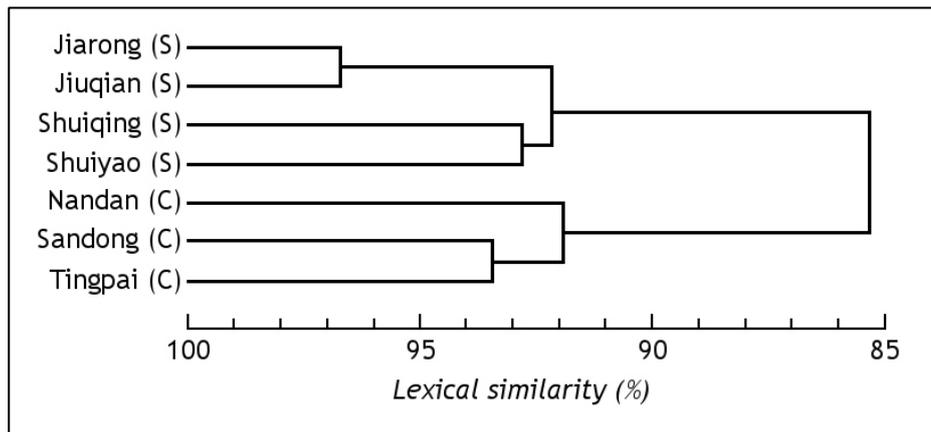
Figure 5: *Dendogram showing lexical similarity clustering (plotted with Gabmap), S=Southern Sui, C=Central Sui.*

A closer examination of the data shows that there are a number of words particularly characteristic of Southern Sui for which alternative forms are usually used in Central Sui. Some of these words are listed in Table 23. There are several cases in which Shuiqing appears to use a word more typical of Central Sui than Southern Sui. One possible explanation for this is that the speaker on whom the Shuiqing data is based had some knowledge of "standard (Sandong) Sui" and thus may occasionally have reported a Sandong word rather than the word used more commonly in his own village.

It should be noted that many of the "Southern Sui" words *are* present in Central Sui, but their meaning or usage is different. For example [ɕju³], the general word for 'tooth' in Southern Sui, is used in Central Sui to mean 'wisdom tooth' (Zeng & Yao 1996:248). The word [saːu³] is used in all Sui dialects to mean 'to cook (dishes)' but in Central Sui it also means 'to fry lightly in oil'. Southern Sui's [kʰaːŋ¹] 'to fry lightly in oil' is used in Central Sui to mean 'to roast (meat)' (Zeng & Yao 1996:106). Thus many of these lexical differences actually represent semantic narrowing, broadening or shift. The intricacies and subtleties of such instances of semantic divergence are complex enough for an entirely separate study.

There is, of course, lexical variation within Southern Sui itself. Some examples are listed in Table 24. Again, many of these cases are due to semantic shift. For example, the word most often used for 'black' in Jiarong and Shuiyao, [qɐm⁵], is used to mean 'dark (red)' in Central Sui varieties (Li 1965:170). Other lexical differences are due to borrowings. For example, both [maːu⁶] 'hat' and [tjɛm⁶] 'table' are Bouyei (Libo dialect) words (Wu et al. 2007:473; CNU 1985:95), the former probably an old loan from Chinese. Bouyei speakers are in close contact with Sui speakers in Shuiyao, Shuiqing and Jiarong.

Table 23: *Examples of lexical items characteristic of Southern Sui (highlighted in grey).*

	Central Sui			Southern Sui			
	ND	SD	TP	SQ	JR	JQ	SY
'mountain' [41]	tin²	nu²	nu² tin²	pja¹	pja¹	pja¹	pja¹
'insect'	ta¹, nui²	nui²	nuːi²	ta¹	ta¹	ta¹	ta¹
'praying mantis'	ʔe³	ʔẽ³	ʔẽ³	zaːi³	zai³	zaːi³	zaːi³
'tooth (human)'	vjan¹, heu³	vjan¹	vjen¹	ɕu³	ɕju³	ɕju³	ɕju³
'to fry (lightly in oil)'	tsaːu³	saːu³	saːu³	kʰaːŋ¹	kʰaːŋ¹	kʰaːŋ¹	kʰaːŋ¹
'to chat with friends for fun'	fjen³	ɕaːn³, fjaːn³	fiɛn³	qeu⁶	qɛːu⁶	qɛːu⁶	qɛːu⁶
'dark (at night)'	niŋ⁵	ⁿdən⁵	ⁿdɪŋ⁵	-	hʊp⁷	χʊp⁷	xəːp⁷
'silly, stupid'	tshun⁴ʼ*	pən¹*, ɕaŋ², nja⁵	nɐŋ⁴, ɕɐn²	ʔwa³	ʔba³	ʔwa³	ʔwa³
'beautiful'	ʔdaːi¹ɣak⁸	kiŋ³	kɪŋ³	qʰan⁵	qʰɐn⁵	qʰɐn⁵	qʰɐn⁵
'clever'	tshuŋ³min²*	ɕai¹	kin¹	kiŋ³	kɪŋ³	kiŋ³	kiŋ³, kwaːi¹
'smelly'	n̥u¹	n̥u¹	n̥u¹	ʔȵin¹	ʔȵin¹	ʔȵen¹	ʔȵin¹
'narrow'	ʔȵap⁷	ʔnjap⁷	ʔnjɐp⁷	ʁep⁷	ʁɛːp⁷	ɣɛːp⁷	ʁɛːp⁷
'(particle, completed action)'	-	ljeu²	ljeu²	ljeu²	ɣaːŋ³	ɣaːŋ³	ɣaːŋ³
'skin (human)'	pi²	pi²	pi²	pi²	ʁa¹	ʁa¹	ʁa¹
'knee'	qam⁴ qu⁵	qam⁴ qu⁵	qok⁷ qu⁵	qam⁴ qu⁵	qɔ³ tsɔːŋ⁵	ku³ tsoɐŋ⁵	qa³ ʈɔːŋ⁵
'to open' [42]	ŋ̊ai¹	ŋ̊ai¹	ŋ̊ei¹	ŋ̊ai¹	tsjaːŋ⁴	tsjaːŋ⁴	tsɛŋ⁴
'to love (child)'	ʔbjum¹	mjət⁷, ᵐbjum¹	mjɛt⁷	ᵐbjəm¹	ɣei¹	ɣɛi¹	ɣei¹
'key'	mai⁴fuŋ³	mai⁴ fuŋ³	mɛi⁴ foŋ³	mai⁴ fuŋ³	hɛi³ si²	mɛi⁴ si²	hɛi³ si²

* denotes a modern Chinese loan word.

---

[41] [tin²] more commonly means 'rock' and [nu²] means 'hill' or 'earth mountain' in Southern Sui dialects. [pja¹] also occurs in some varieties of Central Sui, but it means 'cliff' or 'rock'.

[42] In Jiuqian [ŋ̊ei¹] means "to open (a door) slightly" in contrast with [tsjaːŋ⁴] which means "to open (a door) wide".

Table 24: *Examples of lexical variation within Southern Sui. Non-cognate words are highlighted in grey.*

	Central Sui			Southern Sui			
Gloss	ND	SD	TP	SQ	JR	JQ	SY
'hat'	n̠on^4	n̠on^4	n̠ɔn^4	maːu^6	n̠uɜn^4	n̠oɐn^4	maːu^6
'black'	ʔnam^1	ʔnam^1	ʔnɐm^1	ʔnam^1	qɐm^5	ʔnɐm^1	qɐm^5
'table'	hi^3	çi^3	hi^3	tjəm^6	tjɛm^6	çi^3	tjɛm^6
'to vomit'	kuːn^5	kun^5	kuːn^5	ⁿdaːk^7	taːk^7	kuːn^5	ⁿdaːk^7
'to look for'	tʰaːu^3	tʰaːu^3	tʰaːu^3	tʰaːu^3	la^6	tʰaːu^3	la^6
'far'	ʔdi^1	ʔdi^1	-	qaːi^1	haːn^4	χaːn^4	qaːi^1
'near'	pʰjai^5	pʰjai^5	pʰɛi^5	ɣau^3	pⁱʰɛi^5	pʰjei^5	ɣou^3
'slowly'	fuːn^1	faːn^1	fɐɐn^1	maːn^6, ɣai^1	faːn^1	faːn^1	maːn^6, ɣɛi^1
'fine hair on body'	tsən^1 n̠ik^7	tsən^1 n̠on^5	tsən^1 n̠aːŋ^3	tsən^1 n̠on^5	tsən^1 hwa^3	tsən^1 n̠aːŋ^3	tsən^1 n̠on^5

6 Conclusions

Southern Sui speech varieties are both phonologically and lexically divergent from Central Sui. Furthermore, Southern Sui displays high internal consistency, both in terms of phonological innovations and lexical similarity, across a wide geographical area. These phonological and lexical differences could account for much of the reported difficulty in comprehension of Central Sui among Southern Sui speakers. In terms of cultural practices there is also a clear divide between Southern Sui speakers, most of whom celebrate *Maox*, and the rest of the Sui community, most of whom celebrate *Dwac*. Taken together, the linguistic and cultural evidence indicates that Southern Sui should be viewed as a distinct, fourth dialect of Sui.

Data in this paper also contribute to Proto-Sui and Proto-Kam-Sui reconstructions, confirming the likelihood of PS *f-, a series of PS palatalised onsets, and a hitherto unreconstructed PKS initial which developed into x- in Southern Sui.

Further research, particularly on Sui as spoken in the Pandong, Yang'an and Central dialect areas, is needed to determine genetic relationships between all Sui varieties. Thorough reconstructions of both Proto-Sui and Proto-Kam would be a valuable first step towards a more solid investigation into the historical development of Kam-Sui languages as a whole.

References

Baxter, W. (白一平) and L. Sagart (n.d.). 2011. Baxter-Sagart Old Chinese reconstruction (Version 1.00). Online at http://crlao.ehess.fr/document.php?id=1217. Accessed (15.08.2011).

Castro, Andy & Gu Chaowen. 2010. Phonological innovation among Hmong dialects of Wenshan. *Journal of the Southeast Asian Linguistics Society*, Volume 3.1:1-39.

Central University for Nationalities (CNU), Number 5 Research Unit of the Minority Languages Research Institute (ed.) 1985. *Zhuang-Dong yuzu yuyan cihui*

ji.《壮侗语族语言词汇集》 *[Lexicon of Tai-Kadai languages.]* Beijing: Central University for Nationalities Publishing House.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le Maître Phonétique* 30:24-27.

Diller, Anthony V. N. 2008. Introduction. In Diller, Edmondson and Luo (eds.) (2008).

Diller, Anthony Van Nostrand, Jerold A. Edmondson and Yongxian Luo (eds.) 2008. *The Tai-Kadai languages*. Abingdon, Oxon: Routledge.

Ding Bangxin (丁邦新) (ed.) 2005. Li Fanggui quanji 5: Mohua jilüe, Shuihua yanjiu. 《李方桂全集5：莫话记略·水话研究》 [Complete works of Li Fang-kuei volume 5: A brief record of Mak, Research on Sui language.] Beijing: Qinghua University Publishing House.

—— 2008. Li Fanggui quanji 6: Shuihua cihui. 《李方桂全集6：水话词汇》 [Complete works of Li Fang-kuei volume 6: Sui lexicon.] Beijing: Qinghua University Publishing House.

Edmondson, Jerold A. and David B. Solnit (eds.) 1990. *Comparative Kadai: Linguistic Studies beyond Tai* (revised edition). Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Edmondson, Jerold A. and Yang Quan. 1988. Kam-Sui resonant initials and tones. In Edmondson and Solnit (eds.) (1990).

Edmondson, Jerold A., John H. Esling, Jimmy G. Harris and James Wei. 2004. A phonetic study of Sui consonants and vowels. *Mon-Khmer Studies: A Journal of Southeast Asian Linguistics and Languages* 34: 47-66.

Guangxi Zhuang Autonomous Region Minority Languages Commission (广西壮族自治区少数民族语言文字工作委员会, GZARMLC) (eds.) 2008. *Guangxi minzu yuyan fangyin cihui.* 《广西民族语言方音词汇》 *[Glossary of minority languages and dialects of Guangxi.]* Beijing: Nationalities Publishing House.

Guizhou Minority Affairs Commission and Guizhou Minorities Research Institute (eds.) 2008. *Guizhou "Liu shan liu shui" minzu diaocha ziliao xuanbian: minzu yuyan juan.* 《贵州"六山六水"民族调查资料选编·民族语言卷》 [*Guizhou "Six Mountains and Six Rivers", selections from minority survey data, minority languages volume.*] Guiyang: Guizhou Nationalities Publishing House.

Harbsmeier, Christoph (ed.) (n.d.) Thesaurus Linguae Sericae: an historical and comparative encyclopedia of Chinese conceptual systems. University of Heidelberg. Online at http://tls.uni-hd.de/. Accessed (15.08.2011).

Heggarty, Paul. 2010. Beyond lexicostatistics: How to get more out of 'word list' comparisons. *Diachronica* 27:2, 301-324.

Institute of Language and Culture for Rural Development (ILCRD, Mahidol University) and the Kam-Tai Institute (Central University of Nationalities). 1996. *Languages and cultures of the Kam-Tai (Zhuang-Dong) group: A word list*. Bangkok: Mahidol University.

Johnson, Michael. 1998. Farwestern Hmongic. ms.

Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, Texas: SIL International. Online at http://www.ethnologue.com/. Accessed (10.06.2011).

Li Fang-kuei (李方桂). 1965. The Tai and the Kam-Sui Languages. *Lingua*. 14:148-179.

—— 1977. *Shuihua yanjiu* 《水话研究》 *[A study of the Sui language: Texts and translations (I)]*. Publications Editorial Committee of the Department of History and Linguistics at the Central Research Institute (Taipei) 中央研究院历史语言研究所出版品编辑委员会, reprinted in Ding Bangxin (ed.) (2005).

—— 2008. *Shuihua cihui* 《水话词汇》 *[A study of the Sui language: Glossary (II)]*. Published posthumously in Ding Bangxin (ed.) (2008).

Nerbonne, John, Rinke Colen, Charlotte Gooskens, Peter Kleiweg and Therese Leinonen. 2011. Gabmap - a web application for dialectology. Center for Language and Cognition, University of Groningen. ms. (http://www.gabmap.nl/wp-content/uploads/2011/05/Gabmap-long-2011-jan-07-rev-mei.pdf, accessed 22 July 2011)

Peiros, Ilia. 1998. *Comparative linguistics in Southeast Asia*. Canberra: Pacific Linguistics.

Pittayaporn, Pittayawat. 2009. *The Phonology of Proto-Tai*. Ph.D. dissertation. Department of Linguistics, Cornell University.

Pulleyblank, Edwin G (浦立本). 1991. Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin. Vancouver: University of British Columbia Press.

Sandu County Education Bureau and Sandu Sui Research Institute (ed.) 三都水族自治县教育局，三都水族自治县水族研究所. 2007. *Shuizu wenhua jinxiaoyuan xiaoduben* 《水族文化进校园小读本》*[Sui culture: a reader for primary school]*. self-published.

Shangguyin chaxun (SGYCX) 《上古音查询》 [Old Chinese database search] (n.d.). Online at http://www.eastling.org/oc/oldage.aspx. A searchable database produced by the Shanghai Higher Education Electronic Comparative Linguistics Research Institute上海高校比较语言学E-研究院. Accessed (15.08.2011).

Shi Lin (石林). 1997. *Dong-Taiyu bijiao yanjiu.* 《侗台语比较研究》 *[Comparative studies in Kam-Tai languages.]* Tianjin: Tianjin Ancient Documents Publishing House.

Shuiyu Diaocha Baogao (SDB)《水语调查报告》[Sui Language Research Report]. 1956. Unpublished ms. Guiyang, Guizhou: Zhongguo Shaoshu Minzu Yuyan Diaocha Diyi Gongzuodui 中国少数民族语言调查第一工作队 [Chinese Minority Language Research Working Group].

Stanford, James N. 2007. Dialect contact and identity: A case study of exogamous Sui clans. PhD dissertation, Michigan State University.

—— 2008. A sociotonetic analysis of Sui dialect contact. *Language Variation and Change* 20(3):409-450.

—— 2009. "Eating the food of our place": Sociolinguistic loyalties in multidialectal Sui villages. *Language in Society* 38(3):287-309.

—— in prep. A 50-year real-time comparison of regional dialect variation in the Sui language. ms.

Thurgood, Graham. 1988. Notes on the Reconstruction of Kam-Sui. In Edmondson and Solnit (eds.) 1990.

Weng Jialie (翁家烈) (ed.) 2001. *Guizhou sheng zhi: Minzu zhi.* 《贵州省志·民族志》 *[Guizhou provincial almanac: minorities almanac.]* Guiyang: Guizhou Nationalities Publishing House.

Wei, James Xuecun and Jerold A. Edmondson. 2003. *Sui, Chinese, Thai, English dictionary*. Bangkok: Mahidol University Press.

—— 2008. Sui. In Diller, Edmondson and Luo (eds. 2008).

Wu Wenyi (伍文义), Wil C. Snyder (辛维) and Liang Yongshu (梁永枢). 2007. *Survey of the Guizhou Bouyei language*. SIL Language and Culture Documentation and Description 2007-001. Dallas: SIL International. Online. URL: http://www.sil.org/silepubs/abstract.asp?id=49147. Accessed on 16.08.2011.

Xia Yongliang (夏永良). 1989. Libo *xian Dongliu Shuiyu diaocha baogao.* 《荔波县洞流水语调查报告》 [*A survey report of Dongshui Sui, Libo county.*] In Guizhou Minority Affairs Commission et al. (2008).

—— 1992. Sandong Shuiyu de yinwei yu yufa. 《三洞水语的音位与语法》 [The phonology and grammar of Sandong Sui.] in Guizhou Minority Affairs Commission et al. (2008).

Zeng Xiaoyu (曾晓渝). 1994. *Hanyu Shuiyu guanxici yanjiu.*《汉语水语关系词研究》 *[Study of Mandarin-Sui related words.]* Chongqing: Chongqing Publishing House.

—— 2004. Hanyu Shuiyu guanxi lun—Shuiyu li Hanyu jieci ji tongyuanci fenceng yanjiu.《汉语水语关系论—水语里汉语借词及同源词分层研究》 [Discussion of the relationship between Mandarin and Sui—a stratified study of Mandarin loanwords and cognates in Sui.] Beijing: Commercial Press.

—— 2010. Dongtai Miaoyao yuyan de Han jieci yanjiu. 《侗台苗瑶语言的汉借词研究》 [The study of Chinese loanwords in Kam-Tai and Miao-Yao languages.] Beijing: Commercial Press.

Zeng Xiaoyu (曾晓渝) and Yao Fuxiang (姚福祥). 1996. *Han-Shui cidian.*《汉水词典》 *[Chinese-Sui dictionary.]* Chengdu: Sichuan Nationalities Publishing House.

Zhang Junru (张均如). 1980. *Shuiyu jianzhi.* 《水语简志》 *[A Sketch of the Sui Language.]* Beijing: People's Publishing House.

Zhang Zhenjiang (张振江). 2008. Libo Shuiyao Shuizu: Guizhou Libo Shuiyao xiang diaocha yu yanjiu. 《荔波水尧水族：贵州荔波水尧乡调查与研究》 [The Sui of Shuiyao: Survey and research of Shuiyao township, Libo, Guizhou.] Beijing: Intellectual Property Rights Publishing House.

Appendix A. Notes on data sources

NL, PY, RJ: 314 words, listed in order to highlight certain sound changes. Li Fang-kuei only records supposed cognates for each gloss and gives very few alternative forms. On the few occasions when he does, he specifies neither semantic distinctions nor differences in usage.

SN, YK: Over 1,800 words, collected in 2001 and 2002. YK data was elicited from four speakers, all male and over 65 years of age. SN data was elicited from two speakers, one male, 74 years old, and one female, 50 years old. Most speakers recorded were either government officials or teachers. Two alternatives are often given for the same gloss, with no indication of semantic differences or usage patterns.

SQ: Around 2,400 words with lengthy English definitions. This data appears virtually identical to that given in CNU (ed.) (1985). The informant, Yao Fuxiang (姚富祥), was a Sui intellectual who was well acquainted with "standard Sui" (i.e. Central Sui as spoken in Sandong district) pronunciation[43], thus for some words he lists two pronunciations, one for his own village and one for Sandong (although he does not indicate which word belongs to which dialect).

ND: 4,480 words with narrow definitions, collected from a speaker from Liuzhai district. The Sui in Liuzhai migrated from Sandu county in the 1930s and 40s (GZARMLC 2008:780). Their use of [ɣai²] for "1S I, me" and a low rising tone for Tone 6 indicates that this community originated from south-west of Sandong in the present-day Tingpai ("Pyo") area.

TP: Around 500 words collected from an 18 year old girl born and raised in a village near Tingpai township. PY and ND data indicate that her speech is virtually identical to older speakers in terms of phonetic features.

JQ: Over 1,000 words collected from a 30 year old man born and raised in a village east of Jiuqian township. His speech is typical of older speakers and does not exhibit any sound changes (such as ɕ- >hj-) observed among younger speakers in his village (including his own sister).

SY: Around 600 words collected from a 78 year old man born and raised in a village next to Shuiyao township. While living in this village for three months, the author observed that speakers under the age of 40 almost uniformly exhibit the following regular sound changes: nd- > l-; mb- > v-; hw- > f-; ʔb- > ʔm-; and ɕ- > çj- or hj-.

JR: Around 600 words collected from four male speakers, aged 18, 24, 41 and 71, born and raised in Laliang village, about 3km from Jiarong township. The two youngest speakers exhibited the following sound changes: nd- > l-; hw- > f-; tsj- > ʈ-; and sj- > ɕ-. This particular dialect was unique in having a voiceless lateral fricative, ɬ-, instead of the usual pre-palatal ɕ- observed in other locations. A speaker who knows the area well said that over half of Sui speakers in Jiarong district pronounce ɕ- as ɬ-. Thus younger speakers' sj- > ɕ- is part of a mini chain shift (ɕ- > ɬ-, sj- >ɕ).

---

[43] Indeed, he co-authored a Sui dictionary based on Sandong dialect (Zeng & Yao 1996).

Appendix B. Other diachronic variation in Sui varieties.

The data considered for this study show some additional variation in the realisation of certain proto-initials but none of them are consistent or widespread enough to suggest a dialect split other than Central/Southern. In brief, we note that:

1. PKS *ʔ- is usually retained across the board, although on occasion it acquires a velar nasal, for example in Shuiyao and Shuiqing *ʔu:m³ 'to hold (a baby)' > [ʔŋʊm³]. Libo and Sandu dialects of southwestern Mandarin often pronounce Late Middle Chinese (and modern standard Mandarin) initial [ʔ] as[ŋ], e.g. LMC *ʔan¹ 安 'peace' > [ŋan³³] and LMC *ʔaj⁴ 爱 'to love' > [ŋai²¹] (see Zeng 2010:43-47);

2. All prenasalised voiced stops have merged with preglottalised stops in Nandan;

3. The phoneme which is pronounced [ɕ] in most dialects is transcribed as [hj] (or just [h] before -i) in Rongjiang and Nandan. Thurgood reconstructs this as PKS *hj- (for example *hjit⁷ 'morning'). Interestingly, the author has observed that while older speakers in Shuiyao and Jiuqian retain a clear pre-palatal [ɕ] in these words, younger speakers tend to pronounce the same sound as [çj], [xj] or [hj] (auditorily these are all very close and the relative lack of friction often makes it difficult to distinguish between them), suggesting a ɕ > hj sound change, contrary to directionality of change which would indicate hj > ɕ;

4. There are some instances of alternation between ɕ- and lj- (l-). For example: [ɕuŋ¹] (Central Sui) 'to boil' is pronounced as [ɬjɔ:ŋ¹] (JR) or [ljɔ:ŋ¹] (JQ, SY); [ɕa³] (SD) 'daughter-in-law' is pronounced [ɬa³] (JR), [le³] (ND) or [li:ɛ³] (TP); and [ɕo³] (SD) 'very (post-adjectival intensifier)' is pronounced [ljɔ³] (TP);

5. In most cases, the lateral in *phl- has become a palatal e.g. PKS *phla:t⁷ 'blood' > [pʰja:t⁷]; PKS *phlai⁵ 'near, close'> [pʰjai⁵]. Further examples can be seen in [pʰja:m¹] 'to disappear', [pʰja:u¹] 'to warm (by a fire)' and [pʰjuŋ¹] 'steam' (although they were not reconstructed by Thurgood). There is, however, sporadic deletion of the [j], e.g. PKS *phla:t⁷ 'blood' > [pʰa:t⁷] (SQ, SY, JQ); PKS *phlai⁵ 'near, close'> [pʰɛi⁵] (TP); [pʰja:m¹] 'to disappear' (most dialects) is pronounced as [pʰa:m¹] in SY; and [pʰjʊŋ¹] 'steam' (ND, TP) is pronounced as [pʰɔ:ŋ¹] in JQ, SY and JR.

Finally, there are some regular sound changes specific to Tingpai, and others specific to Jiuqian and Jiarong, which show they have diverged slightly from Central Sui and Southern Sui respectively. In particular:

1. PKS *mpr-, which becomes mbj- in most varieties (while retaining the original tone category), undergoes lenition in TP and PY, becoming mj-, e.g. *mpra:ŋ¹ > [mjaŋ²], *mpra¹ > [miɛ²]. In these cases, it appears that the fully voiced onset has resulted in a switch of tone category (possibly indicating that the change in initial occurred before the voiced onset induced tone split);

2. The deletion of [w] in PKS *ɓw- (*ɓwa:ŋ¹ 'thin, flat' > [ʔba:ŋ¹]) results in a merger with *ɓ- (also > ʔb-) in most Sandong Sui varieties. In TP, however (but not in PY), this seems to have initiated a mini-chain shift, with *ɓ- consistently becoming ʔm-, for example *ɓa:n³ 'village' > [ʔma:n³] (TP), [ʔba:n³] (elsewhere); *ɓun⁵ 'well' > [ʔmən⁵] (TP), [ʔbən⁵] (elsewhere); and *ɓun¹ 'sky' > [ʔmən¹] (TP), [ʔbən¹] (elsewhere).

3. In most varieties of Sandong Sui, the voicing from both the rhotic release in PKS *thr-/*tr- and the lateral release in PKS *thl- have been transferred to the beginning of the word in the form of a nasal (some sort of metathesis), for example: *thram^5 'low, short' > [^{n}dam^5]; *tra:i^5 'dry field' > [^{n}da:i^5]; *thla^1 'eye' > [^{n}da^1] (this is an unconditional merger). In Jiuqian and Jiarong, only the lateral has undergone this metathesis, whereas the rhotic release in *thr-/*tr- has been deleted entirely, thus *thram^5 'low, short' > [tam^5] (JQ, JR); *tra:i^5 'dry field' > [ta:i^5] (JQ, JR); but *thla^1 'eye' > [^{n}da^1] (JQ, JR). Thus *thr- and *tr- have merged entirely with *t- and *d-. The same is true across the board for palatalised alveolar stops, which emerged from the same series of initials (how this came about is unclear). A full list of correspondences for this significant sound change is given in Table 25. Neither Thurgood (1988) nor Zeng (1994, 2004) offer an explanation for the development of palatalised alveolar onsets in Sui.[44] Data presented here indicates that palatalised alveolar onsets can reasonably be reconstructed for Proto-Sui.

---

[44] Zeng (1994:35, 37, 40) does raise the possibility of PS palatalised alveolar initials *dj-, *^{n}dj- and *^{n}tj-. She decides not to reconstruct them due to inconsistency across the lects between ^{n}d- and ^{n}dj-. However, her data are extremely limited (and she is not helped by the fact that the words [^{n}djeu^1] 'we (inclusive)' and [^{n}da:u^1] 'we (exclusive)' are mixed up in her data under the same gloss).

Table 25: *Correspondences for PKS *thr-, *tr-, *thl-, with *d- and *t- for comparison. Words with voiceless alveolar stops are shaded in grey. Words with palatalised alveolar stops are framed in double lines.*

Gloss	PKS	PS	Central Sui			Southern Sui				
			ND	SD	TP	SQ	JR	JQ	SY	SN
'low, short'	*thram^5	*^{n}t-	ʔdam^5	^{n}dam^5	^{n}dɐm^5	^{n}dam^5	tɐm^5	tɐm^5	^{n}dɐm^5	^{n}dam^5
'body'	*thrun^1	*^{n}t-	ʔdən^1	^{n}dən^1	-	^{n}dən^1	tən^1	tən^1	^{n}dən^1	^{n}dən^1
'hot'	-	*^{n}t-	ʔdu^3	^{n}du^3	^{n}du^3	^{n}do^3	tu^3	tu^3	^{n}du^3	^{n}du^3
'correct, yes'	-	-	-	^{n}dum^3	^{n}dʊm^3	-	tʊm^3	tom^3	^{n}dəm^3	-
'locust'	*thrak^7	*^{n}t-	-	ʔdjak^7	^{n}djak^7	ʔdjak^7	tjɐk^7	tjɐk^7	^{n}djɐk^7	ʔdjak^7*
'short (length)'	*thrin^3	*^{n}t-	ʔdjən^3	^{n}djən^3	^{n}djen^3	^{n}din^3	tɪn^3	tjɪn^3	^{n}din^3	^{n}din^3
'dry field'	*traːi^5	*^{n}t-	ʔdaːi^5	^{n}daːi^5	-	ʔdaːi^5	taːi^5	taːi^5	^{n}daːi^5	^{n}daːi^5
'we (incl.)'	*traːu^1	*^{n}t-	ʔdaːu^1	^{n}daːu^1	^{n}daːu^1	^{n}daːu^1	taːu^1	taːu^1	^{n}daːu^1	^{n}daːu^1
'firewood'	*trit^7	-	ʔdjət^7	^{n}djət^7	-	^{n}dit^7	tjɪt^7	tjət^7	^{n}djɪt^7	^{n}djət^7
'to buy'	*trai^3	*^{n}t-	ʔdjai^3	^{n}djai^3	^{n}djɛi^3	^{n}djai^3	tjɛi^3	tjɛi^3	^{n}djɛi^3	^{n}djai^3
'chest (body)'	*tak^7	-	tak^7	tak^7	tak^7	tak^7	tɐk^7	tɐk^7	tɐk^7	tak^7
'to weave'	*tam^3	*t-	tam^3	tam^3	tɐm^3	tam^3	tɐm^3	tɐm^3	tɐm^3	tam^3
'small bowl'	*duːi^4	*d-	tui^4	tui^4	tuːi^4	tui^4	tui^4	tui^4	tui^4	tui^4
'to pass'	*da^6	*d-	ta^5	ta^6	ta^6	ta^6	ta^6	ta^6	ta^6	ta^6
'bamboo hat'	-	*t-	tjum^1	tjum^1	tjʊm^1	tim^1	tjʊm^1	tjom^1	tjɪm^1	tjum^1
'long time'	-	-	ɬaːŋ^1	-	tjeŋ^1	-	tjeŋ^1	tjeŋ^1	tjeŋ^1	tjaŋ^1
'full, satiated'	-	*t-	tjaŋ^5	tjaŋ^5	tjeŋ^5	tjaŋ^5	tjeŋ^5	tjeŋ^5	tjeŋ^5	tjaŋ^5
'pond'	*thlam^1	*^{n}t-	-	^{n}dam^1	-	^{n}dam^1	^{n}dɐm^1	^{n}dɐm^1	^{n}dɐm^1	^{n}dam^1
'eye'	*thla^1	*^{n}t-	ʔda^1	^{n}da^1	^{n}da^1	^{n}da^1	^{n}da^1	^{n}da^1	^{n}da^1	^{n}da^1
'fragrant'	-	-	ʔdaːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1	^{n}daːŋ^1
'thorn'	-	*^{n}t-	ʔdun^1	^{n}dun^1	^{n}dʊn^1	^{n}dun^1	^{n}dʊn^1	^{n}dʊn^1	^{n}dun^1	^{n}dun^1
'to see'	-	*^{n}t-	ʔdo^3	^{n}do^3	^{n}dɔ^3	^{n}do^3	^{n}dɔ^3	^{n}do^3	^{n}dɔ^3	^{n}do^3
'we (excl.)'	-	-	ʔdiu^1	^{n}diu^1	^{n}diɛu^1	^{n}djeu^1	^{n}deu^1	^{n}diu^1	^{n}djeu^1	^{n}diu^1
'to scrub'	-	-	-	-	^{n}djen^1	-	^{n}djɛn^1	^{n}djɛn^1	-	-
'dark'	-	-	niŋ^5	^{n}dəŋ^5	^{n}dɪŋ^5	-	-	^{n}djəŋ^5	^{n}diŋ^5	-

* Li (1965) gives [^{n}djak^7] for LN.

SERIAL VERB CONSTRUCTIONS AND MODAL DOUBLING IN HONG KONG HOKKIEN

Yin Ling Cheung
Nanyang Technological University
<yinling.cheung@nie.edu.sg>

Abstract
This is the first study that investigates Modal Doubling in Hong Kong Hokkien. I will argue that Modal Doubling exists only with resultative serial verb construction. Specifically I will discuss how a Larsonian structure for resultative serial verb construction can be adopted to facilitate the presence of Modal Doubling. My study will address the following questions: Do two modal elements involve in Modal Doubling? How many modality readings are allowed? Is e^7 the only modal element that allows doubling? Does Modal Doubling only appear with serial verb constructions? If yes, which type of serial verb constructions co-exists with Modal Doubling?

Keywords: Hokkien, serial verb construction, modals

Introduction
This discussion concerns modal doubling in Hong Kong Hokkien. Hokkien, also known as Southern Min, is a language spoken in southern China, namely, in Fujian province. It belongs to Sino-Tibetan language family. As a result of migration, Hokkien is also used among Chinese in many Asian cities / countries such as Hong Kong, Indonesia, Malaysia, Philippines, Singapore, and Taiwan. This paper only focuses on the Hokkien variety spoken in Hong Kong. The number of speakers is approximately 1 million. Hong Kong Hokkien belongs to the variant spoken in Xiamen, a subgroup of Southern Min. At issue is a construction which consists of two instances of the same modal which I will argue exists only with resultative serial verb construction. Modal Doubling is an interesting problem in Hong Kong Hokkien because doubling does not exist in other Southern Min languages. All the examples in this paper are taken from a Hong Kong-born Chinese who has Hokkien as her first language. In this paper I will specifically discuss how a Larsonian structure for resultative serial verb construction can be adopted to facilitate the presence of Modal Doubling in Hokkien.

What is Modal Doubling?
It is a special feature in Hong Kong Hokkien, meaning the speaker is 100% certain that the subject will do the action. In this paper I will show that such construction is restricted to one single type of modal element e^7 and is dependent on co-occurrence with resultative

serial verb construction. Example (1) illustrates Modal Doubling in Hong Kong Hokkien. e^7 is the modal element being doubled.[45]

(1) *i ^{1}* *e ^{7}* *pa ^{8}* *e ^{7}* *si ^{1}* *he ^{8}* *chia ^{5}* *ga ^{4}-chua ^{5}*
     Prn.3s MOD1 hit (V1) MOD2 die (V2) DEM CL cockroach
     'S/he will kill this cockroach.'
     (The speaker is 100% certain that the subject will do the action.)

Research Questions
In this section I will describe the properties of Modal Doubling by addressing the following questions:

  i.   Do two modal elements involve in Modal Doubling? How many modality readings are allowed?
  ii.   Is e^7 the only modal element that allows doubling?
  iii. Does Modal Doubling only appear with serial verb constructions? If yes, which type of serial verb constructions co-exists with Modal Doubling?

    As indicated in example (1), Modal Doubling is compatible with the resultative type of serial verb construction. Hong Kong Hokkien has seven different serial verb constructions, which will be discussed in section 3. In (2), the predicate comprises two verbs and the construction yields a single event of interpretation, that is, the subject will lift this box of books. There is no coordinating or subordinating element intervening the verbs. The predicate uses one tense and shares one argument.
    None of other serial verb constructions (direct object sharing, directional, instrumental, double object, causative, sequential) allow the presence of Modal Doubling. This is to say, it cannot interact with serial verb in the types such as direct object sharing (3), directional (4), instrumental (5), double object (6), causative (7), and sequential (8).

Resultative

(2) *i ^{1}* *e ^{7}* *dao ^{5}* *e ^{7}* *ki ^{8}* *ji ^{8}-ʌ ^{4}-zi ^{4}*
     Prn.3s MOD1 take MOD2 rise DET-CL-books
     'S/he will lift this box of books.'
     (The speaker is 100% certain that the subject can do the action.)

Direct object sharing

(3) * *i ^{1}* *e ^{7}* *zi ^{5}* *e ^{7}* *mʌ ^{8}* *tsia ^{8}*
     Prn.3s MOD1 cook MOD2 meat eat
     Intended: 'S/he will cook and eat meat.'

---

[45] The Hokkien text could be traced back to the 16th century. One example is the *Doctrina Christiana en letra y lengua China* written by the Spanish Dominicans in the Philippines after 1587. Another example is a script of a play written in Ming Dynasty called *Romance of the Lychee Mirror* (1566 AD), one of the earliest Southern Min colloquial texts. Xiamen University has developed a romanization system based on Pinyin, which has been published in a dictionary called *Minnan Fangyan Putonghua Cidian* (閩南方言普通話詞典), which is used to teach the language to non-native speakers of Chinese. This romanization system is known as Pumindian. This system is used in this paper.

Directional

(4)* *i¹*   *e⁷*   *mui⁵*   *hə²*   *e⁷*   *lai²*
    Prn.3s MOD1 buy   fish   MOD2 come
    Intended: 'S/he will buy a fish and bring it.'

Instrumental

(5)* *i¹*   *e⁷*   *te¹*   *do¹*   *e⁷*   *tai¹*   *he²*
    Prn.3s MOD1 take   knife   MOD2 cut-up   fish
    Intended: 'S/he will cut up a fish with a knife.'

Double object

(6)* *i¹*   *e⁷*   *kia¹*   *ji⁸*   *bun⁴*   *e⁷*   *tse²*   *hi² gua³*
    Prn.3s MOD1 send   NUM.ONE   CL   MOD2 book   giveprn.1s
    Intended: 'S/he will send a book to me.'

Causative

(7)* *i¹*   *e⁷*   *hi⁸*   *ngaan⁷-ah¹*   *e⁷*   *kia⁵*   *tsau⁶*
    Prn.3s MOD1 make   children   MOD2 fear   bird
    Intended: 'S/he will make the children afraid of the bird.'

Sequential

(8)* *i¹*   *e⁷*   *lip⁸*   *he⁸-yi⁴*   *e⁷*   *kua¹*   *he⁸*
    Prn.3s MOD1 enter   cinema   MOD2 see   movie
    Intended: 'S/he will go into the cinema to see a movie.'

These facts suggest that the resultative type of serial verb construction plays an important role in Modal Doubling. I will, in the next section, explain why the resultative verb construction presents a unique structural configuration which distinguishes it from direct object sharing, directional, instrumental, double object, causative and sequential serial verb constructions.

Serial Verb Constructions in Hong Kong Hokkien
This section considers some constructions in Hong Kong Hokkien involving two verbs (V1 and V2), which could be taken to be the serial verb constructions. Seven types of serial verb constructions are identified, namely, direct object sharing, resultative, directional, instrumental, double object, causative, and sequential. An example of each type is shown in (9) – (15).

(9)   *i¹*   *dao⁵*   *ji⁸-ʌ⁴-zi⁴*   *ki⁸*   (resultative)
    Prn.3s take   DET-CL-books   rise
    'S/he lifts this box of books.'

(10)   *i¹*   *zi⁵*   *mʌ⁸*   *tsia⁸*   (direct object sharing)
       Prn.3s   cook   meat   eat
       'S/he cooks and eats meat.'

(11)   *i¹*   *mui⁵*   *hə²*   *lai*   (directional)
       Prn.3s   buy   fish   come
       'S/he buys a fish and brings it.'

(12)   *i¹*   *te¹*   *do¹*   *tai¹*   *he²*   (instrumental)
       Prn.3s   take   knife   cut-up   fish
       'S/he cuts up a fish with a knife.'

(13)   *i¹*   *kia¹*   *ji⁸*   *bun⁴* *tse²*   *hi²*   *gua³*   (double object)
       Prn.3s   send   NUM.ONE CL   book   give   prn.1s
       'S/he sends a book to me.'

(14)   *i¹*   *hi⁸*   *ngaan⁷-ah¹*   *kia⁵*   *tsau⁶*   (causative)
       Prn.3s   make   children   fear   bird
       'S/he makes the children afraid of the bird.'

(15)   *i¹*   *lip⁸*   *he⁸-yi⁴*   *kua¹*   *he⁸*   (sequential)
       Prn.3s   enter   cinema   see   movie
       'S/he goes into the cinema to see a movie.'

A-not-A question formation test helps us detect that different types of serial verb constructions in Hokkien fall into two groups: group A – resultative, direct object sharing, and directional serial verb constructions – that does not allow A-not-A question formation on V2 and; group B – instrumental, double object, causative, and sequential serial verb constructions – that allows A-not-A question formation on V2.

First, I will demonstrate A-not-A question formation test on group A type of serial verb constructions, namely, resultative, direct object sharing, and directional.

(16) shows how the verbs *dao⁵* 'take' (V1) and *ki⁸* 'rise' (V2) behave with respect to A-not-A question formation.

Resultative
(16) a. *i¹*   *dao⁵-m⁷-dao⁵*   *ji⁸-ʌ⁴-zi⁴*   *ki⁸*
         Prn.3s   take-NEG-take   DET-CL-books   rise
         'Will she lift this box of books?'

     b. * *i¹*   *dao⁵*   *ji⁸-ʌ⁴-zi⁴*   *ki⁸-m⁷-ki⁸*
          Prn.3s   take   DET-CL-books   rise-NEG-rise
          [No reading available]

     c. * *i¹*   *dao⁵-m⁷-dao⁵*   *ji8-ʌ⁴-zi⁴*   *ki⁸-m⁷-ki8*
          Prn.3s   take-NEG-take   DET-CL-books   rise-NEG-rise
          [No reading available]

(17) shows how the verbs *zi5* 'cook' (V1) and *tsia8* 'eat' (V2) behave with respect to A-not-A question formation.

Direct object sharing

(17) a. i¹   zi⁵-m⁷-zi⁵   mʌ⁸   tsia⁸
         Prn.3s   cook-NEG-cook meat   eat
         'Will s/he cook the meat (instead of the vegetable) to eat?'

     b. * *i ¹*   *zi ⁵*   *mʌ ⁸*   *tsia ⁸-m ⁷-tsia ⁸*
         Prn.3s   cook   meat   eat-NEG-eat
         [No reading available]

     c. * *i ¹*   *zi ⁵-m ⁷-zi ⁵*   *mʌ ⁸*   *tsia ⁸-m ⁷-tsia ⁸*
         Prn.3s   cook-NEG-cook meat   eat-NEG-eat
         [No reading available]

     (18) shows how the verbs *mui⁵* 'buy' (V1) and *lai²* 'come' (V2) behave with respect to A-not-A question formation.

Directional

(18) a. *i ¹*   *mui ⁵-m ⁷-mui ⁵*   *hə ²*   *lai ²*
         Prn.3s   buy-NEG-buy   fish   come
         'Will s/he BUY fish and bring it?'

     b. * *i ¹*   *mui ⁵*   *hə ²*   *lai ²-m ⁷-lai ²*
         Prn.3s   buy   fish   come-NEG-come
         [No reading available]

     c. * *i ¹*   *mui ⁵-m ⁷-mui ⁵*   *hə ²*   *lai ²-m ⁷-lai ²*
         Prn.3s   buy-NEG-buy   fish   come-NEG-come
         [No reading available]

     To sum up the results in (16), (17) and (18), group A type of serial verb constructions does not allow A-not-A question formation on V2 (i.e. 16b, 17b, and 18b) and on both V1 and V2 (i.e. 16c, 17c, and 18c).
     Next, I will demonstrate A-not-A question formation test on group B type of serial verb constructions, namely, instrumental, double object, causative, and sequential. Group B type of serial verb constructions allows A-not-A question formation on V2.
     (19) shows how the verbs *te1* 'take' (V1) and *ta1* 'cut-up' (V2) behave with respect to A-not-A question formation. This question formation can target V1, as in (19a), and V2 in (19b), thus identifying it as the head of the clause, not with both V1 + V2 in (19c).

<u>Instrumental</u>

(19) a. *i¹*   *te¹-m⁷-te¹*   *do¹*   *tai¹*   *he²*
      Prn.3s   take-NEG-take   knife   cut-up   fish
      'Will s/he use a knife (instead of a pair of scissors) to cut up a fish?'

    b. *i¹*   *te¹*   *do¹*   *tai¹-m⁷-tai¹*   *he²*
      Prn.3s   take   knife   cut-up-NEG-up   fish
      'Will s/he use a knife to cut up a fish (instead of a chicken)?'

    c. * *i¹*   *te¹-m⁷-te¹*   *do¹*   *tai¹-m⁷-tai¹*   *he²*
      Prn.3s   take-NEG-take   knife   cut-up-NEG-up   fish
      [No reading available]

    (20) shows how the verbs *kia¹* 'send' (V1) and *hi¹* 'give' (V2) behave with respect to A-not-A question formation.

<u>Double object</u>

(20) a. *i¹*   *kia¹-m⁷-kia¹*   *ji⁸*   *bun⁴*   *tse²*   *hi²*   *gua³*
      Prn.3s   send-NEG-send   NUM.ONE   CL   book   give   prn.1s
      'Will she SEND a book to me?'

    b. *i¹*   *kia¹*   *ji⁸*   *bun⁴*   *tse²*   *hi²-m⁷-hi²*   *gua³*
      Prn.3s   send   NUM.ONE   CL   book   give-NEG-give   prn.1s
      'Will she send a book to ME?'

    c. * *i¹*   *kia¹-m⁷-kia¹*   *ji⁸*   *bun⁴tse²*   *hi²-m⁷-hi²*   gua³
      Prn.3s   send-NEG-send   NUM.ONE CL book   give-NEG-give   prn.1s
      [No reading available]

    (21) shows how the verbs *hi⁸* 'make' (V1) and *kia⁵* 'fear' (V2) behave with respect to A-not-A question formation.

<u>Causative</u>

(21) a. *i¹*   *hi⁸-m⁷-hi⁸*   *ngaan⁷-ah¹*   *kia⁵*   *tsau⁶*
      Prn.3s   make-NEG-make   children   fear   bird
      'Will s/he MAKE the children afraid of the bird?'

    b. *i¹*   *hi⁸*   *ngaan⁷-ah¹*   *kia⁵-m⁷-kia⁵*   *tsau⁶*
      Prn.3s   make   children   fear-NEG-fear   bird
      'Will s/he make the children AFRAID of the bird?'

    c. * *i¹*   *hi⁸-m⁷-hi⁸*   *ngaan⁷-ah¹*   *kia⁵-m⁷-kia⁵*   *tsau⁶*
      Prn.3s   make-NEG-make children   fear-NEG-fear   bird
      [No reading available]

(22) shows how the verbs lip^8 'enter' (V1) and kua^1 'see' (V2) behave with respect to A-not-A question formation.

<u>Sequential</u>

(22) a. *i^1 $\underline{lip^8\text{-}m^7\text{-}lip^8}$ $he^8\text{-}yi^4$ kua^1 he^8*
       Prn.3s   enter-NEG-enter cinema see   movie
       'Will s/he GO INTO THE CINEMA to watch a movie?'

   b *i^1 lip^8 $he^8\text{-}yi^4$ $\underline{kua^1\text{-}m^7\text{-}kua}$ he^8*
       Prn.3s   enter   cinema see-NEG-see   movie
       'Will s/he go into the cinema to WATCH A MOVIE?' (instead of falling asleep in the cinema.)

   c.* *i^1 $\underline{lip^8\text{-}m^7\text{-}lip^8}$ $he^8\text{-}yi^4$ $\underline{kua^1\text{-}m^7\text{-}kua^1}$ he^8*
       Prn.3s   enter-NEG-enter cinema see-NEG-see   movie
       [No reading available]

Data analysis

Let's move on to analyze Hokkien serial verb constructions by adopting the proposals from Law (1996) and Larson (1991). Law (1996:200-1), in his analysis of Mandarin, presents two types of serial verb constructions, as in (23) and (24). Examples (25a-b) illustrate the structure in (23); (26a-b) are the examples using the structure in (24). Please note that (25) and (26) are adopted from Law (1996:200). When it comes to Hong Kong Hokkien, I suggest that serial verb constructions in group A (resultative, direct object, directional) belong to the structure in (23) and those in group B (instrumental, double object, causative, sequential) belong to the structure in (24).

(23)   NP1 [$_{VP}$ V1 NP2 [$_{VP}$ V2]]

(24)   NP1 [$_{VP}$ [$_{VP}$ V1 NP2 [$_{VP}$ V2 NP3]]

(25) a. *Ta   song-le   yi-   ge   xiangzi lai*
        Prn.3s   send-PERF   NUM.ONE   CL   suitcase come
        'He sent over a suitcase.'

   b. *Ta   na-le   nei-ben   shu   zou*
        Prn.3s   hold-PERF   DEM-CL   book   go
        'He took away that book.'

(26) a. *Ta   na   dao   qie-le   rou*
        Prn.3s   hold   knife   cut-PERF   meat
        'He cut the meat with a knife.'

   b. *Ta   na   yaoshi   kai-le   men*
        Prn.3s   hold   key   open-PERF   door
        'He opened the door with a key.'

The structure in (23) suggest that the first verb takes as complement the VP headed by the second verb such as *lai* 'come' *and zou* 'go'. In (24), the VP headed by the second verb such as *qie* 'cut' and *kai* 'open' in (26) is an adjunct to the VP headed by the first verb. V2 can move to V1 producing the structure of [NP1 V1 V2 NP2] in the case of (23) but not in (24). When the trace left behind by movement of the second verb to the position immediately following the first verb is properly governed by the first verb, Empty Category Principle is not violated (Law, 1996, p. 205). By contrast, in (24) with the second VP as an adjunct to the first VP, the movement of the second verb to the position immediately following the first verb is prohibited. The order in [NP1 V1 V2 NP2] is possible in Hokkien for group A type of serial verb constructions as shown in (27) – (29).

(27) i^1 dao^5 <u>ki^8</u>t ji^8-Λ^4-zi^4 t (resultative)
Prn.3s   take   rise   DET-CL-books
'S/he lifts this box of books.'

(28) i^1 zi^5 <u>$tsia^8$</u>t $m\Lambda^8$ t (direct object sharing)
Prn.3s   cook   eat   meat
'S/he cooks and eats meat.'

(29) i^1 mui^5 <u>lai^2</u>t $h\partial^2$ t (directional)
Prn.3s   buy   come   fish
'S/he buys a fish and brings it.'

One possible weakness with regard to Law's (1996) proposal is that he seems to suggest a tripartite structure consisting of V1, NP2, and V2. On the other hand, Larson (1991, p. 202) gives a more convincing explanation by suggesting "when NP receives a thematic role from a predicate, the two must appear as sisters and form a constituent that is itself sister to NP." Larson (1991, pp. 201-2), drawing data from English, suggests that secondary predicates are daughters of V', appearing in the configuration shown in (30), which is a resultative structure.

(30)   VP   (Larson 1991:202)
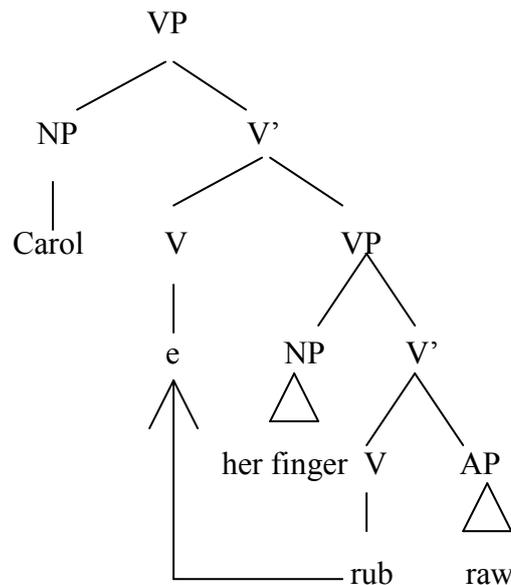
Under Larson's proposal, in (30), *her finger* receives a theta-role from *rub* and a theta-role from *raw*. *Rub* and *raw* therefore appear as sisters under a V' which is predicated of her *finger*. Here I will explain how a Larsonian structure for resultative serial verb construction can be adopted to facilitate the presence of Modal Doubling in Hokkien.

Recall example (1), only resultative serial verb construction is compatible with Modal Doubling. Resultative serial verb construction presents a unique structural configuration which distinguishes it from the other serial verb constructions, in particular, direct object sharing and directional, which both belong to group A type of construction. In what follows, I will give evidence supporting why the resultative serial verb construction is different from direct object sharing and directional serial verb constructions in terms of syntactic properties.

The resultative serial verb construction is different from direct object sharing and directional serial verb constructions in two ways: First, using the aspectual marking test, V1 in resultative serial verb construction does not allow the postverbal perfective marker le^8 to follow V1, while for both direct object sharing and directional serial verb constructions perfective marker *le8* can follow V1, as indicated in examples (31) – (33).

(31)   * *John*   *pa^8*(V1)   <u>*le^8*</u>   *ga^5-chua^5*   *si^1*(V2)   [resultative]
         John   hit   ASP   cockroach   die
         Intended: John has killed cockcroach.

(32)   *i^1*   *zi^5*(V1)   <u>*le^8*</u>   *m$_A$^{8}*   *tsia^8*(V2)   [direct object sharing]
         Prn.3s   cook   ASP   meat   eat
         'S/he eats meat she cooked.'

(33)   *i^1*   *mui^5*(V1)   <u>*le^8*</u>   *hə^2*   *lai^2* (V2)   [directional]
         Prn.3s   buy   ASP   fish   come
         'S/he bought a fish and brought it.'

Second, the bare noun test shows that the resultative serial verb construction is different from direct object sharing and directional serial verb constructions. Bare nouns with a weak existential interpretation are restricted to postverbal, that is, a position to the sister of V^0 and they undergo object shift. Bare nouns, in the position of NP2, in serial verb constructions in group A manifest a different behavior. They produce a grammatical result with direct object sharing and directional serial verb constructions, but ungrammatical with resultative serial verb construction. The fact that a bare noun NP2 in (35) and (36) in object sharing and directional serial verb constructions is grammatical indicates that the NP2 is in a position sister to V^0 in those cases. That is, they remain in the base position as object of V1. In the resultative serial verb construction in (34), the fact that a bare noun NP2 is ungrammatical indicates that it is not in the same structural position as object sharing and directional serial verb constructions, that is, NP2 cannot be the structural object (sister to the head) of V1.

(34)   * *John* (NP1)   *pa^8*(V1) <u>*ga^4-chua^5*</u>(NP2)*si^1*(V2)   [resultative]
         John   hit   cockroach   die
         Intended: John hits and kills cockroach.

(35) *John* (NP1) *zi⁵* (V1) *mʌ⁸* (NP2) *tsia⁸* (V2) [direct object sharing]
John cook meat eat
John cooks and eats meat.

(36) *John* (NP1) *mui⁵* (V1) *hə²* (NP2) *lai²* (V2) [directional]
John buy fish come
John buys and brings fish.

After reviewing the results in the aspectual marking test and the bare noun test, it is evident that resultative serial verb construction shows a unique configuration which distinguishes it from direct object sharing and directional serial verb constructions despite the fact that superficially these three constructions belong to group A type of structure as suggested earlier.

Using Larsonian Structure to Account for Resultative Serial Verb Construction
This section illustrates how the Larsonian structure can be adopted to explain resultative serial verb construction in Hokkien. The tree in (37) is an example of the construction with resultative serial verbs. In (37), we have VP as the sister to V^0, the first verb *dao⁵* 'take' takes the VP with the second verb *ki⁸* 'rise' as the complement.

(37)

```
                    VP
                  /    \
               NP₁      V'
                |      /   \
              John   V⁰     VP
                |         /    \
                |       NP₂     V
                |        △      |
              dao⁵  ji⁸-ʌ⁴-zi⁴  ki⁸
```

John take this-box-of-book rise
'John lifts this box of books.'
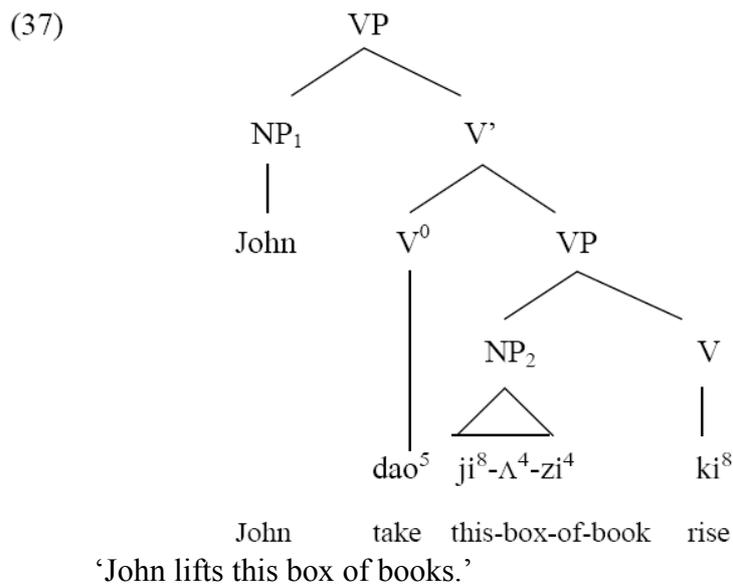
The structure in (37) allows two positions for the modal, as in (38):

(38)
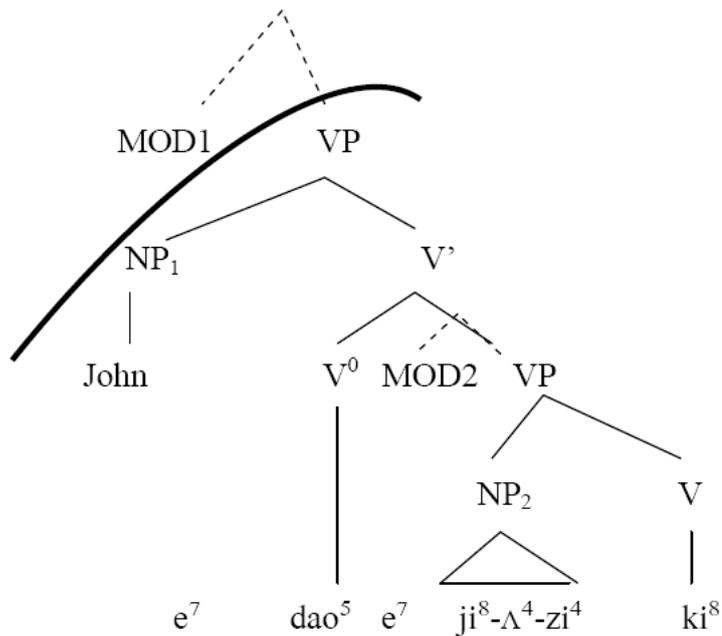

```
        John MOD       take MOD this-box-of-book  rise
```
'John will lift this box of books.'

   (39) demonstrates how the Larsonian structure is relevant to explain Modal Doubling with resultative serial verb construction in Hokkien. If we understand Modal Doubling as a relation between heads, MOD1 and MOD2 – either as the result of movement or the result of an agree operation – then the structure in (38) presents the right structural relation between MOD1 and MOD2, since MOD2 is in the checking (complement) domain of MOD1. Given that VP is a complement to V^0 and Modal Doubling is a result of head to head movement, this suggests that the two modal elements have to be the same, as depicted in (39).

(39)
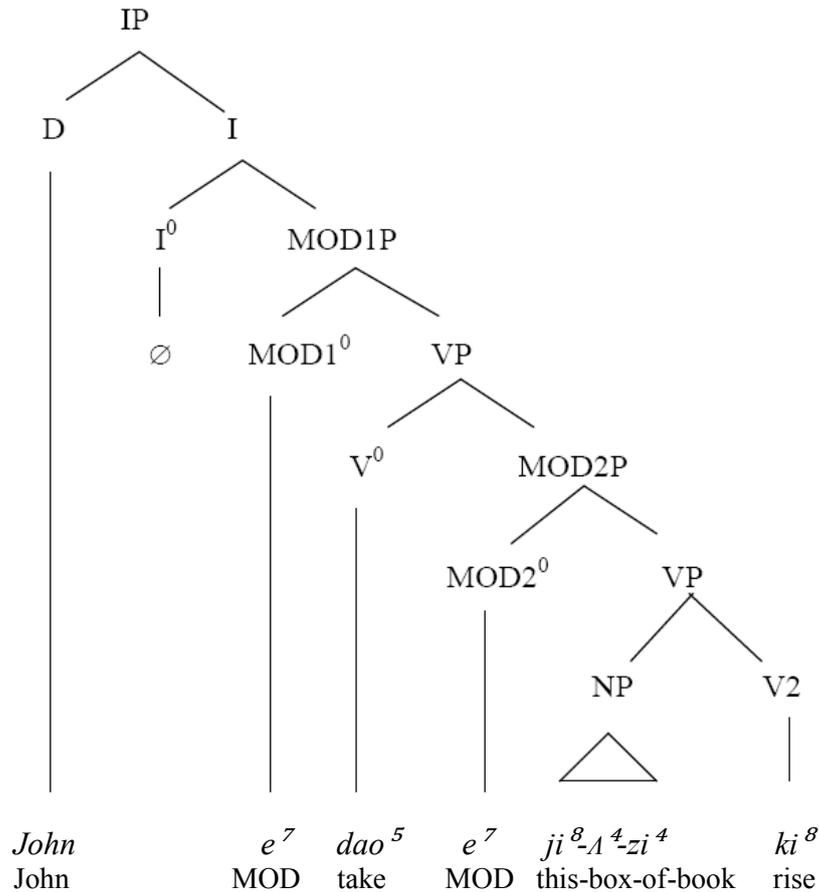
```
                              IP
                          ____|____
                         D         I
                         |       __|__
                         |      I⁰    MOD1P
                         |      |    ___|___
                         |      ∅  MOD1⁰    VP
                         |      |        ___|___
                         |      |       V⁰    MOD2P
                         |      |       |    ___|___
                         |      |       |  MOD2⁰    VP
                         |      |       |    |    __|__
                         |      |       |    |   NP    V2
                         |      |       |    |   /\    |
                       John    e⁷    dao⁵  e⁷  ji⁸-ʌ⁴-zi⁴   ki⁸
                       John    MOD    take  MOD this-box-of-book rise
```

'John will lift this box of books.'
(The speaker is 100% certain that the subject will do the action.)

If only the lower copy is kept, as in (40), then there is what has been called covert movement.

(40) *John* *e⁷* *dao⁵* *e⁷* *ji⁸-ʌ⁴-zi⁴* *ki⁸*
      John   ~~MOD~~ take   MOD   this-box-of-book rise
      'John will lift this box of books.'
      (The speaker is 100% certain that the subject will do the action.)

If only the upper copy is kept, as in (41), then a regular case of overt movement with a trace in the base position obtains, with the interpretation remaining the same in both cases, that is, the speaker is 100% certain that the subject will do the action.

(41) *John* *e⁷* *dao⁵* *e⁷* *ji⁸-ʌ⁴-zi⁴* *ki⁸*
      John   MOD   take   ~~MOD~~ this-box-of-book rise
      'John will lift this box of books.'
      (The speaker is 100% certain that the subject will do the action.)

Conclusion

In this paper I have identified a unique construction in Hong Kong Hokkien that presents two instances of the same modal, which I call Modal Doubling. I show that such construction is restricted to one single type of modal element e^7, and is dependent on co-occurrence with only resultative serial verb construction. The main properties in e^7 modal doubling are as follows: there are two modal elements in Modal Doubling with only single modality reading allowed; e^7 is the only modal that allows doubling; Modal Doubling can only appear with resultative type of serial verb construction. The A-not-A question formation tells us that these seven types of serial verb constructions fall into two group: with resultative, direct object sharing, directional in group A and; instrumental, double object, causative, and sequential in group B. Law's (1996) structure in (23) can account for group A type of constructions, while (24) can be used to explain the constructions in group B. The main difference between (23) and (24) is that in the former V1 takes the VP as the complement; in the latter the VP headed by V2 is an adjunct to the VP headed by V1. I have proposed a Larsonian structure to explain resultative serial verb construction as in (38). The reason being resultative serial verb construction fits into Larson's proposal with VP as a sister to V_0. The Larsonian structure for resultative serial verb construction allows for a structural position for both modal heads, in that way accounting for modal doubling with resultative serial verb construction in Hokkien. Given that VP is a complement to V^0 and modal doubling is a result of head to head movement, this suggests that the two modal elements have to be the same. When only the lower copy is kept, there is a covert movement. When the upper copy is kept, a regular case of overt movement with a trace in the base position obtains, with the interpretation remaining the same in both cases, that is, the speaker is 100% certain that the subject will do the action.

Acknowledgements

References

Larson, Richard. 1991. Some issues in verb serialization, in C. Lefebre (ed), *Serial verbs: grammatical, comparative, and cognitive approaches*, John Benjamins, Amsterdam, pp. 185-210.

Law, Paul. 1996. A note on the serial verb construction in Chinese, *Cahiers de Linguistique-Asia Orientale* **25**, 199-233.

Further readings

Baker, Mark. 1991. On the Relations of Serialization to Verb Extensions, in C. Lefebvre (ed), *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, John Benjamins Publishing Company, Amsterdam,/Philadelphia, pp.79-102.

Barbiers, Sjef, Frits Beukema, and Wim van der Wurff. (Eds.) 2002. *Modality and its Interaction with the Verbal System*. John Benjamins, Amsterdam.

Bybee, Joan L. 1995. 'The semantic development of past tense modals in English', in J. Bybee and S. Fleischman (eds.) *Modality in grammar and discourse*, John Benjamins, Amsterdam, pp. 503-517.

Bybee, Joan, and Suzanne Fleischman. 1995. *Modality in Grammar and Discourse,* John Benjamins, Amsterdam.

Carstens, Vicki. 2002. Antisymmetry and word order in serial constructions, *Language* 78, 3-50.

Chan, Alice Yin Wa. 2000. Chinese serial verb constructions with experiential aspect marker-syntactic representations and semantic interpretations, *Communications of COLIPS* 10, 69-97.

Dai, John Xiang-ling. 1990. Syntactic constructions in serial verb expression in Chinese, in B. Joseph and A. Zwicky (eds), *When verbs collide: papers from the 1990 Ohio State Mini-Conference on serial verbs*. Department of Linguistics, The Ohio State University, Columbus, pp. 316-339.

Hagemeijer, Tjerk. 2001. Underspecification in Serial Verb Constructions, in N. Corver and H. van Riemsdijk (eds), *Semi-lexical categories: The function of content words and the content of function words*, Mouton de Gruyter, Berlin, pp.415-451.

Hale, Ken. 1991. Misumalpan Verb Sequencing Constructions, in C. Lefebvre (ed), *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, John Benjamins Publishing Company, Amsterdam,/Philadelphia, pp.1-36.

Hwang, Jya-Lin 2000. *On Grammaticalization in Serial Verb Constructions in Chinese.* Ph.D. thesis. Department of Linguistics, University of Hawaii.

Law, Paul and Veenstra, Tonjes. 1992. 'On the Structure of Serial Verb Constructions', *Linguistic Analysis* 22, 185-217.

Lefebvre, Claire. (Ed). 1991. *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, John Benjamins Publishing Company, Amsterdam,/Philadelphia.

Lefebvre, Claire. 1991. Take Serial Verb Constructions in Fon, in C. Lefebvre (ed), *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, John Benjamins Publishing Company, Amsterdam,/Philadelphia, pp.37-78.

Li, Charles and Thompson, Sandra. 1981. *Mandarin Chinese: A Functional Reference Grammar*, University of California Press, Berkeley.

Li, Chen Ching. 1979. *A Syntactic and Semantic Analysis of Taiwanese Modality,* Ph.D dissertation, University of Hawaii, Hawaii.

Li, Yafei. 1991. On deriving serial verb constructions, in C. Lefebre (ed), *Serial verbs: grammatical, comparative, and cognitive approaches*, John Benjamins, Amsterdam, pp. 103-136.

Li, Yafei. 1998. Chinese resultative constructions and the uniformity of theta assignment hypothesis. *New approaches to Chinese word formation*, ed. by J. Packard, Berlin, Mouton de Gruyter, pp. 285-310.

Muysken, Pieter and Veenstra, Tonjes. 2005. Serial Verb Constructions, in M. Everaert and H. van Riemsdijk (eds), *The Blackwell Companion to Syntax*, Blackwell, Oxford.

Palmer, Frank. 1986. *Mood and Modality*, Oxford University Press, London.

Veenstra, Tonjes. 1996. Serial verbs in Saramaccan: predication and creole genesis. The Hague, HAG.

QUANTIFIERS IN THAI

Kitima Indrambarya
Kasetsart University
<fhumkmi@ku.ac.th>

0 Abstract

Thai linguists (Savetamalaya 1989, Stein 1991, Deephuengton 1992) have investigated Thai words denoting quantities in Thai, which appear within the noun phrase. However, linguists have different views about the categorization of such words. In many cases, they have been unsystematically classified either as adjectives or as adverbs.

As with Prasithrathsint (2010), this paper proposes that quantifier is a word class in Thai and identifies the syntactic characteristics of quantifiers as different from other word classes. The study reveals that quantifiers appear within a noun phrase in which a numeral functions as head. Only words which precede a noun are identified as quantifiers in Thai. Post-nominal words are adjectives. Quantifiers in Thai may co-occur sequentially and can be subclassified. Thai quantifiers are, for example, *lǎay* 'several', *ʔìik* 'more', *khêɛ* 'just', and *tâŋ* 'as much' *thúk* 'every'.

Keywords: Thai language, quantifiers, classifiers

1 Introduction

Quantifiers are defined in a dictionary as 'linguistic forms that express a contrast in quantity' (The Free Dictionary 2010). English quantifiers, such as *only, many, every, all, and two* are found to precede and modify nouns. Quantifiers are not commonly treated as a separate word class. In English, some are identified as determiners (e.g. *few, many, every*) and others (e.g. *two, first*), adjectives. Savetamalaya (1989) and Deephuengton (1992) have looked into words semantically denoting quantities in Thai and agree that these words appear within a noun phrase. Such noun phrase can be shown to be headed by *náŋsǔ* 'book' in (1).

(1)   [_{NP} *náŋsǔ*   *ʔìik*   *khêɛ*   *sǎamsìp*   *lêm*   *thâwnán* _{NP}]
        book   more   just   thirty   CL   only
        'Just another thirty books'

Within the embedded phrase in which the classifier noun *lêm* is head, the words *ʔìik* 'more', *khêɛ* 'just', *sǎamsìp* 'thirty', *thâwnán* 'only' were found to be semantically identified as quantifiers in Savetamalaya (1989) and Deephuengton (1992). Yet, syntactically they were unsystematically classified as adjectives or as adverbs.

This paper aims at identifying the syntactic characteristics of quantifiers as different from other word classes and subclassifying them.

2 Previous Analyses

Previous studies define quantifiers semantically as words denoting quantities and numbers. Deephuengton (1992), Savetamalaya (1989), and Stein (1981) agree that quantifiers appear within a noun phrase as in (1). Moreover, Deephuengton (1992) and Savetamalaya (1989) agree that quantifiers co-occur with a classifier noun head. However, they disagree on the type of phrase in which quantifiers appear.

In Deephuengton's analysis (1992:229), quantifiers appear within a quantifier phrase as pre-modifiers and post-modifiers of a head noun, as illustrated in (2a). The classifier *lêm* is the noun head for this quantifier phrase.

(2) a [$_{QP}$ *ʔìik* *khɛ̂ɛ* [$_{NumP}$ *sǎamsìp* $_{NumP}$] ***lêm*** *thâwnán* $_{QP}$]

More	just	thirty	CL	only
Pre-modifier	Pre-modifier	Pre-modifier	Head	Post- modifier
Adverb	Degree	Quantifier	Noun	Adverb

'Just another thirty books'

On the other hand, quantifiers in Savetamalaya (1989) precede a noun in a noun phrase, as in (2b). The classifier *lêm* is the noun head for this noun phrase.

(2) b [$_{NP}$ *ʔìik* *khɛ̂ɛ* *sǎamsìp* ***lêm*** *thâwnán* $_{NP}$]

More	just	thirty	CL	only
Quantifier	Quantifier	Quantifier	Head	modifier
Adjective	Adverb	Adjective	Noun	Adverb

'Just another thirty books'

Other linguists (Upakitsilapasarn 1995, Starosta 1994, and Indrambarya 1994), Savetamalaya (1989) and Deephuengton (1992) did not treat quantifiers as a separate word class in Thai. As illustrated in (2a) and (2b), Deephuengton (1992:229) and Savetamalaya (1989:164) accounted for the syntactic category for quantifiers differently. Deephuengton (1992:229), considered numerals to be quantifiers while Savetamalaya (1989:165) treated them as adjectives. The pre-nominal modifier *ʔìik* 'more' was viewed as an adverb in Deephuengton (1992:229), but an adjective in Savetamalaya (1989:164). The pre-nominal modifier *khɛ̂ɛ* 'just' was treated as a degree word, unspecified for its syntactic word class in Deephuengton (1992:229) but an adverb in Savetamalaya (1989:164). The post-nominal modifier *thâwnán* 'only' in (2) was treated as an adverb in both analyses. Upakitsilapasarn (1995:87) also treated words denoting quantities in various positions as adverbs. Previous analyses did not provide a systematic way of identifying words denoting quantities in Thai.

2 Data collection

In this study, data of standard Thai were collected. The data were drawn mainly from Thai newspapers, local television news and talk shows between January and February 2008. The Thai concordance program of the Department of Linguistics at Chulalongkorn University was used to find relevant examples.

3 Analysis
The analysis is done within the framework of Lexicase Dependency Grammar, in which each sentence has only one level: surface structure. This section discusses the analysis of the study. It is divided into three subsections: 1) identifying Thai quantifiers; 2) classification of Thai quantifiers; and 3) syntactic ordering of Thai quantifiers.

3.1 Identifying Thai Quantifiers
The so-called quantifiers in Thai are commonly referred to words semantically expressing quantities as previously discussed in section 2. While these words always co-occur with nouns, they can either precede or follow nouns. This is demonstrated in the example (2) repeated here as (3).

(3)   *ʔìik   khɛ̂ɛ   sǎamsìp   lêm   thâwnán*
        more   just   thirty   CL   only
        'Only another thirty books'

There are four words expressing quantities found in (3). They are *ʔìik* 'more', *khɛ̂ɛ* 'just', *sǎamsìp* 'thirty' and *thâwnán* 'only'. While the first three words precede the noun *lêm* 'classifier for books', the word *thâwnán* 'only' follows the classifier noun.

       The following subsections will discuss different syntactic categories of words expressing quantities in Thai and syntactically identify quantifiers in Thai.

3.1.1 Numerals as Nouns
Numerals behave differently from other words expressing quantities. More specifically, numerals and classifiers behave similarly as nouns. This section illustrates that numerals behave like nouns. First of all, numerals can stand alone while other words denoting quantities cannot. Consider (4) – (5) in which words in questions are used to answer the questions on how many minutes and (6) in which the words in questions are used to answer the question on how many pieces of clothes.

(4)   a   *yîisìp   naathii*
            twenty   minute
            'twenty minutes'

       b   *yîisìp*
            twenty
            'twenty'

(5)   a   *thúk   naathii*
            every   minute
            'every minute'

       b   **thúk*
            every

(6) a   ***raaw***   *sìp*   *phɯ̆ɯn*
         approximate   ten   piece
         'Approximately ten pieces of clothes'

     b   * ***raaw***
         approximate

     c   *sìp*
         ten
         'ten'

The examples above show that the numeral   *yîisìp* 'twenty' acts differently from *thúk* 'every' and *raaw* 'approximate'. The numeral *yîisìp* 'twenty' in (4b) can stand alone but the words like *thúk* 'every' in (5b) and *raaw* 'approximately' in (6b) cannot stand alone. The latter two must co-occur with a classifier, as in (5a) or with a numeral and a classifier, as in (6a). Hence, it is evident that numerals like *yîisìp* 'twenty' behave differently from so-called quantifiers or words denoting quantities.

   Moreover, numerals act like nouns, just as classifiers do. They can function as the object of a preposition. Consider (7) and (8).

(7) a   *nɯ̀ŋ*   <u>*nay*</u>   *rɔ́y*
         one   in   hundred
         'One in a hundred.'

     b   *lûatlaay*   <u>*bon*</u>   ***phɯ̆ɯn***   *níi*
         pattern   on   CL for cloth   this
         'Pattern on this article of clothing.'

Example (7) shows that the numeral *rɔ́y* 'hundred' in (7a), just like the classifier *phɯ̆ɯn* 'classifier for cloth' in (7b) can be objects of prepositions *nay* 'in' and *bon* 'on', respectively. Further, like other nouns, numerals can be followed by a determiner, a subclass of adjectives, such as *níi* 'this' in (8a) and *ʔɯ̀ɯn* 'others' in (8b) in certain contexts.

(8) a   *sɔ̆ɔŋrɔ́y*   <u>*níi*</u>   *khɔ̆ɔŋ*   *thəə*
         two hundred   this   belong   you
         'This two hundred is yours.'

     b   *sɔ̆ɔŋrɔ́y*   <u>*ʔɯ̀ɯn*</u>
         two hundred   other
         'Other two hundred.'

It is evident that a numeral behaves differently from other words expressing quantities. Hence, a numeral is analyzed as a noun. Numeral nouns are, for example, *sǎamphan* 'three thousand', *sìp* 'ten', *khrɯ̂ŋ₁* 'half'. In the next sections, words expressing quantities which appear before and after a noun are discussed.

3.1.2 So-called quantifiers as Adjectives

Unlike numerals, other words expressing quantities which are commonly referred to as quantifiers share certain similarities. They cannot stand alone and must co-occur with nouns. Moreover, they cannot co-occur with the negation word *mây* 'not'. So, words preceding nouns such as *ʔìik* 'more' and *khɛ̂ɛ* 'just' cannot stand alone and must co-occur with a following noun, as shown earlier in (5b) and (6b). Similarly, those following nouns such as *thâwnán* 'only' and *khrɨ̂ŋ₂* 'half' cannot stand alone and must co-occur with a preceding noun, as exemplified in (9b), (9c), (10b) and (10c).

(9) a   *kìi*   *nâa*
          How many   page
          'How many pages?'

     b   *yîisip*   *nâa*   **thâwnán**
          twenty   page   only
          'Twenty pages only'

     c   ***thâwnán***
          only

     d   * *yîisip*   *nâa*   *mây*   **thâwnán**
          twenty   page   not   only

(10) a   *kìi*   *kilo*
          how many   kilograms
          'How many kilograms?'

     b   *kilo*   **khrɨ̂ŋ₂**
          kilo   half
          'One and a half kilogram'

     c   * *khrɨ̂ŋ₂*
          half

     d   * *kilo*   *mây*   **khrɨ̂ŋ₂**
          kilo   not   half

Moreover, example (9d) and (10d) shows that *thâwnán* 'only' and *khrɨ̂ŋ₂* 'half' cannot appear with the negation word *mây* 'not'.

Word distribution plays a significant role in deciding the syntactic category of a word in question. As a head-initial language, Thai has modifiers following its head. This analysis, following Prasithrathsint (2010), defines Thai adjectives as words which appear after a preceding noun. Another characteristic of an adjective is that it cannot co-occur with the negation word *mây* 'not', as in (11b).

(11) a   *phûuyĭŋ*   ***sŭaysŭay***
          woman   beautiful
          'A beautiful woman'

  b   **phûuyĭŋ*   *mây*   ***sŭaysŭay***
          woman   not   beautiful

The adjective *sŭaysŭay* 'beautiful' cannot be negated with *mây* 'not'. Since the words expressing quantities such as *thâwnán* 'only' and *khrɯ̂ŋ₂* 'half' appear after a noun and behave just like adjectives, they are regarded as adjectives. More examples of adjectives expressing quantities are shown in (12) and (13).

(12)   *ʔìik*   *khɛ̂ɛ*   *nâa*   ***dɯaw***   ***thâwnán***
        more   just   page   one   only
        +Q   +Q   +N   +AJ   +AJ
        'Just one more page.'

The words *dɯaw* 'one' and *thâwnán* 'only' in (12) and the word *sèet* 'residue' in (13) appear after the noun *nâa* 'classifier for books; page' and *dɯan* 'month' respectively. Hence they are analyzed as adjectives.

(13)   *phɯaŋ*   *khɛ̂ɛ*   *dɯan*   ***sèet***
        only   just   month   residue
        +Q   +Q   +N   +AJ
        'Just over a month.'

    In the next section, words expressing quantities which appear before a noun are identified as quantifiers in Thai.

3.1.3 Quantifier as a Word Class in Thai

As a right-branching language, dependents stand to the right of their head. A verb precedes its complements. A preposition precedes a noun complement. As for modifiers, adverbs follow verbs. Adjectives follow nouns. As discussed earlier, words denoting quantities in Thai can either precede or follow nominals, i.e. numerals, classifiers, and common nouns. While those that follow the nouns can be regarded as adjectives, those that precede the nouns cannot. Unlike previous analyses (Upakitsilapasarn 1995; Thonglor 2002; Bandhumedha 1979; Savetamalaya 1989; Indrambarya 1994) that view words denoting quantities as quantifiers and set them into subclasses of adjectives or adverbs, this study identifies quantifiers syntactically. Words expressing quantities which appear before a noun are the only group of words in the Thai language which precede nouns. Their distinct characteristic deserves its own position as a word class in Thai.

    I then propose that quantifiers in Thai are identified as words which precede a noun head in an endocentric construction as with Prasithrathsint (2010). Quantifiers are, for example, *khɛ̂ɛ* 'just', *raaw* 'approximately', *tàaŋ* 'each', *kìi* 'how many', *sàk* 'approximately', *pramaan* 'approximately' *tháŋ* 'altogether', *baaŋ* 'some', *lăay* 'several', *tʃʰat* 'almost', *kɯ̀ap* 'almost', *kwàa* 'over'. Employing syntactic criteria to identify

quantifiers, this study is able to capture a better generalization for the Thai language. Only words which precede nouns are identified as quantifiers. Any words following nouns are identified as adjectives.

The tree structure in (14) shows that the numeral head noun *sɔ̌ɔŋ* 'two' takes three dependent sisters. It is preceded by the quantifier *khɛ̂ɛ* 'just' and is followed by the classifier noun *wan* 'day' and the adjective *thâwnán* 'only'.

(14)

khɛ̂ɛ	*sɔ̌ɔŋ*	*wan*	*thâwnán*
just	two	day	only
+Q	+N	+N	+AJ

'Just only two days.'
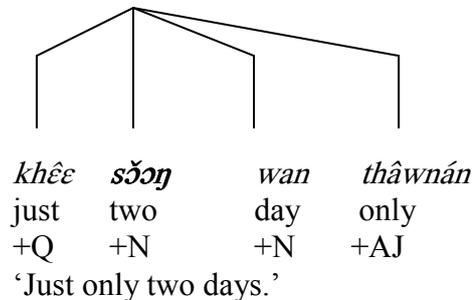
This section has identified words preceding nouns as quantifiers. The following section will discuss the type of phrase that contains quantifiers.

3.1.4 Phrase Containing Quantifiers
It is widely accepted that quantifiers appear within a noun phrase. Consider (1) repeated here.

(1) [NP *náŋsɯ̌*   *ʔìik*   *khɛ̂ɛ*   *sǎamsìp*   *lêm*   *thâwnán* NP]
          book   more   just   thirty   CL   only
     'Just another thirty books'

(1) is a noun phrase headed by the noun *náŋsɯ̌* 'book'. Quantifiers appear within its embedded phrase. It is necessary to determine which word is head of this embedded phrase and what kind of phrase contains quantifiers. A head must be obligatory while all others could be optional. So, to maintain the meaning 'Just another thirty books,' obligatoriness test is then used to find out which word is the head of the phrase: a quantifier, a numeral noun or a classifier noun. The result is shown in (15).

(15) a   *khɛ̂ɛ*   *sǎamsìp*   *lêm*
          just   thirty   CL
          'Just another thirty (books).'

     b   *ʔìik*   *sǎamsìp*   *lêm*
          more   thirty   CL
          'Another thirty (books).'

     c   **ʔìik*   *khɛ̂ɛ*   *lêm*
          more   just   CL
          'Just another (books).'

d   *ʔìik*   *khɛ̂ɛ*   *sǎamsìp*
    more   just   thirty
    'Just another thirty (books).'

Example (15) illustrates that quantifiers such as *ʔìik* 'more' and *khɛ̂ɛ* 'just' can be left out, as in (15a) and (15b), respectively. Example (15c) shows that the numeral *sǎamsìp* 'thirty' is obligatory and cannot be left out. The example (15d) shows that the classifier noun *lêm* 'classifiers for books' can be left out within the context. The obligatoriness of numerals provides evidence that numerals, rather than quantifiers or classifiers, are heads of the phrase containing quantifiers. Furthermore, numerals also carry the meaning of the whole phrase in (15).

Unlike previous analyses, this study proposes that the phrase containing quantifiers is a noun phrase headed by a noun, in most cases, a numeral, as illustrated with the tree diagram in (16).

(16)

[_{NP} *ʔìik*   *khɛ̂ɛ*   *sǎamsìp*   *lêm*   *thâwnán* _{NP}]
     more   just   thirty   CL for books   only
     +Q   +Q   +N   +N   +AJ
     'Just only another thirty books.'
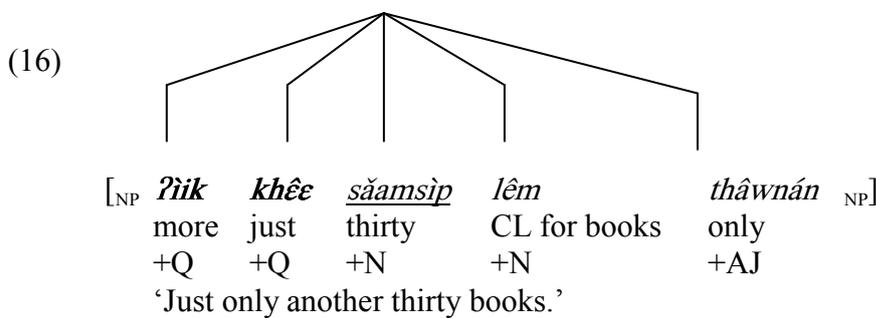
In (16), the numeral noun *sǎamsìp* 'thirty' is the head of the phrase. It is preceded by the quantifiers *ʔìik* 'more' and *khɛ̂ɛ* 'just'. Example (15c) would be acceptable only if the meaning were 'just one more book'. The classifier *lêm* 'classifier for books' will be the head of the phrase when a numeral is not present, as shown in the tree diagram (17).

(17)

[_{NP} *ʔìik*   *khɛ̂ɛ*   *lêm*   *diaw*   *thâwnán* _{NP}]
     more   just   CL for books   one   only
     +Q   +Q   +N   +AJ   +AJ
     'Just only one more book.'
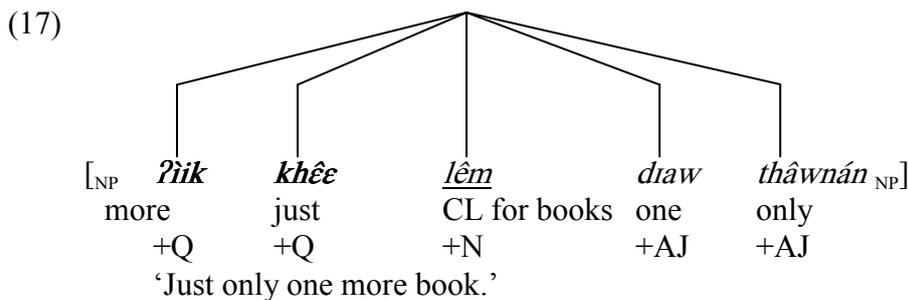
Note that when a classifier is the head of the noun phrase containing quantifiers, it implies 'one' in number. This is illustrated in the following set of examples.

(18)

```
              Ɂìik        pramaan       sǎam        dɨan
              more        approximate   three       month
              'Approximately another three months'
```
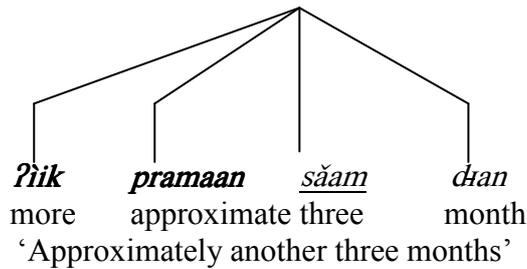
Example (18) shows that a numeral, when present, is the head of the phrase containing quantifiers. Thus, *sǎam* 'three' is the head of this noun phrase. When a numeral is not present, a classifier noun like *dɨan* 'month', as in (19), or a common noun like *tʃâat* 'nation', as in (20), acts as a noun head. Again, it is 'one' in number.

(19)

```
              Ɂìik        pramaan       dɨan
              more        approximate   month
              'Approximately one month'
```

(20)

```
              tèeláɁ      tʃâat
              more        nation
              'Each nation'
```
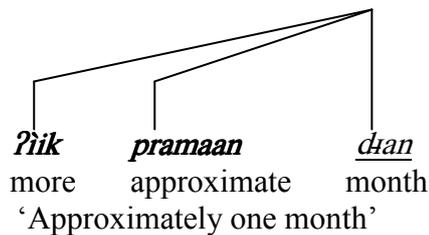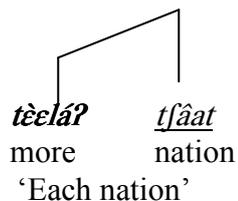
This section has illustrated that the phrase containing quantifiers is a noun phrase in which a numeral noun functions as the head. Without the presence of a numeral, a classifier or a common noun acts as the head noun. The next section will discuss the syntactic characteristics of Thai quantifiers.

3.1.5 Syntactic Characteristics of Quantifiers

This section discusses the characteristics of Thai quantifiers. Thai quantifiers can be identified syntactically with features as [-[*mây* ___], [+[___ N]] (Prasithrathsint et al. 2011). Syntactically they can be characterized as 1) preceding a noun; 2) unable to co-occur with *mây* 'not'; and 3) co-occurrence with other quantifiers.

3.1.5.1 Preceding a Noun

Quantifiers are found to precede a head noun i.e. a numeral, a classifier or a common noun in an endocentric construction.

(21)

```
        ┌──────────┬──────────┐
        │          │          │
      t͡ɕìat      rûam      sɔ̌ɔŋphan
     almost     almost    two thousand
      +Q         +Q        +N
```
      'Almost two thousand'

As shown, in (21), quantifiers *t͡ɕìat* 'almost' and *rûam* 'almost' are sister dependents which precede the head noun *sɔ̌ɔŋphan* 'two thousand'. Quantifiers may also precede classifiers or common nouns, as illustrated earlier in (19) and (20).

3.1.5.2 unable to Co-occur With the Negation word *mây* 'Not'

Only verbs in Thai may co-occur with the negation word *mây* 'not'. Quantifiers may not be preceded and negated by *mây* 'not', as illustrated in (22) – (23).

(22)   *<u>mây</u>   **ʔìik**   caan
           not   more   plate
          'Not another plate.'

(23)   *<u>mây</u>   **pramaan**   sǎamsìp   t͡ɕûamooŋ
            not   approximate   thirty   hour
          'Not approximately thirty hours.'

3.1.5.3 Co-occurrence Among Quantifiers

All quantifiers function as sister dependents of a head noun that follows. Quantifiers may co-occur. Detailed analysis is given in section 3.3.

(24)   **kìap**   **thúk**   <u>sìi</u>   t͡ɕûamooŋ
           almost   every   four   hour
           +Q   +Q   +N   +N
          'Almost every four hours.'

(25)   **ʔìik**   **phiaŋ**   **khɛ̂ɛ**   **pramaan**   <u>hâa</u>   t͡ɕán
           more   only   just   approximate   five   level
           +Q   +Q   +Q   +Q   +N   +N
          'Only about another five levels.'

Sentence (24) and (25) exemplify that a series of quantifiers may precede the noun heads *sìi* 'four' in (24) and *hâa* 'five' in (25).

This section has discussed the syntactic categories of so-called quantifiers. Numerals are identified as nouns. Those following nouns are analyzed as adjectives. Those preceding nouns are identified as quantifiers in Thai. Quantifier becomes a new syntactic category in the Thai language. This section also proposes that quantifiers appear in a noun phrase headed by a numeral, a classifier or a common noun.

3.2 Classification of Thai Quantifiers

Thai quantifiers may be syntactically classified into five classes in terms of positions in which they appear. They are, namely, 'one' quantifiers, 'incomplete' quantifiers, 'estimated' quantifiers, 'near-complete' quantifiers, and 'excessive' quantifiers.

3.2.1 'One' Quantifiers

'One' quantifiers express the 'one, each'. They include *tèɛláʔ* 'each', *tàaŋ* 'each' and *láʔ* 'each'.

(26)   *tèɛláʔ*   *lêm*
        each   book
        'Each book'

(27)   *wan*   **láʔ**   *saam*   *khráŋ*
        day   per   three   time
        Three times a day.'

3.2.2 'Incomplete' Quantifiers

'Incomplete' quantifiers express the quantities that are more or less than expected. They include *ʔìik* 'more' *phıaŋ* 'only' and *khɛ̂ɛ* 'just'. These quantifiers may co-occur and reverse order among them as in (28a-f).

(28) a   **ʔìik**   **phıaŋ**   **khɛ̂ɛ**   *sìp*   *khon*
          more   only   just   ten   person
          'Just 10 more people'

    b   **ʔìik**   **khɛ̂ɛ**   **phıaŋ**   *sìp*   *khon*
          only   just   more   ten   person
          'Just 10 more people'

    c   **phıaŋ**   **ʔìik**   **khɛ̂ɛ**   *sìp*   *khon*
          more   just   only   ten   person
          'Just 10 more people'

    d   **phıaŋ**   **khɛ̂ɛ**   **ʔìik**   *sìp*   *khon*
          only   more   just   ten   person
          'Just 10 more people'

    e   **khɛ̂ɛ**   **ʔìik**   **phıaŋ**   *sìp*   *khon*
          just   more   only   ten   person
          'Just 10 more people'

    f   **khɛ̂ɛ**   **phıaŋ**   **ʔìik**   *sìp*   *khon*
          just   only   more   ten   person
          'Just 10 more people'

Further, 'incomplete' quantifiers may precede other classes of quantifiers. In (29), the incomplete quantifier *ʔìik* 'more' precedes 'near-complete' quantifiers *tâŋ* 'as much' and *kìap* 'almost'.

(29)   **ʔìik**   *tâŋ*   *kìap*   *sɔ̌ɔŋ*   *pii*
          more   as much   almost   two   year
          'As much as almost two years more'

3.2.3 'Estimated' Quantifiers

'Estimated' quantifiers express estimated number. They are, for example, *pramaan* 'approximate', *sàk* (*sák*) 'approximate', *raaw* 'approximate', *raawraaw* 'vaguely approximate', *kìi* 'how many', *baaŋ* 'some', *rɨan* 'about', *náp* 'about', *lǎay* 'a few', *lǎaylǎay* 'several', and *naanaa* 'several'. These quantifiers may co-occur. Examples are shown in (30) and (31).

(30)   **baaŋ**   *khon*
          some   person
          'Some people'

(31)   **pramaan**   *sák*   *sǎamrɔ́ɔy*   *naathii*
          approximate   approximate   three hundred   minute
          'Approximately 300 minutes'

3.2.4 'Near-complete' quantifiers

'Near-complete' quantifiers express the perspective of speaker toward the full amount. They are, for example, *tʃîat* 'almost', *yìap* 'step on', *kìap* 'almost', *kìapkìap* 'almost', *khɔ̂ɔn* 'almost', *rûam* 'almost', *khâw* 'getting into'. Moreover, this set of quantifiers include words expressing the full amount like *tháŋ* 'all', *puaŋ* 'all', *muan* 'all', *sǎaraphát* 'all kinds of' *sǎaraphan* 'all kinds of', *tâŋ* 'as much', *thúk* 'every', *thúkthúk* 'every', as shown in (32) - (33). Again, quantifiers in this class may co-occur.

(32)   **kìap**   *tháŋ*   *sǎam*   *wan*
          Almost   all   three   day
          'Almost three days long'

(33)   **thúk**   *sìi*   *tʃûamooŋ*
          every   four   hour
          'Every four hours'

3.2.5 'Excessive' quantifiers

'Excessive' quantifiers express the quantities that exceed the full amount. They include *kwàa* 'over', as in (34).

(34)   **kwàa**   *yîisìp*   *naathii*
          over   twenty   minute
          'Over twenty minutes'

This section has illustrated the five subclasses of Thai quantifiers most of which may co-occur. The next section will discuss the syntactic ordering of co-occurring quantifiers.

3.3 Syntactic ordering of Thai Quantifiers

Quantifiers may co-occur and precede a following noun head. Each class of quantifier may co-occur in a particular order. The following diagram shows the linear ordering of Thai quantifiers.

> One < incomplete < estimated < near-complete < excessive

That is, 'one' quantifiers are in the foremost position. Then 'incomplete' quantifiers precede 'estimated' quantifiers which precede 'near-complete' quantifiers. 'Excessive' quantifiers appear last. Co-occurrences of quantifiers are shown in examples (35a) – (38a). The wrong order of quantifiers could result in unacceptability of each sentence, as illustrated in (35b) - (38b) and (38c).

(35) a   *náŋsɨ̌*   **ʔìik**   **sák**   **kìap**   *kâaw*   *nâa*
          (book)   more   approximate   almost   nine   page
                        incomplete   estimated   near-complete

      'Almost approximately nine pages more'

      b   **náŋsɨ̌*   **sák**   **ʔìik**   *kìap*   *kâaw*   *nâa*
          (book)   approximate   more   almost   nine   page
                        estimated   incomplete   near-complete

(36) a   *náŋsɨ̌*   **ʔìik**   **tâŋ**   **kwàa**   *rɔ́ɔy*   *nâa*
          (book)   more   as much   more   hundred   page
                        incomplete   near-complete   excessive

      'As much as over a hundred pages more'

      b   * *náŋsɨ̌*   *ʔìik*   **kwàa**   **tâŋ**   *rɔ́ɔy*   *nâa*
          (book)   more   more   as much   hundred   page
                        incomplete   excessive   near-complete

(37) a   *phûutʃom*   **ʔìik**   **kìap**   *sɔ̌ɔŋrɔ́y*   *khon*
          (audience)   more   almost   two hundred   person
                        incomplete   near-complete

      'Almost another two hundred people'

      b   **phûutʃom*   **kìap**   **ʔìik**   *sɔ̌ɔŋrɔ́y*   *khon*
          (audience)   almost   more   two hundred   person
                        near-complete   incomplete

(38) a

	khon	***láʔ***	***ʔìik***	***pramaan***	***rûam***	*săamsìp*	*naathii*
	person	per	more	approximate	almost	thirty	minute
		one	incomplete	estimated	near-complete		

'Approximately another thirty minutes for each person'

b

	*khon	***ʔìik***	***láʔ***	*pramaan*	*rûam*	*săamsìp*	*naathii*
	person	more	per	approximate	almost	thirty	minute
		incomplete	one	estimated	near-complete		

c

	*khon	*láʔ*	*ʔìik*	***rûam***	***pramaan***	*săamsìp*	*naathii*
	person	per	more	almost	approximate	thirty	minute
		one	incomplete	near-complete	estimated		

The linear order of different types of quantifier when appearing within a noun phrase can be shown in the following table.

Noun Phrase								
Head	Modifiers							
Noun	Noun Phrase							
	Modifiers					Head	Modifiers	
	Quantifiers	Quantifiers	Quantifiers	Quantifiers	Quantifiers	Noun	Noun	Adjectives
	One	Incomplete	Estimated	Near-Complete	Excessive	Numerals	Classifiers	
nánsɯ̌ 'book'		*phıaŋ* 'only'	*pramaan* 'approximate'	*kìap* 'almost'		*kâaw* 'nine'	*nâa* 'page'	*khrɯ̂ŋ* 'half'
phûutʃom 'audience'		*ʔìik* 'more'		*kìap* 'almost'		*sɔ̌ɔŋróy* 'two hundred'	*khon* 'person'	
khon 'person'	*láʔ* 'each'	*ʔìik* 'more'	*pramaan* 'approximate'		*kwàa* 'over'	*săamsìp* 'thirty'	*naathii* 'minute'	

Table 1: Noun phrase with sequential order of quantifiers

4 Conclusions

In this paper, so-called quantifiers in Thai are identified as quantifiers, adjectives and numeral nouns. Numerals are identified as nouns for they can stand alone and behave just like nouns, in contrast to other words denoting quantities. Words denoting quantities which follow noun heads are analyzed as adjectives since they share the same characteristics as

adjectives. Only those which precede nouns are identified as quantifiers in Thai. Being the only class of words which precede a noun head, I argue that quantifiers stand as a separate word class in Thai. Quantifiers can be subdivided into five classes syntactically, based on their co-occurrence property. They are, 'one', 'incomplete', 'estimated', 'near-complete', and 'excessive' quantifiers.

Reference

Bandhumedha, Banchob. 1979. *Laksana phasa thai (Characteristics of Thai).* Bangkok: Khurusapha.

Deephuengton, Phawadee. 1992. *A Study of Quantifier Phrases in Thai.* Kansas Working Papers in Linguistics. Vol.17 No.1-2

Indrambarya, Kitima. 1994. *Subcategorization of Verbs in Thai: A Lexicase Dependency Approach.* University of Hawaii Ph.D. Dissertation.

Prasithrathsint, Amara. 2010. *Lak Samkhan Nay Kanjamnaek Chanitkham Nay Phasa Thai (Main Principles of Word Classification in Thai).* In Prasithrathsint, Amara, Rungrojsuwan, Sorabud and Sankaburanurak, Sarawanee (eds.). Selected Articles from the Pan-dialectal Grammar of Thai Symposium on Diversity of Languages in Thai Society, Standard Thai, Thai Regional and Social Dialects: Issues and New Discoveries (in Thai). Bangkok: ASP.

Prasithrathsint, Amara, Indrambarya, Kitima and Chaijaroen, Nattawut. 2011. *Waiyakorn Phasa Thai Matrathan (Standard Thai Grammar).* A full report of research on Pan-dialectal Grammar.

Savetamalaya, Saranya. 1989. *Thai nouns and noun phrases: A lexicase analysis.* University of Hawai'i Ph.D. Dissertation.

Starosta, Stanley. 1994. *Identification of Word Class in Thai.* Ms. Revised version of a Paper presented at the Fifth International Conference on Thai Studies at the Center for Southeast Asian Studies, School of African and Oriental Studies in London. July 1993.

Stein, Mark Jeffrey. 1981. *Quantification in Thai.* Ph.D. Dissertation University of Massachusetts Amherst.

Thonglor, Kamchai. 2002. *Lak Phasa Thai. (Principles of Thai).* Bangkok: Ruamsarn.

Upakitsilapasarn, Phraya. 1995. *Lak Phasa Thai (Principles of Thai Language.)* Bangkok: Thai Wattana Panich.

The Free dictionary by Farlex. 2010. http://www.thefreedictionary.com/quantifier [online]. Accessed on December 25, 2010.

TONAL OCP AND CONSONANT-TONE INTERACTION IN THAI[*]

Seunghun J. Lee
Central Connecticut State University
<juliolee@gmail.com>

Abstract

In Thai, high tone is not realized in syllables with unaspirated onsets or codas, but falling tone is. Previous studies have argued that contour tones in Thai are combinations of level tones. If falling tone is regarded as a composite of a high tone and a low tone, it is puzzling that falling tone in fact is realized with unaspirated onsets. This paper will reexamine the status of Thai contour tones, propose OCP constraints on tonal features, and analyze the permission of falling tone with unaspirated onsets as an emergence of the unmarked effect.

Keywords: contour tones, consonant-tone interaction, optimality theory (OT)

1. Introduction

This paper aims to examine the status of contour tones and consonant-tone interaction in Thai, and to propose tonal constraints referring to the Obligatory Contour Principle (OCP). The basic observation comes from the report that high tone in Thai does not co-occur with unaspirated voiceless and voiced stop onsets, while falling tone does. Since falling tone is assumed to be composed of high tone and low tone in Thai, the asymmetry regarding consonant-tone interaction in high tone and falling tone needs to be accounted for. The analysis will show that falling tone is allowed after an unaspirated onset because the OCP constraint strictly dominates the constraint that bans high tone from co-occurring with unaspirated onsets.

The table in (1) shows the five tones in native Thai words with syllables beginning with various types of onsets. While all five tones are realized in syllables with aspirate or sonorant onsets, only four tones appear in syllables with unaspirated onsets. As the heavy-lined box shows, in the variety of Thai reported by Ruangjaroon, high tone does not appear with unaspirated stops. I will report recent findings regarding this point in section 5.

(1)   Tone and consonants in Thai (Ruangjaroon, 2006: 39-48)

Onsets	low	mid	high	falling	rising
a. unaspirated	[kù:]	[kū:]	N/A	[kû:]	[kǔ:]
	'holler'	1st pers. sg.	N/A	'borrow'	'uncle'
b. aspirates	[pʰà:]	[pʰā:]	[pʰrá:]	[pʰâ:]	[pʰǎ:]
	'cut'	'take'	'a knife'	'clothes'	'a cliff'
c. sonorants	[lò:]	[lā:]	[lá:]	[lâ:]	[lǎ:]
	'last'	'a donkey'	'exhausted'	'chase'	'yard'

This cooccurrence restriction between high tone and unaspirated onsets raise a range of issues in understanding Thai tone. What is the status of contour tones in Thai? What kind of tonal features can describe the Thai tone system? What is the domain of OCP constraints in Thai? These questions will be addressed in section 2. The rest of the paper is organized as follows. Issues related to consonant-tone interaction will be introduced in section 3. Based on the discussion, an analysis of the permission of unaspirated onsets with falling tone will be presented in section 4. A discussion on relevant points will be given in section 5.

2. Tonal features

Thai tone has been extensively studied in various areas. Production studies examine the shape of pitch in tonal categories (Bradley, 1911, Abramson, 1962, Intrasai, 2001, Morén and Zsiga, 2006, Thepboriruk, 2010). The results of these studies show that phonetic shapes of the Thai tones have undergone a major change since the early 20th century. A recent study on Thai tone (Teeranon, 2007) adds data from perception experiments and discusses a change in tonal shapes for high tone.

Different tonal feature systems have been proposed in order to describe the tonal patterns since the early studies on tone (Woo, 1969 and following studies). In particular, whether contour tones are formed from level tones or whether they form a single unit has been a source of debate (Anderson, 1978: 146-161). In this paper, I will assume that contour tones in Thai are composed of level tones which are associated with a mora following Woo (1969) and subsequent proposals. I will also assume that Thai syllables satisfy a minimality requirement of two moras in syllable weight, as reported in two separate studies by Leben (1971) and Gandour (1974).

I claim that two features [upper] and [lower] are sufficient in describing Thai tone.[46] These two features capture the three level tones in Thai; H tone [+upper, -lower], M tone [-upper, -lower], and L tone [-upper, +lower]. Contour tones are represented as combinations of level tones as in (2).

---

[46] The proposal regarding the basic architecture of tonal features in Yip (1980, 1989) has an advantage in accounting for the Chinese tone system. Since Thai lacks tone sandhi and other evidence for Yip's tone system, I will leave the question open to future research whether Thai tone should also be analyzed with Yip's feature system.

(2)   Contour tones

a.   Falling tone   b.   Rising tone

[+upper, -lower]   [-upper, +lower]   [-upper, +lower]   [+upper, -lower]

H   L   L   H
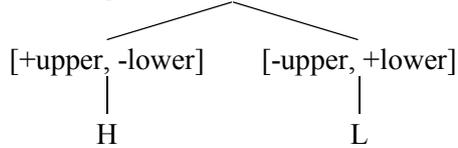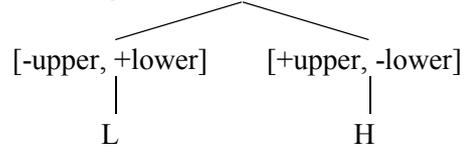
      This representation of contour tones will be used to account for the ban of H tone in syllables with unaspirated stops, and the permission of falling tone after the same consonants. The analytic force behind the pattern is the OCP constraints on tonal features: OCP-[upper] and/or OCP-[lower].

2.1 Tonal feature constraints[47]

A series of faithfulness constraints and markedness constraints are proposed in (3) and (4). These constraints conspire with consonant-tone interaction constraints. The faithfulness constraints force the preservation of the tonal identity between corresponding moras in the input and output. The markedness constraints in (4) ban the presence of identical tonal features from occurring in the same syllable.

(3)   Faithfulness constraints

a. IDENT [+UPPER]   Correspondent segments associated to a mora have identical values for the tonal feature [+upper]. If x is a mora in the input and y is a mora in the output from $x \Re y$, x is [+upper], then y is [+upper].

b. IDENT[-UPPER]   Correspondent segments associated to a mora have identical values for the tonal feature [-upper]. If x is a mora in the input and y is a mora in the output from $x \Re y$, x is [-upper], then y is [-upper].

c. IDENT[+LOWER]   Correspondent segments associated to a mora have identical values for the tonal feature [+lower]. If x is a mora in the input and y is a mora in the output from $x \Re y$, x is [+lower], then y is [+lower].

d. IDENT[-LOWER]   Correspondent segments associated to a mora have identical values for the tonal feature [-lower]. If x is a mora in the input and y is a mora in the output from $x \Re y$, x is [-lower], then y is [-lower].

---

[47] I specially thank Paul de Lacy for the discussion regarding these constraints.

(4)   Markedness constraints

    a. OCP[UPPER]   Assign a violation mark, if for two moras in the same syllable, both moras are associated to [αupper], and have different values for [lower].

    b. OCP[LOWER]   Assign a violation mark, if for two moras in the same syllable, both moras are associated to [αlower], and have different values for [upper].

       The assumptions about contour tones in Thai and the proposed constraints in (3) and (4) are the starting point for analyzing falling tone with unaspirated onsets when high tones are banned in the same type of syllable. Before discussing an OT analysis of this phenomenon, section 3 describes consonant-tone interaction in Thai.

3. Consonant-tone interaction in Thai

The interaction between consonants and tone has been a focus of several studies (Hyman and Schuh, 1974, Hombert, 1978 and following studies). Recently, consonant-tone interaction has become the topic of several major studies. Bradshaw (1999) proposes a multi-planar theory of features, Tang (2008) presents a detailed study of phonetics and phonology of this interaction, and Lee (2008) proposes 'extended tone bearing unit' theory in which he argues that it is an asymmetry in the constraint system that results in consonant-tone interaction within an OT analysis.

       Thai is one such language that shows restrictions on tone as a result of certain consonant types. The data in this paper mainly comes from the variety of Thai reported in Ruangjaroon (2006). Ruangjaroon reports that high tone does not occur with unaspirated stops. This is interesting because falling tone does occur with these stops. If falling tone is a composite of a high tone and a low tone, the difference between falling tone and high tone needs to be accounted for. Other studies on the relationship between consonants and tone in Thai will only be mentioned where it is relevant (Rischel, 1986, Tumtavitikul, 1992, Roengpitya, 2000).

       The patterns of consonant-tone interaction in Asian languages differ from those in African languages. In African languages, consonants block the spreading of tone. Voiced stops block H tone spreading in Xitsonga (Lee, 2009c), and voiceless stops block L tone spreading in Bade (Tang, 2007). In Asian languages, consonants typically restrict the type of tones that can co-occur in a given syllable (Lee, 2009b, a).

       In Thai, consonants in both onset and coda positions neutralize tone in the output. In the literature, open syllables and syllables with sonorant codas are classified as unchecked syllables, and syllables with stop codas are called checked syllables. There is no restriction on tones when an unchecked syllable has an aspirate or sonorant onset (5). An unaspirated stop onset in an unchecked syllable is not realized with a high tone as in (6).

(5) Tones in unchecked syllables (Ruangjaroon, 2006: 39-48)

	Onsets		**low**	**mid**	**high**	**falling**	**rising**
a.	**aspirates**	CV:	[pʰà:] 'cut'	[pʰā:] 'take'	[pʰrá:] 'a knife'	[pʰâ:] 'clothes'	[pʰǎ:] 'a cliff'
		CVS	[kʰùn] 'muddy'	[kʰūn] 2ⁿᵈ sg. pers.	[kʰún] 'familiar'	[kʰôn] 'thick'	[kʰǔn] 'fatten'
b.	**sonorants**	CV:	[lò:] 'last'	[lā:] 'a donkey'	[lá:] 'exhausted'	[lâ:] 'chase'	[lǎ:] 'yard'
		CVS	[màn] 'persistent'	[mān] 'greasy'	[nán] 'that'	[mân] 'engage'	[mǎn] 'sterile'

(6) Unchecked syllables with an unaspirated stop (Ruangjaroon, 2006: 39-48)

	low	**mid**	**high**	**falling**	**rising**
a. CV:	[kù:] 'holler'	[kū:] 1ˢᵗ sg. pers.	////////	[kû:] 'borrow'	[kǔ:] 'uncle'
b. CVS	[bòn] 'complain'	[bōn] 'on'	////////	[bân] 'a portion'	[bǔn] 'civil'

Checked syllables with aspirate or sonorant onsets can appear only with high or low tone (7). Checked syllables do not occur with mid or rising tone. It is also the case that checked syllables with long vowels can never be realized with high tone (7c-d). I argue that the surface falling tone in (7d) is a variant of a high tone that occurs in a checked syllable with a long vowel. The H tone associates with the first mora of the vowel in [máàk] 'very' and the second mora is realized with a phonetic L tone. The surface tone of such a syllable is a falling tone (see also Wong-opasi, 1992: 455). This high-low contrast in checked syllables is lost when a word begins with an unaspirated stop as in (8). As shown in (6), high tone is blocked in such syllables, so the surface tone is limited to low tone. Different effects on fundamental frequency (F0) from pre-vocalic and post-vocalic consonants have already been reported in Hombert et al. (1979). This different effect is also found in Thai; prevocalic unaspirated stops block H tone. Post-vocalic unaspirated stops, however, block M tone, but not H tone, as shown in (7).

(7) Checked syllables with aspirate or sonorant onsets (Ruangjaroon, 2006: 51-58)

		Onsets	**low**	**mid**	**high**	**falling**	**rising**
a.	**CVO**	aspirates[48]	[sàk] 'tattoo'	//////	[sák] 'wash'	//////	
b.		sonorants	[lòp] 'hide'	//////	[lóp] 'erase'	//////	
c.	**CV:O**	aspirates	[fà:t] 'acidulous'	//////		[fáàt] 'to eat'	//////
d.		sonorants	[mà:k] 'an areca palm'	//////		[máàk] 'very'	//////

---

[48] I assume that fricatives are [+spread glottis] as well (cf. Vaux, 1998).

(8)   Tone in checked syllables with an unaspirated onset (Ruangjaroon, 2006: 55, 60)

	unaspirated	low	mid	high	falling	rising
a.	CVO	[bòk] 'on land'				
b.	CV:O	[pà:t] 'to cut'				

3.1 Consonant-tone interaction constraints

For the analysis of the patterns in (5)–(8), the following markedness constraints, similar to the ones proposed in Lee (2008), will be introduced. The constraint *[-SPREAD GLOTTIS]/H in (9a) generally bans any segments that are specified with [-spread glottis] occurring next to a mora associated with H tone. The *[-SPREAD GLOTTIS]/H-ONSET constraint in (9b) is a specific constraint that only restricts unaspirated onsets from occurring before a H tone vowel. These two constraints interplay in the analysis of patterns found in checked syllables (7-8).

(9)   Consonant-tone interaction constraints

   a.   *[-SPREAD GLOTTIS]/H   (cf. Lee, 2008)
        Assign a violation mark to any syllable that has a segment with [-spread glottis] and the immediately adjacent mora is associated with a high tone.
   b.   *[-SPREAD GLOTTIS]/H-ONSET
        Assign a violation mark to any syllable that has a segment with [-spread glottis] and the immediately following mora is associated with a high tone.

A detailed OT analysis of Thai tone will be presented in section 4 based on the patterns of consonant-tone interaction. The main discussion of the analysis relates to the OCP constraints and the representation of mid tone in the Thai language.

4.   Analysis of consonant-tone interaction in Thai

The analysis of restriction of high tone in Thai is couched in Optimality Theory (Prince and Smolensky, 1993/2004). There are two main questions that should be answered. How do unaspirated onsets block high tone? How are falling tones allowed in the same environment? The blocking of high tone is due to the consonant-tone markedness constraint *[-SPREAD GLOTTIS]/H (*[-S.G.]/H), which is violated when a syllable with a [-spread glottis] segment is associated with a high tone. The more specific version of this constraint *[-SPREAD GLOTTIS]/H-ONSET (*[-S.G.]/H-ONS) plays an important role when it interacts with other constraints in the analysis of checked syllables.

The analysis of the Thai data should be able to account for how H tone is mapped to Falling tone in certain syllable types. In turn, the analysis will show what markedness constraints ban tones in certain syllable structures. The change of an underlying tone is the result of faithfulness constraints (the IDENT constraints in (3)) being strictly dominated by markedness constraints on consonant-tone interaction. In Thai, the consonant-induced ban of tones in syllables with an unaspirated stop is due to the markedness constraint that bans H tone on a mora immediately adjacent to an unaspirated consonant.

4.2 Unchecked syllables: the ban of H tone and the permission of HL tone

An OT analysis should account for the fact that a hypothetical input with H tone and an unaspirated onset such as /kú:/ does not surface faithfully. A faithful candidate [kú:] would violate the markedness constraint *[-S.G.]/H-ONS and it would satisfy all the faithfulness constraints. The optimal mapping of this input is [kū:] with mid tone because it minimally violates the lower ranked faithfulness constraint IDENT[+UPPER].

(10) Hypothetical input /kú:/

hypothetical input		optimal output
H		M
/\	→	/\
/k u u/		[k u u]

The tableau in (11) shows that *[-S.G.]/H outranks IDENT[+UPPER][49]. That is the reason why this hypothetical input is realized with a mid tone in the output. A low tone candidate in (11b) will be ruled out by IDENT[-LOWER]. The candidate (11d) has the H tone associated with the first mora changing to a mid tone. This candidate violates OCP[LOWER] because both mid tone and high tone share the feature [-lower]. The two other faithfulness constraints IDENT[+LOWER] and IDENT[-UPPER] are not included in the tableau because there is no feature in the input that these constraints can refer to. This is a case of emergence of the unmarked with high tone inputs.

(11) Ranking: *[-s.g.]/H >> Ident[+upper]

	H /\ /p a a/	OCP [LOWER]	IDENT [-LOWER]	*[-S.G.]/H ONSET	IDENT [+UPPER]
a. ☞	M /\ [p ā ā]				*
b.	L /\ [p à à]		W *		*
c.	H /\ [p á á]			W *	L
d.	M H | | [p ā̄ á]	W *			*

---

[49] There is another relevant candidate, in which the [-s.g] feature changes to [+s.g.]. A candidate such as [pʰaa] with an aspirated onset will satisfy all the constraints. This candidate is ruled out by the ranking in which the IDENT[SPREADGLOTTIS] constraint dominates the IDENT[+UPPER] constraint. In all the following tableaux, the candidate that changes [spread glottis] feature will not be shown assuming the relationship between IDENT[S.G.] and IDENT[+UPPER].

The presence of *[-S.G.]/H-ONS predicts that falling tone should not occur with unaspirated onsets either. However, an underlying falling tone /kûː/ is realized faithfully in the output as [kûː] 'borrow' as in (12). The optimal output does violate *[-S.G.]/H-ONS, but this output does not violate OCP[UPPER] and other faithfulness constraints.

(12) Optimal mapping /kûː/ → [kûː] 'borrow'

input		optimal output
H L		H L
| |	→	| |
/k u u/		[k u u]

The change of an input HL tone to an output L tone or ML tone is not optimal in Thai. The faithfulness constraint IDENT[-LOWER] is violated in an output mapping with an L tone (13a). In case of an unintended output ML tone, the candidate violates OCP[UPPER] as in (13b).

(13) Unintended outputs

a. input		unintended output	b. input		unintended output
H L		L	H L		M L
| |	→	∧	| |	→	| |
/k u u/		[k u u]	/k u u/		[k u u]
		Violating			Violating
		IDENT[-LOWER]			OCP[UPPER]

The aim of the analysis is to show that the proposed constraint system accounts for all the tonal occurrences and non-occurrences. The pivotal parts in the analysis are the OCP constraints on tonal features and consonant-tone interaction constraints. The OCP constraints strictly dominate consonant-tone interaction constraints, which results in the faithfull mapping between the HL input and the HL output.

The tableau in (14) represents the discussion so far. An unaspirated onset input with a falling tone is realized faithfully. Lowering of the first H in the contour tone to M as in (14b) violates the OCP[UPPER] constraint. Changing the whole syllable to an L tone as in (14c) violates the IDENT[-LOWER] constraint. Changing the whole syllable to a M tone as in (14d) can avoid the violation of the OCP[UPPER] constraint. However, the candidate in (14d) violates IDENT[+LOWER].

(14) OCP[upper], Id[+lower], Id[-lower] >> *[-s.g.]/H-Ons >> Id[+upper]

	H L | | /pa a/	OCP [UPPER]	IDENT [+LOWER]	IDENT [-LOWER]	*[-S.G.]/H ONSET	IDENT [+UPPER]
a. ☞	H L | | [pá à]				*	
b.	M L | | [pā à]	W *			L	W * (H→M)
c.	L \\ [p à à]			W * (H→L)	L	W * (H→L)
d.	M \\ [p ā ā]		W * (L→M)		L	W * (H→M)

4.2 Checked syllables: contour tones and mid tone

The analysis also needs to account for checked syllables. In checked syllables, the presence of a stop coda limits the tone to low and high only. When the onset of a checked syllable is an unaspirated stop, only low tone is found. In checked syllables with a short vowel (CVO, 15a), contours tones do not occur. Contour tones may be banned from occurring due to independent reasons: the higher ranked NOCONTOUR constraint, which assigns a violation mark on a monomoraic segment that is associated with more than two tones (cf. Gordon, 2002). In checked syllables with a long vowel (CVVO, 15b), H and LH tones are banned since the ranking in which the markedness constraint on consonant-tone co-occurrence does not allow a H tone mora to be adjacent to an unaspirated stop.

(15) Checked syllables and tone

	(C ≠ unaspirated)	L	H	LH	HL	M
a.	CVO	✓	✓	*	*	*
b.	CVVO	✓	*	*	✓	*

Analytical challenges regarding checked syllables in Thai are the ban on M tones and the permission of H tone on CVO syllables (shaded cells in (15)). The checkmarks in (15) indicate tonal patterns that surface in checked syllables with unaspirated onsets. The permission of H tone on CVO syllables is due to the relatively low ranking of the general markedness constraint *[-S.G.]/H. Changing the H tone input to M tone or L tone violates the tonal faithfulness constraints. Contour tones on a single mora are not allowed due to the NOCONTOUR constraint.

(16) [lóp] 'erase'

	H | /l o p/	No Contour	Ident [-lower]	*[-s.g.]/H Onset	Ident [+upper]	*[-s.g.]/H
a. ☞	H | [l ó p]					*
b.	M | [l ō p]				W * (H→M)	L
c.	L | [l ò p]		W * (H→L)			L
d.	L H ∨ [l ǒ p]	W *				*
e.	H L ∨ [l ô p]	W *				L

(17) [fáàt] 'to eat'

	H | /f a a t/	Ident [-lower]	Ident [+upper]	*[-s.g.]/H	Dep-T
a. ☞	H L | | [fá àt]				*
b.	M \\ [fā āt]		W * (H→M)		L
c.	L \\ [fà àt]	W * (H→L)			L
d.	H \\ [fá át]			W *	L

Checked syllables with long vowels have two realizations: L tone or HL tone. The HL output may have two different sources. If we assume an HL input, the input is realized faithfully in the output; no markedness constraint is violated by the faithful output. If a H tone input is assumed, then the output is realized with a falling contour HL by inserting a L tone after the input H tone. In OT, an analysis does not need to provide a unique underlying form for each output, a property called Richness of the Base (McCarthy, 2002: 68). Thus, the surface HL tone in the CVVO syllables comes from either an H tone input or an HL tone input.

Checked syllables do not permit M tone, while H tone and L tone are allowed in CVO and CVVO syllables. In (18), analogous to *[-S.G.]/H, I propose that there is the markedness constraint *[-S.G.]/M, which bans moras associated with an M tone from being immediately adjacent to unaspirated stops. This constraint has a positionally conditioned version *[-S.G.]/M-CODA, which does not permit an unaspirated post-vocalic consonant to be adjacent to an M tone. This constraint is constructed similar to Ruangjaroon's *[v̄]-[-SG] constraint (Ruangjaroon, 2006: 13)[50]. Since M tone does not appear in the output, the context-sensitive *[-S.G.]/M-CODA constraint should dominate the general consonant-tone interaction constraint *[-S.G.]/H. Thus, the grammar constructed so far predicts that an input M tone is realized with an H tone in the output.

(18) CVO with M tone: not permitted in Thai

	M | /CVO/	NO CONTOUR	*[-S.G]/M -CODA	IDENT [-LOWER]	*[-S.G.]/H
a. ☞	H | [CVO]				*
b.	M | [CVO]		W *		L
c.	L | [CVO]			W * (M→ L)	L
d.	L H \✓ [CVO]	W *			*
e.	H L \✓ [CVO]	W *			L

The mid and high tones share the feature [-lower], which suggests that the constraints *[-S.G.]/M and *[-S.G.]/H could be a single constraint *[-S.G]/[-LOWER], which assigns a violation mark when moraic segments associated with a [-lower] tone is adjacent to unaspirated stops. This constraint *[-S.G]/[-LOWER], however, would have banned H tone in CVO syllables, contra to what the Thai data shows. Thus, the *[-S.G.]/M constraint was used in the analysis.

In checked syllables with an unaspirated onset, only low tone syllables are permitted. This is due to the higher ranked contextual markedness constraint *[-S.G.]/H-ONS, which bans a H tone mora from being adjacent to an unaspirated consonant.

---

[50] Chen (2007) argues against Ruangjaroon's (2006) OT analysis of consonant-tone interaction. I will not go into details of this debate.

(19)   Checked syllables with unaspirated onsets and tone

	(C = unaspirated)	L	H	LH	HL	M
a.	CVO	✓	*	*	*	*
b.	CVVO	✓	*	*	*	*

The ranking of the constraints discussed in this paper is shown in (20). The markedness constraints NOCONTOUR and *[-S.G.]/M-CODA account for the patterns in CVO syllables. The NOCONTOUR constraint does not permit contour tones on a mono-moraic segment, so surface contour tones are not allowed. *[-S.G.]/M-CODA does not permit M tone on CVO syllables. The blocking of H tone in CV(S) syllables with an unaspirated onset is due to the ranking between *[-S.G.]/H-ONSET and IDENT[+UPPER], which allows the change of H tone while it bans unaspirated onsets from being adjacent to a mora associated with a H tone. Even so, HL tone surfaces in Thai, which is possible due to the higher ranked OCP and IDENT[-LOWER], which does not permit HL to become ML or L tone.

(20)   Constraint ranking

NOCONTOUR   OCP   ID[-UP]   ID[-LO]   ID[+LO]   *[-S.G.]/M-CODA

*[-S.G.]/H-ONSET   IDENT[S.G.]

IDENT[+UPPER]

*[-S.G.]/H

DEP-T
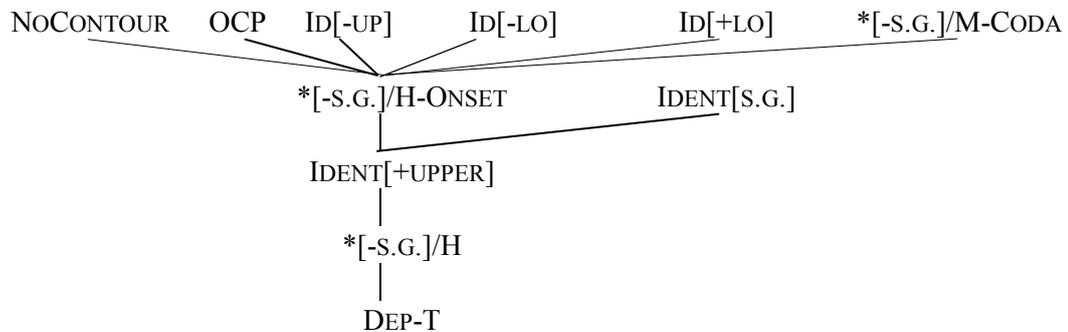
5. Discussion

5.1. Loanword phonology of Thai

The consonant-tone interaction data discussed in this paper is based upon native words of Thai. Loanwords in Thai, unlike native words, seem to be less concerned about observing the consonant-tone coocurrence restriction. A reviewer pointed out that Thai loanwords of the English words 'gas' and 'dad or father' are [káːt] and [páː], respectively. Both examples have high tone in syllables with unaspirated initial consonants. These words look like counterexamples to the analysis proposed in this paper. What these words show is that we need to consider the asymmetry between native phonology and loanword phonology in order to explain these loanword data. For example, in Korean, responses in loanword phonology may be different from those in native phonology. The prohibition of aspirated consonants in coda position is uniformly resolved by deaspiration, in the native Korean grammar. The same prohibition in the loanword grammar, however, may be resolved by inserting a vowel (Kang, 2003). Speakers of Japanese even show lexical stratification in their own native grammar (see Ito and Mester, 1995). Thus, it is not surprising that loanword phonology in Thai could demonstrate different behavior from native phonology. For more detailed discussions on this topic, see Kenstowicz and Suchato (2006), which is a corpus-based study on Thai loanword phonology.

5.2. A possible merger of high tone

There could be a phonetic explanation why the so-called high tone in Thai might be tolerated in recent loanwords reported in Kenstowicz and Suchato (2006). An examination of three acoustic studies of Thai tones (Bradley, 1911, Abramson, 1962, Potisuk et al., 1994), Teeranon (2007: 4) suggests that the phonetic characteristics of high tone in current Standard Thai have changed; "the high tone height has changed from high to mid, and its direction from falling to rising". Moreover, Teeranon's perceptual study reports that there is a generational difference in Thai tone perceptions. The older Thai speakers perceive high tone as a level tone, while younger speakers perceive high tone as a contour tone. When F0 movement is greater, younger speakers are less certain in distinguishing the difference between a high tone and a rising tone. There are also discrepancies between results of acoustic experiments of high tone and perceptual experiments of high tone; while acoustic studies suggest high tone is a contour tone, perceptual experiments do not seem to support these findings. In fact, the change of high tone and a possible merger with rising tone might be due to the perceptual similarities between these two tones (Morén and Zsiga, 2006, Thepboriruk, 2010). Even so, a note of caution should be added because perceptual studies can have noise in the results due to the subconscious linguistics nature of tone.

The tolerance toward the cooccurrence between high tone and unaspirated stops in recent loanwords in Thai could be due to the change of the characteristics of high tone. The Thai phonology may still enforce the consonant-tone cooccurrence restriction even though there is a change of acoustic properties of high tone.

5.3. Phonetics of tones in connected speech

Phonetic realizations of tones do not necessarily conform to their phonological behavior. For example, if unaspirated stops were physiologically incompatible with high tone, there should be no human language that can have unaspirated stops with high tone. This phonetic view of consonant-tone interaction is challenged because languages such as Mandarin Chinese do have unaspirated stops with high tone. Aiming to explain phonetic realizations of tones in consonant-tone interaction is beyond the scope of this paper. However, it is worth noting that tonal change is reported to occur in connected speech.

In connected speech, Thai exhibits tone neutralization of underlying lexical tones. Tingsabadh and Deeprasert (1997) report that both falling and rising tones lose their contour in connected speech; falling tones become similar to high tones, and rising tones become like low tones. Realizations of level tones in connected speech do not differ from citation forms. The neutralization of the falling tone to a high tone is interesting because it suggests that the low tone part of the falling tone might be as a result of phonetic implementation rather than a phonological low tone (see the paragraph below (6)). Phonologically, the neutralization of the falling tone to a high tone could be interpreted as a requirement for all tone to align at the left edge rather than the right edge, which would also explain the neutralization of the rising tone to a low tone in connected speech.

6. Conclusion

This paper addresses several points regarding various aspects of Thai phonology. First, syllables in Thai require minimum weight of at least two moras. Second, contour tones in Thai are combinations of two level tones. Third, two tone features are sufficient in describing Thai tone: [upper] and [lower]. Fourth, the OCP constraint blocks ML or HM

contour tones from emerging. Fifth, consonant-tone interaction constraints have a general and a position-specific version.

These claims are supported by the patterns of interaction between consonant and tone. H tone is blocked in a syllable with unaspirated onsets, whereas falling tone is not. The blocking of H tone unaspirated stops occurs due to the markedness constraint dominating the faithfulness constraint IDENT[+UPPER] (see (3)). The faithful mapping of HL tone surfaces because OCP and IDENT[-LOWER], IDENT[+LOWER] outrank the markedness constraint *[-S.G.]/H-ONSET. The faithful mapping of falling tone is due to the dominance relationship between the OCP[upper] constraint and the constraint *[-S.G.]/H-ONSET which in turn allows the marked HL form as an optimal output.

Another context-sensitive markedness constraint *[-S.G.]/M-CODA is proposed in order to explain the ban of mid tone in checked syllables. In general, H tone is permitted in checked syllables with unaspirated onsets (CVO syllables) because the general constraint *[-S.G.]/H is ranked lower than other constraints that rule out competing candidates. In Ruangjaroon's analysis, H tone is preserved in checked syllables because of the undominated markedness constraint MAX-H. Although MAX-H captures the generalization regarding the permission of H tone in CVO syllables, I claim that it is not necessary to posit this constraint. As shown in section 4, there is no constraint that would ban an H tone from occurring in a CVO syllable while allowing another tone to surface.

This analysis also differs from Ruangjaroon in terms of the formalism of the consonant-tone interaction constraint. The constraint *[-S.G.]/H locally assigns the violation when a mora associated with an H tone is adjacent to a segment specified with the feature [-spread glottis]. The markedness constraint *[-S.G.]/M also belongs to the family of these constraints. These consonant-tone interaction constraints can be context-sensitive as I have proposed. In particular, it is necessary to posit the *[-S.G.]/H-ONS constraint and the *[-S.G.]/M-CODA constraint; these constraints are positionally sensitive. Thus, the advantage of this analysis over Ruangjaroon's is that it does not require constraints that assess violation non-adjacently. The positional asymmetry in consonant-tone interaction is captured using general markedness constraints and specific markedness constraints.

As mentioned in section 1, this paper is based on a variety of Thai tone reported in Ruangjaroon (2006). Her data provides empirically interesting observations as well as theoretically challenging puzzles. Results from a production experiment in Perkins (2009) reports slightly different patterns from Ruangjaroon's observations. Perkins (2009), in particular, comments that rising tone is underrepresented in the Thai lexicon. These latter studies suggest that the tonal grammar in Thai may look different from Ruangjaroon. As suggested by Perkins (p.c.), it could be that Ruangjaroon holds a view of the categorical grammar, while the others assume the gradient grammar. Further investigation in this area is much needed in order to understand the dynamics of consonant-tone interaction in Thai.

References

Abramson, Arthur S. 1962. The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics* 28:2, Part III.

Anderson, Stephen R. 1978. Tone features. In *Tone: A Linguistic Survey*, ed. Victoria Fromkin, 133-176. New York: Academic Press.

Bradley, Cornelius. 1911. Graphic analysis of the tone-accents of the Siamese language. *Journal of the American Oriental Society* 31:3, 282-289.

Bradshaw, Mary. 1999. A Crosslinguistic Study of Consonant-Tone Interaction, Department of Linguistics, Ohio State University: Ph. D. Dissertation.

Chen, Tsung-Ying. 2007. Problems of Analyzing Consonant-tone Interaction in Thai: A reply to Ruangjaroon. *Taiwan Journal of Linguistics* 5:2, 65-84.

Gandour, Jack. 1974. On the Representation of Tone in Siamese. *UCLA Working Papers in Phonetics* 27, 92-117.

Gordon, Matthew. 2002. A typology of contour tone restrictions. *Studies in Language* 25:3, 423-462.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality, and tone. In *Tone: A Linguistic Survey*, ed. Victoria Fromkin, 77-111. New York: Academic Press.

Hombert, Jean-Marie, Ohala, John J., and Ewan, William G. 1979. Phonetic explanations for the development of tones. *Language* 55:1, 37-58.

Hyman, Larry, and Schuh, Russell. 1974. Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* 5, 81-115.

Intrasai, Jiraporn. 2001. Topics in Thai tonology: an acoustic investigation and some pedagogical applications, University of Delaware: Ph. D. Dissertation.

Ito, Junko, and Mester, Armin. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *Papers in Optimality Theory*, eds. Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, 181-210. Amherst: GLSA Publications.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20, 219-273.

Kenstowicz, Michael, and Suchato, Atiwong. 2006. Issues in loanword adaptation: a case study from Thai. *Lingua* 116:7, 921-949.

Leben, W. 1971. On the segmental nature of tone in Thai. *Quarterly Progress Report* 101, 221-224.

Lee, Seunghun J. 2008. Consonant-Tone Interaction in Optimality Theory, Rutgers University: Ph. D. Dissertation.

Lee, Seunghun J. 2009a. Neutralization to high tone by consonants in Mulao. *Journal of Southeast Asian Linguistics Society* 2, 73-84.

Lee, Seunghun J. 2009b. Effects of glottal stops on tone in Burmese. In *Proceedings from the Main Session of the Forty-third Annual Meeting of the Chicago Linguistic Society*, eds. Malcolm Elliott, James Kirby, Osamu Sawada, Eleni Staraki and Suwon Yoon, 199-213. Chicago, IL: The Chicago Linguistic Society.

Lee, Seunghun J. 2009c. H tone, depressors and downstep in Tsonga. In *Selected Proceedings of the 38th Annual Conference on African Linguistics: Linguistic Theory and African Language Documentation*, eds. Masangu Matondo, Fiona McLaughlin and Eric Potsdam, 26-37. Somerville, MA: Cascadilla Proceedings Project.

McCarthy, John J. 2002. *A Thematic Guide to Optimality Theory*: Research Surveys in Linguistics. Cambridge: Cambridge University Press.

Morén, Bruce, and Zsiga, Elizabeth. 2006. The Lexical and Post-Lexical Phonology of Thai Tones. *Natural Language & Linguistic Theory* 24:1, 113-178.

Perkins, Jeremy. 2009. Consonant-Tone Interaction and Glottalization in Thai. Ms. Rutgers, The State University of New Jersey.

Potisuk, Siripong, Gandou, Jackson T., and Harper, Mary P. 1994. F_0 correlates of stress in Thai. *Linguistics of the Tibeto-Burman Area* 17:2, 1-27.

Prince, Alan, and Smolensky, Paul. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Basil Blackwell.

Rischel, Jørgen. 1986. Can "The great tone split" in Thai be phonetically explained? *Annual Report of the Institute of Phonetics, University of Copenhagen* 20, 79-98.

Roengpitya, Rungpat. 2000. Can a perceptual experiment reflect tonogenesis in Tai? In *Tai Studies Proceedings, The International Conference on Tai Studies (Thailand July 29-31, 1998)*, ed. Somsonge Burusphat, 361-369. Nakhon Prathom, Thailand: ILCRD, Mahidol University.

Ruangjaroon, Sugunya. 2006. Consonant-Tone Interaction in Thai: An OT Analysis. *Taiwan Journal of Linguistics* 4:2, 1-66.

Tang, Katrina E. 2008. The Phonology and Phonetics of Consonant-Tone Interaction, UCLA: Ph. D. Dissertation.

Tang, Katrina Schack. 2007. Voice-Tone Interaction in Bade. Paper presented at *the Conference on the Typology of Tone and Intonation (TIE), Berlin, Germany*.

Teeranon, Phanintra. 2007. The change of Standard Thai high tone: an acoustic study and a perceptual experiment. *Journal of Theoretical Linguistics, The Slovak Association for the Study of English (SKASE)* 4:3, 1-16.

Thepboriruk, Kanjana. 2010. Bangkok Thai Tones Revisited. *Journal of Southeast Asian Linguistics Society* 3:1, 86-105.

Tingsabadh, K.M.R., and Deeprasert, D. 1997. Tones in standard Thai connected speech. In *Southeast Asian Linguistics Studies in honor of Vichin Panupong*, ed. A. S. Abramson, 297-308: Chulalongkorn University Press.

Tumtavitikul, Apiluck. 1992. Consonant onsets and tones in Thai, University of Texas at Austin: Ph. D. Dissertation.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29:3, 497-512.

Wong-opasi, Uthaiwan. 1992. The interplay between tone, stress, and syllabification in Thai. In *Papers from The First Annual Meeting of The Southeast Asian Linguistics Society 1991*, eds. Martha Ratliff and Eric Schiller, 441-481: Arizona State University Program for Southeast Asian Studies.

Woo, Nancy. 1969. Prosody and Phonology, MIT: Ph. D. Dissertation.

Yip, Moira. 1980. The Tonal Phonology of Chinese, MIT: Ph. D. Dissertation.

Yip, Moira. 1989. Contour tones. *Phonology* 6, 149-174.

VOLUNTARINESS AND SPONTANEITY IN THAI

Makoto Minegishi
Research Institute for Languages and Cultures of Asia and Africa (ILCAA),
Tokyo University of Foreign Studies, Japan
<mmine@aa.tufs.ac.jp>

Abstract
By examining the occurrence of Thai verbs in negative, imperative, causative, and passive constructions, this paper claims that the most fundamental semantic contrast in Thai verbs is between voluntariness and spontaneity, rather than between transitivity and intransitivity. On the basis of this semantic distinction, we further assume that the most basic sentence structure in Thai is serial verb construction, which we attempt to analyze linearly, that is, avoiding binary and hierarchical representation. This enables us to determine that the causative and passive constructions are anti-parallel to each other, as entailed in serial verb constructions.

Keywords: voluntariness, spontaneity, affectedness, serial verb construction

Introduction
This paper is a preliminary report on the examination of Thai verbs in terms of voluntary and spontaneous contrast. We will first examine verbs in minimal syntagma; i.e., co-occurrence with the negator and the imperative marker. We will then extend our examination to larger syntactic constructions, such as serial verbs, causatives, and passives.[51] Since constituents and constructions make a cline, redefining constituents affects the description of constructions that accommodate the constituents. Thai, whose canonical word order is SVO, is a "pro-drop" language; the presence or absence of subject or object is conditioned by the preceding context, which means that distinguishing transitive from intransitive verbs is practically impossible in actual texts. We may then assume that the transitive-intransitive contrast is irrelevant in verb classification. If so, it is worthwhile to find another semantic property that is preferably constant and not affected by the presence of object, and relevant to syntactic constructions that accommodate verbs. In this paper, we will discuss the following two issues:

1. Examining verbs based on the semantic contrast: voluntariness vs. spontaneity
2. Attempting a linear and dynamic analysis of serial verb constructions

1. Previous studies on verbal subclassification and transitivity

---

[51] I would like to express my gratitude to the anonymous reviewers for helpful comments and suggestions on the early version of this paper, which was presented in SEALS XX originally titled "Verb Subclassification in Thai Based on Syntactic Phenomena". The remaining faults are all my own. This research has been supported by the Global Center of Excellence Program "Corpus-based Linguistics and Language Education (2007–2012)" under the auspices of the Japan Ministry of Education, Culture, Sports, Science and Technology.

As is the case in other isolating languages without verbal inflection and nominal case marking, Thai verbs are classified in terms of syntactic distribution in sentences and their semantic properties. Many verbs that can co-occur with direct objects are generally classified as transitive verbs. In addition, there are verbs that have no direct objects, which are classified as intransitive verbs, stative verbs, or adjectives according to the theoretical frameworks of each description.

One of the earliest attempts at classifying Thai verbs was made by Noss (1964: 114–129) based on American linguistic distributionism. He classified "predicatives" that function as predicators into modal verbs, adjectives, transitive verbs, and completive verbs; verbs in the third sub-group that co-occur with direct objects, irrespective of their semantic roles, are also referred to simply as "verbs."[52] He also stated that all transitive verbs occur both with and without objects, and both with and without subjects, which is now called the "pro-drop" feature. On the basis of his description, we may assume, by adopting neutral expressions concerning semantic roles, that Thai verbs can be used as both two-place verbs and one-place verbs. Having or not having an object should then be a description of the syntactic or discourse environment rather than the semantic properties of a verb. Thai verbs have both two-place usage with objects and one-place usage without explicit objects; therefore, the dichotomy transitive versus intransitive as a semantic property of verbs does not make sense.

Noss also stated that semantic roles assumed by the subject and the object in the subject-verb-object construction are not restricted to agent and patient. Although he found that "[t]he meaning of the verb-object construction is that the referent of the object is the goal of the action designated by the verb. The meaning of the subject-verb-object construction is that the referent of the subject is the actor initiating action toward that goal," he remarked that "[...] the subject may refer either to the actor or the goal...," (*ibid.*: 123). The same subject-verb-object constructions, hence, may vary with respect to their transitivity and the semantic roles of subject and object.

Hopper and Thompson (1980) proposed the transitivity hypothesis, claiming that transitivity, as a discourse-determined notion, is a crucial relationship in language. They proposed 10 properties, which we would re-group into three categories: (1) properties concerning the grammatical meaning of a sentence (A. Participants, C. Aspect, D. Punctuality, F. Affirmation, and G. Mode), (2) properties regarding the agent (B. Kinesis, E. Volitionality, and H. Agency), and (3) properties concerning the object (O) (I. Affectedness of O, and J. Individuation of O). They claimed that the topmost parameter A, that is—two participants involved—is the most important, since "no transfer at all can take place unless at least two participants are involved." Two-place constructions such as *"Susan left Jim"* have higher transitivity than one-place constructions such as *"Susan left."*

Although the above properties are not meant for classifying verbs, we can assume that verbs with more transitivity properties have higher transitivity than those with fewer. It should be noted, however, that parameters regarding the agent, such as kinesis, volitionality, and agency are relevant to voluntary activities in general. In many cases, we can determine the values of the parameters concerning voluntary activities regardless of the

---

[52] By his definition, Noss's "adjectives" as a sub-group of predicatives are equivalent to "adjectival verbs," analyzed by Prasithrathsint (2000).

number of participants. We may assume that, concerning the above English examples by Hopper and Thompson, *Susan's leaving* is a voluntary action regardless of the presence or absence of the object *John*. It may be the case, then, that the constant properties of voluntariness, rather than transitivity, can be attributed to verbs themselves because they are independent of syntactic environments.

Tsunoda (1985), in response to Hopper and Thompson's discussion, focused on the affectedness of the objects. He claimed that higher affectedness of the patient indicates higher transitivity, and that higher transitivity is marked with nominative-accusative case marking, whereas lower transitivity is marked with various case markings in different languages, such as nominative-dative and dative-nominative.

In what follows, we will show that in the semantic analysis of Thai verbs, both voluntariness and affectedness are important. We will further show that the direction of affectedness is important—in most cases, the agent affects the patient, but there are also cases in which the agent, or experiencer, is affected by the object that assumes the role of source, locus of source, or cause.

2. Methods for examining semantic features in Thai verbs

As part of the Thai–Japanese electronic dictionary project, preliminary study of Thai verbs is in progress which aims at analyzing semantic features of verbs listed as the most basic (approximately 300) verbs in the linguistic questionnaire published by the Institute for the Studies of Languages and Cultures of Asia and Africa in 1966. In the initial stage of the project, we examined to what extent Tsunoda's transitivity scale of two-place predicates (*ibid.*: 388, Table I) is valid. We used the following method:

1. Choose Thai verbs corresponding to the two-place verbs found in Tsunoda's table of affectedness to apply the following tests regarding the agent (A)'s voluntariness.
2. Check whether the intention is negated with *mây* (negative marker).
3. Check whether the verb can be in an imperative construction with *sìʔ* (imperative final particle).
4. Check whether the verb can be used in passive construction with *thùuk* (passive marker).

Among the above, 2 and 3 concern the agent's voluntariness, while 4 is to examine the patient's affectedness.

It should be noted here that the number of verbs that appear in Tsunoda's table is limited since his table is designed for cross linguistic examination of canonical and non-canonical case marking patterns of constructions with two participants. It is therefore natural that intransitive verbs, and verbs followed by an "adjunct", are not found in the table, which should be taken into consideration in order to examine general semantic properties in Thai verbs.

3. Results

3.1 Voluntary vs. spontaneous contrast in verbs

Table 1 shows the results of the examination. Note that the x-y coordinates are swapped as compared to Tsunoda's original table for the convenience of page layout. The three columns on the left correspond to the types, their meanings, and examples of verbs given

by Tsunoda. In addition, the leftmost column contains the abbreviations Vv and Vs; Vv denotes the voluntary verb, Vs, the spontaneous verb, respectively. We will show below that the Vv versus Vs contrast is the most fundamental one in Thai verbs. The columns on the right are corresponding Thai verbs, the results of our test as to the negated intention, imperative construction, and passive construction. "Yes" or "No" in the right three columns shows the results of the examination. "(No)" in parentheses means that the result is generally negative, although there are exceptional cases with "Yes." It should be noted that the order of rows is changed from Tsunoda's original arrangement to reflect the contrast in the results. The topmost row has all affirmative results, whereas the bottom row has all negative ones. The rows containing parenthesized "(No)" are placed in the middle. Table 1 shows clear contrast between Type 1A plus 1B and Type 2A, while Type 3, 4, 5, 6, and 7 that contain exceptional cases show gradience. Tsunoda claimed that verbs in Type 1A, having direct effect on the patient, are prototypical transitive verbs with the highest transitivity. Affectedness of the patient is less in Type 1B, Type 3, etc., and the least in Type 7. Verbs in Type 7 therefore have the least transitivity.

Table 1: *Voluntary and spontaneous verbs*

Type	Meaning	Example	Thai	Negated Intention	Imperative	Passive
1A Vv	Direct effect on patient	Kill...	*khâa (kill), ʔùn (warm)*	Yes	Yes	Yes
1B Vv		Hit...	*tii (hit), tèʔ (kick), chon (collide)*	Yes	Yes	Yes
3 Vv	Pursuit	Search...	*rɔɔ (wait for), hǎa (find)*	Yes	Yes	(No)
2B Vv	Perception	Listen...	*duu (watch), faŋ (listen)*	Yes	Yes	No
4 Vs	Knowledge	Know...	*rúu (know), khâwcay (understand), cam (remember), luuum (forget)*	(No)	(No)	(No)
5 Vs	Feeling	Love...	*rák (love), chɔ̂ɔp (like), klìat (hate), kròot (get angry), klua (afraid of)*	(No)	(No)	(No)
6 Vs	Relationship	Possess...	*mii (have), mǔan (resemble), khláay (change), khàat (lack), pen (be)*	(No)	(No)	No
7 Vs	Ability	Capable...	*dây (able), pen (skilled), kèŋ (skilfull), thon (endure)*	No	(No)	No
2A Vs	Perception	See...	*hěn (see), dâyyin (hear), cəə (meet)*	No	No	No

The first thing to be observed in the results is that Thai verbs appear in the subject-verb-object construction in all types of Tsunoda's classification, which shows that the word order is canonical. There is an exception, that is, Type 7, for which Thai verbs are mostly in the serial verb construction. Among Type-7 verbs, however, *thon* (endure) is an exceptional two-place verb. Table 1 explicitly shows that in spite of Tsunoda's claim, Thai

verbs show no difference in case marking as to the scale of transitivity based on affectedness of the patient. In fact, the degree of affectedness among Tsunoda's types is not always clear, except for Type 1A, where the result of physical damage by an action can be observed, and Type 1B, where a patient is physically affected. For the other types, it would be practically impossible to tell how and to what extent the patient is affected. Tsunoda's classification, then, reflects patterns of non-canonical case marking gradually changing according to the types of verbs, rather than the degree of affectedness.

Thai verbs listed in Table 1 can be roughly classified into two groups: the first group includes Types 1A plus 1B, 3, and 2B denoted each as Vv, and the second group, Types 4, 5, 6, 7, and 2A, denoted as Vs. Among these, Type 1A plus 1B and Type 2A shows clear-cut contrast since no exception is observed in the results. Type-1A-plus-1B verbs can therefore be regarded as prototypical Vv, and Type-2A verbs as prototypical Vs.[53] The distinction is made based on the results of intended negation and imperative construction; the results of passive construction can be ignored since the passive is limited in Thai, even for verbs with a direct effect on the patient.

Vv's listed above include the Type-1A verbs *khâa* (kill) and *ʔùn* (warm); Type-1B verbs *tii* (hit), *tèʔ* (kick), and *chon* (collide); Type-3 verbs *rɔɔ* (wait) and *hăa* (seek); and Type-2B verbs *duu* (watch) and *faŋ* (listen). These are intentional verbs denoting direct effect on the patient, pursuit, or perception, which needs a human subject in principle. Vv's denote actions that are controllable by the agent. When negated, Vv's imply that the agent's intention to do something is also negated. Further, since Vv's indicate voluntary action, these verbs can be used in the imperative construction.

On the other hand, the Vs's listed above include the Type-4 verbs *rúu* (know), *khâwcay* (understand), *cam* (remember), and *luuum* (forget); Type-5 verbs *rák* (love), *chɔ̂ɔp* (be fond of), *klìat* (hate), *kròot* (be angry with), and *klua* (be afraid of); Type-6 verbs *mii* (have), *mŭan* (be the same as), *khláay* (resemble), *khàat* (lack), and *pen* (become); Type-7 verbs *dây* (be capable), *pen* (be learned), *kèŋ* (be skilled), and *thon* (endure); and Type-2A verbs *hĕn* (see), *dâyyin* (hear), *and cəə* (meet). These verbs denote knowledge, feeling, relationship, ability, and perception, which do not represent real kinetic action, but state, change of state, existence, and various natural phenomena, processes, or accidental events that are beyond or irrelevant to human control. It should be noted that many of Vv's listed in Table 1 are also stative verbs which denote either states or change of states in appropriate contexts. We assume that the feature 'spontaneity' together with the direction of 'affectedness' is more important than 'state' since the former is relevant to analyzing syntactic constructions as shown below.

3.2 Direction of affectedness

Vv's denote the agent's voluntary involvement in the action. In other words, Vv's assume that the agent can control his kinetic action affecting the patient. Accordingly, in the case of the Type-1A verbs *khâa* (kill) and *ʔùn* (warm), and Type-1B verbs *tii* (hit), *tèʔ* (kick), and *chon* (collide), it is the patient that is affected by the agent. On the other hand, Vs's indicate "passive" involvement in uncontrollable events. In cases in which Vs has a human subject,

---

[53] We propose to call the latter verbs "spontaneous" verbs in contrast to the former "voluntary" ones, because we would like to pose the contrast as a "polar contrast," whereas the terms intransitive, unintentional, and passive all have an unmarked connotation in contrast to the marked transitive, intentional, and active.

the subject assumes the role of experiencer, rather than agent, perceiving the stimulus given by the existence of the object. Thus, in the case of the Type-2A verbs *hěn* (see), *dâyyin* (hear), and *cǝǝ* (find), it is the subject, not the object, that is affected by the existence of the object. As a result, Table 1 shows that not only the existence of affectedness, but also the direction of affectedness is important.[54]

The following examples (1) to (3) illustrate the contrast between Vv and Vs. Arrows (→ and ←) denote the direction of affectedness.

(1) **1A/B (Vv)**   *dɛɛŋ*   *tii*   *lék*
                      Daeng   hit   Lek
                      'Daeng hit Lek.'

In (1) the agent (A) voluntarily affects the patient (P); thus, the direction of affecting is A → P.

(2) **2A (Vs)**   *dɛɛŋ*   *hěn*   *mǎa*
                      Daeng   see   dog
                      'Daeng saw a dog.'

In (2), the experiencer (A) is involuntarily affected by the source (P); therefore, the direction of affecting is A ← P.

It should also be noted that Vv's include the Type-2B verbs *duu* (watch) and *faŋ (listen)*, which indicate the agent's voluntary involvement in perceiving, which would not be always successful in that perception is not perfectly controlled by the agent's intention.

(3) **2B (Vv)**   *dɛɛŋ*   *duu*   *nǎŋ*
                      Daeng   watch   movie
                      'Daeng watches the movie.'

In (3), the agentive experiencer (A) voluntarily participates in perception, and he is spontaneously affected by what he perceives at the same time; affecting is hence bidirectional: (A → P) and (A ← P).

3.3 Exceptional cases

As is shown above, prototypical Vv's have direct physical effects on patients; therefore, they can be used to form passive and imperative statements. Vs's with human subjects concern mental activities: perception, sentiment, feeling, relationships, etc. There are, however, exceptions to the results of our examination, as follows.

First, some Vs's for mental activities can be treated as voluntary when they appear in the negated intention construction; among Vs's, for example, *lɯɯm* (forget) with the negative marker means that its subject intends, or wishes, not to forget, although it is evident that holding on to memories is not under one's direct control. Similarly, uncontrollable Vs such as *taay* (die) can be used in the imperative construction, although it

---

[54] Direction of affectedness in the lexical level matters not only in languages in mainland Sotuheast Asia such as Thai, but also in Japanese. Imaizumi (2001) analyzes the direction of affectedness in Japanese within the framework of Lexical Conceptual Structure.

is impossible to do so literally; what we can do is initiate some voluntary action that may cause the desired change of state. Human beings may wish for something unfavorable not to happen or for something favorable to happen, although we know that we cannot control whatever spontaneous events may happen. Thus, when Vs's appear exceptionally in intentional or imperative constructions, their "spontaneous" property can be regarded as overridden by the constructions.

Second, some verbs, such as *pə̀ət* (open), *pìt* (close), and *ʔɔ̀ɔk* (produce) can be used both as Vv and Vs, according to the subject with which they co-occur. Refer to Noss (1964: 124) and Sakamoto (1985) for such verbs. Compare the following examples (4a) to (4c).

(4a)   *pratuu*   *nán*   *pə̀ət*
       door   that   open
       'That door is open.'

(4b)   *dɛɛŋ*   *pə̀ət* *pratuu*   *nán*
       Daeng   open door   that
       'Daeng opened that door.'

(4c)   *pratuu*   *nán*   *dɛɛŋ*   *pə̀ət*
       door   that   Daeng   open
       'As for that door, Daeng opened it.'

The verb *pə̀ət* (open) in (4a) preceded by [-Human] subject is Vs, while the same verb in (4b) and (4c) following [+Human] subject is Vv. Since there are limited situations that can be seen as either an event irrelevant to any human agentive force, or the result of some agentive force, the number of such verbs is relatively small. We denote these verbs with "Vv-s" to show that they permit both voluntary and spontaneous usages. Another group of verbs to be classified as Vv-s are verbs denoting motions such as *pay* (go), *maa* (come), etc. since motions can be either voluntary, spontaneous, or out of control. It might be the case that, similar to the transitive-intransitive contrast, the distinction between Vv–Vs should be attributed to the constructions in which they appear, rather than the properties of the verbs. We assume for the present that the Vv-Vs contrast is part of the verbal properties because the number of Vv-s verbs is limited; most verbs can be classified as either Vv or Vs without consideration of their environments.

3.4 Semantic roles of complements for Vv and Vs

The fact that the two-place construction is canonical in Thai does not mean that the nominative-accusative case marking is presumed to the construction, which is a common assumption in English grammar. In Thai, most Vv's are followed by objects that play the semantic role of patient (or product), for example, *tii* (hit), *mɔɔŋ* (look at), *duu* (watch), and *tham* (make). Some verbs may be followed by a locus, for example, *nâŋ* (sit) and *nɔɔn* (lie down). Other verbs denoting "movement" take objects denoting a goal, for example, *pay* (go) and *maa* (come), which could be included in the locus. Some verbs take objects denoting an instrument or means, for example, *kin* (eat with) and *pay* (go by). These semantic roles are common in different "isolating languages." Because the semantic roles of the direct objects of Vv's vary according to the meanings of the verbs, it is preferable to use the rather traditional term "complement" instead of objects.

Vs's with human subjects, on the other hand, are generally not followed by a patient. Although verbs denoting perception, such as *hěn* (see) and *dâyyin* (hear) may seem to take the patient as a complement, they should, however, be regarded as having locus, or source of stimulus, since there are also similar sentiment verbs such as *klua* (afraid of) and *kròot* (angry with). This analysis more plausibly allows other feeling verbs to be taken into consideration, such as *pùat* (ache) and *cèp* (be sore), which take locus or source of stimulus or cause, but not patient. These perceptive or feeling verbs co-occur with [+Human] subjects and complement expressing locus, or source of stimulus, which functions as a cause to affect the human mind or feelings.

3.5 Summary of the voluntary and spontaneous contrast
Table 2 summarizes the contrast between Vv and Vs. Since the methods adopted here concern the agent's controllability, [+Human] subject with Vv is generally expected. Other properties mostly co-vary with [+Human].

Table 2: *Voluntary and spontaneous contrast*

	Voluntary Verbs (Vv)	Spontaneous Verbs (Vs)
Subject	+Human	+/-Human
Voluntariness	+Voluntary	-Voluntary
General meaning	Voluntary action	Spontaneous (change of) state
Kinesis	+Kinetic	-Kinetic
Mental activity	-/+Mental	-/+Mental
Controllability	+(limited) Control	-Control
Direction of affectedness	Subject → Complement	Subject ← Complement
C's semantic roles	Patient, (Product), Locus (Goal), Instrument	Locus, or Source of Stimulus, or Cause

It should be noted that because the data given above is based on previous studies in transitivity, which focus on verbs mostly for human activities, verbs with non-human subjects are not examined extensively in this paper. We can predict that since non-human subjects cannot be volitional, they must be Vs's in principle. In fact, Thai has verbs with non-human subjects that appear in two-place constructions, such as verbs for describing phenomena in the outer world without human beings involved. Since the subject-verb-object construction is canonical in Thai, verbs that denote natural process or accidental events with two participants appear in the transitive construction. Such verbs must have low transitivity even if they take direct objects, for which further elaborated examination would be necessary.[55]

---

[55] Iwasaki (2002), for example, discusses "putatively intransitive" constructions [N1 V N2] for "proprioceptive-state" expressions. Takahasi (2007) also discusses "intransitive verbs with direct objects" whose [direct] objects denote body parts. On the basis of our discussion, they would be re-grouped into two-place verbs: (a) Vs's with non-human subjects whose direct objects assume patient or product, for example, *ʔɔ̀ɔk phǒn* (bear fruit), *loŋ* (take root, rain), and *khûn sanǐm* (get rusty); (b) Vs's with human subjects whose objects assume locus or source of stimulus, for example, *tòk cay* (fall heart=be surprised), *thanàt mɯɯ khwǎa* (be skillful with the right hand), *saʔàat taa* (be clean to the eyes), and *cèp taa* (to have sore eyes); (c) Vs's with non-human subjects whose objects assume locus or source of stimulus, for

4. Analyzing serial verb constructions

4.1 Basic assumptions

We have seen in the above that the semantic contrast between voluntary and spontaneous is relevant to constructions denoting negated intention, imperative, and passive. We claim the distinction is fundamental in Thai because it plays a significant role in forming other syntactic constructions, such as serial verb constructions, as well. The serial verb construction (or verb serialization) is defined by Bisang (1991:509) as "the unmarked juxtaposition of two or more verbs or verb phrases (with or without subject and/or object), each of which would also be able to form a sentence on its own". Bisang's definition which also appears in Bisang (1995) is made on the basis of his analysis of converbs and serial verbs in East, Southeast, and South Asian languages. It should be noted that according to his definition, passive and causative construction in Thai and other mainland Southeast Asian isolating languages can be regarded as part of the serial verb constructions.

On the basis of the semantic distinction in the verbs above, we will attempt to analyze serial verb constructions that accommodate verbs. We start with two basic assumptions.

1. Assume serial verb construction (hereafter abbreviated as "SVC"), not single verb construction, as basic in Thai and other isolating languages in mainland Southeast Asia.
2. Regard SVC as an open-ended concatenated structure, rather than a hierarchical (embedded, or binary-branching) structure.

In what follows, we will only point out what types of constructions reflect the Vv–Vs contrast. Close examination of each construction by comparing with alternative hierarchical analysis is yet to be done.

Thai is a "verby" language that allows a structurally endless concatenation of verbs or verb phrases such as (5).

(5) *dəən khâam*   *(saphaan lɔɔy)*   *pay kin fàŋ nóon*
     walk cross   (pedestrian bridge)   go   eat   side that
     'To walk over (the pedestrian bridge) to eat on that side'

If a series of verbs denotes successive events, as in (5), you could add more verbs as far as your memory permits. It should be noted here that in (5), the verbs all share the same covert subject, and they are all Vv's. Concatenations, for example, with alternative appearance of Vv and Vs, such as "Vv+Vs+Vv+Vs…" are not acceptable, which suggests that the Vv–Vs contrast plays some role in forming verb serialization. It should also be noted that in (5), verbs are so arranged as to be in accord with the temporal flow of events.

---

example, *klay taa* (be out of sight), etc. These are Vs's in that the state or change of state is out of the subject's control. Moreover, there are Vs's in group (c) such as *pìak* (get wet) in *sûa pìak fŏn* (clothes got wet in the rain) whose object *fŏn* (rain) is not part of clothes but the cause of the change of state, which suggests that the semantic roles of objects of Takahasi's "intransitive" verbs are not restricted to body parts.

This explains why the grammatical tense, which expresses the particular temporal point, is nonexistent in Thai; Thai serialized verbs together show the direction of change, or successive events.

4.2 Linear approach to the language structure

For more than a half century, the hierarchical analysis, which assumes binary-branching structures in every level of linguistics, has attracted linguists because it appears to be unique to the human language, making binary branching the most important feature. Even in recent studies of Thai, many linguists have tended to assume the hierarchical structure in SVC's, and in other constructions such as causative and passive ones. We will attempt here to examine another perspective for analyzing Thai syntax.

We claim that an SVC, such as in (5) above, is a simple open-ended concatenation of relatively homogeneous verb phrases, rather than a hierarchical multi-embedded structure. Applying hierarchical binary branching such as X-bar structure would bring unnecessary complexity to describing an SVC, which would be analyzed as a series of complicated clauses, each of which consists of only a verb and empty pro. This is because binary branching or phrase structure grammar (PSG) assumes the clause to be a fixed domain in which each constituent occurs in a fixed order and a fixed number of times. What we point out is that concatenation, or multiple coordinating constitutes, are out of the scope of X-bar syntax. Suppose an SVC consists of multiple embedded clauses. It is then practically difficult to process in the level of language performance. See Chomsky (1965: 12–13) regarding his remark on linguistic performance, stating, "repeated nesting contributes to unacceptability" while "multiple-branching constructions [i.e., concatenation] are optimal in acceptability," and Jackendoff (1977: 50–51) who referred to "coordination" as "one obvious exception to the theory of phrase structure" and left it without further consideration. Since then not many works have been done for concatenations.

Phonology provides a good example to contrast a binary branching and a linear analysis. The binary-branching analysis is suitable for tone, such that a syllable is divided into onset and coda, then the latter is further divided into a vowel nucleus and a final consonant. Note that every constituent in the syllable domain is fixed concerning the order and number of occurrence. By so dividing, one can refer to different levels of the syllable when necessary, that is, referring to a whole syllable as a fixed domain to which each tone is assigned, or referring to the coda as a domain for tonal contour, or referring to the onset concerning the historical change of tonal contour, etc.

On the other hand, the linear analysis is suitable for describing an open-end construction such as a domain of accent, which consists of several syllables, the number of which is not predictable, or can be extended in case of forming compounds. The linear analysis of the Japanese accent system was initiated by Shiro Hattori, and has been succeeded and established by a leading phonologist, Zendo Uwano, with his extensive study of Japanese dialects over 30 years. According to his analysis, in short, a Japanese word consists of concatenated syllables, that is, CV-CV-CV-CV…, where C denotes a consonant, V a vowel, respectively. Instead of assigning high or low tone as a property to each syllable, assuming accent patterns with an accent nucleus to one syllable (or mora), all the possible contour patterns can be neatly described. Refer to Uwano (1999), (2007) for an overview of Japanese dialectal varieties in accent system.

4.3 Linear analysis of serial verb constructions

We attempt to apply the linear analysis to SVCs. The basic strategy for analysis inspired by that of the Japanese accent is as follows.

1. Assume linear structures as far as possible, avoiding "trees" or embedded structures.
2. Think of "adjacent" constituents, that is, referring the immediate precedent constituent to the next one, avoiding referring to elements far in the distance, backward, or forward.

These are important issues, especially when we are to consider linguistic performance.

On the basis of semantic contrast between Vv and Vs, we analyze verbal concatenation as follows. Note that we are not concerned with pre-verbal modal auxiliaries or directional verbs such as *pay* (go) and *maa* (come) that have no participant of their own.

In Thai, we may assume that the main verb, or the "core" of a predication, is the verb preceded by the negative marker. The main verb can also be testified in polar questions with *máy* (the interrogative final particle) since the core verb serves as an answer replying to the question. In the following examples, the underscored <u>Vv</u> or <u>Vs</u> denotes that the verb is the main verb. First, both Vv and Vs can be the main verb in a single verb construction. (6) is an example with Vv.

(6) *(mây)*   <u>*kin*</u>   *(khâaw):*   (NEG) <u>Vv</u> (N)
     (NEG)   eat   food
     'Will eat, or will not eat (food)'

In cases starting with Vv, you can continue to add Vv repeatedly.

(7) *(mây)*   <u>*pay*</u> *kin*   *(khâaw):*   (NEG) <u>Vv</u> Vv (Vv…)
     (NEG)   go   eat   (food)
     'Will go to eat, or will not go to eat (food)'

In the above (7), the first Vv remains as the main verb, as the negative marker indicates. In the following, however, if Vs follows Vv, either the first Vv or the last Vs may be negated. In (8) the first Vv preceded by the negative marker remains as the main verb. The negative marker negates the intention to "eat".

(8) *mây kin   khâaw   yə́ʔ:* NEG-<u>Vv</u> Vs
     NEG eat   food   much
     'not to eat much intentionally (for fear of gaining weight)'

In (9) below, the last Vv is negated, which shows that the main verb "moves" to be the last constituent, denoting that irrespective of the eater's intention, the result of "eating" is "not much". The "movement" of the core verb cannot be explained from the hierarchical perspective which defines the main verb in terms of the syntactic position of the verb; because the subject-verb-object (SVO) as basic word order assumes right branching, where

the main verb should be fixed as the first and topmost verb in the hierarchy. The hierarchical analysis is not appropriate for analyzing (9) since it needs additional "operations", for example, "right dislocation" of *mây* (NEG) to explain such constructions.

(9) *kin   khâaw   mây   yə́ʔ:*   Vv NEG-V̲s̲
     eat   food   NEG   much
     'eat but not much (because of poor appetite)'

The concatenative perspective enables us to define the "core" verb regardless of its position in a construction. Since each verb is treated equally in a concatenated construction, the "core", or the focus of predication can be assigned to the first or the last verb as if the phonetic stress can be assigned to the initial or the last syllable in a free accent language.

Similar analysis shows that the so-called causative construction is analyzed as an SVC, whose main verb is the first Vv [*hây* (give)] with a direction of affectedness →, that is, affecting the event.[56]

(10) *chán   hây   →   [dɛɛŋ   tii   lék]:*   V̲v̲ N Vv (Causative)
      1stPRON   CAUS   Daeng   hit   Lek
      'I made Daeng hit Lek.'

Similarly, the so-called passive construction is analyzed as an SVC, whose main verb is the first Vs [*thùuk* (get hit)] with a direction of affectedness ←, that is, affected by the event.

(11) *chán   thùuk ←   [dɛɛŋ   tii]:*   V̲s̲ N Vv (Passive)
      1stPRON   PASS   Daeng   hit
      'I was hit by Daeng.'

Notice first that (10) and (11) share the meaning that the subject's action is not kinetic, describing merely that the subject has affecting the object in (10), and that it has been affected by the action of the object in (11). Namely, the two sentences are structurally "antiparallel" in that their construction is the same SVC, but with opposite directions of affectedness. These causative and passive phrases are hence not individual constructions, but part of the SVC.

---

[56] In analyzing causative and passive constructions as concatenation, we assume that the main verb of each construction is the one that can be negated. For example, in the following causative case, the main verb is *sàŋ* (order) which can be negated.

   mɛ̂ɛ   (mây dây)   sàŋ   hây   lûuk   pay   rooŋ-rian
   mother(NEG PAST)order CAUS son   go   school
   'Mother (did not) order her son to go to school.'

The rest of SVC's, (*hây lûuk pay rooŋ-rian)* are conjoined, not embedded to the main verb. For now, I am not sure whether or not the main verb is always identified in terms of negation and polar question. Further work is necessary for justifying these assumptions.

The following (12) illustrates that the subject suffered from the other's action given in SVO construction with two participants. It also suggests that the subject in Thai passive construction does not always correspond to the object of the active construction.

(12) *chán   thùuk   [dɛɛŋ   khamooy   ŋən]*
     1st.PRON PASS   Daeng   steal   money
     'I had my money stolen by Daeng' = My money was stolen by Daeng.

Based on the above analysis, we claim that in Thai sentence structure, it is causative versus passive structure, rather than active versus passive structure that is the meaningful contrast.[57]

Compare the following SVC with antiparallel structures denoting perceptive meaning with each direction of affectedness, → and ←.

(13) *chán   faŋ   →← [dɛɛŋ   rɔ́ɔŋ phleeŋ]:*   <u>Vv</u> N Vv (Perceptive)
     1stPRON listen   Daeng   sing song
     'I listened to Daeng singing a song.'

(14) *chán   dâyyin   ←   [dɛɛŋ   rɔ́ɔŋ phleeŋ]:*   <u>Vs</u> N Vv (Perceptive)
     1stPRON hear   Daeng   sing song
     'I heard Daeng singing a song.'

(13) and (14) are structurally similar to (10) and (11), respectively, in terms of their constructions.

In sum, SVC is almost an almighty construction, entailing the causative, passive, perceptive, etc. Hence, we need not pose individual constructions in describing these expressions.

5. Conclusion

Starting with an examination of the transitivity hypothesis by Hopper and Thompson, and of the successive proposal by Tsunoda, we have attempted to examine Thai verbs by means of their occurrence in syntactic environments. Since Thai verbs that take direct objects can occur with or without objects, it is practically impossible to distinguish transitive and intransitive verbs without considering syntactic environments. It would hence be more appropriate to say that such verbs can be used in both transitive and intransitive constructions, rather than regarding the contrast to be part of the semantic properties attributed to verbs. By examining verbs in negated intention, imperative and passive constructions, we claim that the semantic Vv–Vs distinction as a polar contrast is the fundamental one in Thai verbs. Since for most verbs, the distinction between Vv and

---

[57] Traditional Japanese grammar observes the similarity of causative and passive constructions, which is similar to that in Thai given in (12). In Japanese, the causative sentence *"taroo-ga* (nominative) *jiroo-ni* (dative) *okane-o* (accusative) *nusuma-se-ta"* (steal-CAUS-PERFECT) meaning "Taro made Jiro steal the money" has the same morphosyntactic structure as the passive one *"taroo-ga* (nominative) *jiroo-ni* (dative) *okane-o* (accusative) *nusuma-re-ta"* (steal-PASSIVE-PERFECT) meaning "Taro's money was stolen by Jiro" except that the former has a causative suffix *"se"* while the latter has a passive one *"re,"* so they could be analyzed as antipallarel constructions with different directions of affectedness. Noda (1990) clearly illustrated the morphosyntactic "symmetry" of these constructions in Japanese.

Vs can be determined regardless of the presence or absence of objects, the distinction can be more important than that between transitive and intransitive.

Although the canonical construction in Thai is the subject-verb-object construction, the semantic roles of two participants vary according to the semantic properties of verbs. Prototypical Vv assumes human subjects attempting volitional actions that are controllable in general. On the other hand, Vs assumes either human or non-human subjects getting involved in natural phenomena, process, or accidental events that are beyond human control. Consequently, with Vv, the subject assumes the role of agent and the object, patient; the former affects the latter, whereas with Vs, the subject assumes the role of experiencer, and the object, the locus of stimulus that functions as a cause affecting the subject. The direction of affectedness is hence important in addition to the controllability. We claim the semantic distinction is fundamental in Thai because it is reflected in syntactic constructions. We have attempted to describe SVC as a non-hierarchical concatenation of verbal constituents, by showing constructions that reflect the Vv–Vs contrast. Based on the definition by Bisang (1991), we have analyzed causative and passive constructions as part of SVC. These two constructions are antiparallel to each other, sharing similar characteristics as SVC but with different directions of affectedness denoted by their main verbs. Furthermore, we have shown that the perceptive construction can be analyzed similarly to causative and passive constructions as part of SVC.

As a preliminary report of our semantic analysis of Thai verbs, only a limited number of verbs are examined in this paper. Also not many works have been done in analyzing SVC as concatenative construction. Further examinations of basic verbs including one-place verbs are necessary to properly analyzing syntactic constructions in Thai.

References

Aikhenvald, Alexandra Y. and R. M. W. Dixon (eds.) (2006) *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford: Oxford University Press.

Bisang, Walter (1991) "Verb serialization, grammaticalization and attractor positions in Chinese, Hmong, Vietnamese, Thai and Khmer", in Seiler H, Premper W, eds. *Partizipation: das sprachliche Erfassen von Sachverhalten,* pp. 509-62. Tübingen: Narr. ,

Bisang, Walter (1995) "Verb serialization and converbs: differences and similarities", in Haspelmath, Martin, and König E. (eds.) *Converbs in Cross-Linguistic Perspective,* pp. 137-88. Berlin: Mouton de Gruyter.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.

Hopper, Paul J., and Sandra Thompson (1980): "Transitivity in Grammar and Discourse." *Language* 56(2):251-99.

Imaizumi, Shinako (2001): "The Role of AFFECTED in Lexical Causative Alternations in Japanese." *Journal of Japanese Linguistics* [Indiana University] 17:1-28.

Institute for the Studies of Languages and Cultures of Asia and Africa (ed), (1966): *Linguistic Questionnaire for Asian and African Languages* (Part I).[in Japanese, Ajia-Afurika Gengo Chosahyo Jo] Tokyo: Institute for the Studies of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.

Iwasaki, Shoichi (2002) "Proprioceptive-state expressions in Thai", *Studies in Language* 26:1.33-66.

Iwasaki, Shoichi, and Preeya Ingkaphirom (2005): *A Reference Grammar of Thai*. New York: Cambridge University Press.

Jackendoff, Ray (1977): *X-bar Syntax: A Study of Phrase Structure.* Cambridge: The MIT Press.

Minegishi, Makoto (2004): "Southeast Asian Languages—A Case for Caseless?" In Peri Bhaskararao and K.V. Subbarao (eds), *Non-Nominative Subjects Vol. 1*, pp. 301-317. Amsterdam: John Benjamins.

Minegishi, Makoto, and Akagi Osamu (2009): "Development of Electronic Dictionary for Analyzing Linguistic Data," *Proceedings of Chula-Japan Linguistics Conference 2008*, Tokyo University of Foreign Studies, pp.11-16.

Noda, Hisashi (1990): "Nihongo-no judooka-to shiekika-no taishoosei [The symmetry between passivization and causativization in Japanese]." *Bungei Gengo Kenkyuu [Studies in Language and Literature: Language]* 19: 31-51.

Noss, Richard B. (1964): *Thai Reference Grammar.* Washington, D.C.: Foreign Service Institute.

Prasithrathsint, Amara (2000): "Adjectives as Verbs in Thai." *Linguistic Typology* 4: 251-271.

Sakamoto, Hinako (1985): [in Japanese] "Subcategorizations of Thai Verbs." *Journal of Asian and African Studies* 30:177-192.

Takahashi, Kiyoko (2008): "Post-Verbal Noun for a Part in Thai." *Papers from the 14th Annual Meeting of the Southeast Asian Linguistics Society 2004*, Vol. 2, pp. 63-72. Canberra: Pacific Linguistics.

Thepkanjana, Kingkarn (1986): *Serial Verb Constructions in Thai*. Ph.D. Dissertation, University of Michigan.

Tsunoda, Tasaku (1985): "Remarks on Transitivity." *Journal of Linguistics* 21: 385-396.

Uwano, Zendo (1999): "Classification of Japanese Accent Systems." In Kaji Shigeki (ed), *Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics,* pp.151–178, Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.

Uwano, Zendo (2007): "Two-pattern Accent Systems in Three Japanese Dialects." In T. Riad and C. Gussenhoven (eds), *Tones and Tunes: Typological Studies in Word and Sentence Prosody,* pp.147–165, Berlin: Mouton de Gruyter.

LANGUAGE IDEOLOGIES OF ETHNIC ORTHOGRAPHY IN A MULTILINGUAL STATE: THE CASE OF ETHNIC *THÁI* ORTHOGRAPHIES IN VIETNAM[58]

Yukti Mukdawijitra[59]
Thammasat University, Thailand
<iamyukti@yahoo.com>

Abstract

Contrary to the standard view on the imperative role of language homogenization to nation-state formation, this essay examines how multilingualism is vital to nation-state formation. Approached through language ideology framework, this ethnographic and historical research explores everlasting politics of orthography in the ethnic Thái case in Vietnam. Corresponding to local dialects, Thái orthographies represent pre-modern political formation of Thái sub-groups (Tai Dam, Tai Don and Tai Daeng). This diversity continues to colonial and post-colonial regimes. Consequently, while the state promotes national script to facilitate nation building, Thái sub-ethnic groups negotiate to maintain their orthographies in contemporary Vietnam.

Keywords: language ideology, multilingualism, orthography

Introduction

Based on documentary, archival and ethnographic research conducted in Vietnam in 2002-2005, this essay examines how the Vietnamese state and the Thái negotiate national and ethnic identity through the formation of policies and practices regarding multilingualism. Contrary to scholarship depicting ethnic minorities as internally homogeneous, this essay suggests how Vietnam's multilingual policies emerge from the complex historical and political interactions of the state, the ethnic minority communities, and the internal ethnic subgroups. As linguistic and orthographic characteristics of the Thái are diverse, this essay analyzes the ongoing politics of different Thái dialects and orthographies.[60] Such politics present the dilemmas of Vietnam's ethnic policies which, on the one hand, retain and

---

[58] I am grateful to anonymous reviewers of JSEALS for their advices to improve the draft. However the shortcoming of this paper is due to my limitation. If this article benefits anything, I devote it to Cầm Trọng (1934-2007), a great Vietnamese-Thái intellectual who devoted his life to preserve Thái scripts in Vietnam.

[59] Assistant professor of anthropology at Faculty of Sociology and Anthropology, Thammasat University, Thailand.

[60] In addition, I studied three major Tai Dam ancient texts and unpublished manuscripts such as *Kwaam To Muang, Song Chu Son Saaw*, and *Kwaam Paaw Khwan* under Cầm Trọng's supervision. Upon being relatively fluent in Tai Dam (Black Tai), I explore spoken and written dialects of the Tai Don (White Tai) and the Tai Daeng (Red Tai) by myself. Because the differences between these dialects are very obvious in both spoken sounds and writing system I can identify the differences of individual consonants, vowels, and meanings. However, I am not as fluent in those dialects as I am in Tai Dam.

promote ethnic diversity and ethnic traditional cultures, and yet, on the other hand, encourage the unified nationhood.

This essay is divided into six sections. First of all, I present my framework on language ideologies and how this framework benefits the understanding of the role of ethnic orthographies to the formation of the modern nation-state. Next, I introduce briefly who the Thái are and from where Thái orthographies came. The following three sections are a historical process of Tai language ideologies. They are periodized into the pre-French period, the period of the French rule and the early Vietnamese modern nation-state, and the contemporary period. Finally, I present a concluding remark on language ideologies of Thái orthographies in the Vietnamese context.

Language Ideologies: A Theoretical Background

Language is central to the formation of the modern nation-state. James Scott has argued that it is necessary for states to eliminate the complexities of the civil society in order to establish legibility, rationalization, standardization, and simplification (Scott 1998). If a state is founded out of a divers society, in Scott's view, such state will gradually reduce the diversity and thus homogenize the society. According to Scott, "of all state simplification, then, the imposition of a single official language may be the most powerful, and it is the precondition of many other simplifications" (1998: 72). Similarly, Ernest Gellner suggests that the process of homogenization of culture and language is integral to the formation of nation states (Gellner 1983). Using the case of French, Eugen Weber shows that the state had a longstanding interest in eradicating local languages and dialects, thus promoting the standardization of French, itself a Parisian French dialect (Weber 1979:67-94). Moreover, Benedict Anderson has argued that "print-capitalism," the mass production of printed materials, is a critical process for the formation of modern nation states (Anderson 1991). According to Anderson, a nation is a new kind of community imagined by people in a state, which emerges out of the sense of "we-ness" acquired and shared by people through the mass distribution of printed materials. In this regard, in places where a tradition of literacy has not been established, a standard national literacy is installed, thereby marginalizing the diversity of vernacular dialects.

However, language ideology approaches provide alternate theoretical grounds for the ethnographic examination of the roles of language in the nation-state formation. Anthropologists and linguists have long debated whether language objectively exists by itself or is embodied within power relations and human actions (Williams 1977; Hymes 1974; Voloshinov 1973 [1920s]). Recently, studies focusing on the ideology of language have begun to examine how the political economy, ranging from the face-to-face to the global level, intersects with ideas and uses of language (Gal 1989; Woolard and Schieffelin 1994; Schieffelin et al. 1998; Kroskrity 2000a). These processes can be referred collectively as "ideologies of language," summarized by Woolard and Schieffelin (1994:57-58) as "a set of beliefs about language articulated by users as a justification of perceived language structure and use. Additionally, as the term ideology suggests, this concept points to the relevance of social changes and power relations."

In response to Benedict Anderson's influential proposals on the relationship between language and nationhood, Susan Gal remarks that "it is clear that not only communities but also languages must be imagined before their unity can be socially accomplished" (1998:325). Language ideologies are thus the primary condition of the nation-state. In addition, Gal notes, "'language ideology' encourages analysis to encompass both social

interaction ... and state policy" (1998:319). In this sense, the language ideology approach covers the unit of analysis ranging from everyday linguistic ideas and practices to the state's linguistic policies.

Correspondingly, along with the recent treatment of ideology in the Gramscian notion of "hegemony" (e.g. Williams 1977; Eagelton 1991), rather than focusing on the dominant role of the state, the language ideology approach suggests that "different ideologies construct alternate, even opposing realities; they create differing views arising from and often constituting different social positions and subjectivities within a single social formation" (Gal 1998: 320). Accordingly, noted Gal, "It is the recognition that hegemony is never complete and that, in any social formation, ideologies--including linguistic ideologies--are multiple and at odds that renders the achievement of domination problematic, often fragile, and makes the semiotic aspects of its constant construction important to explore" (1998:323). Thereby, as Benedict Anderson has been criticized by Michael Silverstein, the language ideology approach points out the shortcoming of the approach to language and nationalism that "suppress[es] all the contestation and social history." To Silverstein, Anderson "takes its [the "we-ness"] meaning to be the straightforwardly and uniformly presupposed order of imaginable homogeneity-of-identity in the discursive-equal-discoursed-about spatiotemporal envelop of "the nation" (that is, the linguistic community informed by hegemonic standard) in which its speaker feel they reside" (Silverstein 2000:124). In this manner, apart from politicizing the ideas and practices of language, the language ideology approach views the politics of the ideas and practices of language as divergent, dynamic, processual, and complex.

As linguistic differentiation within and between social groups is political and processual, Susan Gal suggests characteristics of three semiotic processes for the investigation of sociolinguistic difference: (1) iconization "linguistic differences that index social contrasts are reinterpreted as icons of the social contrasts" (1998:328); (2) recursiveness "the projection of an opposition salient at one level of relationship onto some other level. Thus, the dichotomizing and participating process that was involved in some understood opposition (between groups or identities) recurs in distinctions made within the group, creating subcategories that mimic the original contrast" (Gal 1998:328); (3) erasure "occur[s] when an ideology simplifies a sociolinguistic field, forcing attention on only one part or dimension of it, thereby rendering some linguistic forms or groups invisible or recasting the image of their presence and practices to better fit the ideology" (Gal 1998:328). These political and processual characteristics of language are also found in the case of Thái orthographies in Vietnam.

The politics of orthography, which is essential to nation-state formation, is integral to language ideologies. Viewed through the language ideology approach, Woolard notes, "Orthographic systems cannot be conceptualized as simply reducing speech to writing but rather are symbols that themselves carry historical, cultural, and political meanings" (1998: 23). Not only is the orthographic choice involved intimately with the political process (Schieffelin and Doucet 1998), but orthographies in many cases are also the "weapon of the weak" (Scott 1985) used by social groups as their alternate or opposing means for transcription. Examples can be found in the new writing system of, for instance, the Apache in the United States (Basso and Anderson 1973), the orthography of Hmong and Khamu in Laos (Smalley et al. 1990), and the religious script in Nigeria (Probst 1997).

Although standardization of national language and orthography seems to be the most significant means of nation-state formation, many states have long since encountered the dilemmas involved in maintaining a national standard language while retaining minority languages or local dialects. They are, for example, the former Soviet Union (Lewis 1972), China (Zhou 2003), India (Zaman 1984), the United States and Canada (Handler 1988; Craith 1996; Ricento and Burnaby 1998), countries in Southeast Asia (Chee 1990; Esman 1990), and countries in Africa (Mansour 1993).

As the ideologies embodied in nation-state formation are complex and divisive, this essay sheds light on the complex interaction in the historical interplay between language, ethnic identity, and Vietnamese modern nation-state formation. The ethnographic and historical study offered by this essay will provide a case study of the way in which the diverse ethnic languages, literacies, and orthographies play role in the nation-state formation.

The Thái and Thái Orthographies

According to official records, in addition to the Kinh majority, Vietnam is comprised of 53 ethnic minorities, making up approximately 14% of Vietnam's overall population. Ranked as the second largest ethnic minority population (Asian Development Bank 2008:1), living in the northwest upland region, ethnic Thái (referred in Vietnamese) are inhabiting an area covering more than 30% of the landmass of northern Vietnam.[61] The Thái nationality can be divided into three main subgroups, *Tai Don* (White Tai or *Thái Trắng* in Vietnamese), *Tai Dam* (Black Tai or *Thái Đen* in Vietnamese), *Tai Daeng* (Red Tai or *Thái Đỏ* in Vietnamese); each subgroup has its own dialect and orthography. The populations of Tai Don, Tai Dam, and Tai Daeng (including Tai Do, Tai Hang Tong, Tai Thanh, and Tai Daeng listed in *Ethnologue* 2010) in 2002 numbered approximately 490,000, 764,000 and 190,000 respectively (*Ethnologue* 2010 [2002]).

Sharing the same Indic origin as Tai-Lao scripts in Southeast Asia, the Thái scripts were derived from "proto-Tai scripts" and arrived in Vietnam as early as the sixteenth century (Hartmann 1987). Unlike the Laotians or the Siamese, however, the Thái in Vietnam did not adopt Buddhism. Thái traditional literacy served the Thái elites to maintain class-based chiefdoms.[62] Even so, informants I interviewed recalled that villagers in the period before the modern state had access to the Thái traditional literacy to a certain degree. Although very few (mostly male) commoners were literate in the traditional Thái communities, literate villagers such as ritual experts, local poets, and singers disseminated the practices of the scripts in both religious and secular contexts. In this way, although Buddhism was not established in Thái traditional society, the script the Thái adopted from their Buddhist neighbor became an important means of communication for both elites and commoners.

---

[61] The Tai language used in northwestern Vietnam belongs to a branch of the "Southwestern Tai language" whose close language-kin include Lao, Siamese, and Lue (Li 1977). In Vietnam, another major related linguistic branch of the Tai language family is the "Central Tai language" branch spoken by, for example, the Tày and the Nùng who settle in northeastern Vietnam. Throughout this article, I use the term Thái to refer to the Tai Dam, Tai Don and Tai Daeng in northwestern Vietnam who are recognized collectively by the Vietnamese as the Thái (*người Thái*). When referring to the Thái subgroups, I use local ethnonyms, which are Tai Dam, Tai Don and Tai Daeng.

[62] On Thái chiefdom, see Condominas (1990) and Cầm Trọng (1978: 240-265).

Figure I: *Map of Tai Dam, Tai Don, Tai Daeng Settlements*

Thái Orthographies in Pre-French Period
Pre-French Thái chiefdom politics formed by the patrilineal marriage system among the Thái elites. This political formation helped in maintaining not only political community but also linguistic and orthographic community. I argue that the language and the ethnic group considered "Thái" is a result of the political formation that created "linguistic boundaries," meaning geographical areas where a set of linguistic characteristics had been practiced. These linguistic boundaries had corresponded with Thái ethnic and subethnic identification.

Seen from the Thái tradition of literacy, each different Thái orthography has long been preserved to represent a form of Tai dialect used by a subgroup of the Thái, and each community of orthography was ruled by a Thái ruling clan. For example, the Tai Dam's *Kwaam To Muang* (Tai principality chronicle) was recorded in Tai Dam orthography and language and inherited through the Tai Dam patrilineal descended *Lo Kam* clan. The text was read to the public at the ruling class funeral. In this manner, I argue, the Tai Dam elites established an "imagined community," an imagination of "we-ness" in which a group of people living in a large area shared as they were belonging to the same group of people.

Such Tai Dam we-ness originated from the use of Tai Dam dialect and orthography. In this area the Tai Dam maintained patrilineal marriage relations within the *Lo Kam* clan.[63]

Not merely did the Thái subgroups speak different dialects, but they also used various systems of writing system to convey their distinctive spoken dialects. The spoken and written dialects were located within particular regions.[64] Similar to the spoken dialects, various types of Thái orthographies can be classified roughly by the different dialects of speaking. The correlation of writing systems and spoken dialects suggests a conclusion that the differences in Thái writing systems correlated with the Thái ethnic boundaries dividing them into several subgroups. This correlation leads me to argue that the different Thái groups probably identified themselves through the dialects spoken and the orthographies written, among other cultural characteristics. Linguistic forms and writing systems were perhaps used as a partial, if not the whole, index of ethnic identity. Each of these writing systems was a constructive part of a "community of literacy," where the population of each community of literacy speak, write, read, recite, and perform using a similar kind of orthography, practical writing, and literature.

Based on my collections of Thái orthographies, I propose that the Thái in Vietnam had at least seven orthographies, which correlated to the sub-groups and spoken dialects of the Thái in five regions, as presented in Table I.[65] Prior to the French period, the northern part of northwestern Vietnam was the location where the Tai Don orthography was used as the linguistic ideological index of the Tai settling in Lai Châu and Phong Thổ, the two major Tai Don principalities. The Tai Dam orthography was identified with the Tai chiefdoms located southward, covering a relatively large area ranging from the westernmost Điên Biên Phủ, toward central Sơn La, and Nghĩa Lộ in the easternmost of the Tai chiefdoms. On the southern region, the Tai Daeng orthography was related to the

---

[63] A similar pattern of linguistic, orthographic, literary, ethnic and political boundary was also practiced by other Thái subgroups. Please find details in Yukti (2007: 97-140).

[64] Gedney's accounts on the Thái phonology (Gedney 1989) can be used as a starting point to understand the heterogeneity of Thái orthographies found in the pre-French period. The location of the dialects can still be found in present day. When traveling to the country of the Thái in northwestern Vietnam, as a speaker of Tai Dam, I found that the dialect spoken in central Sơn La, southern Điên Biên Phủ, western Yên Bái, southern Phong Thổ, and southern Than Uyên were the most understandable to me. By contrast, I hardly understood the Tai Don dialects spoken in Lai Châu, central Phong Thổ, western Sơn La and the Tai Daeng dialects spoken in eastern Sơn La, western Hòa Bình, western Thanh Hóa, and western Nghệ An. When native Tháis of different dialects met, they also had difficulty in understanding each other. Although, for example, the Tai Dam dialect is not completely unintelligible to a Tai Don speaker, in a long conversation on normal daily life issues the Tai Don speaker would have difficulty in understanding the Tai Dam dialect. This difficulty is also found among the Tai Dam and the Tai Daeng speakers when they try to communicate across Tai dialects. The Thái dialects are thus almost mutually unintelligible.

However, Gedney's accounts on Thái phonology probably need to be reinvestigated. My collection of Thái orthographies shows that the Thái orthographies are more diverse than what Gedney suggests regarding Thái phonology. According to Cầm Trọng, the Vietnamese-Tai Dam scholar, there are eight different types of Thái orthographies in Vietnam (Cầm Trọng 2002:809-810). Louis Finot lists five types of Vietnamese Thái scripts (Finot 1917). Disagreeing with both previous accounts, Michel Ferlus suggests that Thái scripts in Vietnam are comprised of four types (Ferlus 2546 BE: 276). Still, the variation corresponds with the divisions of Thái subgroups I found, i.e. the divisions of three major Thái scripts, Tai Dam, Tai Don and Tai Daeng. Nevertheless, this essay cannot provide a more accurate evidence and systematic research on variation of Thái phonology. A more elaborated study on Thái dialects may help explaining the great variation of Thái orthography.

[65] Please find details of the comparison of Thái letters in Appendix located at the end of this essay.

Tai residing in several Tai principalities in Thanh Hóa Province, and it was also used in the western part of Hòa Bình Province in Mai Châu District. Quỳ Châu and the Pao River (Sông Cà) were the southern and southernmost locations where the Quỳ Châu orthography and the Lai Pao orthography were found.

Different Thái sub-ethnic groups maintained distinctive ways of writing. The differences were represented in letter shapes, correspondences between phonemes (of both consonants and vowels) and script, special symbols, direction of writing (horizontally or vertically), and choice of phonemes (voiced or voiceless initial consonants). As a result of these differences, in pre-French Thái society, a native writer of Tai Dam, for instance, would find difficulties in reading Tai Don and Tai Daeng texts in the same way as those native of other Thái communities of literacy would.[66]

Table I: *Locations of Thái Orthographies*

Thái Orthographies	Thái Regions	Thái Ruling Clans
I. Tai Don (Central)	Lai Châu and northwestern Sơn La	Đeo (or Đèo in Vietnamese)
II Tai Don (Northeastern)	Phong Thổ	
III Tai Dam	Điên Biên Phủ, central Sơn La, Nghĩa Lộ	Lò Cầm (or Bạc Cầm, Cầm, and Cầm Ngọc in Vietnamese)
IV Tai Daeng (Central)	Thanh Hóa	Sa (or Hà, Hà Công in Vietnamese) and Hoàng
V Tai Daeng (Northern)	Mai Châu (Hòa Bình) (and eastern Sơn La?)	
VI Quỳ Châu (Southern Tai Daeng?)	Southern Thanh Hóa and Qùy Châu, northern Nghệ An	no information
VII Lai Pao (Southern Tai Daeng?)	Southern Nghệ An	no information

---

[66] I already provided the detail of the differences between the pre-modern Thái orthographies in Yukti (2007: 141-182).

Figure II: *Map of Tai Dam, Tai Don, Tai Daeng orthographies*

The language ideologies of the Thái in the pre-French period was thus a process of the iconization and recursiveness which demarcated sub-ethnic identities. Through the iconization, we can assume that the linguistic and orthographic features of different Thái languages were practiced over generations and were treated by the different Thái groups as if the index of each Thái group. For instance, Tai Don dialects used their vowel system which has fewer diphthongs than Tai Dam dialects; therefore, Tai Don dialects created a system of writing to represent their vowels. In terms of orthographies, for example, by writing a hook under an arch to represent phoneme /high k/, the Tai Daeng differentiated their orthography and ethnic identity from the Tai Dam whose letter /high k/ has a line crossing an arch. By practicing these different features the different Thái groups recognized their linguistic differences and created ethnic boundaries between each group. In other words, the linguistic and orthographic practices became an icon recognized by different Thái groups.

The process of recursiveness divided a Thái group into smaller subgroups by projecting different features of language and orthography applied for dividing the main subgroups into those features of languages and orthographies in the smaller subgroups. An example of the recursive process is the division of the Tai Don into the Tai Don of Lai and the Tai Don of Phong Thổ. According to my ethnographic and historical research, the Tai of Lai Châu and the Tai of Phong Thổ identify themselves as "Tai Khao" or "Tai Don." In term of political formation, the elites of the two areas claim that they descended from the same clan of Đèo or Điêu. In terms of language, the distinctive features of consonants and tones used to iconize the Tai Dam, the Tai Don, and the Tai Daeng were also applied to the

distinction between the Tai Don of Lai and the Tai Don of Phong Thổ. I assume that the similar process also occurred to the Tai Daeng regions in which several Tai Daeng dialects and orthographies were found. Nonetheless, it is remarkable that the Tai Dam did not develop their internal recursive linguistic ideology. The Tai Dam community of language and orthography was thus relatively more homogenous than the other Thái communities. It is likely that Tai Dam political system and literary tradition were more solid than those of other Tháis; and thus the Tai Dam were able to maintain their homogenous language and orthography.

As the division of Thái subgroups corresponded with the division of Thái chiefdoms formed through ruling patrilineal, the linguistic iconization and recursiveness of Thái languages, orthographies, and identities related intimately with the socio-political formation of the Thái division of sub-ethnic groups. To form the Thái polity, the Thái imagined and created not only their political, social, and ethnic boundaries but also their linguistic and orthographic boundaries. As northwestern Vietnam has been inhabited by diverse ethnic groups, the northwest residents, including the Thái, were multilingual. Due to the fact that the Thái were the hegemonic power of the northwest, however, Thái dialects became the dominant means of communication in this region. Moreover, the demarcation of sub-ethnic Thái boundaries into Tai Don, Tai Dam, and Tai Daeng, corresponding to the politics of the Thái clans and chiefdoms, determined the sub-division of Thái linguistic and orthographic communities.

Thái Orthographies under the French Rule and the Vietnamese Nation-State
The Thái did not isolate themselves from other ethnic groups, both in the region and in the globe, and thus the Thái community of language was never monolingual. Consequently, the choice of orthography within Thái society has always been complicated and politically contested. The Thái used not just in their own various kinds of Thái orthographies but also in the regional and globally influential states' orthographies. Lao script, Chinese characters, Vietnamese ancient *nôm* and modern *quốc ngữ* and French are thus not unfamiliar to the Thái. When they needed to communicate with surrounding more influential states, pre-French colonial Thái had to use those more powerful states' languages and orthographies. In addition to their Thái orthographies and dialects, evidences taken from many texts of Thái literature show that some Thái elites had skills in Vietnamese, Chinese, Lao, and French.

Language ideologies involved in Thái orthographies were even more complicated under the periods of French colonization and the Vietnamese nation-state formation. In the French period, the Thái had to learn Vietnamese and French. Two kinds of Roman-based orthographies--*quốc ngữ* and French--were introduced to the Thái. In addition, the French created a Roman-based orthography to write Thái. By doing so, the colonial regime iconized the diverse Tháis under one standard Thái language and orthography. However, French-based Thái orthographies did not succeed.

After the fall of French colonial power in Vietnam, the policies promoting mass education for ethnic minorities were crucial in the process of Vietnamese state formation. Three autonomous zones were founded, one of which was the Northwest Autonomous Zone, founded in 1955, on the first anniversary of the Điện Biên Phủ victory. Mass education in the ethnic minority areas was implemented after the northern regime was founded in 1954. Democratic Republic of Vietnam urged the ethnic peoples, particularly those who resided in the autonomous zones, to learn Vietnamese in *quốc ngữ*, while

supporting them to learn their own ancient scripts or newly invented *quốc ngữ*-based ethnic scripts.[67]   Since then, influenced by the Soviet Union and the People's Republic of China's multilingual policies intended to promote mass literacy of ethnic minorities together with teaching the national language (Lewis 1972; Zhou 2003), the Vietnamese state invented quốc ngữ-based orthographies for the ethnic minorities. The state allowed several ethnic groups, particularly the Thái, the Tày, the Chinese and the Cham, the use of their own scripts, which were not quốc ngữ-based scripts, at school (Thanh Ha 1968; Trần Trí Dõi 1999).[68]

During the 1950s-1960s, in the Northwest Autonomous Zone, Tai Dam script was chosen by the Vietnamese government to be promoted over other Thái scripts in Vietnam. Under this new regime of language, the Vietnamese state's ideology of language characterized the Tai Dam script as it is "used by the largest number of people, . . . simple, beautiful, currently used and representing ethnic identity" (TTLTQG-3 1954: 6), "more advanced," "accurate" and "wealthy in literature of every kind" (TTLTQG-3 1956: 38-41) than other Thái orthographies. However, I argue that it is more likely because the Tai Dam actively participated in the liberation of the northwest in the early 1940s-1950s the Tai Dam script was thus the proper choice.[69]   As pointed out by Kroskrity, "The imposition of a state-supported hegemonic standard will always benefit some social groups over others" (2000b:8); the choice of script implemented in the Thái community reflects the close relations between the Party and a division of the Tai Dam.

Later, however, the Vietnamese state attempted to create a quốc ngữ-based orthography, claiming that the quốc ngữ-based script is "quicker to learn and more accurate," "will bridge Thái language and the national script much more easily," "no longer makes Thái officials and civilians have the sense of division between different regions," "will facilitate the officials and civilians of other ethnic groups in the zone to learn Thái script much faster" (TTLTGQ-3 1971:34). Thái languages and orthographies have not been taught in schools since 1975 when the autonomous zones were dissolved. After the unification of northern and southern Vietnam, the Tai Dam orthography and other Thái orthographies were thus subordinated to Vietnamese language and orthography. To paraphrase Susan Gal's approach to language ideologies (1998:327-9), while the Vietnamese nation-state makes Tai Dam orthography become "recursive" in opposition to other Thái orthographies and the Vietnamese national orthography, the other Thái orthographies were almost completely "erased."

After the country was unified in 1975, discussions among ethnic Thái regarding the possibility of bringing the script back to schools and which script is the proper version are

---

[67] Aside from ethnic Thái, many ethnic groups, e.g. Tày, Chinese, Cham, and Khmer, also have their own scripts. While the Tày script is based on Chinese, the Cham script and Khmer script in southern Vietnam are derived from Arabic and Indic scripts, respectively. As European missionaries were historically working in the highland areas of Vietnam, other ethnic groups, such as the Hmong, adopted different versions of roman-derived scripts (Smalley et al. 1990).

[68] Drawn mainly from the National Archive III, I provide a detailed discussion of Vietnam's language and orthography policies implemented in the Northwest Autonomous Zone in 1955-1975 in Yukti (2007:260-313).

[69] Please find details of Thái politics and how the Tai Dam became the leading group in the Vietnamese revolution in Yukti (2007: 214-239).

still divisive.[70]   In summary, similar to many cases occurring cross-culturally worldwide, language ideologies in the case of the Thái in Vietnam are sites of "not only multiplicity and contestation but also clashes or disjunctures in which divergent ideological perspectives on language and discourse are juxtaposed, resulting in conflict, confusion, and contradiction" (Kroskrity 2000b:13).

Contemporary Debates on Thái Orthographies
The Thái in the present day still debate the proper way to preserve their orthographies. The debate is whether or not the Thái should standardize Thái orthographies to empower Tai ethnic identity. Many Thái scholars disagree and urge the Thái to preserve only each subgroup's local orthography. "Workshop on the Preservation and Digitalization of Tai Scripts" conference held on November 15th-16th, 2005 at Hanoi provides a clear picture of such debate.[71]   Sponsored by UNESCO, the Programme for Thái Studies of Hanoi National University and the Institute of Information Technology, the Academy of Science and Technology, Thái delegates were invited from major Thái provinces including Điện Biên Phủ, Sơn La, Lai Châu, and Hòa Bình. International scholars from Japan, Thailand, and the United States were also invited. The main objectives were to call attention to roles, feasibilities, and benefits of the digitization of Thái scripts. In order to support the use of the digitized Thái script, UNESCO required that there must be at least one million Thái who will use the script. Another main purpose of the meeting was thus to urge local Tháis to reach an agreement on the standardization of Thái orthography.

After two days of academic presentations by Vietnamese, Thái, and international academics, as well as users of the Thái scripts, delegates from the Thái regions presented their reflections on the meeting. A prominent female delegate from Lai Châu noted that "Aside from worrying about what benefit the future of Thái script would have for Vietnam's modern life, born as half Tai Dam, half Tai Don, I don't know whether a standard Thái script would better be based on Tai Dam or Tai Don." A group of district officials representing Hòa Bình Province's Tai Daeng (nowadays referring to themselves as Tai Don) which successfully carried out the teaching of a standard a Thái script (which had been recently created by a group of Thái scholars working in Hanoi) reflected that "After adjusting the script to write our dialect, it is convenient for us to learn the standard Thái script." A young female teacher, wearing her "traditional" Tai Dam attire and hair bun representing her marriage status, insisted that "Intellectuals of Sơn La agree only on reviving the ancient Tai Dam script." Although she herself is a daughter of Lò Văn Mươi, an active leader of the movement to standardize Thái orthography during the 1950s-1960s, she did not agree with the use of a standard Thái. Instead, she had brought a computerized version of Sơn La's Tai Dam orthography, which she had helped create to, present at the meeting.

The meeting reflected main debates regarding Thái orthographies in contemporary Vietnam very well. First of all, it reflected the common interest shared among Tháis in many regions that Thái literacy is tied intimately with Thái orthography and it plays a key

---

[70] How the Thái orthographies and Thái traditional literacies are still significant to Thái villagers in present day is extensively discussed in my ethnographic study of Thái religious and secular textual performances (Yukti 2007:314-400).
[71] Some information on the conference and current situation on Thái scripts are presented at http://www.huesoft.com.vn/chuthaivietnam/TINTUC/.

role in the preservation of Thái ethnic identity and culture. While struggling to catch up with the national Vietnamese language and script, the Thái hope they could simultaneously preserve their ethnic identity and culture by preserving Thái orthographies and literacies. In this regard, the orthography that each Thái subgroup maintains "iconicizes" each subgroup's "imagined community" defining each Thái ethnic boundaries of spoken dialect and orthography against one another. The boundaries were drawn as if a Thái group was monolingual, although it is evident that multilingualism were practiced throughout Thái region. Within each Thái group, the linguistic and orthographic features iconized as the ethnic identity of a Thái group also recurred to iconize the sub-divisions of most of the Thái subgroups, except the Tai Dam.

Second, as different regions of the Thái have their own version of Tai orthography, they still disagree regarding the standardization of the Thái scripts. The maintenance of local diversity and the local hesitation to accept the standard Thái orthography echoes the agitation voiced by local intellectuals and authorities against Thái script reformation policies several decades ago in the Northwest Autonomous Zone (1955-1975). Each Tai proto-imagined community of language and orthography maintained its pre-colonial politico-linguistic boundaries even in the later eras. In the colonial period, different factions of the Thái established diverse relationships with Chinese bandits, the French, and Vietnamese anti-colonial organizations, including the Vietnamese Communist Party. The diverse Thái spoken and written dialects played different key roles in each Thái faction. After the 1950s, the intimate affiliation between a branch of the Tai Dam and the Vietnamese state led the Vietnamese state to grant its favor to Tai Dam language and orthography, while languages and orthographies of other Tháis were undermined.

Consequently, the attempts to homogenize Thái languages and orthographies led to counter-agitation by the diverse Tháis whose dialects and orthographies were under threat of erasure. The dilemma whether it would be better to accept a common Thái identity and the standard Thái orthography or to insist on developing each region's local orthography remains among the Thái community nowadays. This dilemma, however, prevents the Thái from creating a pan-Thái identity encompassing Thái subgroups and thus benefits the Vietnamese state enabling it to implement a "divide and rule" administrative strategy toward the different Thái subgroups.

Third, by granting the ethnic minorities the right to preserve and to use their languages and orthographies, the Vietnamese state has sought to implement multilingual policies. However, as demonstrated in the essay, Thái languages and orthographies have gone through an ever-shifting processes of being iconized, subordinated, and erased. Between the 1950s-1960s, the Vietnamese state iconized Tai Dam language and orthography as the language and script of the Thái, even though the languages and orthographies of the Thái in different regions are diverse. In the 1970s, Thái languages and orthographies, parallel to other ethnic languages and scripts, were recursively projected and subordinated to Vietnamese and quốc ngữ, the national language and orthography. From the 1980s to the present day, the multilingual policies were revived; thereby, nowadays Thái orthographies and traditional literacies are commonly practiced among Thái villagers, male and female alike in villages. Recently, as the Vietnamese government also acknowledges the use of ethnic minorities scripts in schools, Thái scripts are taught in schools located in Thái regions. Nonetheless, although the Vietnamese state preserves Thái languages and orthographies, the preservation is selective and thus, in the same way as

Susan Gal puts it, "rendering some linguistic forms or groups invisible or recasting the image of their presence and practices to better fit the ideology" (1998: 328). As the Tai Dam language and orthography are still promoted for preservation, other Thái dialects and orthographies are susceptible to erasure.

Nonetheless, Vietnam remains a multilingual state. Since the early years of the resistance war against the French and the Japanese in the 1940s, Thái languages and orthographies, particularly the Tai Dam language and orthography, had largely benefited from the formation of the anti-colonial movement. Consequently, Thái languages and scripts were significant to the transitional period of the integration of the Thái into the nation-state, particularly during the 1950s and 1960s. The state continues to consider the shortcomings of using the national language particularly in the educational sector in ethnic minority areas where ethnic languages still play key roles. Nowadays, a period when the Vietnamese nation-state is relatively stable, the multilingual policies are still being maintained or even expanded. On the one hand, the Vietnamese state's multilingual policies are a means for national integration and cultural assimilation. On the other hand, a side effect of its policies is that the ethnic peoples can to a certain degree preserve their culture and identity. Thái villagers thus continue to practice both traditional and emerging genres of Tai textual performances.

Last but not least, within the complex relationship between the Thái ideologies of language and the formation of Vietnamese nationhood, the ethnic peoples participate actively in policy making and implementation, both at the national and local level. At the local level, as suggested by the Tai Dam delegates from Sơn La and the Tai Daeng delegates from Hòa Bình attending UNESCO meeting noted above, Thái intellectuals and authorities initiated the Thái script classes and the computerization of the Thái script. At the national level, led by the Programme for Thái Studies, Thái scholars collaborate with the central governmental organizations to preserve and develop the use of Thái orthography.

The digitization of Thái script meeting ended without any agreement whether or not a standard Thái script already used in some region should be adopted by the entire Thái community. In 2006, however, a computerized Tai Dam script created in Sơn La was recognized by Unicode (Lò Luận 2010 [2009]). Still, the meeting did make certain gains. It marked the very first time since the termination of the autonomous zone that the Thái gathered to discuss the future of Thái orthographies. It assured the Thái that the Vietnamese state remains open to acknowledging and encouraging the preservation and the use of Thái orthography. The UNESCO meeting sheds light on the ever-shifting language ideologies of the Thái orthographies in the context of the rise of the Vietnamese nation-state. Not only did the meeting demonstrate that the diversity of the Thái still plays important role in their internal ethnic politics, but also it reveals that ethnic diversity continues to play a role in Vietnamese state policy decisions.

Conclusion

Reflected in a nuanced history of the language ideologies in the case of the Thái in Vietnam, the Vietnamese nation-state emerges from contestations of ethnic minorities, colonizers, and the ethnic majority over the ideas and practices of language and orthography. The ever-shifting linguistic ideologies regarding ethnic Thái in Vietnam have undergone processes of the linguistic change in various regimes of language. The case of Thái orthographies demonstrates that Vietnam has long maintained multiple literacy

policies where traditional literacy and standard national literacy co-exists in the ethnic minority communities. The language ideology approach provides a theoretical perspective for a linguistic anthropological examination to unravel the convergence of ideas and practices of the ethnic peoples, the sub-ethnic groups, and the state revealing how ethnic languages and orthographies take part in nation-state formation.

References

Anderson, Benedict R. O'G. 1991. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Asian Development Bank. 2008. *Reducing Poverty of Ethnic Minorities through Natural Resource Management.* Manila, Philippines: Asian Development Bank.

Basso, Keith H., and Anderson, Ned. 1973. A Western Apache Writing System: The Symbols of Silas John. *Science* 180:1013-1022.

Cầm Trọng. 1978. *Người Thái ở Tây Bắc Việt Nam*. [Ethnic Thái in Northwestern Vietnam]. Hà Nội: Khao học xã hội.

Cầm Trọng. 2002. Chữ Thái ở Việt Nam. [Thái script in Vietnam.] In *Văn hóa và lịnh sử các dân tộc trong nhóm ngôn ngữ Thái Việt Nam*. [Culture and history of Thái speaking ethnic groups in Vietnam] Chương trình Thái học Việt Nam, chủ biên. Pp. 809-826. Hà Nội: Văn hóa thông tin.

Chee, Tham Seong. 1990. *Multi-Lingualism in Singapore: Two Decades of Development*. Singapore: Census of Population.

Condominas, Georges. 1990. *From Lawa to Mon, from Saa' to Thai: Historical and Anthropological Aspects of Southeast Asian Social Spaces*. S. Anderson, M. Magannon, and G. Wijeyewardene, transl. Canberra: Departnemt of Anthropology, Australian National University.

Craith, Maire Nic (ed.). 1996. *Watching One's Tongue: Issues in Language Planning*. Liverpool: Liverpool University Press.

Đồng Trọng Im (sưu tầm). 2004. *Chữ cái Thái Trắng Mường So*. [White Thái script of Muong So.] Manuscript.

Eagleton, Terry. 1991. *Ideology: An Introduction*. London: Verso, 1991.

Esman, Milton J. 1990. Language Policy and Political Community in Southeast Asia. In *Language Policy and Political Development*. Brian Weinstein (ed.) Pp. 325-340. Norwood: Ablex Publication.

Ethnologue: Languages of the World. 2010 [2002]. http://www.ethnologue.com/show_country.asp?name=VN

Ferlus, Michel et Trần Trí Dõi. 1997. *Présentation de l'Écriture Lai Pao dé Thai de Tương Dương (Nghệ An)*. Văn Hóa Nghệ Thuật (157):90-92.

Ferlus, Michel. 1988. Langues et Écritures en Asie du Sud-Est: les Écritures d'origine Indienne et Leur Adaptation aux Langues de l' Asie du Sud-Est Continentale: les Consonnes. 21st International Conference on Sino-Tibetan Languages and Linguistics, Lund University, Sweden, 1988.

Ferlus, Michel. 1993 Phonétique et Écriture du Tai de Qui Châu (Vietnam). C.L.A.O. 17(1):87-106.

Ferlus, Michel. 2546BE (2003) Akson thai: ton kamnoet chak india lae akson khom. [Thai script: its India and Khmer origin.] In *Yot Sede kap thai sueksa*. [George Cedes and Thai Studies]. Pp. 269-311. Krungthep: Sun Manusayawithya Sirindhron. (In Thai.)

Finot, Louis. 1917. Recherches sur la Littérature Laotienne. *Bulletin de l'École Française d'Extrême-Orient* 17(5):1-174.

Gal, Susan. 1998. Multiplicity and Contention among Language Ideologies: A Commentary. In *Language Ideologies: Practicea and Theory*. B.B. Schieffelin, K.A. Woolard, and P.V. Kroskrity, eds. Pp. 317-331. New York: Oxford University Press.

Gedney, Willam J.1989. A Comparative Sketch of White, Black and Red Tai. In Robert J. Bickner, John Hartmann, Thomas John Hudak and Patcharin Peyasantiwong (eds.) *Selected Papers on Comparative Tai Studies*. Pp. 415-461. Michigan: Center for Southeast Asian Studies, Michigan University.

Gellner, Ernest. 1983. *Nations and Nationalism*. Oxford: Blackwell.

Hà Công Tín (sưu tầm và biên soạn). 1998. Sách tự học tiếng Thái cổ truyền Mai Châu. [A handbook for learn-it-yourself ancient Thái script of Mai Châu.] Manuscript.

Handler, Richard. 1988. *Nationalism and the Politics of Culture in Quebec*. Madison: University of Wisconsin Press.

Hartmann, John F. 1987. *Dating White Thai and Black Thai Scripts*. Ruam botkhwam prawattisat Thai. 5:121-139. (In Thai).

Hymes, Dell. 1974. Way of Speaking. In Richard Bauman and Joel Sherzer (eds.) *Explorations in the Ethnography of Speaking*. Pp. 433-451. Cambridge: Cambridge University Press.

Kroskrity, Paul V. (ed.) 2000a. *Regimes of Language: Ideologies, Polities, and Identities*. Santa Fe: School of American Research Press.

Kroskrity, Paul V. 2000b. Regimention Languages: Language Ideological Perspectives. In Paul V. Kroskrity (ed.) *Regimes of Language: Ideologies, Polities, and Identities*. Pp. 1-34. Santa Fe: School of American Research Press.

Lefèvre-Pontalis, Pierre. 1892. Étude sur quelques Alphabetes et Vocabulaires *Thais*. *T'uong pao*, vol. III.

Lewis, E. Glyn. 1972. *Multilingualism in the Soviet Union: Aspects of Language Policy and its Implication*. Mouton: The Hague.

Li, Fang-kuei. 1977. *The Handbook of Comparative Tai*. Honolulu: University Press of Hawaii.

Lò Luận. 2010 [2009] http://www.huesoft.com.vn/chuthaivietnam/TINTUC/?f=TINTUC& MaBaiViet=22&q=

Mansour, Gerda. 1993. *Multilingualism and Nation Building*. Clevedon: Multilingual Matters Ltd.

Martini, François. 1954. Romanisation des Parlers 'Tay Du Nord Viêtnam. *Bulletin de l'École Française d'Extrême-Orient* 46(2):555-572.

Probst, P. 1997. The Letter and the Spirits: Literacy and Religious Authority in the History of the Aladura Movement in Western Nigeria. In Brian Street (ed.) *Cross-Cultural Approaches to Literacy*. Pp. 198-219. Cambridge: Cambridge University Press.

Ricento, Thomas, and Barbara Burnaby, eds. 1998. *Language and Politics in the United States and Canada: Myths and Realities*. Mahwah: Lawrence Erlbaum Association.

Schieffelin, Bambi B., and Rachelle C. Doucet. 1998. The "Real" Haitian Creole: Ideology, Metalingualistics, and Orthographic Choice. In *Language Ideologies: Practice and Theory*. B.B. Schieffelin, K.A. Woolard, and P.V. Kroskrity, eds. Pp. 285-316. New York: Oxford University Press.

Schieffelin, Bambi B., Kathryn A. Woolard, and Paul V. Kroskrity, eds. 1998. *Language Ideologies: Practice and Theory*. New York: Oxford University Press.

Scott, James C. 1985. *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven: Yale University Press.

Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Conditions have Failed*. New Haven: Yale University Press.

Silverstein, Michael. 2000. Whorfianism and the Linguistic Imagination of Nationality. In Paul V. Kroskrity (ed.) *Regimes of Language: Ideologies, Politics, and Identities*. Pp. 85-138. Santa Fe: School of American Research Press.

Silvestre, J. 1918. Les Thai Blancs de Phong-Tho. *Bulletin De L'Ecole Française D'Extrême-Orient* 18(4):1-56.

Smalley, William, Chia Koua Vang, and Gnia Yee Yang. 1990. *Mother of Writing: The Origin and Development of a Hmong Messianic Script*. Chicago: University of Chicago Press.

Thanh Ha. 1968. The Language of National Minorities and the Creation Or Improvement of their Scripts. Vietnamese Studies 15:121-136.

Trần Trí Dõi. 1999 *Nghiên cứu ngôn ngữ các dân tộc thiểu số Việt Nam*. [Researches on ethnic minority scripts in Vietnam.] Hà Nội: Đại học Quốc gia Hà Nội.

TTLTQG-3 (Trung tâm lưu trữ Quốc gia III). 1954 Hồ sơ số 7053 "Biên bản Hội nghị nghiên cứu chữ Thái ngày 26/7/1954 của Khu Giáo dục Tây Bắc." [Folder number 7053, Report on the official meeting on Thái script research, July 26th, 1954, Department of Educational Affairs of the Tây Bắc Autonomous Zone.]

TTLTQG-3 (Trung tâm lưu trữ Quốc gia III). 1956 Hồ sơ số 7042 "Kế hoạch, chỉ thị, biên bản về việc hội chữ Thái và mẫu chữ Thái năm 1956 của UBHC Khu T.T. Thái Mèo và Sở giáo dục." [Folder number 7042, Project, instruction and report on the standard Thái script in 1956, Administration Committee of Thái-Mèo Autonomous Zone and the Department of Education.]

TTLTQG-3 (Trung tâm lưu trữ Quốc gia III). 1971 Hồ sơ số 7188 "Báo cáo tổng kết công tác giảng dạy chữ Thái trong 15 năm qua của Sở giáo dục năm 1971." [Folder number 7188, Summary report on the instruction of Thái script in the past 15 years, Department of Education, 1971.]

Voloshinov, V.N. 1993. *Marxism and the Philosophy of Language*. Translated. Cambridge: Harvard University Press.

Weber, Eugen. 1979. *Peasants into Frenchmen: The Modernization of Rural French 1870-1914*. Stanford: Stanford University Press.

Williams, Raymond. 1977. *Marxism and Literature*. Oxford: Oxford University Press.

Woolard, Kathryn A. 1998. Introduction: Language Ideology as a Field of Inquiry. In Kathryn A. Woolard and Paul V. Kroskrity (eds.) *Language Ideologies: Practice and Theory*. Bambi B. Schieffelin. Pp. 3-47. New York: Oxford University Press.

Woolard, Kathryn, and Bambi B. Schieffelin. 1994. Language Ideology. *Annual Review of Anthropology* 23:55-82.

Yukti Mukdawijitra. 2007. *Ethnicity and Multilingualism: The Case of Ethnic Tai in the Vietnamese State*. Ph.D. dissertation, Department of Anthropology, University of Wisconsin-Madison.

Zaman, Mukhtar. 1984. *The Language Policy of India: Problems of its Implementation and their Solution*. Islamabad: National Language Authority.

Zhou, Minglang. 2003. *Multilingualism in China: The Politics of Writing Reforms for Minority Languages 1949-2002*. Berlin: Mouton de Gruyter.

Appendix

This appendix exhibits a collection of Thái scripts in four tables: Table 1: Tai Dam Consonants, Table 2: Tai Don Consonants, Table 3: Tai Daeng Consonants, and Table 4: Tai Dam, Tai Don, Tai Daeng Vowels. Some remarkable differences are:

(1) Tai Don consonants have six more letters than other Thái consonants because Tai Don dialect has phonemes low-high pʰ, low-high cʰ, and low-high x that are not found in other Thái dialects.

(2) Tai Don Lai Châu consonants and Tai Don Phong Thổ consonants differ only in the form of the low-high kʰ and low-high x. Interestingly, they are similar to the low-high kʰ of Tai Daeng Thanh Hóa and Tai Daeng Mai Châu scripts.

(3) Tai Daeng consonants from various regions differ from one another a great deal in their forms and location of vowels. For instance, although the vowel ɛ of Tai Daeng Lai Pao looks similar to the vowel ɛ of other Thái, Lao Pao ɛ comes after the initial consonant. However, other Thái ɛs, except Quỳ Châu ɛ, are located before the initial consonant.

(4) The most distinctive one is the Tai Daeng script of Quỳ Châu which is written vertically from right to left, while other Thái scripts are written horizontally from left to right. Quỳ Châu vowel locations are thus distinctive.

(5) In terms of vowels, in general, while i, ɛ, ɔ, ɨ, and u are relatively similar in all Thái scripts, some are very different, like e, o, ə, and ɨa. However, Tai Don scripts are distinctive in their smaller number of diphthongs in comparison to other Thaí scripts.

Table 1: Tai Dam Consonants

Tai Region	Taidam Điện Biên Phủ, Sơn La, Nghĩa Lộ	
Phonetic	SIL Tai Dam	TTLTQG-3, KTTTB
low p	(glyph)	(glyph)
high p	(glyph)	(glyph)
low pʰ	–	–
high pʰ	–	–
low b	(glyph)	(glyph)
high b	(glyph)	(glyph)
low m	(glyph)	(glyph)
high m	(glyph)	(glyph)
low ʋ	(glyph)	(glyph)
high ʋ	(glyph)	(glyph)
low f	(glyph)	(glyph)
high f	(glyph)	(glyph)
low t	(glyph)	(glyph)
high t	(glyph)	(glyph)

Table 1: Tai Dam Consonants (continued)

Tai Region	Taidam Điện Biên Phủ, Sơn La, Nghĩa Lộ	
Phonetic	SIL Tai Dam	TTLTQG-3, KTTTB
low tʰ	૨	૨
high tʰ	૫	૨
low d	∩	∩
high d	૫	૫
low n	ʋᴜ	ʋᴜ
high n	૫	૫
low s	x	x
high s	૫	૫
low l	૫	૫
high l	૫	૫
low c	૫	૫
high c	૨	૨
low cʰ	–	–
high cʰ	–	–

Table 1: Tai Dam Consonants (continued)

Tai Region	Taidam Điện Biên Phủ, Sơn La, Nghĩa Lộ	
Phonetic	SIL Tai Dam	TTLTQG-3, KTTTB
low ɲ	ʋɤ	ᴎᴎ
high ɲ	ᴎ	ᴎ
low j	✓	✓
high j	✓	✓
low k	n	u
high k	α	α
low kʰ	ʌ	ʌ
high kʰ	3	3
low x	-	-
high x	-	-
low ŋ	ʋɤ	ʋɤ
high ŋ	6	6
low ʔ	θ	θ
high ʔ	9	9
low h	ɯ	ɯ
high h	ʋ	ɯ

Table 2: Tai Don Consonants

Tai Region	Lai Châu				Phong Thổ			
Phonetic	Lefevre-Pontails (1892)	Martini (1954)	TTLTQG-3, KTTTB	Ferlus (1988)	Silvestre (1918)	Martini (1954)	TTLTQG-3, KTTTB	Đồng Trọng Im (2004)
low p						–		
high p					–	–		
low pʰ			–			–	–	
high pʰ			–	–			–	
low b					–	–		
high b			–	–	–	–		
low m						–		
high m						–		
low ʋ					–			
high ʋ					–	–	–	
low f						–		
high f					–	–		
low t								
high t					–			

Table 2: Tai Don Consonants (continued)

Tai Region	Lai Châu				Phong Thổ			
Phonetic	Lefevre-Pontails (1892)	Martini (1954)	TTLTQG -3, KTTTB	Ferlus (1988)	Silvestre (1918)	Martini (1954)	TTLTQG -3, KTTTB	Đồng Trọng Im (2004)
low tʰ	[script]	[script]	[script]	[script]	–	[script]	–	[script]
high tʰ	[script]	[script]	[script]	–	–	[script]	–	[script]
low d	[script]	[script]	[script]	[script]	–	–	–	[script]
high d	[script]	[script]	[script]	–	[script]	–	–	[script]
low n	[script]	[script]	[script]	[script]	[script]	–	[script]	[script]
high n	[script]	[script]	[script]	[script]	[script]	–	[script]	[script]
low s	[script]	[script]	[script]	[script]	[script]	–	[script]	[script]
high s	[script]	[script]	[script]	[script]	[script]	–	[script]	[script]
low l	[script]	[script]	[script]	[script]	[script]	–	–	[script]
high l	[script]	[script]	[script]	[script]	[script]	–	–	[script]
low c	[script]	[script]	[script]	[script]	[script]	–	–	[script]
high c	[script]	[script]	[script]	[script]	[script]	–	–	[script]
low cʰ	–	[script]	–	–	[script]	–	–	[script]
high cʰ	[script]	[script]	–	–	–	–	–	[script]

Table 2: Tai Don Consonants (continued)

Tai Region	Lai Châu				Phong Thổ			
Phonetic	Lefevre-Pontails (1892)	Martini (1954)	TTLTQG-3, KTTTB	Ferlus (1988)	Silvestre (1918)	Martini (1954)	TTLTQG-3, KTTTB	Đồng Trọng Im (2004)
low ɲ							–	
high ɲ							–	
low j								
high j						–		
low k						–		
high k						–		
low kʰ			–				–	
high kʰ			–		–		–	
low x	–		–				–	
high x	–		–	–	–		–	
low ŋ				–	–			
high ŋ						–		
low ʔ						–	–	
high ʔ				–		–	–	
low h						–		
high h					–	–		

Table 3: Tai Daeng Consonants

Tai Region	Tai Daeng			
Phonetic	**Thanh Hóa** (Silvestre 1886 c.f. Finot 1917: 16)	**Mai Châu** (Hà Công Tín 1998)	**Quy Châu** Ferlus 1993	**Lai Pao** Ferlus and Dõi 1997
low p				
high p				-
low pʰ	-	-		-
high pʰ	-	-	-	
low b				
high b				-
low m				
high m			-	
low ʋ			-	
high ʋ				-
low f				-
high f	-			-
low t				
high t		-		-

Table 3: Tai Daeng Consonants (continued)

Tai Region	Tai Daeng			
Phonetic	Thanh Hóa (Silvestre 1886 c.f. Finot 1917: 16)	Mai Châu (Hà Công Tín 1998)	Quy Châu Ferlus 1993	Lai Pao Ferlus and Dõi 1997
low tʰ	ʒ	ʒ	ʒ	-
high tʰ	-	ᴎ	-	-
low d	∩	∕)	-	ᴦ
high d	ʊ	ᴎ	ɡ)	ᴡ
low n	ᴜ	ᴡ	-	ᴚᴍ
high n	ʉ	ᴎ	ᴢᴄ	ᴍ
low s	ʈ	x	ʒ	-
high s	ᴎ	-	ᴃ	-
low l	ᴣ	ʋ	ᴤ	ᴣᴄ
high l	ʋ	y	-	ᴣ
low c	ʊ	ʃ	ᴢᴊ, ᴄ	-
high c	ʒ	ᴢ	-	-
low cʰ	-	-	-	ᴄᴍ
high cʰ	-	-	-	-

Table 3: Tai Daeng Consonants (continued)

Tai Region	Tai Daeng			
Phonetic	**Thanh Hóa** (Silvestre 1886 c.f. Finot 1917: 16)	**Mai Châu** (Hà Công Tín 1998)	**Quy Châu** Ferlus 1993	Lai Pao Ferlus and Dõi 1997
low ɲ				
high ɲ	-			-
low j				-
high j	-		-	-
low k				
high k	-		-	
low kʰ				
high kʰ				-
low x	-	-		
high x	-	-		
low ŋ	-	-		
high ŋ			-	
low ʔ				-
high ʔ	-		-	-
low h			-	
high h				

Table 4: Tai Dam, Tai Don, Tai Daeng Vowels

Phonemes	Tai Dam	Tai Don Lai Châu	Tai Don Phong Thổ	Tai Daeng Thanh Hóa	Tai Daeng Mai Châu	Tai Daeng Quy Châu	Tai Daeng Lai Pao
i							
e							
ɛ							
a							
aː							
ɨ							
ə							
u							
o							
ɔ							
aɨ							
ai							
ao							
ia							
ɨa							
ua							

Table 4: Tai Dam, Tai Don, Tai Daeng Vowels (Continued)

Phonemes	Tai Dam	Tai Don Lai Châu	Tai Don Phong Thổ	Tai Daeng Thanh Hóa	Tai Daeng Mai Châu	Tai Daeng Quy Châu	Tai Daeng Lai Pao
oa	−	−	−	−	−	−ᵗ	−
an	− ᵗ	− ᵗ	−	✓ − ι	−ᵩ	−	ᵭ , ̅
am	ᵭ ᵂ ⁄	ᵭ , ᵂ , ⁄	−	ᵭ ᵂ ⁄	ᵕ	ō̅ ₂	−ᵕ
ap	−	−	−	−	−	−	⁀
at	−	ᵿ	−	ᵿ	⁄	−ᵧ	−
aŋ	−	−	−	−	⁄	− ᵐ	− ᵹ̅
oi	−	−	−	−	−	−	⌢

A 50-YEAR COMPARISON OF REGIONAL DIALECT VARIATION IN THE SUI LANGUAGE[72]

James N. Stanford
Dartmouth College
< *James.N.Stanford@Dartmouth.edu*>

Abstract
This Sui dialect geography study conducted new fieldwork to examine changes among regional dialects across a time span of 50 years. The new field results were compared to an unpublished 1950s Sui dialect survey, *Shuiyu Diaocha Baogao*. The results provide new insights about this particular Tai-Kadai language and also new perspectives for the study of dialects and physical space in other small, rural indigenous communities across Southeast Asia.

Key words: Dialects, Dialectology, Sui, Tai-Kadai, Sociolinguistics

1. Introduction

Prior work has provided a great amount of progress in understanding the structure of Tai-Kadai languages and their historical/comparative relationships (e.g., Edmondson & Solnit 1988, 1997; Diller et al. 2008). But dialect geography tends to be understudied in Tai-Kadai research, especially in small, rural communities. This is unfortunate since physical space has long been viewed as an important aspect of human language (e.g., Bloomfield 1933:476; Auer & Schmidt 2010). Recent dialect geography in other parts of the world, such as the Labov et al. (2006) *Atlas of North American English* and Kretzschmar's analysis of the *Linguistic Atlas of the Middle and South Atlantic States* (e.g., 2009:64-145), have provided new knowledge about characteristics of human language from the perspective of well-known languages. New dialect geography research is now needed for understudied Tai-Kadai languages. In the same way that structural or historical analyses of a particular Southeast Asian language can shed light on area languages, Sui dialectology can provide insights for area languages as well.

The present study conducted new fieldwork on regional dialect variation among the indigenous minority Sui people of Guizhou Province, China. The new fieldwork was compared to an unpublished 1956 survey of the same region: the handwritten manuscript

---

Shuiyu Diaocha Baogao[73] "Report on Investigations of the Sui language" (henceforth SDB). That early study surveyed 16 speakers representing 16 Sui locations. Now approximately 50 years later, fieldwork was conducted with 33 speakers representing 17 locations. The results show that the basic geographic distinctions among Sui dialects have remained quite stable over 50 years. Although some details have changed and new variables have emerged, Sui dialects have generally retained the same long-term regional patterns.

The results not only provide new knowledge about Sui dialects in particular, but also generalizations about the role of space for Tai-Kadai and other rural indigenous minorities of Southeast Asia. Specifically, the Sui results echo the overall stability of long-term dialect patterns found in the *Atlas of North American English* (Labov et al. 2006) and other work on large, majority languages. Despite the cultural, linguistic, and geographic contrasts, this tiny, rural Tai-Kadai community maintains dialect boundaries just as consistently.

Organization of the paper. The remainder of the paper is organized as follows: Section 2 presents background information about Sui. Section 3 describes research methods of the study. Section 4 presents maps and tables to compare the results of the present study with the 1956 SDB study. Conclusions are given in Section 5.

2. Background

The Sui people. The Sui ethnic group is concentrated in a relatively small area of Guizhou Province, a southwestern province of China. About 93% of Sui people live in Guizhou (Wei & Edmondson 2008:585). Moreover, within Guizhou, the Sui ethnic group is centred in Sandu Sui Autonomous County. This county is the only autonomous Sui region, and it is widely considered to be the cultural and linguistic center of the society. Sandu County was home to 189,128 Sui people as of 2000 (China National Census Bureau 2003; Andy Castro p.c.). Figure 1 shows the location of Guizhou Province within China, and Sandu County within Guizhou.

The fieldwork for this project is based on the author's field research trips in August 2010, August 2006, and August 2005. Additional linguistic and cultural background on Sui is based on the author's four years in Guizhou (1999-2003), where he learned to speak Sui.

Although many Sui people seasonally migrate to large Chinese cities for industrial work, the villages in rural Sandu County continue an agrarian lifestyle with distinctive Sui practices. The Sui people strictly practice clan exogamy: A wife and husband must be from different clans, and the wife moves permanently to the husband's village at the time of marriage (i.e., patrilocal clan exogamy).

The Sui language. Sui is a tonal, largely monosyllabic Tai-Kadai language, a family whose more well-known members are Thai and Lao (Edmonson & Solnit 1988; Diller, Edmondson & Luo 2008; Burusphat, Wei & Edmondson 2003). Other Sui research includes Li (1948, 1965), Zhang (1980), Zeng & Yao (1996), Edmondson et al. (2004), Castro (2011), and Stanford (2007a-c, 2008a-b, 2009).

---

[73] Acquired from Jerold Edmondson, University of Texas-Arlington.

Figure 1: Sandu Sui Autonomous County in Guizhou Province, China.
[Map Source: Evans Map Room, Dartmouth College Library/lmh]

Sui is the primary language spoken in rural Sui villages. It is the first language acquired by children in the community, although many Sui people are able to speak Chinese to outsiders (Southwest Mandarin). Older women are usually monolingual in Sui. The present study focuses on native Sui words, although Chinese loanwords are noted whenever they appear in place of native lexical items in the interviews. Although Sui is the primary language of oral communication in rural areas, written communication is conducted in Chinese. A Sui alphabetic writing system was developed by Chinese scholars (cf. Zeng & Yao 1996:262; Luo 1992:153-55), but it is rarely used in daily life (cf. Zhou 2003:133-6). There is also a set of character-based Sui shamanistic symbols for very limited domains of use (Wei 2007; Luo 1992:147-152).

Sui tones. Table 1 shows tones of unchecked syllables for speakers in a village in the Sandong Township region.[74] The tone values in Table 1 are based on the author's research and Zeng & Yao (1996), presented with Chao's (1930) 1-5 pitch scale. For example, 33 is a mid-level tone, 52 is high-falling tone, 13 is a low-rising tone, etc. Tone numbers are written as superscripts, e.g., [fa³] 'cloud' indicates the syllable bears Tone 3.

---

[74] Possible regional variation in checked tones awaits future study. The present study focuses on unchecked tones since the clearest regional variants are in the unchecked tones (Tones 1 and 6).

Table 1: Tones in Sandong Township (unchecked syllables)

Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6
13	31	33	52	35	55

Tone 6 is 24 in some regions, and subtle regional variation in Tone 1 is also reported (discussed below). In all dialects, Tone 6 is high level 55 for recent Chinese loanwords having Mandarin Tone 3.

Sui dialects. Sui is typically divided into three major dialect divisions (e.g., SDB 1956:137 and Zhang 1980). The Sandong dialect encompasses the central part of Sui society. Two smaller dialects, Pandong and Yang'an, are found in the northwest and west, respectively. All three dialects are considered mutually intelligible. Speakers report that Yang'an dialect is the most divergent of the three dialects. In addition to these three dialects in prior literature, Castro (2011) has recently suggested that there is a fourth Sui dialect in the southern region. As discussed below, Castro's analysis is consisted with the results of the present study.

Besides these major dialect divisions, fine-grained local distinctions are found in phonological and lexical variables. As with Smith & Johnson's (1986) findings for the clan-oriented Nganhcara people of Australia, Sui dialect features are primarily phonological and lexical, rather than morphosyntactic.

3. Methods

Locations. SDB (1956) sampled 16 speakers across the Sui region, and the present study recorded 33 speakers representing locations across a similar area. The fieldwork included as many of the original SDB locations as possible, as well as additional data points in the more populous central valley region to provide a higher level of detail. Figure 2 shows the locations represented in the present study (black circles labeled with uppercase letters A-Q) and the SDB locations. The SDB locations are represented as gray triangles that are numerically labeled with the original SDB numbers of those locations (#1-9). Note the distance scale in the map which shows how tiny this region is in comparison to more well-known dialect geography (e.g., Labov et al. 2006). For reference, a few location names are given in Chinese Pinyin Romanization, following contemporary Chinese administrative names and boundaries. A full list of location names in Chinese and Sui is provided in Table 2.

The placemarkers in Figure 2 were positioned according to the coordinates of villages in Google Earth, and then the latitude/longitude coordinates were transferred to GIS mapping software (Evans Map Room, Dartmouth College). Where specific village names were not available, a Sui informant helped mark a map with the most likely location (points D, H, K, N).

Between the eastern township of Bajie and the central region (Shuilong/Zhonghe/Sandong), there is a mountainous, less-populated area. That area was not sampled in either SDB or the current study. Likewise, the northwest region between Sanhe and Jichang/Yanghe is also relatively mountainous and not sampled in either study. In addition to the locations shown in Figure 2, SDB also surveyed five people from other nearby counties: Two locations further west in Dushan County (SDB #7a and #9c), one in Rongjiang County east of Sandu (#8), and two in Libo County south of Sandu (#3a and #4). Those five locations were not investigated in the present study since the author's

contacts are primarily located in Sandu County (locations C-Q) and in the Yanghe/Jichang region (locations A-B). Moreover, Sandu County is considered the linguistic, cultural, and administrative center of Sui society, so the focus on central Sandu County is appropriate for the study. Since the central region of Sandu County has a relatively high concentration of villages, the present study collected additional points in that area (C, D, G, H, J, K, O) beyond the three points in SDB (#1, 2, 6).
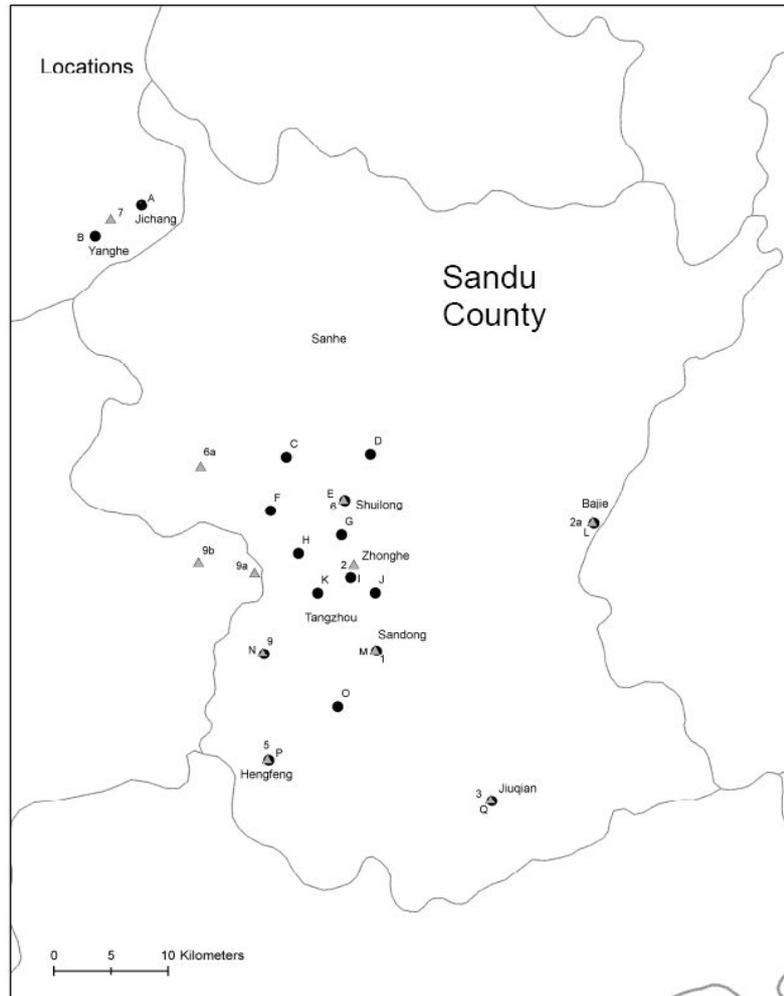
Figure 2: The locations of the present study are black circles ● labeled with an uppercase alphabet letter A-Q. Locations of SDB (1956) are gray triangles ▲ labeled with numbers (1-9). Where both studies have samples representing the same location, a gray triangle is placed on top of a black circle (e.g., P and 5).
(Map source - Evans Map Room, Dartmouth College Library/lmh)

A significant Sui population is also found in Sanhe, the county seat of Sandu County (pop. 9,032 according to Liu et al. 2002). However, rather than being a part of rural, traditional Sui society, Sanhe is a rapidly growing city with a diverse collection of Sui people from all over the county, as well as ethnic Chinese people from other parts of Guizhou and China as a whole. The city is a recent phenomenon, almost tripling in size from 1992 to 2002 (cf. Luo 1992:773; Liu et al. 2002:898), and it receives significantly more influence from the Chinese language than rural Sui areas do. This city would be

interesting to study in terms of linguistic contact and sudden social changes, but it is not included in SDB or the current study.

Table 2: Informants and locations represented in the present study

33 informants (gender, age):	17 regions represented:	Sui toponym and Chinese toponym	Nearest SDB 1956 location
Female, 25 years old	A	Sui: $t^ha{:}u^5$ Chinese: Jichang	7
Male, ~35	B	Sui: $t^ha{:}u^5$ Chinese: Yanghe	7
Female, 28	C	Sui: $miu^1\ ja^1\ (ʔju^1)$ Chinese: Miaocao	6a/6
Female, 47	D	Sui: $^ndi^5$ Chinese: Dixiang	6
2 female speakers: 35, 40 5 male speakers: 24, 28, 30, 40, 45	E	Sui: $ljoŋ^2$ Chinese: Shuilong	6
Female, 39	F	Sui: $ɣɔŋ^3$ Chinese: ?	9ab/6a
Male, 30	G	Sui: $pa{:}ŋ^4$ Chinese: Xiyang	2
Female, 15	H	Sui: $mɔŋ^6\ (ʔjɔ^3)$ Chinese: Yangmeng, Tangzhou	9a
Male, 22	I	Sui: $ku^3\ jin^6$ Chinese: Guyin	2
Male, 35	J	Sui: $lja{:}i^6$ Chinese: Jiahua	2
Female, 36	K	Sui: $p^ha{:}n^1\ (ʔjɔ^3)$ Chinese: Tangzhou	9a/2
Female, 31	L	Sui: $ʔnja^1$ Chinese: Bajie	2a
6 female speakers: 29, 36, 38, 40, 41, 41 4 male speakers: 23, 27, 39, 42	M	Sui: $tɔŋ^6$ Chinese: Sandong	1
Female, 37 Male, 24	N	Sui: $ja{:}ŋ^2$ Chinese: Yang'an	9
Female, 33	O	Sui: $^ndɔŋ^1$ Chinese: Shuidong	1
Male, 37	P	Sui: $ʔɔŋ^1$ Chinese: Hengfeng	5
Male, 24	Q	Sui: $tɕu^3\ ɕɛn^1$ Chinese: Jiuqian	3

Informants. Table 2 lists the informants and locations. Note that it was possible to include multiple speakers for locations E and M since recordings from those two locations were available from a prior study (Stanford 2008a).

Ages. The informants interviewed for SDB were middle-aged adults and young adults (22-46 years old), with the exception of a 17 year-old. Likewise, the present study targeted middle-aged and young adults in the same age range: All informants analyzed in the present study were 22-47 years old, with the exception of a 15 year-old.

Gender and mobility. For the SDB 1956 project, there were 15 men and 1 woman. The paucity of female informants in SDB reflects traditional dialectology's focus on NORMs (Non-mobile, Older, Rural, Men) (cf. Chambers & Trudgill 1998). More recently, modern dialectology has come to recognize the importance of gender diversity, and the present study includes women. It was not possible to include a male and female speaker from each location, but the study has an overall balance of 17 women and 16 men. Women's speech is crucial in any study of Sui, especially since a gender-related factor (exogamy) plays a major role in social organization. It would therefore be far too simplistic to suggest that a Sui village can be represented as having a single dialect. After all, the married women in any village have necessarily come from other clans, and many clans have distinctive dialect features. At the same time, there is a strong emic notion that a given village has a dominant dialect, namely, the dialect of the local men, children, and teenagers.[75] The present study reports on the dominant dialect in each location, while recognizing the actual linguistic complexity of each village due to in-marriage.

In the present study, 11 recordings were conducted in or very near the home village of the informants (E, I, J, and M). Other informants were recorded in locations other than their home villages. SDB does not indicate the setting of their interview sessions, and it is likely that some of their informants were recorded in locations other than their home villages, such as students. Of course, permanent lifelong residents are the preferred representatives of the dialect of a given location. However, the effects of mobility on a Sui individual's dialect features are believed to be quite limited. Prior research strongly shows that, due to clan ideology and loyalty, Sui speakers' dialects are highly stable across the lifespan, regardless of mobility. Stanford (2008a) provides quantitative results showing that in-married women maintain the phonological features of their original dialects to a very high degree, even after decades in the husband's village. As for lexical variation, in two villages where in-married women were in daily contact with the local dialect of the husband's village, both the non-mobile residents and the in-married women categorically used the variants of their home village in all 226 recorded tokens of lexical variables in free speech (Stanford 2009:292). In addition, ethnographic interviews show that Sui clan ideology encourages individuals to carefully maintain their original dialects (Stanford 2009). There is a strong Sui notion of loyalty to one's original home clan and village, and this loyalty is linguistically constructed as each speaker continually uses the father's dialect features, regardless of any later mobility.

Three determining factors were involved in the choice of informants and locations. First, it was not possible to personally visit all of the SDB regions due to cultural constraints. The author, a Westerner who has learned to speak Sui, has personal contacts in many but not in all of the regions of SDB.

---

[75] See Stanford 2008b for an investigation of Sui child dialect acquisition.

A second factor in selecting informants is related to women's mobility. Since Sui women traditionally marry in their late teenage years or early 20s and then move to the husband's village, most adult female potential informants in a given village were raised elsewhere. In fact, owing to the Sui dialect stability across the lifespan as discussed above, these exogamous customs provide an opportunity to interview informants representing locations that would otherwise be difficult to access.

Thirdly, Sui society as a whole has become more mobile than it was at the time of SDB. Each year, many Sui people leave their villages to travel to Chinese cities for migrant labor opportunities where they interact with other Sui people, Chinese people, and other minorities. Other Sui people have opportunities to teach or work in Sui towns, rather than spending their lives primarily in their home villages. Informants representing locations Q, P, N, G and C were recorded in a local town (Zhouqin) where they were working. The speaker representing location B was a migrant laborer recorded in the city of Duyun outside of the Sui area.

Interviews. The interviews were conducted in spoken Sui by the author, who occasionally also spoke Chinese with bilingual informants. Interviews consisted of asking informants to identify common everyday pictures, objects, actions (e.g., standing/sitting), eliciting antonyms of given words, and counting. This interview protocol produced about 90-110 words from each informant. The speech style in this study was more conversational than a typical word-list style. Rather than simply reading through a list of words, informants examined each picture or object and then identified it. The overall interview approach follows Chambers' (1992) dialect acquisition research, where a picture identification task is used so that informants are not influenced by hearing the word in advance. Some of the informants, especially older women, were monolingual in Sui and non-literate, so the study was designed to allow for their participation. By contrast, the SDB interviews apparently only used bilingual informants who could recognize words in Chinese. All interviews were recorded with an Edirol R-09HR digital recorder or digitized from analog cassette recordings on a Marantz recorder (locations E and M).

4. Results

SDB's overall dialect boundaries were found generally intact after 50 years. Among the data reported in SDB, there are 18 regional variables that can be tested against the current study: 14 variables in the SDB maps and 4 other variables in SDB's data tables. In this section, key representative examples from the 18 items are discussed along with maps and tables. Due to space limitations, it is not possible to discuss each of the 18 items individually in detail, but Table 5 (section 4.5) provides a full list and summary of the result for each item.

Note that the discussion of these features is not intended to imply that these particular variables are sufficient to characterize Sui dialects (cf. Nerbonne & Heeringa 2010:550). Britain notes that traditional dialect geography has long been criticized for portraying boundaries as "abrupt, discrete, and invariable" when the reality is far more complex (Britain 2002:629; cf. Kretzschmar 2009:66ff; Nerbonne 2009:187-89). The present study is designed as a real-time comparison of a set of Sui dialect features that can be examined across two studies, rather than a comprehensive regional dialect description of Sui.

4.1 Regional Contrasts

Contrasts in the word 'boat'. Figure 3a shows SDB's results for the word 'boat'. Horizontal lines represent the pronunciation [lua]; vertical lines represent [ʔda], and blank space represents [lwa]. Figure 3b shows that the same contrasts for 'boat' appear 50 years later in the present study. Because isoglosses can sometimes be misinterpreted as implying greater uniformity than the data may support, the maps of the current study are presented without isoglosses.

Comparing Figures 3a and 3b, it is clear that the three-way regional contrast in 'boat' has remained stable over the 50-year period. (Note from Figure 2 that SDB #6 corresponds to location E, SDB #2 corresponds to location I, and SDB #3 corresponds to location Q.) The current study transcribes the vowels slightly differently, which is most likely due to modern availability of acoustic analysis (Stanford 2007c) and different analysis of glides. SDB analyzes the glide as part of the onset [lw-], following a particular phonemic analysis; SDB shows the regional contrast as [lwa] versus [lua]. The present study phonetically transcribes the same regional contrast as [lua] versus [luə]. Acoustic analyses show a contrast in [-a] versus [-ə]. In addition, the [luə] region generally has a slightly longer and fronted [u]. These differences in [u] are probably reflected in SDB's transcription of a glide to represent the same regional contrast.

In addition, the current study finds a Chinese loanword [suən²] for 'boat' in the northwest region, locations A and B. The nearest large Chinese city is Duyun, located about 30 kilometers northwest of locations A/B. Speakers in locations A/B reported that they have more Chinese loanwords than people living in central Sandu County.

Finally, as noted in Section 2, Castro (2011) proposed that there is a fourth Sui dialect, a "southern dialect," not just the traditional three dialects of earlier work (e.g., Zhang 1980). Note that Castro's analysis is supported by the distinctive southern regional variants of 'boat' in SDB (vertical lines in Figure 3a) and in the current study (point Q in Figure 3b).

East-west contrasts. Among the contrasts in 'boat', two of the variants ([lua] and [luə]) are related to a more unified diphthongal contrast that is found throughout the lexicon. There are two diphthongal variables, symbolized here as (ua) and (ia), and they pattern geographically in an east-west contrast. The (ua) variable is realized with the regional variants [-ua] versus [-uə]. The (ia) variable is realized as [-ia] versus [-iə]. The two diphthong variants have the same regional distribution: Speakers who use the [-ua] variant of (ua) use [-ia] for (ia). Speakers who use the [-uə] variant of (ua) use the [-iə] variant for (ia).

The current study and SDB find the same diphthong contrasts in the same east-west geographic distributions, although SDB transcribes the two variant pairs respectively as [-wa] versus [-ua], and [-ja] (sometimes [-jɛ]) versus [-ia]. SDB includes the glides as part of the onset for the eastern variants, matching their overall phonemic analysis of the language. For the present sociophonetic study, the vowels of all variants are written phonetically. In addition, acoustic analysis (Stanford 2007c) suggests the final vowel of the western variant is more centralized than some of SDB's impressionistic transcriptions show. Regardless, it is clear that the regional line of east-west contrast in these two diphthongs remains the same after 50 years. For SDB, Figure 4 shows the (ia) contrast, and Figure 3a shows the (ua) contrast (horizontal lines versus white space). Figure 5 shows the results of the current study for both diphthong variables (ua) and (ia).
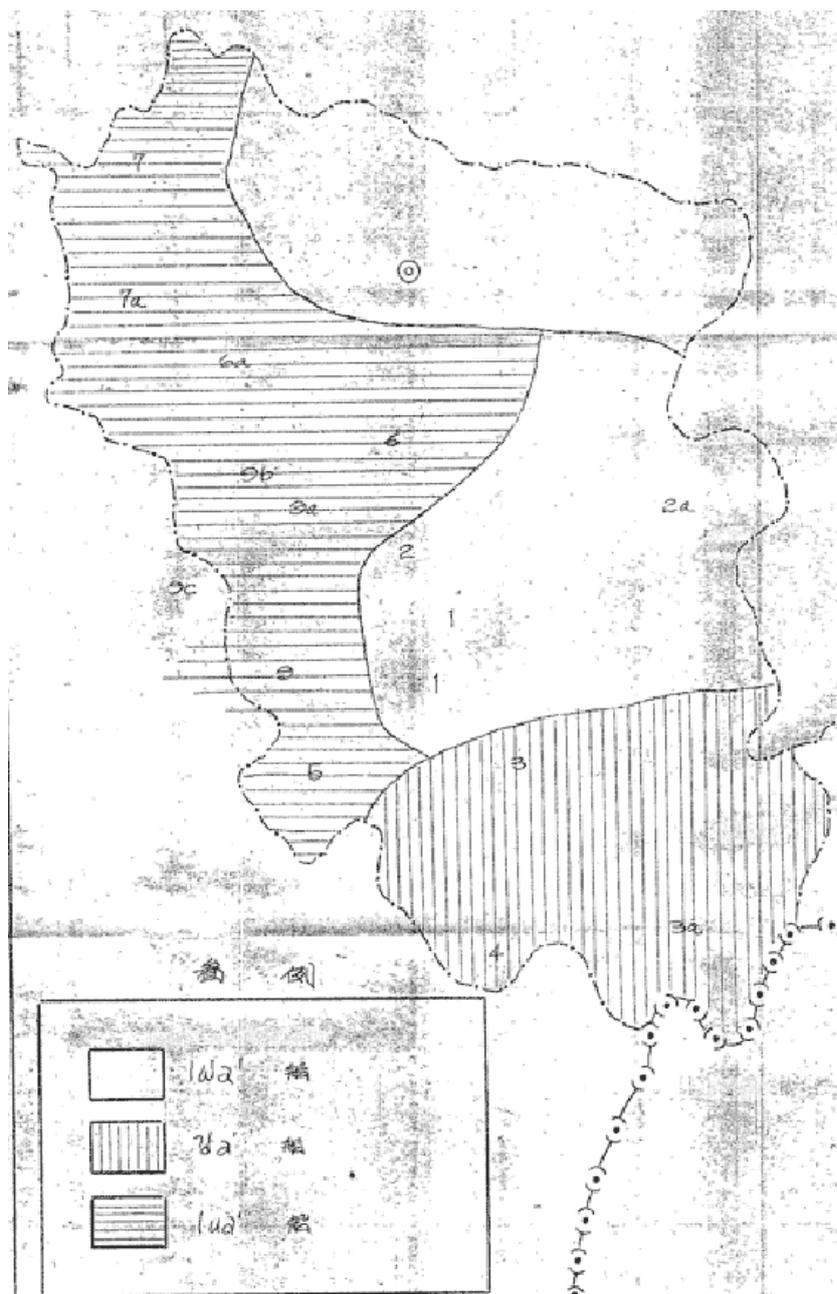
Figure 3a: The word 'boat' in SDB (1956, unpublished ms., map 6).
Isoglosses for three variants of the word 'boat': [lwa], [lua], [ʔda].
Numbers 1-9 represent informant locations. The small concentric circle symbol represents
the county seat, Sanhe, but no data was collected there for SDB nor the present study.
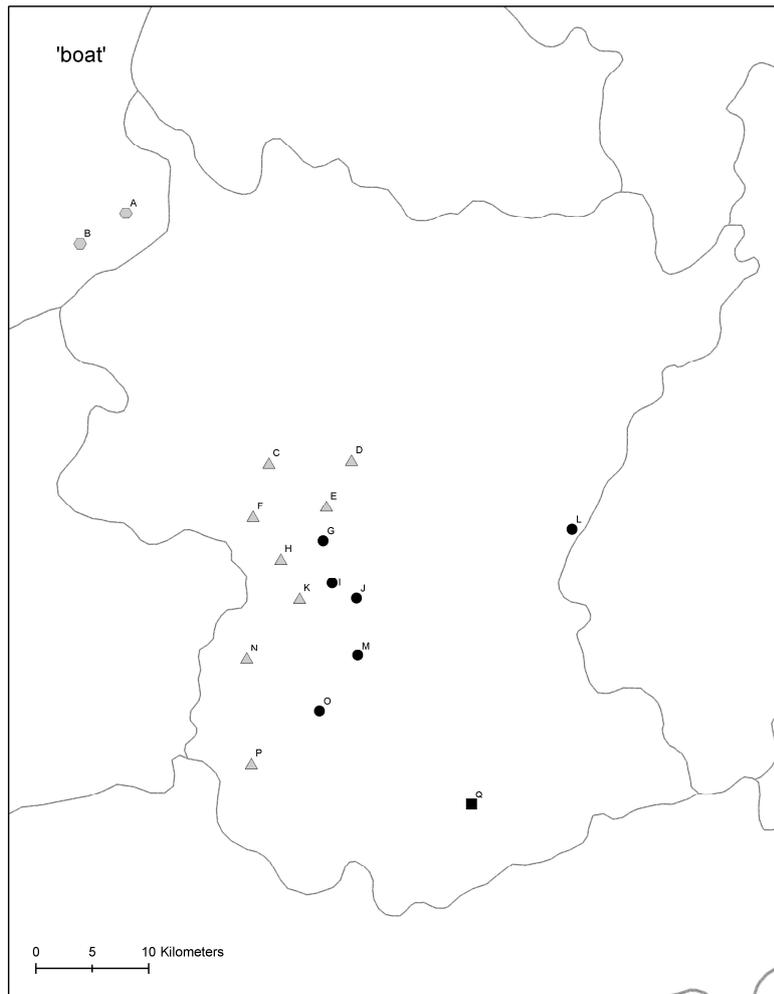
Figure 3b: The word 'boat' in the current study: ● = [lua^1], ▲ = [luə^1], ■ = [ʔdua^1], ⬡ = [suən^2] (Chinese loan) (Map source - Evans Map Room, Dartmouth College Library/lmh)
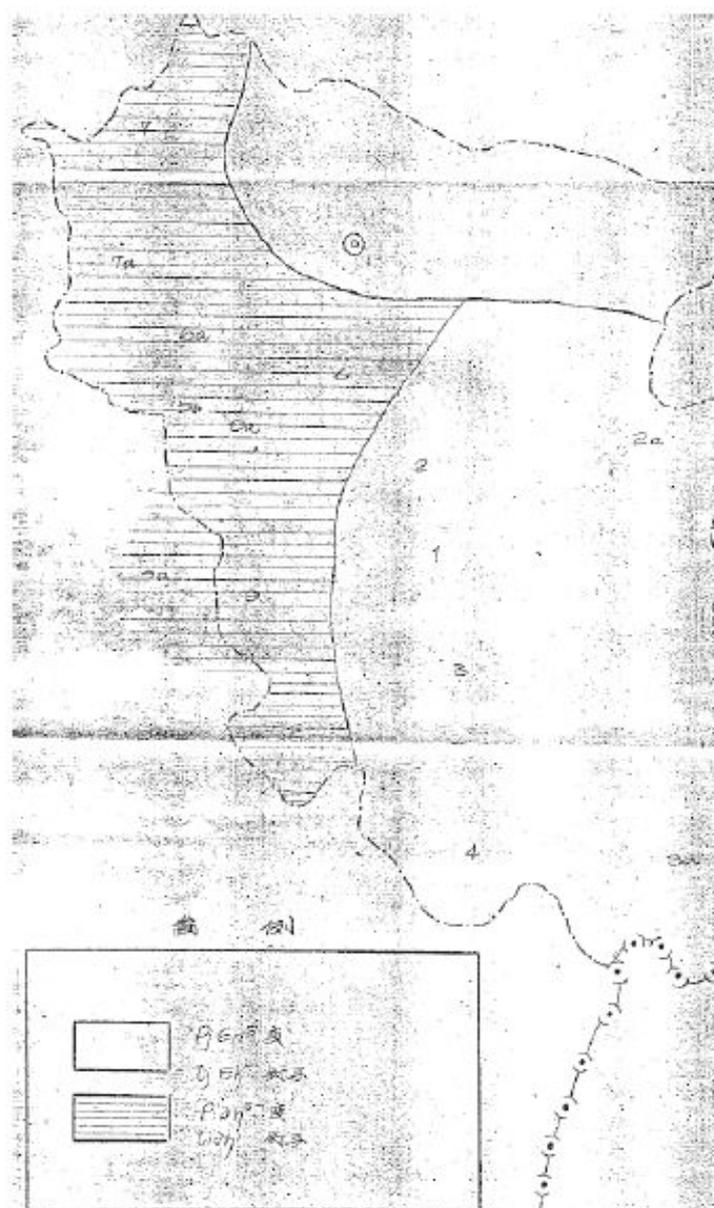
Figure 4: An SDB map showing the east-west contrast in diphthongs (SDB map 1)
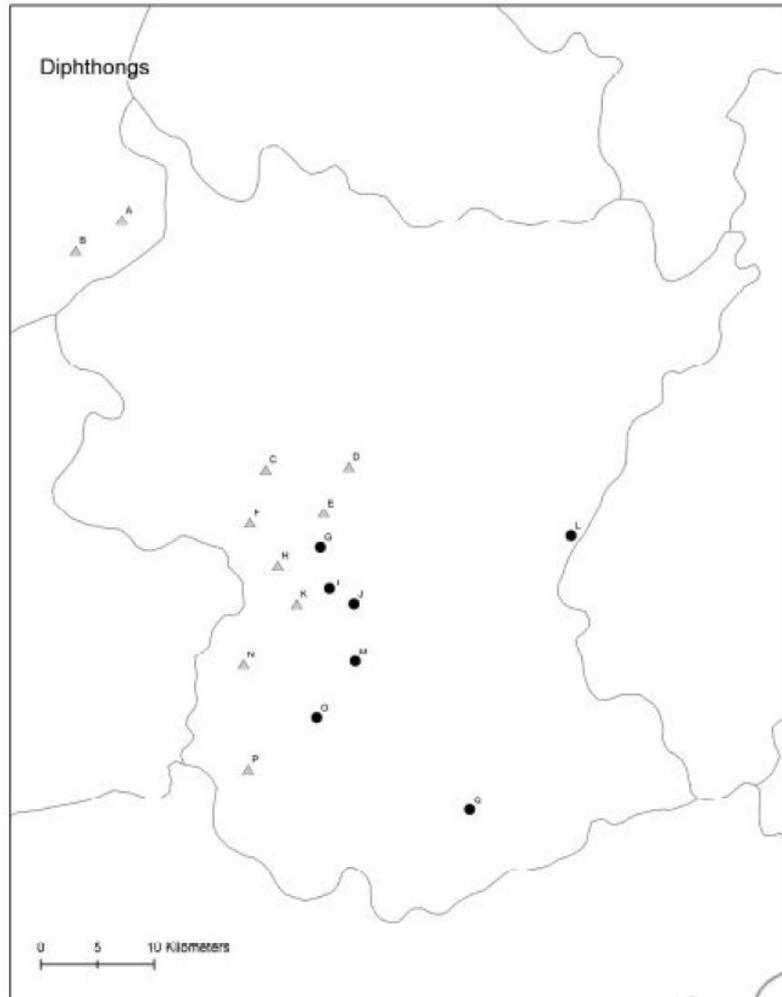
Figure 5: Diphthong variants in the current study. (ua) has two variants: ● = [-ua],
▲ = [-uə]. The results are exactly the same for the two variants of (ia): ● = [-ia], ▲ = [-iə].
(Map source - Evans Map Room, Dartmouth College Library/lmh)

Another east-west contrast. The following two variables also have an east-west contrast: the presence/absence of preglottalized consonant onsets, e.g., 'nose' [ʔnəŋ] versus [nəŋ],[76] and the presence/absence of voiceless nasal onsets, e.g., 'dog' [m̥a] versus [hwã]. These two variables pattern together in both SDB and the current study: Speakers who produce a preglottalized consonant in words like 'nose' [ʔnəŋ] use a voiceless nasal for words like 'dog' [m̥a]. Speakers who pronounce 'nose' as [nəŋ] pronounce 'dog' as [hwã]. Figure 6 shows the results of the current study for preglottalized consonant onsets and voiceless nasal onsets. SDB's maps have the same result (cf. SDB maps 2, 4, 7, not shown here due to limited space).

Distinctions between northwest, west and central. In addition to the east-west contrasts described above, both studies report a distinction between the northwest (locations A and B) and west (location N), as seen in the Table 3 results for 'sky', 'arm', 'straight', and 'water buffalo'.

---

[76] See Edmondson et al. (2004) for acoustic analysis of Sui preglottalized consonants.

Table 3: Examples of distinctions between northwest, west, and central. Results are the same in the current study and SDB except where noted.

	Northwest (A/B)	West (N)	Central (all other locations)
'arm'	εin^1	$k^h in^1$ SDB reports [kin^1]	$t\varepsilon^h in^1$
'sky'	van^1 (SDB reports [ban^1])	man^1	$?ban^1$
'straight'	$x^j a\eta^2$	$xa\eta^2$	$\varepsilon a\eta^2$ Exception: location F = [$j\varepsilon n^2$] in the current study.
'water buffalo'	qui^2	kui^2 SDB reports [qui^2]	kui^2
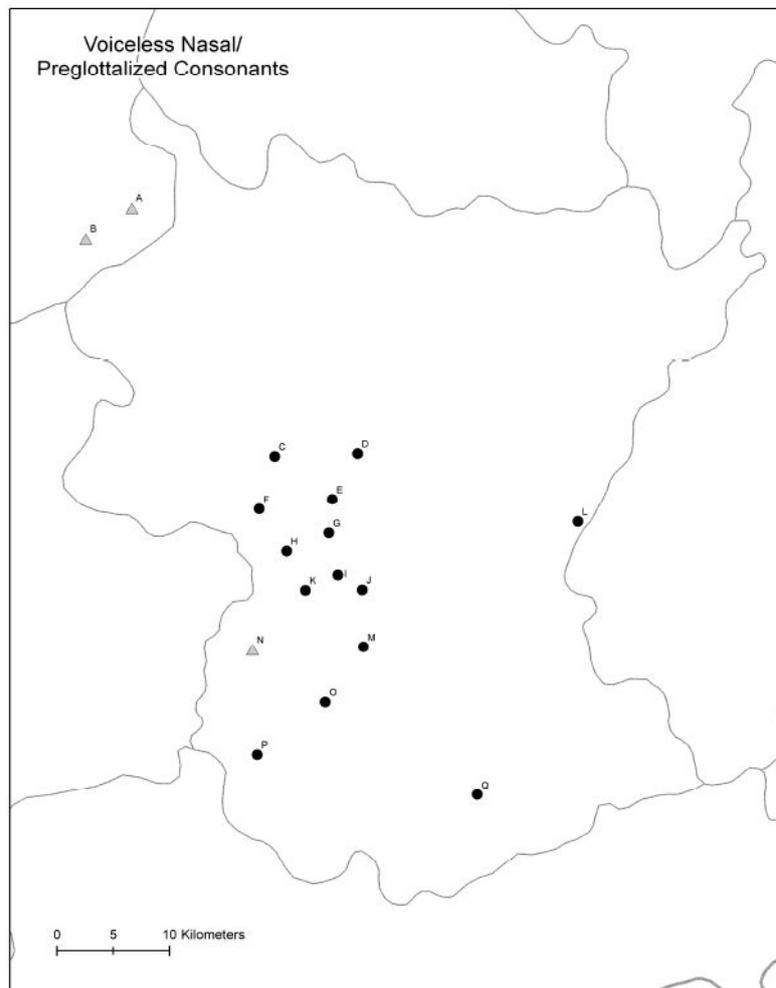
Figure 6: Results of the current study for preglottalized consonants and voiceless nasals.
● = [+preglottalized consonants, +voiceless nasals];
▲ = [-preglottalized consonants, -voiceless nasals]
(Map source - Evans Map Room, Dartmouth College Library/lmh)

Two other variables in SDB and the current study highlight the west/northwest/central contrast: lateral onsets replace [f-] in location N, and [xʷa-] replaces [fa-] in locations A/B.

Summary of major regional contrasts. The overall results confirm a stable three-way distinction of major Sui dialects, as proposed in SDB (1956:137) and other work (Zhang 1980). The "northwest" locations correspond to the "Pandong dialect" region as it is called in SDB and other literature. The "west" corresponds to the Yang'an dialect. "Central" corresponds to the Sandong dialect. In addition, Castro's (2011) notion of a fourth (southern) dialect is supported by the fact that both SDB and the current study confirm distinctive southern variants for 'boat' and 'salt' in point Q, a distinction that has remained stable over 50 years.[77]

4.2 Complexities

As dialectologists have found in many other studies, regional results are often less tidy than the above examples, and the results for a few well-behaved regional variants do not necessarily imply that a regional contrast strictly holds (Nerbonne & Heeringa 2010:550). For the word 'market' (Figure 7), both SDB and the current study show a stable but complex pattern: Location N unexpectedly patterns with the central group in this case.

Other results show similar complexities. The current study's finding for 'diligent' reflects SDB's result in general, but there are discrepancies (Table 4).

Table 4: Results for 'diligent'

	Northwest (A/B)	West (N)	Location Q[78]	Elsewhere
'diligent' in SDB	jak⁷	xak⁷		kʰak⁷
'diligent' in the current study	xʲak⁷	xak⁷ for one speaker in this location kʰak⁷ for the other speaker	xʲɐk⁷	kʰak⁷

1ˢᵗ Singular. The 1ˢᵗ Singular pronoun is worthy of special note since it is a highly salient variable that informants often mention in ethnographic interviews. The regional contrasts in 1ˢᵗ Singular are well known among Sui people, and they use the variants as shibboleths and common topics of overt comment, often playfully imitating different regions. Figure 8 shows the results for 1ˢᵗ Singular. The results in Figure 8 are consistent with SDB for all locations shared by the two studies. SDB does not include the two variants that the current study found for locations K and O.

Note that the [ɛi²] variants are clustered together. This is a more populous central valley region. This distribution might therefore be modeled as a case of long-term dialect-leveling (Trudgill 1986:98ff.), or perhaps contagious diffusion (Bailey et al. 1993) originating from one of the central locations. In the populous central valley region, most

---

[77] Further supporting Castro's (2011) "fourth dialect," the present study found distinctive southern variants for other words as well: 'cloud' [wa], 'bull' [mo], 'seedling' [tɕa], 'Sui New Year' [ta], 'sharp' [tɕʰa]. But these distinctions go beyond the current study since those words are not available in SDB for comparison.

[78] The vowel transcription [ɐ] is based on a suggestion by Andy Castro (p.c.).

other variants seem to be leveled out, while different variants exist in the more distant parts of the valley and in comparatively remote mountainous regions.
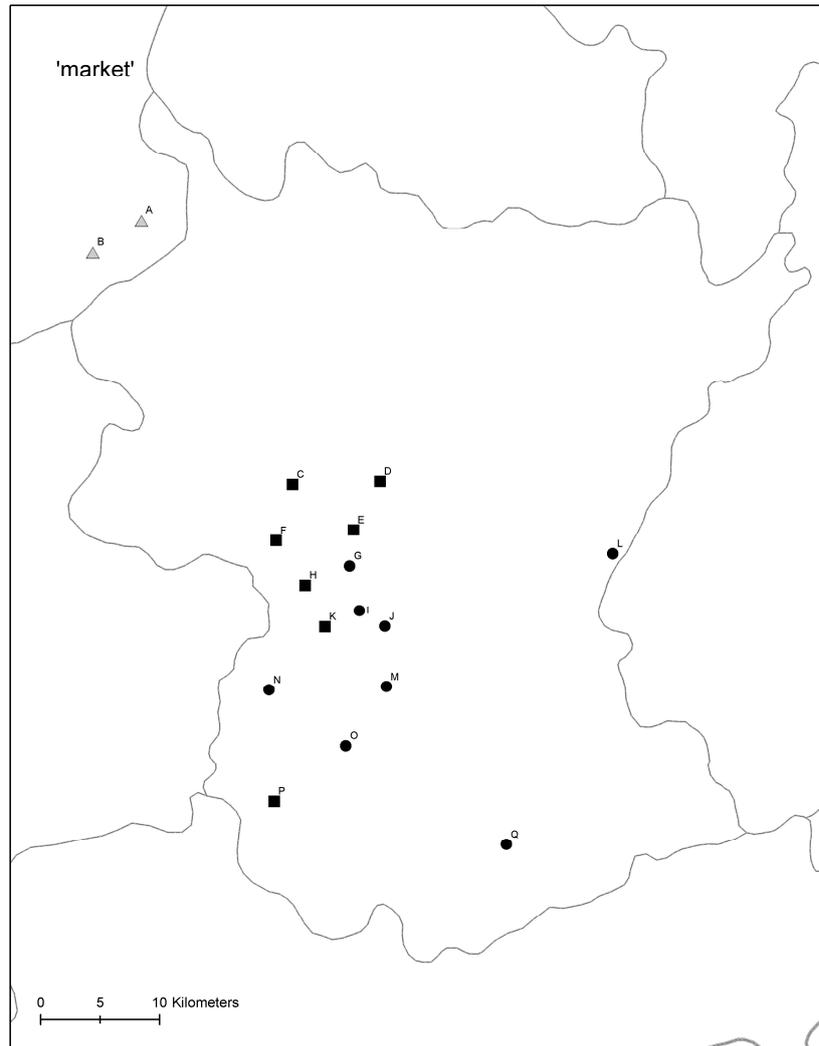
Figure 7: Results for 'market'. ● = $[q\varepsilon^4]$; ■ = $[t\varepsilon\varepsilon^4]$; ▲ = $[k\varepsilon^4]$
(Map source - Evans Map Room, Dartmouth College Library/lmh)

On the other hand, locations H and K are near the central $[\varepsilon i^2]$-speaking region, yet H and K have divergent 1st Singular forms, $[\kappa\varepsilon i^2]$ and $[\mathtt{?}j\varepsilon^2]$ respectively. The difference may be due to terrain. In the terrain map (Figure 9), note that there is a north-south line of low mountains separating K from I and J. This low line of mountains has a steep eastern face, rising over 100 meters vertically in only about 250 meters of east-west distance (Google Earth). Standing near I and looking toward K, this line of mountains is a formidable barrier. It is reasonable to suppose, then, that the lack of 1st Singular $[\varepsilon i^2]$ in locations K and H could be related to these mountains. Location G also appears to be on the west side of the same line of mountains, but it is actually quite accessible from the east side (there is a lower-elevation path).

Other variants are not so easily explained. The 1st Singular contrast between J and M does not appear to be a matter of terrain. Since there are no written or oral historical

records tracing clan settlement at such a level of local detail, it may not be possible to speculate further about the regional distribution of 1st Singular.
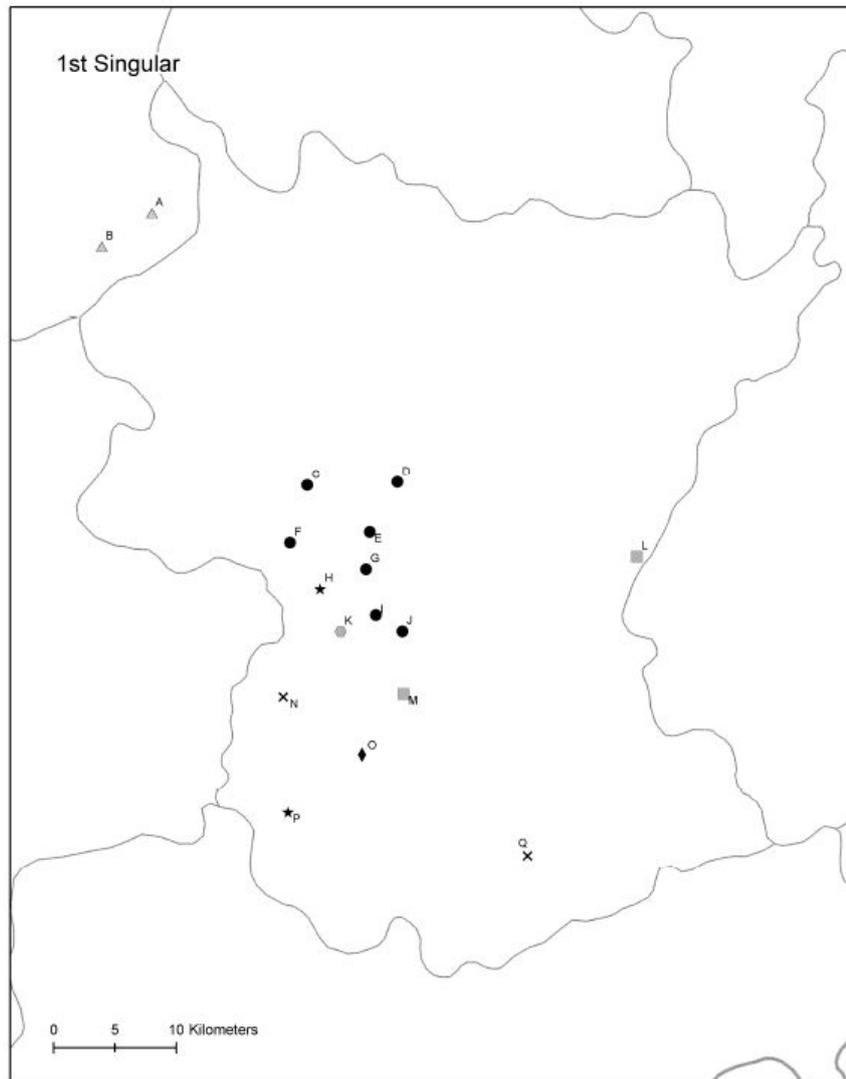
Figure 8: 1st Singular in the current study. ●=[εi²]; ▲=[ai²]; ★=[ʁεi²]; ⬡=[ʔjε²]; ■=[ju²]; ×=[ʔjiu²]; ◆=[i²]
(Map source - Evans Map Room, Dartmouth College Library/lmh)
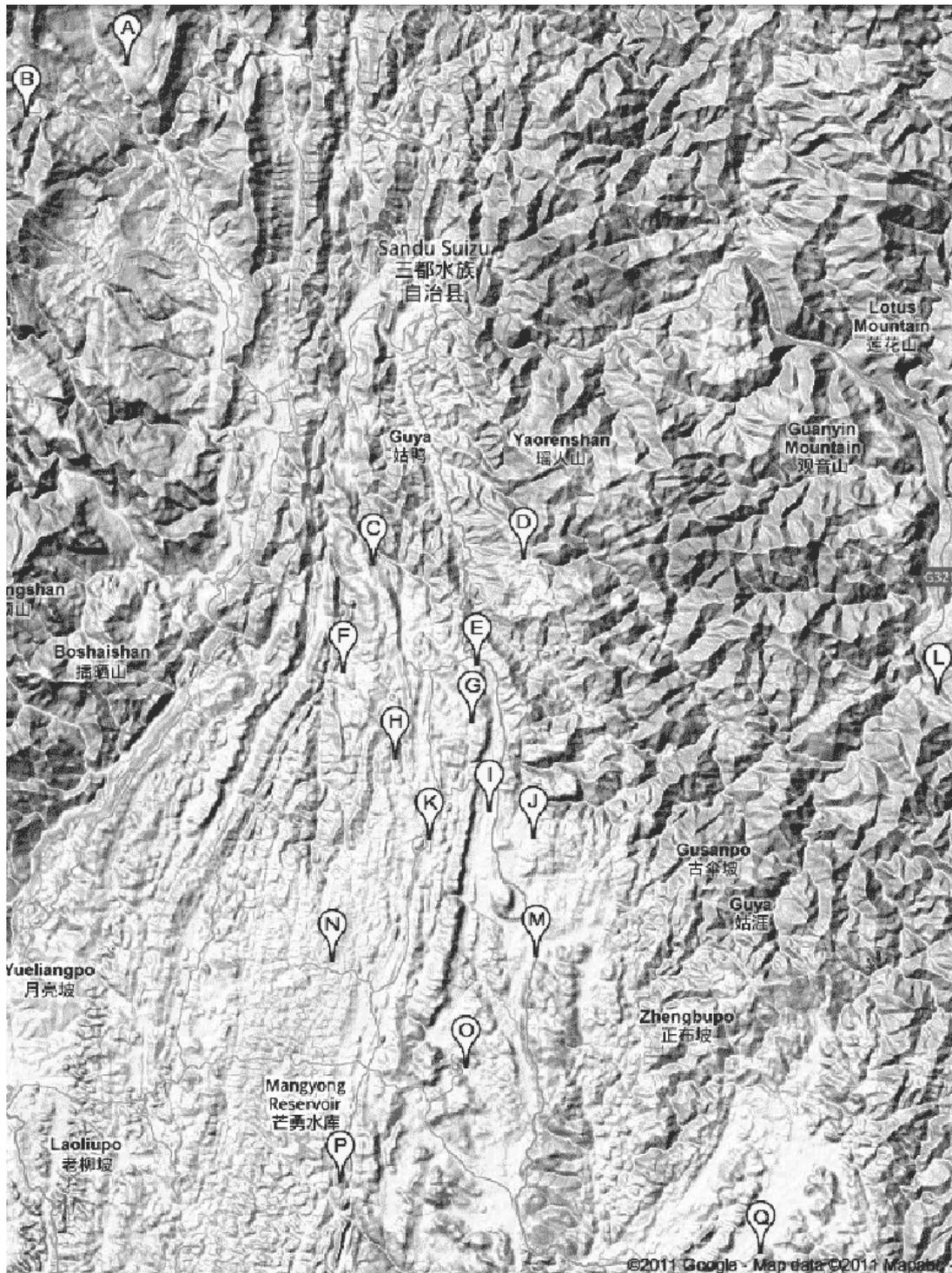
Figure 9: Terrain of the study region. (© Google – Map data © Mapabc)

4.3 Tone Variation

Two of the six Sui tones show regional variation: Tone 6 and Tone 1. Tone 6 is highly salient and easy to distinguish as a categorical variable. As Figure 10 shows, Tone 6 is a high-level 55 in some regions and low-rising 24 in other regions. This variable has been stable for 50 years; SDB shows the same regional pattern. The centralized regional

distribution of the 24 variant suggests that Tone 6 could plausibly be viewed in terms of dialect-leveling or contagious diffusion, much like the [εi²] form of 1st Singular discussed above.

Tone 1 is a far more subtle variable than Tone 6. SDB reports Tone 1 as low-rising in every location, but other studies find a low-falling variant in some locations (Li 1948; Edmondson et al. 2004; Stanford 2008a; Xia 2008:262[1988]). In the present study, results for Tone 1 were inconclusive: The Tone 1 regional pitch contrasts are subtle and vary considerably with speaker style and context, unlike Tone 6. For Tone 6, a single speaker can adequately represent the clear, categorical variant of any given location, but a single speaker is not sufficient to represent the subtle behavior of Tone 1. This result shows the limits of a non-variationist dialectology approach to tone. In fact, SDB appears to have overlooked variation in Tone 1. Using multiple speakers, the variationist analysis in Stanford (2008a) showed a clear difference in Tone 1 between Shuilong and Sandong (locations E and M). An understanding of Tone 1 variation across other regions awaits further variationist work.
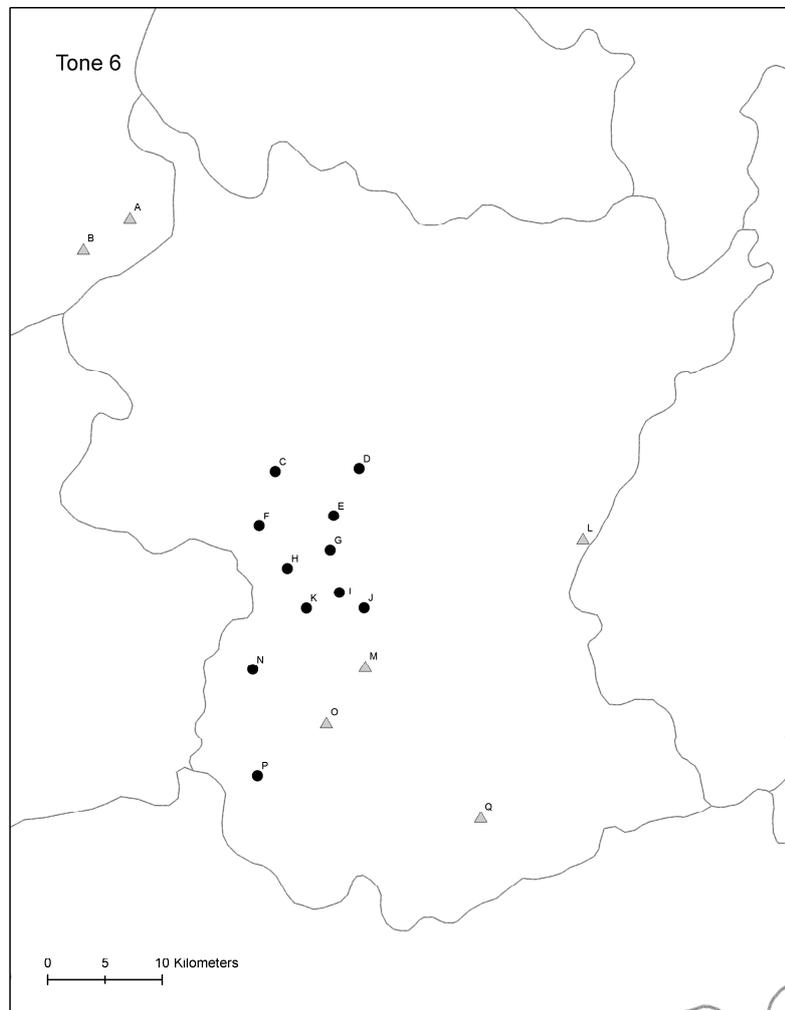
Figure 10: Tone 6. ●=24; ▲=55
(Map source - Evans Map Room, Dartmouth College Library/lmh)

4.4 Change in Progress

A change in progress appears to be occurring in the word 'to do'. SDB reports two variants, [hɛ^4] and [fɛ^4]. The present study finds those variants as well, although the distributions differ slightly. However, the present study also observes a very different new variant [li^4], which is not reported in SDB at all. The new variant appears in locations I, J, M and Q. While [li^4] commonly means 'to build' in all regions investigated, the young and middle-aged speakers of the south are using this word in place of 'to do': for example, *li niman* 'do something' (lit. 'build something'). As in some other languages, the original word for 'to do' can have a coarse, euphemistic interpretation in Sui, so younger Sui speakers in locations I, J, M and Q and perhaps other areas are beginning to substitute it with [li^4]. Since this lexical variant of 'to do' did not appear in SDB, and since speakers now report that it is growing among young and middle-aged speakers, it may be analyzed as a change in progress.

4.5 Summary

Table 5 provides a summary of the 50-year comparison between SDB and the current study. The majority of the variants patterned the same for the two studies. Some minor differences were observed, but they are likely due to fieldworkers' transcriptions in the two studies (cf. Bosch & Scobbie's 2009 discussion of fieldworker isoglosses for Scottish Gaelic).

Table 5: Summary of the 50-year comparison. References for SDB data are given in brackets.

FEATURE	50-YEAR COMPARISON BETWEEN SDB AND THE CURRENT STUDY
Part I: Phonological features	
1. (ia)	Same distribution for SDB and the current study. See Figure 5. [SDB ref: map 1]
2. (ua)	Same distribution for SDB and the current study. See Figure 5. [SDB ref: map 6 and map 17]
3. Preglottalized consonant onsets	Same overall distribution, but location F is not preglottalized in the current study. See Figure 6. [SDB ref: map 2]
4. Voiceless nasal onsets	Same overall distribution, but location F patterns with the voiceless nasal group. See Figure 6. [SDB ref: map 4 and map 7]
5. Tone 6: 55 versus 24	Same results for both studies. See Figure 10. [SDB p. 23]
6. Lateral replacement	Same results for both studies: Location N has a lateral onset in place of fricative onsets in 'to sit' [lui], 'tail' [lat], 'thread' [la:n], and 'sheep' [lɛ]. All other locations have a fricative onset [x-], [h-] or [f-] for these words. [SDB p. 62, 70, 81]
7. [fa-] → [x^{w}a-]	Same results for both studies. The northwest region around locations A and B has [x^{w}a-] in place of [fa-]. [SDB p. 61-62, 75]

Part II: Cognate lexical items	
8. 'sky'	Same overall distribution. Both studies find [ʔban^1] in most regions but other variants in two locations: The northwest region around locations A and B has [van^1] in current study and [ban^1] in SDB. The western region around N has [man^1] in both studies. See Table 3. [SDB ref: map 3]
9. 'diligent'	Similar overall pattern but some differences between the two studies. See Table 4. [SDB ref: map 5]
10. 'arm'	Same, except that Location N in the current study is [k^{h}in] not [kin]. See Table 3. [SDB ref: map 10]
11. 'market'	Same results for both studies. See Figure 7. [SDB p. 85]
12. 'straight'	Similar results: The current study finds [x^{j}aŋ^2] for locations A and B; [xaŋ^2] for N; [jɛn^2] for F; [ɕaŋ^2] elsewhere. SDB is the same except no [jɛn^2] variant is reported. See Table 3. [SDB ref: map 11]
13. 'smile'	Same. All locations have [ku^1] except the western region around Location N, which is [kɔ^1] in both studies. [SDB ref: map 14]
14. 'water buffalo'	In both studies, the northwest region around Locations A and B has [qui^2] for 'water buffalo', while all other regions have [kui^2]. Exception: SDB has [qui^2] in Location N. See Table 3. [SDB p. 67, 83]
Part III: **Non-cognate lexical items** (such that at least one location has a non-cognate variant)	
15. 1st Singular	Same results for both studies in all shared fieldwork locations, although SDB lacks variants corresponding to those found in location O and K. See Figure 8. [SDB pp. 124, 128]
16. 'do'	A new variant [li^4] has emerged in locations I, J, M Q. See discussion above. [SDB ref: map 13. SDB p. 85 also lists an additional variant [vɛ^4] for location N.]
17. 'boat'	Same overall distribution, but locations A and B have a Chinese loanword in the current study. See Figures 3a-b. [SDB ref: map 6]
18. 'salt'	'salt' [ʔdua^1] has a regional vowel variation that is similar to 'boat' (Fig. 3a-b), and it is stable across the two studies. However, the present study also finds 'salt' has the onset [k-] in location Q, and location N has [pa:u^4] for 'salt' (a Chinese loanword). [SDB ref: map 13]

5. Conclusion

The results show consistency across 50 years for the 18 variables studied, strongly suggesting a high degree of long-term stability of Sui dialect distinctions. Even across very short distances of 5 km between locations (recall the scale in Figure 2), the dialect distinctions have been consistently maintained over at least half a century. Minor transcription differences between the two studies are attributed to the lack of acoustic analysis in SDB (1956) and possible effects of fieldworker differences (cf. Bosch & Scobbie 2009). A change in progress was noted in the word 'to do', where young and middle-aged speakers in some locations of Sandu County are replacing this word with the word for 'to build'. An increase in Chinese loanwords was noted in the northwest region (locations A and B), e.g., 'boat', but other regions did not show this type of change.

Besides the results for Sui, this study is also meaningful with respect to other small indigenous Tai-Kadai communities in Southeast Asia. This study shows that the dialect boundaries of such a rural, agrarian community can be stable and distinct over half a century. Unlike large-scale studies of vast geographic areas, many clan-oriented communities like Sui are quite small. Even so, stable dialect boundaries endure. Labov et al. (2006:303) determine that North American English dialect boundaries reflect the "enduring influence of the original regional patterns" of the English-speaking settlers. While there are no Sui records to provide detail about the earliest settlements in Sandu County, this study nonetheless shows the presence of enduring patterns across 50 years. Sui dialect boundaries are lasting evidence of a rich sociolinguistic history with hints of a fascinating future. This type of dialect behavior is likely true of many other small, rural indigenous communities in Southeast Asia.

References

Auer, Peter & Jurgen Erich Schmidt (eds). (2010). *Language and space: An international handbook of linguistic variation: Theories and methods*. Berlin/New York: De Gruyter.

Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. (1993). Some patterns of linguistic diffusion. *Language Variation and Change* 5:359-90.

Bloomfield, Leonard. (1933). *Language*. New York: Henry Holt.

Bosch, Anna & James Scobbie. (2009). Fine-grained morphological variation in Scottish Gaelic: Evidence from the Linguistic Survey of Scotland. In James Stanford & Dennis Preston (eds), *Variation in indigenous minority languages*, 347-368. Amsterdam: Benjamins.

Britain, David. (2002). Space and spatial diffusion. In J.K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds), *The handbook of language variation and change*. Malden, MA/Oxford: Blackwell. 603-37.

Bucholtz, Mary. (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7(3):398-416.

Burusphat, Somsonge, Xuechun Wei & Jerold Edmondson. (2003). *Sui (Shui) Chinese-Thai dictionary*. Bangkok: Mahidol University.

Castro, Andy. (2011). Southern Sui: a fourth Sui dialect. Journal of the Southeast Asian Linguistics Society (JSEALS), 4.2:XX.

Chambers, J. K. (1992). Dialect acquisition. *Language* 68(4):673-705.

Chambers, J.K. & Peter Trudgill. (1998). *Dialectology* (2nd ed). Cambridge: Cambridge University Press.

Chao, Y.R. (1930). A system of tone letters. *La Maitre Phonetique* 45:24-27.

China National Bureau of Statistics, Office of Population and Social Science Statistics and the National Ethnic Affairs Commission, Office of Economic Development [Guojia tongjiju renkouheshehuikejitongjisi, Guojia minzushiwu weiyuanhui jingjifazhansi] (ed). (2003). *2000 nian renkou pucha Zhongguo minzu renkou ziliao, shang ce* [Year 2000 population census data for China's ethnic groups, vol. one]. Beijing: Nationalities

Diller, Anthony, Edmondson, Jerold & Luo, Yongxian. (2008). *The Tai-Kadai languages*. London/New York: Routledge.

Edmondson, Jerold A., Esling, John H., Harris, Jimmy G., & James Wei. (2004). A phonetic study of Sui consonants and tones. *Mon-Khmer Studies* 34:47-66.

Edmondson, Jerold A. & David B. Solnit (eds). (1988). *Comparative Kadai: Linguistic studies beyond Tai.* Dallas, Texas: SIL and the University of Texas at Arlington.

Edmondson, Jerold A. & David B. Solnit (eds). (1997). *Comparative Kadai: The Tai branch*. Dallas: SIL and the University of Texas at Arlington.

Kretzschmar, William A., Jr. (2009). *The linguistics of speech*. Cambridge: Cambridge University Press.

Labov, William. (2008). The mysterious uniformity of the Northern Cities Shift. Presented at *Methods in Dialectology XIII*, University of Leeds, England.

Labov, William, Sharon Ash & Charles Boberg. (2006). *The Atlas of North American English*. Berlin/New York: Mouton de Gruyter.

Li, Fang Kuei. (1948). The distribution of initials and tones in the Sui language. *Language* 24:160-67.

Li, Fang Kuei. (1965). The Tai and Kam-Sui languages. *Lingua* 14:148-79.

Liu, Daichun, Cheng Bangjia & Chen Yingfang. (2002). *Qiannan Buyizu Miaozu Zizhizhou 2000 nian renkou pucha ziliao huibian* [Compilation of the 2000 census results of Qiannan Autonomous Prefecture of Buyi and Miao Nationalities]. Guiyang, China: Guizhou Province 5th Census Office.

Luo, Shiwu (ed). (1992). *Sandu Shuizu Zizhixian zhi* (Sandu Sui Autonomous County journal). Guizhou: Guizhou People's Publishing.

Nerbonne, John. (2009). Data-driven dialectology. *Language and Linguistics Compass* 3(1):175– 198.

Nerbonne, John & Wilbert Heeringa. (2010). Measuring dialect differences. In Peter Auer & Jurgen Erich Schmidt (eds), *Language and space: An international handbook of linguistic variation: Theories and methods.* Berlin/New York: De Gruyter 550-67.

Shuiyu Diaocha Baogao (SDB) [Sui Language Research Report]. (1956). Unpublished ms. Guiyang, Guizhou: Zhongguo Shaoshu Minzu Yuyan Diaocha Diyi Gongzuodui [Chinese Minority Language Research Working Group]. Acquired from Jerold Edmondson, University of Texas-Arlington.

Smith, Ian, & Steve Johnson. (1986). Sociolinguistic patterns in an unstratified society: The patrilects of Nganhcara. *Journal of the Atlantic Provinces Linguistic Association* 8:29-43.

Stanford, James N. (2007a). Sui adjective reduplication as poetic morpho-phonology. *Journal of East Asian Linguistics* 16(2):87-111.

Stanford, James N. (2007b). Lexicon and description of Sui adjective intensifiers. *Linguistic Discovery* 5(1):1-27.

Stanford, James N. (2007c). Dialect contact and identity: A case study of exogamous Sui clans. PhD dissertation, Michigan State University.

Stanford, James N. (2008a). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change* 20(3):409-450.

Stanford, James N. (2008b). Child dialect acquisition: New perspectives on parent/peer influence. *Journal of Sociolinguistics* 12(5):567-96.

Stanford, James N. (2009). "Eating the food of our place": Sociolinguistic loyalties in multidialectal Sui villages. *Language in Society* 38(3):287-309.

Stanford, James & Dennis Preston (eds). (2009). *Variation in indigenous minority languages*. Amsterdam/Philadelphia: Benjamins.

Trudgill, Peter. (1986). *Dialects in contact*. Oxford/New York: Basil Blackwell.

Wei, Shifang. (2007). *Shuishu changyong zidian* [Dictionary of common words in Shui writing]. Guiyang, China: Guizhou Nationalities Publishing.

Wei, Xuechun & Jerold Edmondson. (2008). Sui. In Anthony Diller, Jerold Edmondson & Yongxian Luo (eds.), *The Tai-Kadai languages*. London/New York: Routledge. 585-95.

Xia, Yongliang. (2008[1988]). Duyunshi Yanghexiang Pandong Shuiyu diaocha [Duyun city Yanghe village Pandong Sui language investigation]. In Liang Yongqu (ed.), *Guizhou liushanliushui minzu diaocha ziliao xuanbian [Selections from Guizhou nationalities investigations.]* Guiyang, China: Guizhou Nationalities Publishing. 254-65.

Zeng, Xiaoyu & Fuxiang Yao. (1996). *Han-Shui cidian (Chinese-Sui dictionary)*. Chengdu: Sichuan Minorities Publishing.

Zhang, Junru. (1980). *Shuiyu jianzhi (A sketch of the Sui language)*. Beijing: Minorities Press.

Zhou, Minglang. (2003). *Multilingualism in China: The politics of writing reforms for minority languages 1949-2002*. Berlin/New York: Mouton de Gruyter.

CITATION FOR MAPS: Evans Map Room, Baker/Berry Library, Dartmouth College / lmh. Data sources: DIVA-GIS, James Stanford, Linguistics Department, Dartmouth College.

PROTO KHASIAN AND KHASI-PALAUNGIC

Paul Sidwell
Australian National University (Canberra) &
Centre for Research in Computational Linguistics (Bangkok)
<paulsidwell@yahoo.com>

Abstract

The present paper sketches a reconstruction of proto-Khasian, the immediate ancestor to the Khasi, Lyngngam, Pnar and War languages of Meghalaya, India. Furthermore, it is argued that the Khasian branch of Austroasiatic (AA) is best characterised as a sister of Palaungic, and thus probably reflects the historical migration of a single community out of the upper Salween-Mekong hills area and into the Brahmaputra valley before the arrival of present-day Tai and Tibeto-Burman groups. The paper also addresses various counter views offered by Daladier in her companion paper in this issue arguing that diversification within Khasian, since its arrival in the Brahmaputra region, is sufficient to account for the apparent diversity within the group, and cannot be held to invalidate the hypothesis of the genetic unity of Khasian.

Key words: Khasi language, proto-language, classification

Introduction

The Mon-Khmer Languages Project (MKLP),[79] which has been actively in train since 2007, is contributing to our understanding of the history and development of the AA languages in various ways, not the least of which is the progressive reconstruction of branch level proto-languages. This paper specifically reports on progress in the reconstruction of proto-Khasian by the project, and new results and lines of enquiry that arise from that reconstruction. Importantly, since it is also a vital part of the MKLP's role, that effort has stimulated scholarly debate over the importance and nature of reconstruction. In particular, Anne Daladier, a scholar with substantial first hand experience with Khasian languages, has challenged the idea of proto-Khasian, suggesting that the group is the outcome of linguistic convergence among a number of diverse AA groups that independently settled on the Meghalaya plateau. Daladier's approach places importance on identifying lexical and structural differences among Khasian languages, and attempts to reconcile linguistic history with the understandings that various Khasian peoples have concerning their own origins.

This disagreement became apparent in connection with Sidwell presenting details of his proto-Khasian reconstruction at the 21st SEALS meeting at Kasetsart University, Bangkok, in May 2011. Out of the discussions that Sidwell and Daladier had arising from this, it was decided that each would present their views in this edition of JSEALS, and this paper constitutes Sidwell's contribution. Given that the root of the differences lays in

---

[79] The MKLP was funded from 2007 to 2011 by the National Endowment for the Humanities (Washington). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily represent those of the National Endowment for the Humanities

theoretical and methodological approaches, the first part of this paper provides a brief overview of this writer's understanding of the theoretical issues at stake and how they relate to the Khasian question.

Historical spread of Austroasiatic
The MKLP recognizes, consistent with the broad scholarly consensus represented in more than half a century of literature (essentially since Thomas & Headley 1970) that the AA languages form a dozen or so branch-level groupings. Each of these branches reflects a genealogical unity; descendents of what are modelled as a dozen proto-languages, each defined by linguistic innovations that emerged as they separated linguistically; this was the process of the break-up of proto-AA.

These languages themselves diversified, creating the diversity we observe within branches that we see today. These vary from complex linguistic communities with dozens of languages spread over mountainous territories (e.g. Bahnaric, Palaungic and others) to simple (internally levelled) branches (e.g. Khmer, Mon). This is a crucial point: the great age of AA, at least 4000 years since distinct branches began to form (see Diffloth 2005, Sidwell & Blench 2011 for discussion) means that, unless other factors apply, individual branches are inevitably internally diverse. The apparent exceptions such as Khmer and Mon, which are each dominated by a single language (or dialect chain), have well understood explanations: each was the language of a highly organized state (Angkor, Dvaravati) that imposed linguistic uniformity by public administration and cultural dominance, effectively undoing the natural effect of time on diversification. There is no evidence of other AA groups having such a level of cultural development in classical times.

There are other small AA branches, such as Nicobarese, Pearic and Mangic, all with very small populations. The first are a refuge community isolated on small islands, unable to support large scale growth and diversification, and in effect self-levelling; as (reported by de Roepstorff 1875) Nicobarese speakers were conversant in all dialects of the island chain (except Shompen). This is quite understandable when population size has never been more than a few thousands. The other groups may have had larger populations and ranges, since they are now discontinuous internally, but are today reduced to rump communities by language and culture shift, a result of the levelling effects of the success of Khmer and Chinese respectively.

Otherwise, the great diversity in AA languages is among the upland dwelling communities that are emblematic of AA culture. It can be suggested that the adoption of swidden rice farming (an extension of originally cultivating wet rice on river flats and later dyked fields) allowed expansion into and along mountainous regions which were not cursed with malarial mosquitoes and other barriers to the growth of small-scale societies. Far from being refugia, the mountains of SEAsia provided ecologically rich and mercifully temperate habitable zones (Yunnan in particular is known is one of the world's great ecological "hot spots", and the source of many domesticated species). Thus, AA speakers were able to spread along the Annamite Chain and the uplands along the upper Mekong, Salween and Brahmaputra valleys, and even to the hilly areas of Eastern India.

Certainly AA speakers did come to inhabit and even dominate lowlands, but all linguistic, archaeological and historical evidence indicates that in SEAsia this was essentially a First Millennium process associated with the growth of larger scale societies and their emerging agricultural and hydrological economies. As for the Brahmaputra

region, we have very little to inform us of the situation prior to the Second Millennium. Today the upper Brahmaputra is a patchwork of largely Tibeto-Burman and Tai speaking communities while Khasian speakers occupy the neighbouring Meghalaya hills. The present stage of knowledge concerning how this came about is summed up nicely by Post:

> The Brahmaputra Valley is an area about whose history a little more can be said: Prior to 1000 BC, it is difficult to conjecture about the cultural-linguistic composition of the area, although there is at least a possibility of Austroasiatic predominance (Kakati 1995; Diffloth 2005). From 1000 BC to 400 AD we find the South-westward spread of Bodo-Garo, most likely from an initial position in the Northern Burmese/North-East Indian hill regions, where "Sal" languages such as Tangsa are spoken in great variety to this day (DeLancey 2012). From 400 AD to the present, we find the North-eastward spread of the Eastern Indo-Aryan languages Bengali and Assamese (Baruah 1960 [1933]). From 1200 AD, we find the arrival and subsequent decline of Ahom (Tai) from the Northern Burmese Shan states, plus smallcommunities of later Tai arrivals from the same area such as Khamti, Aiton, Phake and Khamyang (Morey 2005).
> Post (2011: 216-17)

Assuming that folklore/folkhistory is of secondary value in objectively understanding historical events hundreds or thousands of years past, the most parsimonious interpretation of the geographical, linguistic and historical facts in our possession is that Khasian speakers arrived in the region of the Meghalaya hills as speakers of a single proto-language or the close equivalent of a chain of closely related dialects, who subsequently diversified over time to form the branch-level grouping of languages recognised today.

In the text that follows will see that all the Khasian languages which have been treated in this study show common innovations, lexical and phonological, consistent with diversification from a proto-Khasian mother tongue. This contrasts with the approach of Daladier (this issue) which highlights differences between Khasian languages as indicative of descent from distinct, conservative AA languages that have come to resemble each other as a result of close proximity (more specific remarks on Daladier's thesis towards the end of this paper).

Comparative reconstruction and language contact

The approach taken is a contemporary take on the neo-grammarian "branching with modification" paradigm, which has developed over two centuries of scholarship, and enjoyed tremendous success in its adoption by evolutionary biology in the 19th Century. To this day it remains a robust paradigm, complemented with extensively well developed methodological resources, collectively known as the "Comparative Method".

The neo-grammarian approach has been challenged, and this can be characterized as pointing out the lack of a perfect parallel between linguistic and biological evolution. Essentially, the equivalent of horizontal gene transfer, which plays a real but marginal role in the evolutionary biology of multi-cellular organisms, does play a substantial role in linguistic evolution, since there are no ultimate limitations on the borrowing of linguistic structures. Proto-languages may diverge, but remain in contact over a long period of time, with sound changes, and other innovations radiating - wave-like - in conflicting and/or overlapping patterns that cannot be represented in a simple branching tree model (see, for example, Heggarty, Maguire & McMahon 2010 for a recent discussion and further references). However, these facts do not negate the reality of proto-languages, language families, or the capacity to reconstruct proto-languages. Patterns of correspondences

emerging from such dialect chains (or "linkages", see Ross 1988 and passim) may still show sufficient regularity to allow reconstruction with high reliability, and this is indeed the experience of the highly successful *Oceanic Lexicon Project* at the Australian National University.

Recent decades have seen substantial theoretical progress achieved in relation to modelling linguistic prehistory. Particular attention is drawn to, for example, Nichols (1997 and passim.) empirically demonstrating a global tendency for what Nichols calls "spread zones" versus "accretion zones". The salient point being that local areas of high diversity, or "accretion zones" form by several mechanisms. They can form by the diversification of one or more phylogenetic grouping, and/or by the accretion of various related or unrelated groups arriving in what we might call a refuge area.

We can consider the possibility of contact conditioned convergence within a refuge area being so extensive that it becomes impossible to distinguish between a phylogenetic unity and a language area. Nichols explains:

> Languages long in contact can retain their discrete identities but come to resemble each other in sound structure, lexicon, and/or grammar.
> (Nichols 1997:367)

However, in the present case of the AA languages of the Meghalaya hills, it is rather unlikely that such contact is the primary explanation for their numerous similarities. The languages are lexically and structurally very close by AA standards: the Khasian lexicostatistics reported by Sidwell (2009) and Daladier (this issue) show that the group is are comparable to branches such as Katuic, with the lowest cognate scores in the low 40s percentage, 20 points above the cognacy counted for the next closest AA branch (in this Palaungic). So lexically it is not just a normal AA branch, but is more coherent than many. Additionally, Khasian shows specific affinities to the Palaungic branch (discussed later below) suggesting that these two groups separated rather later than the general dispersal of AA branches. Considering these facts, it is evident that hardly more than a couple thousand years has passed since the Khasian settlement of the Meghalaya hills. On the other hand, an explanation for diversity among Khasian languages that relies upon a convergence model must necessarily show evidence consistent with great antiquity and continuity of the languages, while similarly reconciling this with a coherent model of AA branching - the approach must have general applicability.

It is theoretically possible that a group of languages could share substantial basic vocabulary due to borrowing, and show consistently regular correspondences because the forms had changed very little, and/or had completely replaced the lexicon in the recipient languages. But in such a case we would have to ask what such a data set actually represents. This problem arose in Altaic studies, with Doerfer (1963) asserting that Mongolian and Tunguisic had borrowed their basic vocabularies from proto-Turkic, but this still becomes a claim that the languages none-the-less descend from a reconstructable common source.

And if we assert that in a particular case the basic vocabulary is largely replaced by contact, we should also find similar extensive borrowing throughout other sub-systems of the language(s). Thus, where languages show important differences in, for example, cultural vocabulary, phonology, morphosyntax, etc. but largely share basic vocabulary, within which are found productive-predictive correspondences, a convergence explanation

is less likley, and common ancestry can be reconstructed on the basis of the lexical evidence.

Reconstruction of proto-Austroasiatic

Concretely, we can straightforwardly demonstrate the common ancestry of the AA phylum by invoking lexical comparisons such as the following:

gloss	Mundari	Khasi (orth.)	Palaung	Khmu'	Viet. (orth.)	Katu	Semelai	Nancowry
'nose'	*mũ*	*khmut*	*muh*	*muh*	*mũi*	*muh*	*muh*	*moəh*
'thigh'	*bulu*	--	*blu*	*bluʔ*	--	*malaːw*	*blu*	*pulɔː*
'eye'	--	*khmat*	--	*mat*	*mắt*	*mat*	*mɔt*	*mət*
'day'	*siŋgi*	*sngi*	*səɲi*	*sɲiʔ*	*ngày*	*taŋaːj*	*ʈɲi*	--

And hundreds of similar comparisons are readily extracted from reference works such as Shorto (2006), and the regularity of the sound correspondences are already established and do not need to be demonstrated here.

As we discussed at the end of the preceding section, it is this kind of lexical agreement that demonstrates the unity AA, regardless of any amount of diversification, restructuring, or other innovation that has taken place among any of the daughter languages. Thus, despite claims by sceptical writers such as Maspero (1912), Sebeok (1942) and others, who sought to deny the unity of AA by citing various differences between the languages, no argument of that kind invalidates a claim on genetic unity in the face of such agreements in vocabulary.

Several thousands of lexical comparisons, such as compiled by Shorto (2006) and others, permit us to establish regular phonological correspondences, both in terms of phonemes and phonotactics, such that we can propose a reconstruction of pAA phonemes as follows:

Prevocalic Consonants	*p*	*t*	*c*	*k*	*ʔ*
	b	*d*	*j*	*g*	
	ɓ	*ɗ*			
	m	*n*	*ɲ*	*ŋ*	
	w	*r, l*	*s, j*		*h*

Vowels (long):	*iː*			*uː*
		eː	*əː*	*oː*
		ɛː	*aː*	*ɔː*
		iə		*uə*

Postvocalic Consonants	*p*	*t*	*c*	*k*	*ʔ*
	m	*n*	*ɲ*	*ŋ*	
	w	*r, l*	*s, j*		*h*

Vowels (short):	*i*			*u*
		e	*ə*	*o*
		ɛ	*a*	*ɔ*

With a basic word template: (C)CVC, without tones, or open syllables.

This represents a minor modification of the scheme offered by Shorto (2006) in that the low front vowels /ɛ, ɛː/ are added and Shorto's /ai, ɯə/ diphthongs are removed. The proto-Khasian system is readily derived from this scheme without any unrealistic or typologically odd processes needing to be reconstructed.

Proto-Khasian

I have now made available online a preliminary proto-Khasian reconstruction (sealang.net/ monkhmer). The reconstruction is characterized as preliminary because it is based upon rather sparse data, yielding only some 800+ etymologies and proto-forms. I have considerable confidence that the reconstruction is broadly reliable for 2 reasons:

- Some 600 of the etyma have deeper AA etymologies based on the data of Shorto (2006), and;

- Data for all four coordinate sub-branches of Khasian are used.

Specifically the sources utilized are:

- Standard Khasi as represented in published dictionaries

- Lyngngam (Nagaraja 1996, Daladier this issue)

- War (Daladier this issue), Amwi (Weidert 1975)

- Pnar (Bareh 2010, Choudhary2004, Daladier this issue)

And additional insights were obtained by reference to the survey of Pnar (Jaintia) dialects spoken in Bangladesh compiled by SIL affiliated researchers (Brightbill et al. 2007), and various other related publications that are widely available. I am quite willing to acknowledge that these sources are pitiable compared to what ought to be applied to the task, but I take the attitude that we should not hesitate to start using what is readily at hand, and then build on the results as new data is available, and welcome any scholarly discussion that said work generates by way of response.

It is crucial for comparative reconstruction that data is taken into account that represents the genetic diversity within the group being investigated. This often takes the approach of identifying criterion languages, that is, languages that trace back to more than one coordinate branch, so that comparisons using these sources can be reasonably projected back to the proto-languages. This necessarily requires a model of the internal classification of the group, and although various sources have been vague or even contradictory in respect of Khasian, the scale of the problem is not large. Generally, sources have reported four main linguistic sub-divisions within the group: Khasi, Lyngngam, Pnar, War. In Sidwell (2009) I analysed both lexicostatistical and phonological data to suggest that there are two basic sub-branches, essentially Lyngngam-Khasi-Pnar and War-Amwi. The latter is particularly distinguished by distinctive changes in vocalism that includes raising and fronting of **a** in many environments, and related restructuring of diphthongs.

My 2009 scheme effectively assumed that Lyngngam-Khasi-Pnar constituted a group which does not share the characteristic War-Amwi phonological innovations. But common failure to share an innovation is not a reason to sub-group, and consequently I revisited the question and reported on my results to the May 2011 SEALS meeting in Bangkok. More careful examination of the data at hand reveals that, at least from a phonological viewpoint, Standard Khasi, Lyngngam, and Pnar are readily derived directly from proto-Khasian by only a modest number of language specific changes, suggesting that the most parsimonious

explanation is to posit the four approximately equidistant sub-branches, consistent with those already commonly reported in the literature.

Broadly speaking, I find that Pnar is somewhat phonologically closer to the proto-language, while Standard Khasi and Lyngngam each show various specific innovations, while War languages are dramatically more innovative. I hesitate to make strong claims about Lyngngam at this stage because I have only a few hundred lexical items to consider, and these may be contaminated with Standard Khasi to an extent that I cannot yet judge.

Comparative Phonology

Comparative phonology begins with characterizing the systems of the languages being compared. The phoneme inventories reported in the literature include the following (note: loan phonemes in bracketed):

Lyngngam (Nagaraja 1996):

/	p^h	t^h	c^h	k^h			i, iː	ɨ	u	
	p	t	c	k	ʔ		e	ə	o	
	b	d	ɟ	g				a, aː		
	m	n	ɲ	ŋ						
	w	r, l	j							
		s			h					/

Standard Khasi (Rabel 1961):

/	p^h	t^h		k^h			i, iː	ɨ	u, uː	
	p	t		k	ʔ			ə		
	b	d	ɟ				ɛ, ɛː	a, aː	ɔ, ɔː	
	m	n	ɲ	ŋ			ie, ia		uo	
	w	r, l	j							
	(f)	s	ʃ		h					/

Pnar (Bareh 2010)

/	p^h	t^h	c^h	k^h			i	ɨ	u	
	p	t	c	k	ʔ		e		o	
	b	d	ɟ				ɛ		ɔ	
	($b^ɦ$	$d^ɦ$	$ɟ^ɦ$)				ia	a		
	m	n	ɲ	ŋ						
	w	r, l	j							
		s			h					/

Amwi (Weidert 1975)

/	p^h	t^h	c^h	k^h			i		u	
	p	t	c	k	ʔ		e	ə	o	
	b	d	ɟ				ɛ	a	ɔ	
	m	n	ɲ	ŋ			ia		ua	
	w	r, l	j							
		s	ʃ		h					/

Note: Daladier advises that Amwi shows an additional palatal affricate [tʒ] which was not distinguished by Weidert.

The phonotactic template is characteristically generally:
$(C_{-1}V_{-1})C_1(C_2)V_1(C_3)$,

In which:

- the sequence $C_1(C_2)$ includes a diverse range of clusters, including falling sonority sequences (e.g. Standard Khasi examples from Henderson 1989-90: *bti* 'to lead', *btʰi* 'sticky', *dkar* 'tortoise' etc.),
- the presyllable vowel V_{-1}, lacking phonemic value, is realized as *i* or *ə* only,
- word final C_3 shows no contrast in voicing (although symbols *b, d, g* maybe used in written Khasi), and
- vowel length is only distinctive in the a subset of the languages (such as Standard Khasi, and even then it is distributed rather asymmetrically in the system).

Obstruents

Henderson (1976, 1989-90) provides crucial commentary on the interpretation and significance of Standard Khasi phonotactics and initial clusters in particular. She notes that there is a strong tendency to avoid homorganic clusters and, "In fact, there seems in Khasi to be a deliberate dissimilation of voicing in initial clusters, especially when the cluster is of two stops" (1989-90:62). Other Khasian varieties do not seem to be bound by this tendency, as we can see in such forms as Pnar and Amwi /tput/ 'revenge', Amwi /ktiaŋ/ 'afraid' and others. I interpret this as indicating that dissimilation of voice is a peculiarity of Standard Khasi and need not be reconstructed for the proto-language.

Also, Henderson makes an important point about the distribution of velar stops. Despite the tendency for voicing dissimilation, there are sequences such as *kti* 'hand', *kpa* 'father' but none with /g/. This correlates with the lack of a /g/ generally in Standard Khasi, and it is also absent in Pnar and War varieties. A voiced velar /g/ is reported for Lyngngam, but a direct reflex of this segment is not found in corresponding etyma in other Khasian languages. It appears to occur as a result unpacking of nasals, and as prefix of recent origin. Consider the following comparisons:

Gloss	Lyngngam	S.Khasi	Pnar	War-J.[80]	War-KN	Amwi
'to hear'	*sŋgu*	*sŋew*	*sɲiaw*	*snao*	--	--
'to stop'	*sŋgeʔ*	*sŋeʔ*	--	--	*(səŋit)*	--
'water'	*gum*	*ʔum*	*ʔum*	*ʔum*	*ʔam*	*ʔam*
'that'	*gteʔ*	--	*taj*	*tai*	--	*tə*

The Lyngngam form with initial velar is isolated; there is no etymological basis for positing an initial velar so it looks like a fused prefix, or is perhaps a sandhi form. These considerations allow us to suggest that there was no ***g** in proto-Khasian, consistent with a general devoicing of stops, and the fact that no velar implosive is reconstructable for pAA. In fact, this was suggested by Haudricourt some four decades ago:

---

[80] War-J is the Jaintiapur dialect spoken in Bangladesh, from Brightbill et al. (2007). War KN is Daladier's Kudeng Nongtalang dialect, apparently quite similar to Amwi. Otherwise examples cited are from the standard sources listed above.

Enfin une troisième langue présente les mêmes correspondances, le Khasi actuellement parlée dans l'Inde, dans le massif montagneux au sud de l'Assam et probablement originaire de Birmanie.

khasi		langues conservatrices
vous	*phi*	*pe* (Boloven)
tisser	*thaaiñ*	*taañ* (Köho, Bahnar)
crabe de terre	*thaam*	*taam* (Köho), *ktaam* (Bahnar)
poisson	*khaa*	*kaa* (Köho, Bahnar)
enfant	*khuun*	*kɔɔn* (Köho, Bahnar), *kuon* (Boloven)
flèche	*khnam*	*kam* (Köho, T'eng)
tigre	*khlaa*	*klaa* (Bahnar), *kluə* (Boloven)

Malheureusement, on ne trouve pas d'exemple probant de la correspondance attendue des occlusives sourdes non-aspirées khasi avec les occlusives sonores des autres langues. Mais on trouve trace d'une troisième série d'occlusives représentées actuellement par des sonores. Le fait que cette série manque du g (qui ne se trouve que dans les emprunts récents), suggére qu'il s'agit d'une ancienne série préglottalisée ; il y a un bon exemple : le nom du paddy kba, qui est en Mon et en Bahnar 'ba avec un 'b préglottalisé. (Haudricourt 1965:164)

The general claim is that within Khasian there was a restructuring of oral stops (occupying the C_1 position), such that Austroasiatic voiceless stops became aspirated, and implosives became plain voiced stops. Haudricourt could not find unambiguous examples indicating the fate of Austroasiatic plain voiced stops, but these were relatively infrequent in proto-Austroasiatic, so it is not so surprising, and on typological grounds we would expect a merger of implosives and voiced stops much as has happened in other Austroasiatic branches. However, looking among the etymologies compiled by Shorto (2006), we can find various examples of Khasi voiceless stops from original plain voiced stops, such as:

570 *dɔŋ *to pull:* Khmer **tùəŋ** *to pull back and forth*, Stieng **dɔŋ** *to pull, push*, Stieng dialect **dɔŋ** *to pull down*, Khasi ***tong*** *to draw [water &c.]*; ~ Kuy **nthàŋ** *to drag, pull.*

1357 *dəm *to lodge for the night, to roost.* Old Mon ***dum*** */døm/*, Modern Mon **tə̀m** *to lodge for the night*, Khmer **tùm** *daṁ* *to perch, to sit*, Jeh **dàm**, Halang **dɒm** *to sleep away from home*, Khmu' **dɯm** *to stay overnight*, Palaung **dəm** *to lodge*, Praok **tum** *to lodge, alight*, Khasi ***dem*** (!) *to alight, to stoop, to lie down.*

Theoretically such restructuring should have resulted in the general loss of plain voiceless stops from the C_1 position. Yet they do occur, somewhat infrequently among AA etyma which are indicative of original plain voiceless stops, and thus in apparent violation of Haudricourt's generalization. Consider:

Gloss	Lyngngam	S.Khasi	Pnar	Amwi	Shorto PMK
'mushroom'	--	*tit*	*tit*	*tet*	*1903 *ptis 'fungus'*
'to blow'	--	*put*	*put*	--	*1023 *puːt 'to blow'*
'to bite'	*kap*	*kap*	--	--	*1231 *kap 'to bite'*

Examples such as those above show that it is difficult to argue that any phonological conditioning might be involved, so we have a small problem - an apparent violation of the neogrammarian principle, in which a sound change has failed to apply to a portion of the vocabulary. If we did not have external comparisons to inform our analysis, we might

suggest that these represent vestiges of the Austroasiatic plain voiced series, but in this case we compelling evidence that they reflect voiceless stops that we would otherwise expect to see reflected as aspirates.

This is not a serious theoretical problem; it is widely recognized that sound changes do not always propagate fully over a given speech community, or through the entire lexicon that are eligible to reflect a given sound change. For example, the study of the Great English Vowel Shift by Ogura showed that the, "… processes of the development of ME *iː* and *uː* have propagated themselves gradually from morpheme to morpheme." (Ogura 1987:45) Absent any specific evidence for another explanation, it may simply be that the 'exceptions' listed above result from a failure of the Khasian aspiration shift to apply to all eligible forms before it ceased to be productive. Given the modest extent of the phenomenon, and the indicative nature of the external comparisons, my approach is to straightforwardly treat these as plain voiceless stops in proto-Khasian.

Taken together then, the above changes in labial, apical and velar stops can be summarized as follows:

pAA >   pKhasian
*ɓ-/*b-, * ɗ-/*d-,   *b-, *d-
*p, *t-, *k-/*g-   *pʰ-, *tʰ, *kʰ- (and occasionally *p-, t-, k-*)

More complex are the developments in respect of palatals and oral fricatives, which we now review.

Proto-Austroasiatic *ɟ, both prevocalic and preconsonantal, is preserved as a voiced stop or affricate in Lyngngam, Khasi and Pnar (the sources tend to record the phoneme as /ɟ/ and the phone as /ʤ/) tending to devoice in War languages. Examples:

Gloss	Lyngngam	S.Khasi	Pnar	War-J.	War-KN	Amwi	Shorto PMK
'sour'	ɟu	brɟɛw	ɟaw	ʤao	--	--	*ɟuʔ
'soft'	--	ɟɛm	--	--	--	cem	*[k]ɟəm
'long,tall'	ɟiroŋ	ɟrɔŋ	ɟrɔŋ	ʤaroŋ	(kərɔŋ)	(kərɔŋ)	*ɟruŋ
'rice'	əɟa	ɟa	ɟa	ʤa	tʒi	ci	--
'cold'	--	--	kɟam		ktʒam	kcam	--

The development of pAA prevocalic and preconsonantal ***s-** and ***c-** into pKhasian patterns into three correspondence sets. The first indicates that pAA ***s-** continues as [s], occasionally palatalised to [ʃ] in Standard Khasi by assimilation. Examples:

Gloss	Lyngngam	S.Khasi	Pnar	War-J.	War-KN	Amwi	Shorto PMK
'hair'	sɲjək	ʃɲjuʔ	sɲoʔ	snuʔ	suʔ	suʔ	*suk
'five'*	san	san	san	san	san	san	*sən
'fruit'	--	sɔʔ	sɔʔ	soʔ	--	sɔʔ	*s[u]k
'leaf'	slaː	sla	sla	sla	sli	sli	*slaʔ
'blood'*	snam	snam	snam	snam	rnə[81]	--	*snaːm

---

[81] The War form with initial rhotic and no final nasal is quite likely to be cognate: rhotacization of s > r occurs sporadically in War, as does loss of final nasals. Daladier (this issue) provides further examples of both.

At the same time pAA ***c-** appears to have paralleled the other voiceless stops, developing into an aspirate /cʰ/ in proto-Khasian, with various reflexes in the modern languages: [ʃ] in Standard Khasi (orthographic *sh*), [cʰ ~ ʧ] in Pnar (the variation appears to be merely notational), [cʰ ~ ʧ ~ ʃ] in War/Amwi ([ʃ] apparently before [i, j]), and [c~ ʧ] in Lyngngam (where [c] is Nagaraja's phonemic notation). My interpretation is that the notation *c, cʰ, ʧ* in these various sources is reflecting an aspirated palato-alveolar [ʧʰ], reflecting proto-Khasian ***cʰ**. Examples:

Gloss	Lyngngam	S.Khasi	Pnar-R	Pnar-J.	War-KN	Amwi	Shorto PMK
'bone'	ʧʔiaŋ	ʃʔieŋ	ʧʔiŋ	--	ʃʔiaŋ	--	*cʔaːŋ
'to pinch'	--	ʃʔit	--	--	--	--	*cʔiːt
'sharpen'	--	ʃut		cʰut	--	--	*suːt (*cuːt[82])
'to sit'	maʧoŋ	ʃoŋ	ʧoŋ	cʰɔŋ	--	-	--
'village'	ɟnoŋ	ʃnɔŋ	ʧnoŋ	cʰnɔŋ	ʧnoŋ	cʰnɔŋ	--

(Pnar-R = Ralliang dialect, Daladier (this issue); Pnar-J = Jaintia dialect, Brightbill et al. 2007)

However, examples such as those above are few, and occasionally problematic: for example the 'village' word in Lyngngam is recorded with a voiced palatal by Daladier, but with a voiceless [c] by Nagaraja, yet is is clearly an infixed reflex of the root for 'to sit' so histotically is must be *cʰnɔŋ.

Additionally there are examples of pAA ***c-** reflected as [s]:

Gloss	Lyngngam	S.Khasi	Pnar-R	Pnar-J.	War-KN	Amwi	Shorto PMK
'bird'	ʔəsim	sim	sim	sim	ksɛm	ksem	*cim
'grandchild'	--	ksiew	--	--	--	hnsəw	*cuːʔ
'dog'	ksu	ksɛw	ksaːu	ksaw	kseːa	ksia	*cɔʔ
'year'	snɪm	snɛm	--	snem		snim	*cnam

This creates something of a problem for historical reconstruction. The situation can be summarised as follows:

pAA		pKhasian		Modern
*ɟ	>	***ɟ**	>	ɟ ~ ʤ (and devoiced reflexes tʒi, c etc.)
*c	>	***cʰ**	>	cʰ ~ ʧ⁽ᵗʰ⁾ (and ʃ in Khasi, and before [i,j] in War/Amwi)
			>	s
*s	>	***s**	>	s (and rare ʃ allophones in Khasi)

Frankly this is a problem. Both apparent outcomes of proto-Khasian ***cʰ-** occur before high, back and low vowels, and in prevocalic and preconsonantal positions,

---

[82] Shorto reconstructs initial **s** based on Khasi [ʃ] and South Bahnaric reflexes [s, ch], the latter regularly reflect both pAA ***s** and ***c**, in this case the Khasian reflexes indicate that ***c** is the correct pAA reconstruction.

although the absolute numbers of examples are low, so it is difficult to say whether or not there is a specific conditioning environment for the apparent split/merger. The best suggestion I have at present is that pAA **c-** split, shifting to pK **s-** in most environments through an aspirated affricate stage, remaining an affricate before oral/glottal stops. Examples above such as 'to sit' **cʰɔŋ** and 'village' **cʰnɔŋ** reflect Khasian lexical innovations, and thus do not have the same segmental collocational restrictions. One apparent counter example above, 'sharpen', remains to be explained.

Sonorants

The nasals, liquids and glides in Khasian languages are mostly unchanged and present no special problems for reconstruction. The only really notable change is that word final pAA **-l** is generally reflected as a nasal, except after the diphthong **ia** where it is lost completely. At the same time final pAA **-r** is retained. This is quite an odd pattern, since it is areally common for both finals **-l**, **-r** to be merged to **-n**. I interpret this as another diagnostic Khasian innovation that evidences the genetic unity of the group. E.g.:

Gloss	Lyngngam	S.Khasi	Pnar	War-J.	Amwi	Shorto PMK
'near'	ɟəŋaːn	ɟan	--	ʥan	--	1727.A *ɟal
'thick'	--	rbɛn	rbɛn	--	rben	1768.A *[t]ɓəl
'hail'	--	pʰria	pʰria	--	pʰrua	1791.B *priəl
'cucumber'	--	kʰia	sakʰia	--	--	1710.A *[t]kiəl

Vowels

My reconstruction of the vocalism of proto-Khasian is still in progress, but I intend to present/circulate a detailed account at the Northeast Indian Linguistic Society meeting at Guwahati in February 2012. In short, it appears that the four vowel inventories tabled above can be readily derived from the following proto-vowel inventory:

Proto-Khasian Vowels:

```
/    i, iː   ɨ      u, uː
     (e)   ə, əː  (o)
     ɛ, ia  a, aː  ɔ, ɔː    /
```

Broadly speaking, we can make the following points about the historical development of the vocalism:

- The mid-vowels /e, o/ are not clearly attested, and seem to have variously merged with their higher and lower neighbours, except perhaps before final glottals.
- The etymological short-long distinction was lost in War languages, to some extent in Standard Khasi and Pnar, and seems to be well retained in Lyngngam - presenting a kind of East-West cline.
- The etymological diphthong **ia** is well preserved, while pAA **ua** seems to have been lost completely, and reintroduced variously by later sound-changes.

- Standard Khasi is particularly marked by mergers of front and central vowels, e.g. orthographic *e, ie* frequently correspond to historical central vowels; there are numerous illustrations of these in the etymologies of Shorto (2006).
- War languages frequently show a raising of **a* and **aː*, often to [i], plus dissimilation in diphthongs/vowel-glide sequences (e.g. compare Amwi *ksia* 'dog' with Pnar *ksaw*).

It is also notable that, despite having gone through a general devoicing of prevocalic obstruents, no Khasian language has apparently undergone tonogenesis or vowel splitting characteristic of AA languages historically in contact with Tai languages (see Huffman 1985 for a discussion of this type of vowel restructuring). I take this as indicative that the devoicing change in Khasian occurred before entry of Tai speakers into the same region, and after it split off from its nearest AA relatives (see discussion below). This is consistent with it being associated with the formation of proto-Khasian two or three thousand years ago, and not with it being a more recent areally conditioned shift.

Relation to Palaungic

How does the Khasian branch fit into the AA tree? The most recent published view of Diffloth (2005) is that Khasian is the highest branching node within a Northern-Mon-Khmer or "Khasi-Khmuic" family, as illustrated in the figure below.
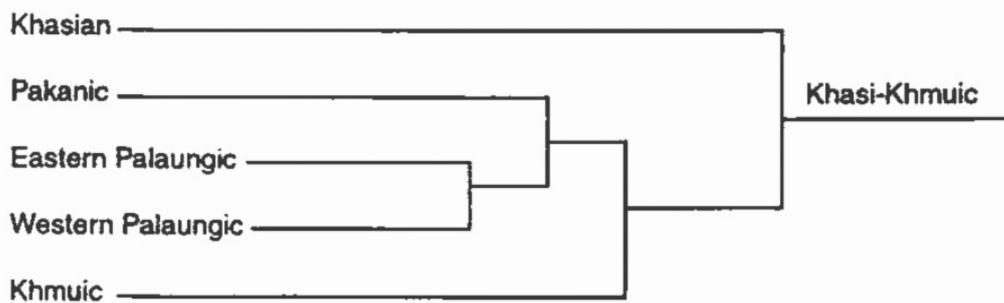
No comprehensive explanation of this classification has been offered, so the details of this proposal are difficult to analyse, but we can say that it comes on the back of a century of tradition of grouping Khasian with Palaungic. Schmidt (1904, 1906) treated Khasi as a sister of his "Salween" group, identifying various isoglosses in his limited data set. Later, the numerous lexicostatistical studies that informed AA studies in the 1960-70s also consistently suggested a Northern family including Khasian (e.g. Huffman (1978) counted 26% between Khasi and Palaung, the next highest inter-branch percentage being 24%).

In this paper I suggest that there is no compelling evidence to support a Northern Mon-Khmer family that includes either Khmuic or Pakanic (the latter I refer to as 'Mangic'). Instead, I present evidence that Khasian and Palaungic probably form a sub-family within AA. The evidence is of two types:

1. there is a disproportionally high number of exclusive isoglosses connecting the two groups; and
2. within the basic vocabulary there are non-trivial shared innovations.

A reasonable explanation for these facts posits an internal separation for Khasi-Palaungic that occurred later that their separation from other AA branches. This suggests a grouping with the following structure (incorporating analyses of Palaungic presented by this writer at the 20th SEALS meeting in Zurich in May 2010):
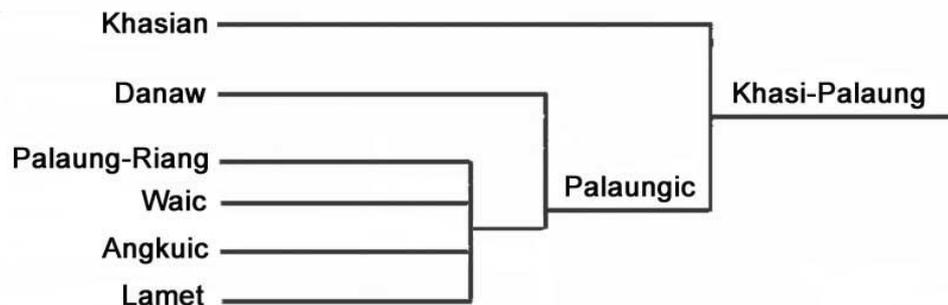
The list below sets out cognates that are restricted to Khasian and just one other branch which occur in the Swadesh 200 list, extracted from Shorto's (2006) data set of over 2000 comparisons. I restrict the comparison to items on the 200 list as a rough control on the fact that Shorto's data is somewhat skewed towards Mon, and to a lesser extent Bahnaric, due to the manner of its compilation (see Sidwell 2006 for discussion).

The raw facts are that within this set, we find the following numbers of cognates with other branches are as follows: Palaungic-6, Mon-6, Bahnaric-3, Khmuic-2, Nicobarese-1, Katuic-1. Other branches are zero. Also, in the subsequent sub-section, we note additional isoglosses with Palaungic made possible by inclusion of War/Amwi data.

Branch level isoglosses with Khasian extracted from Shorto (2006):

Palaungic-Khasian

'to burn'   **544 *taŋ** *to roast, bake.* Palaung **təŋ** *to roast, steam, heat*, Riang-Lang ˉ**taŋ** *to bake in open dish*, Praok **toŋ** *to broil*, Lawa Bo Luang **taŋ**, Lawa Umphai, Mae Sariang **tɔŋ** *to broil, grill*, Khasi *thang to burn, roast, cremate.*

'dirty'   **189 *laʔ liːʔ; *luːʔ liːʔ** *careless, slovenly.* Palaung **la li** *to be untidy, careless in*, Khasi *lali slovenly, sluttish, dirty.*

'to eat'   **1373 *ɓaam** *to chew.* Palaung **bam** *to chew, munch*, Khasi *bam /baːm/ to eat.*

'to rain'   **539 *ɟuːŋ** *rain, to rain.* Palaung **ɟuŋ** *(to) rain*, Riang-Lang _**cuŋ** *rain*, (?) Khasi *jung to urinate.*

'warm'   **1000 *taːt** *hot, to warm.* Palaung **tat** *to be hot*, Riang-Lang ˉ**tat** *to warm oneself at*, Khasi *thad to dry in sun, to bask in sun.*

'worm'   **541 *ɲaŋ** *larva.* Riang-Lang _**ɲaŋ** *caterpillar*, Khasi *'niang /ɲaŋ/ worm, insect.*

Mon-Khasian

'to float'   **1642 *ber** *to float.* Mon **pè** *to ride low in the water*, Khasi *per to float.*

'fruit'   **293 *c[u]k** *(to bear) fruit in clusters.* Khasi *soh fruit, to bear fruit, to cling, adhere*; ~ Mon **həcak** *cluster of fruit.*

'many'   **1545 *[c]haːy** *to multiply.* Late Middle Mon *[rān] chāy*, Modern Mon **[ràiŋ] chai** *to become widespread, flourish, prosper*; ~ Khasi *kyrhai abundant*; ~ *byrhai many.*

'name'   **1107 *[k]hu(ə)t** Modern Mon **khut** *to name.* Khasi *khot to call, summon, denominate.*

'rotten'   **148 *[s]maʔ** *rotten.* Mon **[ph] hmaʔ** *to be rotten*, Khasi *sma to have a bad smell.*

'to split'   **1786 *riəl** *to cut up, dismember.* Mon **rèa rey** *to cut up*, Khasi *ria small, broken*; ~ Khasi *pharia to split into small pieces.*

Bahnaric-Khasian

'to flow' **878 *huɐc** *to flow.* Central Rölöm **hɷac**, Biat **hɔːc** *to flow*, Bahnar **hɔːc** *[water] to carry away; to unroll, flow out*, Khasi **hoit** *to flow out, seep out; ~* Bahnar **təhɔːc** *to dispose of by throwing into stream to overflow.*

'man (male)' **692 *t₂nraaŋ** *man, male.* West Bahnar **kədraːŋ**, Khasi **shynrang**.

'sharp' **1100 *suut** *to sharpen; sharp.* Sre **sout**, Chrau **soːt**, Biat **choːt** *sharp*, Khasi **shut** *to sharpen.*

Khmuic-Khasian

'to know' **592 * [ ] n₁ [əə] ŋ** *to know.* Kammu-Yuan **nɛ̀ːŋ**, Thin **nɯɯŋ, hnɯɯŋ, ʔnɯɯŋ** *to know*, Khasi **nang** *to be able, to know.*

'not' **1297 *ʔam** *not.* Kammu dials. **àm**, Thin **ʔăm**, Khasi **em, ym**.

Nicobarese-Khasian

'belly; guts' Khasi **rwieng** *intestines of bird or fowl*, Nancowry **wíaŋ** *belly, stomach*

Katuic-Khasian

'near' **1727 *ɟal** *near.* Khasi **jan** *to be near; ~* Kuy **ɲchàl** *near.*

The above compilation indicates that unique Khasian-Palaungic cognates are well represented, being found at least as frequently as unique Khasian-Mon cognates, and twice as frequently as unique Khasian-Bahnaric cognates, despite the fact that Shorto worked with very limited Palaungic data, while at the same time having access to extensive Mon and Bahnaric dictionaries, and a century of comparative work utilizing those languages. This is much more significant than a simple lexicostatistical count, because it is indicative of unique lexical innovations or retentions rather than gross percentage differences. However, even more important than these, are specific innovations within particular sets, discussed below.

Khasian-Palaungic Innovations:

Gloss	pKhasian / S-Khasi	Palaungic	Mangic	Khmuic (Chuang)	Munda	Shorto PMK
1) blood	*snaːm / snam	*snaːm / naːm¹(Riang)				-
			saːm (Bolyu)			(*saːm 'to bleed') *ɟhaːm
				maːm	mã̃jã̃m (Santali)	
2) claw/nail	*trcʰiːm / tɨrsim	*rnsiːm / r°mhim¹(Riang)				-
						-
			maːi¹³ti⁵⁵ (Bolyu)			-
					rama (Santali)	-
				tmʰmɔːŋ		*tm[uə]ŋʔ
3) hair	*cʰɲok / ʃɲuʔ	ɲok (Danaw)	suk⁵³ (Bolyu)			*suk
	*cʰuk / suʔ (War/Amwi)	*suk / huk¹(Riang)				
					u'b (Santali)	
			glə̆ʔ			

	Khasian	Palaungic	(Mangic)		Munda	Reconstruction
4) man/ husband	tərma: (War), trmɛ (Amwi)	*-meʔ / kᵊrmeʔ²(Riang)	qɔ³¹pɔ³³ (Bolyu)	gleʔ	hɔɽ (Santali)	- - - -
5) rain	*slap / slap slɛ(Amwi) sla: (War)	*sᵊlɛʔ / saléʔ(Lamet), kᵊlɪ¹(Danaw)	maʔ (Mang) qɔ⁵⁵ (Bolyu)	kma ʔ	ɟargɛ (Santali)	- *gmaʔ -
6) swim	*ɟɲi: / ɟɲi:	*ŋɔj / ŋɔj²(Riang)		kljɔːŋ	pajraʔ (Santali)	*[l]ŋuj -
7) two	*ʔa:r / ʔa:r	*lʔa:r	mbi⁵⁵ (Bolyu)	ba:r	bar(Santali)	*ʔa:r *ɓa:r
8) water	*ʔum / ʔum	*ʔo:m / ʔom¹ (Riang)	(ʒum Mang?) nde⁵³(Bolyu)	ʔom daʔ(Cuang 'water source') ʔɔːk(Pray) paj(TaiHat) paʔaŋ (Pong)	daʔ(Santali)	*ʔ[o]m *ɗaak

The table above sets out data for eight semantic fields within the same basic word list in which we can point to specific Khasian-Palaungic innovations, and demonstrate that these are not especially shared with either Khmuic or Mangic. Commentary follows.

1) Ferlus (2009) reconstructs a proto-Vietic root *sa:m 'to bleed' (e.g. Viet. *tươm* 'ooze, exude', which has a direct cognate in Mangic, and is uniquely infixed in Khasian and Palaungic. It is not clear how this connects to other Austroasiatic forms which indicate a prevocalic /h/ (note regular loss of /h/ in Khmuic).

2) The fingernail/claw etymon unique to Khasian-Palaungic is apparently morphologically complex, including a non-trivial -r- affix, strongly indicative of a common innovation.

3) Remarkably, in both Khasian and Palaungic, a subset of languages shows the apparent intrusive palatal nasal.

4) Terms for man/husband are diverse and apparently rather unstable in AA languages, hence the lack of an indicative reconstruction by Shorto. This makes the finding of a common form in Amwi and Palaung, including -r- refix, very significant.

5) The War/Amwi forms *sla:, slɛ* 'rain' appear incongruous in the Khasian context, as the other members of the group have *slap* or similar. However, the Palaungic comparisons strongly suggest a proto-form *slaʔ* (or similar) such that Khasian reflexes with rhyme /ap/ as Sandhi or pun forms.

6) These 'swim' isoglosses are compared by Shorto to Bahnar *ŋəːj* '[fish] to come to the surface, float' but forms such as Acehnese *laŋuə* 'float' rather suggest Chamic borrowing into Bahnar of an accidental look-alike (it is hardly reasonable to propose Khasian or Palaungic contact with Chamic or Malayic).

7) Forms for 'two' are problematic. Nicobarese also shows a reflex without the otherwise expected initial labial (e.g. Nancowry *ʔã*), but in that case it is possible to suggest an independent explanation, since Nicobarese, Acehnese, Moklen form a area that tends to reduce implosives to glottal stop. In any case, it is not clear whether the forms with initial glottal stop are innovative or archaic (Shorto suggested origin via a reduction of hypothetical **biʔaːr).

8) The root **ʔ[o]m* 'water' replaced proto-Austroasiatic **ɗaːk,* perhaps by development from a root meaning 'to bathe' (cf. proto-South Bahnaric *ʔum 'bathe'). The same etymon does occur in Khmuic, but restricted to Khmu Cuang, Khang, and Bit (Bit may be Khmuic or Palaungic, sources conflict). Other Khmuic languages have diverse forms for 'water', e.g. Iduh *paj*, Ksingmul *hɔːt*, Mlabri *wək*, Pray *ʔɔːk*. This suggest borrowing replacement in Khmu Cuang etc. Additionally, Mang has the curious *ʒum* form, which suggests *rum by regular correspondence, which may be speculatively compared to Waic *rʔom 'water' (Diffloth 1980).

Conclusion

As I have argued, ever so briefly above, it is apparent that the Khasian branch of AA is a coherent phylogenetic sub-grouping of languages, most closely related to the Palaungic branch with AA. Although some details remain obscure, and the task is far from complete, it is possible to offer a model of proto-Khasian phonology and lexicon.

Emerging genetic studies, such as Langstieh et al. (2004), show that the bulk of Khasian speaking peoples form a coherent population with closest genetic relations among Northern Mon-Khmer groups further east. At the same time, Lyngngam speakers have a genetic profile consistent with Garos whom they neighbour to the West, suggesting a westward language shift.

Taking all of the above together, I offer the schematic diagram below (previously presented at the 2011 SEALS meeting), which approximates geographically some historical features of the development of the Khasian group. The diagram is to be read as follows: Pnar emerges more or less directly out of proto-Khasian, with no major restructurings or migrations. Standard Khasi emerges from a dialect that is marked by particular vowel developments of central vowels and diphthongs. The War sub-branch has the most extensive loss of vowel length distinctions, and restructures in a manner that raises the low central vowel and dissimilates diphthongs. Lyngngam reflects a pKhasian dialect spoken by a previously Garo population, although no extensive Garo linguistic influence is apparent.

Schematic Reconstruction of Khasian history
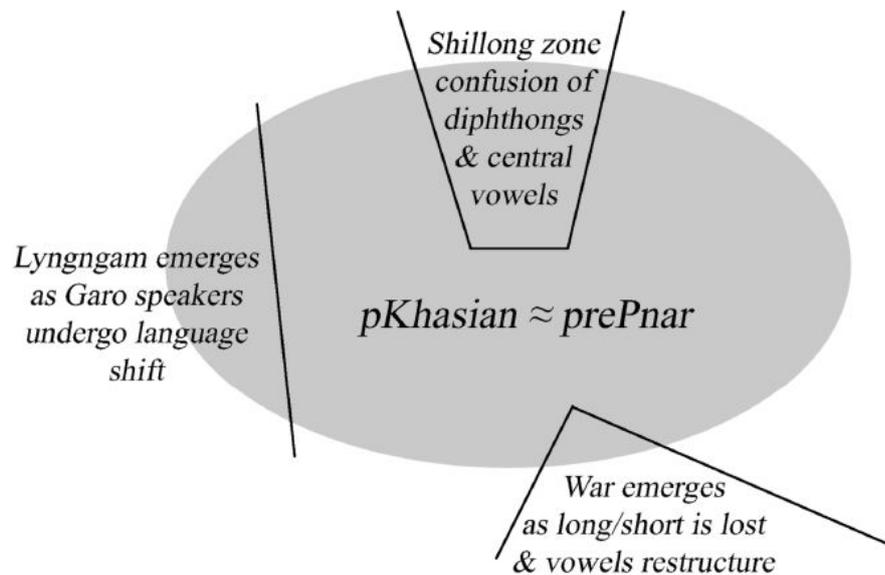
Additionally, the Khasian family has a rich internal and external language contact history, but this only makes the task of comparative historical reconstruction more complex and more interesting, and arguably more important.

Postscript: remarks on Daladier's "The Group Pnaric-War-Lyngngam and as Khasi as a branch of Pnaric".

It is very gratifying that Anne Daladier (henceforth D) has taken the time and effort to respond to my paper "Proto Khasian And Khasi-Palaungic" by setting out her own perspectives and results. It also shows a great generosity of spirit that she further agreed to allow this response discussion in the same issue of JSEALS. This is a great realization of our editorial policy that "JSEALS welcomes articles that are topical, focused on linguistic (as opposed to cultural or anthropological) issues, and which further the lively debate that characterizes the annual SEALS conferences." Additionally, the data tables provided by D are incredibly useful and valuable, and I resolutely commend this demonstrated commitment to empiricism and data sharing.

Coming to the point of this short commentary, I must admit that I have difficulties with D's theoretical approach, which appears to me to be antithetical to the comparative method as I understand and practice it. My main problem with her paper is that, as far as I can tell, her theoretical approach is not clearly laid out or characterized, and this really makes it difficult to assess any of her claims, both internally and externally.

The role of theory in science is paramount, being the highest form of knowledge in science. By contrast facts are often cheap and plentiful in the real world, but lack value without an adequate framework with which to understand and use them. Therefore, in this commentary I will on focus on the theoretical issues at stake in this debate.

In my paper I have taken the limited facts concerning Khasian languages at my disposal and offered an analysis on the basis of the theory we call the *comparative method* (henceforth CM), a demonstrably robust theory that has served linguistics well for a

century and a half (and subject to continuous improvement and elaboration since the pioneering efforts of 19thC scholars such as Rask, Verner, Grimm, Schleicher etc.). Important general features of the CM, relevant for this discussion, include the following assumptions:

- all languages change over time, modifying, creating, acquiring and losing features;

- languages tend to diversify over time in a branching manner,

- characterization of features variously as innovative or conservative implies branching relations which may be represented schematically, with nodes representing (proto-)languages connected by sequenced changes,

- branches can cease, or merge to form mixed languages, which will show characteristic structural asymmetries (i.e. mixing is not homogenous),

- the effect of conducting a comparative analysis is that historical changes correlate with specific branching or contact events among languages.

Presently, I have done this with the very limited Khasian data at my disposal, suggesting a preliminary reconstruction of proto-Khasian. I am quite willing to accept that these results may be incorrect, but I do not find a refutation of my thesis in D's paper. Rather, I find a collection of claims which may or may not be true, but which do not appear to constitute a model of linguistic change and reconstruction that can be tested in accordance with the precepts of the CM. This is not in itself a rejection of her results - I expect that they are very useful - but so far I see no inconsistencies between the thrust of my thesis and the facts presented in the preceding pages by D.

Daladier's subgrouping hypothesis?

The broad consensus among comparative linguistis, since at least the *Linguistic Survey of India* (Grierson 1905 etc.) and the work of Schmidt (1904), has been that the Austroasiatic languages of Meghalaya hills constitute a single phylogenetic grouping (a view which is also implied by the frequent characterisation of them as merely "Khasi dialects"). Against this view, D offers the perspective of the Khasian languages ("PWL" in her nomenclature) as a "group on another converging trade route in Assam at the beginning of our era or a little before." Not a phylogenetic unity, but an areal grouping of four "conservative AA groups", namely Pnaric, War, Lyngngam, and Khasi. The four groups are represented by "conservative varieties" (variously named), additionally we are told that there are numerous "composite varieties" or "mixed varieties" (also variously named). This linguistic mélange is explained as the consequence of "intricate migrations".

The concepts of conservative and mixed varieties are not well defined by D beyond the assertion that the conservative varieties are not the products of more recent mixing. A couple of crucial implied claims arise:

- that D is able to distinguish conservative and mixed varieties, and

- that conservative varieties, so distinguished, are informative of the ancestral forms of the languages that gave rise to the four modern groups.

One of many consequences of these claims is that citation of data from "mixed varieties" by this writer in support of his reconstruction is criticized as illicit by D. Common sense suggests that the four conservative languages (as so characterised by D) represent what the

mainstream would call four distinct branches of Austroasiatic, which due to proximity came to share some innovations, creating the appearance of a single "Khasian" branch.

What are the implications for the diversification of Austroasiatic? What innovations are associated with each node in the branching tree such that we can distinguish innovative from conservative features? This is really the crux of my difficulty with D's claims. No such model is provided; rather various illustrative examples of differing features in the conservative varieties are given, and many comparisons in other Austroasiatic branches are offered, to show how different they from each other. The problem I have with this kind of discussion is that, absent an explicit model of phylogeny, no amount of "difference" can constitute a counter-argument to my (or any particular) proto-Khasian hypothesis. Tremendous innovation may have occurred in various Khasian languages over the past couple thousand years. The accrual of innovations will be asymmetrical as Khasian has internally branched and diversified, replacing archaisms in some languages, but not all, with the result inevitably that some will show up as isoglosses/parallels in other Austroasiatic branches. In that general sense, D's evidence of differences does not falsify my theory by contradicting its predictions.

A couple of example from D's paper:

> Lyngngam 'lips' *ləmɔr* is cognate with Santali *ləmer* 'to move the lips'. Pnaric and War have *tʃəntur*.

> Pnaric and Lyngngam **kʰlaɲ* 'grease, fat, marrow' is related to Shorto (2006:928) **klaɲ* in Khmeric, Monic and Vietic: Khmer *klaɲ*, Surin *khlaɲ*, Mon *klɔ̰iŋ*, Nyakur *kəlíɲ*, Ruc *kluŋ²*.

> War *ləʔɔt* is probably related to Shorto (2006:1879) **lʔus* 'fat, in Palaungic and in North and Central Aslian, Semai *lʔuːs*, Temiar *lɛnʔɔs*. (/t/ in rhymes in War, is often found in MK cognates with /s/).

In such cases we are confronted with a diversity of wider isoglosses; such that in the former it is Pnaric and War that share a particular external isogloss, while in the second it is Pnaric and Lyngngam that share an external isogloss. This is quite unremarkable. The proto-AA lexicon was complex, much of it was transmitted into proto-Khasian, and subsequently there have been numerous replacements, semantic shifts, dialect borrowings, and reanalyses that yielded the lexical distributions such as illustrated above. Consequently, citations such as these do not contradict the proto-Khasian hypothesis, but a substantial compilation should provide the basis of an internal classification of Khasian that would reveal any nested branching relations.

The same applies to D's listing of phonological facts, such as:

> War is the only language of PWL which has a phonemic opposition between two fricatives /tʃ/ and /tʒ/. /tʒ/ is in fact a devoiced palatal affricate with a smaller quantity of friction than /tʃ/.

And likewise various phonological differences are specified by D for the different languages. This in itself is valid and useful information, but it is not evidence that these varieties come from distinct branches. They are still readily analysed in terms of regular developments from proto-Khasian phonemes, and in practice the same logic could be applied to the syntactic and morpho-syntactic data in the context of a reconstruction.

And the very term "conservative" is used in ways that to my mind rather suggest an exactly opposite interpretation. For example, when D writes that "Unlike most MK languages but like Nicobarese and Munda, PWL still have productive suffixes like the depreciative -siʔ on verbs." she is implying that suffixation is an archaic AA feature. Quite the reverse is true: The transformative work of Donegan & Stampe (2004 and passim.) demonstrates that suffixation as such is innovative in AA, and it is uncontroversial that the suffixal systems of Munda, Khasian and Nicobarese languages (more properly characterised as *clitics* in the latter two groups) are not cognate.

The significance of the "mixed varieties" is also ambiguous. Without information to the contrary, I am happy to accept D's account of mixed varieties, but I see no counter argument to my thesis in this. It seems straightforward that if a feature is attested in a mixed variety, it may still be treated as inherited from proto-Khasian and can be invoked to justify reconstruction of a proto-Khasian form, subject to usual caveats.

The bottom line is that claims about features being conservative or innovative are inherently dependent upon an explicit theory of descent, but no such theory is articulated in D's paper.

Daladier's approach to morphology

Another important aspect of D's paper is reflected in her novel approach to morphology, which plays a role in her wider etymological analyses. In this, and other papers, D is articulating a theory in which words can be segmented in diverse ways, and sometimes also conflating roots that are otherwise considered distinct. Several examples:

'blood'

> I analyse *snaːm* 'blood' in PWL as a reduced form of an AA word formative *san* into *sn-* combined with *ʔam* with lenition of the glottal. More generally, the AA names of 'blood' may be analysed with different structures combining with *ʔ[o]m* 'water' where the vowel specializes into /a/. To me these different structures are like *koraŋ* and *ʧənraŋ* 'male elder analysed before, not cognates in a strict sense but extensions from a common cognate, here of water, body liquids.

Here we see segmentation between rime and preceding segments. D is challenging the analysis of Ferlus (2011), which accounts for reflexes found in Khasian, Palaungic and Mangic by invoking a nominalisation of the root *saːm* 'to bleed' which is uncontroversial attested in Austroasiatic. D's proposal is less elegant, and poses serious phonological problems. She posits a root *ʔaːm* 'blood' which is not directly attested, but is supposed to account for forms with variously prevocalic *h-, ɲ-, j-, ø-,* without explaining the bases of the supposed phonological changes.

'male; horned being'

D offers an extensive commentary comparing forms that she relates to a hypothetical masculine figure meme. E.g.:

> *koraŋ* 'male' in Lyngngam is connected to *goraŋ* 'male, household ancestor' in Bonda (South Munda). I analyse *koraŋ, goraŋ* as a combination of two AA word formatives *kur* 'clan descent' and *rVŋ* as a word formative for horned beings in an abstract metaphorical sense. A corresponding grammaticalized word formative *rVŋ* is suffixed or prefixed in many AA cognates related to 'horned beings' in a concrete or in a metaphorical sense,

including males and things or beings related to the Underworld (see Daladier 2007); *rVŋ* is found in *tʃənraŋ 'male' in Pnaric and in *roŋbaʔ* 'representative village man, male adult' in War.

These data reflect several roots:

- Munda root *kor 'man' represented at Pinnow (1959) A311, and for which no Mon-Khmer cognate is cited by Shorto (2006) nor otherwise known to this writer;

- AA root reconstructed as *d₂raŋ 'horn' by Shorto (2006), on the basis of reflexes such as Old Mon *draŋ*, Sora *ˈdeˑraŋ*. There is a variant with schwa main-vowel in Palaungic *krəŋ, and Vietic *k-rəŋ, which agree well with Khasian cognates such as Lyngngam *ʔəreŋ*;

- A root reconstructed as *t₂nraaŋ 'man, male' by Shorto (2006), on the basis of West Bahnar *kədraaŋ*, Khasi **shynrang**. Lyngngam *koraŋ* is clearly cognate, but we no firm basis for considering the *k-* presyllable indicative of the proto-form, given the general tendency to analogically level presyllables to *k-*, and the Khasi and Pnar indication of a palatal.

I see no basis for explaining the later with reference to the two preceding roots. The Munda root *kor is phonologically simple, and the initial velar so generic and unmarked that an unconstrained eye could imagine seeing it in hundreds of words. The supposed association with 'horn' is unconfirmed by citation of attested phrasal or other constructions that might link it in the manner indication.

'eat, chew, food'

> Pnaric has *bam* 'eat', War *bua, ba, bu* 'eat, food'. Pnaric has *saː* as an imitative of *bam*, with many cognates in MK, and War has *sum* imitative *sam* for the small wild plants which produce edible seeds near the village, see below. Lyngngam *b(ənn)aŋ* 'eat' is probably related to *ba with a shift from *baŋ* 'tasty' in Pnaric and in War; *baŋ, bam* 'eat' is probably related to Palaungic: Palaung *bam* 'chew'', Munda, Bonda (rɔri) bai 'eat (much)', *bun* 'eat' in Parengi (South Munda) (Bhattacharya 1968), South and North Munda have 'eat belly-fully' *bʔe* in Sora, *bi* in Santali, Mundari, Ho, Korku.

The above is readily analysed:

- The basic AA root for eat is *caʔ, which survives in, e.g. Khasi *bsa* 'to feed', which is hardly an imitative.

- In various AA language terms for 'chew', 'bite', 'suck', 'hold in mouth' and such have replaced reflexes of *caʔ to become the normal words for 'eat'.

- Refering to Shorto (2006) entries 1375 and 1376 we see numerous variants throughout AA on the pattern (*k/g*)*ɓVm* with meanings related to 'hold in mouth', clearly 'baby-words' with sound-symbolic formation, indicative of the origin of Khasian forms *bam, baŋ* etc. and Palaung and Katu *bam* 'chew'.

Proto-AA *ɓaʔ 'paddy' (see Shorto 2006 entry 120) is simply a word denoting an edible grain and there is no basis for connecting it to the various 'eat, chew etc.' words with final nasals. Yet, remarkably, based on this and several other examples D reconstructs an AA

final *-m* morpheme for "edible beings". This includes such non-viable etymologies as deriving proto-AA **kta:m* 'crab' from Sanskrit *karkata + -m.*

Concluding remarks

D's paper is quite rich, containing much useful factual data that is sorely welcomed. However, I have reservations concerning here analyses and theoretical principles. Her hypothesis that the various Khasian languages are distinct conservative AA languages rather than a coherent phylogenetic grouping is not well formed. Her richly illustrated text shows differences between the languages does not explain how those differences were inherited or innovated in a manner at variance with my proto-Khasian hypothesis. Secondly, D's method of etymological analysis, which posits morphemes by splitting words at final consonants, rimes, or medial clusters to facilitate comparisons within specific semantic fields has no basis in the known morphological properties of AA languages and is at variance with prevailing views concerning how AA etymology is to be carried out.

Finally, I do not wish my remarks to be taken as being too harsh. I absolutely commend D's paper to readers, and urge them to consider it thoroughly and carefully. My own proto-Khasian hypothesis is rather preliminary and unproven, but my present view is that while it is weak on data it is methodologically strong. By contrast, I am delighted to defer to D in matters of fact concerning the AA languages of Meghalaya and surrounds, but as I hope to have made clear, I do not accept the theoretical underpinnings of her comparative analytical approach, and disagree with a proportion of her finding. I hope that my commentary above contributes to an ongoing dialogue that can help to improve our mutual understanding and advance the generally the historical investigation of the languages in question.

References

Bareh, Curiously. 2010. *A Descriptive Analysis of the Jowai and Rymbai Dialects of Khasi.* PhD thesis North-Eastern Hills University, Shillong, India.

Brightbill J., Amy, K., Seung, K. 2007, The War-Jaintia in Bangladesh: a soscio-linguistic survey, SIL electronic report

Choudhary, Narayan Kumar. 2004. *Word Order in Pnar.* MA thesis, Jawaharlal Nehru University, India.

Diffloth, Gérard. 1980. *The Wa Languages.* Linguistics of the Tibeto-Burman Area. Vol. 5.2. Berkeley: University of California.

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In: Sagart, Laurent , Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics.* Routledge/Curzon. pp79-82.

Doerfer, Gerhard. 1963. "Bemerkungen zur Verwandtschaft der sog. altaische Sprachen", (Remarks on the relationship of the so-called Altaic languages), In *Türkische und mongolische Elemente im Neupersischen, Bd. I: Mongolische Elemente im Neupersischen*, 51–105, Wiesbaden, Franz Steiner Verlag.

Donegan, Patricia and David Stampe. 2004. Rhythm and the Synthetic Drift of Munda, The Yearbook of South Asian Languages and Linguistics. Berlin and New York, De Gruyter. pp 3-36.

Ferlus, Michel. 2009. Le mot "sang" en austroasiaque. *Mon-Khmer Studies*, 38:25-39.

Georg, Stefan, Peter A. Michalove, Alexis Manaster Ramer, and Paul J. Sidwell. 1999. "Telling general linguists about Altaic." *Journal of Linguistics* 35:65–98.

Haudricourt, André-Georges. 1965. Mutation consonantique en Mon-Khmer. *Bulletin de la Société Linguistique de Paris,* 60:160-72.

Heggarty, Paul, Warren Maguire & April MacMahon. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3829-3843.

Henderson, E. 1989-1990, "Khasi clusters and Greenberg's universals", in *The Mon-Khmer Studies Journal*, vol. 18-19, pp. 61-66.

Henderson, Eugénie J.A. 1976. Vestiges of Morphology in Modern Standard Khasi. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies, Part 1.* Honolulu, University of Hawaii (Oceanic Linguistics, Special Publication, No. 13). pp.477-522

Huffman, Franklin E. 1978. *On the centrality of Katuic-Bahnaric to Austroasiatic.* Communication à la 2-ième Conférence Internationale sur les langues Austroasiatiques, Mysore (Inde) In *Proceedings of the International Congress of Anthropological and Ethnological Sciences*, 10.2:78-79.

Langstieh B.T. et al., "Genetic diversity and relationships among the tribes of Meghalaya compared to other Indian and Continental populations". *Human Biology.* Aug 2004;76(4):569-90

Maspéro, Henri. 1912. Etude sur la phonétique historique de la langue annamite. Les initiales. *Bulletin de l'Ecole Française d'Extrême Orient* 12.1-27.

Nagaraja, K. S. 1996. The Status of Lyngngam. *Mon-Khmer Studies* 26:37-50.

Nichols, Johanna. 1997. "The Epicenter of the Indo-European Linguistic Spread". In *Archaeology and Language I: Theoretical and Methodological Orientations*, Roger Blench and Matthew Spriggs (eds.), London: Routledge.

Ogura, Meiko. 1987. Historical English Phonology: a Lexical Perspective. Tokyo; Kenkyusha.

Post, Mark W. 2011. "Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against "Sinosphere" and "Indosphere"." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.198-221.

De Roepstorff, 1875. [2nd ed.] Vocabulary of dialects spoken in the Nicobar and Andaman islands. Calcutta: [no publisher].

Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian languages of Western Melanesia*. Pacific Linguistics. Canberra: Australian National University.

Schmidt, Pater Wilhelm. 1904. Grundzüge einer Lautlehreder Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayrischen Akademie der Wissenschaft,* 1.22.3:677-810.

Schmidt, Pater Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie, Braunschweig,* 5:59-109.

Sebeok, Thomas A. 1942. An examination of the Austro-Asiatic Language family. *Language*, 18: 206-217.

Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary.* Canberra, Pacific Linguistics 579.

Sidwell, Paul & Roger Blench. 2011. The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In *The Dynamics of Human Diversity*. N. J. Enfield (ed.). Canberra: Pacific Linguistics. pp.1-30

Sidwell, Paul. 2006. *Preface* to Shorto 2006 *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics 579; pp. vii-xxv.

Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich, Lincom Europa.

Thomas, David & Robert Headley 1970. More on Mon-Khmer subgroupings. *Lingua* 25:398-418.

Weidert, Alfons K. 1975. *I tkong Amwi. Deskriptive Analyse eines Wardialekts des Khasi*. Wiesbaden: Harrassowitz.

THE GROUP PNARIC-WAR-LYNGNGAM AND KHASI AS A BRANCH OF PNARIC[83]

Anne Daladier,
LACITO-CNRS Paris
<A.DALADIER@wanadoo.fr>

Abstract
Historical information shows that East Meghalaya is a late refuge land with settlement in several groups of Lyngngam, Pnar, Khasi and War. A diversity of mixed languages in the context of migrations and two sucessive lingua francas (Pnar and Standard Khasi since the British colonisation) is a prominent feature of this group but there are still more conservative varieties of War, Lyngngam and Pnar. Pnar still is the main spoken language and is closer to War and to Lyngngam than Khasi. Negations and pronominal systems in Pnar, War, Khasi and Lyngngam have distinct morphological isoglosses in Mon-Khmer and in Munda. After four centuries of Pnar Lingua Franca in Meghalaya, Khasi lexicon appears to be Pnaric (including its cardinal system). After centuries of Pnar Lingua Franca, Lyngngam and War have integrated different Pnar sub-lexicons in their lexicons but Lyngngam and War are not Pnaric in origin as shown by different isoglosses in Munda (especially Juang for Lyngngam and Sora for War) not shared by Pnaric. Pnaric and War show significative morphological differences but Lyngngam differs from Pnaric and War in even more important morphological and systemic features. More data and analyses are needed to set the branching of Pnaric, War and Lyngngam and the branching of this Pnaric-War-Lyngngam (PWL) group in the AA tree.

Key words: Lyngngam, Pnar, Khasi, War language, morphology

1. Introduction
First, I would like to thank Paul Sidwell who urged me to precise my views on a classification of the so-called Khasian group different from his own and to publish side by side our papers. In a first draft, which I had sent him a few weeks after he had sent me his own first draft, I had made an attempt to answer too many questions raised by his article and I will present a more restricted answer. I will show that it is premature to offer a precise tree of this group and its branching in the AA family. Different unsolved questions raised by the intricate situation of Meghalaya as a refuge land, by AA classification and by comparative methods should be solved one after the other. Further data are requested, not only on lexical cognates but also on AA pre-cardinal systems, AA verbal and deictic bases, AA sesquisyllabic structures and AA negation systems.

As surprising as it may seem, the long tradition of Khasi as a main conservative language with Pnar, War and Lyngngam as its offshoots has never been grounded on linguistic descriptions of Pnar, War and Lyngngam or even on precise (well located)

comparative data. As will be shown, it appears that till now the central language of the group is Pnar and not Khasi; Pnar appears to still be the main spoken language in Meghalaya and still is closer to War and to Lyngngam than Khasi. I will show here that Khasi appears to be an innovative Pnaric language. Standard Khasi (S. Khasi) can be called "conservative" from an AA viewpoint as Shorto's comparative entries show. However, as a Pnaric language, S. Khasi is not a conservative language from the view point of the classification of this group here called Pnaric-War-Lyngngam (PWL).

The settlement of Meghalaya by Pnar groups in the East and by Lyngngam groups in the West might have started in the 15th century, some two centuries after the arrival of the Moguls and the Tai Ahom and their wars in Assam and in the Gulf of Bengal (see §2). Garos might have settled in West Meghalaya and Karbis around North Eastern Meghalaya also in this period for the same reasons. The settlement in Meghalaya has two unification periods for MK languages, firstly under Pnar lingua franca, secondly under Standard Khasi. Mixed varieties of all kinds: Pnar-Khasi, War-Pnar, War-Khasi, Pnar-Lyngngam, Lyngngam-Pnar-Khasi, Pnar-War-Khasi-Karbi, Lyngngam-Garo, Khasi-Assamese, Pnar-Bengali etc. actually are an important feature of the languages spoken in the Khasi constitution of Meghalaya. In the centre-West and in the North, it is not always clear whether a variety is Pnar-Khasi or Khasi-Pnar and the burden of classifying such mixed varieties would be pointless from a classification viewpoint. Some mixed varieties have interesting remains as will be shown below. From the viewpoint of Pnaric, War and Lyngngam classification, War and Lyngngam have what I call, "conservative" varieties, that is varieties having gone through some internal evolution with their own innovations and with less influence from the two successive lingua franca. The eastern Pnar varieties are very close and may be called conservative though the written Jowai Pnar is somewhat influenced by S. Khasi. These Pnaric, War and Lyngngam conservative varieties have preserved different specific phonological, lexical and morpho-syntactic features which defines them. Some of these features are briefly sketched here. On the other hand, Lyngngam and War conservative varieties have also now acquired different large segments of Pnar and Standard Khasi lexical elements. For example, conservative War has lost *dʒ* (remaining in Khasi, Pnar and Lyngngam) except in loans like *dʒiŋmut* 'meaning' from S. Khasi and has a phonological opposition of fricatives *tʃ/tʒ* (see §3). The War-Khasi mixed varieties spoken in the South Khasi Hills have lost most of the conservative lexical, morphological and syntactic features of conservative War but have got the Khasi voiced fricative, lost in conservative War. Thought the opposition of fricatives is an innovation, it is an isogloss of conservative War, in the relative sense given here to the opposition of conservative and mixed varieties of PWL. I aim at showing that Pnar, War and Lyngngam conservative varieties, as delimited in §2, are very conservative AA languages, using lexical and morphological comparisons with AA reconstructions offered by Shorto (2006) and by Pinnow (1959) and also with a few personnal hypotheses.

Wars have migrated in Meghalaya in many different groups, coming mostly from what is now Bangladesh, on Khasi lands during the colonial period but before that period and after the partition with Bangladesh on Pnar lands, in fixed villages. Lyngngam groups speaking conservative varieties have been jhum cultivators without fixed villages until very recently. This situation is described and illustrated with a map of conservative and mixed varieties in Meghalaya and a detailed map of War presented in Daladier (2012 b).

More accurate maps of Lyngngam varieties and an overview of isolated groups outside Meghalaya are in preparation. A summary is presented in §2.

The linguistic consensus since the Linguistic Survey of India is not that the MK "languages" of Meghalaya constitute one single phylogenetic grouping but that S. Khasi is the main language of this group. For Grierson (1927) Khasi is the only "language" of this group, Pnar (Synteng), War, and Lyngngam being "dialects" of S. Khasi. The only comparative data on which this claim is based is the word lists of some 80 words plus 160 grammatical elicited items gathered by Roberts and presented in Grierson (1927). Most of these items are conjugated verbs in the various tenses and moods of English, nominal declination according to Latin cases and basic elicited sentences plus two pages of the New Testament translated in Pnar, War and Lyngngam. There are no tenses in PWL, as in most MK languages. The lexical list shows interesting unrelated cognates but is too small to draw any conclusion. The data of Roberts used by Grierson, especially the glossed translations, show that S. Khasi and Jowai Pnar are lexically very close but that Khasi, War and Lyngngam are not mutually understandable. In his preface, Roberts (1891/1995[2]: xiv-xv) claims his scientific affiliation to Max Müller's comparative methods and after quoting Khasi-Hebrew cognates writes "In this work, the dialect of Cherapoonjee is taken as the standard because it is the purest, as universally acknowledged by the natives (…) Such ugly barbarisms as (…) *ham klam* and *ri shïar tha* for *wat krɛn* should be avoided". In place of "barbarisms", quoted data are genuine utterances in Pnar and in War. *wat krɛn* 'don't speak!' is Khasi while *ham klam* 'don't speak' is Pnar and *ri shïar tha* sounds like Kudeng War: *ri, ʃiar tə* 'leave it, speak no more' (see §4 negation *tə*). The Welsh missionaries first settled in Cherra poonjee (alias *Soʔra* 'bringing fruit') where Roberts fixed a Khasi orthography, a westernized grammar of Khasi and a Khasi-British Bible style.

Diffloth (2005) uses the label Khasian for the group of S. Khasi, Pnar, Lyngngam and War "varieties" or "languages" suggesting without showing it, phylogenetic unity of a group centered on Khasi. This viewpoint is made more precise in Diffloth (2011:1100) who analyses Khasian into: Khasi, Synteng, Lyngngam, Nongtung, Amwi (War), Bhoi. He does not give data on these groups and no derivation tree is presented for his new Khasian group and its six sub-groups. Synteng is an alternative name for Pnar, used formerly when it was spoken in the area of Sutnga. At first Bhoi was a Pnar northern territory of the Manar clan (*bhoi* 'north' in Pnar). Bhoi is now an administrative unit where Karbi (Bodo), Pnar, Khasi and War speak different kinds of mixed varieties in addition to a few Pnar and War conservative varieties. Nongtung is one of the northern mixed varieties. Amwi (War) is only one of the three War conservative groups (see §2). On the SIL web site 2011, Amwi is equated with War in Meghalaya and Nongtalang is listed as a Pnar (or Synteng) dialect. The word list in Kudeng Nongtalang War of annex 2 and a sample text in Daladier 2012b show that Nongtalang is a sister dialect of Amwi if compared with Grierson (1927) and Wiedert unpublished lexicon available at SEALANG web site.

Brightbill et al. (2007) propose a soscio-linguistic study of unclassified varieties of what they call "jaintia" containing Pnar and War conservative varieties together with mixed varieties Pnar-War and War-Pnar. They include in their comparative word list a Lyngngam variety spoken in Bangladesh and also S. Khasi. This list is very difficult to use without knowing Pnar and War as one does not know whether a word is Pnar or War, for example Noksia and Jaintiapur are mostly Pnar speaking villages. These varieties cannot

be used to show that War is Khasian. War from Aliachora near Tripura border in Bangladesh is relevant for classification as very close to conservative Nongtalang-Nongbareh varieties. Data on Lyngngam should also be based on conservative varieties for classification purposes. Lankymma Lyngngam presented in annex 2, though influenced by S. Khasi appears to be much less influenced by Pnar than Langrin and Nongstoin; Tra Lyngngam which I am currently documenting appears to be the most conservative.

Shorto (2006) marvellous documentation has to be completed by documentation on War, Pnar and Lyngngam conservative varieties. Many of the specificities of War and Lyngngam can be related to AA cognates different from Pnaric (or Khasi when Khasi differs from Pnar) or to AA grammatical morphology.

The Grammar of Amwi War by Weidert (1975) is not accurate in several phonetic, phonological and morpho-syntactic important respects. For example, he did not recognize the labio-dental voiced fricative of War not found in Pnar and Khasi and long vowels, all but /ɨ/, several diphtongues in addition to /ia/ and /ua/ and the opposition between two fricatives, isoglosse of conservative War (see War phonetics in §2). The analysis presented here is based on some 500 lexical and morphological elements in Pnar, War, Lyngngam conservative varieties and S. Khasi (some 320 elements presented in six tables and two annexes).

In annex 1, I have listed cognates between the Luce (1965) comparative wordlist (245 words) of Old and Modern Mon and six Palaungic languages with S. Khasi, Ralliang Pnar Kudeng War, and Langkyma Lyngngam. The link between Palaungic and S. Khasi, acknowledged by Luce (1985) and Shorto (2006), can be made more precise with Pnar and War data. Using the Shorto (2006) MK phonology correspondances, I have found some 23% lexical similarity between Palaungic and Pnaric or War, especially with some specific isoglosses between War and Palaung, Riang and Wa. Old Mon, Modern Mon and Pnaric or War have some 20% similarity. Using this comparison and annex 2 comparative word list (250 words) with Shorto (2006), specific isoglosses between Palaungic and PWL appear to be few and not significatively more than with Monic. Similarities between Pnaric and Palaungic usually also involve Khmuic. These results fit another comparison previously made on Annex 2 between S. Khasi, East and West Pnar, War and Lyngngam which I have compared with AA cognates analysed in Shorto (2006). This list includes most of the elements of an AA comparative Swadesh list, kindly sent to me by Paul Sidwell (I have removed some items leading to ambiguities in PWL). Comparing the PWL sample in annex 2 with this AA sample, 45% have cognates with either: 1) Munda, 2) Central MK, Monic or Aslian or 3) both MK and Munda. PWL data of annex 2, show only around 44% similarity between Kudeng War and Standard Khasi and around 51% similarity between Langkymma Lyngngam and S. Khasi. Dissimilarities are still important according to the growing influence of S. Khasi. War and Lyngngam are closer to Pnar than to Khasi: around 65% similarity between War and East Pnar conservative varieties and around 68% similarity between Langkymma Lyngngam and West Pnar. Conservative East Pnar and S. Khasi have around 74% similarity. The around 84% similarity between the western most and easternmost varieties of Pnar: Ralliang and Nobosopoh, shows the remaining extension of Pnar in the West Khasi hills, though the syntax of western Pnar varieties are now very influenced by S. Khasi. The analysis of lexical elements divergent from Pnar in War and in Lyngngam reveals interesting connections with different AA groups, see §5 and §6.

Several cognates listed in the table of unique pair sets Palaungic-Khasian and in the table of shared innovations of Paul Sidwell appear to have isoglosses with other AA languages or to raise questions on reconstruction methods. I give a few exemples here.

Pnaric has *baːm* 'eat, food', War *bua, ba, bu* 'eat, food' related by Shorto (2006) to Palaung *bam* 'chew'. In addition to the data of Shorto, this cognate is probably also connected to Munda, Bonda (*rɔri*) *bai* 'eat (much)', *bun* 'eat' in Parengi (South Munda) (Bhattacharya 1968); South and North Munda have 'eat belly-fully' *bʔe* in Sora, *bi* in Santali, Mundari, Ho, Korku. Following a suggestion of Shorto (2006) *bam 'eat, chew' is probably related to AA **[kV]bam* 'to hold in one's mouth, to close the lips. Katuic: Kuy *bɔm* and Bru *kuboːm* 'hold in one's mouth', Khasi *kbum* to close one's lips.

Nobosopoh Pnar has *saː* 'eat' and Ralliang Pnar has kept *saː* as an imitative of *bam*, a much older and widespread element for 'eat' in AA.

Shorto (2006: 1298) analyzes **ʔ[o]m* 'water' in Kammu-Yuan, Khasi and Palaungic. Adding new data this cognate extends to Khmuic, Mangic and Aslian, see §8.

*kɲiaŋ generic term for insects and larva in PWL is cognate with Palaung and Riang-Lang *ɲaŋ* 'larva'. *kəníaŋ* 'to fly' in Bahnaric, Sedang is probably related to a MK cognate *kənjaŋ 'flying insects, larva'.

snaːm 'blood' raises difficult reconstruction questions and cannot be considered as a specific isogloss in PWL and Palaungic for reasons analysed in §8.

AA *suk 'hair' analysed by Shorto (2006) can have an infixation *[-ɲ-]* found in Pnaric, Lyngngam *sɲɔk 'hair' and Danaw *ɲok¹* 'hair'. This infixation of *-ɲ-* is probably related to the *-n-* infixation found in Bahnaric, Sre *sənoː* , Stieng *sonɔːk* 'hair', see also *n-* in Khmuic, Mal *nsoːk*. Interestingly this *n* affixation does not appear in War *suʔ* and in Palaung, Riang and Wa groups (see annex1 and 2). AA *suk 'hair is found without *n* affixation in most of the AA languages, in Aslian, Bahnaric, Katuic, Khmeric, Khmuic, Mangic, Monic, Nicobaric, Palaung, Vietic and probably in Munda. Sora *uʔu* 'hair' is probably related to *suk with a reduction of an expressive reduplication as in Vietic, Ruc *uçuk³*.

Specific isoglosses occur between War and Palaungic subgroups: Palaung, Riang and Wa. They are interesting because they appear to be late innovations, for example:

War *ksjaŋ*, Palaungic *jaŋ* 'dung' (see annex1) is a recent shared innovation as Pnaric and Lyngngam have the widespread unrelatated AA cognate *ʔec 'dung' found in Santali *ic*, Remo *ig*, Khmer *ʔaːc*, Modern Mon *ʔik*, Kammu-Yuan *ʔjak*, Svantesson (1983), Pnar *ʔait⁷*, Khasi and Lynngam *ʔeit⁷*.

'male' *tərmaː* in Kudeng War and *kərmeʔ* in Riang are isoglosses of War and Riang but not of Pnaric and Palaungic, see §8.

AA *bar 'two' (*ʔbar/ʔbaːr/ʔba/ubar*) > *ʔaːr/ʔuːr/ʔəːr/ʔɔ̃* in PWL. Most Munda and MK languages have derived forms from *bar for the cardinal 'two', see M.P.I. data bank on numbers (2011). *ʔaːr* 'two' is probably a very recent isogloss in Pnaric, War and Riang for reasons summarized in §4.2.

Interestingly, Pnaric and Palaungic cognates with no War connection in the word list of Luce, in annex 1, only concern the names of the rice, Pnaric *kuba* 'paddy' and *kʰoː* 'rice grain (non husked)', two widespread AA names, while War has innovated *hətʒi* 'paddy' ('cooked rice' *tʒi* < *dʒa* in Pnaric) and *rhija* 'rice grain (non husked)' from the name in Pnaric *kʰo rhijum*, a mountain variety brown or white, one of the Pnar varieties cultivated on hills. Untill very recently the staple food of Wars was millet, grown on edges of tree and creeper plantations.

2. Overview of historical and geographic data on Pnar, War, Khasi and Lyngngam languages and their mixed varieties

Baruah (1985: 35) mentions Chinese commercial and diplomatic contacts with Assam through a Northern route (Patkoi Range). Coedès (1989:61) describes Chinese extended contacts with the Khmers, the Mons and Indonesia especially through a Southern sea route, both starting in the second century B.C., from Chinese sources. A Hinduised Pnar kingdom was perhaps already settled on a converging trade route in Assam at the beginning of our era or a little before at the convergence of these two trade routes. This kingdom might have attracted from different areas small groups practising jhum cultivation along rivers, like the Wars and the Lyngngams.

Barua (1985) mentions a Bhauma Kara dynasty in 737 AD with a maternal side as important as the paternal one, suggesting Hindu and Pnar allied families (*kara* might be related to $k^h ara$ 'mankind' in Pnaric, something like: "Bauma of the the Pnaric mankind").

The PWL in Assam and Bengal have been at a crossing point of two different influence routes: a) a north-eastern Sino-Tibeto-Tai-AA influence route linking Assam, upper Burma and Yunnan, b) a south-eastern Sino-austro-Thai-AA influence route linking the gulf of Bengal, lower Burma, Indonesia and the gulf of Tonkin. A Sino-Tibeto-Tai contact situation with Pnar (or Pnaric) in Assam might be associated with an early Hinduised kingdom in Assam.

In the 13th century Assam is invaded by the Moguls; the Tai Ahom cross the Patkoi range, reach Assam probably in this period and push away the Moguls in Bangladesh. Shadap-Sen (1981) summarizes Ahom chronicles starting in the 15th century, describing alliances and fights with Pnar kingdoms, also battles with the Koch and the moves of the Pnar who leave the area of Gauhati and settle first in Nowgong in North Assam, in the North of Shillet in Bangladesh and in East Meghalaya. We know from the Ahom chronicles that most of the Pnar kingdom was still settled in Assam and in Bangladesh in the 15th century. This chronicles describe Pnar kingdoms, called Jaintia, in the Assam plains between the Kupili and the Kalang rivers and in the adjacent "Jaintia hills" of Meghalaya and Cachar hills, with their capital in Nowgong. When the Pnars extended their territory over the Jaintia Hills in Meghalaya they settled their capital in Jaintiapur in Shyllet. The Pnars already had an elaborate administration. As stated by Baruah (1985), they had merchants actively trading both with Bouthan and Tibet and with the Moguls of Bengal. Pnar peasants were settled in fixed villages growing rice on wet rice fields among other edibles which were exported. Pnar trades with Moguls used four routes, three land routes including one across Meghalaya and one sea route. The Pnars settled in the East of Meghalaya first around Sutnga and spread in the Jaintia Hills. Shadap-Sen (1981) states that in the 18th century rich Khasi merchants set three Khasi districts, west of the Pnar lands (now called East Khasi Hills). In the 17th and 19th century, Pnar clans from Jowai settled in the North of Meghalaya and in the West of the Khasi districts (now called West Khasi Hills). They allied with some Lyngngam groups in Langrin and Nongstoin and created various other doloiship districts all over the so-called West Khasi Hills.

It is possible to trace back from actual clan connections among Wars in Meghalaya that many Wars were settled in fixed villages in the hills of Bangladesh up to Tripura border; there are still groups of Wars in North Assam, Cachar, Manipur and Bangladesh. Wars traditionally cultivate for trade betel nut, citrus, paan leaves and peper in plantations on hill slopes.

Wars and Lyngngams never had any kingdoms and urban centres as opposed to Pnars and Khasis. Up to very recent times, Lyngngams have been *jhum* cultivators without fixed villages, feeding on millet, tubers, job's tear, some mountain rice and raising some poultry and pigs in a deprived area of West Meghalaya and Bangladesh, West of Garo territories.

Kharakor (1951) gives a map showing the repartition of MK communities in Meghalaya at the arrival of the British, which shows that Khasi population and territories in their "doloiships" (traditional districts based on clan covenants): Sohra, Mylliem and Khyrim were a minority compared to Pnar "doloiships". The districts corresponded to different oral sub-groups. Some of these sub-groups have faded or mixed with other sub-groups, as described below.

PWL has arround 1,300,000 speakers in Meghalaya according to the 2006 electoral roles. Khasi mixed varieties are all the varieties closer to S.Khasi than to conservative varieties of Pnar, War, Lyngngam or other TB or IA languages. S. Khasi and Khasi mixed varieties are spoken by around 400,000 speakers. Most Khasis stay in Shillong (267,000 persons in 2001 including different communities). Different War-Khasi or Khasi-War varieties now very close to S. Khasi are spoken in South Khasi Hills by some 120,000 speakers. These speakers are the descendants of a former War community from Bangladesh who settled on Khyrim Khasi lands.

Conservative Pnar and Pnar composite varieties are spoken by around 700,000 speakers. The main dialects of Pnar in the Jaintia Hills and in the West Khasi Hills now regroup former different East and West dialects listed as "doloiships" by Kharakor (1951). The Pnar East group is mostly located in the Jaintia Hills with isolated groups remaining in Assam (in the Karbi Anglong), in the North Cachar Hills and in Jaintiapur in Bangladesh. The East group has a morpho-syntax very different from Khasi. It has a standard written Jowai Pnar orthography. The main subgroups of East Pnar are: 1) Sutnga, Narpuh, Lakadong, 2) Ralliang, Shangpung, 3) Jowai, Rymbai, 4) Nongbah, Nongjini, Nartiang, 5) Mynso, Thadbamon. East Pnar also has different Pnar composite varieties in the northern Bhoi district. West Pnar dialects have different loans from Lyngngam and are now variously influenced by S. Khasi to some important extent in their morpho-syntax. West Pnar still has subgroups: 1) Rambrai, Myriaw, Mawïang; 2) Nobosopoh; 3) Maram (Maharam). The fact that Rymbrai is no more in contact with Lyngngam groups, the fact that small isolated Lyngngam groups are found in Pnar speaking Nobosohpoh and Ma(ha)ram districts, and the fact that conservative Lyngngams did not have doloiships, indicates that Lyngngam groups were already settled in this western area of the Meghalaya plateau before the settlement of Western Pnar doloiships. Langrin is a Lyngngam-Pnar composite variety spoken in a doloiship created by allied Pnar and Lyngngam clans and Nongstoin, now the main city in the west, mostly has Pnar composite varieties speakers.

Conservative Lyngngam and Lyngngam composite varieties are spoken mainly in Meghalaya in the so-called West Khasi Hills perhaps by around 70,000 speakers and in Bangladesh by around 2,500 speakers. Lyngngam and Garo have inter-maried in some villages where they speak mixed varieties Garo-Lyngngam or Lyngngam-Garo.

Lyngngam has several conservative subgroups: Tra, Langkymma, Rongrin and Dygir and several composite varieties influenced by Pnar including Langrin and Nongstoin. Nagaraja (1993) presented a first survey of Langrin, showing that it has typical features of what appear to be Pnar, such as a negation in *re/rje* (see Table 3), that he could not

recognize as such since he did not know Pnar, but which he rightly compares with Mnar, which turns out to be a Jowai dialect of Pnar and not a Bhoi language.

Conservative War has around 65,000 speakers according to clan names in the 2006 electoral roles (Wars settled in Bangladesh are allowed to vote in India) and an estimation of both isolated groups and War speakers who stay in Jowai and in Shillong. It is mainly spoken in the South East of Meghalaya but also in Bangladesh and in the Bhoi district. There are also small War groups all over Assam (not studied) and small conservative groups on the Eastern hills in Bangladesh. I classify conservative War into three main groups: Nongtalang-Nongbareh, Amvi and Satpator, on the basis of lexical word lists, phonological features, and morpho-syntactic features, especially pronominal and deictic systems, negative assertive features and other grammatical markers. Amvi and Satpator groups settled first in the Jaintia Hills (a group of Amwi settled also in the North, in Jirang). Dialectal variation inside the Nongtalang-Nongbareh group is more important and Pnar influence less than in the Amvi area. Pnar influence in Amwi is less important than in Satpator. There are varieties of War close to Nongbareh-Nongtalang in South-East Bangladesh, on the border with Tripura especially at Mawlavi and Brahman Baria.

3. Overview of some phonological features of S.Khasi and conservative Pnar, Lyngngam and War varieties

There are no major differences in the phonotactics of Pnar, War, Lyngngam and S. Khasi but there are major differences in their intonation systems, Khasi being the most innovative with no or little word accentuation and final accentuation of utterances.

Haudricourt (1965) accounts for transformation of voiceless stops in MK. The transformation of voiceless stops into aspirated voiceless stops occurs in PWL but is partial: it occurs sometimes differently in Khasi, in Pnar and in War. For example *kʰia* 'cucumber in S. Khasi and in Pnar but *[t]kuəʔ* in War are cognate with AA **[t]kiəl* 'cucumber' reconstructed by Shorto (2006:1710). Pnar and Khasi have *kəllut* 'deaf' and War *kʰlɔt* 'deaf'. Partial process of aspiration of voiceless stops is found in different groups, especially Khmuic and Aslian groups, for example in Khmuic, for 'fish' Ksinmul has *ka* and Mal *khaː*; for 'hand', Mlabri has *tiʔ*, Mal *thɨ*, in Aslian, for hand, Temiar has *tiːk*, Semelai *thi*. Another related point is that this transformation of voiceless stops is infrequent in pre-syllables in *kV* and *tV* and does not occur in prefixes like causative prefixes *p-, pən-, təm-* in PWL.

There are different devoicing features in Pnar, in War and in Khasi and Lyngngam has /g/ corresponding to a secondary voicing. This /g/ might result from an influence of Garo; it is found in Garo loans like *daːogɛp* 'duck', perhaps also an influence from Kherwar Munda groups, still settled in Bangladesh close to West Meghalaya. Lyngngam still has specific isoglosses with Kherwar languages, see §5. This /g/ is not a result of unpacking nasals and not a prefix of recent origin in: *gte* 'that'. Deictic pronouns do not occur without a pronominal base in Lyngngam as in Pnaric and War, as shown in §4, in annex 2 and below. Third person pronoun feminine *ka* in Pnaric, *kə* in War precedes the deictics expressing 'that one (fem.) more or less far' and 'this one'. Langkymma Lyngngam has /g/ in *gtu* 'that one far' and in *gteːʔ* 'that one very far away but still in view' also in *gni* 'this one', where /g/ is a short form of *gʲu* 'she, her', see table 3 in §4, which shows that /g/ and /k/ may alternate in Lyngngam pronouns; the third person plural is *kiju*. Hence /g/ in 'she, her' might be a sandhi form (see alternation *k/g* in the names of mother,

female, elder sister in annex 2). Some words with /g/ like *əgap* 'mouth', which is not a loan from Garo (see Burling 2005) but differs from Pnar and War, remains to be accounted for.

Table 1: Some deictics in PWL

gloss	Khasi	R. Pnar	N. Pnar	K. War	L. Lyngngam
this one fem.	ka＝ne	ka＝ni	ka＝neːu	kə＝nə	g＝ni
that one fem. far, very far, out of view	ka＝tu, ka＝taj, ka＝ta	ka＝tu, ka＝taj, ka＝te	ka＝paːu , ka paːu ha ɟŋej	kə＝tə, kə＝tun, kə＝tutun	g＝ tu, g＝ teːʔ

Pnar still has a voiceless palatal affricate /tʃ/, sometimes pronounced as a stop /c/, lost in S. Khasi. Rabel (1961) has shown that the voiced palatal stop and the voiced palatal affricate are allophones in Khasi. The same can be said in Jowai Pnar.

Pnar, War and Lyngngam have long vowels as in Khasi and also many diphthongs, which may vary in sub-groups of each language, see Annex 2 and table 6 for examples. There are independent phonological innovations in Pnar, in War, in Lyngngam and in Khasi. For example, in Pnar the loss of /m/ or /b/ in onset position of monosyllabic words is frequent, as in *mi > wi* 'one', *ba > wa* 'dependency marker'. This loss is only found in loans from Pnar, like *mi* 'one'< *wi, wej* 'one' in Pnar, in Khasi and in Lyngngam. Hence Khasi and Lyngngam cardinal systems are derived from Pnar, see table 6 in Annex 2. Generally speaking, Pnar cognates are more conservative than Khasi from an AA view point. War has lost the voiced palatal affricate of Pnar and Khasi but it has a voiced labio-dental affricate lost in Pnar and Khasi.

Dialectal variation is more important in conservative War than in conservative Pnar, especially in Nongbareh-Nongtalang; there are important differences in vowel nasalisation, diphthongs, vowel length and vowel change and also in pronominal features especially. These variations can be explained by different recent migrations in Meghalaya of isolated groups from Assam and Bangladesh.

As analysed by Henderson (1976) there are violations of the dissimilation law of Greenberg in Khasi like *dpej* 'ash'; it also happens in West Pnar influenced by Khasi, for example *dpeɲ* 'hearth, ash' in Nobosohpoh Pnar but *tpaj* in Ralliang Pnar. *pu.seːa* 'ash' (*seːa* 'red') in War, *əpaːo* 'ash' in Lyngngam are cognate with Shorto (2006: 2034) **buh; *buəh* 'ash, dust'. Dissimilation in *dpej* might be an innovative feature of Khasi.

Different clusters of consonants in onset position are found in Pnaric, War and Lyngngam as a prominent feature of this group. Rather than considering them as random presyllables, I analyse some of them as reductions of frozen word-formatives, kind of classifiers, prefixed, see §8. Unlike most MK languages but like Nicobarese and Munda, PWL still have productive suffixes like the depreciative *-siʔ* on verbs.

Table 2: War consonants and PWL vowels

War Consonants	labial	labio-dental	dental	alveolar	palatal	velar	glottal
voiceless stops	p		t			k	ʔ
voiced stops	b		d				
aspirated voiceless stops	pʰ		tʰ			kʰ	
voiceless fricatives				s	ʃ		h
voiced fricatives		v					
voiceless affricates					ʧ tʒ		
lateral				l			
flap				r			
nasals	m		n	ɲ		ŋ	
semi-vowels	w				j		

PWL vowels	front	Centre	Back
close	i	ɨ	u
mid close	e		o
mid open	ɛ	ə	ɔ
open	a		

War, like Pnar, Lyngngam and Khasi has long vowels (all but /ɨ/) and has many diphtongs which vary inside dialectal groups (especially inside the Nongtalang-Nongbareh group).

/e, o/ are not infrequent, especially before a glottal and inside diphthongues for exemple in War *eʔ* 'can', *beʔ* 'run in pursuit of something, *deːu* 'fairie' *boːu* many, much, *ʧəproːu* person. /eː/ also appears in rhymes like *kreː* 'millet' in Pnar (Khasi *kraj* 'millet'), see also *pero* 'brothers and sisters from the same mother' in War.

War has nasalized vowels; Nongbareh village and Lamin have a greater tendency to nasalize vowels than other War sub-groups. Nongtalang village replaces the current alveolar fricative by a dental fricative.

There are several phonetic differences between War, Khasi and Pnar which are isoglosses of conservative War dialects.

In War, the voiced fricative palatal occurs only in words which have been recently borrowed from Khasi, like *ɟiŋmut* 'meaning' with the Khasi innovative prefix *ɟiŋ* for abstract nouns (used in a different way in Pnar).

War is the only language of PWL which has a phonemic opposition between two fricatives /ʧ/ and /tʒ/. /tʒ/ is in fact a devoiced palatal affricate with a smaller quantity of friction than /ʧ/. This feature appears when both sounds are analyzed using PRAAT software in opposing words, for example in *tʒia* 'to happen' as opposed to *ʧia' (hi)* 'fish bone', *tʒi* 'cooked rice' as opposed to *ʧi!* a depreciative interjection. This innovative opposition of affricates is an isogloss of conservative War: Amvi, Nongtalang-Nongbareh and Satpator War. /tʒ/ is frequent in War and appears systematically in words cognate to words with a voiced palatal (or voiced affricate) in Pnar and in Khasi like *dʒan* 'near' in Pnar and in Khasi, *tʒan* 'near' in War; *tʒi* 'cooked rice' in War, *dʒa* in Pnar and in Khasi.

4. Khasi as a Pnaric language and Pnar, War, Lyngngam specific AA features in negation systems, pronominal systems and number systems

Annex 2 shows that Khasi lexicon is very close to Pnar with a few more IA and Mogul loans and a few innovations. Using mainly Shorto (2006) and additional information on Munda as reference works, different conservative AA cognates cognates are found in Pnaric, in War and in Lyngngam as will be shown in §5 and §6.

Table 3 shows that the S. Khasi system of negation is very simplified morphologically and semantically compared to the Pnar system. Pnar, War and Lyngngam negation systems are complex with different AA isoglosses.

PWL cardinal systems are presented in annex 3; the S. Khasi system is borrowed from Pnar. A common pre-cardinal number system is more conservative

Table 6 shows that the Khasi personnal pronouns mostly correspond to the non focus Pnar sequence of pronons. Table 6 also shows that PWL has developed a gender opposition for second and third person pronouns with different morphologies in Pnaric, in War and in Lyngngam, which show different AA isoglosses.

Table 7 shows that Pnaric and War secondary renewal of their third person pronouns into different gender/number particles marking nouns is not shared by Lyngngam.

4. 1. Comparative morphology of PWL negations and AA negations with semantic shifts

Table 3: The different assertive negation systems of Pnar, War, Lyngngam and Standard Khasi from a morphological and semantic viewpoint.

Table 2 PWKL Assertive Negations systems	East Pnar	Kudeng N. War	Langkyma Lyngngam	S. Khasi
plain negation	V **re** (post verbal)	**hn-V**	V **(ən)ɟi** (post verbal)	. S PRO-**ˀm** V . S PRO-n **ˀəm** V (potential)
did not happen but might happen	V **ˀɛm.re** cop. Neg	**ˀaː. tʒu** V NEG. CONS	∅	**ˀəm ʃim** V NEG NEG-PAST
not yet but should happen	V **puˀ.re**	**ˀaː.tʒu.pʰuˀ** V	V **(ən)traˀj**	S PRO-**ˀm put** V
emphatic not yet	V **ˀɛm. puˀ.re**	∅	∅	∅
neg future	∅	V **tə**	∅	∅
not anymore when it has taken place	V **deː. re**	V **tə. wan** N FUT. come	V **(ən)ɟɛt**	∅
impossible (not physical)	V **jeˀ… re**	**man. jə. tə** become.it. NFUT		**bə.m.laˀ.ban.loŋ** V DEP.NEG.ABIL.COM.BE V
cannot stop doing			V **ənɟi**.rə NPAST. CONT	
absolute willingness	**Vdam … deː. re**	∅		
it is inappropriate to	**hɔj.re**	**hɔj.tə**	**dəu. ənɟi**	
not knowing if it will take place	**jeˀ…dam**	∅		
mirative neg	∅	**mə**	∅	∅
never (deliberately)	∅	**kaˀ**		
never in past	**hi …re**	∅		
never in future	**ɟo …re**	∅		

unwillingness to answer or unknown answer		tɔʔ. laʔ know. INDEF P		
denial	tɔʔ.re	tɔʔ.tə		
emphatic denial	hɔʔ	hɔʔ		
precative do not	bɔj V re	bɔj V tə	V re	
precative do not anymore	bɔj V deːɪ. re	bɔj V tə.wan	V re.dɨn	
prohibitive	ham V	V tə knə.ə Amvi War	V re.eʔ	wat V
harsh prohibitive	dam V	dam V Nongtalang War V tə klə.ə Amvi War	V re.eʔ	
emphatic confirmation	ɔ.re	tɔʔ		
Neg in complement clause	wɔm	∅	∅	bɔm
without doing	khlɛm	khlɛm	khlim	khlɛm

The negation system of War is analysed in details in Daladier (2011a). Milne (1921) describes many negators in Rumai Palaung in different chapters of her grammar and in her dictionary.

Kruspe (2004) analyses complex negations in Semlai, Aslian. A complex assertive negation system is outlined in Kammu by Svantesson (1983:78). Shorto (1971) analyses four assertive negations in Old Mon. Complex assertive negation systems are also described in Munda, especially in conservative South Munda languages, like Gutob, Juang, Kharia, Sora but also in Kherwar, in Bodding (1929), Ramamurti (1931) and in Anderson ed. (2008).

As in PWL, many South Munda languages have rich negation systems expressing tenses and assertive modalities not always expressed in their positive declarative systems. Some negations may be attached to pronoun clitics. They may express emphatic denials and affirmations. They have emphatic forms combined with copula verbs. War even uses several other grammaticalized verbs like *wan* 'come', *beʔ* 'chase', *tɔʔ* 'know' in its modal negations, denials and emphatic affirmations (see Daladier 2012a). Most of the semantic assertive features and most of the negative morphology used in PWL negation systems are also found in different groups in Munda. Some of the morphology and semantic features are found in conservative MK languages.

bɔj is a prohibitive marker in Semlai (Aslian) as in Pnar and in War.

ʔa in they negative past, *ʔa tʒu* 'not up to now' of War is most probably to Sora *ʔa* and to *ʔaʔ* in Sre Bahnar which can combine with different particles to express 'not yet', emphatic negation and prohibitive negation, see Ramamurti (1931) and Manley (1972).

The assertive negative future *tə* in War is most probably related to the denial particle *-ted* and present future *-tə* in Sora and the prohibitive use of War *tə* in *bɔj tə* with *taː* prohibitive in Kammu (Svantesson 1983:78).

re main negation in Pnar and prohibitive particle in Lyngngam might be related to South Munda negations: *(a)r, ar* main negation in Gorum, Remo and Gtaʔ, to *ɔra* 'not indeed' in Bonda (Bhattacharya 1968). Gutob combines *ar-* and *ur-* with other elements to produce the values of past, future, 'not yet' negations and negative ability, see Anderson

ed. (2008). Juang combines *ari* and *ɟena*, Patnaik (2008). In North Munda, Ho has *auri* 'not yet', see Anderson (2008).

As analyzed by Shorto (2006:1297) Khasi main negation *ʔəm, -ʔm* is cognate with *ʔam 'not' in Kammu Thin and *ʔUu*. The simplification of Khasi negation system as compared to Pnar, War and Lyngngam is shown by the many blanks in Table 3 and by the fact that *ʔəm* is found in specialized uses in PWL. *ʔəm* is suffixed in PWL *khlɛm 'without something, without doing something' and in *b/w -əm* negation of a dependent clause in Pnar and in Khasi. Pnar also has *ʔɛm* in *pu ʔɛm re* 'not yet'. *um* main negation in Kharia, see Peterson in Anderson ed. (2008) and *ama* in Juang might be related to this MK negation unless they are borrowed from an IA negative particle *ma* derived from *ma* prohibitive in Sanskrit.

hɔ emphatic negation in Pnar and in War is probably related to *ɔhɔ* emphatic negation in Santali, see Bodding (1929). War also has *hən-* plain negation prefixed to verbs which might be related to *hə* negative particle 'not' in Old Mon and to the strengthening negative particle *hwaʔ* in lit. Mon, Shorto (1971).

ənɟi actualized negation and *ənɟɛt* negation future or potential in Lyngngam (with *-t* probably related to *tə* negation future or potential in War) is probably related to *ɟena* main negation in Juang (see Patnaik in Anderson 2008 ed.) and to *ɟa* 'no one' in Bonda see Battacharya (1968); *kaɟe kaɟe* 'no one' in Mawroh Pnar (Pnar spoken in the Bhoi district) with *ka* indefinite pronoun might be related to this negation. In East Pnar *ɟo …re* 'never (in the future) combines its plain *re* negation with *ɟo* probably also related with Lyngngam and Munda negations.

kaʔ expresses 'never' with deliberate intent in War and might be related by metathesis to *-ak* negation past in Sora, Anderson (2008) in Anderson ed.(2008) . *ka* is the main negation in Mundari and in Palaung, see Osada (2008) in Anderson ed. and Milne (1921). Mundari and Palaung both have rather simple negation systems.

Complex negation systems are analysed in conservative South Munda languages in Anderson ed. (2008). Svantesson (1983) states that Khmuic sub-groups have rich and diverse negation systems.They might be vestiges of conservative AA features. A thorough comparison of Khmuic negation systems with the complex systems of Pnar, War and Lyngngam remains to be done.

4.2. Pnaric and War cardinal systems. Conservative pre-cardinal number system in PWL

PWL cardinal systems are presented in table 6 in annex (first hand data). This table shows that there are two systems in PWL, a Pnar one and a War one. the Khasi system of cardinals is derived from the Pnar one. 'One' *wei* < *mi* has *w* < *m* ; *w* < *m* and *w* < *b* in onset position is a specific innovation of Pnar. The Lyngngam cardinals are also borrowed from Pnar but they have in addition two suffixed classifiers for people which depend on numbers up to 'ten'. The two number classifiers for people in Langkymma Lyngngam, *-re* and *-də*, are connected with Gtaʔ *-re* and *-de* (South Munda) with *de* also suffixed to 'five' as a vestige of a quinary base, see Zide (1978). *-r* is found in people's name in War and in Munda and *ra is found in Bahnaric as a classifier for people, see Jacq and Sidwell (2000).

While PWL cardinal systems are rather innovative, with Bodish and Pnaric mutual loans, see Daladier 2010-to appear, Matisoff (1997) and Annex 6, a common pre-cardinal system using "groupings" in PWL appear to be very conservative in AA. I have compared

first hand data on "grouping"[84] number systems still used in Pnaric, War and Lyngngam with AA cardinal systems. PWL grouping number words together with their numeration bases four, five and twenty appear to be very conservative as they have vestigial cognates over the whole AA family. Isoglosses confirms historical data and show that PWL groups were settled in Assam and Bangladesh around the beginning of our era at a crossing point of two South-eastern and North-eastern routes of influence: a) Sino-Austronesian-Tai-AA with the Hinduised kingdoms related to Fu-nan and b) Sino-Tibeto-Tai-AA with the upper Burma route to Yun-nan (Daladier 2011- to appear).

The PWL cardinal numeral *ʔ*ar 'two' has the same origin than the PWL conservative counting unit *bar*/*bʰar* used in PWL to count a measure of two *pəntrɔʔ* of betel nuts. This unit *bar* has probably been used in AA before it was transformed into cardinal 'two' like five other counting units using different numeration bases.

The names of six PWL counting units: *mɔn, bar, pɔn, soŋ, kti/taː* and *kuri* are widely found again as cardinal names in AA with values related to their numeration base in PWL e.g. *mi/mɔn /m- ,*'one' is also used for 'one' as a counter of fives in *m-sun* 'one-five' in Mon, *sun*/*san/soŋ/saŋ* 'five' both in Munda and in MK on all the AA geographic era (Daladier 2010-in prep). PWL counting units are very close to the reconstructed roots of many numbers of decimal cardinal systems by Zide (1978), by Jenner (1976) and by Thomas (1976), in Munda and in MK. Coedès (1942) and Jenner (1976) show that the former number systems in Old Khmer used quadrennial and quinary bases and vestiges of such bases are also found in Aslian and in Old Mon cardinals. Zide (1978) shows that cardinal number systems in Munda use quadrenial, quinary and vigesimal bases. These numeration bases are also found in TB cardinals and in PWL counting measures. AA cardinal number systems are late comers compared to "grouping" number systems and have probably emerged under contacts with Hindu and Chinese trades and more locally in the Assam corridor with Tai and Bodish trades, around the beginning of our era.

miː//ʃi (or *wiː//tʃi*) represents a PWL innovative contrastive pair. In English, 'one' has different mathematical uses which are disambiguised in Pnaric-War-Lyngngam with *mi* and *tʃi*. *mi* is mainly used as cardinal one, *ʃi/tʃi* is mainly used to count 'one' for measure units and to count 'one' for powers of ten in cardinals. For example, in War *ʃi pʰuːa* 'ten', lit. 'one-ten', *ʃi swaʔ* 'one hundred'. *mi* expresses cardinal 'one' in 'one leave'; *ʃi swaʔ mi* 'one hundred one'. I relate *tʃi* as a counter of tens in PWL to Fu-nan Chinese *tsʔiet* 'eight' via a Thai innovation for 'ten' see Benedict (1942) and MK loans in Old Mon, in Bahnaric and Katuic groups for the cardinal 'ten', see Jenner (1976) and Thomas (1976).

The common use of *ʔaːr* in Pnaric and in Riang (not found in other Palung-Wa languages) is significative of a late contact, perhaps only trade contacts, as most cardinals are different in Riang or Palaungic and in PWL. PWL cardinals have more similarities with O. Mon, which has *baːr* 'two', and with Old Khmer cardinals, see Jenner (1976). Interesting from the view point of other contacts is the specific isogloss for 'three' in PWL, Nicobarese and Palaung-Wa: *laː* in PWL, *lɛ* in Palaung, *loi*/*luɛ²* in Wa, *lohe* in Lemet and *loːe*/*luːe* in Central Nicobar, see Luce (1985); it might be a common loan from Chinese ternary length measure *li* used since the Shang dynasty up to modern times with shifts in its value, 300 *bu* or 360 *bu* (Iprah 1998).

---

[84] The pre-cardinal number notion of "grouping" is analyzed by Menninger (1969) from both a cognitive and historical point of view.

4.3 PWL Personal pronoun systems and their different AA cognates. Renewals into deictic bases and into gender/number particles

Table 3: PWL pronominal systems (Ralliang-Pnar and Kudeng-War double sequence pronouns)

Person pronouns	Pnar focus and object pronouns	Pnar plain pronouns	K. War plain pronouns	K. War focus and object pronouns	Langkyma Lyngngam	S. Khasi
1 S	ŋa	ˀɔ	ŋə	njɛ	nə	ŋa
2 F	pʰa	pʰɔ	hə	eˀahə	pʰe	pʰa
2 M	me	mi	m̥	eˀam̥	mi	me
3 F	ka	kɔ	kə	eˀakə	gɟu	ka
3 M	ˀu	ˀo	ˀu/ˀo	eˀaˀ/eˀaˀo	ɟu	ˀu
1 P	ˀi	ˀi	ʔi	iʔi	haj	ŋi
2 P	pʰi	pʰi	hi	ihi	pʰjaːo	pʰi
3 P	ki	ki	jə/ˀi	ijə	kiju	ki

Table 4: AA personal pronouns Pinnow (1965)

	PWL	Munda	Palaungic	Monic, Khmeric	Aslian	Nicobarese
1 sing. *iŋ	+	+	-	+	+	+
2 sing. *me	+	+	+	+	+	+
2 plur. *pe	+	+	+	+	+	+
3 plur. **ki/ku	+	+	-	+	+	- 3 S k(V)

Personal pronoun systems are interesting as they usually bear conservative features in different language families. This is also the case in AA, and in PWL especially, as will be illustrated in details with first person pronouns and tables 4 and 5. War and Pnar still have two person pronouns series, one for plain pronouns and one for salient and non subject forms, a feature also found in some Munda languages as described by Pinnow (1965).

The plain forms for 'I' in Pnar is *ˀɔ* which can be related to Khmuic, Mlabri *ʔo*, Khmu *ʔoʔ*, Palaungic, U*ʔò*, Rumai *ʔàw*, De'ang *ʔo*, Danaw *ʔoʔ*, Monic, Nyakur *ʔɤ́j*, Mon *ʔoa*, Mangic, Paliu *ʔaːu⁵⁵*, Bugan *ɔ̌³¹*, Bahnaric, Sedang *ʔaw*, Jru *ʔaːj*, Jeh *ʔau*.

The salient form for 'I' in war is *ɲɛ* which can be related by metathesis to Munda forms, Santali *iɲ* Juang *aɲ*, Korku *aɲ*, Mundari, *aiɲ*, Sora *ɲen*, and to MK forms with loss of the initial glottal, Aslian Semai *ʔɛɲ*, Semelai *ʔəɲ*, Bahnaric, Bahnar *ʔiɲ*, Sre *ʔaɲ*, Tampuan *ʔaɲ*, Khmeric, Surin *ʔaɲ*, Khmuic, Ksinmul *ʔaɲ*, Mal *ʔəɲ*, Pearic Chong-H *ʔiɲ*, Kasong *ʔiɲ*, Pear-B *ʔiɲ*.

Pinnow (1965) reconstructs *iŋ 'I' in PAA with *ɲ*- in place of *ŋ*- by assimilation to *i*-. This form is found in all AA groups except Palaungic and Nicobarese. Accordingly, I also relate to this cognate the plain form for 'I' *ŋə* in War, the salient form for 'I' in Pnar *ŋa*, the

unique forms for 'I' in Khasi *ŋa* and the unique form in Lyngngam *nə*, rather than relating them to a TB borrowing *ŋa* 'I'. The similarity may be due to a Himalayan contact area. In addition, 'we' *ŋi* in Khasi is probably related to this proto-form, perhaps as a former dual form, probably related to *ɲiː* in Bahnar.

As suggested by Shorto (2006), other AA forms for 'we' in *ʔiʔ, ʔi, iʔ* might include AA **ʔiiʔ* 'person' to express 'we' as 'we persons'; that would explain the double sequence in War *iʔ* 'we' and *iʔi* 'we-persons' as a salient form. Kammu has *ʔiˀ* 'we' svanteson (1983). Kharia has *iŋ/ɲi* 'person' which shows how person pronouns and infixed forms of 'person' in pronominal bases may be related. Shorto (2006) relates to this **ʔiiʔ* 'person' root the forms for 'I' in Kammu-Yuan *ʔò*, Lawa Umphai *ʔauʔ* to which can be added the forms 'I' *ˀɔ* in Pnar, and all the related forms in Khmuic, Monic, Mangic and Bahnaric mentioned earlier. These forms could be related to AA **iŋ* 'I' as combined salient forms in **iŋ *ʔiiʔ* transformed into **ʔo* or *oʔ* after loss of *iŋ* and diphthonguisation of /i/. Then these forms would fad to the value of a plain 'I' as in Pnar, Palaungic, Khmuic, Monic, Mangic and Bahnaric. Khasi and Lyngngam both have differently innovated focus pronouns by prefixation: in S.Khasi *ma-* and in Langkymma Lyngngam *sə-*. In addition to its salient pronouns, Ralliang Pnar has two kinds of topic pronouns suffixed with *-te* and *-se* (personnal unpublished documentation).

PWL languages have no dual but, except for this feature, Pnar singular person pronouns are closer to Kammu, see Svantesson (1983) than to Palaungic. In addition, Kammu has a pronominal element *joˀ* 'each other' cognate with *ja-* in Pnaric and Lyngngam and *jaˀ-* in War, prefixed to verbs expressing 'each other' and action performed in succession together.

The personal pronoun system of Palaung, except for the isogloss of 'I' with plain pronoun in Pnar, also found in Khmuic, mostly differs from the different pronoun systems in PWL. Its morphology differs and it has dual forms and no gender opposition. The same can be said for the negation system, see Milne (1921:17 and 107-110), though Palaung still has a rich assertive negation system, as in Khmuic, South Munda and PWL.

Table 4 shows that S. Khasi personal pronouns are the same than Pnar focus personal pronouns except for we, us innovative *ŋi*, as Khasi has renewed Pnar and War *ʔi* as a diminutive pronoun for babies, children, dear ones, small things.

The divergent second person singular feminine and second person plural in War is most interesting as it is a specific isogloss with Aslian, perhaps a loan from Indonesian as indicated by Shorto (2006:1436), Semai *hɛːʔ*, Temiar *haːʔ*, Mah Meri *hiʔ* singular and plural.

The masculine/feminine opposition in second and third person pronoun singular is an innovation of a few AA languages. This opposition is realized in Pnaric with simple shifts of the AA second person plural as a second person feminine singular and an opposition of vowels for the third person singular feminine, masculine and plural.

Lyngngam has a different marking for third person feminine, masculine singular and for plural with *-ɟu/ju* suffixed for the third pronoun feminine to /g/secondary voicing of /k/, instead of Pnaric and War /a/ and suffixed to AA /ki/ for the plural. This suffix *-ɟu* is very interesting because it is used in Korku to mark femine gender in kin terms as shown by Zide (2008).

Pnaric and War have inovated a renewal of their gender and number opposition of third person pronouns as gender and number particules preceding nouns. As shown in table

5, S. Khasi has nearly the same system than Pnar, War is more innovative with more specific oppositions. Lyngngam does not have such a system. In Lyngngam *ɟoŋ-* is a loan of S. Khasi *dʒiŋ-* in words like *dʒoŋmut* 'meaning' and in Khasi those words are marked with *ka* feminine.

Table 5: Third person pronouns renewed as gender/number particles before nouns in Pnaric and War but not in Lyngngam

gender/number particles	S. Khasi	Pnar	War	Langkyma Lyngngam
masc. sing.	ʔu	ʔu	ʔu	ʔu with abstract nouns in ɟoŋ-
fem. sing.	ka	ka	kə	∅
plur.	ki	ki	ʔi	∅
	ʔi diminutive	ʔi inalienable possession	ʔi mass. terms and abstract nouns	∅
	∅	∅	ki shared feature of a set of people	∅

Using again their gender/number oppositions in their pronominal person systems Pnaric, War and Lyngngam have developed very precise deictic pronominal bases as shown in Annex 2 and Table 1. This secondary grammaticalisation is very interesting as Pnar, War and Lyngngam did not develop any kind of verbal bases. Tense values are not expressed verbaly in PWL but different tense values are expressed in deictics and in some negation particles in the languages of this group. S. Khasi is the only language of this group which has inovated a kind of verbal base with a pronoun refering to a lexical subject preceding the verb and with a suffixation of negative and potential particles to the subject.

Object and benficiary argument pronouns in S. Khasi and in Lyngngam are plain pronouns marked for these semantic role with particle *ja* in S. Khasi and with *sə* in Lyngngam before the argument. The marking of these roles by *ja* in Pnar and *haʔ* inWar is optional and marks some salience, or definiteness, and accordingly the marking of these roles construct with focal pronouns.

To conclude this section, Khasi is Pnaric, Pnaric and War show significative morphological differences but Lyngngam differs from Pnaric and War in even more important morphological and systemic features.

5. Specific isoglosses of Lyngngam with Kherwar and with South Munda languages
Lyngngam has no specific isoglosses with Palaungic but more phonetic, lexical and morphological similarities with Munda than Pnaric and War: lexical similarities with Kherwar (North Munda) are probably due to relatively recent contacts in the gulf of Bengal. Lexical and morphological similarities with Juang especially, among South Munda languages might have originated earlier. A few lexical and morphological examples of isoglosses with Munda are given here.

(ə)*dɔm* 'hill' in Lyngngam contrasts with Pnaric *lɔːm* 'hill' and with War *pdeŋ* 'hill'. (ə)*dɔm* 'hill' in Lyngngam is probably cognate with Santali *ɖomgi* 'hillock'.

Pnaric *lɔːm* 'hill' is probably cognate with Palaungic: Lawa *plaum* 'hill', see Shorto (2006:1369) *bn₂uum* 'hill'.

War *pdeŋ* 'hill' is probably cognate with Sora *bədoːŋ* 'hill'

Lyngngam 'lips' *ləmɔr* is cognate with Santali *ləmer* 'to move the lips'. Pnaric and War have *tʃəntur* analysed by Shorto (2006: 1629) as cognate with **sdur* 'lips' Central Aslian Sakai *səntur* 'lips', Palaungic, Riang-Lang *-tor*.

Lyngngam 'go' *diʔ/denniʔ*, unrelated with Pnaric and War **laː* 'go', is probably related to 'come' in Juang (South Munda) *ḍen* 'come'.

Khasi and Pnar have *kpoʔ* 'belly' and War *pɔʔ* 'belly' which can be related to the MK cognate **boʃ ʃk* 'belly' that Shorto (2006:358) analyses in Katuic, Khasi and Nicobarese and further analyses after Benedict (1942) as a borrowing from Archaic Chineese *piôk* 'belly', also found in TB as **puʾk* and **buʾk*.

Lyngngam has *khlao* 'belly' perhaps connected with Bahnaric, Sedang *kléa*, Bahnar *klaːk*, Jru *klaːk*, Nhaheun *klaːk*

Lyngngam *maːo.dar* 'rock' (lit. stone.rock) is probably related to Santali *diri* 'stone', departing from Pnaric and War **mu.sjaŋ* 'rock' (lit. stone.rock) which contain the cognate AA **ʃ ʃmuː* 'stone', also found in PWL.

Shorto (2006) reconstructs **rmit, rmiːt, rmiət* 'Curcuma species, yellow' in Aslian, Bahnaric, Monic, Palaungic, Khmeric and Katuic. Pnar and War *[tʃɛ]rmit* 'Curcuma' is related to this cognate. Khasi *ʃənraj* is an innovation but it has *lmit-lmit* very yellow, an attribute name with reduplication of **rmit* 'turmeric'. Bahnaric 'yellow' is also often expressed as an attribute name of turmeric with **ʃ ʃrmit*, as in Sre *rmit*, Stieng *rmɨt*, Tampuan *krmɨt*, Jru and Nhaheun *hmeːt* 'yellow'. The Pnar and War names for 'yellow' are unrelated.

Lyngngam *tʃʔiaŋ tim* 'Curcuma' (lit. 'bone liver') contains the name of the bone and *tim* 'liver' probably for both its yellow colour and for its reproduction features without fecundation according to PWL world views. The connection between the name of the curcuma and the name of the bone is found in Bonda (South Munda). Bonda has *sik saŋ* 'bone' and *saŋsaŋ* 'turmeric'. Sora, Gorum, Kharia have *saŋsaŋ* 'turmeric', Remo, Mundari have *sasaŋ* 'turmeric'. The morphological relationship between the bones and the ginger and turmeric rhizomes reflects especially the way they reproduce from their rhizomes-roots, which link them to the clan reproduction properties of the bones in Munda and in MK, see Daladier (2007). It also reflects its lasting properties when dry.

Lyngngam *kap* 'bite' and its cognates in Munda are analyzed in §6.

Minor syllable *ɟ(V)* is found in MK and in Munda e.g. Korku *ɟumuː* 'stone' and in Lyngngam but rare in Pnaric and not found in War which has lost the voiced palatal plosive and replaces this presyllable by a voiceless velar plosive, for example *ɟəroŋ* 'long' in Pnaric, *ɟiroŋ* in Lyngngam, *kəroŋ* in War. Lyngngam has *ɟba* 'paddy', Pnaric **kŭba* 'paddy'; **ʃ ʃba* 'paddy' is found in Munda and in MK; Lyngngam **[ɟə]ra:* 'millet', Pnaric **kra:* 'millet' (see §6). Presyllabe *l(V)* found in Munda, and perhaps in Nicobaric as *ɛl-*, not found in Pnaric and in War, is found in Lyngngam, e.g. 'nose': Lyngngam *ləːmu*, Khasi *kmut'*, Pnar *kmun/kmut'*, Nicobaric, Car *ɛl-mɛh* 'nose'.

6. Specific isoglosses of War with Northern and Southern MK groups and with Munda

'bite' is expressed with three different cognates in PWL, two AA, one IA.

War *hit* is related to Shorto (2006: 957) **ki:t/kiət* 'bite' in Mon, Palaungic, Khmuic and Vietic: Danaw *kiə̠t³*, Plang *kiaⁱ³³*, Wa *kiat*, Monic, Mon *kit*, Nyakur *kⁱiⁱ*; Khmuic, Mal *khe:t*, Vietic Malieng *ka:t⁴ ³*, Thavung So *kâ:t* (this cognate is related by Shorto to proto-Austronesian **ki:tki:t*)

Lyngngam *kap/kənnap* 'bite' (with indefinite or with definite object) is related to the very widespread AA **kap* 'bite' reconstructed by Shorto (2006:1231) for MK and Munda: Sora *ka:b*, Juang *keḍab*, Korku *kab*. Aslian, Jahai *kap*, Jah-Hut *kap*, Kensiw *kap*, Semai *kap*, Temiar *kab*, Bahnaric Cua *ka:p*, Jeh *kap*, Bahnar *kap*, Jru *kap*, Nhaheun *kap*, Sedang *ka*, Sre *kap*, Tampuan *kap*, Katuic, Bru *kap*, Katu *kap*, Kui *kap*, Nge' *kap*, Pacoh *kap*; Nicobaric, Car *kap*, Nancawri *kə́p*; Vietic, Ruc *kəʌm⁴ ,*

Pnaric uses *ɟkap* as an expressive in adition to **dajt* 'hit'. **dajt* 'hit' does not have any known AA cognate and is most probably an early borrowing of the intransitive form *ḍaçyate* of Sanskrit *ḍamç* 'bite, sting'. Pakrit has *ḍakiba* 'to bite, sting', *ḍakijjay* 'is bitten'. Assamese has inovated a split between *kamur* 'sting' and *dak* 'bite' derrived from Pakrit, wich has only one transitive and intransitive form, Joshi Tamuli p.c.. The Pnar form is closer to the Sanskrit form than to the Assamese one. Interestingly the innovation *kamur* 'bite' in Assamese might be a borrowing from AA **kap* 'bite', see Vietic, Ruc *kəʌm⁴* 'bite'.

War *fʰiɛn* 'cooked, ripe, mature (for things and beings), to know (as a way to master a technique)' is connected to **ciən* 'cooked' reconstructed by Shorto (2006:1137) in Mon, Khmer, Bahnaric, Khmuic, Palaungic, Vietmuong and South Munda, see Shorto (2006). For example: Katuic *cɛ:n* 'to be ripe, cooked', Mon *həcin* 'to cook', Nicobarese *iʃi:an* 'cooked', Kharia *isin*, Sora *ə'sin*, Remo *isiŋ* 'to boil, bake'.

*ʔi 'cooked, ripe, mature' in Pnaric and Lyngngam is unrelated with *fʰiɛn* 'cooked, ripe, mature.

'millet' Pnar k*re:*, Khasi k*raj*, Lyngngam *jəraj*. Pnaric *kraj/kre:* 'millet' is also found in TB languages, Martine Mazaudon p.c.. War has *kraʔ* 'medicinal herbs' and *tʒʰan* 'millet' not found in Pnaric for edible grains or medicinal herbs and probably related to a kind of millet in Santali (Munda). Santali has two main names for different millets related with War and with Pnaric: *ɟanhe* 'paspallum'; *ɟaŋa ira* 'panicum' and *kukra* 'Setaria Glauca', a kind of wild millet. Korku has *koro* 'millet' and Sora *kɔrɔ'j* 'millet' (big millet cultivated like rice). In Monic, Nya Kur has *khraj*/*hah* for two kinds of chumps and cluster plants: Accorus Tatarinivi and Philanthus Taxidifolius. The names of millets in PWL are connected to Mon and Munda and not to the main roots in MK reconstructed by Shorto (2006: 1447; 1834) **skuəj* and **d[b]aw*.

MK **lʔus* 'fat, oil' and MK **kləŋ* 'grease, fat, marrow' in PWL

War *ləʔɔt* is probably related to Shorto (2006:1879) **lʔus* 'fat, grease' in Palaungic and in North and Central Aslian, Semai *lʔu:s*, Temiar *lɛnʔɔs*. (/t/in rhymes in War, is often found in MK cognates with /s/)

Pnaric and Lyngngam *k^hlaɲ* 'grease, fat, marrow' is related to Shorto (2006:928) *kləɲ* in Khmeric, Monic, Palaungic and Vietic: Khmer *klaɲ*, Surin *khlaɲ*, Mon *klɔiŋ*, Nyakur *kəlíɲ*, Ruc *kluɲʔ*.

War has *phɲaj* 'image of a person, shadow' which can be related to *[b]ɲa:j/*[m]ɲaj:* 'person, human being' reconstructed for North Bahnaric and Viet Muong by Shorto (2006): Bahnar *bəɲa:j*, Halang *ɲa:j*. Pnaric has an unrelated cognate for 'image of a person, shadow', Pnar *trad*, Khasi *trud*.

War *maʔ* 'see, look at' from AA *mat* 'eye' is related to the extended cognate *mat* 'see, see in divination practices' in Munda, see *mat.war* in next cognate.

Pnar *jo*, Lyngngam *ɟɔ*, *muɟɔ*, Nobosopoh Pnar *ɟi*, Khasi *ʔi* 'see, look at' is cognate with South Munda, Juang *jo*, Kharia *ɟoʔ*, Bonda *ɟu* and Sora *giɟa* 'see, look at', Santali *ijo!* 'see!' Palaungic De'ang *ɟi*, U *jò*, Wa *jauʔ* and with Bugan *jə³¹*. Shorto (2006:158) *[s]jeeʔ* 'to see' in Palaungic, Khasi, South Bahnaric, Khmuic is extended by PWL and Munda data to *ɟɔ*.

AA *war* 'incantation, flow' is found in War *var*, *war* 'incantation, to break a spell', and in *wa/war* in Munda languages, Santali *səwa* and *maswar* 'worship', Santali *matwara*, Mundari *matwar* 'transe, visions under intoxicating beverages' (see Bodding 1932-5) with *mat* 'eye, to see' used as word formative or as attribute name in words related to divination practices like Amwi War *tmat* 'egg'. *War*, *k^hɛr-war* (see *k^hɛr-* in §8) and *Wa* are probably related to the cognate *war* 'incantation' as 'people of incantations'.
Pnaric has an unrelated cognate *məntɔr* 'incantation, to break a spell'

War has *priŋ* 'black, dark, burnt colour' related to Bahnaric, Sedang *práŋ* 'black'.

Pnaric *jɔŋ* 'black' and Lyngngam *ɘnoŋ* is related to Shorto (2006:654) *sjuəŋ* 'black' South Bahnaric, Sre *soaɲ*, Jru (West Bahnaric) *ʔjoŋ*, Jeh (North Bahnaric) *ʔɲu:ŋ*, Stieng (South Bahnaric) *ɲo:ŋ*, Halang (North Bahnaric) *ɲu:ŋ*, Kammu-Yan (Khmuic) *jiaŋ*

War *səlaŋ* 'white', *[sə]laŋ*, is cognate with South Bahnaric Sre *laŋ* 'white' (in something white and light) and Chrau *səlaŋ* 'clear, light'; Vietic: Muong Koi *tla:ŋ³* 'white'. Pnar, Khasi and Lyngam *leʔ* 'white' is cognate with MK *[kV]la:k* 'white, clear', Katuic Bru *kla:ʔ*, Khmuic: Khmu *klɔ:k*, Ksinmul *luək*, Mal *kluak*, Vietic: Thavung So *lɔk²*.

In War, *k^hlə ʃmɛn* 'star' literally means 'flower of fire' with *k^hlə* 'flower' and *ʃmɛn* 'fire' in War. *ʃmɛn* 'fire' in War is probably a shift from *smʔan* 'fire' reconstructed by Shorto (2006) in Palaungic, Riang Lang *səkmən*, Rumai *smèn*, Plang *sa³¹ məiŋ⁵¹*, De'ang *siman*, Wa *simʔuiŋ*, Khmuic, Mlabri *sam mon;* Monic Nyakur *chəmén*, literary Mon *saman;* Bahnaric, Stieng *səme:ɲ* Sre *səmaɲ*

'Fire' in Pnar *dɨŋ*, in Khasi *diŋ* and in Lyngngam *ədɔɲ* might be related as *[t]ʔuɲ* to Shorto (2006:885) *ʔuɲ* in Bahnaric perhaps connected to 1872 *ʔus* 'fire' found in Aslian, Bahnaric, Katuic and Khmuic.

7. PWL cognates with Aslian and/or Munda or with Khmuic, Aslian and Munda
ma: 'eldest maternal uncle' in PWL and in Munda
Pnar, Khasi, War and Lyngngam use differently as address terms or as kin terms *koɲ* related to Shorto (2006: 893) *[ ]ku:ɲ* 'mother's brother' (widespread in MK and in Munda) and *ɲi:* 'in-laws' in War and in Pnar', *ɲe:u* 'maternal uncle' in War related to

Shorto (2006:58a) *ɲi:* 'maternal uncle' in Katuic and in S. Khasi. In addition, Pnar has *ma:* 'eldest maternal uncle', Lyngngam *mama:* 'eldest maternal uncle' and War (Nongbareh and Amwi) *mama:* 'eldest maternal uncle'. This cognate, not mentioned for other MK languages by Shorto, seems to have specific cognates in Munda. Sora has *ma:ma:ŋ* 'maternal uncle', with duplication of *ma:* as in Lyngngam and War. *kuma 'uncle, address term for uncle' is found all over the kherwari group, Santali *kuman*.

War and Pnaric *bəroʔ*, Lyngngam *prɔk* 'all', PWL *bəroʔ seems to be a specific isogloss with *[sə]bəroh* 'all' in Semelai.

MK **[ ]mu:ɲ* 'tooth'
Pnar, War and Aslian have /lə/ presyllable while Khmeric has /t/: Pnar and War *ləmɛn* 'tooth', Jah-Hut *ləmoɲ*, Semai *lmu:ɲ*, Semelai *ləmɔɲ*, Khmer *tmiɲ*, Surin *thmɛɲ*. *mu:ɲ 'tooth' is found in Lyngngam *əmɔɲ*, in Aslian, Temiar *mɔɲ*, and in Mangic, Mang *miɲ*[5] [1], Paliu *maɲ*[31]. Shorto (2006) reconstructs MK **liɲ, *li:ɲ, *lu:ɲ* 'gums' and relates to it *muɲ* 'tooth'. Apparently Khasi *bniat* 'tooth' is not found in AA.

AA **[t₁]kiəl* 'cucumber' reconstructed in Shorto (2006: 1710) relates cognates in Khasi, Palaungic, Katuic, Nicobarese and Mundari. It extends to PWL and Kherwar: in PWL, War *tkuə* 'cucumber (wild and cultivated)', Pnaric *kʰia* and in Kherwar, Santali, Mundari *tahɛr*, Ho *taer* 'wild cucumber cucumis sativus' (Bodding 1932-7).

mrat 'all species of "moving beings" including people' in Pnaric and in War is probably related to *mraʔ 'person' reconstructed by Shorto (2006:183) in Khmuic, in North Aslian and also in Sora *mar-ən* 'man, male, person in compounds.

PWL *ʔaŋ* 'open the mouth' in annex 1 is related to AA *ʔa:ŋ* analysed by Shorto (2006:484) in Palaungic, South Aslian, Khmuic and Sora in Munda.

PWL *pʰlaŋ* 'thatching grass' in annex 1 is cognate with AA *[p]laŋ* 'thaching grass' analysed by Shorto (2006:749) in Khasi, Khmer, Katuic, Khmuic, Palaungic, Viet Muong, Aslian and South Munda.

PWL **[ ] lit* 'to lick, tongue': Pnar *təllen*, Khasi *tʰɛlliɛt*, Lyngngam *təllɔjt*, War *kʰlitʔ* are related to Shorto (2006:320) **l(n)t₁aak* 'tongue', in Khmer, Bahnaric, Katuic, Khmuic, Palaungic, Khasi and Aslian.

A number of War, Pnar and Lyngngam words presented here in annexes are cognate with entries of Shorto (2006) unrelated with Khasi; in a more general way, data from these language extends the conservative character of this group and its connection to the times of MK and Munda separation.

8. Some open questions for defining relevant cognates in AA comparisons: extension or change of cognates in long term language change and the analysis of sesquisyllabic structures

Independently of the arguments I have been using so far to advocate a PWL group as opposed to a Khasian group, I wish to raise questions about the notion of cognate used in

classification. As is well known, current word lists used in lexico-statistics have irrelevant items which prove to be ambiguous in some languages while relevant oppositions are not always stated in glosses (multiple English terms may correspond to single cognates in some languages). This appears to be true especially in AA languages, which appear to share many cultural features and share especially different ways of naming by attribute according to common world views. For example, as sketched in Table 3, PWL languages express together modalities, subjectivity, aspect and assertive values in negative particles and there is no cognitive negation in the sense of the Predicate Calculus or its recent avatars (see Daladier 2010). To name but a few, knowledge terms and emotion terms in general, often are non iconic with English terms and this question should be better settled in relevant word lists for AA comparisons.

I will address now a more intricate question about a new hypothesis on variable structures containing core cognates and the evolution, or areal change, of cognates on large time scales in unwritten languages. PWL languages have impressive initial clusters of consonants and sesquisyllabic structures are a prominent AA feature. I have tried to understand how the apparently random distributional forms of initials or minor syllables of AA cognates might be related to a coherent underlying morpho-phonological system. I have come to the hypothesis that many structure-groups of minor syllables are frozen reduction forms of a complex system of word-formatives and that these reduced forms may combine with different regular affixes to core cognates. In other words, I hypothecize that many AA sesquisyllabic structures might be analyzed as vestiges from a complex AA system of kinds of classifiers, usually prefixed but sometimes suffixed in Munda and PWL, combined with AA roots. I reconstruct nine nominal formatives to connect PWL cognates to AA reconstructed cognates by Shorto. This analysis takes up and develops the enlightening analysis of Shorto (1963) who shows especially how several minor syllables may be combined in Palaung-AA cognates (Daladier in Prep.). I cannot present this complex system here but I will try to to give hints for this hypothesis on three exemples: the names of male elder, water and blood.

Shorto (2006:1708) reconstructs *krku:l* 'descent group', Khmer *trəko:l*, Old Mon *kirku:l*/*kərku:l* 'family, clan', Gölar Bahnar *khul* 'family, clan'. For Shorto, this cognate is hardly derived from *kula* 'family, caste' in Sanskrit. The Brahmanic Sanskrit *kula* 'caste, family' might be derived with a shift from an AA cognate *kur* 'clan, descent group.' One may question whether *krku:l* contains an expressive duplication as *ma:* in *mama:* 'eldest maternal uncle' in PWL and South Munda. I reconstruct **kur* 'clan, descent group' in Munda and in PWL **kur* 'maternal clan descent'. **kur* 'clan, descent group' used as an attribute name and word formative for 'person' is widely found in Munda as shown by Pinnow (1959:311), Korku *ko:ro:* (*kor-ku* 'the men' with *-ku* plural from third pers. plur., see table 5), Birhor *hɔr* (*bir- hɔr* 'forest people'), Ho *ho*, 'person', Mundari *koɽa*, Kharia *ko:r*, Santali *hɔɽ* 'person'. As a word formative *kur* >*kʰər-* 'people' appears in different Munda and PWL autonyms like *Kʰerwar* for the group: Santali, Mundari and Ho, *kʰər-rim* (lit. people of the tradition) a group of Khasi, *kʰɛr-vi* (lit. people from the river *Vi*) which is the autonym of War *ʔam vi* 'river *Vi*' (Amwi) people. *War* and *kʰɛr-war* are probably related to the same cognate *war* 'flow, incantation'. As a word formative, **kur* is used in different AA names with vowel or ending oppositions for different kinds of person which are related but not cognates like *koɽa* 'boy' and *koɽi* 'girl' in Santali, *kuɽa* 'husband', *kuɽi* 'wife' in Mundari, *kurim* 'male and female spouse' in Pnar, see annex 2. As seen in §7,

mama: 'eldest maternal uncle' in PWL, *ma:maŋ* 'uncle' in Sora has prefix *ku* in Kherwar: *[ku]ma:* 'uncle' which I analyse as a presyllable, reduced form, of the word formative *kur* 'clan'. This *[ku]* occurs also in eddibles species like *kukra* 'Setaria Glauca', a kind of wild millet in Santali (see AA *kra:* 'millet' in §5).

Shorto (2006:699) reconstructs **dₐraŋ* 'horn' connecting as cognates word structures such as: *ruŋ, krɛaŋ, draŋ, ʔrəŋ, cəndruŋ, kəmrəŋ* 'horn'. Shorto (2006:692) reconstructs **tₐnra:ŋ* 'male' connecting *kədra:ŋ* 'male' in West Bahnar and *ʃənraŋ* 'male' in Khasi as cognates. *koraŋ* 'male' in Lyngngam is connected to *goraŋ* 'male, household ancestor' in Bonda (South Munda). I analyse *koraŋ, goraŋ* as a combination of two AA word formatives **kur* 'clan descent' and *rVŋ* as a word formative for horned beings in an abstract metaphorical sense. A corresponding grammaticalized word formative *rVŋ* is suffixed or prefixed in many AA cognates related to 'horned beings' in a concrete or in a metaphorical sense, including males and things or beings related to the Underworld (see Daladier 2007); *rVŋ* is found in **ʧənraŋ* 'male' in Pnaric and in *roŋbaʔ* 'representative village man, male adult' in War. Rather than relating *koraŋ, ʧənraŋ, kədra:ŋ* to **tₐnra:ŋ* 'male' and positing an independent cognate **dₐraŋ* 'horn' both with random initial forms and/or minor syllables, I consider *raŋ* 'horn' both as a cognate which may construct with various pre-syllables in a structure *[ ]raŋ* and as a grammaticalized word formative *rVŋ* which constructs in different word structures, where it combines with different relevant affixes and one or possibly two minor syllables. 'male' can be expressed by metaphor as a horned being like in Pnaric, Lyngngam, Munda and in some Bahnaric groups as in *Rengao* 'the men' as horned ones.

In PWL as in dfferent MK languages and in Munda, *san* 'great' is also used as a nominal word formative. As a frozen word formative, *san* would have various reduced forms used in many different cognates in PWL and in AA as a presyllable structure with variants: [s/sV/sn/n/ʃV/cV], see examples below in 'blood'. In PWL, *san* can still be used as a word formative, preposed or postposed, like Pnaric *basan* 'elder, important person' and *raŋbaʔ raŋsan* imitatives 'male elder, grown up male', War *roŋbaʔ* (imitative *roŋsan*) 'male elder, grown up'. *san* 'great' is also found in Korku. *san* is found both as a prefix and suffix in PWL and in Munda, in words which denote especially important beings or important eddibles, for example in War *roŋsan* 'male elder' and *sən.ʧar* name of a higher ancestor (*ʧar* is also the proper name of a river); Pnaric *raŋsan* 'male elder' and *kraj san* (lit. great millet) 'Paspalum sanguinale'; Sora has *sadai san* 'red gram', *ganga sa'an* ' a kind of millet' and also *sanna-kinadɛn* 'leopard' (*kinadɛn* 'tiger').

Analysing the presyllabes *s(ə), sn, n, ʃ(ən), ʧ(ən)* in PWL as frozen reductions of an AA word formative *san*, these presyllables are frozen remains of one word formative. They should not be confused with phonological reductions of AA roots for example, Khasi *ʃʔieŋ* < Pnar *ʧʔieŋ*. The same remark applies for *[tV]* and *[kV]* when they represent two different presyllables reduced from two different word formatives and a set of four word formatives with initial *k* for animate beings which incarnate on earth (including especially some animals, some body parts and some kin terms, see Daladier (2005)). These presyllables remains should not be confused with phonological reductions like /tr/>/kr/ in MK languages, analysed by Shorto (1971). Phonological derivations like *k < kʰ* (see Haudricourt 1965) with *k < kh < h* in onset position in some War words like *hi* 'fish'< **ka* 'fish' (with raising of /a/ into /i/ is frequent in War rhymes (Pnaric *kha* 'fish'); in onset

position *h* < *s* in Palaungic (see Diffloth 1980). Both kind of reduction: *k*< *kh*< *h* and *s* < *h* occur in presyllables of different AA languages.

To conclude this example, I do not consider Lyngngam *kuraŋ* 'male elder' and Pnaric *tʃənraŋ* 'male elder' as direct cognates, rather as two related elements containing the same core cognate **rVŋ* 'horn, horned' but constructing with two different word formatives in a structure *[ ]rVŋ*. In *[tʃən]raŋ*, I consider *tʃən-* as a frozen form related to an AA word formative **san* ' great'. In *[ku]raŋ*, I consider *ku-* as a frozen form related to an AA word formative **kur* 'clan descent'. *kuraŋ* and *tʃənraŋ* express a male elder with two attributes, as an important horned one in *tʃənraŋ* and as a clan descent horned one in *kuraŋ*. I do not consider that these two forms can be phonologically derived from the same proto-cognate but I consider that they are two different extensions of the same attribute name 'horn, horned one'.

'Male' can also be expressed as a being with a penis. *tərmaː* 'male (people and animals)' in War might be related to *krmeʔ* (man, husband) in Riang analysed with two different minor-syllables plus AA *-r-* 'people' infix (see §5) and a root: *[k]-r-meʔ, [tə]-r-rmaː*, as beings with a penis, following a suggestion of Shorto (2006). *meʔ* 'penis' occurs in Bahnaric: Jru and Nhaheun have *meʔ klɔː* 'male' and Shorto (2006) questions whether *meʔ klɔː* should be related to Khmer *mèː* 'penis'.

Ferlus (2011) analysis of *snaːm* 'blood' in Palaung and in Pnar-Khasi, proposes an infixation in AA **saːm* 'bleed, ooze' of a nominalizing infix *-n-*. The attempt at reconstructing partially an AA cognate of 'blood' by Ferlus is far from simple as he himself states: "These reflections on 'blood' lead us to reconsider nothing else than both the method and the process in the linguistic reconstruction". In conclusion he offers to reconstruct only two AA forms for blood from *saːm* bleed: *s-rn-am* for Palaungic and Khasi, *pN-saam* for many MK languages, perhaps *c/ɟ-saam* for Mon and Khmer, leaving unexplained or "random" structures in [ʔV/kV/V]- **saːm* 'blood' which occur in Vietic and in Katuic. In the analysis of Ferlus (2011) the root *saːm* 'to ooze, bleed' is the verb form from which the word 'blood' in the Austroasiatic languages is derived, except for Vietnamese and for some Vietic languages". Unfortunatly the proposed proto AA verb *saːm* 'to ooze, bleed', attested in Vietic, does not have any related element in S. Khasi and more generally in PWL and in Palaung, precisely the languages where the alleged *-n-* nominalizing infix of *s-n-aːm* blood from AA *saːm* 'bleed' is found. Ferlus does not mention any 'ooze, bleed' cognate in any Palaungic language either. More precisely, in Palaung, Milne (1932) analyses *hnaːm* 'blood' and *hnaːm hlaːi* 'bleed' literally 'blood flows' or *hnaːm huwaːr* 'blood flows' (*h* < *s* in onset position in Palaungic). Pnaric and War express the meaning of 'to bleed' as 'blood emerges, issues, raises or appears': in Pnar and in S. Khasi *snaːm miʔ* 'bleed' literally 'blood rises' (*miʔ* is found in *miʔsŋi* 'East' lit. (place where) 'rises sun'; in War: *ʃlɔʔ rnə* 'bleed' literally 'blood emerges'. Also contradicting with the hypothesis of Ferlus, there is no general *-n-* nominalizing infix in PWL but instead a conservative lack of noun/verb morphological distinction with a functional nominal and verbal use of most lexical roots. There are a few specialized prefixes for some nominal values like *noŋ-* producing agentive nouns and *dʒiŋ-* for abstract nouns and for transitive and causative verbal uses *pən-* and *təm-.* In addition, in different AA languages, 'blood' and 'bleed' are one single cognate, as in Santali *maːjaːm* 'blood, bleed' (Bodding 1932-7). Systematic morphological oppositions for nouns and verbs are most unlikely to be proto-AA.

The observation by Ferlus of a connection between 'ooze' as a process of producing a body liquid and blood production is interesting. Ferlus looking at all the AA available data extends the cognate of 'blood' to 'liquids which ooze' to reconstruct and relate two proto-forms and possibly a third one, concluding that there is no single proto-form for 'blood' in AA.

A different hypothesis would be to consider that here also minor syllables are not random and that they are frozen remains of reconstructible AA word formatives, some kind of complex underlying system of classifiers. These word formatives would also combine with affixes like causative prefixes. This hypothesis would not break the laws of phonological and historical reconstructions but enlarge them in places where apparently there is no logical explanation for phonological derivations. The names of blood would be analysed in variable sesquisyllabic structures which would specialize as 'blood' in structures with different attributes combining with a common root 'water, body liquids'.

Shorto (2006:1298) reconstructs *ʔ[o]m 'water' in Kammu-Yuan, Khasi and Palaungic. Adding new data, this cognate extends to: PWL, War *ʔam*, Pnar *ʔum*, Lyngngam *gum*; Khmuic, Khmu *ʔom*, most probably also Wa *rɔm* 'water', Aslian, Jah-Hut *tɔm* 'water' and Mangic, Mang *gom*[5] [1] 'water'; so this cognate of water extends to PWL, Palaungic, Khmuic, Mangic and Aslian.

Pinnow reconstructs (1959:2) *daʔ* 'water', Shorto (2006:274) reconstructs pre-Proto-Mon-Khmer *dàːk* 'water'. This AA cognate for 'water' might be a very widespread AA loan from Sanskrit *udaka* 'water, rituals performed with water', with different IA forms including *dak* 'water' (see Turner 1969). The widespread AA use of *daːk* 'water' might be due to the influence of Hinduism and Budhism in SE Asia and to the importance of water in *daksina* purification rituals. In PWL *dak* does not mean 'water' but 'sign, omen' in rituals and has extended its meaning to letter, alphabet. In PWL, *dak* has kept the notion of ritual of the Sanskrit *udaka* without the value of 'water, wet' of the Sanskrit root *ud*. Such loans in important AA cognates are not rare, for example, in many AA languages a word derived from IA *Brahman* like *brah* or *prah* (*praː* in War), is used as an adress term for great ancestors. In my opinion the original AA word for water is *ʔ[o]m and it has been replaced by *daːk for 'water' while *[ ]ʔ[o]m has extended into different related cognates like different body liquids inside different structures perhaps also 'to bath' as suggested by Shorto (2006:1426; 1417). *ʔ[o]m 'water' might be connected with *huum, *hum, *ʔum, *sum, p-hum 'to bath' with the vowel specializing in /u/ in Mon, North Bahnaric, Khmuic, South Aslian, Palaungic, Khasi, Vietic; The hypothesis of Shorto can be extended with data in North and South Munda: Santali, Mundari, Ho *um* 'bathe', Sora *uma* 'bathe').

In PWL many kinds of juices and body liquids in addition to 'water, river' are named with Pnaric *ʔum*, War *ʔam* like in War: *ʔam* 'river' *ʔam kiŋ* 'semen', *ʔam mat* 'tears', *ʔam meŋ* 'mucus', *ʔam pʔur* 'sweat', *ʔam ŋap* 'honey', *ʔam soʔ* 'fruit juice'.

Pnaric and Lyngngam have *snaːm* 'blood'. War has *rnə*, which I analyse as a reduced form of *rnaːm* from *snaːm* by spirantisation of initial *s* into *r* . Spirantisation of *s* in onset position is frequent in War, as in War *ran* 'five', Pnar *san* 'five'; War *rea* 'red', Pnar *sea* 'red'.

I analyse *snaːm 'blood' in PWL as a reduced form of an AA word formative *san* into *sn-* combined with *ʔam* with lenition of the glottal. More generally, the AA names of 'blood' may be analysed with different structures combining with *ʔ[o]m 'water' where the vowel specializes into /a/. To me these different structures are like *koraŋ* and *tʃənraŋ*

'male elder analysed before, not cognates in a strict sense but extensions from a common cognate, here of water, body liquids. I take up the analysis of *h < s* of Diffloth (1980) for Palaungic generalized to other groups by Ferlus (2011). I relate a set of pressyllables to the word formative *san* 'great' in (a). In (b), (c), (d), (e), (f), relating the presyllables to different combinations of word formatives. in (b) I consider [kV] as a reduced form of an indefinite element in set of four word formatives. The causative *pən-* prefix combines with [s/h] in (g). I have no solution for a word formative reducing to *mV*, it might be a variant of the causative prefix in (d), (e), (f):

Presyllables derived from *san*: [sn/s/h/hn/*n*V]

(a) [sn/s/h/hn/nV] **ʔaːm* 'blood'; [sn/hn]**ʔaːm > snaːm* 'blood' in PWL, Palaungic, U *sanàm*; Palaung *hnam*; [n] **ʔaːm* >Riang and Wa *naːm*, Plang *nam*[55],Rumai *n̥àːm*, Aslian; *[s]ʔaːm* >saːm, Mangic; Paliu saːm[5] [3]; Vietic Malieng asaːm[5] [1], Ruc açam[1]; *[h]ʔaːm* >haːm 'blood' in Katuic Nge' *haːm*; Mangic Mang *ham*[5] [3] Pearic, Kasong *hâːm*

(b) *[kV/ʔV] [s/h] **ʔaːm* 'blood' in Katuic Katu*ʔəhaːm*, Kui *ʔəhaːm*, Pacoh *ʔəhaːm*

(c) [(i)ɲ] [s/h] **ʔaːm* 'blood' in Katuic, Bru *ŋhaːm;* Munda, Juang *iɲam*

(d) [m/mV] [hV] **ʔaːm* 'blood' Aslian, Semelai *maham*; Bahnaric, Sedang *məheám*, Sre *mhaːm*, Stieng *mham*; Nicobaric, Car *ma-haːm*; Pearic, Chong-H *məhaːm*, Pear-B *môham*;

(e) [m/mV] **ʔaːm* 'blood' Khmuic, Khmu *maːm*, Ksinmul *miəm*, Mal *miam*, Mlabri *maːm*; Munda, Korku *majam*, Santali *maːjaːm* and Mundari *maĕom*, Hoffman and Van Emelen (1924-1930).

(f) *[m] [iɲ]*ʔam* 'blood' '*miɲam* 'to bleed' in Sora (Ramamurti)

(g) AA p(ən)- [s/h] **ʔaːm*   'blood' (with *p* causative 'make body liquid as blood') Bahnaric, Bahnar *phaːm*, Jeh *phaːm*, Jru *phəːm*, Nhaheun *phaːm*, Tampuan *phaːm*

Conclusion
Pnar, War and Lyngngam conservative varieties and S. Khasi are not mutually understandable; they still are four different languages, getting closer under the growing pressure of the "elite" status of S. Khasi. They have many composite varieties due to the successive influence of Pnar and Khasi as Lingua Franca, which will merge sooner or later into some Khasian. War, Pnar and Lyngngam conservative varieties still have very different morpho-syntactic systems, typologically conservative from an AA viewpoint, which could not be shown here in general, but which could be exemplified on their negation and pronominal systems. Pnar, War and Lyngnam have distinctive lexical and morphological similarities with Munda and MK groups. Specific isoglosses of War and Lyngngam with two sets of AA languages over their whole lexicons cannot be reconstructed because of Pnaric loans in different portions of their lexicons. Lyngngam has no specific similarities with Palaungic groups but it has more specific isoglosses with Juang and Kherwar (Munda) than War and Pnaric.

Pnaric, War and Lyngngam show intricate connections with different Northern and Southern AA groups and SE Asia areal contacts with TB and IA, Fu-nan Chinese in some

cardinal loans and perhaps loans from archaic Chinese in AA as indicated by Benedict (1972) quoted by Shorto (2006). Such kinds of intricate connections appear in some cognates like MK *bo[k]* 'belly' reconstructed by Shorto in Katuic and Nicobarese and extending to Pnaric and War *[k]po?* 'belly and principle of clan generation', further analysed by Shorto after Benedict (1972) as a borrowing from Archaic Chineese *piôk* 'belly', also found in TB as *pu'k* and *bu'k*. Lyngngam has *khlao* 'belly' reconstructed by Shorto for Bahnaric and Kherwar languages. This kind of split of cognates inside PWL, current between War and Pnaric and between Lyngngam and Pnaric suggests intricate moves, split and temporary settlements of small groups of Jhum cultivators and semi-settled agriculturists as opposed to permanent settlements in Hinduized kingdoms as for Pnaric, Monic and Khmeric.

When Hinduised Pnar kingdoms were already settled in fertile plains of Assam having trade routes and contacts with Tibet, China and the Gulf of Bengal, groups of War cultivators might have been shifting along small rivers in East Assam and Lyngngams might have been shifting in West and South West Assam perhaps in contact with Kherwar groups.

Lyngngam, Pnaric and then War groups have settled in that order in the refuge land of Meghalaya with Pnar and post-colonial S.Khasi successive lingua franca producing many recent mixed languages. A Pnaric influence on War and Lyngngam lexicons is an important feature of this group but each group still has many specific isoglosses in different northern and southern AA groups. They have probably been spoken in different places before Pnars and Wars met in the Assam Coridor, Pnars and Lyngngam in Meghalaya.

Pnaric is closer to Khmuic than to Palaungic. Lyngngam might have been in contact rather recently with Kherwar groups; there are still Kherwar groups settled between the South of Nepal and Bangladesh near the Brahmaputra. Wars may have been in contact with Palaung, Riang and Wa groups South-East of Assam between Tripura state and the Burma border before being pushed West by different conflicts induced by the settlement of the Ahom and the Moguls and by neighbouring promiscuities with smaller TB groups. There are still War groups near Jiribam South-East of Meghalaya and other War groups South East of Bangladesh speaking conservative War varieties.

Examination of conservative Pnar and relatively conservative War and Lyngngam varieties brings new light on the classification of the PWL group and perhaps also on sub-grouping inside AA. More data on Lyngngam conservative varieties are being collected and more comparative data have to be analysed to precise the origins of this group and the position of Lyngngam in the PWL tree. My guesses are that Pnaric, War and Lyngngam are sister groups, Khasi being a pnaric sub-group and that PWL might have originated before the Khmuic branching, not so far from Monic and Aslian branchings, being in contact with early conservative Munda groups before their South-Western migrations.

References:

Anderson, G. ed., 2008, *The Munda languages*, Routledge Language Family Series

Azemar, H. 1865, *Dictionnaire stieng*, Bro'lam: mission apostolique

Baruah, S.L. 1985. *A comprehensive history of Assam*, Delhi: Manoharlal pub.

Bhattacharya, S. 1968, *A Bonda dictionary*, Poona University Press

Benedict, P. 1942, « Thai-Kadai and Indonesian: a new alignment in south-eastern Asia", *American Anthropologist*,Vol. 44, 4

_____ 1972, *Sino-Tibetan, a Conspectus*, Cambridge: CUP

Brightbill J., Amy, K., Seung, K. 2007, The War-Jaintia in Bangladesh: a soscio-linguistic survey, SIL electronic report

Coedès, G. 1942, *Inscriptions du Cambodge*, II, Hanoi : EFEO

_____ 1989, Les états Hindouisés d'Indochine, Paris: De Boccard

Bodding, P.O. 1929, *Materials for a Santali Grammar*, II, Dumka

_____ 1932-7, *A Santal Dictionary*, (5 vol.), repub. Gyan

Bradley, D. 1997, "Tibeto-Burman Languages and Classification" in *Papers in SE Asian Linguistics* 14, Pacific Linguistics

Daladier, A. 2005, "Kinship and Spirit Terms Renewed as Classifiers of « Animate » Nouns and their reduced Combining Forms in Austroasiatic", BLS 28 S, Berkeley University Press.

_____ 2007, « Éléments cosmogoniques et vocabulaire austroasiatique du groupe môn-khmer du Meghalaya. Interactions avec le monde indien » Bulletin d'Études Indiennes 22, Paris : Collège de France, 337-389

_____ 2010, "Quand le non être n'est qu'un autre de l'être : négation-TAM en Kudeng War. " In: Floricic and Lambert Eds., *Les énoncés non susceptibles d'être niés*. Paris, CNRS Editions: 42-66.

_____ 2010-in prep, "Counting techniques with their "grouping" systems in Pnar, War, Khasi and Lyngngam and their relation to Austroasiatic decimal cardinals" presented at NEILS 5

_____ 2011 "Khasi" in Bovini, Busuttil and Peyraube eds., *Dictionnaire des langues*, Paris:PUF, 1121-1129

_____2012 a « Graded Active and Passive values » in Serial Constructions in Kudeng War" in: Gwendolyn Hyslop, Stephen Morey and Mark W. Post eds., North East Indian Linguistics 4, Cambridge University Press India, 373-403

_____ 2012 b "A multi-purpose project for the preservation of War oral literature" in: Gwendolyn Hyslop, Stephen Morey and Mark W. Post eds., North East Indian Linguistics 4, Cambridge University Press India, 167-194

_____ (in prep.) A Grammar of War

Diffloth, G., 1980, « The Wa languages », *Linguistics of the Tibeto-Burman Area* 5.2, Berkeley University

_____ 2005, "The contribution of linguistic paleontology to the homeland of Austroasiatic"
   in the peopling of East Asia, Sagart L., Blench R., Sanchez-Mazas A., Routlege

_____ 2011, "Les langues Austro-Asiatiques" in Bonvini, Busuttil and Peyraube eds., *Dictionnaire des langues*, Paris:PUF, 1099-1102

Engstrand, L., Widén, M., Widén, P., Ràw, K. and Svantesson, J.-O. 2008, "A checklist of Kammu plant names", *Mon-Khmer Studies Journal*, vol. 38, pp. 223-248.

Grierson, G.,A., 1927, Linguistic Survey of India, Vol. IV, Motilal, and vol.1 repub. Gyan

Haudricourt, G. 1965 "Les mutations consonantiques des occlusives initiales en Mon-Khmer", BSLP, 60 repris dans: *Problèmes de phonologie diachronique*, 1972, SELAF

Henderson E., 1976, "Vestiges of Morphology in Standard Modern Khasi" in *Austroasiatic Studies II*, P. Jenner, L. Thomson, S.Starosta, the University Press of Hawaï

Hoffmann, J., Van Emelin, A. 1924-1930 *Encyclopedia Mundarica*, (16 vol.), repub. Gyan

Jacq, P. and Sidwell, P. (2000). *A comparative West Bahnaric dictionary*, Languages of the world, 21. München: LINCOM Europa.

Jenner, P., (1976). « Les noms de nombre en Khmer » in Diffloth, G. and Zide, N. eds. *Austroasiatic Number Systems*, (special issue), Linguistics : 39-61

Kharakor, S. (1951). *Ki hun ki ksiew u Hynniew Trep*, Shillong, Ri Khasi Press.

Luce, G.H. 1985, *Phases of Pre-Pagan Burma*, Oxford University Press

Luce, G.H. 1965, "Danaw, a dying Austroasiatic language", Lingua 14

Matisoff, James. 1997. *Sino-Tibetan Numeral Systems: prefixes, proto-forms and problems.* Canberra: Pacific Linguistics

Menninger, K. (1969). [1934]. Number words and number symbols. A cultural history of numbers, Cambridge Mass.: M.I.T. Press

Milne, L. 1921, *An Elementary Palaung Grammar*, Oxford: Clarendon Press

Milne, L. 1931, Palaung-English and English Palaung Dictionary, Rangoon Spdt Gov. Printing house, Burma

Nagaraja, K.S. (1996). "The status of Lyngngam", *Mon-Khmer Studies* 26: 37-50.

Pinnow, H. J. 1959, *Einer Historischen Lautlehre der Kharia-Sprache*, Wiesbaden : Otto Harrassowitz

_____ 1965, "Personal Pronouns in Austroasiatic", Lingua 14

Ramamurti, G.V. 1931, *A Manual of the Soṛa (Savara) Language*, Madras Government Press

Ramamurti, G. V. 1933, *Sora English Dictionary*, repub. Mittal

Shadap-Sen N., C. (1981). The origin and early history of the Khasi-Synteng people, Calcutta: Firma KLM

Shorto, H.L. 1971, *A Dictionary of the Mon Inscriptions*, Oxford University Press

_____ 1963, "The Structural Patterns of Northern Mon-Khmer Languages", in H.L.Shorto (ed.) *Linguistics comparison in South-East Asia and the Pacific*, 45-61

_____ 2006, *A Mon-Khmer Comparative Dictionary*, Pacific Linguistics

Sidwel, P. 2009, Classifying the Austroasiatic languages: history and state of the art, München: LINCOM Studies in Asian Linguistics, 76

Singh, U. Nissor, 1906, *English Khasi Dictionary*, repub. Delhi: Mittal

Svantesson, J-O, 1983, *Kammu Phonology and Morphology*, Lund CWK cleerup

Thurgood, G. 1988, "K- Prefixes in Kam-Sui and Kadai: Some Notes*", Festschrift for Tatsuo Nishida*, Kyoto: Nishido

Turner, R. (1999).[1966]. *A comparative dictionary of the Indo-Aryan languages*, Delhi:Motilal Barnasidas

Weidert, A. 1975, I tkong Amwi. Deskriptive Analyse eines Wardialekts des Khasi, Wiesbaden: O. Harrassowitcz

Zide, N. 1978, *Studies in the Munda numerals*, Mysore: CIIL

Annex 1.
PWL cognates with the Mon and Palaungic comparative list (245 words) of Luce (1965).
P = Pnar, K= S. Khasi, *=Pnaric, L= Lyngngam, MM Modern Mon (O. Mon if unspecified)

Gloss	P or K or L	War	O.Mon or M.Mon	Danaw	Riang White/ Black	Palaung	Wa
one	*mi	mi	moj				
two	*ˀaːr	ˀə̃r	baːr		ʔar		
three	*laː	laː					luɛ²
five	*san	*san	msuːn				
six	P hndru	tʰroːu	turow				
I	P ˀɔ			oʔ	oʔ		
we	P ʔi	ʔi, iʔi			iʔ		
male		tərmaː			kərmeʔ		
child	*kɔn	hun	kon		kon		
gd. father (*ancestor*)	L. taolaŋ P. thaːo	wɔʔ	lwaˀ	ta	taʔ	ta	taʔ
gd. mother (ancestress)	P. (jaːo bej)	jaːo		jaʔ	jaʔ	jaʔ	jaʔ
gd. child	P. ksu	hənsow	cow				kɔnsaəʔ
hair	P. ṣɲiːuʔ	suʔ	sok	ɲok	huk	huɯʔ	haɯk
eye	*kʰmat	mat	mat				
nose	*kʰmut	muʔ	muʔ			muʔ	
breast	P. iɛmbu	buŋ		bu	buʔ		
thight	P. blu	ploːu		pluʔ	pluʔ	blau	
tail	*tdoŋ	tdoŋ		toŋta			
arm, hand	P. ti	taː	tej	ti	ti	ɗai²	taiʔ¹
nails	P. tərsim	snɛm	MM sanem		rahnim		
bone	P. tʃjeŋ	ʃʔjaŋ		kanaŋ⁴	tsənʔaŋ	kaʔaŋ	saʔaŋ²
skin	P. sneʔ	sniaʔ	MM snaːm				
blood	*snam	rnə	chim		nam	hnam²	nam²
dung		ksjaŋ		jaŋ⁴	jaŋ	iəŋ²	aiŋ²
dog	P ksaw	kseːa		tso¹	sʼoˀ		soʔ¹
tiger	kʰla	kʰli	klaʔ				
rat, mouse	P. kʰneː	kʰnaː	kni	kané			
crab	*tʰam	tʰəː	gatam	kətan²	kətam		tam
fish	*kʰaː	hiː	kaˀ		kaʔ	ka²	kaʔ
snake	*bsaŋ	psɛŋ		paθen⁴	həŋ		
bird	*sim	ksɛm	kincem		sʼim	sim²	ʃim²
fowl	*sʔiar	sʔi			jɛr	iər	
eagle	*kʰlijaŋ	kʰlijaŋ			klaŋ	klaŋ²	klaŋ²
louse	*ksiː	ksaː	MM caj	tsi¹	sʼiʔ	sai²	ʃiʔ
leaf	P. slaː	sliː	sla	la¹	laʔ	hla³	laʔ
mushroom	*tit	tɨt	ptis	tət³	tis	diːh	tiːh
thatch grass	*pʰlaŋ	pʰlaŋ		plaŋ⁴	plaŋ	plaŋ²	plɔŋ²
medicine	*dawaj	dawaj	ga uaj				

paddy	*kŭba			ba¹			
rice (non husked in PWL)	*kʰoː			ko¹	koʔ	rakau²	ŋ-gauʔ
taro	*tʃriːw	tʃroːu	krow	karo¹	s'əroʔ	s'əroʔ	krauʔ
sessanum		ləɲiː	MM laŋau	loŋ²ŋaʔ⁴	ləŋaʔ		ŋaʔ
ginger	*sʔiŋ	sʃiaŋ		katsaŋ⁴	kəs'iaŋ	ʃiaŋ²	ʃi⁴ kiːŋ²
turmeric	P. tʃɛrmɨt	tʃɛrmit	mit	k'amət²	rəmit		
earth, ground	*ktieʔ	kteʔ	ti	kate¹	kəteʔ	kaɗai²	deʔ³
sun	*sɲi	ɲuŋa	tŋej		s'əɲiʔ	saŋai	ʃi⁴ŋaiʔ³
year	*snɛm	snɨm	cnam			snam²	num²
water	*ʔum	ʔam			om	om², um²	rɔm²
bathe	*sum	sɛm	p-huːm		hum		həm
house post		kroŋ		k'araŋ	kənraŋ		
stone	*muː	ʃmea	tmo'	kamu¹	s'əmoʔ	mau²	ʃi⁴mauʔ³
house		sni	sɲiʔ				
boat	*lieŋ	liaŋ	dluŋ	gluŋ	tsən-luaŋ		
bitter	*kʰtaŋ	kʃaŋ	kataŋ	tsaŋ⁴	tsaŋ	saŋ²	soːŋ²
deaf	*kʰlɔt	kʰlɔt		klət³	lut	lut¹	lət³
new	*tʰmeː	tʰmaː	tumi	k'amɛʔ⁴	tən\meʔ	kamai²	
old (things)	*rim	sərɛm	tinrem		trim	aprim²	
ripe, cooked		ʃʔiɛn	cin		s'in	s'iːn²	ʃiːn
this (one fem)	*(ka.)ni	(kə.)nə	MM na'	ni²	ni		
bite	*dait	hit	MM kit	kiət³	kak		kiːt¹
burry	*tʰap	tʰɛp	tip				
carry on back	*baʔ	baʔ	baː	bɔʔ⁴	paʔ	bau	puʔ¹
fly	*hɛr	pʰear	par		pər		
mourn	*jam	ɲə	jam		jam	jaːm²	
open mouth	*ʔaŋ	ʔaŋ			ʔaŋ	ʔaŋ	ʔaŋ²
plait, weave	*thaŋ	thɛː	taːn	taːn²	taŋ		taŋ²
send, conduct	*pʰaʔ	pʰaʔ		p'uʔ	p'uʔ		
swell	*ʔat	ʔat		ut³	as	ah³	
tie, fasten	*kdɔʔ	kdɔʔ	dak	tôk¹	tuk		
want		kwa	gwa^?				

Annex 2.

Comparative word list of S. Khasi, West Pnar and conservative varieties of Pnar, War and Lyngngam (first hand data)

Gloss	S. Khasi	Ralliang Pnar	Nobosohpoh Pnar	Kudeng Nongtalang War	Amarsang Langkymma Lyngngam
head	kʰliʔ	kʰleː	kʰleːu	kʰleːa	kʰliʔ
ear	ʃkɔr	tʃkur	tʃkur	taraŋ	ləːkur
nose	kmutˀ	kmun/kmutˀ	kmutˀ	mərkoŋ	ləːmut
brains	ɟabieŋ	ɟabeŋ	ɟabeŋ	tʒi. kʰleːa	ɟaliʔ
stomach, belly	kpɔʔ	kpɔʔ	kpɔʔ	pɔʔ	kʰlaːo
knee	kʰɔʔsiːu	kʰasu	kʰɔʔseːu	kʰleːa hənsija	ʔɛŋ maːo khu
hair	ʃɲːuʔ	sɲeːuʔ	sɲɔːuʔ	suʔ	səɲɔk
leg, foot	kɟat	kɟat	kɟat	nea	kəɟat
hand, arm	kti	ti	kteːu	taː	ktɛj
tooth	bniat	ləmɛn	bniɛt	ləmɛn	əmɔɲ
eye	kʰmat	kʰman/kʰmat	kʰmat	mat	kʰmat
skin, bark	sniɛp	sneː	snɛp	sneːa/sɲiɛʔ	səniʔ
tongue	tʰəlliɛt	təllen	tlliɛt	kʰlɨtˀ	təllɔjt
word	ktiɛn	kteŋ	ktiɛn	tkoŋ	ktɔn
mouth	ʃəntur	kteŋ	ktiɛŋ	tkoŋ	əgap, ləmɔr
lips	ʃəntur	tʃəntur	tʃəntur	tʃəntur	ləmɔr
bone	ʃʔjeŋ	tʃʔeɲ/tʃʔeŋ	tʃʔiŋ	ʃʔjaŋ	tʃʔjaŋ
neck	rəŋdaŋ	radaŋ	raŋdaŋ	rdaŋ	kraŋ
blood	snam	snam	snam	rnə	snam
heart (blood channel)	kloŋ.snam	kloŋ.snam	kloŋ.snam	kloŋ.rnə	kloŋ.snam
liver	dɔʔ nutˀ	nɔtˀ	nɔtˀ	ktim	ənɔt
lungs	tɔr	tɔr	tɔr	tɔr	ətɔr
intestines	sniɛr	snɛr	snir	nɔr	snɔr
grease, fat, marrow	kʰlɛɲ	kʰlaɲ	kʰlaɲ	ləʔɔt	lɛɲ.taŋdoŋ
nerves, veins, roots of plants	tʰiɛtˀ	tɨtˀ	tɨtˀ	ʃɨtˀ	ətʰɔtˀ
tree	dieŋ	deŋ	deŋ	tvea	ədeaŋ
leaf	slaː	slaː	slaː	sliː	slaː
forest	kʰlaːo	kʰlɔː	kʰlaːo	kərmeːa	ləːu tʰap
sacred grove	kʰlaːo kəntəŋ	kʰlɔː kəntəŋ	kʰlaːo kəntəŋ	kərmeːa kəntəŋ	daːo tʃiga
playing ground (IA)	madan	madan	madan	madan	madan
water, river	ʔum	ʔum	ʔum	ʔam	gum
river, stream	waʔ	waʔ	waʔ	waʔ	əpɔr kmaːo (big river)
paddy field	ləŋkʰa	pəntʰɔr	ktiʔ (earth, land)	pəntʰɔr	laŋkʰa
paddy	kŭba	kŭba	kŭbja/kbje	hətʒi	ɟba
seed	səmbaj	səmbeː siaŋ	səmbje	tʒusbaj ʃʔjaŋ	kwit
rice (non husked)	kʰaw	khoː	kʰaw	rhija	əkʰaːo

cooked rice, food	ɟa	ɟa	ɟia	tʒi	əɟa
millet	kraɟ	kre:	kraɟ	tʒhan	ɟəraɟ
taro	ʃriːw	tʃriːw	tʃriːw	tʃroːw	tʃrəːw
ginger	sʔiŋ	sʔiŋ	sɔʔ ɔjn	sʔiaŋ	əɟin
turmeric	ʃənraɟ	tʃɛrmɨn, tʃɛrmit	tʃɛrmit	tʃɛrmit	tʃʔiaŋ tim
vegetable	ɟhur	ɟhur	ɟhur	tʒia	ɟhur/laʔur
banana	kait˺	ladaːo	lɔŋkait˺	ləde:a	əkait˺, lɔŋkait˺
betel nut	kwaɟ	kwəɟ	kwəɟ	kwɔɟ	kwi
paan leave	təmpʰeːu	tʰaŋpʰeːu	tʰaŋpʰeːu	pətʰa	thəmpu
medicinal herb	kənbat	kəmbat	kəmbat	kraʔ	kəmbatj
trad. doctor	paʔ han	kombiratɟ	kombiratɟ	kombiratɟ	kobira
salt	mluʔ	bloʔ	məlluʔ	pnuʔ	mllək
liquor	kiat	kiat	kiat	rə	kiat
flower	səntʰeːu	səntʰu:	səntʰeːu	kʰlə	səntʰəʔ
earth	ktieʔ	ktieʔ	ktieʔ	kteʔ	ktiʔ
earth (mud)	khəndeːu	khəndeːu	khɔndeːu	ktʒu	kmjaŋ
star	kʰlur	kʰlur	kʰlɔr	kʰlə ʃmɨn	kʰlɔr
sky	bneŋ	bneŋ	bnej	pʰlijaŋ	brɛj
sun	sɲiː	sɲiː	sɲeːu	ɲuŋa	sŋɛj
moon	bnaːj	bnaːj	bnɛj	pnu:	bni
cloud	lʔɔʔ	lʔɔʔ	lʔɔʔ	ləmpɛm	ləʔɔʔ
wind	kʰərwait	lʔɛr	lʔɛr	sərə:	ləʔir
rain	slap˺	slam/slap˺	slap˺	sla:	slap˺
stone	maːo	mu:	maːo	ʃme:a	əmaːo
rock	maːosjaŋ	musjaŋ	maːosjaŋ	musjaŋ	maːodar
hill	lɔːm	lɔːm	lɔːm	pdeŋ	ədɔm
fire	diŋ	dɨŋ	deŋ	ʃmɛn	ədɔŋ
ashes	dpej	tpaj	dpeŋ	puse:a	əpaːo
cave	krɛm	krɛm	krɛm/krɔʔ	krɛm/kraŋ	krɔŋ
village	ʃnoŋ	tʃnoŋ	tʃnoŋ	tʃnoŋ	ɟnoŋ
footpath	lənti	luti (kɟat)	luti (kɟat)	rhɛn	tuwar
bridge, ladder	ɟiŋkieŋ	jɛnkʰleiɲ	pərnɔn	ləʔu	pərnɔn
boat	liəŋ	lɛiŋ	lɛiŋ	liaŋ	liəŋ
egg	pəlləŋ	pəlleiŋ	pəlleiŋ	hun sʔi ('hen child'), tmat	pəlləŋ
egg divination	khan pəlləŋ	tmat		tmat	khan khaw de: (rice divination)
fish (alive)	(dɔʔ) kʰaː	(dɔʔ) kʰaː	kʰaː dɔʔ	hi:	əkʰaː
tiger	kʰlaː	kʰlaː	kʰlaː	kʰli:	kʰlaː
rat	kʰnaj	kne:	kʰnaj	kne:	kʰnaj
dog	kseːu	ksaːo	ksaːo	ksiːa	ksu:
crab	tʰam	tʰam	tʰam	tʰə	tʰam
spider	tʰapbawa	niaŋ.tʰawa	niampada	niaŋ.pɔʔ.nɨm	ənjam
mosquito	kaiɲ pərɟoŋ	maʔɲ	maʔtj	təŋkrɔŋ	ɟəkaŋ
insects, larva	kʰɲaŋ	kʰɲaŋ	kʰɲaŋ	kʰɲaŋ	kʰɲaŋ
bird	sim	sim	sim	ksɛm	əsim
fowl	ʔiar	sʔiar	sʔiɛr	sʔi	sʔɛr
duck	han	rappasa	hɛn	rəppəsə	daːogɛp (<Garo)
pig	sniaŋ	rniaŋ	rniaŋ	rniaŋ	sniaŋ

cow (Garo macu)	masi	masi	maseːu	məsaːo	məsə
turtle	dkar	lakaŋ	dkar	ləkaŋ	gəntat
frog, toad	ɟakoitˀ, hənrɔʔ	kʰrɔʔ	ɟakoitˀ, hənrɔʔ	kʰrɔʔ	haruʔ
dove	parɔ	rəpəti	rəpəti	rəpətə	tɔŋhur
horn	reŋ	reŋ	reŋ	reŋ	əreŋ
colour	roŋ	roŋ	roŋ	roŋ	əroŋ
ritual ceremony	roŋ	roŋ	roŋ	roŋ	əroŋ
white	ba.lieʔ	waˀ.leʔ	liʔ	səlaŋ	əlliʔ
red	saːo	soː	saːo	saːo	ənsaːo
black	ɟoŋ	ɟoŋ	tʃoŋ	priŋ	əɲoŋ
ancient, old (person)	thəmmɛn	thəmmɛn	thəmmɛn	tʃɛrkiaŋ	thəmmin
ancient, old (times, things)	rɨm	rɨm	rɨm	sərɛm	rim
new (thing)	thəmmaj	thmmɛ	thəmmi	thəmmaj	thəmmaj
bachelor fem.	samla, kʰənraːo	kəndrɔʔ	səmla	dərɔːu	kəllɔt
bachelor masc.	kʰənraːo	kəndrɔʔ	tʃəndraŋ	dərɔːu	kɔraːo
stop (transitive)	sŋeʔ	sŋeʔ		səŋit	səŋeʔ
hear, feel	sŋeːu	sŋjaːo	sŋja	sãʔ	səŋu
sad	sŋeːu siʔ	sŋjaːo diaːu	sŋja diaːu	sãʔdiaːu	səŋu siʔ
will, chosen	mɔn	mɔn	mɔn	mɔn	mɔn
beautiful, nice	ithənnat	miɛt	miɛt	miɛt	mərriaŋ
smell good	sma baŋ	sma baŋ	sma baŋ	rhiaŋ baŋ	aːu ŋur
smell bad	sma iːu	sma tuŋ	sma ejt/tuŋ	rhiang hɔʔ	aːu ktʃaʔ
respect	burɔm	burɔm	burɔm	burɔm	burɔm
to clean	pənkʰuitˀ	pənkɔjtˀ	phatˀ	tʃəlliŋ	phatˀ
cold	kʰriat	kɟam	kʰriɛt	ktʒam	təŋŋam
hot (pleasant)	ʃit	tʃhɛt	tʃhit	dɔt	əlup
prepare (food)	kʰreʔ	tʃɛt	tʃɛt	teŋ	kreaʔ
blow (fire)	pərsat	sluː	sleːu	phɛt	p(ən)nur
vital principle, breath	mənsiɛm	mənsɛm	mənsim	hənsua	ʔənsɔm
breath, inhalate	riŋ mənsiɛm	riŋ mənsim	than mənsim	rəŋ hənsua	riŋ hənsɔm
breath exh.	pənhiɛr mənsiɛm	pənhiɛr mənsɛm	pənhiɛr mənsim	pənnɔʔ hənsua	pənhiar hənsɔm
incarnated image, shadow	rəɲeːu	rəŋuː	rəŋuː	kɲaŋ	rɲəː
dark	(ba.)dum	(waˀ.)dum	dum	dum	ɲoŋna
dead, die	ɟap	ɟap	ɟap	jip	əɲap
alive, life	ʔim	ʔɛm	ʔim	pʔɛm	əɲnɨm
tasty	(ba.)baŋ	(waˀ.)baŋ	baŋ	baŋ	səŋəʔ
difficult	(ba.)ʔe	(waˀ.)ʔe	ʔe	saːr	ʔe
painful	(ba.)paŋ	(waˀ.)kɟut	kʔit	ram	əŋkʰɛt
enough	da biaŋ	da biaŋ	bjaŋ laʔ	də biaŋ	koitləʔ
all	ba.roʔ	waˀ.roʔ	bəroʔ	bəroʔ	prɔk
tired	(ba.)thait	(waˀ.)ŋaʔ	ɟraj	tʃriɛm	əmphi
little bit	kʰəndiat	kʰaɟiɛk	kʰəndiɛp	ʃi dit (time) ʃi tieŋ (things)	ədɛt
forget	klɛt	məlliŋ	rap	plijaŋ	bəlliŋ
remember	kənmaːo	kənmoː	tamɟa	kəmmoː	təmmaː

without manners	kʰlɛm akɔr	kʰlɛm akɔr	ɛn akɔr rej	kʰlɛm akɔr	kʰlim akɔr
slowly, peacefully (IAsuk)	suki	suki	suki	suki	əmɔn
sleep	tʰiaʔ	tʰiaʔ	tʰiaʔ	tʰiaʔ	eniŋ
walk	jaitˈ	kap	diʔ di kɟat (go by foot)	kap	d(ənn)iʔ bə kɟat
see, look at	ʔi	jo	ɟi	maʔ	ɟɔ/muɟɔ
ripe, cooked, mature	(ba.la)ʔi	(wa.)ˀi	ʔɔjtˈ	ʃhiɛn	ənnajt
cook	ʃɛt	tʃʰɛt	tʃʰɛt	teŋ	tʰ(ən)niŋ
love	jeit	maja	oŋʔit	məjə	pəndaj
heavy	(ba.)kʰia	(waˀ.)kʰia	kʰia	stuʔ	kniaʔ
sharp	(ba.)nɛp	(waˀ.)nip	njɛp	nɨp	əntaʔ/təllɔn
dry (intr.)	rkhiaŋ	rkhiaŋ	rkhiaŋ	rhiaŋ	rieŋkhɔŋ
small	(ba.)rit	kʰiɛn	rit	sbiɛt (tʃrit 'short')	duʔdɨt
big	(ba.)kʰraːo	(waˀ.)heʔ	heʔ	meːa	kəmbuʔ
long	ɟərɔŋ	ɟərɔŋ	ɟərɔŋ	kərɔŋ	ɟirɔŋ
right hand	kti mun	ti mun	kteːu ka mon	taː mun	ktəj təmmun
left hand	kti djaŋ	ti djaŋ	kteːu ka njaŋ	taː djaŋ	ktəj təmmɨŋ
North	ʃa.tej	bhoj	ha.siju	nu (upward)	tə.təj
South	ʃa.tʰi	dkar	ha.tʰeːu	ʃə (downward)	tə.tɔʔ
East, evening.sun	miʔ.ŋi	miʔ.sŋi	mɔjt.sŋeːu	miʔ.sŋi	mejt.ŋej
West, morning.sun	sɛp.ŋi	sɛp.sŋi	sɛp.sŋeːu	sɛp.sŋi	sip.ŋej
this one fem.	ka=ne	ka=ni	ka=neːu	kə=nə	g=ni
that one fem. ±remote	ka=tu/taj/ta	ka=tu/taj/te	ka=paːu /paːu ha ɟŋej	ka=tə/tun/tutun	g= tu/teːʔ
upward	ʃa.ɟərɔŋ	tʃa.ɟərɔŋ	ha.kənɟoŋ	nu	ətaʔ
downward	ha.pɔʔ	tʃa.pɔʔ	ha.tban	ʃə	həpoʔ
here	haŋ=ne	tʃei=ni	hən=neːu	ti=nə	hən.niʔ
there ±remote	haŋ=taj/tajtaj/ta	tʃei=taj/tu/ta	hən=taːu/taːu ɟŋej	tʃaː= tə/tun	ətu/ətəʔ/təɟiŋi
downward close	ha.pɔʔ haŋ=ne	tʃa.pɔʔ ɟaŋ	ha.tban haɟan	ʃə ʃon	həpoʔ ɟəŋan
evening	ɟan.miɛt	ɟaŋ.miɛt	mən.mɨt	tʒan.smɨt	ɟun.mɔt
dark /full night	miɛt/səŋaː	miɛt/səŋaː	mɨt/səŋaː	smɨt/ləmaʔ	mɔt/səŋeːu
morning	stɛp	stɛp	stɛp	rəti	ənsiʔ
today	kanə ka sŋi	məntu	nawein ka sŋjeːu	laʔ.hənlə	hənta
now	mənta	tʰadtʰeː kattə	nawei	daŋ.jə (as I am speaking) kat.nə, kat.lə	hndə
tomorrow	laʔ.ʃaj	mən.stɛp	laŋ.stɛp	laʔ.hənti	rə.tip
yesterday	mən.hənnɨn	ənnɛm	mən.mɨn	daŋ.hmmɨt	mɔʔ naːo
true	ʃiʃa	sə?kiɛn	məndɛj	tʃənnam	dəu/hnnan

false	lamlɛr	tɔʔ.re (not so)	tɔk.rɛj (not precise)	tɔʔ. tə (not so)	dəu.ɟi (not true)
one	wej	wiː	wiː	mi	əwə
two	ʔaːr	ʔaːr	ʔaːr	ʔə̃r	ʔair
I	ŋa	ˀɔ/ŋa	ˀɔ/ŋa	ŋə/njɛ	nə
we	ŋi	ʔi/ʔi	ʔi/ʔi	ʔi/iʔi	haj
they	ki	ki	ki	jə/ijə	kiju
know, be aware	tip	tip	tip	tɔʔ	wata
understand	sŋjeːo.thuʔ	sŋjaːo.thuʔ	sap.kʰan	sã̃ʔ.thuʔ	səŋu.thuʔ
think	prɔkʰanthe	pərkʰan	pərkʰat	pərkʰat	pərkʰat
take	ʃim	tʃim	tʃim	lum	tənnɔm
become, grow, be	man	man	man	man	ənsan
be, have	dɔn (exist)	ʔim (exist)	ʔɛm	ʔaˀ (have)	ʔim
give	ʔaj	ʔaː	ʔaː	ʔaj	(ənn)aj
go	leitˀ	laj	diʔ/laj	lea	d(ənn)iʔ
come	wan	wan	wan	wan/laʔ	l(ən)ar
enter	ruŋ	ŋaitˀ	ŋaitˀ	ŋit	s(ənn)aʔ
sit, stay, reside	ʃoŋ	tʃoŋ	tʃoŋ	ʃkeːa	matʃoŋ
fly	hɛr	hɛr	heːar	pheːar	kʰəndej
eat	bam	bam	bat/saː	bua	b(ənn)aŋ
speak	krɛn	klam	krɛn	ʃiɛr	kraʔ
cure, heal	pənkɔjtˀ	pənbhaʔ	pənkɔjtˀ	sumar/pənbhaʔ/ hjaʔ	pənkɔjtˀ
illness	(ɟiŋ-)paŋ	kdʒun/kdʒutˀ	kdʒutˀ	ktʒoːu	ɟoŋɟiŋhaj
drum (on the ground)	ksiŋ	kseɲ	kseɲ	nakraː	kseɲ
cloth	ɟain	that	ɟeɲ	deːa	əɟaŋ
tell, narrate	kʰana	parɔm, kʰana	(mən)jatʰu	perɔm, kʰənə	pərtɔk, khnaː
narrative/ritual narrative	kʰanataŋ kʰanataŋ	parɔm/kʰanataŋ	kʰana/kʰanataŋ	perɔm/kʰanataŋ	pərtɔk
dance	ʃat	tʃʰat	tʃʰat	karaj	tʃ(ənn)at
sing	ruaj	ruwaj	rivej	rwaj	rəŋwi
(traditional) music < IA	sur	sur	sur	sur	sur
burry, intern bones	thɛp	thap	thap	thɛp	thɨp
person	briːu	bruː	breːu	tʃəproːu	(b)rəː/brɑːu
mat. grandmother	kmjerad	bej.pun	mej.pun	jaːo	bej gjaːo
grandmother	kjaːo	kjaːo	kjaːo	jaːo	gjaːo
mat. grandfather	kparad	pa.pun	pa.pun	wɔʔ	thaːo/pathaːo
eldest mat. uncle	kɲi	maː	kɲi	mamaː, ɲeːu, ɲukoŋ	mamaː
elder brother	koŋ	san pun baʔ pun	baʔ deŋ	bai.pun	hmɨn kʰɔn kɔraŋ
youngest brother	duʔ	lɔj baʔ duʔ	baʔ duʔ	bai.duʔ	hbu kʰɔn kɔraŋ
elder sister	koŋ	san pun baʔ pun	koŋ heʔ	bai.lɔm	hmɨn raːo kmaːo
youngest sister	duʔ	lɔj	koŋ deŋ	bai.duʔ	hbu raːo kmaːo

		baʔ duʔ			
mother	kmjeː	bɛj	mɛj	maː	əbɛj/gmaːo
father	kpaː	paː	paː	paː	paː
friend	lɔk	paralɔk	paralɔk	p(ə)rəlɔk	marlɔk
wife	tŋa	kurim	lɔk	lɔk	kʰɔn.thaːo
husband	tŋa	kurim	lɔk	lɔk	kɔraŋ
child	kʰɔn	kʰɔn	kʰɔn	hun	kʰɔn
female	khəntʰaː	khəntʰaː	khəntʰaː	həntʰaː	raːo kmaːo
male	ʃənraŋ	tʃənraŋ	tʃənraŋ	tərmaː	kɔraŋ
I	ŋa	ʔɔ, ŋa	ʔɔ, ŋa	ŋə, ɲɛ	nə
house	iŋ	juŋ	tʃuŋ	(ʔu masc.) sni	iɲ/əjɛɲ
household	ka jiŋ ka sɛm	ka juŋ ka sɛm	ka tʃuŋ ka sim	(kə fem.) sni	(b)rə.jɛɲ loŋ jɛɲ loŋ sim
clan, fem.descent	kur	ɟat (specy), kur	ɟait (specy), kur	kur	əkur
rel. on the father's side	kʰa	kʰa	kʰa	kʰa	əkʰa
bite	dajt	dajt		hit	k(ənn)ap
breast, milk	buɲ	jɨmbu		buɲ	həmbəw
drink	diʔ	diʔ		deʔ	d(ənn)ajt˺
feather	snɛr	snɛr		suʔ ksɛm	snir
full	dap	dap		dap	əndap
good (to eat)	baŋ	baŋ		baŋ	səŋʔə
happy (IA suk)	suk, həmɛn	suk, kmɛn		suk, təmpaŋ, kmɨn	pəlləː
kill (cause to die)	pəniap	pəniap		pənjip	pənŋiap
lie down, sleep	thiaʔ	thiaʔ		thiaʔ	ənniŋ
louse	ksi	ksi		ksaj	ksəj
many	bun	bun		boːu	əbɔn
meat (class. for animals meat, for some tubers, and used as surname)	dɔʔ	dɔʔ	dɔʔ	dɔʔ	mim
dog	ksoːu	ksaːu		kseːa	ksuː
name	kərteŋ, jɛr	pərtuit		twijaŋ	kərteŋ
round	pəllun	pəllun		təpəllun	pəllun
sand	ʃʔjap˺	ʃʔjap˺		ʃʔjap˺	ɟʔɛp
smoke (in fire place)	tdɛm	tdɛm		tdɛm	ənthək
raise up, stand	jəŋ	jeːɲ		reŋ	ɲəŋ
swim	ɟɲi	jimpa	tʃimpu	rijaŋ	ŋ(ən)naj
cry, weep	jam	jam		ɲə	waːo
tail	tdoŋ	tdoŋ		tdoŋ	kdoŋ
want, desire	aŋud	aŋud		kwaʔ	kəndur
mustard	tərso (IA)	janɛm		nɨm	həriŋ

Annex 3
Cardinal numbers in Pnar, Khasi, War and Lyngngam: first hand
Jowai Pnar (JP), Ralliang Pnar (RP), Standard Khasi (SK), Langkymma Lyngngam (LL),
Kudeng Nongbareh-Nongtalang War (KW), Nongbareh village War (NW), and Thangbuli
Amwi War (TW).

	JP	RP	SK	LL	KW	NW	TW
1	wiː//t͡ʃi	wiː//t͡ʃi	wej //ʃi	əwə //t͡ʃə	miː//ʃi	miː//ʃi	miː //ʃi
2	ˀaːr	ˀaːr	ˀaːr	ˀair	ˀɔ̃/ˀɔ̃r	ˀũə	ˀũ/ˀũr
3	lɛː	lɛː	laj	laj-re	laː/laj	laː	laː/lɛː
4	sɔː	sɔː	saːo	saːo-re	reːa	riːa	siːa
5	san	san	san	san-də	ran	ran	san
6	hnru	hndru	hnriːu	hərəː	t͡ʰroːu	t͡ʰroːu	t͡ʰroːu
7	hnɲaːo	hnɲaːo	hnnjeu	hnju-re	hnt͡ʰlaː	hnt͡ʰlaː	hnt͡ʰlaː/hnt͡ʰlɛː
8	pʰraː	pʰraː	pʰraː	pʰraː-re	hmpʔɔ̃	hmpʔũə	hmpʔũ
9	kʰndeː	kʰndeː	kʰndaj	kʰndaj-re	hnʃʔaː	hnʃʔaː	hnʃʔeː
10	t͡ʃi phaːo	t͡ʃi phaːo	ʃi pheːu	t͡ʃə phuː	ʃi phuːa	ʃi phuːa	ʃi phuːa
11	khat wiː	khat wiː	khat wej	khat wə	ʃi phər mi	ʃi phər mi	ʃi phər mi
12	khat ʔaːr	khat ʔaːr	khat ʔaːr	khat ʔaːr	ʃi phər ʔɔ̃	ʃi phər ʔũə	ʃi phər ʔũ
15	khat san	khat san	khat san	khat san	ʃi phən ran	ʃi phən ran	ʃi phən san
20	ʔaːr phaːo	ʔaːr phaːo	ʔaːr pheːu	ʔaːr phu	ʔɔ̃r phuːa	ʔũər phuːa	ʔũr phuːa
31	lɛː phaːo wiː	lɛː phaːo wiː	laj pheːu wej	laj phu wə	laj phuːa miː	laː phuːa miː	laː phuːa mi
100	t͡ʃi sphaʔ	t͡ʃi sphaʔ	ʃi sphaʔ	t͡ʃə sphaʔ	ʃi swaʔ	ʃi swaʔ	ʃi swaʔ
1000	t͡ʃi had͡ʒar	t͡ʃi had͡ʒar	ʃi had͡ʒar	t͡ʃə had͡ʒar	ʃi had͡ʒar	ʃi had͡ʒar	ʃi had͡ʒar

JSEALS style guide for authors

Papers should be submitted electronically as .DOC or .RTF files. Initial drafts should be prepared with minimal formatting, and where possible with just one Unicode compliant font.

Ensure that your pages settings are as follows:

Margins on A4 paper:
Top   2.5 cm
Bottom   4.0 cm
Left   2.5 cm
Right   2.5 cm
Header   1.25cm
Footer   3.0 cm
Gutter margin: 0.5 cm
Font – 12 point Times New Roman. If your version of Times New Roman does not accommodate the characters you require, you may substitute with DoulosIPA (freely available from sil.org). For non-roman fonts please consult with the Managing Editor.

Line spacing: 14 pt exactly in normal text paragraphs

Mirror margins set ON

Image files for Tables or Figures should also be provided separately. Where possible, image files should be *greyscale* or *black & white*, .GIF or .TIF or .JPG, 300 DPI, 10cm wide.

Do not insert page numbers, or other header or footer text.

Do not link references, headings, tables or figure numbers in WORD.

Use footnotes, not endnotes.

All papers must have an abstract of approximately 100 words, plus 3 keywords (preferably consistent with the keywords used at http://sealang.net/sala/).

If you have any quiries, please consult the website at www.jseals.org or contact Paul Sidwell via <paulsidwell@yahoo.com> or <paul.sidwell@anu.edu.au>.

Journal of the Southeast Asian Linguistics Society
Volume 4-2

December 2011