



JOURNAL OF THE SOUTHEAST ASIAN LINGUISTICS SOCIETY

Volume 3.2
December 2010

JSEALS

Journal of the Southeast Asian Linguistics Society

Managing Editor:

Paul Sidwell (Pacific Linguistics, Canberra)

Editorial Advisory Board:

Mark Alves (USA)

George Bedell (Thailand)

Marc Brunelle (Canada)

Gerard Diffloth (Cambodia)

Marlys Macken (USA)

Brian Migliazza (USA)

Keralapura Nagaraja (India)

Peter Norquest (USA)

Amara Prasithrathsint (Thailand)

Martha Ratliff (USA)

Sophana Srichampa (Thailand)

Uri Tadmor (Indonesia)

Justin Watkins (UK)

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). Hard copies may also be ordered (availability subject to reasonable demand).

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 annually. Papers can be submitted to the Managing Editor, electronically (paul.sidwell@anu.edu.au or paulsidwell@yahoo.com) at any time.

Contents

Editorial	ii
Papers	
Postpositions and Relational Nouns in Lai <i>George Bedell</i>	1
Phonology of the Stieng Language: a Rime Study <i>Noëllie Bon</i>	22
The Differential Development of Proto-Southwestern Tai * <i>r</i> in Lao and Thai <i>Garry Davis</i>	49
The Austroasiatic Vocabulary for Rice: its Origin and Expansion <i>Michel Ferlus</i>	61
A Preliminary Study of Early Changes of Verbal Negators in Thai <i>Kiyoko Takahashi</i>	77
<i>klah</i> in Contemporary Khmer: Quantitative and Qualitative Plurality <i>Joseph D. Thach and Denis Paillard</i>	93
Child Acquisition of Vietnamese Classifier Phrases <i>Jennie Tran</i>	111
Pulling Out All the Stops in Vietnamese: A Delineation Between Native and Non-Native Vietnamese Speech for Voice Onset Time <i>Alina Twist, Jessica Shamoo Marx, Jessica Bauman, Allison Blodgett</i>	138
Book Notice	151

Editorial

Welcome to JSEALS Volume 3.2. Only a brief editorial this time around, as things seem to be bedding down fairly well with the running of JSEALS. Most importantly, I would like to express my deep appreciation for the editorial assistance given through 2010 by Marc Brunelle (University of Ottawa) - thank you Marc.

Readers will notice incremental improvements in this and coming issues. In this issue you will note that received and accepted dates are now being provided for each paper. Beginning in 2011 we will also be insisting on 100 word abstracts and three key words for each paper. And there are some revisions to the submission guidelines at the end of this issue - please check. Naturally we welcome any suggestions that will help us to continue improving JSEALS.

The 2011 SEALS meeting is shaping up to be a memorable get-together. The host institution - Kasetsart University in Bangkok - is offering beautiful modern meeting rooms in a great campus setting. Acharn Kitima Indrambarya, Chair of the Organizing Committee, is doing a superb job, and we look forward to seeing everybody there this coming May 11-13 (check jseals.org for updated information from time to time).

Paul Sidwell (Managing Editor)
December 2010

POSTPOSITIONS AND RELATIONAL NOUNS IN LAI

George Bedell
Payap University
<gdbedell@yahoo.com>

Introduction

This discussion¹ concerns the lexical category (word class) system in Lai, a Tibeto-Burman language belonging to the Kuki-Chin subgroup, spoken primarily in the townships of Hakha and Thantlang, Chin State, Myanmar. Lai is sometimes called (Hakha) Chin. At issue is a construction which consists of a postposition preceded by a word which we will argue belongs to a category ‘relational noun’, the combination corresponding to a single preposition in other languages, such as English. This makes it possible for Lai to get by with a very small number of monomorphemic postpositions. All the examples in this paper are taken from the Lai translation of the gospel according to Matthew, in *Lai Baibal Thiang* (1999). The parenthesized numbers accompanying each example are the standard chapter and verse reference. Examples are cited in the orthography used in the source, which is standard Lai orthography. Lai distinguishes short from long vowels and has a restricted tone system. Vowel length is not consistently represented in the orthography and tone is not represented at all. Otherwise, the orthography is relatively straightforward; syllable final h represents a glottal stop or glottalized sonorant. An outline of Lai phonology appears in Peterson (2003), and of Lai orthography in Bedell (1999a).

Compound postpositions

Consider the syntactic analysis of Lai words like *chungah* in sentence (1).

- (1) *Na-n pum cu muih-nak chung-ah a um ko lai* (6:23)
2pl body that dark-ness inside-p 3 be emph fut
‘your bodies will be in darkness’

At a superficial level, *chungah* appears to correspond to the English word ‘in’ in the gloss given. English ‘in’ is a preposition; therefore a first guess might be that *chungah* is a Lai preposition.² Or rather, since the term preposition incorporates the word order of languages like English, *chungah* ought to be a Lai postposition, since it follows its noun phrase complement: *muihnak chungah* ‘in darkness’. Unlike English ‘in’, *chungah* has two parts: *chung* and *ah*; in the analysis to be defended here, these are syntactically distinct, the former a noun and only the latter a postposition. Thus we take the relevant portion of the syntactic structure of (1) to be as in (i).

¹ An earlier version of this paper was presented to the 5th International Symposium on Languages and Linguistics ‘Pan-Asiatic Linguistics’ hosted by Vietnam National University of Social Sciences and Humanities in Ho Chi Minh City, Vietnam, November 16-17, 2000.

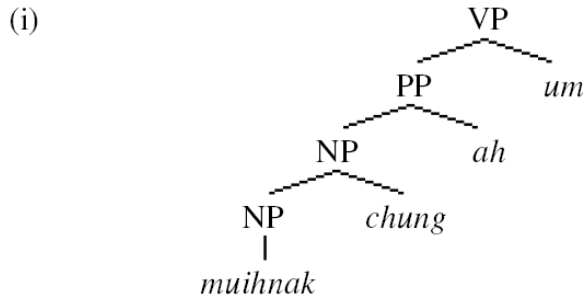
² The traditional view, as exemplified by Hay-Neave (1953), classifies such words as prepositions.

George Bedell. 2010. Postpositions and Relational Nouns in Lai.

Journal of the Southeast Asian Linguistics Society 3.2:1-21.

Copyright vested in the author.

Received 29/12/09, revised text accepted 1/12/10



In example (1), the PP *muihnak chungah* is the complement of the location verb *um* ‘be’, and specifies a place. Such PPs can also be used with motion verbs such as *paih* ‘throw’ in which case they refer to motion toward a place, as in (2). The PP complement in (2) is *mei chungah* ‘into the fire’. In such usage *chungah* corresponds to English ‘into’ rather than ‘in’.

- (2) *Mei chung-ah paih a-n si lai* (3:10)
 fire inside-p throw 3pl be fut
 ‘they will be thrown into the fire’

A parallel construction with *chungin* rather than *chungah* is illustrated in (3).

- (3) *kaa chung-in a chuak-mi thil cu* (15:18)
 mouth inside-p 3 come.out-rel thing that
 ‘things which come out of the mouth’

Here the PP *kaa chungin* ‘out of the mouth’ is complement to the motion verb *chuak* ‘go/come out’. The meaning of *chungin* in (3), which corresponds to the compound preposition ‘out of’ in English, differs from *chungah* in (2) in that the PP refers to motion away from, rather than toward, a place. Clearly the directionality of motion belongs to the meaning of *ah* versus *in* rather than to *chung*. *Chung* itself refers only to a place in abstraction from a location or direction.

The postposition *i*

Chung occurs without *ah* or *in* in examples like (4) or (5).

- (4) *muih-nak chung i a um-mi hna nih khan* (4:16)
 dark-ness inside p 3 be-rel pl by that
 ‘those who were in darkness’
- (5) *mei chung i paih awk-ah tom hna u* (13:30)
 fire inside p throw should-p gather pl pl
 ‘gather them to be thrown into the fire’

In (4) there is a locative meaning with the verb *um* and in (5) a directional meaning with the verb *paih*, just as in (1) and (2). The difference is that in (4) and (5), the verb has been

nominalized with the relative marker *mi* or the auxiliary noun *awk*.³ In this situation *ah* is replaced by *i*, which neutralizes not only the locative versus directional ambiguity of *ah*, but also the directional contrast between *ah* and *in*. Both directionality and its orientation must be inferred from the meaning of the main verb (*um* versus *paih*). The same effect is found when an overt head noun is present with a relative clause, as in (6).

- (6) *na mit chung i a um-mi thing-tan kha* (7: 3)
 2 eye inside p 3 be-rel log that
 ‘the log which is in your eye’

There are also examples like (7), in which a PP appears as a noun complement.

- (7) *na mit chung i thing-tan kha* (7: 5)
 2 eye inside p log that
 ‘the log in your eye’

Chung with its complement NP can modify a noun without *i*, as in examples like (8) or (9).

- (8) *lo chung belh bia-na-bia sullam kha* (13:36)
 field inside weed parable meaning that
 ‘the meaning of the parable of weeds of the field’

- (9) *tu-run chung tuu kha* (26:31)
 flock inside sheep that
 ‘sheep of the flock’

On occasion it can also be found modifying a relativized verb as in (10) or an infinitive verb as in (11).

- (10) *Ca Thiang chung trial-mi kha* (19: 4)
 writing holy inside write-rel that
 ‘what is written in the Holy Scripture’

- (11) *Van-cung Pen-nak chung luh hi* (19:23)
 heaven kingdom inside enter this
 ‘to enter into the Kingdom of Heaven’

Example (11) is to be compared with (12) containing the corresponding finite verb.

- (12) *Van-cung Pen-nak chung-ah a-n lut dih lai* (7:21)
 heaven kingdom inside-p 3pl enter all fut
 ‘they will all enter into the Kingdom of Heaven’

³ For further discussion of *awk*, see Bedell (1998).

That *luh* ‘enter’ in (11) is nominalized is shown in three distinct ways: (i) it does not agree with its subject; (ii) it is followed by the demonstrative *hi* ‘this’; and (iii) it appears in the ‘stem II’ form rather than the ‘stem I’ form *lut*, as in (12).⁴

In examples like (8) to (11), *chung* occurs without *ah*, *in* or *i*. The latter also occur alone as in examples like (13) to (16).

- (13) *Judea ram Bethlehem khua ah a si lai* (2: 5)
 Judea land Bethlehem town p 3 be fut
 ‘it will be in the city of Bethlehem of the land of Judea’

In (13) *ah* corresponds to English ‘in’ and takes a noun phrase referring to a place as its complement.

- (14) *arfi zoh-thiam-mi hna kha ni-chuah-lei in Jerusalem khua ah*
 star watch-expert-rel pl that sun-come.out-direction p Jerusalem town p
khan a-n ra (2: 1)
 that 3pl come
 ‘astrologers came from the east to the city of Jerusalem’

In (14) *in* and *ah* correspond to English ‘from’ and ‘to’, again taking noun phrases referring to places as their complements and, with the motion verb *ra* ‘come’, indicating motion away from or toward those places.

- (15) *Israel-mi cu Babilon ram ah sal ah a-n kal-pi hna* (1:11)
 Israel-person that Babylon land p slave p 3pl go-accom. pl
 ‘they took the Israelites to the land of Babylon as slaves’

- (16) *Babilon i sai i a-n kal-pi hna hnu-ah* (1:12)
 Babylon p slave p 3pl go-accompany pl after-p
 ‘after they took them to Babylon as slaves’

In (15) the first *ah* is the same as in (14), but the second illustrates one of the several additional uses (or meanings) of this postposition. As shown in (16), *ah* in either use can become *i* in a nominalized clause.

Deictics

Lai has a set of four deictic particles, used somewhat like English articles, which take a preceding noun phrase complement.⁵

- hi:* this, near me
kha: that, near you
khi: that, over there (visible)
cu: that, over there (not visible)

⁴ For analysis of Lai verb stem alternation, see Lehman (1996) and Kathol and Van Bik (2000).

⁵ For analysis of Lai deictic particles, see Bedell (1999b).

In examples like (17) and (18), phrases headed by *chung* but without *ah* or *in* may be followed by these deictics.

- (17) *ka lung chung hi ngaih-chiat-nak in a khat tuk* (26:38)
 1 heart inside this sad-ness p 3 fill much
 ‘inside my heart is filled with sorrow’

- (18) *na-n chung cu lih le sual-nak in a khat* (23:28)
 2pl inside that lie and bad-ness p 3 fill
 ‘your insides are filled with lies and evil’

These examples, more clearly than (8) to (11) above, show *chung* without any kind of pre- or postpositional meaning. The most natural English glosses might ignore it altogether: ‘my heart is filled up with sorrow’ or ‘you are filled with lies and evil’. The meaning of *chung* in examples like these is simply the interior of something. Since the phrases it heads are subjects, no postposition is needed, or allowed.

When the Lai deictic particles have a postpositional phrase as complement, they appear with a suffixed *n*: *hin*, *khan*, *khin* and *cun*. All four are illustrated in examples (19) through (26) with the postposition *ah*. These may be compared with (13) to (16) above; an additional relevant example is seen in (14). Note that *ah* often has a temporal sense.

- (19) *kan pawng-kam khua ah hin* (14:15)
 1pl near town p this
 ‘to the villages near us’

- (20) *zarh-khat chung ni hmasa-bik ni khuadei kate ah hin* (28: 1)
 week inside day before-most sun dawn time p this
 ‘toward dawn of the first day of the week’

- (21) *cu-ka in a-a hlat lem lo-mi hmun ah khin* (8:30)
 that-place p 3-rr far quite not-rel place p that
 ‘at a place not far from there’

- (22) *a-a thawk kate ah khin* (18:24)
 3-rr begin time p that
 ‘when he began’

- (23) *cu-ka hrawng i a um-mi khua ah khan* (11: 1)
 that-place vicinity p 3 be-rel town p that
 ‘to the villages around there’

- (24) *Judah ram a uk-tu hna mit ah khan* (2: 6)
 Judah land 3 rule-rel pl eye p that
 ‘in the eyes of the rulers of Judah’

- (25) *tlangbawi ngan Kaiafa innpi ah cun* (26: 3)
 priest big Caiaphas palace p that
 ‘in the palace of the high priest Caiaphas’

- (26) *Biaceih Ni ah cun* (11:22)
 judgment day p that
 ‘on the day of judgment’

This is true also with *in*, except that *in* itself does not appear, only the suffixed deictic.

- (27) *ka-n ram hin kir ko* (8:34)
 1pl land this go.away emph
 ‘go away from our land’

- (28) *belh cu a-n pumh hna i a-n khangh hna bantuk hin*
 weed that 3pl gather pl and 3pl burn pl just.as this (13:40)
 ‘just as weeds are gathered and burned’

- (29) *Biak-inn khan* (24: 1)
 worship-house that
 ‘from the temple’

- (30) *tu-khal nih tuu cu meheh sin-in a thren hna*
 shepherd by sheep that goat from-p 3 separate pl
bantuk khan (25:32)
 just.as that
 ‘just as a shepherd separates the sheep from the goats’

- (31) *a-n lung-thin tak-tak cun* (15: 8)
 3pl heart true that
 ‘from their true heart’

- (32) *a chimh hna bang cun* (26: 19)
 3 tell pl as that
 ‘as he had told them’

That examples like (27) to (32) contain *in* is clear from their meaning, and from the fact that if *hin*, *khan* or *cun* is removed, *in* will reappear. Also *bang* ‘as’ and *bantuk* ‘just as’ require *in* as adverbial clause conjunctions.

If we accept this condition (requiring the suffixed forms *hin*, *khan*, *khin* or *cun* of the deictic particles) as diagnostic for Lai postpositions, Lai has two other postpositions: *nih* ‘by’ and *he* ‘with’.

- (33) *a kaa chung i a lut-mi nih* (15:11)
 3 mouth inside p 3 go.in-rel by
 ‘what goes into one’s mouth’
- (34) *a kaa chung-in a chuak-mi tu nih hin* (15:11)
 3 mouth inside-p 3 go.out-rel emph by this
 ‘what comes out of one’s mouth’
- (35) *a thli-te in na-n tuah-mi a hmu-tu na-n Pa nih khan* (6: 4)
 3 secret p 2pl do-rel 3 see-rel 2pl father by that
 ‘your father who sees what you do in secret’
- (36) *lam pawng i a thru-mi mitcaw pa-hnih hna nih khan* (20:30)
 road side p 3 sit-rel blind.man two pl by that
 ‘two blind men sitting by the roadside’
- (37) *keimah ruang-ah a nunnak a sung-mi nih cun* (10:39)
 I because-p 3 alive-ness 3 lose-rel by that
 ‘anyone who loses his life because of me’
- The postposition *nih* is used with the subjects of transitive verbs, and is therefore often called the Lai ergative case marker.
- (38) *na unau he khan rem-nak va tuah* (5:24)
 2 brother with that peaceful-ness away do
 ‘go and settle with your brother’
- (39) *Zebedi fa-pa hna nu kha Jesuh sin-ah cun a fa-pa hna*
 Zebedee son pl mother that Jesus to that 3 son pl
he khan a-n ra (20:20)
 with that 3pl come
 ‘the mother of the sons of Zebedee came to Jesus with her sons’
- (40) *amah he cun rem-nak tuah i zuam* (5:25)
 he/she with that peaceful-ness do and make.haste
 ‘go quickly and settle with him’
- (41) *nanmah ca he kanmah ca he cun ka-n ngeih-mi hi a*
 youpl for with we for with that 1pl have-rel this 3
zaa lai lo (25: 9)
 suffice fut not
 ‘there will not be enough for both you and us’

The postposition *he* indicates accompaniment or, as in (41), conjunction. These four Lai words: *ah*, *in*, *nih* and *he*, together with *i* as a variant of *ah*, are the only postpositions in the language, by this criterion.⁶

In contrast to *chung* alone as illustrated in (17) and (18) above, *chungah* requires the suffixed deictics as in (42) and (43), while *chungin* requires them with the suppression of *in*, as in (44) to (46).

- (42) *phung-bia chung-ah hin zei nawl-bia dah a ngan*
 custom-word inside-p this which command-word q 3 big
bik-mi a si? (22:36)
 sup-rel 3 be
 ‘which command is the greatest in the law?’

- (43) *mitsur-dum chung-ah cun rian-truan awk-ah khan a thlah hna* (20:2)
 grape-garden inside-p this work should-p that 3 send pl
 ‘he sent them to work in the vineyard’

- (44) *nangmah chung hin hruai-tu pa-khat ... a chuak lai* (2: 6)
 youpl inside this lead-rel one 3 come.out fut
 ‘a leader will emerge from among you’

- (45) *Jesuh cu ti chung cun a chuak* (3:16)
 Jesus that water inside that 3 come.out
 ‘Jesus came out of the water’

- (46) *Jesuh cu lawng chung khan a hung chuak* (14:14)
 Jesus that boat inside that 3 up come.out
 ‘Jesus came up out of the boat’

The reasons for assuming that (44) to (46) involve *chungin* rather than just *chung* are the same as for examples like (27) to (32) above.

Genitives

The contrast between examples (17) and (18) on the one hand, and (19) to (26) on the other, shows that *chung* and *ah* do not behave syntactically in the same way, and therefore belong to different word classes. A second difference between them involves genitive constructions. Lai nouns can be accompanied by a possessor particle (or prefix), which agrees with a genitive noun phrase in person and number. These particles are homophonous with the subject agreement particles (or prefixes) which accompany intransitive verbs.

<i>ka</i>	‘my’	<i>kan</i>	‘our’
<i>na</i>	‘your’	<i>nan</i>	‘your’
<i>a</i>	‘his/her/its’	<i>an</i>	‘their’

⁶ Nahuatl represents a language more extreme in this respect than Lai. According to Bedell (1995) it has no prepositions at all.

They freely occur with *chung*, whether or not it is followed by *ah* or *i*, as illustrated in (47) to (53).

- (47) *a chung cu mi-thi ruh le a thu-mi pei a khat*
 3 inside that person-dead bone and 3 decay-rel emph 3 full
cu (23:27)
 that
 ‘its inside is filled with dead people’s bones and decay’

A parallel example with *nan* is (18) above.

- (48) *ram pa-khat cu a chung-ah anmah le anmah a-a*
 land one that 3 inside-p they and they 3-rr
do-mi bu a-n um ahcun (12:25)
 oppose-rel group 3pl be if
 ‘a land, if there are groups inside it who oppose each other’
- (49) *a-n chung-ah hram a thlak lo ca-ah* (13:21)
 3pl inside-p root 3 put.down not because
 ‘because they have no roots inside them’
- (50) *na chung-ah ka tuah-mi khua-ruah-har-nak vialtehi* (11:23)
 2 inside-p 1 do-rel miracle all this
 ‘all the miracles I performed in you’
- (51) *na-n chung i bia a chim-tu na-n Pa Thlarau* (10:20)
 2pl inside p word 3 speak-rel 2pl father spirit
 ‘the spirit of your father who speaks inside you’
- (52) *a-n chung i a um-mi hna nih a-n ziaza a-n thlen*
 3pl inside p 3 be-rel pl by 3pl ways 3pl change
hle i lo ca-ah (11:20)
 at.all not because
 ‘because the people in them had not changed their ways’
- (53) *hi hna hi a-n chung i khua-chia a ngei-mi a-n si* (8:28)
 this pl this 3pl inside p devil 3 have-rel 3pl be
 ‘they had demons inside them’

Pronouns

By contrast, a Lai postposition must have a full noun phrase complement and cannot be accompanied by a possessive agreement particle.

**ka ah*
kan in
na + i

nan *nih*
a *he*
an

This restriction does not apply to pronouns, which are not particles but full noun phrases. Lai has two sets of pronouns which reflect the same person and number features as the possessor agreement particles above.

<i>kei</i> 'I'	<i>kannih</i> 'we'
<i>nang</i> 'you'	<i>nannih</i> 'you'
<i>an</i> 'he/she/it'	<i>annih</i> 'they'
<i>keimah</i> 'I'	<i>kanmah</i> 'we'
<i>nangmah</i> 'you'	<i>nanmah</i> 'you'
<i>am</i> 'he/she/it'	<i>anmah</i> 'they'

The occurrence of pronouns with *nih* and *he* is illustrated in (54) to (59).

- (54) *nangmah nih na tuah-nak in va-a tuah ko* (27: 4)
 yousg by 2 deed p away-rr do emph
 'you go do your own deeds'
- (55) *nannih nih man pek lo-in na-n hmuhi, man pek*
 youpl by payment give without 2pl get and payment give
lo-in va pe ve hna u (10: 8)
 without away give emph pl pl
 'you, having received without pay, give without pay'
- (56) *an* *nih, 'ka duh lo,' tia* *a leh* (21:29)
 he/she by 1 want not quot 3 answer
 'he answered, "I won't"'
- (57) *am* *nih cun nannih chinchin cu thomh lo-in a-n*
 he/she by that youpl more that dress without 3-2
um hna hnga maw? (6:30)
 be pl would q
 'would he leave you without being clothed?'
- (58) *annihnih cun, 'mi nih ka-n thang-thrat hna seh,' ti in a-n*
 they by that person by 1pl praise pl imp quot p 3pl
tuah (6: 2)
 do
 'they do it saying, "let people praise us"'

- (59) *ka Pa pen-nak chung-ah nanmah he a thar in ka din*
 1 fatherkingdom inside-p 2pl with 3 new p 1 drink
thran hlan-tiang (26:29)
 again until
 ‘until I drink it new with you in my father’s kingdom’

A parallel example with *he* is (40) above.

Lai pronouns do not occur as complements to the postpositions *ah*, *in* or *i*. The reason for this is that pronouns refer to people and all human noun phrases are so restricted. Lai has a word *sin*, which serves to relate people to location and motion.

- (60) *Emmanuel ti cu ‘Pa-thian ka-n sin-ah a um,’ ti-nak a si*
 Emmanuel quot that God 1pl n-p 3 be quot 3 be (1:23)
 ‘Emmanuel means, “God is with us”’
- (61) *a mang ah Bawi-pa van-cung mi pa-khat a sin-ah*
 3 dream p lord heaven person one 3 n-p
a lang (1:20)
 3 come.down
 ‘in his dreams an angel of the Lord came down to him’
- (62) *a-n pa cu a fa-pa a dang pa-khat sin-ah khan a kal*
 3pl father that 3 son 3 other one n-p that 3 go (21:30)
 ‘their father went to his other son’
- (63) *Pa-thian sin-in maw a si, mi-nung sin-in dah?* (21:25)
 God n-p q 3 be human n-p q
 ‘is it from God, or from men?’
- (64) *nupi thri-tu a-n sin-in a-a thren can kha* (9:15)
 wife marry-rel 3pl n-p 3-rr separate time that
 ‘when the bridegroom is separated from them’
- (65) *thetse ram i Johan sin i na-n va kal ah khan* (11: 7)
 desert land p John n p 2pl away go p that
 ‘your going out to John in the desert’
- (66) *na sin i ka thih ve ding a si hmanh-ah* (26:35)
 2 n p 1 die also must 3 be even-p
 ‘even if I must die with you’
- (67) *na-n sin a-n rat ahcun* (7:15)
 2pl n 3pl come if
 ‘if they come to you’

- (68) *mi zapi sin kha a-n phak thran tik-ah* (17:14)
 person many n that 3pl arrive again time-p
 ‘when they again arrived at the crowd’

Although *sin* corresponds directly to no English word, preposition or otherwise, its syntactic properties are the same as those of *chung*. It is usually followed by *ah* or *in* in the sense either of location or of motion toward or away from. In nominalized clauses it may be followed by *i* as in (65) and (66), or occur alone as in (67) and (68). It may be followed by deictic particles as in (62) or (68), or preceded by possessor agreement particles as in (60), (61), (64), (66) or (67).

Relational nouns

The class of Lai postpositions, as we have seen, is quite small. The word class to which *chung* and *sin* belong is rather larger. As suggested at the outset, we take them to be nouns, which is consistent with their syntactic properties: heading complements to postpositions, heading modifiers of other nouns, being modified by noun phrases, and being able to be accompanied by possessor agreement particles. That being said, it is clear that *chung* and *sin* represent a significant subclass, which have some properties not shared by all nouns. We take this to a result of the kind of meaning that they have. They do not refer to objects of a simple sort, but rather to abstractions defined in relation to objects. Thus, as we saw, *chung* means the interior of something, and *sin* means the place of a person. Such nouns do not normally occur in subject or object positions in a sentence, but rather in adverbial function, so that they most often are accompanied by a postposition. This is not a syntactic restriction, as is clear from examples like (17) and (18). The subclass may be called ‘relational nouns’.

Relational nouns in Lai include words for spatial relations like *chung* and *sin*. *Cung* in examples (69) to (71) refers to the relation of being above something, and corresponds to English ‘on’ or ‘off’.

- (69) *mi dang cung i na hman-mi tah-nak-te kha Pathian nih na*
 person other on p 2 use-rel measure that God by 2
cung-ah a hman ve lai
 on-p 3 use also fut
 ‘the measure that you use on others God will use on you’ (7: 2)

- (70) *Jesuh cu tlang cung khan a rung trum* (8: 1)
 Jesus that mountain on that 3 down come
 ‘Jesus came down off the mountain’

- (71) *vawlei cung mi-phun vialte kha* (24:30)
 earth on people all that
 ‘all peoples on earth’

Similarly, *tang* in (72) and (73) refers to the relation of being below something, and corresponds to English ‘under’.

- (72) *arpi nih a fa-le cu a thla tang-ah a huh hna*
 hen by 3 children that 3 wing under-p 3 protect pl
bantuk in (23:37)
 just.as p
 ‘just as a hen protects her children under her wing’

- (73) *kei zong ka cung-lei bawi nawl tang i a um-mi ka*
 I too 1 above lord command under p 3 be-rel 1
si (8: 9)
 be
 ‘I too am under the command of my superiors’

Hmai and *hnu* in (74) to (76) refer to the relations of being in front of or in back of something, and correspond to English ‘before’ or ‘behind’.

- (74) *Mari cu zapi hmai-ah ning-zah-ter a duh lo ca-ah (1:19)*
 Mary that many front-p shame 3 want not for- p
 ‘because he did not want to shame Mary before people’

- (75) *a hmai i a kal-mi hna le a hnu in a zul-mi*
 3 front p 3 go-rel pl and 3 behind p 3 follow-rel
hna cu (21: 9)
 pl that
 ‘those who went in front of him and those who followed behind him’

- (76) *na-n hmai khua ah khan rak kal u law (21: 2)*
 2pl front town p that prf go pl and
 ‘go to the village in front of you’

Cung, *tang*, *hmai* and *hnu* in (69) through (76) show the same syntactic properties as *chung* and *sin* in earlier examples. Other similar words include *leng* ‘outside’, *lak* ‘among’, *kam* ‘near’, *kiang* ‘near’ *lei* ‘toward’, *pawng* ‘near’, *pin* ‘beyond’ and *velchum* ‘around’.

Lai relational nouns are not limited to spatial relations, but are also used for temporal relations. *Hlan* in (77) to (81) refers to the relation of temporally preceding something, and corresponds to English ‘before’.

- (77) *atu i a nung lio-mi a-n thih dih hlan-ah hin (24:34)*
 now p 3 alive prog-rel 3pl die all before-p this
 ‘before all those who now are living have died’

- (78) *zei-zong vialte hi a dongh-nak a-n phak hlan-ah cun*
 what-too all this 3 end 3pl arrive before-p that (5:18)
 ‘before the end of all these things arrives’

- (79) *zei-cah-tiah na-n Pa nih cun na-n hal hlan-in na-n herh-mi*
 because 2pl father by that 2pl ask before-p 2pl need-rel
cu a hngalh cia ko (6: 8)
 that 3 know already emph
 ‘because your father already knows what you need before you ask’

- (80) *na-n hlan i a rak um-mi profet hna kha (5:12)*
 2pl before p 3 prf be- rel prophet pl that
 ‘the prophets who were before you’

- (81) *siangpahrang bantuk in ka rat na-n ka hmuhhlan cu*
 king as p 1 come 2pl 1 see before that (16:28)
 ‘before you see me coming like a king’

Hnu in (82) to (86) refers to the relation of temporally following something. Like the corresponding English ‘after’, it is used also for a spatial relation as in (75) above.

- (82) *ni-thum hnu-ah (26:61; 27:63)*
 day-three after-p
 ‘after three days’

- (83) *Jesuh cu thih-nak in a thawh thran hnu-ah (27:53)*
 Jesus that death p 3 rise again after-p
 ‘after Jesus rose again from the dead’

- (84) *ka hnu-in a ra lai ding pa (3:11)*
 1 after-p 3 come fut must man
 ‘the man who will come after me’

- (85) *cu hnu ni-ruk ah (17: 1)*
 that after day-six p
 ‘six days later’

- (86) *khua-khat hnu khua-khat na-n dawi hna lai (23:34)*
 town-one after town-one 2pl chase pl fut
 ‘you will chase them from town to town’

Tiang in (87) to (90) refers to the relation of temporal extension up to something, and corresponds to English ‘until’.

- (87) *Abraham in David tiang-ah khan (1:17)*
 Abraham p David until-p that
 ‘from Abraham until David’

- (88) *Noah lawng chung i a luh ni tiang khan* (24:38)
 Noah boat inside p 3 enter day until that
 ‘until the day Noah entered the ark’
- (89) *Herod a thih tiang kha* (2:15)
 Herod 3 die until that
 ‘until Herod is dead’
- (90) *ni-hin ni tiang hin* (11:23)
 day-this day until this
 ‘until today’

Hlan, *hnu* and *tiang* behave like other relational nouns, and often have noun phrase complements. But they also typically may have clause complements as in (77), (78), (79), (81), (83) and (89). When they do, an intransitive verb in the complement clause appears in an infinitive form: *thih* in (77) or (89), and *thawh* in (83), in contrast to *thi* and *tho*. In such uses, they correspond to English conjunctions rather than prepositions, and resemble English ‘before’, ‘after’ and ‘until’ in this respect also.

The Lai word *tik* may be analyzed as a relational noun, though it almost invariably has a clause complement. *Tik* in (91) to (98) refers to the relation of something happening (or being so) at a moment in time, and corresponds to the English conjunction ‘when’.

- (91) *Josef cu a hun i hlauh tik-ah* (1:24)
 Joseph that 3 up rr wake when-p
 ‘when Joseph woke up’
- (92) *na-n hmuhtik-ah ra ka chim thran u* (2: 8)
 2pl find when-p come 1 tell again pl
 ‘when you find him come again and tell me’
- (93) *a-n daw-tu hna lawng na-n dawt hna tik-ah hin* (5:46)
 3-2 love-rel pl only 2pl love pl when-p this
 ‘when you love only those who love you’
- (94) *Jesuh cu inn chung i a luh tik-ah khan* (9:28)
 Jesus that house inside p 3 enter when-p that
 ‘when Jesus went into the house’

(95) and (96) illustrate *tik* preceded only by a deictic, and (97) and (98) illustrate it within an interrogative phrase.

- (95) *khi tik-ah* (7:22)
 that time-p
 ‘at that time’

- (96) *cu tik-ah* (9: 3)
that time-p
'at that time'
- (97) *cu ni le cu can cu zei-tik ah dah a si lai, ti kha*
that day and that time that what-time p q 3 be fut quot that
aho-hmanh nih a-n hngal lo (24:36)
no one by 3pl know not
'no one knows when the day or the hour will be'
- (98) *zei-tik set ah dah arfi kha a chuah ti kha fiangte in*
what-time just p q star that 3 come.out quot that clearly p
a hal hna (2: 7)
3 ask pl
'he asked them exactly when the star appeared'
- The relational noun *tik* contrasts with the ordinary noun *can*; in (99) to (106); *can* refers to time as an entity.
- (99) *cu can-ah cun* (24:10)
that time-p that
'at that time'
- (100) *cu can cu* (24:22)
that time that
'that time'
- (101) *cu can har-nak cu* (24:29)
that time difficult-ness that
'that time of troubles'
- (102) *cu can i a um-mi har-nak cu* (24:21)
that time p 3 be-rel difficult-ness that
'the troubles at that time'
- (103) *inn ngei-tu nih mi-fir a rat lai can kha hngal*
house own-er by person-steal 3 come fut time that know
seh-law (24:43)
imp
'if the householder knew when the thief would come'
- (104) *a ni si-seh, a can si-seh, na-n hngalh trung lo ca-ah*
3 day be-imp 3 time be-imp 2pl know emph not for-p
ralring-te in um ko u (25:13)
careful p be emph pl
'be careful, because you do not know the day or the time'

- (105) *mitsur thei lawh can kha a phak tik-ah* (21:34)
 grape fruit pick time that 3 arrive when-p
 ‘when the time to pick the grapes arrived’

- (106) *can dongh tiang-in ka um lai* (28:20)
 time end until-p 1 be fut
 ‘I will be here until the end of time’

Though in examples like (99) (or (64) above), *can* may on occasion be interchangeable with *tik*, this is not the case in (100) to (106).

The Lai words *ca*, *ruang*, *awk* and *bang* may also be analyzed as relational nouns. *Ca* in (107) to (110) refers to the relation of being for some purpose, and corresponds to the English preposition ‘for’ or the conjunctions ‘in order to’ or ‘because’.

- (107) *Pathian nih a thim-mi hna ca lawng ah a si* (19:11)
 God by 3 choose-rel pl for only p 3 be
 ‘it is only for those God has chosen’

- (108) *a ka chung-ah cun kan pa-hnih ca i a za ding*
 3 mouth inside-p that 1pl two for p 3 suffice must
tangka na hmuh lai (17:27)
 coin 2 find fut
 ‘in its mouth you will find a coin which is enough for both of us’

- (109) *na-n nun-nak ding ca i na-n herh-mi rawl le ti* (6:25)
 2pl live must for p 2pl need-rel food and water
 ‘the food and water you need in order to live’

- (110) *khuachia nih a tuah ca i a holh kho lo-mi pa pa-khat*
 demon by 3 do for p 3 speak can not-rel man one
kha (9:32)
 that
 ‘a man who could not speak because a demon possessed him’

Additional examples of *ca* with *ah* are (41), (49), (52) and (74) above. *Ruang* in (111) to (114) refers to the relation of being caused by something, and corresponds to English ‘for’ or ‘because’.

- (111) *ka-n thlacam a sau ruang-ah hin Pathian nih a kan theih-piak*
 1pl prayer 3 long because-p this God by 3 1pl hear-ben
deuh lai (6: 7)
 comp fut
 ‘because our prayer is long God will hear us better’

- (112) *mi-thra* *cu* *mi-thra* *a-n* *si* *ruang-ah* *a* *rak*
 person-good that person-good 3pl be because-p 3 prf
cong-lawmh-tu *paoh* *cu* (10:41)
 receive-er all that
 ‘all who receive a good man because they are good’
- (113) *mi-pa* *pa-khat*, *khuachia* *nih* *a* *tuah* *ruang-ah* *a* *mit* *a caw*
 man one demon by 3 do because-p 3 eye 3 blind
i *a* *holh* *kho* *lo-mi* *kha* (12:22)
 p 3 speak can not-rel that
 ‘a man who was blind and could not speak because a demon possessed him’
- (114) *hi* *pa* *thah-nak* *ruang* *i* *dan-tat-nak* *cu* *kanmah* *le* *ka-n*
 this man killing because p punishment that we and 1pl
te-fa *hna* *cung-ah* *tlung* *ko* *seh* (27:25)
 child pl on-p fall emph imp
 ‘may the punishment for this man’s death be on us and our children’

Examples like (110) and (113) show that the relations of purpose and cause sometimes overlap. An additional example of *ruang* is (37) above. *Bang* and its variant *bantuk* refer to the relation of similarity; see examples (28), (29), (32), (72) and (81) above. *Awk* refers to the relation of purpose; see example (43) above as well as Bedell (1998).

Finally it may be of interest that the Lai word *nak* which serves to indicate the standard of comparison corresponding to English ‘than’ is also a relational noun, always accompanied by the postposition *in*, as in (115) to (120).

- (115) *amah* *cu* *keimah* *nak-in* *tam-pi* *in* *a* *ngan* *deuh* (3:11)
 he/she that I than-p much p 3 big comp
 ‘he is much greater than I’
- (116) *mi* *cu* *tuu* *nak-in* *a* *let* *tam-pi* *in* *a* *sung*
 person that sheep than-p 3 time many p 3 valuable
deuh! (12:12)
 comp
 ‘a man is many times more valuable than a sheep’
- (117) *profet* *nak-in* *bia-pi* *deuh* *a* *si-mi* *pei* *n-an* *hmuh* *can*
 prophet than-p important comp 3 be-rel emph 2pl see prf
hi! (11: 9)
 this
 ‘you have seen something more important than a prophet’

- (118) *na pum a ning-pi in hel chung i tlak nak cun na*
 2 body 3 whole p hell inside p fall than that 2
kut-ke pa-khat khat sungh cu a thra deuh vingvan (5:30)
 hand-foot one one lose that 3 good compints
 ‘it is much better to lose a hand or foot than for your whole body to fall into hell’

- (119) *a hmuhtik-ah sawm-kua le pa-kua a tlau lo-mi cung nak*
 3 find when-p ninety and nine 3 lose not-rel over than
khan hi tuu pa-khat cung-ah hin aa lawm deuh (18:13)
 that this sheep one over-p this 3-rr happy comp
 ‘when he finds it, he is happier over this one sheep than over the ninety-nine
 which he did not lose’

- (120) *dum ngei-tu nih cun hmasa nak tam deuh in sal*
 gardner own-er by that before than many comp p servant
dang a thlah thran hna (21:36)
 other 3 send again pl
 ‘the owner of the vineyard sent again more servants than before’

Conclusion

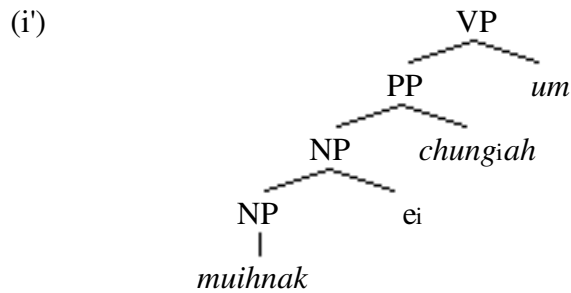
Our analysis with relational nouns applies not just to the variety of Lai constructions listed here, but is in fact a common phenomenon in other languages as well.⁷ English too has compound prepositions and conjunctions which seem to contain a nominal element:

‘on top of’
 ‘in back of’
 ‘inside of’
 ‘on account of’
 ‘in order to’

In these, the initial preposition serves to case mark the relational noun which follows it. In English the following ‘of’ serves to case mark the genitive noun phrase which follows it; in Lai, there is no genitive case marker. In Lai orthography, a relational noun is written as one word with a following *ah* or *in* (but not with *i*).⁸ It would be possible to take this as reflecting a morphological structure as in (i’).

⁷ The term ‘relational noun’, in the sense used in this discussion, appears often in the literature of Mesoamerican languages. See for example Dayley (1985, pp. 152-59), Bedell (1995) and Lillehaugen and Munro (2006).

⁸ For discussion of Lai orthographic conventions, and in particular word division, see Bedell (1999a).



But there is no evidence other than orthographic practice for such a structure.

Abbreviations

1	first person
2	second person
3	third person
ben	benefactive suffix
by	agentive postposition (or ergative case marker)
comp	comparative particle
emph	emphatic particle
fut	future particle
imp	imperative particle
n	noun
p	postposition
pl	plural suffix or particle
prog	progressive particle
q	interrogative particle
quot	quotative particle
rel	relative suffix
rr	reflexive or reciprocal particle
sg	singular

References

- Bedell, G. 1995. 'The Category P in Nahuatl', *ICU Language Research Bulletin* 10:79-92.
- Bedell, G. 1998. 'Nominal Auxiliaries in Lai', *Papers from the 8th annual meeting of the Southeast Asian Linguistics Society (SEALS)*, Kuala Lumpur, Malaysia, 20-22 July 1998, 11-24. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Bedell, G. 1999a. "'Word Combination" in Lai', *Papers from the 9th annual meeting of the Southeast Asian Linguistics Society (SEALS)*, University of California, Berkeley, 21-23 May 1999, 201-219. Tempe, Arizona: Program for Southeast Asian Studies, Arizona State University.
- Bedell, G. 001. 'The Syntax of Deixis in Lai', *Linguistics of the Tibeto-Burman Area (LTBA)* 24:2. 157-171.
- Dayley, J. P. 1985. *Tzutujil Grammar*, Berkeley and Los Angeles: University of California Press.

- Hay-Neave, D. R. 1953. *Lai Grammar and Exercises*, Rangoon: Ministry for Chin Affairs.
- Kathol A. and K. Van Bik. 2001. 'Lexical Constraints and Constructional Overrides: on the Syntax of Verbal Stem Alternations in Lai', presented at Linguistic Society of America (LSA), Washington D. C.
- Lehman, F. K. 1996. 'Relative Clauses in Lai Chin, with Special Reference to Verb Stem Alternation and the Extension of Control Theory', *Linguistics of the Tibeto-Burman Area* (LTBA) 19:1. 43-58.
- Lillehaugen B. D. and P. Munro. 2006. 'Prepositions and Relational Nouns' in a *Typology of Component Part Locatives*. Downloaded from http://en.wikipedia.org/wiki/Relational_noun, November 2010.
- Peterson, D. 2003. 'Hakha Lai'. *The Sino-Tibetan Languages*, ed. Thurgood and LaPolla, 409-426. London and New York: Routledge.
- (D. Van Bik) 1999. '*Matthai sinin Thawngthra*' (*the Gospel according to Matthew*), *Lai Baibal Thiang* (*The Holy Bible in Lai*), Bible Society of Myanmar, centennial edition.

PHONOLOGY OF THE STIENG LANGUAGE: A RIME STUDY⁹

Noëllie BON

*Université Lyon 2, Laboratoire Dynamique du Langage
<noellie.bon@hotmail.fr>*

1 Introduction: metalinguistic data

1.1 About the Stieng language

The Stieng language is a minority language belonging to the South-Bahnaric sub-group of the Mon-Khmer group in the Austro-Asiatic family¹⁰. It is spoken both in Vietnam (Binh Duong and Dong Nai provinces) and Cambodia (Kratie, Mondulhiri and Kompong Cham provinces).

The exact number of speakers is currently unknown but the community may account about 3,500 members in Cambodia and over 50,000 members in both countries¹¹.

According to Krauss' criteria (2006), Stieng is definitely endangered as the language has not been transmitted to the current generation. In Cambodia, there is a dominant bilingualism with Khmer (the official language) in Kratie and Kompong Cham provinces; and with Bu-nong (also known as Phnong and Mnong), a related language, in Mondulhiri.

Previous work on Stieng consists of lexicons and dictionaries compiled by French missionaries and administrators during the French Protectorate period¹². Then, during the 70s, 80s and 90s, Haupers & al (SIL) produced manuscripts, articles and a dictionary based on a dialect spoken in Vietnam (Stieng Bulo).

1.2 Data and Fieldwork

1.2.1 Four sets of data

This paper is based on four sets of data: two first hand data sets and two second hand data sets.

⁹ Work in progress. Follows up a presentation titled "A phonology of the Stieng language: Toward vocalic subgroupings", within the SEALS XIX Conference in Ho Chi Minh City, Vietnam (28-29 May 2009 - Bon, 2009a). This paper is also an updated version of chapter VI, MA thesis (Bon, 2009b:82-154).

¹⁰ Classification from The Mon-Khmer Languages Project, Directed by P. Sidwell, SEAlang Projects.

¹¹ Dang Nghiem Van, 1993. However, according to a recent survey from ICC - SIL (report in progress), the stieng population might be as much as 9000 in Cambodia (personal communication, Philip Lambrecht, 2010).

¹² Azémar (1886); Morice (1875), Gerber (1937); Morere (1932). We should note that Azémar (1886) is the first published dictionary on a minority language of « Cochinchina ». This dictionary was used as a basis to the founder work of Mon-Khmer linguistics published by Schmidt in 1905 (personal communication: Gérard Diffloth, 2009).

1.2.1.1 First hand data

a) *Bon 2007-2009*: a word list that I collected in Summer 2007 and winter 2009 among speakers of a Stieng dialect in Tro Peang Ron village (also known as Kbaal Snuol - Snuol commune, Snuol district, Kratie province, Cambodia). This data set represents the data I focus my analysis on in the present paper.

b) *Bon 2010*: a word list that I collected in 2010 among speakers of a different dialect of Stieng in the village of Dey Kraham (Pii Thnu commune, Snuol district, Kratie province, Cambodia). In the present paper, I occasionally use this recent data set for comparison purposes.

1.2.1.2 Second hand data for comparison purposes

a) *Haupers & Haupers 1991*: a Stieng dictionary compiled among speakers of the Bulo dialect of Stieng (Vietnam) between 1960 and 1975.

b) *Sidwell 2000*: Sidwell's reconstruction of Proto-South Bahnaric (PSB).

1.2.2 Fieldwork

I conducted 2007 and 2009 data collection sessions with two main speakers, both women (mother and daughter of 66 and 45 years old each). The corpus contains 1270 lexical entries collected on the basis of Greenberg's List, the EFEO List completed by Marie Martin and direct elicitation.

Then one should underline that the Stieng variety spoken in the area of Snuol is definitely influenced by the Khmer language: the contact between both communities started around the XVII^e Century¹³. Thus, many loan words from Khmer are part of the lexicon of these speakers. Consequently, one difficulty of the study was to recognize these borrowings as Stieng speakers who also speak Khmer, keep some features of their Stieng pronunciation and of their regional accent in Khmer (Snuol).

1.3 Framework

This paper aims to present my work in progress about the rime patterns of the variety of Stieng spoken in Kbaal Snuol, in both a synchronic and a diachronic point of view.

I start with briefly introducing the most salient particularities of the Stieng phonology such as word and syllabic canon and consonant systems (part 2.). The aim of the following section is to establish the native phoneme inventory of my data (07-09) (part 3.). Finally, I present the different occurring and non-occurring rime patterns and asymmetries, underlying their direct relation with vowel shifts and splits from Proto-South Bahnaric to modern Stieng.

2 Phonological features of the language

2.1 Word and syllabic canon

One areal and typological feature of Stieng is the syllabic and word structure: simple words can be monosyllabic or sesquisyllabic (i.e. one syllable and a half). The latter are composed of two types of syllables: one weak syllable, which is light, with a non-

¹³ Personal communication: Mathieu Guérin (2009).

phonemic vowel¹⁴, and one main syllable which is heavy, with a phonemic vowel - with the stress falling on the main syllable. Both type of syllable can be open or closed.

The figure below illustrates the word and syllable canon of the language:

(C(C) -ə(C)). C(C) -V(C)
(weak syllable). main syllable

Figure 1: *Word and syllable canon of Stieng*

The onsets of both types of syllables (weak and main syllable) can be a single or a sequence of two consonants. The nuclei of main syllables can be short, long vowels or diphthongs. The coda of both type of syllables are exclusively single consonants.

2.2 Consonants of Stieng

An areal feature of Sino-Tibetan and Southeast Asian languages is that consonant inventories are clearly different depending on their position in the word and in the syllable. Considering these differences, Smith (1975 in Smith 1989-1990:108) highlights the necessity of studying consonants within different subsystems depending on their position.

In Stieng, only a subset of the consonant system occurs in weak syllables. As I am not dealing with weak syllables but only with main syllables in the present paper, here are presented only the consonant subsystems of main syllables.

2.2.1 Initial consonants of Stieng

The initial consonant system (table 1) is composed of 30 units. It is different from the final consonant system (see table 2) concerning stops: it contains not only unvoiced stops but also voiced, aspirate, glottalized, and prenasalized. Within my current research, the phonological status of prenasalized remains hypothetical (Bon 2009b:77-80) and glottalized stops might come from borrowings from Khmer.

Table (1): *Initial consonants of main syllable subsystem of Stieng*

		Labial	Alveolar	Palatal	Velar	Glottal
Stop	- voiced	p	t	c	k	ʔ
	+aspirate	ph	th	ch	kh	
	+ voiced	b	d	ɟ	g	
	glottalized	ɓ	ɗ			
	prenasalized	mb	nd, ndʰ, nt		ŋg, ŋk	
Fricative			s			h
Nasal		m	n	ɲ	ŋ	
Liquid		w	l, r	j		

¹⁴ The nucleus can be deleted in realization. The result is that a weak syllable can be reduced to a single consonant. I have argued this point in Bon 2009b (chapter IV, 34-37). As the language admits also consonant clusters as onset of monosyllables, my analysis aims to determine of which type are the words that begin with consonants sequences: are they monosyllables with cluster onsets or sesquisyllables with reduced weak syllable?

The presence of voiced stops suggests that Stieng is a conservative language in a genetic and areal point of view and explains the absence of register distinction¹⁵. Here are some minimal pairs justifying the stop voicing distinction in Stieng:

	MP	St. Bon09 ¹⁶	Gloss
(1) a.	[p] vs. [b]	pu:	<i>to suck at</i>
b.		bu:	<i>someone</i>
(2) a.	[t] vs. [d]	tə:p	<i>turtle dove</i>
b.		də:p	<i>to claim one's due</i>
(3) a.	[c] vs. [ɟ]	caŋ	<i>to prune wood</i>
b.		ɟaŋ	<i>to throw</i>
(4) a.	[k] vs. [g]	kuŋ	<i>corn cob without the grains</i>
b.		guŋ	<i>stairs</i>

2.2.2 Final consonants of Stieng

Table (2): Final consonants of main syllable subsystem of Stieng

	Labial	Alveolar	Palatal	Velar	Glottal
Stop	p	t	c	k	ʔ
Fricative		s			h
Nasal	m	n	ɲ	ŋ	
Liquid	w	l, r	j		

The final consonant system contains 15 units: four stops, four nasals, one fricative, two liquids, two semi-consonants and two glottals.

Remark: a genetic feature of Mon-Khmer is the special realization of final –s: it is pronounced with a very weak friction and a palatalization: - [jç]. We will see below that final /s/ patterns with the palatals in terms of the overall rime inventory.

¹⁵ Ferlus (1979), Huffman (1976), Sidwell, (2000). One characteristic of Mon-Khmer languages is the loss of the initial stop voicing distinction: there was a confusion of voiced and unvoiced stops. Due to this phenomenon, vowels have developed a register distinction opposing vowels preceded by an original unvoiced consonant and vowels preceded by an original voiced consonant. Note that usually ‘*South-Bahnaric languages do not have registers*’ (Sidwell 2000: 6). However my second data set collected in Dey Kraham (2010) shows some distinctions that are likely to be interpreted as registers. Further data collections are planned to clarify that point in late 2010.

¹⁶ Examples are labelled as following: St. Bon09: my data (2007-2009) ; St. Bon10: my data (2010); St. Hau91: Hauper’s data (1991); *PSB Sid00: Sidwell’s reconstruction of Proto-South Bahnaric (2000).

2.3 Vowels of Stieng

Table (3): *Phonetic vowel inventory of Stieng*

	Front			Central		Back	
High	i	i:	ia	ɨ	ɨ:	u	u: ua
Mid-High	e	e:		ə	ə:	o	o:
Mid-Low	ɛ	ɛ:		ə	ə:	ɔ	ɔ:
Low				a	a:	ɑ	ɑ:

According to table (3), the phonetic vowel inventory of my Stieng data (07-09) includes 24 units.

However, this inventory does not reflect the reality: the full set never occurs either in open syllable or before any given final consonant insofar as vowel distribution is closely related to the final context.

Some gaps may be due to the small corpus available but some others are significant. A universal fact among languages of the world is that final consonants usually have many effects on the vowel. But more importantly, in the modern Stieng language (as well as other South-Bahnaric languages), the distribution of vowels depending on the final context reflects current or historical shifts in length and quality which pattern with consonants natural classes and structural hierarchies, and create structural gaps. So I aim to describe vowels depending on the way they rime with their final context to identify these structural gaps, with an attempt to give a historical explanatory account.

3 Vocalic systems

The aim of this section is to establish the native phoneme inventory of my data (07-09), and to get rid of vowels that are only on loan words, expressives or any other exceptions. As mentioned above, I aim to describe the rimes of the Stieng dialect of Kbaal Snuol in a comparative point of view, mostly regarding previous works done on the phonology of the language, that is Haupers' descriptions of Stieng Bulo (Haupers 1969 and Haupers & Haupers 1991) and Sidwell's analyses within his reconstruction of Proto-South Bahnaric (2000). Thus I start with presenting the different vocalic charts suggested in the aforementioned publications.

3.1 Vocalic systems of Stieng in the literature

3.1.1 Haupers 1969, Haupers & Haupers 1991

Haupers ends out with the following inventories:

Table (4): Stieng Bulo vowel phonemes inventory – Haupers 1969:132-133

	Front		Central		Back	
High		ii iə	ʊ	ʊʊ		uu uə
Mid	ê	êê	ɔ	ɔɔ	ô	ôô
Low	e	ee	a	aa	o	oo

Table (5): Stieng Bulo vowel phonemes inventory– Haupers 1991:vi

	Front		Central		Back	
High	i	iê	ʊ , ɨ	ʊʊ , ʊɔ	u	uô
Mid	ê	êê	ɔ	ɔɔ	ô	ôô
Low	e	ee	a	aa	o	oo

3.1.2. Sidwell 2000

Based on Haupers & Haupers 1991, Sidwell suggests the following inventory for native Stieng vowel phonemes:

Table (6): Native stieng vowel phonemes inventory – Sidwell 2000:30

	Front		Central		Back	
High	i	iə			u	uə
Mid	e	e:	ə	ə:	o	o:
Low	ɛ	ɛ:	a	a:	ɔ	ɔ:

The lack of high central vowels is justified by the following statements:

a) The three long high central vowels listed by Haupers & Haupers (1991) as /ɨ/, /ʊʊ/ and /ʊɔ/ (that is /ɨ/ and /iə/¹⁷) are rare and occur only in loan words from Khmer and Vietnamese or in nursery words. Consequently, Sidwell decides to ignore them in his reconstruction.

b) Concerning the short high central vowel listed as /ʊ/ by Haupers & Haupers (that is /ɨ/), Sidwell ends out to analyse it as an allophonic realisation of /e/ after labial and velar initials:

¹⁷ The IPA transcription of /ʊʊ/ remains unclear.

/e/ > [i]/ Lab. Vel. ____
[e]/ elsewhere (Sidwell 2000:30)

The consistency of both statements within my data is tested respectively parts 3.2.2 and 4.4.

3.2 Vocalic systems of Stieng in my data (2007-2009)

3.2.1 Vocalic phones inventories

Here are the vowel phones which occur in my data:

Table (7): *Vocalic phones inventory in open rimes – Bon 2007-2009*

	Front		Central		Back	
High		i:		ɨ:		u:
Mid-high		e:		ə:		o:
Mid-low		ɛ:		ɜ:		ɔ:
Low				a:		ɑ:

Table (8): *Vocalic phones inventory in closed rimes – Bon 2007-2009*

	Front		Central		Back	
High	i	ia	ɨ	ɨ:	u	ua
Upper-Mid	e	e:	ə	ə:	o	o:
Lower-Mid	ɛ	ɛ:	ɜ	ɜ:	ɔ	ɔ:
Low			a	a:	ɑ	ɑ:

Remark: italicized vowels have a hypothetic status. Both inventories look quite odd compared to Haupers' and Sidwell's especially because they include four height levels for central and back vowels while only three or two would be expected.

Thus the next sub-section aims to clarify the status of these particular segments (italicized), identifying whether they only occur in loan words from Khmer or they show evident contextual variation.

Identifying whether a word is a loan word or a cognate is difficult by looking only at the Khmer translation. Thus I compared the words in question with Haupers' data, Sidwell's reconstruction of Proto-South Bahnaric and occasionally my data set (Bon 2010) when the latter was offering consistent help. Besides identifying the loan words, comparison was useful for identifying some contextual variation phenomenon as well.

3.2.2 Patterning of vowel phones

3.2.2.1 Open rimes

a) Front vowels /e:/ and /ɛ:/ in open rimes

The front vowel /e:/ occurs only in two words that can be loan words from Khmer (examples (5)). Then, words with /ɛ:/ actually correspond to words with rime -ej from Haupers and Bon10 that are reconstructed *e: by Sidwell (examples (6)):

	<i>St. Bon09</i>	<i>St. Bon10</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Khmer</i>	<i>Gloss</i>
(5) a.	phe:	phe:	---	---	phe:	otter
b.	de:	de:	---	---	te:	no
(6) a.	pɛ:	pej	pej /pe:/	*pe:	ɬæj	three
b.	che:	chej	chej /cəhe:/	*kəse:	ksæ	rope, cord
c.	kɛ:	kej	kej /ke:/	*nəke:	sna:n	horn
d.	bə.be:	bej	bej /be:/	*be:	popɛ:	goat

One should note that in Chrau, a related language, *e: shifted to /ɛ:/ in open rimes as well (Sidwell 2000:47).

So far, the status of /e:/ in my data 09 remains hypothetical but it seems consistent to maintain it in the inventory as according to Sidwell's analysis it occurs in Haupers' lexicon as well as in the pre-Stieng vocalic system:

*e:	>	e:	(Sidwell 2000:47-50)
/e:/	→	[ej] / ____#	(Sidwell 2000:30)

b) Central vowels /i:/ and /ə:/ in open rimes

As mentioned in part 3.1.2.1, Sidwell found out in Haupers' data, that /i:/ and /ə:/ occur in loan words from Khmer and Vietnamese or nursery words, both in open and closed rimes, so he does ignore them in the native vowel system of Stieng (Sidwell 2000:29).

This statement is consistent with my data where /i:/ and /ə:/ occur only in presumed loan words from Khmer, and so does /ə:/ . Consequently, I remove /i:/ and /ə:/ from the phonological inventory.

The following list records all examples of these three phones in my data:

	<i>St. Bon09</i>	<i>St. Bon10</i>	<i>Khmer</i>	<i>Gloss</i>
(7) a.	chi:	chi:	chə:	tree, wood
b.	krə.ɬi:	krə.ɬi:	krəɬæj	water buffalo
c.	prə.ɬi:	prə.ɬi:	pteji	spinach
d.	rə.sɪ:	rə.sɪ:	rɪsej	bamboo
(8) a.	krə.bə:	krə.bə:	krəpə:	crocodile
b.	smə:#gə.na:	smə:#gə.na:	smaə#khnɪə	to be equal to
c.	rə:	rə:	rə:	to disassemble
(9) a.	jə:	jəw	cuə	to believe
b.	lə:	ləw	lə:	above
d.	pə.jə:	pə.jəw	phɲə	to throw

The status of loan words is less clear for /ə:/, however given that the pre-Stieng vocalic system did not list *ə: in open rimes (Sidwell 2000:47), it would be plausible to omit it in the phonological system.

c) *Back vowel /ɑ:/ in open rime:*

/ɑ:/ occurs in one single word which could be a loan word from Khmer:

	<i>St. Bon09</i>	<i>Khmer</i>	<i>Gloss</i>
(10)	ga:	kəo	<i>dumb</i>

Thus /ɑ:/ is omitted from the inventory for the present study.

3.2.2.2 Closed rimes

a) *Central Vowels in closed rimes:*

Firstly, the central short vowel transcribed [ɪ] occurs only in one word which is a loan from Sanskrit: bɪs ‘snake poison’.

Secondly, the central long vowel [ə:] mostly occurs in presumed loan words from Khmer in closed rimes, which is consistent with Sidwell’s statement¹⁸. There are nevertheless some exceptions that are omitted in the present paper¹⁹. Here are listed some examples:

	<i>St. Bon09</i>	<i>Khmer</i>	<i>Gloss</i>
(11) a.	kə:t	kaət#ko:n	<i>be born</i>
b.	pə:c ; bə:c	baoc	<i>remove the hair from</i>
c.	bə:k	baək	<i>to drive</i>
d.	n.cə:m	cɛŋ#caəm	<i>eyebrow</i>
e.	mə:n	mə:n	<i>ten thousand</i>
f.	pə:ŋ	paəŋ	<i>to glide</i>
g.	gə:l	kə:l	<i>bumpy</i>
h.	hə:j	haoj	<i>already</i>

So it is plausible to remove it from the phonologic inventory.

Thirdly, the central long vowel [ɛ:] has a restricted distribution in closed rimes: it precedes only nasal -ŋ and rhotic -r, found in only four examples in the corpus. Two examples look like loan words from Khmer (examples (12)) but two other ones are definitely not loans (examples (13)):

¹⁸ See again 3.1.2. (a) and above: central vowels in open syllable.

¹⁹ That are: mbə:j ‘to dream’ which remains unclear ; pə:k ‘to open’, də:k ‘insolent’ and chə:c ‘sour, bitter’ which I presume to have transcription mistakes after comparison with my data 2010, Haupers’ data and Sidwell’s PSB reconstruction.

	<i>St. Bon09</i>	<i>Khmer</i>	<i>Gloss</i>
(12) a.	pə.sɪr	sɑːsəə ; təsəə	to compliment
b.	hɪr	haə#poa	to change color
(13) a.	cɪr	trɒnuŋ	backbone (fish)
b.	dɪŋ	tiɛl	blunt

So the hypothesis that [ɪ] only comes from Khmer borrowings is not plausible²⁰. Nevertheless, I omit this vowel from the phonological system, for the present study as it is occurring in a very limited number of examples.

Finally, the status of the short central vowel [ə] remains unclear. This vowel corresponds to /ʉ/ (/ i/) listed by Haupers apart from some rare exceptions that are omitted here. According to Sidwell (2000:30) this vowel [ɪ] might not belong to the native vocalic system of Stieng but is rather an allophone of /e/ after labials and velars:

/e/ > [ɪ] / Lab. Vel. ____
[e] / elsewhere (Sidwell 2000:30)

Note that this analysis assumes that [e] before palatals c, ɲ is an allophone of /ə/:

/ə/ > [e] / ____ c, ɲ
[ə] / elsewhere (ibid.)

In a broader perspective, interpretation problem related to [ə] seems to be dependant of the development of *i reconstructed by Sidwell (2000:49). I will give an in-depth demonstration to define the status of [ə] part 4.4.

b) Back vowels in closed rimes

First, long back vowel [ɑ:] occurs only in presumed loan words from Khmer or Pali. Thus I omit it in the present study:

	<i>St. Bon09</i>	<i>Khmer</i> ²¹	<i>Gloss</i>
(14) a.	ʔɑ:r	ʔɑ:	glad
b.	ʔa.sɑ:r	ʔak#sa:	letter (alphabet)
(15) a.	cha:k	cha:ʔ	bald
b.	ka:k	ka:ʔ	to freeze
c.	sa:k	sa:ʔ	to peel

²⁰ Note that [ɪ] is in complementary distribution with [ə:] in closed rimes as [ə:] never occurs in rhotic and nasal palatal rimes. Consequently we could think that [ɪ] is an allophone of /ə:/ before -ɲ and -r. However both final consonants cannot really be grouped into natural classes and there is no articulatory or historical reason that can explain why such a phenomenon happens in these particular contexts and not in the others.

²¹ Remark: Khmer final -r was recently lost. According to Ferlus (1992:72, in Phal Sok (2004:120)) the total loss of final -r might have happened during the XXe century.

Second, short back vowels are characterized by a lower realization: [ɑ] corresponds to /ô/ (/ɔ/) described by Haupers and to *ɔ reconstructed by Sidwell. Consequently, [ɑ] is transcribed /ɔ/:

	<i>St. Bon09</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(16) a.	ʃɑk /ʃɔk/	ʃɔk /ʃɔk/	*ʃɔk	<i>long time</i>
b.	klɑʔ /klɔʔ/	klɔʔ /kəɔʔ/	*kəɔʔ	<i>navel</i>
c.	kə.tah /kətəh/	kətəh /kətəh/	*gətəh	<i>breast</i>
d.	sɑŋ /sɔŋ/	sɔŋ /sɔŋ/	*sɔŋ	<i>straight</i>

Then, [o] and [ɔ] both correspond to /ô/ (/ɔ/) described by Haupers and to *u reconstructed by Sidwell. Regardless three words that I omitted in the present study²², [o] and [ɔ] are in complementary distribution according to the final consonant:

/o/ → [ɔ] / ____ dental ; velar
[o] / elsewhere

	<i>St. Bon09</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(17) a.	sɔt /sot/	sot /sot/	*sut	<i>honeybee</i>
b.	sɔr /sor/	sor /sor/	*sur	<i>porc</i>
c.	blɔk /bəɔk/	bɔk /bəɔk/	*bəɔk	<i>elephant tusk</i>
d.	tɔŋ /toŋ/	toŋ /toŋ/	*tuŋ	<i>steal</i>

This interesting distribution of the short back vowels makes sense in a diachronic point of view. Indeed, according to Sidwell, Proto-South Bahnaric vowel *u was affected by a lowering phenomenon to /o/ (Sidwell 2000:49). From this perspective, my data may show another stage of lowering as a feeding phenomenon: after the lowering of *u to /o/, there is currently a conditioned lowering of /o/ to [ɔ] and a systematic lowering of /ɔ/ to [ɑ] which might be a shift in progress, to ensure distinction and avoid homophony:

/o/ → [ɔ]/____ dental; velar
/ɔ/ > [ɑ]

3.2.3 Phonological vowel systems

Many segments discussed above can be removed from the system unless new data come contradict this decision, within further data collection. So far, my charts can be reshaped as following:

²² The words in question are: hɔp 'Never mind !' which is an expressive ; jon 'to offer' which might be a loan word from khmer cu:n ; and gon 'to draw water' for which I do not have any hypothesis.

Table (9): *Vowel system in open rimes – Bon 2007-2009*

	Front		Central		Back	
High		i:				u:
Mid		e:				o:
Low		ɛ:		a:		ɔ:

Table (10): *Vowel system in closed rimes – Bon 2007-2009*

	Front		Central		Back	
High	i	ia	ə		u	ua
Mid	e	e:	ə	ə:	o	o:
Low	ɛ	ɛ:	a	a:	ɔ	ɔ:

Basically, the above inventories are quite consistent with Sidwell's native phoneme inventory (Sidwell 2000:30).

4 A rime study of Stieng

4.1 Presentation and general statements

One of the essential problems in describing the segmental phonology of Stieng is the collocation restrictions on vowels and finals within rimes. Indeed, it appears that *vowels are or have been shifting in patterns that correlate with natural classes and structural hierarchies, creating structural gaps*²³.

In this section I aim to describe the patterns of occurring and non-occurring rimes, identifying structural gaps and their correlation with vowel's current and historical shifts.

4.1.1 Inventory of rimes:

Table 11 next shows the different occurring and non-occurring rimes.

4.1.2 General statements: three types of rimes

Rimes of Stieng can clearly be divided into three different types: open, glottal, and closed. These different types underline that there is a requirement for the main syllable to be heavy (while as already mentioned, the weak initial syllable can be light):

- open rimes are always long ;
- glottal rimes (with final /h/ or /ʔ/) always have a short vowel and diphthongs.
- and closed rimes with other consonants can have either short, long vowels and diphthongs.

Obviously the open rimes are in complementary distribution with glottal rimes. Then, vowel length is contrastive only in closed rimes with non-glottal finals.

Next section is a description of the rimes patterns and asymmetries found out between the rimes looking at length opposition (4.2.), articulation point (4.3.) and height (4.4.) features.

²³ Note of the reviewer.

One should already note that Sidwell (2000:47-48) argued for a chain of shifts in Stieng concerning the high vowels that is:

- a shortening of *u: and *i:, respectively to /u/ and /i/
- a lowering of *u and *i, respectively to /o/ and /e/.

These changes account for many of the synchronic asymmetries.

Table (11): Rimes of Stieng - Bon09

Nuclei \ Coda				Labial			Dental				Palatal				Velar		Glottal		Zero
				p	m	w	t	n	r	l	c	ɲ	j	s ²⁴	k	ŋ	ʔ	h	∅
S H O R T	F R O N T	High	i	---	im	---	it	in	ir	---	ic	ip			ik	---	--	ih	
		Mid	e				---	---		---	ec	eɲ	---		---	---	eʔ	eh	
		Low	ɛ														ɛʔ	ɛh	
	C E N T	High	ə	əp			ət	ən	ər						əŋ	--			
		Mid	ə	əp	əm		ət	ən	ər	əl					ək	əŋ	əʔ	əh	
		Low	a	ap	am	aw	at	an	ar	al	ac	aɲ	aj	as	ak	aŋ	aʔ	ah	
	B A C K	High	u	up	um		ut	un	ur	ul	uc		uj	us	uk	uŋ		uh	
		Mid	o	---	om	---	ot	on	or	ol	oc	oɲ	oj	os	ok	oŋ	oʔ	oh	
		Low	ɔ	---	am		---	an	ar	al	---				ak	aŋ	aʔ	ah	
L O N G	F R O N T	High	i:															i:	
		Mid	e:	e:p	---	---	e:t	e:n	--	e:l					e:k	e:ŋ			e:
		Low	ɛ:	ɛ:p		---	ɛ:t			---					ɛ:k	ɛ:ŋ			ɛ:
	C E N T	Mid	ə:	ə:p	ə:m		ə:t	ə:n	--	ə:l	ə:c	---	ə:j	ə:s	ə:k	ə:ŋ			
		Low	a:	a:p	a:m	a:w	a:t	a:n	a:r	a:l	a:c	a:ɲ	a:j	a:s	a:k	a:ŋ			a:
	B A C K	High	u:																u:
		Mid	o:	o:p	o:m		o:t	o:n	---	o:l	o:c	o:ɲ	o:j	o:s	o:k	o:ŋ			o:
		Low	ɔ:	ɔ:p	ɔ:m		---	ɔ:n	ɔ:r	---	ɔ:c	---	---	ɔ:s	ɔ:k	ɔ:ŋ			ɔ:
Diph.	Front	High	ia	iap	iam	iaw	iat	ian	iar	ial					iak	iaŋ		iah	
	Back		ua			uat	uan		ual	uac	uaɲ	uaj	uas						

--- : existing rime in Haupers and/or Sidwell's reconstruction

e: : rime with a remaining hypothetical phonological status

²⁴ Remark: as already mentioned, /s/ is realized with a weak friction and palatalized: - [jç]. Then, given its behavior within the rime patterns, it is consistent to list it in the palatal natural class.

Remark: rimes that only occur in presumed Khmer loan words as well as expressives and onomatopoeia are omitted from this table²⁵.

4.2 Length opposition asymmetries

4.2.1 High vowels /i/ and /u/

According to their distribution within the rimes, the high vowels /i/ and /u/ are in complementary distribution with their long equivalents /i:/ and /u:/: the short ones occur only in closed rimes while the long ones occur only in open rimes. Consequently, length opposition of high vowels /i/ and /u/ does not exist at all.

There are nevertheless two possible interpretations concerning the length opposition of high vowels /i/ and /u/ in closed rimes.

4.2.1.1 No length opposition

In his reconstruction of Proto-South Bahnaric, Sidwell (2000:29) suggests that the lack of long high vowels /i:/ and /u:/ in closed rimes could be due to diphthongization or shortening of both vowels. His proposal is to consider that there were a shortening of *i: and *u: respectively to /i/ and /u/²⁶ (with nevertheless a split before final glottal -h): '[...] *[T]he restructuring of Stieng which eliminated long high vowels is clearly explained as a shift from long to short*' (Sidwell 2000:47):

*i: > e / ____ h²⁷
i / elsewhere (ibid.)

*u: > o / ____ h
u / elsewhere (ibid.)

4.2.1.2 Length opposition

We can suggest to interpret the diphthongs /ia/ and /ua/ as the long equivalents of /i/ and /u/, respectively coming from the pre-Stieng long high vowels *i: and *u:. Accordingly, length opposition for high vowels does exist in closed rimes: /i/ vs. /ia/ and /u/ vs. /ua/.

Then, looking at the rimes patterns, we notice that:

-/i/ and /ia/ are in opposition in all rimes except before palatals (see part 4.3.4)

-/u/ and /ua/ are in opposition in rimes with dental and palatal rimes but not in labial (see part 4.3.4.) and velar rimes where /ua/ does not occur.

4.2.2 Mid and Low Front vowels /e/ and /ɛ/

First, short vowel /e/ and /ɛ/ have quite a marginal status, according to their limited distribution within the rimes: /e/ only occurs in palatal and glottal rimes and /ɛ/, only in

²⁵ Omitted words are listed Appendix I.

²⁶ Sidwell specifies that 'it is not clear that it is the source of all short high vowels' (Sidwell 2000: 29).

²⁷ Note that in my data (09), *i: did not lower to /e/ before -h but shortened to /i/ as elsewhere (e.g.: *pi:h > pih 'knife'). The lowering occurred for the back vowels though (e.g.: *tu:h > toh 'bean').

glottal rimes. I will discuss the status of [e] in palatal rimes more precisely part 4.3.2. and explain the limited distribution of /e/ part 4.4.

Concerning /ɛ/, it seems that the vowel comes from a split of PSB *ə in glottal rimes:

*ə > ɛ / ____ ʔ, h (ibid.)

So in my data, as well as in Hauper's, the limited distribution of /ɛ/ in glottal rimes can be consistently explained by the fact that /ɛ/ merged in the system in this particular glottal context, probably quite recently.

According to Sidwell (2000:48): *'/ɛ/ is marginal phoneme in the modern languages contrasting reliably with [ə] only before glottals [...]'*. This statement is consistent with my data²⁸.

Second, as /e:/ and /ɛ:/ never occur either in palatal or glottal rimes, there is no length opposition for mid and low front vowels in closed rimes.

4.2.3 Other vowels

Other vowels all show some length oppositions in closed rimes. Here are the different length opposition sets from the most frequent to the less frequent: /a/ vs. /a:/; /ə/ vs. /ə:/; /o/ vs. /o:/; /ɔ/ vs. /ɔ:/.

/a/ and /a:/ are present and in oppositions in all closed rimes, with any given final consonant. One should note that, according to Sidwell's reconstruction (2000:47), both occurred in the Proto-South Bahnaric vowel system and did not shift or split.

Length opposition is attested in every velar rime for every vowels mentioned above (/a/ vs. /a:/; /ə/ vs. /ə:/; /o/ vs. /o:/; /ɔ/ vs. /ɔ:/).

There is only one length opposition with the labial semi-consonant: –aw vs. –a:w. In a general point of view, rimes with semi-consonant -w are very limited and found only with nuclei /a/; /a:/ and /ia/. Consequently, one should address the possibility of these rimes to be triphthongs in a synchronic point of view. However, this is not the purpose of the present paper so I move on to the articulation point asymmetries.

4.3 Articulation point asymmetries

4.3.1 Front vowels

4.3.1.1 General remarks

In a general point of view, front vowels occur in a limited number of rimes in my data: in a diachronic point of view, as already specified (4.2.2.), that limitation can be explained by the fact that /e/ and /ɛ/ are new in the modern language and come from the following shifts and splits:

e < *i
 *i:/ __h
 *ə / __s

²⁸ e.g.: lɛʔ 'finished completed' vs. rə.ləʔ 'water melon'; dɛh 'give birth' vs. dəh 'near'.

ε < *ə / __ ʔ, h

4.3.1.2 Front vowels in palatal rimes

There are evident asymmetries of front vowels in palatal rimes:

The only front vowels which can occur in palatal rimes (that is with –c and -ɲ) are [e] and [i]. The status of [e] in palatal rimes is clarified in the next subsection (4.3.2.).

Then, front vowels never occur in rimes with fricative –s which might be due to its palatalized realization: [jç]. Note that *s is likely to have shifted to /h/ after *ə (with *ə > e / __s):

*s > h / e__ (Sidwell 2000:39)

4.3.1.2 Special case of [i] in palatal rimes

In my data, [i] can occur before –c and -ɲ whereas it cannot in Haupers' data and Sidwell's PSB reconstruction. Thus, here I aim to give an explanation of this collocation looking at the correspondences with Haupers' data and Sidwell's reconstruction. The following list records all examples of these two rimes in my data:

	<i>St. Bon09</i>	<i>St. Bon10</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(18) a.	pic	pec	pik /pik/	*səpi:k	<i>civette cat, weasel</i>
b.	ɟic	ɟec	ɟik /ɟik/	*ɟi:k	<i>to hoe weed; to dig up</i>
c.	ric	rec	---	---	<i>to grow up, to flower</i>
d.	wic	wec	---	---	<i>nocturnal bird</i>
(19) a.	prɪɲ	preɲ	---	---	<i>vegetal oil</i>
b.	tɪɲ	teɲ	tiɲ /tiɲ/	*nəti:ɲ	<i>bone</i>
c.	hiɲ	heɲ	---	---	<i>bullfrog</i>
d.	səm.liɲ	səm.leɲ	---	---	<i>voice</i>
e.	ndə:r#wɪɲ	ndər#wɛɲ	---	---	<i>big scorpion</i>

According to Sidwell's reconstruction (2000:39) there was a palatalization of final *k after long *i: in Stieng:

*k > c / *i:__

Reminder:

*i: > i (Sidwell 2000:29 and 47).

Strangely, Hauper's data do not reflect this split as the proto final velar is still a velar in the modern language:

*ɟi:k > ɟik

*səpi:k > pik

*... (Sidwell 2000:83-84)

So I presume that there is a mistake related to some transcription confusion, knowing that final –k is transcribed –c in Haupers & Haupers 1991 and the correct analysis may be:

*k > c / *i__ (ibid.)

Examples:

	<i>St. Bon09</i>	<i>St. Bon10</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(20)	a. bec	bec	bec /bæc/	*bik	to lie down
	b. tec	tec	tec /tæc/	*tik ~ *tæc	deaf

This analysis of palatalization of *k after long *i: fits in my data, except for one single example where the velar is not palatalized:

	<i>St. Bon09</i>	<i>St. Bon10</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(21)	kəm.lik	plek	pəlik /pəlik/	*pəni:k ~ *pəli:k	shoulder

Concerning the nasal velar, there is no palatalization of *ŋ either after long *i: and short *i in Stieng, according to Sidwell's reconstruction (2000:39): *ŋ > ŋ (see example 19.b. above).

One should note that [i] never occur in nasal velar rimes in my data.

So according to these different statements, I propose four interpretations:

a) Rimes -ic and -iŋ do not exist in my data, they are transcription mistakes and they should be revised respectively as -ik and -iŋ.

b) There is a synchronic palatalization of velars after /i/ - omitting the single example of rime -ik (see example (21)) - in my data.

c) In the dialect of Stieng reflected by my data, there was a diachronic palatalization of velars after *i:, followed by a shortening of *i:.

*i:k > ik > ic
 *i:ŋ > iŋ > iŋ

d) In the dialect of Stieng reflected by my data, there was a diachronic palatalization of velars after long *i:.

*i:k > ic
 *i:ŋ > iŋ

4.3.2 Central vowels

According to the rime table (4.1.1.), there is no short central vowel /ə/ in palatal rimes in my data.

In Haupers & Haupers' (1991) and Yeem's (1977) data, [e] is the only front vowel occurring in palatal rimes. Consequently, Sidwell interprets [e] as an allophone of /ə/ in that particular context:

/ə/ > [e]/ _ c, ɲ
 [ə]/ elsewhere (Sidwell 2000:30)

In my data, [e] is not the only front vowel occurring in palatal rimes as [i] can also occur before palatals –c and -ɲ (see above). However the status of these rimes with [i] remain unclear. Then Sidwell’s interpretations fit in my data and I assume that [e] is indeed an allophone of /ə/ before –c and -ɲ:

	<i>St. Bon09</i>	<i>St. Hau91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(22) a.	tec /təc/	tec /təc/	*tik ~ *təc	deaf
b.	bec /bəc/	bec /bəc/	*bik	to lie down
c.	dec /dəc/	dec /dəc/	*dəc	slave
d.	dɛɲ /dɛɲ/	dɛɲ /dɛɲ/	*dɛɲ	bamboo tube
e.	ʃɛɲ /ʃɛɲ/	ʃɛɲ /ʃɛɲ/	*ʃɛɲ	sew
	...			

4.3.3. Back vowels

Paralleling front vowels asymmetries with palatal rimes, we could predict less back vowels in labial rimes but it is not the case, apart from rimes with –w where back vowels never occur (see 4.3.4. below).

Then, in my corpus, there are many gaps with the short low back vowel /ɔ/ which may be meaningful as the proto-equivalent *ɔ was restricted to velar and glottal rimes (Sidwell 2000:48).

Recall the presumed lowering feeding phenomenon happening with back vowels (end of 3.2.2).

*u	>	o
/o/	→	[ɔ]/__dental ; velar
*ɔ	>	ɔ
/ɔ/	>	[ɑ]

4.3.4. Semi-consonant rimes and diphthongs asymmetries:

In a general point of view, a limited variety of vowels occur in semi-consonant rimes.

A noticeable asymmetry is that front vowels never occur with the semi-consonant palatal –j and similarly back vowels never occur with the semi-consonant labial –w:

*j / V [+ant; +high]	___
*w / V [+post; +high]	___ ²⁹

According to Swantesson (1988:72 in Sidwell 2000:13), this asymmetry is quite frequent among Mon-Khmer languages.

Another asymmetry is that front diphthong /ia/ never occurs in palatal rimes and back diphthong /ua/ never occurs in labial rimes.

4.4. Height asymmetries: the special case of [ə]

As announced part 3.2.2, we have to clarify the status of the central high vowel that I transcribed [ə] in my data, transcribed /ʊ/ (/ɪ/) in Haupers.

²⁹ Here * refers to “prohibited”.

According to Sidwell (2000:30), in Haupers' data, this vowel [ɪ] might not belong to the native vocalic system of Stieng but is rather a variant of /e/ after labials and velars:

/e/ > [ɪ]/ Lab. Vel.
[e]/ elsewhere³⁰ (ibid.)

This allophony is likely to be related to the development of the proto-front vowel *i (Sidwell 2000:49). However, both in Haupers data and my data, [ɪ] can also occur after palatal ch- and ʃ-, dental t- and s-, and glottal ʔ-:

	<i>St. Bon 09</i>	<i>St. Hau 91</i>	<i>*PSB Sid00</i>	<i>Gloss</i>
(23) a.	təp	---	*kətip	cockroach
b.	sən	sen ; sin /sen/	*sin	to cook
c.	chət	chit /cəhet/	*kəsit ~ *kəset	to die
d.	ʃəl	ʃil /ʃel/	*ʃil	deer
e.	kʔəp	kʔip /kʔep/	*kərʔip	centipede

Similarly, and regardless rimes with palatal -c, -ɲ (see rule in footnote 23), it appears that [e] can occur after labials and velars both in Haupers' data and my data:

	<i>St. Bon 09</i>	<i>St. Hau 91</i>	<i>*PSB Sid. 2000</i>	<i>Gloss</i>
(24) a.	beh	beh /beh/	*bəʂ	snake
b.	geh	geh /geh/	*gih ~ *geh	to snap
c.	pih ³¹	peh /peh/	*pih ~ *pi:h	knife
d.	---	keh	---	trigger
e.	---	kheh	---	to be poisonous
f.	---	geʔ	---	to be small

That said, according to my data, there is actually a complementary distribution between [ə] and [e] depending on the final context³²: [e] occurs only before glottal -ʔ, -h and [ə] occurs elsewhere. As [ə] occurs in the unconditioned environment I propose the following rule, in a strict synchronic point of view:

/ə/ → [e]/___ glottal
[ə] / elsewhere

But, for comparative purposes, another analysis is plausible:

³⁰ Reminder: /ə/ > [e] / ___c, ɲ [ə]
/ elsewhere

³¹ Reminder : the rule *i: > e/___h built by Sidwell (2000:47) with Haupers' data is not consistent with my data where *i: is only shortened to /i/.

³² I omit the rime -əh about which I am not quite sure comparing the words in which it occurs with Sidwell's reconstruction, Haupers data and my new data set (10). So I assume that there might be some transcription mistakes.

*i > /e/
 /e/ → [e]/___ glottal
 [ə] / elsewhere

With:

/ə/ > [e]/ _____ c, ɲ
 [ə]/ elsewhere

One should note indeed that in Chrau, a related South Bahnaric language, the system doesn't list the short front vowel /e/ (Thomas & Luc 1966 in Sidwell 2000:24). However the high front vowel /i/ has exactly the same distribution as /e/ from my data.

Then Sidwell gives the following interpretation for Chrau short vowels [i, ɪ, ə] (2000:25), which fits with mine:

/i/ > [i]/___ ʔ, h
 [ɪ]/elsewhere (Sidwell 2000:25)

/ə/ > [i]/ _____ c, ɲ
 [ə]/ elsewhere (ibid.)

As already mentioned, this problem is likely to be related to the development of *i.

One finally should note that Sidwell also found out that *i split to /o/ before labials in Stieng Haupers' data:

*'Recalling that I reconstructed the split of *i to [e] and [ɪ] in Stieng, it appears that a further rule is required, namely *i split a third way, to [o] before labials. Consulting the Stieng lexicon, one finds no example of [e] or [ɪ] before labial terminals.'*
 Haupers (2000:48-49)

However, this statement is not consistent with my data, [ə] occurring before labial finals:

Chrau	St.Hau91	St. Bon09	Köho	gloss	*PSB Sid00
sim	com	(paj#penar)	si:m	bird	> * [ʔ] sim
kəʔip	kəʔop	kəʔəp	kəʔip	centipede	> * kəʔip
cip	---	təp	tip	cockroach	> * kətɪp

5. Conclusion

We have seen that rimes first pattern depending on the length feature of vowels, into three kinds of rimes: open, glottal, closed.

In closed rimes, we notice a tendency of avoidance for vowel and consonant of same articulation point to co-occur especially with the semi-consonants (back and front vowels respectively with palatal and labial finals³³).

³³ That is not true with the fronted realization of /ə/ before palatals -c and -ɲ where, on the contrary, there is an assimilation of articulation point to [e].

Then the main point of this paper is that historical vowel shifts conditioned by the final context account for the synchronic distribution of vowels. Noticeably, the synchronic distribution of the front and back vowels is a consequence of a chain shift that involved a lowering of *u and *i to /o/ and /e/, and a shortening of *u: and *i: to /u/ and /i/ (argued by Sidwell 2000).

Thus vowels coming from *i and *u might have passed through different neutralization phenomenon, at different stages of the history of Stieng, as specified by Sidwell and outlined in this paper: ‘*There has been various neutralisations of /i/ and /u/ in the histories of Chrau and Stieng most often realized as [ɨ].*’ (Sidwell 2000: 49)

My data reflect the development of *i and *u in a split, respectively to [e] and [ə]; and [o] and [ɔ].

The paths the vowels took might not be the same from one Stieng dialect to another as we have seen by comparing my own data with Haupers’, unless my data reflect a more recent stage of Stieng evolution, as I conducted fieldwork about 50 years after Haupers. Nevertheless, it seems that in my data, vowels took paths that are more similar to Chrau’s, a related language, as we have seen with the shift of *e: in open rimes and the development of *i.

Finally, some vowel shifts or splits operating currently in the dialect of Kbaal Snul (my data 07-09) are visible, such as the lowering of /ɔ/ to [a] as a feeding phenomenon to compensate the lowered realization of /o/ before dentals and velars.

References:

- Azémar, H. 1886. *Dictionnaire Stieng*. Saïgon, Imprimerie coloniale.
- Bon, N. 2009a. 19th *Annual Meeting of the Southeast Asian Linguistics Society*. University of Social Sciences and Humanities USSH). Vietnam National University, Ho Chi Minh City. Paper read: A phonology of the Stieng language: toward vocalic subsystems. (28-29 may).
- Bon, N. 2009b. *Une introduction à la langue stieng: profil sociolinguistique et principes phonologiques*. Mémoire de Master 2. Lyon, Université Lyon 2.
- Ferlus, M. 1992. Essai de phonétique historique du khmer. *The Mon-Khmer Studies Journal* 21:57-89.
- Ferlus, M. 1979. Formation des registres et mutations consonantiques dans les langues Mon-Khmer. *The Mon-Khmer Studies Journal* 8 :1-76.
- Gerber, T. 1937. *Lexique franco-stieng*. Saïgon, Imprimerie du Théâtre.
- Haupers, R. 1969. Stieng phonemes. *The Mon-Khmer Studies Journal* 3:131-137.
- Haupers, R. 1962. Word-final syllabic in Stieng. *VHNS* 11:846-848.
- Haupers, R. & Diêu, B. 1967. *Stieng phrase book*. Santa Ana, CA, SIL.
- Haupers, R. & Haupers, L. 1991. *Stieng-English Dictionary*. Compilers, SIL.
- Huffman, F.E. 1976. The register problem in fifteen Môn-Khmer languages, *AS* 1:575-90.
- Krauss, M. 2006. Classification and Terminology for Degrees of Language Endangerment. M. Brenzinger (Ed.), *Language Diversity Endangered*. Berlin-New York, Mouton de Gruyter, pp. 1-8

- Matisoff, J. 1973. Tonogenesis in Southeast Asia. L. M. Hyman, (Ed.), *Consonant Types and Tone*. Southern California Occasional Papers in Linguistics 1: 71-95. Los Angeles, UCLA.
- Morere, M. 1932. *Essai de vocabulaire francais-stieng*. Saigon, J. Viêt et Fils.
- Morice, A. 1875. *Etude sur deux dialectes de l'Indochine, Les Tiams et les Stiengs*. Paris, Maisonneuve.
- Phal, S. 2004. L'amuissement du r final en khmer: allongement et diphtongaison. *The Mon-Khmer Studies Journal* 34:113-136.
- Sidwell, P. 2000. *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra, Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Sidwell, P. 1993. *The Proto-Bahnaric Vowel System: a new theory*. La Trobe University Working Papers in Linguistics 6:137-142
- Smith, K. 1989-1990. The -VC rhyme link between Bahnaric and Katuic. *The Mon-Khmer Studies Journal* 18-19:106-159.
- Swantesson, J.O. 1988. U. *Linguistics of the Tibeto-Burman Area* 11.1:64-133.
- Thomas, D.D. 1965. Checking vowel contrasts by rhyming. *The Mon-Khmer Studies Journal* 2:99-102.
- Yeem, D. 1977. *Stieng, Vietnamese & English Thesaurus*. Dallas, Texas, Summer Institute of Linguistics.

Appendices

II. List of exceptions

Table A: *Loan words and expressives*

<i>Rime</i>	<i>Bon09</i>	<i>Bon10</i>	<i>Khmer</i>	<i>gloss</i>
ɛ:c	cɛ:c(#gə.na:)	ceəc(#gə.na:)	ca:c	<i>to split (1)</i>
	trə.ɬɛ:c	tə.ɬeəc	traɬac	<i>to split (2)</i>
	re:c(#da:k)	reəc(#da:k)	re:c(#tək)	<i>to carry water</i>
	wɛ:c	wəc	wɛ:c	<i>ladle</i>
ɛ:r	kɬɛ:r	kbeər#gəna:	kɬæ	<i>near</i>
ej	sbej	sə.bej	spej	<i>mustard</i>
	pə.rej	pə.rej	bərej	<i>cigaret</i>
ɛt	ɟɛt	ɟət	Ø	<i>ten</i>
	grɔt#gret	grɔt#grat	kokrim#kokrəm	<i>rough</i>
ɛw	(paj#)mew	(pa:j#)mɛo	cma:	<i>cat</i>
ɛ:ɲ	re:ɲ	reəɲ	kra:ɲ	<i>plaited</i>
ɔp	hɔp	hɔp	kadaoj	<i>never mind ! (expressive)</i>
ɑ:k	cha:k	cha:ʔ	cha:ʔ	<i>blad</i>
	ka:k	ka:ʔ	ka:ʔ	<i>to freeze</i>
	sa:k	sa:ʔ	sa:ʔ	<i>to peel (with hand)</i>
ɑ:r	ʔɑ:r	ʔɑ:r	ʔɑ:	<i>pleased</i>
	ʔa.sa:r	ʔaʔ.sa:r	ʔak#sa:	<i>letter (alphabet)</i>
ɑ:	ga:	ga:	kɔo	<i>dumb</i>
uap	cuap	cuap	cuəp	<i>to meet</i>
	guap	---	pkuop	<i>to save</i>
uak	buak	---	puəʔ	<i>group</i>
	muak	---	muəʔ	<i>hat</i>
uaɲ	klap#kluap	---	laʔa:#laʔac	<i>Echo word - beautiful</i>
	cuaɲ	---	cuəɲ	<i>bell</i>
iac	wiac	keɲ	wiec	<i>curved</i>

Table B: Correspondence problem: presumed transcription mistake in Bon09

Rime	Bon09	Bon10	Hau91	*PSB Sid00	Khmer	Gloss
əh	təh(#piaŋ)	to:h(#piaŋ)	---	---	ḍah#ḍa:j	to serve rice
	kəh	kəəh	(suôt)	---	kəok	dried
	?əh	?əəh	ơh	---	chlaoj	answer
	təm.bəh	təm.bəəh	mbơh	---	prap	to tell, announce
	bəh	bəəh	(lôh)	(*luh)	mək	to come
	ləh	ləəh	lơh		thwə:	to do
əm	gəm	gəəm	gə:m /gə:m/	*gə:m	saoc	to laugh
	khaj#rə.ləm	khaj#rə.ləəm	rləm	---	khæ#wosa:	rainy season
	jəm	ɲəm(#sak)	nhôm	*ɲum	ḍuəŋ	bun
on	kə.ŋon	n.ɲən	---	---	cəpɕu:n	to transport
	han#jon	hən#ɲən	---	---	cu:n	to accompany
	jon	ɲən	---	---	cu:n	to give sth to sb
	gon(#da:k)	gən(#da:k)	---	---	ḍa:ŋ#tək	to draw water

Table C: Long central Vowels

Rime	Bon09	Bon10	Hau91	*PSB Sid00	Khmer	Gloss
ɨr	cɨr	cɨər	crông	*kruŋ	trənuŋ	backbone (fish)
	pə.sɨr	pə.sɨər	---	---	sə:səə ; təsəə	to compliment
	hɨr	hɨər	---	---	haə#poa	to change color
ɨɲ	dɨ:ɲ	?a.dəɲ	---	---	tiel	blunt
ɨ:	chɨ:	chɨ:	jhư	---	chə:	tree, wood
	krə.ḍi:	krə.ḍi:	cơpư	---	krəḍəj	water buffalo, carabao
	prə.ḍi:	prə.ḍi:	bơrbhoe	*rum	pteji	spinach
	rə.sɨ:	rə.sɨ:	‘jar ; rdêy	---	risej	bamboo
ət	kət	kɨət	cơt	---	kaət#kə:n	be born
ɛ:c	pɛ:c, bɛ:c	ḍiəc	---	---	ḍaoc	to remove the hair from
	chɛ:c	Ø	srat ; chot	*sra:t ; *taŋ ~ *btaŋ	cət	sour, bitter
ək	pək	pək	pək	*pək	ḍaək	to open
	bək	(pək)		---	ḍaək	to drive
	dək	dək	ḍưk	---	prohə:n	insolent
əm	n.cə:m	n.ciəm	---	---	cəp#caəm	eyebrow
	da:k#kə.sə:m	da:k#kə.sɨəm	tôc	*brju:h	tək#sənsaəm	dew
əm	sə:m	sɨəm	sôh	*su:h	saəm	wet
ən	blə.hə:n	blə.hɨən	---	---	prohə:n	insolent
	mə:n	mən	mươn	---	mə:n	ten thousand

ə:ŋ	pə:ŋ	pəŋ	---	---	pəəŋ	<i>to glide</i>
<i>Rime</i>	<i>Bon09</i>	<i>Bon10</i>	<i>Hau91</i>	<i>*PSB Sid00</i>	<i>Khmer</i>	<i>Gloss</i>
	(mak#)jə:ŋ	(mak#)jəŋ	---	---	khaŋ#cə:ŋ	<i>Nord</i>
	ʃim.ʔmə:ŋ	ʃim.ʔəŋ	ja ; rsôi	*gəja:	cəmbəəŋ	<i>thatch, straw</i>
ə:l	ḃə:l	ḃʔiəl	---	---	phniəp#ʔaəl	<i>to be surprised</i>
	gə:l	giəl	ndəʔm ləʔn	---	kə:l	<i>bumpy</i>
ə:j	hə:j	hiəj	hufi	---	haoj	<i>already</i>
	mbə:j	mbiəj	mbəʔi, mbufi	*npa:w	jol#səp	<i>dream</i>
ə:	krə.bə:	kə.bəw	cəʔbəʔ	(*ja:)	krəpə:	<i>crocodile</i>
	smə:#gə.na:	sməw#gə.na:			sməə#khniə	<i>to be equal to</i>
	rə:	rəw			rə:	<i>to disassemble</i>

II. Vowel correspondences

Table D: Long vowels

<i>PSB</i> <i>Sidwell (2000:47)</i>	<i>Stieng</i> <i>(Haupers 1991)</i>	<i>Bon</i> <i>(data 2007-09)</i>	<i>Chrau</i> <i>(Thomas & Luc 1966)</i>	<i>Environment</i>
*a: >	a:	a:	a:	
*ə: >	ə:	ə:	ə:	
*i: >	e i	i: i i	i: i: i:	/ __ Ø / __ h / elsewhere
*e: >	e: e: With: /e:/ > [ej] / __ #	ɛ: e:	ɛ: e:	/ __ Ø
*ɔ: >	ɔ:	ɔ:	ɔ:	
*u: >	o u	u: o u	u: u: u:	/ __ Ø / __ h / elsewhere
*uə >	uə	ua	uə	
*iə >	iə	ia	iə	

Table E: *Short vowels*

<i>Sidwell PSB</i> (2000:47)	<i>Stieng</i> (Haupers 1991)	<i>Bon</i> (data 07-09)	<i>Chrau (Thomas & Luc 1966)</i>	<i>Environnement</i>
*a >	a	a	a	
*ə >	ɛ e a ə With: *s > h / e__ /ə/ > [e]/ __c, ɲ [ə]/elsewhere	ɛ e o:ɰ ə With: *s > h / e__ /ə/ > [e]/ __c, ɲ [ə]/elsewhere	ɛ i o ə	/ __glottals / __s / __w / elsewhere
*i >	o e With: /e/ > [i]/Lab.Vel.__ [e]/elsewhere	e e With: /e/ > [e]/__glottal [e]/elsewhere	i i With: /i/ > [e]/__glottal [i]/elsewhere	/ __Labial / elsewhere
*ɔ >	ɔ	ɔ With: ɔ > a	ɔ	/ __C [+ vel/glott]
*u >	u o o	 o ɔ	i i u	/ __C[- nas,-son + dent/pal] / __C[-nas + son, + dent/pal] / elsewhere / __Dent, vel

Table F : Front Vowels - Exceptions

<i>Bon09</i>	<i>Bon10</i>	<i>Hau91</i>	<i>*PSB Sid00</i>	<i>English</i>
cim	ncem	siem, chiem /siəm/	*siəm	<i>to adopt</i>
dim	kə.dəm	diem /diəm/	*diəm ~ *ʔiəm	<i>garlic</i>
(da:k#)mi:	(da:k#)mi:	mi: /mi:/	*mi:wh	<i>rain</i>
nhi:	ni:	ɲi: /ɲi:/	*həɲi:wh	<i>house</i>
gə.nɛ:ŋ	gə.neəŋ	---	*gəniəŋ ~ *gəne:ŋ	<i>tusk, canine tooth</i>

Remark: usually *iə correspond to /ia/ (Bon 09) (including -iaŋ rimes)

Table G: Central Vowels - Exceptions

<i>Bon09</i>	<i>Bon10</i>	<i>Hau91</i>	<i>*PSB Sid00</i>	<i>English</i>
wər	wər	wɔʔr	cf. wər 808 p. 141 'stir'	<i>to crawl ; to go on all fours</i>
gəm	gəəm	gə:m /gə:m/	*gə:m	<i>to laugh</i>
bəh	bəəh	(lôh)	(*luh)	<i>to come</i>
ʔən	ʔəŋ	(geh)	(*geh)	<i>to have</i>
pət	pət	(tang)	(*taŋg)	<i>to close</i>
jəp	ɲəp	nhôp	*ɲup	<i>to shake hands</i>
jəm	ɲəm(#sak)	nhôm	*ɲum	<i>to wear one's hair in a bun</i>
nʔnər	nʔnər	---	(*nir)	<i>catch fish with a shovel-shaped basket</i>
ʔəl		ʔil /ʔel/	*ʔil	<i>deer</i>
ndə:p	ndəəp		*nədəp	<i>to give</i>

Table H: Back Vowels - Exceptions

<i>Bon09</i>	<i>Bon10</i>	<i>Hau91</i>	<i>*PSB Sid00</i>	<i>English</i>
tur	tu:r	to:r /to:r/	*to:r	<i>ear</i>
kos	ko:s	ko:s /ko:s/	*ko:s	<i>to shave</i>
(nak#)dos	(nak#)du:s	do:s /do:s/	*do:s	<i>guilty person</i>
jo:p	ɲəp	ɲop /ɲop/	*ɲup	<i>catch, grab</i>
tə:m#dɔŋ	tə:m#dɔŋ	---	*dɔŋ	<i>coconut palm</i>
pə.ŋət#so:ŋ	pə.ŋət#so:ŋ	ŋət	*pəŋət	<i>to be hungry</i>
tut	tot	tət	*kətət	<i>skin sickness</i>
rəŋ.lo:ŋ#kə:	rəŋ.loŋ#kəw	---	*rənɔŋ	<i>throat</i>

THE DIFFERENTIAL DEVELOPMENT OF PROTO-SOUTHWESTERN TAI *R IN LAO AND THAI³⁴

Garry W. Davis

The University of Wisconsin-Milwaukee

<gdavis@uwm.edu>

0 Abstract

Traditional scholarship assumes that disyllables appeared in Lao only as late loan borrowings, and therefore did not undergo the change of *r- > h- that affected monosyllabic forms around 1500 CE (*rim* ‘edge, shore’ > *him*; *rak* ‘to love’ > *hak*, etc.). The present paper considers a scenario whereby early disyllables containing *r* might have been present at the time of the change, but remained unaffected. It is plausible that *r* originally had two positional variants, a more aspirated and robustly trilled allophone ([r]) that occurred initially in monosyllabic words, and a less aspirated allophone with a short trill or a flap ([ɾ]) that occurred word-medially and word-initially in weakly stressed position. Through two regular sound changes that have analogues in other languages, the more aspirated and robustly trilled allophone of *r* merged with *h* (**rao* > *hao* ‘we’), and the less aspirated flapped allophone of *r* later merged with *l* medially (**sāmṛēt* > *sāmlēt* ‘to complete, to succeed’) and in word-initial position in disyllabic words (**rawaən* > *lawaaən* ‘between’) where the initial syllable was weakly stressed. Standard Thai also inherited PSWT **r*, and preserves it normatively, but colloquial varieties of Thai have meanwhile undergone a parallel but modern and unconditioned, merger of *r* and *l* that was not sensitive to stress patterns or position in a word. Both developments may illustrate different kinds of perceptually based sound changes of the sort identified by Blevins (2004: 32-33).

1 Overview

This paper traces the development of Proto-Southwestern Thai (henceforth PSWT) **r* into modern Lao and Thai, and reconsiders the conventional position that the Lao lexicon still consisted exclusively of monosyllables when it and some other closely related Southwestern Tai dialects underwent the change of **r*- > *h*- (**rṓn* > *hṓn* ‘hot’) sometime between 1400 and 1500 CE (Gagneux 1983).³⁵ Conventional wisdom further holds that the disyllables³⁶ that exist in modern Lao were borrowed from Khmer and other sources

³⁴ The author wishes to thank an anonymous reviewer for many comments that helped to improve this paper. I am also grateful to Thomas J. Hudak for sharing helpful advice on various aspects of Thai linguistics. Any remaining errors are mine alone.

³⁵ As is still the case in the modern language, the phones *-r* and *-l* did not occur in the syllable coda during the period in question, so the change of *r* to *h* operated only in initial position.

³⁶ Monosyllables are monomorphemic words consisting of one syllable. Disyllables are monomorphemic words that are comprised of two syllables (Thai *sāmṓng* ‘brain, head’). Compounds consisting of two monosyllabic words (e.g. Thai *roən rian* ‘building’ + ‘to study, learn’ → ‘school’) are thus not disyllables.

Garry Davis. 2010. The Differential Development of Proto-Southwestern Tai **r* in Lao and Thai.

Journal of the Southeast Asian Linguistics Society 3.2:49-60

Copyright vested in the author

Received 21/7/10, revised text accepted 30/11/10

starting in the 16th century after first passing through Thai, and were therefore not in the language until after the change of *r*-> *h*- had run its course. This putative chronology would explain why disyllables that contained *r* in Khmer have *l*, rather than *h* in Lao (*rawaaj* > *lawaaŋ* ‘between’, not ***hawaaj*; *sămrăp* > *sămlăp* ‘for’, not ***sămhăp*). The Lao forms correspond to cognates in Standard Thai that are normatively pronounced with *r*, but that are also pronounced with *l* in modern spoken Thai varieties. The chronological assumptions implicit in the conventional account thus suggest that the Lao and Thai mergers of *r* and *l* are a common development, or are at least virtually identical parallel developments, a view that will be questioned here.

The Ayutthaya period (starting with the founding of the city of the same name in 1351) was a time when a high level of Khmer-Thai bilingualism existed in and around the Thai capital, and this undoubtedly facilitated the borrowing of large numbers of Khmer words into Thai to the point of greatly expanding the Thai lexicon (Khanittanan 2004:375; Suthiwan & Tadmor 2009:604). These borrowings enriched Thai literature and poetry immensely, and it is here that disyllabic Khmer loan borrowings are first attested in abundance. But there are mounting indications that Khmer linguistic influence on Thai may have been ongoing since the arrival of the Tai peoples in what is today Thailand, when the area was still controlled by the Khmer.³⁷ It is possible that these earlier Khmer influences did not always find their way into written sources until the Ayutthaya period, when writing was suddenly more commonplace and diverse in form. The lack of early attestations of disyllables would be particularly understandable in northern Thailand and in Laos where the early written sources are mostly inscriptions (Lorrillard 2009) rather than literary and poetic writings.

The Thais of the Chao Phraya River basin probably had the most intense and sustained interaction of any of the Tai peoples with the Khmer, and the type of borrowings from Khmer attested during the Ayutthaya period makes it likely that such borrowing had been ongoing for generations – perhaps long enough for the borrowed forms to have spread to closely related Tai languages on the periphery like Lanna Thai (also known as Khammueng or Yuan) and Lao.³⁸ The intensity of the linguistic contact is reflected in the fact that Thai borrowings from Khmer are not limited to open class lexical items, but also include function words (Khanittanan 2004:388) such as *rǎu* ‘or’, *khǎu* ‘to be’, *phrɔ́* ‘because’ and *dooi* ‘through’ that had a profound syntactic influence on how Thai and Lao are spoken. Thai appears also to have borrowed infixation strategies from Khmer that began to operate in native Thai words (*chûay* ‘help’ – *chamrûay* ‘gift given to help’; Varasarin (1984:65), Khanittanan, 387). Varasarin (257) even found evidence that Thai eventually coined its own disyllabic words alongside borrowed Khmer forms (*pràtuu* < *tuu* ‘door’, *tàkrâa* ‘basket’). These early Thai innovations can be recognized because they have no discernable Khmer counterparts, and they are often listed as being etymologically opaque (Suthiwan & Tadmor 2009:616).

The integral nature of these early Khmer loan borrowings was recognized already by Gedney (1947) in his landmark study of the Indic element in the Thai lexicon. Gedney called these early loan borrowings normal or ‘blue’ words because of the extent of their integration into the structure of the Thai language. Most Indic (Pali and Sanskrit)

³⁷ Varasarin (1984:26) notes that in the 14th century a king of Sukhothai had a Khmer inscription prepared, presumably for the benefit of remnant Khmer populations there.

³⁸ Ferlus (1985) estimates that the ‘Khmerization’ of Thai began around 1250.

borrowings in Thai, on the other hand, were regarded by Gedney as learned or ‘red’ borrowings. Later studies confirm that most Indic loan words were acquired via literary borrowing rather than by direct contact with speakers of these languages despite the fact that the Indic lexical component is much larger overall than the Khmer component of Thai (Suthiwan & Tadmor, 604).³⁹ These later Indic borrowings also account for almost all words that contain three or more syllables in both Thai and Lao. It seems plausible to assume that some blue loan words, of which many are Khmer disyllables, could have entered Thai at an early time and reached Lao and northern Thai varieties by 1400 CE. Furthermore, as Varasarin (359 ff.) has shown, a number of syllable-initial clusters in Khmer did not exist in Thai, and thus required the insertion of an epenthetic vowel to make the onsets phonotactically acceptable.⁴⁰ This process had the effect of converting Khmer monosyllables (*spi:ən* ‘bridge’, *sba:əy* ‘to be relaxed’) into Thai and Lao disyllables (Thai *sàpaan*, Lao *sapáan* ‘bridge’; Thai *sàbaai*, Lao *sabàai* ‘comfortable’), which could, in turn, have facilitated the increased borrowing of Khmer disyllables. It is also possible that the borrowing process would have continued for some time, and that later loans from Khmer and Indic loans acquired through literary influence might have arrived in Lao from Thai sources at a later time, meaning we must be able to account for how *r emerged in both early and late borrowings.

The conventional assumption that all disyllables in Thai are exclusively 16th century borrowings from Khmer, and that Lao disyllables must therefore be even later borrowings would become untenable should the (ever more accessible) philological record of Lao, or Lanna Thai (Penth et al. 1997-2006) someday yield clear evidence of early disyllables that contained *r*. If it should turn out that not all Lao disyllables were late borrowings, then we will need an explanation of why modern Lao disyllables have *l* in initial position rather than the expected *h*, given the change of *r* > *h*-. What follows is a hypothetical account of how *r* in early disyllables in Lao and Lanna Thai could have escaped the change of *r* > *h*- to emerge later as *l* in these varieties. The posited changes amount to a claim that *r* in Lao developed along different lines than in Thai and that the circumstances and chronology of *r*’s merger with *l* may differ in the two languages.

2 Background

Lao and Thai are closely related languages that both descend from Proto-Tai via the intermediate subgrouping of Proto-Southwestern Tai (henceforth PSWT). The languages share a high degree of mutual intelligibility, and structural similarity, and use closely related writing systems that ultimately derive from earlier forms of the Indic script (Diller 1996). The numerous linguistic similarities that Lao and Thai share serve to make their points of divergence all the more instructive, including the differential development of PSWT *r in the two languages. Proto-Tai *r (and by extension PSWT *r) is thought to have been a ‘tongue-tip vibrant or trill, which probably required strong breath to achieve’ (Li 1977:142), perhaps not unlike the trilled or flapped *r* pronunciations that are

³⁹ Some Indic borrowing came by way of the Khmer, however, who were themselves already thoroughly Indianized. Khmer also passed along to the Thais a number of Indic loan borrowings, Buddhism, and an Indian-derived script. A handful of later borrowings did arrive via contact with Indian traders (Thai *sàbùu* ‘soap’, *bùrìi* ‘cigarette’) (Suthiwan & Tadmor, 602).

⁴⁰ Varasarin indicates that the Khmer initial clusters *sk*, *st*, *sd*, *sp*, *sb*, *sq*, *sj*, *sn*, *sm*, *sv* and others triggered insertion of an epenthetic vowel in Thai (359 ff.).

normatively prescribed for Standard Thai (also known as Central Siamese). Thus, Standard (written) Thai preserves the original distinction between *r* and *l* intact, while *r* and *l* in spoken Thai have actually been undergoing an unconditioned merger for decades (Gedney 1947, 1966). For many Thai speakers, the use of *r* in spoken language varies according to socioeconomic stratification and by register (Diller 2002, Phootirat 2004).

In the following sections, we explore the possibility that the reflexes of PSWT **r* developed differently in Lao and in Thai, because speakers of these languages (mis)perceived, and then reinterpreted the phonetic variations of the **r* sound they heard in different ways, leading to different phonological restructurings of **r* in the two languages. Indeed, the general historical change of **r* > *h*- at an earlier stage of the Lao language (Diller 1996:459) is acknowledged by most scholars, a change that is also common in various other Tai languages spoken in northern and northeastern Thailand, Burma, Yunnan and elsewhere (Gedney 1966:22). Reference works and teaching grammars often hedge on this point, however, and many simply state that Standard Thai *r* (which continues PSWT **r*) corresponds either to *h* or to *l* in Lao (Enfield 1999:272). Dictionaries such as Kerr (1972) often avoid the issue by noting that the letters <ຮ, r>⁴¹, <ຣ, l>, and <ຮ, h> are ‘interchangeable’ in many Lao words.

In Section 3, we examine monosyllabic and disyllabic words in modern Lao that contain reflexes of **r* in word-initial and medial position. We conclude that the development of **r* into *h* and *l* was a split with merger that occurred in two chronologically ordered stages. The merger of **r* with *h* in some environments, and with *l* elsewhere, correlates in part with differences in lexical stress patterns, because **r* in weakly-stressed initial position became *l* in Lao (rather than *h*).

In Section 4, we demonstrate that developments somewhat analogous to these have occurred in a number of other languages in Southeast Asia and beyond, and we explore possible phonetic motivations for the posited changes. We conclude that Blevins’ (2004) model of listener-centered sound changes (part of what is referred to as ‘Evolutionary Phonology’) is best able to explain how non-distinctive phonetic variation (such as non-phonemic lexical stress in Lao) can help to trigger phonological restructuring, and how subtle differences in such listener-centered reanalysis can have surprisingly differential outcomes, as in the case of **r* in Lao and Thai.

3 Distribution of the Reflexes of PSWT **r* in Modern Lao and Lexical Stress

Lao and Thai are tonal languages that also have a system of non-distinctive lexical stress. Extensive research has established that the stress in a polysyllabic Thai word falls on its final syllable (Peyasantiwong 1986:214-215, Luksaneeyanawin 1998:376). As Peyasantiwong (215) notes, the second syllable of a disyllabic Thai word carries ‘normal’ stress, while the preceding syllable has ‘weak’ stress (*sǎabaan* ‘swear’, *kanwon* ‘worry’)⁴². In polysyllabic words of three or more syllables, stress always occurs on the final syllable as in *prachaachon* ‘populace’, *thaleesaap* ‘lake’, and *thanaakhaan* ‘bank’ (Peyasantiwong, 224). In weakly stressed initial syllables, the length of the vowel is usually shortened with long vowels becoming short, and diphthongs and short vowels emerging somewhat phonetically shorter than their usual length: *sǎabaan* > *sǎbaan*. Although the Lao stress

⁴¹ For simplicity’s sake and to save space, phonemes will be written without brackets. Graphemes cited as examples appear between wedge-shaped brackets (e.g. <r>, Lao <ຮ>).

⁴² Underscoring marks greater stress. Superimposed accent marks indicate tone.

system has been less thoroughly studied, the basic similarity of the Lao stress patterns is confirmed by Enfield (2007)⁴³, and we thus assume in what follows that these stress patterns have been constant in both languages since the inception of the disyllables.

3.1 Changes in Word initial Position in Monosyllabic Words

The correspondence of word-initial PSWT *r to word-initial *h* in monosyllabic Lao words is striking and very regular. The items in the following table are thus merely illustrative, and further examples are legion. Since Standard (written) Thai preserves reflexes of PSWT *r, we cite forms from Standard Thai whenever available to illustrate the original state of affairs. Note that all the cited forms in Table 1 are monosyllabic words, some of which are Khmer loan borrowings. (Varasarin's etymological discussion of the Khmer forms is cited in parentheses following the gloss.)

Table 1: Monosyllabic Lao Words with Word-initial *r- >h-

Standard Thai	Lao	Gloss
<i>rim</i>	<i>hím</i>	edge, shore (329)
<i>ríp</i>	<i>híp</i>	to hurry
<i>rian</i>	<i>hían</i>	to learn (328)
<i>rɛɛŋ</i>	<i>hɛɛŋ</i>	to start
<i>râi</i>	<i>hai</i>	field
<i>rao</i>	<i>háo</i>	we
<i>rak</i>	<i>hak</i>	to love
<i>raan</i>	<i>hâan</i>	store, shop
<i>ríap</i>	<i>híap</i>	even, in order (328)
<i>rɯa</i>	<i>húa</i>	boat
<i>rúam</i>	<i>huam</i>	to combine, join (329)
<i>rúu</i>	<i>húu</i>	to know

Despite the above examples of monosyllabic Khmer borrowings that underwent the change of *r*- > *h*-, monosyllabic Khmer borrowings can also be found that did not undergo the change, because they did not arrive in Lao until after this change was complete. These include forms such as *lám* 'to dance' (cf. Thai *ram*) and (*ŋɛ́n*)*lǎan* 'coin' (Thai *rǎan*). There is even the interesting case of the Khmer form *rɔəp* > Thai *ráp*, Lao *lap* 'to receive' (Varasarin 1984:330). It occurs in Lao both as *hap* (via the change of *r*- > *h*-) as well as in the doublet *lap* that presumably resulted due to later re-borrowing. The arrival of these later monosyllabic borrowings with initial *r*- filled the structural gap that had opened when initial *r*- became *h*- in Lao monosyllables.⁴⁴

While only some Khmer (blue) monosyllables arrived in Lao too late to participate in the change, Indic (red or literary) borrowings almost always arrived too late to do so, as in Lao *lat* 'state' (Thai *rát*) and *lôok* (*labàat*) 'epidemic' (Thai *rôok*). These facts make clear that borrowing continued over a span of centuries, and that any account of the borrowing process must be able to explain these differential outcomes.

⁴³ Enfield (2007:139) states that, "Non-monosyllabic Lao words and phrases are stress final, resulting in pre-final elements being reduced."

⁴⁴ During the time when there was a structural gap of *r* initially in the monosyllables, *r* would still have occurred in stop + *r* clusters and in the disyllables.

3.2 Changes in Word-Initial Position in Disyllabic Words

A number of Lao words all of which are of Khmer origin have **rv(v)-* > *lv(v)-* in word-initial position. The examples follow a distinct pattern in that they are all disyllabic words that consist of a single morpheme where the initial syllable of the word carries only weak stress in accordance with the pattern of lexical stress discussed above. The following forms are merely illustrative; further examples abound.

Table 2: Disyllabic Lao words with Word-initial **r* > *l*
(The Khmer origin of these forms is confirmed by Varasarin, 331-332.)

Standard Thai	Lao	Gloss
<i>ràbàat</i>	<i>labàat</i>	to spread
<i>ràbiap</i>	<i>labiap</i>	in order
<i>ràbèət</i>	<i>labèət</i>	to explode
<i>rádòp</i>	<i>labǒp</i>	system
<i>rádàp</i>	<i>ladǎp</i>	level, stage
<i>rádom</i>	<i>ladòm</i>	to persuade
<i>rákəŋ</i>	<i>lakəŋ</i>	bell
<i>ránâat</i>	<i>lanâat</i>	xylophone
<i>rálúk-dâai</i>	<i>lanúk</i>	to recollect
<i>ràwàəŋ</i>	<i>lawaaŋ</i>	between
<i>ráwəŋ</i>	<i>lawán</i>	danger, watch out
<i>ráwəeŋ</i>	<i>lawéəŋ</i>	to suspect, doubt
<i>raəŋwan</i>	<i>láəŋwán</i>	reward, prize
<i>ráyá</i>	<i>láyá</i>	distance

Initial *h* and *l* (< *r*) alternate in at least one pair of related Lao forms that make use of infixing. Where **r-* stood in initial position in a monosyllable, it surfaces as *h* in *hîap* ‘in a line, tidy’ (< **rîap*). In its related disyllabic form, however, in which the initial syllable is weakly stressed, we get *labiab* ‘order’ < **rabiab* (Thai *rabiap*), which contains the now unproductive loan infix *-ab-*. In the related pair of words *lalúk* ‘to be mindful of’ and *lamúk* ‘to reminisce’, we get the expected initial *l-* (< **r-*) in both words, because both forms are disyllabic with a weakly stressed initial syllable.

3.3 Changes in Medial Position in Disyllabic Words

As stated above, neither *r* nor *l* occurs in syllable-final position in Lao, so all occurrences of *r* and *l* in medial position are syllable-initial. In this position, PSWT **r* has been replaced exclusively by *l* in disyllabic Lao words. An illustrative list of some common disyllabic words with medial **r* > *l* appears in Table 3 below. For their Khmer etymologies, see Varasarin (153, 340, 363, 315).

Table 3: *Disyllabic Lao Words with Medial *r > l*

Standard Thai	Lao	Gloss
<i>gamrai</i>	<i>gàmlái</i>	profit
<i>samrap</i>	<i>sămlăp</i>	for
<i>samret</i>	<i>sămlět</i>	succeed
<i>tamrùat</i>	<i>tàmlùat</i>	policeman

3.4 Discussion

Our examination of the Lao lexicon reveals the following generalizations regarding the development of PSWT *r:

- A. In word-initial position in monosyllabic words, *r surfaces as *h* (where it was in word-initial position in a stressed syllable.)
- and,
- B. Elsewhere: *r > *l* (where *r was in word-initial position in a weakly stressed syllable, or in medial position regardless of the stress placement in the word).

4 Motivations for the Posited Changes, Analogs in Other Languages, and Why the Lao and Thai Developments Differ

The change of *r*- > *h*- is a common one in the languages of Southeast Asia. It is documented by Hayes (1982) for Thavung, a Vietic language of Vietnam, where *r* was replaced by *h* word-initially in monosyllabic words as well as syllable-initially in disyllabic ones. More recently, a partially similar change of *r* > *h* has also taken place with surprising speed in the Khmer dialect of the city of Phnom Penh in Cambodia. In Phnom Penh dialect, *r* becomes *h* in initial-position and also when it is the second segment of a cluster (*Cr*- > *Ch*-). Wayland and Guion (2005) argue that they have found a trigger for the change of *Cr*- > *Ch*- in these Cambodian varieties. They begin with the assumption that the voiced alveolar trilled *r* in the Cambodian dialects that undergo this change would have required a high volume of airflow, and that *r* might have become voiceless during fast, colloquial speech in initial position and in clusters. In *Cr*- clusters, the voiceless rhotic could have been reinterpreted as mere aspiration, making *Cr*- clusters into *Ch*- consonant clusters that, Wayland and Guion (78) contend, could trigger falling-rising tone on the following vowel.

But the change of *r*- > *h*- is not restricted to Southeast Asia. Blust (1983) notes in passing a general change of initial *r*- > *h*- in some Portuguese dialects of eastern Brazil, and in various regional dialects of Spain *r has been known to become a variety of other sounds including [h], an alveolar fricative [ɹ], or various types of [l] (Lloyd 1987:348). The most pertinent of the Spanish developments have been documented in Puerto Rican dialect (Navarro 1948). On the island of Puerto Rico, the tense and often voiceless word-initial [r:] of Spanish often becomes [x] in word-initial position, and [l] word-medially. These erstwhile allophones of original *r then merge with the preexisting Spanish phonemes *x* and *l*. (As Hock (1991:124) points out, natives thus often pronounce the name of the island as [pwel.to.xi.ko].⁴⁵

⁴⁵ The periods here mark syllable boundaries. The change of *r > *l* in Puerto Rican dialect occurs only in coda position, however, whereas the Lao developments take place only syllable-initially.

Returning to the Lao developments, we begin with the conventional assumption that initial **r*- at an early stage was strongly trilled. The length of the trill and the robustness of its aspiration were probably greatest in stressed monosyllabic words, rendering the positional variant of **r* in that environment more prominent, longer in duration, and perhaps greater in amplitude. In a fashion similar to the recent developments that Wayland and Guion posit for Cambodian, we contend that with strong aspiration **r* may have tended to lose some of its voicing, and, with a reduction in trilling, the resulting highly aspirated and voiceless phone may have been hard to distinguish from [h], ultimately triggering a phonemic merger of *r* in that environment (**róoŋ* > Lao *hóoŋ* ‘building’) with preexisting *h* (Lao *hǒk* ‘six’).

In weakly stressed word-initial position, we contend that *r* may have been pronounced as a short trill or as a flap and with less tendency to shorten the duration of its voicing (**rabiap* ‘in order’). In medial position – whether the syllable it occurred in was stressed or not -- *r* would have likewise been a short trill or a flap allophone (since the duration of trilled or flapped *r* in medial-position tends cross-linguistically to be phonetically relatively shorter), and the medial environment would have allowed the segment to remain voiced throughout its duration (**sǎmrět* ‘to complete, accomplish’). In medial and in weakly stressed environments, then, we expect that **r* retained a short trill or flapped articulation for some time, but that this pronunciation gradually gave way to a flap that later generations of speakers would reanalyze simply as *l*, since a flap [ɾ] in medial position is sometimes difficult to distinguish from *l*.

The developments we posit for Lao thus seem most analogous to those in Puerto Rican Spanish, because both cases involve a split with merger of **r* with two other phonemes, one a voiceless fricative and the other a voiced liquid. We speculate that the Puerto Rican and Lao developments are both examples of the kind of listener-centered sound changes that Blevins (2004:32-33) addresses within the framework of Evolutionary Phonology. According to her theory, extant sound patterns are rooted in the dynamic of sound change, which is itself determined more by the vicissitudes of perception than production. Sound change, on this model, comes about as listeners either misperceive or reinterpret the phonological structure of speakers’ utterances. Blevins’ framework recognizes a general typology of sound change that consists of three types of reanalysis she calls CHANGE, CHANCE, and CHOICE⁴⁶:

⁴⁶ The terms CHANGE, CHOICE, and CHANCE are written with small capitals to distinguish these technical terms from their common usages.

(S=speaker, L=listener)

CHANGE: The phonetic signal is misheard by the listener due to perceptual similarities of the actual utterance with the perceived utterance.

Example: S says [anpa]

L hears [ampa]

CHANCE: The phonetic signal is accurately perceived by the listener but is intrinsically phonologically ambiguous, and the listener associates a phonological form with the utterance which differs from the phonological form in the speaker's grammar.

Example: S says [ʔaʔ] for /aʔ/

L hears [[ʔaʔ]] and assumes [ʔa]

CHOICE: Multiple phonetic signals representing variants of a single phonological form are accurately perceived by the listener, and, due to this variation, the listener (a) acquires a prototype or best exemplar of a phonetic category which differs from that of the speaker, and/or (b) associates a phonological form with the set of variants which differs from the phonological form in the speaker's grammar.

Example: S says [kakáta], [kǎkáta], [kkáta] for /kakata/

L hears [kkáta], [kǎkáta], [kakáta] and assumes /kkata/

(Blevins 2004: 32-33)

An advantage of Blevins' framework is that it places phonetic variation – and the perception and misperception of this variation by listeners – at the center of why sound change happens. It thus provides a tool to understand the differential development of the reflexes of PSWT *r in Lao and in Thai despite the presumably identical features *r would have had in the early stages of these two closely related and largely mutually intelligible languages. The model also allows us to integrate the phonetic influence of (non-distinctive) features – such as patterns of lexical stress, and the effects of being in medial position – on the perception and misperception of the segments listeners hear.

Accordingly, the development of *r in Lao appears to be an example of what Blevins calls CHOICE. That is, Lao listeners perceived very well the multiple phonetic signals accompanying the variants of *r. In stressed initial-position, the high volume of airflow necessary for the trilled *r may have caused it to become voiceless in fast speech, and listeners apparently associated its phonological form with a set of phonetic variants that differed from the phonological form of the speaker's grammar. That is, listeners associated a partially voiceless and aspirated *r with the already extant phonological segment *h*. In environments where *r tended to be voiced throughout its duration with only a short trill or a flap, the same kind of re-association occurred with listeners eventually associating the features they heard with the already extant (and phonetically very close) phoneme *l*. In this way, lexical stress patterns influenced the phonetic variants of *r that Lao listeners heard, and helped to bring about a reanalysis of *r as either *h* or *l* depending on the phonetic variants that occurred in a particular environment, even where those features were not phonemically distinctive.

While Standard (written) Thai preserves (normatively) the original state of affairs with a phonemic distinction of *l* and *r* (with a trilled or a flapped alveolar articulation of *r*

being acceptable allophones of that phoneme), most varieties of spoken Thai have long since undergone an unconditioned merger of the two phones (to *l*). This merger – in contrast to the Lao developments – reflects a perceptual reanalysis like that described by Blevins under the rubric of CHANGE. The phoneme **r* came to be misheard as *l* by listeners because *r* often occurred in an environment where its phonetics were more ‘*l*-like’, that is, it likely had a short trill or a flap [ɾ] and remained voiced (especially in medial position and in weakly stressed position) throughout most of its duration.

5 Conclusions

An analysis of the modern Lao lexicon indicates that the split of PSWT **r* with merger with **h* and **l* was a regular change with the merger of **r* > *h* predating significantly the merger of **r* > *l*. The early date of the merger of **r* > *h* is made likely by its apparent contemporariness with similar mergers in nearby Lanna Thai, and by the early advent of a new graphemic character (Lao <ꨀ> with the traditional mnemonic *hɔɔ huan*) to write the *h* sound that derived from **r* via this merger (Gagneux 1983). The later merger of *r* and *l* that is observed in the disyllables and in later (mostly literary) borrowings is more difficult to date. Early Lao writings use the letter <ɾ> (<ꨀ> with the traditional mnemonic *rɔɔ rha*) consistently in forms that would have contained *r* historically (as opposed to *l*), but this may be due to conservative orthographic traditions. The work of two French lexicographers, Cuaz (1904) and Gagnard (1912), however, provides nearly incontrovertible proof that the merger of **r* and *l* was complete in Lao by 1900 at the latest.

In Thai, however, the reports of foreign visitors make clear that the phonemic contrast between *r* and *l* was still intact in Ayutthaya as late as 1685, and the process of merger of these phones seems to have begun just prior to 1800 (Harris 1997:10). Unlike the developments in Lao, the Thai merger of *r* > *l* was unconditioned. Furthermore, the implementation of the Thai change depends entirely on the register of the language being spoken; colloquially *r* is replaced everywhere by *l*, but normatively *r* is used in high registers of the language (and always preserved in written language) (Diller 1996, 2002:94). The complex sociolinguistic variation in the use of *r* and *l* in Thai may be due in part to the somewhat later chronological emergence of the change, which apparently coincided with the advent of increasing literacy and the political agenda of the late 19th and early 20th centuries to establish Standard Thai as a national language.

The development of **r* in Lao and Thai thus differs significantly. We tentatively conclude that these changes represent two different kinds of perceptually based reanalysis along the lines of those suggested by Blevins (2004). Blevins’ model shows how phonetic variants of sounds can influence phonological reanalysis and thereby trigger differential outcomes (Lao **r* > *h/l*; Thai **r* > *l*).

The view presented here is necessarily tentative until the existence of early disyllables in Lao and Lanna Thai is confirmed by careful philological research. The primary advantage of our view over the conventional account is that we can accommodate both early and late loan borrowings into Lao by showing how early disyllables may have retained an *r* pronunciation even as *r* become *h* in monosyllables. The surviving occurrences of *r* then at some point merged with *l* so that both early disyllabic borrowings and later monosyllabic ones end up with *l* (*lawaaŋ* ‘between’, (*เงิน*)*lian* ‘coin’ < *r*). The most recent loan borrowings include neologisms that were introduced into Lao during and after the 19th century (mostly via Thai) as calques of foreign technical terms (*thoolasap*

‘telephone’, *thóolalêek* ‘telegram’, etc.). The neologisms were obviously inspired by Sanskrit and Pali root words that contained *r, but their appearance with *l* in Lao (given the fact that *r* and *l* in Lao had already merged) does not represent a sound change as much as a case of Lao speakers equating the *r* in these words with what was by then the only liquid in their phonological system -*l*.

References

- Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Blust, R. 1983. A Note on Hypercorrection in Mongondow. *Bijdragen tot de Taal-, Land- en Volkenkunde* 139 (4): 459-464.
- Cuaz, Marie Joseph. 1904. *Lexique Français-Laoçien*. Hongkong: Société des Missions Etrangères.
- Diller, Anthony. 1996. Thai and Lao Writing. *The World's Writing Systems*, ed. by Peter T. Daniels and William Bright, 457-466. New York: Oxford University Press.
- Diller, Anthony. 2002. What makes Central Thai a National Language? *National Identity and its Defenders: Thailand Today*, ed by Craig J. Reynolds, 71-107. Chiang Mai:Silkworm.
- Enfield, Nick J. 1999. Lao as a national language. *Laos: Culture and Society*, ed. by Grant Evans, 258-291. Chiang Mai: Silkworm Press.
- Enfield, Nick J. 2007. *A Grammar of Lao*. Berlin: Mouton de Gruyter.
- Ferlus, Michel. 1985. Les Emprunts Mòns en Thai et Lao. *Southeast Asian Linguistics Studies presented to André-G. Haudricourt*, ed. by S. Ratanakul, 217-233. Bangkok: Institute for Language and Culture for Rural Development, Mahidol University.
- Gagneux, Pierre-Marie. 1983. Les écritures Lao et leur evolution du XVe au XIX siecles. *Asie du Sud-Est et monde Insulindien* 14 (1-2): 75-96.
- Gedney, William J. 1947. *Indic Loanwords in Spoken Thai*. Ph.D. Dissertation: Yale University.
- Gedney, William J. 1966. *Linguistic Analysis of Tai Languages and Dialects*. Unpublished MS. Ann Arbor: The University of Michigan.
- Guignard, Theodore. 1912. *Dictionnaire Laotien-Français*. Hongkong: Société des Missions Etrangères de Paris. [Reprint 1971, Gregg International Publishers Limited]
- Harris, Jimmy G. 1992. The Consonant Sounds of 17th Century Siamese. *Mon Khmer Studies* 21: 1-17.
- Hayes, La Vaughn H. 1982. Mutation of *r in Pre-Tavung. *Mon Khmer Studies* 11: 83-100.
- Hock, Hans Henrich. 1991. *Principles of Historical Linguistics*. 2nd edition. (Trends in Linguistics: Studies and Monographs, 34.) Berlin: Mouton de Gruyter.
- Kerr, Allen D. ed. 1972. *Lao-English Dictionary*. Washington D.C.:The Catholic University of America Press.
- Khanittanan, Wilaiwan. 2004. Khmero-Thai: The Great Change in the History of the Thai Language of the Chao Phraya Basin. *SEALS 2001*, ed. by Somsong Burusphat, 375-91. Tempe, AZ: Program for Southeast Asian Studies, ASU.

- Li, Fang-kuei. 1977. *Handbook of Comparative Thai*. (Oceanic Linguistics Special Publication, 15.) Manoa: University Press of Hawaii.
- Lloyd, Paul M. 1987. *From Latin to Spanish*. (Memoirs of the American Philosophical Society, 173.) Philadelphia: American Philosophical Society.
- Lorrillard, Michel. 2009. *Scripts and History: the Case of Laos*. Senri Ethnological Studies 74: 33-49.
- Luksaneeyanwin, Sudaporn. 1998. Intonation in Thai. *Intonation Systems A Survey of Twenty Languages*, ed. by Daniel Hirst and Albert Di Cristo, 376-394. Cambridge: Cambridge University Press.
- Navarro, Tomás. 1948. *El Español en Puerto Rico: Contribución a la Geografía Lingüística Hispanoamericana*. Rio Piedras (Puerto Rico): Universidad de Puerto Rico.
- Penth, Hans, Phanphen Khriathai, and Silao Ketphrom (eds.) 1997-2006. *Corpus of Lān Nā Inscriptions* Vol. 1- 13. Archive of Lān Nā Inscriptions Social Research Institute Chiang Mai University: Chiang Mai.
- Peyasantiwong, Patcharin. 1986. Stress in Thai. *Papers from a Conference on Thai Studies in Honor of William J. Gedney*, ed. by Robert Bickner, Thomas Hudak, and Patcharin Peyasantiwong, 211- 230. Ann Arbor: University of Michigan Press.
- Phootirat, Parichart. 2004. *Relationship between Social Class and Variation of Initial (r) in Thai*. Unpublished MS. Sydney: University of Sydney.
- Suthiwan, Titima and Uri Tadmor. 2009. Loanwords in Thai. *Loanwords in the World's Languages: A Comparative Handbook*, ed. by Martin Haspelmath and Uri Tadmor, 599-616. Berlin: De Gruyter Mouton.
- Varasarin, Uraisi. 1984. *Les éléments Khmers dans la formation de la langue Siamoise*. Paris: SELAF.
- Wayland, Ratre P. and Susan P. Guion. 2005. Sound changes following the loss of /r/ in Khmer: a new tonogenetic mechanism? *Mon Khmer Studies* 35:55-82.

THE AUSTROASIATIC VOCABULARY FOR RICE: ITS ORIGIN AND EXPANSION⁴⁷

Michel Ferlus

Independent Researcher (CNRS retired)

<jrmferlus@orange.fr>

1. Introduction

What is noteworthy for rice vocabulary in Far East Asia is its rich paradigm of names according to the preparation phase: “rice (general)/rice on stem”, “rice seedlings”, “unhusked rice/paddy”, “husked rice” and “cooked rice”. We have retained here “paddy” and “husked rice” to represent the two main significant states of the sequence: “raw rice (general, not transformed)” and “elaborated rice (husked, ready to be cooked)”. Austroasiatic languages (henceforth AA), especially Mon-Khmer (henceforth MK) on which this paper focuses, have several sequences of words to so name rice.

The earliest evidences for the use of rice, from about ten thousand years ago, were unearthed in the Yangtse basin region of China (for recent information on these matters, Fuller 2008 and Higham 2002 should be consulted). Before being domesticated, wild rice was gathered and consumed by hunter-gatherer populations, and this gathering phase may have spanned several millennia. Each linguistic area would have had its own vocabulary to name the plant. When rice farming emerged, and then propagated, gatherers gradually became farmers, and they could either borrow the name for rice from the donor population, or keep the word from their own language. This latter case could explain the diversity of names for rice. As a matter of fact, there probably was a combination of both. Improvement in the phases of rice preparation, crushing, husking, cooking in water (gruel or saturated), must have played a role in the specialization and in the diversification of the terminology. Moreover, the ongoing shrinking of the Mon-Khmer language area under the pressure of Chinese, Tibeto-Burman, and finally Thai-Kadai languages led to the extinction of many languages and to a concentration of the surviving ones, and consequently of the names for rice, in a restricted area.

When the culture of rice propagated into areas where the plant did not exist in its wild state, the name for rice in the donor language spread together with the cultivation of the plant. We will see how in an example of this case, one single root spread to the furthest limits of Eurasia.

In line with these considerations, we can easily understand why it is not possible to reconstruct a single sequence for rice names at the Proto Mon-Khmer level.

⁴⁷ I cordially thank Frédéric Pain (Catholic University in Leuven, Belgium), who read over the text with the utmost attention, and Helena Veiga de Oliveira (European Commission, Brussels), who kindly streamlined the English translation.

Michel Ferlus. 2010. The Austroasiatic Vocabulary for Rice: its origin and expansion.

Journal of the Southeast Asian Linguistics Society 3.2:61-76

Copyright vested in the author

Received 19/8/10, revised text accepted 2/12/10

2. The names for rice in Mon-Khmer

We have inventoried the sequence “unhusked rice”, “husked rice” and “cooked rice” in approximately eighty languages and dialects in fifteen groups of the AA family. Of these three terms, only “unhusked rice” and “husked rice”, i.e., basically “raw rice” and “elaborated rice”, were taken into account because we consider these terms to be the most interesting for a scientific study of the history of rice, its domestication and its transformation for consumption. The term “cooked rice” belongs to a different semantic field, and effectively does not interact with the two other terms.

Seven main roots for “paddy/raw rice” and two for “husked rice/elaborated” rice can be highlighted, and their combinations form over a dozen word pairs.

2.1 Main pairs “paddy – husked rice”

Examples of pairs in the most representative languages for our demonstration have been selected, leaving aside some minor languages on purpose.

Note: Latinizations, transliterations and some old unacademic transcriptions are quoted in *italic*. Authors’ transcriptions were as far as possible re-transcribed in phonetics.

	“paddy, raw rice”	“husked rice, elaborated rice”	
(1) Khasi	<i>kba</i> [kba:]	<i>khaw</i> [k ^h aw]	(Bars 1973)
Danaw (Palaungic)	ba:	ko:	(Luce 1965)
Kenieng (Khmuic)	pa:	rko:	(F.)
Semelai (Aslian)	baba:	<i>bras</i> (< AN)	(Kruspe 2004)
Semnam (Aslian)	ba:ʔ	bijə:n	(Burenhult 2009)
PROTO FORM	*k.6a:	*r.koʔ	(Diffloth 2005)
(2) Mnong gar	ba:	p ^h ɛ:j	(Guilleminet 1959)
Stieng	ba:	p ^h ɛ:j	(Hauvers 1991)
PROTO SOUTH BAHNARIC	*6a:	*p ^h ɛ:	(Sidwell 2000)
(3) Da-ang (Palaungic)	^h ŋɔ:	rko:	(F.)
Riang (Palaungic)	ŋoʔ ⁻	koʔ ⁻	(Luce 1965)
Samtao (Angkuic)	^h ŋoʔ	ŋkuʔ	(F.)
*Waic	hŋɔʔ	rŋkoʔ	(Diffloth 1980)
Khmu	^h ŋɔʔ	rŋkoʔ	(F.)
Ksing mul (Khmuic)	ŋɔ: kaw	həko:	(F.)
Hat / Ödu (Khmuic)	ŋa:w	ŋkəw	(F.)
PROTO FORM	*s.ŋɔʔ	*r.koʔ	(F.)

The reconstruction of the presyllable in *s.ŋɔʔ is supported by two languages of the Angkuic branch, Angkú *S’ngaw* [s.ŋaw] (Scott & Hardiman 1900) and Pesin si ŋa [s.ŋa] (Yán & Zhōu 1995).

(4) Laven	cɛh	p ^h ɛ:	(F.)
Lave	cɛh	p ^h ɛ:	(F.)
PROTO WEST BAHNARIC	*cɛh	*p ^h ɛ:	(Sidwell 2003)

The West Bahnaric languages are quite uniform.

(5)	Vietnamese	<i>luá</i> [lua ³]	<i>gəo</i> [ɣa:w ³]	
	Mường	lɔ: ³	kaw ³	(F., T.Tr. Dõi)
	Rục	alɔ: ³	kaw ³	(F., T.Tr. Dõi)
	Arem	alɔ: ²	ŋkɔ: ²	(F., T.Tr. Dõi)
	PROTO VIETIC	*C.lɔ?	*r.ko?	(F.)

In Vietic, “paddy” is a loanword from Chinese : Old Chinese ***C.lu?**> Middle Chinese ***dao**> *dào* 稻 “rice plant” (Sagart 1999: 181). An ancient connection between ***C.lu?** and the name of the taro in AA is possible (see Khmu **sro?** “taro”).

(6)	Somre	ha:l	rək ^h o:	(Martin <i>unp.</i>)
	Chong	ha:j	k ^h o:	(Siripen 2001)
	PROTO PEARIC	*ha:l	*r.k ^h o:	(Headley 1985)

2.2 Pairs in which “paddy” originates from the name of another edible plant

	“paddy, raw rice”	“husked rice, elaborated rice”	
(7)	Suei	srɔ:	həŋkaw (F.)
(8)	Pacoh	trɔ:	asə? (Watson 1979)
(9)	Kantu	arɔ:	amə:j (F.)
	PROTO KATUIC	*srɔ:	- - - (Sidwell 2005)

In Katuic languages the root ***srɔ:** “raw rice” is associated with three roots for “husked rice”, including **asə?** and **amə:j**, which are not attested in other MK groups.

(10)	Old Mon	<i>sro</i> ’ [sroʔ]	<i>sñio</i> ’ [s.ŋuʔ]	(Shorto 1971)
	Middle Mon	<i>sro</i> ’ [srɔʔ]	<i>sñiu</i> [s.ŋuʔ]	(Shorto 1971)
	Modern Mon	spʔ <i>sro</i> ’/swa’	haoʔ <i>sñiu</i>	(Shorto 1971)
	Nyah Kur	c ^h ro:ʔ	hŋkɔ: (<Khmer)	(Theraphan 1984)
	PROTO MON	*sroʔ	*s.ŋəʔ	(Diffloth 1984, Ferlus 1983)
(11)	Old Khmer	<i>srū</i> / <i>srū</i> [sru:]	<i>raŋko</i> [rŋko:]	
	Modern Khmer	srɔw <i>srūv</i>	ʔəŋkɔ: <i>ʔaŋkar</i>	
	PROTO KHMER	*sru:	*rŋko:	(Ferlus 1992)

In Katuic, Khmer and Mon, the word “paddy” originates from the word “taro” (Ferlus 1996). More precisely, the name for taro was selected to name the rice. These two plants, though dissimilar, have a common history, and share the same farming niche which made possible the semantic shift of the word from “taro” to “rice”. Botanists think that rice originally was a wild grass in the taro fields.

(12)	Sora (Munda)	səɔ:	roŋko:	(Zide & Zide 1976)
------	--------------	------	--------	--------------------

Let us note, even if we are unable to give any explanation for this, that the pair in Sora is the same as in Khmer. This may be simply a parallelism, and not a result from a common origin. The root ***r.ko?** “husked rice” is attested in many Munda languages (Gta?, Remo, Gutob, juang).

	“paddy, raw rice”	“husked rice, elaborated rice”	
Hu	θák	ŋk ^h óʔ	(Svantesson 1991)
U	sáʔ	ŋk ^h ù	(Svantesson 1988)
Màn é	hók ³⁵	ta ^{ʔ35}	(Yán & Zhōu 1995)
Lua Phalok	hók	hók ku:	(F.)
Mang	lam ³⁵	bak ³⁵	(Yán & Zhōu 1995)
Bolyu	te ³	ʔja:ŋ ¹	(Edmondson 1995)

2.5 The case of the Thai languages

The vocabulary for rice in the Thai languages originates from MK. The generic term **k^ha:w^{C1}** (exception: **k^ha:w^{C1}** in Siamese) originates from the widespread root ***r.koʔ** “husked rice” in MK.

	“paddy”	“husked rice”
Thai Siamese	k ^h a:w ^{C1} pliak ^{D1} (ข้าวเปลือก)	k ^h a:w ^{C1} sa:n ^{A1} (ข้าวสาร)
Lao	k ^h aw ^{C1} piak ^{D1} (ເຂົ້າ ເປືອກ)	k ^h aw ^{C1} sa:n ^{A1} (ເຂົ້າ ສານ)

The states of rice are named by way of their compounds: **k^ha:w^{C1} pliak^{D1}** “raw rice” is formed on the generic **k^ha:w^{C1}** followed by the specifier **pliak^{D1}** “husk, bark”. Similarly, **k^ha:w^{C1} sa:n^{A1}** “husked rice” in which **sa:n^{A1}** comes from the Proto MK ***sa:l** “to peel”, Kenieng **sa:l**, Khmu **ha:l** (**h**-<***s**-); see Saek **yaw⁶ sa:l²** (Gedney 1993) in which the final ***-l** is preserved. Let us note that the final **-r** (ຣ) in written Siamese is simply a faulty etymology.

2.6 An attempt of synthesis

Six main properly MK roots, and their alternatives, for “paddy/ raw rice” were selected: ***k.ba:** (also ***ba:**), ***s.ŋəʔ**, ***cəh**, ***sroʔ** (also ***srə:** and ***sru:**), ***ha:l** and ***sa:**, the Vietic root ***C.ləʔ** is a loan from Ancient Chinese. These roots which cover “rice” in its most general sense (plant, grain not transformed), which is less specific than “husked rice”, are survivals of the oldest names from a time when rice was still a gathered wild plant. By a universal process, the states of elaborated rice were named by the addition of a determinative to the basic term, as we can observe today in Thin/ Phay, Pong and Maleng Brô. The natural phonetic wear of frequently used words reduced the compound words to their final syllable.

Interestingly enough, there are only two major roots which specify the elaborate state of rice: ***r.koʔ** and its widespread alternatives in across MK, and ***p^hɛ:** restricted to the Bahnaric group. The other terms are too local to be taken into account. One can think that these terms originally meant “to peel, to crush, to husk” although for the moment, this is not confirmed by the MK vocabulary.

The husking of rice with a pestle would be a technique discovered long after rice domestication and which would have spread from its centre of innovation. This discovery could be related to the invention of the large pestle which allows efficient rice husking without crushing the grains (Ferlus 2009). The older manner of rice consumption would have been gruel in which the crushed or roughly husked grain would be cooked in a large

amount of water. Nowadays, gruel is cooked with husked rice. The old manner was replaced by cooking by water saturation, in which rice and water are combined in well defined proportions so that the cooked grains adhere to each other without sticking.

If cooking in gruel is the simplest and the oldest method, cooking by water saturation requires by contrast real know-how, which can only result from an innovation located in time and space, which propagated together with the name for husked rice in the area of AA consumers. We think that the invention of pestle, husking, and cooking by saturation are originally linked processes, which could explain the low number of MK roots for “husked rice”.

Ultimately, the existence of a series of terms to designate the various states of rice, so characteristic for the languages of Eastern Asia would be the consequence of a long relationship between humans and rice. The narrowing of the MK area could make us think that other terms must have existed. An analysis of the map at the end of this paper can give an account of the distribution of the terms exclusively in the latest centuries, after the intrusion of Tibeto-Burman and Thai languages. Beyond that, any attempt at restitution can only be hazardous.

These considerations bear two implications: first of all, we cannot reconstruct a unique root for each state of rice; secondly, the first domestication of rice occurred in the AA linguistic area, somewhere in Central China.

3. Dispersal of one name of rice out its centre of origin

When rice cultivation extended towards areas where the wild plant did not exist, the name for rice in the donor language propagated together with the cultivation of the plant. This is the case for the AA etymon **C.rac* (C.: pre-syllabic consonant) from which daughter forms were generated which scattered in Eurasia (the rationale for this reconstruction is found in §.5).

This root is attested in Chinese by *lì* 糲 < Old Chinese **m.rats* “coarse grain, millet or rice” (Sagart 2003) and in Tibetan by *’bras* “rice (general term)”. We will roughly examine the propagation of the original etymon **C.rac* in both directions, towards the East, the Austronesian (henceforth AN) world, later towards West across the Indian and Iranian areas. Most terms which come from **C.rac* present the same sesqui-syllabic phonetic structure in spite of divergent phonetic evolutions. In the Formosan languages of the Austronesian area, the medial vibrant *-r-* often changed to *l*, *g*, *h* or *#*. The comparison of the earliest forms, Old Chinese **m.rats*, Tibetan *’bras*, Proto AN **b.Ras* (see §.3.1) and Sanskrit/Indus **vr̥jhi* (see §.3.3) shows that the pre-syllable **C-* was probably a labial unit.

General structure:

labial pre-syllable	vibrant <i>r</i> (or <i>l/g/h/#</i>)	vowel	final palatal
1	2	3	4

3.1 Propagation of the word “rice” to the East

The Proto AA etymon **C.rac* “rice (general)” evolved to Proto AN **b.Ras* “husked rice” (cf. Wurm 1975) which split into various forms among which the Malayo-Chamic *běras* is the most significant. We will give here a short list of terms (quoted with the transcriptions

used in the source publications - Ferrell 1969; Revel 1988) for “paddy” and “husked rice” followed by their breaking up according to the four elements in the syllabic structure.

		paddy	husked rice	syllabic structure			
				1	2	3	4
Proto Austronesian			*b.Ras	b.	R	a	s
Formosa: Atayal	pagai		buax	bu		a	x
Rukai	ʔagai		bəráʔ	bə	r	á	t'
Puyuma	rumai		vərás	və	r	á	s
Kavalan	sa:váʔ		bəyás	bə	y	á	s
Paiwan	paday		vat	v		a	t
Luzon: Tagalog	pálay		bigas	bi	g	a	s
Ivatan	paray		paray	pa	r	a	y
Kallahan	pagey		begah	be	g	a	h
Palawan: Batak	paráy		bəgas	bə	g	a	s
Visayas: Hanunóo	pāray		binugas (infix)	bi nu	g	a	s
Borneo: Kadazan	parai		wagas	wa	g	a	s
Kayan	pare		bahah	ba	h	a	h
Malayo-Chamic	padi		běras	bě	r	a	s

Cases of cohabitation between daughter forms resulting from *b.Ras by different ways can easily be observed in the same language, for example in Batak: *paráy* “unhusked rice” and *bəgas* “husked rice”. In the Austronesian area, “unhusked rice” is frequently *padil/paday* and we wonder whether this form might not result as well from *b.Ras! This problem of historical phonetics cannot be solved within the scope of this paper. The Austronesian realm is an area of intense seaborne circulation and it is difficult to follow there the lexical routes. On the contrary, the AA/MK area contrasts by its terrestrial stability.

3.2 Propagation of the word “rice” to the West: the Indian area

During its propagation towards the West, the etymon *C.rac crossed the Indian area (three linguistic families: Munda, Dravidian and Indo-Aryan), where the word underwent many fragmentations, then the Iranian area whence it spread towards the West.

The word is attested in Vedic (c. 1500 BCE) by *vr̥hi* < Indus **vr̥jhi* “rice (general)” which Michael Witzel (1999: 13) allots to a R̥gvedic Para-Munda layer acquired at the time when Indo-Aryans stayed in Punjab. That layer presents morphological similarities with Munda. Some specialists hastily simplified Para-Munda into Munda! The knowledge of the word in Sanskrit spares us the task of seeking the rice vocabulary in the Indo-Aryan languages of India. Let us note henceforth and before any further steps, the similarity in structure between Proto AN *b.Ras, Indus **vr̥jhi* and Latin *oryza*.

		1	2	3	4
Proto Austronesian	*b.Ras	b.	R	a	s
Indus	* <i>vr̥jhi</i>	v	r	ī	jhi
Latin	<i>oryza</i>	o	r	y	za

The vocalic difference is not an obstacle for this pairing, since correspondences between the low vowel **a** and of the high vowels, **i** or **ī**, are attested in more restricted areas:

Waic **ʔmar**, Lamet **mar**, Bahnaric **mir** “dry rice-field”
 Mon **ɗaik** (<ɗa:k) *ɗāk*, Khmer **tik** *dīk*, Old Khmer *dik* “water”
 Mon **^hlaʔ** *sla*, Khmer **slək** *slīk*, Old Khmer *slik* “leaf”

The Dravidian languages attest two types of names for rice. Full forms: Tamil *ariçi* and *virīgi* (Bloch 1925), Telugu *arise* (Burrow 1961) “husked rice” which fits well with **vrījhi*. Truncated forms: Tamil, Telugu *vari*, Tamil, Tulu *ari*, Tulu *bār* “paddy”. The Malagasy *vary* originated from Dravidian (Ottino 1975). In order to explain the difference between full form and truncated form, we have to turn to the Munda languages. It is well known today that the fundamental difference between the Munda languages and the MK languages lies in word intonation (Donegan & Stampe 1983): falling accent in Munda and rising accent in MK. In MK sesqui-syllables there is a rising accent and the pre-syllable is reduced and unstressed while the components of the main syllable are fully realized. In the Munda cognates of the MK sesqui-syllables, the pre-syllable is lengthened and becomes a full syllable with a significant vowel, while the final of the main syllable is simplified. As we believe that the Munda languages acquired these intonational features in contact with the Dravidian languages, we can explain the genesis of the truncated forms and replace their elements in the general structure of the words for “rice”, all the way from Sanskrit until the modern European languages.

Let us examine now a few examples taken from Bloch (Bl) and Burrow (Bu) (the numbers are those in their *Dravidian Etymological Dictionary*). Some of these terms mean “unhusked rice”, others “husked rice”, and others refer to a variety of rice.

		1	2	3	4
Proto Austroasiatic	*C.rac	C.	r	a	c
Old Chinese	*m.rats	m.	r	a	ts
Proto Austronesian	*b.Ras	b.	R	a	s
Indus	* <i>vrījhi</i>	<i>v</i>	<i>r</i>	<i>ī</i>	<i>jhi</i>
Sanskrit	<i>vrihi</i>	<i>v</i>	<i>r</i>	<i>ī</i>	<i>hi</i>
Hindi	<i>biryani</i> (suffix)	<i>bi</i>	<i>r</i>	<i>ya</i>	<i>(ni)</i>
Tamil (Bl)	<i>virīgi</i>	<i>vi</i>	<i>r</i>	<i>ī</i>	<i>gi</i>
Tamil (Bl)	<i>ariçi</i>	<i>a</i>	<i>r</i>	<i>i</i>	<i>çi</i>
Tamil (Bu 178)	<i>arici</i>	<i>a</i>	<i>r</i>	<i>i</i>	<i>ci</i>
Tamil, Telugu (Bu 4306)	<i>vari</i>	<i>va</i>	<i>r</i>	<i>i</i>	
Telugu (Bu 2991)	<i>nivari</i> (pref.) <i>(ni)</i>	<i>va</i>	<i>r</i>	<i>i</i>	
Tamil, Tulu (Bu 178)	<i>ari</i>	<i>a</i>	<i>r</i>	<i>i</i>	
Tulu (Bl)	<i>bār</i>	<i>bā</i>	<i>r</i>		
Malagasy	<i>vary</i>	<i>va</i>	<i>r</i>	<i>y</i>	

One can note the advanced phonetic diversification in Dravidian : however contrary to AN, medial **-r-** is well preserved while the final of the word was simplified in some languages.

3.3 Propagation of the word “rice” to the West: the Iranian area

In his rather old but well documented article, Jules Bloch (1925) links Old Greek *oruza* ὀρυζα (Modern Greek *rúzi* ρύζι) to an Old Persian form which he reconstructs as **wrinjhi*. The Farsi *birinj* passed on to all languages in the Caucasus.

The word was spread throughout the European languages starting from Greek. The syllabic structure is always maintained even if the first element is sometimes erased.

		1	2	3	4
Proto Austroasiatic	*C.rac	C.	r	a	c
Old Chinese	*m.rats	m.	r	a	ts
Proto Austronesian	*b.Ras	b.	R	a	s
Indus	* <i>vr̥jhi</i>	<i>v</i>	<i>r</i>	<i>ī</i>	<i>jhi</i>
Sanskrit	<i>vr̥hi</i>	<i>v</i>	<i>r</i>	<i>ī</i>	<i>hi</i>
Old Persian	* <i>wrin̥jhi</i>	<i>w</i>	<i>r</i>	<i>i</i>	<i>n̥jhi</i>
Farsi	<i>birin̥j</i>	<i>bi</i>	<i>r</i>	<i>i</i>	<i>n̥j</i>
Arabic	<i>'aruz</i>	<i>'a</i>	<i>r</i>	<i>u</i>	<i>z</i>
Old Greek	<i>oruza</i>	<i>o</i>	<i>r</i>	<i>u</i>	<i>za</i>
Greek	<i>rúzi</i>		<i>r</i>	<i>ú</i>	<i>zi</i>
Spanish, Portuguese	<i>arroz</i>	<i>a</i>	<i>rr</i>	<i>o</i>	<i>z</i>
Latin	<i>oryza</i>	<i>o</i>	<i>r</i>	<i>y</i>	<i>za</i>
Italian	<i>riso</i>		<i>r</i>	<i>i</i>	<i>so</i>
Old French	<i>ris</i>		<i>r</i>	<i>i</i>	<i>s</i>
French	<i>riz</i> [ri]		<i>r</i>	<i>i</i>	<i>z</i>
German	<i>reis</i>		<i>r</i>	<i>ei</i>	<i>s</i>
Dutch	<i>rijst</i>		<i>r</i>	<i>ij</i>	<i>st</i>
English	<i>rice</i> [rais]		<i>r</i>	<i>i</i>	<i>ce</i>

In the languages of the Caucasus Jules Bloch (1925: 46) quoted more terms on edible cereals which present the same phonetic structure: Andi *perin̥j*, Agoul *burunz*, Georgian *brič* “(cereal) gruel”, Old Scythian *wru̯jaka* “rye”. It is not uncommon for a term to be used to name a different edible plant in another language. E.g., in Occitan the name for corn, *mil*, originated from the name for millet. We have also noted that in Mlabri the general name for rice, **ju:k**, is reminiscent of Proto Southwestern Tai ***fuak** “(edible) core/rhizome of banana tree”.

3.4 What was the route of diffusion?

How could the words for rice, all from the etymon *C.rac from central China, have extended as far as into Indo-Aryan in The Punjab, as attested by Vedic at the second millennium BC, and which route may it have followed? We could think that this root was transmitted through Munda, however it is not attested there (cf. Zide 1976: 1299-1310) other than marginally in Santali *bādras* “variety of rice grown on Highlands”. The Munda vocabulary for rice is in fact related to that of the Mon-Khmer branch. Therefore the word could not have been transmitted through Munda. On the other hand, the etymon *C.rac is present in Dravidian in a wide variety of daughter forms, a fact that gives evidence for its long age in this language family. We believe that Dravidian may have been the intermediary in the propagation of the name for rice between the AA of Central China and the Indo-Aryan in The Punjab. The role of Dravidian was obscured by the decline of this language family under the pressure from Indo-Aryan and Munda languages.

Another solution is possible, although less likely, namely that the name for rice might have travelled across the northern Himalayas while crossing the Tibetan territory.

The Indian sub-continent is very rich in wild varieties of rice and botanists long believed that there was to be found the oldest cradle of rice domestication, until that place was recognized as being China (de Candolle 1883: 309-311). However, rice cultivated in India belongs to the variety *Oriza sativa indica*, whereas rice cultivated in China is *Oriza sativa japonica*. If the name for rice in India originated in China, this would imply that the

name would have shifted from one variety to the other. This process is not rare in linguistics. We have already mentioned that in Katuic, Mon, and Khmer the name for rice was originally the name for taro; and in the Caucasus, the Persian name for rice became the name for rye. In conclusion, we think that the pre-Neolithic populations of India consumed the wild *indica* obtained by gathering, as well as the rice cultivated; its name might have come from Austroasiatic populations from China before the Munda expansion to India.

The expansion of the word for rice from the Iranian area into Europe is rather difficult to trace. The tables presented in this paper simply compare the forms and do not claim to trace a possible route. More than three-quarters of a century ago, Jules Bloch (1925: 38) wrote:

Les noms européens et presque tous les noms sémitiques du riz se rattachent au grec. D'où venait le mot grec lui-même? A cette question, la réponse qui se présente d'abord est que le mot a dû venir avec le grain, d'Extrême-Orient, et sans doute plus particulièrement de l'Inde, le premier pays d'Extrême-Orient que les anciens ont connu, où d'ailleurs le riz pousse à l'état sauvage et a été cultivé dès une haute antiquité.

Today, the problem stands at the same point.

4. Other possible connections

There are possible matches between the etymon ***C.rac** and names for cereals in some Indo-European languages.

Proto Indo-European ***p^hūr** “wheat”, Greek *pūrós* “grain of wheat”, Sanskrit *pūra* “type of pastry”. Ancient Indo-European dialects ***wruh^hyo-** “rye”, Old English *ryge*, English *rye*, Old High German *rokko*, German *roggen* (Gamkrelidze 1995: 564-8). All these statements fall well enough within the template of the etymon.

		1	2	3	4
Proto Austroasiatic	*C.rac	C.	r	a	c
Proto Indo-European	*p^hūr-	p ^h ū	r	-	
Greek	<i>pūrós</i>	pū	r	ó	s
Ancient I.-E. dialects	*wruh^hyo-	w	r	u	g ^h yo-
Scythian	<i>wrujaka</i>	w	r	u	jaka
Old English	<i>ryge</i>		r	y	ge
Old High German	<i>rokko</i>		r	o	kko

These comparisons may seem audacious, but it should be noted that the phonetic and semantic differences involved are on the same scale as those observed in the Dravidian area.

5. Origin of the name for rice: the etymon ***C.rac**

We will now explain our choice of ***C.rac** as the primordial etymon of the name for rice which then spread out of its original cradle.

Let us first notice that the final **-c** (in **-p -t -c -k**), well represented in MK, corresponds structurally to Old Chinese **-ts** (in **-p -t -ts -k**). Sinologists interpret **-ts** as **-t + suffix -s**. We think that the combination **-ts** is too frequent in Old Chinese to be merely explained by a suffixation process. Chinese might have known the final **-c**, which might

have merged at a very early date with **-ts**. However, we cannot know whether this supposed final was inherited or borrowed.

The most ancient manner of rice gathering consists of taking the ear of rice in one's hand and drawing the hand along the stem up to the top of the plant, a method which is still practised by some hill tribes in Southeast Asia. In several MK languages this particular action is named after the root ***ru:c**: Proto Pong ***pru:c**⁷, Kenieng **ru:c**, Chong **rù:c**, Mường **k^hu:c**⁷ < ***kru:c**, Arem **aruac**, Vietnamese *rút* “to withdraw, extract”; as for Lao ຮຸດ **hu:t**⁸ < ***ru:t** (no final palatals in Thai-Kadai) it is a loan from MK. There is no equivalent word in the European languages, because this manner of gathering is not practicable with corn, where grains adhere firmly to the ear.

We generally observe a tendency in MK to name the actions of removal or scraping along things by words which have the initial **r-** and the final palatal **-c**. E.g., Khmu : **ra:c** “to tear”, **^hra:c** “to remove the bark of a tree”, **^hro:c** “to fray, thin bamboo straps”, **ro:c** “to empty bowels (while gripping between fingers)”, etc.

Further and more precise examples will justify our reconstruction of ***C.rac** “paddy” on which part of this paper is built. Mlabri (Nan, Thailand) attests **krɛc**, in **krɛc ju:k** “harvest rice” (Rischel 1995). However, this language underwent the change **a:/a > ɛ:/ɛ** in one part of its vocabulary. E.g.:

Khmu	Mlabri	
ma:m	mɛ:m	“blood”
ba:r	bɛ:r	“two”
ga:ŋ	gɛ:ŋ	“house”

These correspondences enable us to reconstruct ***krac** on early Mlabri. Nyah Kur, a language closely related to Mon, attests **rɛc** “to reap paddy”. If we compare this form with the Mon **ròt rat** “to reap, cut (standing crops)” (Shorto 1962), ***rac** can be reconstructed (Ferlus 1983, Diffloth 1984).

The reconstructed form ***rac**, shared by early Mlabri (in ***krac**) and Proto Mon, perfectly corresponds to the main syllable of ***C.rac**, and this in spite of the millennia of time span between the two forms. The fact that currently ***rac** indicates the harvest with a sickle is not an obstacle, as this shift in meaning simply represents a technological development.

Summary of some correspondences:

		1	2	3	4
Proto Austroasiatic	*C.rac	C.	r	a	c
Mlabri	*rac		r	a	c
Proto Mon	*rac		r	a	c
Old Chinese	*m.rats	m.	r	a	ts
Proto Austronesian	*b.Ras	b.	R	a	s
Sanskrit	<i>vr̥hi</i>	<i>v</i>	<i>r</i>	<i>ī</i>	<i>hi</i>
Old Persian	<i>*wr̥iŋhi</i>	<i>w</i>	<i>r</i>	<i>i</i>	<i>ŋhi</i>
Latin	<i>oryza</i>	<i>o</i>	<i>r</i>	<i>y</i>	<i>za</i>

Conclusion

The relation of ***rac** “to collect by tearing off the grain (of rice) along the stem with the hand” with ***C.rac** “rice (general)” is that of a verbal base to its nominal derivative.

Consequently, we can affirm that the original meaning of the etymon **C.rac* was simply “The cereal collected by tearing off the grains along the stem with the hand”.

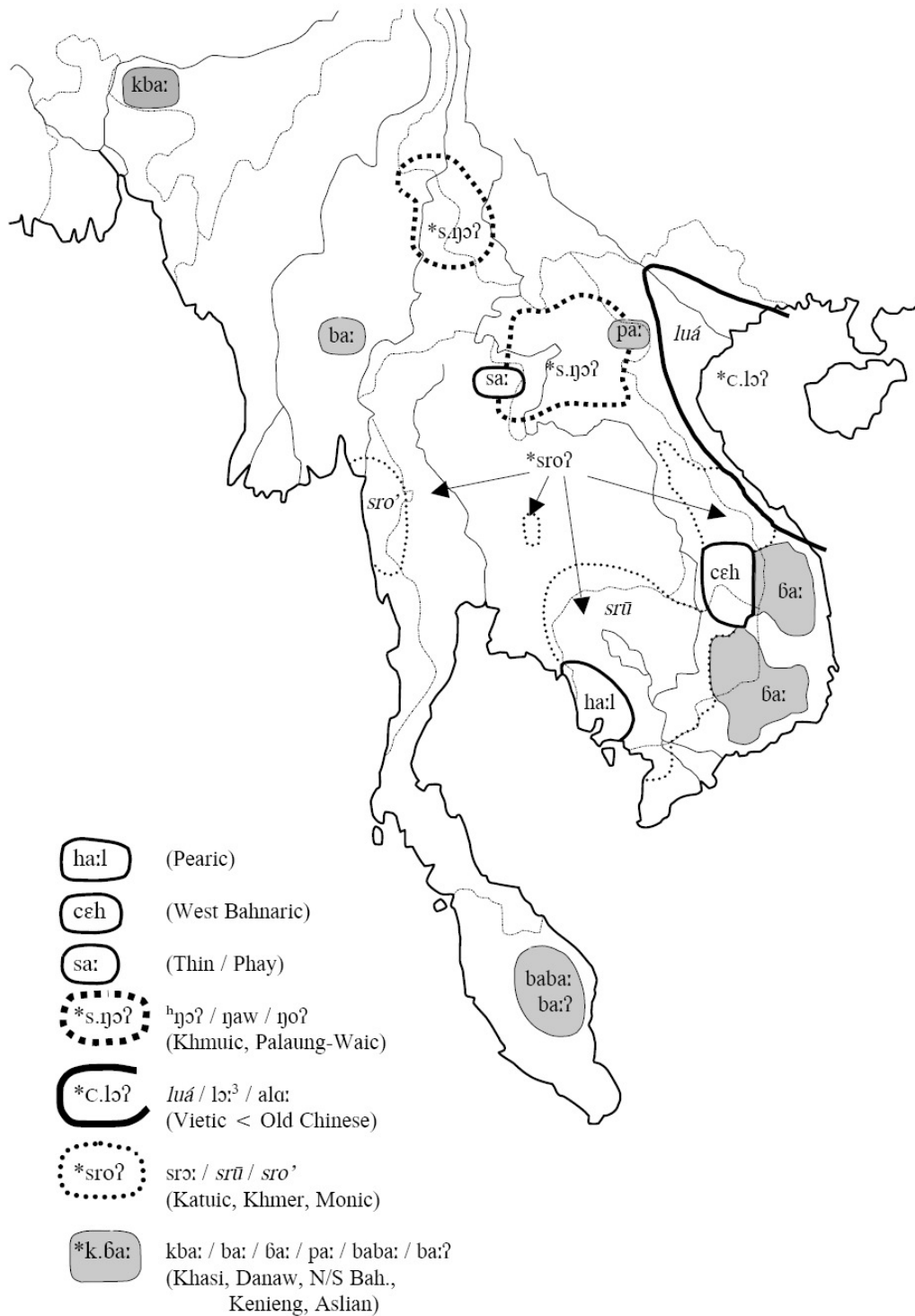


Figure 1: The seven main roots for “paddy/unhusked rice”

References

- Bars, Rev. E. 1973. *Khasi-English Dictionary*. Shillong: Don Bosco.
- Bellwood, Peter. 2005. *First Farmers*. Oxford: Blackwell Publishing. [Chap 7: The Spread of Agriculture into Southeast Asia, 128-145]
- Bellwood, Peter & Colin Renfrew (eds). 2002. *Examining the farming/language dispersal hypothesis*. University of Cambridge.
- Bloch, Jules. 1925. Le nom du riz. *Etudes Asiatiques*, publiées à l'occasion du 25^e anniversaire de l'École française d'Extrême-Orient (PEFEO 19): 37-47. Paris: G. Van Oest.
- Burenhult, Niclas & Claudia Wegener. 2009. Preliminary notes on the Phonology, orthography and Vocabulary of Semnam (Austroasiatic, Malay Peninsula). *JSEALS* 1: 283-312. Pacific Linguistics. The Australian National University.
- Burrow, T. & M.B. Emeneau. 1961. *A Dravidian Etymological Dictionary*. Oxford: Clarendon Press.
- de Candolle, Alphonse. 1883. Origine des plantes cultivées. Paris: Germain Baillièrre et Cie. [Reprint: Editions Jeanne Laffitte, Marseille]
- Diffloth, Gérard. 2005. The contribution of Linguistic Palaeontology to the Homeland of Austro-Asiatic. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), *The Peopling of East Asia*, pp. 77-80.
- Diffloth, Gérard. 1980. The Wa Languages. *Linguistics of the Tibeto-Burman Area* 5(2).
- Diffloth, Gérard. 1984. *The Dvaravati Old Mon Language and Nyah Kur*. Thailand: Chulalongkorn University Printing House.
- Donegan, Patricia Jane & David Stampe. 1983. Rhythm and the Holistic Organization of Language Structure. *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*. Linguistic Department of the Ohio State University.
- Edmondson, Jerold E. 1995. English-Bolyu Glossary. *Mon-Khmer Studies* 24: 133-159.
- Ferlus, Michel. 1983. Essai de phonétique historique du môn. *Mon-Khmer Studies* 12: 1-90.
- Ferlus, Michel. 1988. Les langues austroasiatiques. In Nicole Revel (sous la direction de), *Le riz en Asie du Sud-Est (Atlas du vocabulaire de la plante)*, 3 livrets, 310 p. + 64 p. + 78 cartes. Paris: École des Hautes Études en Sciences Sociales. Articles: 81-94, Vocabularies: 327-31, Maps n° 26 to n° 34.
- Ferlus, Michel. 1992. Essai de phonétique historique du khmer (du milieu du premier millénaire de notre ère à l'époque actuelle), *Mon-Khmer Studies* 21: 57-89.
- Ferlus, Michel. 1996. Du taro au riz en Asie du Sud-Est, petite histoire d'un glissement sémantique. *Mon-Khmer Studies* 25: 39-49.
- Ferlus, Michel. 2001. Les hypercorrections dans le thổ de Làng Lữ (Nghệ An, Vietnam): les pièges du comparatisme. *Quinzièmes Journées de Linguistique de l'Asie Orientale*. Paris, Centre de Recherches Linguistiques sur l'Asie Orientale (CNRS-EHESS), 7-8 juin 2001.
- Ferlus, Michel. 2009. A Layer of Dongsonian Vocabulary in Vietnamese. *Journal of Southeast Asian Linguistic Society* 1: 95-109.
- Ferlus, Michel. Unpublished data on Mon-Khmer languages.

- Ferrell, Raleigh. 1969. Taiwan Aboriginal Groups: *Problems in Cultural and Linguistics Classification*. Academia Sinica Monograph 17. Taiwan.
- Fuller, Dorian Q., Ling Qin & Emma Harvey. 2008. Evidence for a late onset of agriculture in the Lower Yangtse région and challenges for an archeobotany of rice. In Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros and Marie Lin (eds), *Past Human Migrations in East Asia*, pp. 40-83. London & New York: Routledge.
- Gamkrelidze, Thomas V. & Vjačeslav V. Ivanov. 1995. *Indo-European and the Indo-Europeans*, part I *The Text*, part II *Bibliography, Indexes*. Berlin/New York: Mouton de Gruyter.
- Gedney, William J. 1993. *The Saek Language*, edited by Thomas John Hudak. The University of Michigan.
- Gradin, Thôn & Dwight. 1979. *Jeh Vocabulary*, edited by Patrick Cohen. Manila: Summer Institute of Linguistics.
- Gregerson, Kenneth & Marilyn. 1977. *Rengao Vocabulary*. Summer Institute of Linguistics.
- Guilleminet, Paul & R.P. Jules Alberty. 1959. *Dictionnaire Bahnar-Français*, 2 vols. Paris: EFEO.
- Haudricourt, André Georges. 1968. Les arguments géographiques, écologiques et sémantiques pour l'origine des Thai. *Readings on Asian Topics, Papers read at the inauguration of the Scandinavian Institute of Asian Studies*, 16-18 September 1968, Scandinavian Institute of Asian Studies Monograph Series 1: 27-34. Copenhagen.
- Hauvers, Lorraine & Ralph H. 1991. *Stieng-English Dictionary*. Summer Institute of Linguistics.
- Headley, Robert K. 1985. Proto-Pearic and the classification of Pearic. In *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*. Edited by Surya Ratanakul, David Thomas, & Suwilai Premasirat. Bangkok, Mahidol University: 428-478.
- Higham, Charles. 2002. Languages and Farming Dispersals: Austroasiatic Languages and Rice Cultivation. In Peter Bellwood & Colin Renfrew (eds), *Examining the Farming / Language Dispersal Hypothesis*, pp. 223-232. University of Cambridge: McDonald Institute for Archeological Research.
- Isara Choosri. 2007. Investigating Contact-induced Language Change : Cases of Chung (Saoch) in Thailand and Cambodia. Thesis, Mahidol University.
- Jacq, Pascale & Paul Sidwell. 2000. *A Comparative West Bahnaric Dictionary*. Lincom Europa.
- Kosaka, Ruychi. *See* Shintani 2001.
- Kruspe, Nicole. 2004. *A Grammar of Semelai*. Cambridge University Press.
- Luce, Gordon H. 1965. Danaw, a Dying Austroasiatic Language. In Milner, G.B. & Eugénie J.A. Henderson (eds), *Indo-Pacific Linguistic Studies* 1 : 98-129. Amsterdam.
- Maier, J. & Đinh Văn Cao. 1976. *Cua Vocabulary*. Summer Institute of Linguistics, microfiches.
- Martin, Marie A. *Unpublished data* [Samrê, Chong].

- Nguyễn Văn Lợi. 1993. *Tiếng Rục*. Nhà Xuất Bản Khoa Học Xã Hội.
- Nguyễn Văn Lợi, Nguyễn Hữu Hoàn & Tạ Văn Thông. 2008. *Tiếng Mảng*. Nhà Xuất Bản Khoa Học Xã Hội.
- Ottino, Paul. 1975. L'origine dravidienne du vocabulaire du riz et de certains termes de riziculture à Madagascar. *Annuaire des pays de l'Océan Indien* 2: 103-121.
- Paulsen, Debbie. 1989-1990. A Phonological Reconstruction of Proto-Plang. *Mon-Khmer Studies* 18-19: 160-222.
- Pejros, Ilia & Victor Schnirelman. 1998. Rice in Southeast Asia: A regional interdisciplinarity approach. In Roger Blench & Matthiew Spriggs (eds), *Archeology and Langage II*, pp. 379-389. London: Routledge.
- Revel, Nicole (ed.). 1988. Le riz en Asie du Sud-Est (Atlas du vocabulaire de la plante), 3 livrets: articles, vocabulaires et planches. Paris: Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- Rischel, Jørgen. 1995. *Minor Mlabri, A Hunter-Gatherer Langage of Northern Indochina*. Copenhagen: Mudum Tusculanum Press.
- Sagart, Laurent. 1999. *The Roots of Old Chinese*. Amsterdam/ Philadelphia: John Benjamins.
- Sagart, Laurent. 2003. The Vocabulary of Cereal Cultivation and the Phylogeny of East-Asian Languages. *Bulletin of the Indo-Pacific Prehistory Association*, 1 (Taipei papers) 1: 127-136.
- Sagart, Laurent, Roger Blench & Alicia Sanchez-Mazas (eds). 2005. *The Peopling of East Asia*. London & New York: Routledge.
- Sanchez-Mazas, Alicia, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin (eds). 2008. *Past Human Migrations in East Asia. Matching archeology, linguistics and genetics*. London & New York: Routledge.
- Scott, James George & J.P. Hardiman. 1900. *Gazetteer of Upper Burma and the Shan States* 1(1). Rangoon: Superintendent of Government Printing. [voc. Angkú 720-722]
- Shintani, Tadahiko & Ryuichi Kosaka & Takashi Kato. 2001. *Linguistic Survey of Phongxaly, Lao P.D.R.* Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Shintani, Tadahiko. 2008. *The Palaung Language, Comparative Lexicon of its Southern Dialects*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Shorto, Harry L. 1962. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press.
- Shorto, Harry L. 1971. *A Dictionary of the Mon Inscriptions from the sixth to the sixteenth Centuries*. London: Oxford University Press.
- Sidwell, Paul. 2000. *Proto South Bahnaric*. Canberra: ANU, Pacific Linguistics.
- Sidwell, Paul. 2005. *The Katuic Languages*. Lincom Europa.
- Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric*, vol 1 *West Bahnaric*. Canberra: ANU, Pacific Linguistics.

- Siripen Ungsitipoonporn. 2001. A phonological comparison between Khlong Phlu Chong and Wangkraphrae Chong. MA thesis. Institute of Language and Culture for Rural Development, Mahidol University.
- Skeat, Walter William & Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsula*, two vols. London: Franck Cass & Co. Reprint 1966, New York : Barnes & Noble.
- Smith, Kenneth D. 2000. *Sedang Dictionary*. Dallas: Summer Institute of Linguistics.
- Svantesson, Jan-Olof. 1988. U. *Linguistics of the Tibeto-Burman Area* 11(1): 64-133.
- Svantesson, Jan-Olof. 1991. Hu – A Language with an Unorthodox Tonogenesis. In J.H.C.S. Davidson (ed.), *Austroasiatic Languages: Essays in Honour of H.L. Shorto*: 67-79.
- Theraphan L. Thongkum. 1980. *A Bruu-Thai-English Dictionary*. Bangkok: Chulalongkorn University.
- Theraphan L. Thongkum. 1984. *Nyah Kur (Chao Bon)-Thai-English Dictionary*. Bangkok: Chulalongkorn University Printing House.
- Thomas, David & Thổ Sảng Lục. 1966. *Chrau Vocabulary*. Saigon.
- Trần Trí Dõi. See Ferlus *Unpublished data*. [Mường, Rục, Arem]
- Watson, Richard & Sandra W. & Cubuat. 1979. *Pacoh Vocabulary*. Manila: Summer Institute of Linguistics.
- Witzel, Michael. 1999. Substrate Languages in Old Indo-Aryan (Rgvedic, Middle and Late Vedic). *Electronic Journal of Vedic Studies* 5(1): 1-67. [Sanskrit/ Indus]
- Wurm, S.A. & B. Wilson. 1975. *English Finderlist of Reconstructions in Austronesian Languages*. Pacific Linguistics C-33. Canberra: The Australian National University.
- Yán Qíxiāng [颜其香] & Zhōu Zhízhì [周植志]. 1995. *The Mon-Khmer Languages in China and Austro-Asiatic Languages* [中国孟高棉语族语言与南亚语系]. China: The Central University for Nationalities Press. [Mang, Mán É, Pesin]
- Zide, Arlene R.K. & Norman H. Zide. 1976. Proto-Munda cultural vocabulary: evidence for early agriculture. In Philip N. Jenner & als (eds) *Austroasiatic Studies* part II, pp. 1295-1334. Honolulu: University of Hawaii Press. [Sora, Gta?, Remo, Gutob, Juang]

A PRELIMINARY STUDY ON EARLY CHANGES OF VERBAL NEGATORS IN THAI*

Kiyoko Takahashi

Kanda University of International Studies

<kiyoko@kanda.kuis.ac.jp>

0 Abstract

This preliminary study on the evolution of the Thai negative system aims to set forth a hypothesis on part of the early development of the system. It focuses on two erstwhile verbal negators: *bə̀ə mii* and *bə̀ə hə̀əŋ*. The hypothesis is that during the time in which both *bə̀ə mii* and *bə̀ə hə̀əŋ* were used, these two negators were in a marked contrast with respect to the factuality status of the situation described: they were, respectively, irrealis vs. realis negators.⁴⁸

1 The variety of Thai negative expressions

Standard Thai has a number of negators with different forms and functions. Examples (1) to (7) below illustrate a variety of Thai negative expressions containing a negator, which is a single negative morpheme (such as verbal negator *māy* in (1), *mīʔ* in (2) and *hə̀əŋ* in (3)) or a concatenation of two negative morphemes, one of which may be an erstwhile negative

* The present article represents a further development of two conference papers, one of which was given at LSJ136 (the 136th General Meeting of the Linguistic Society of Japan) in Tokyo, June 21-22, 2008, and the other at SEALS19 (the 19th Annual Meeting of the Southeast Asian Linguistics Society) in Ho Chi Minh City, May 28-29, 2009. I wish to express my gratitude to the audiences at these two conferences for thoughtful comments. I am indebted to Bruce Horton and Heiko Narrog for stylistic suggestions and helpful comments on an early draft of this article. Thanks are also due to anonymous reviewers for important suggestions and constructive criticisms of some sections of the article. Any remaining shortcomings are entirely my own responsibility.

⁴⁸ Although a reviewer suggests that the nomenclature ‘counterfactual vs. factual’ would be more appropriate for characterizing the two negators: *bə̀ə mii* vs. *bə̀ə hə̀əŋ*, in this paper I use the nomenclature ‘irrealis vs. realis’ since recently linguists tend to use the nomenclature ‘irrealis vs. realis’ for the notional contrast of ‘non-factual vs. factual’, or of ‘unreal vs. real’ (Palmer 2001: 1). The term ‘irrealis’ was used by Sapir (1930) in his description of Southern Paiute grammar. He notes that the ‘irrealis’ modal suffix in the language indicates that the activity expressed by the verb is ‘unreal’, i.e. either merely potential or contrary to fact (Sapir 1930: 168, 1992: 186). So far many other terms for the concept ‘irrealis’ have been used in linguistics literature. Examples are: ‘manifesting’ (Whorf 1950: 59), ‘nonfactive’ (Hooper 1975: 91), ‘non-factivity’ (Lyons 1977: 795), ‘irrealis-assertion’ (being asserted with doubt, as hypotheses; being weekly asserted) (Givón 1982: 24; 1994: 268), ‘non-actual’ (Chung & Timberlake 1985: 241), ‘non-assertion’ (Bybee et al. 1994: 239; Bybee & Fleischman 1995: 9), ‘nonfactuality’ (being undetermined with respect to its factual status, i.e., is neither positively nor negatively factual) (Narrog 2005: 182, 184), and so forth.

morpheme (such as verbal negator *bòò mí?* in (4)⁴⁹ and nominal negator *mây chây* in (5)⁵⁰) or a composite form consisting of at least one negative morpheme plus other morpheme(s) (such as formulaic verbal negator *hăa ... mây* in (6) and formulaic nominal negator *hăa chây ... mây* in (7)).⁵¹

- (1) **mây** yàak
NEGATIVE want
(I) do not want. [NWRP_EN016]
- (2) năŋ **mí?** dâý còp loŋ troŋ nán
movie NEGATIVE REALIZATION end descend there
The movie does not really end up at that point. [NWRP_EN015]
- (3) ɲaa sǎan ruuu **hòon** hân hòt
ivory big elephant INTERROGATIVE NEGATIVE be worn out
Is it the ivory of a big elephant that would not be worn out? [POET018]
- (4) mǝ? sǝm kɛɛ chaay yǝŋ thîi **bòò mí?**
fit DATIVE man woman RELATIVIZER NEGATIVE
sǝm thaan phêet maa lǎay pii dii nák
copulate come many year good INTENSITIVE
(This vitamin) is fit indeed for a couple who have not had sex for many years.
[NACNS024]
- (5) **mây chây** chaaw rooman
NEGATIVE people Roman
(They) are not the Romans. [NACHM070]
- (6) **hăa** rúu tua **mây**
NEGATIVE (front part) be aware NEGATIVE (rear part)
(He) is not aware. [NACHM]

⁴⁹ I consider *bòò mí?* as compound verbal negator consisting of the erstwhile authentic negative *bòò* and another negative *mí?* on the grounds that the following syntactic structure of a negative expression, which is found in an inscription produced around the 19th century, shows that *bòò* and *mí?* were functioning as a single negator: *bòò [mí?VP1 mí?VP2]* ‘neither VP1 nor VP2’.

⁵⁰ In old days *chây* by itself was a nominal negator (Bradley 1873: 170).

⁵¹ Sample negative expressions in (1) to (7) are derived from the Thai National Corpus (TNC) which is the largest electronic Thai corpus that comprises numerous corpora of various discourse genres (e.g. ‘fiction’, ‘newspaper’, ‘academic’, ‘law’, etc.) [http://www.arts.chula.ac.th/tnc2/]. The English glosses and translations are mine. The sequence of signs following my free translation of each sample expression (e.g. [NWRP_EN016]) is the ID code given to a component corpus that includes the expression. I would like to thank Wirote Aroonmanakun for directing my attention to the public availability of the TNC corpus. Besides, negative expressions in old-day Thai cited in this paper, which are transcribed into phonetic equivalents in modern Thai, are all from *Corpus of Thai Inscriptions* (see the reference section below).

- (7) **hǎa chây** sǐnlapà? **mây**
 NEGATIVE (front part) the arts NEGATIVE (rear part)
(They) are not the arts. [ACHM054]

The syntactic configuration of negative expressions with a simple negator (a single negative morpheme and a concatenation of two negative morphemes) like those in (1) to (5) is relatively simplex; that is, the negator is placed immediately before the negated verb phrase or noun phrase. On the other hand, the syntactic configuration of formulaic negative expressions with a composite negator like those in (6) and (7) is complex; that is, the negated verb phrase or noun phrase is put between the front and the rear parts of the negator. Thus, negative expressions in present-day Thai are quite diverse in form. In order to identify the exact period in which the diversity of Thai negative expressions became conspicuous, I have consulted the corpus of Thai inscriptions (for the details of this corpus, see the reference section at the end of this paper) which contains Thai inscriptions from the end of the 13th century (the Sukhothai dynasty) through the 20th century (the present Ratanakosin dynasty). With this diachronic corpus data, I have learned that Thai negative expressions had been of great variety all the time since the earliest period in the documented history of the Thai language, namely since the end of the 13th century. The number of tokens of negators that I have found in the inscriptions is approximately 550 in total (Takahashi 2008). In examining the inscription discourses, I found a variety of verbal and nominal negators.⁵²

This preliminary study on early development of the Thai negative system focuses on two erstwhile verbal negators, *bòò mii* and *bòò hòòn*, as well as their probable descendants, *mii* / *mí?*⁵³ and *hòòn*. These negators were frequently used in inscriptions produced in the period from the 14th century to the middle of the 19th century (Takahashi 2008). Their possible historical changes are shown in diagrams (8) and (9).

- (8) a. **bòò mii** ‘not exist’, negative existential construction
 [negation of existence]

> b. **bòò mii** VP
 [irrealis negative situation (unwitnessed non-factual situation)]

> c. **mii** / **mí?** VP
 [regular negation]

⁵² For example, *bòò* VP, *pây* VP, *hòòn* VP, *mii* VP, *mí?* VP, *mây* VP, *bòò hòòn* VP, *bòò mii* VP, *bòò mí?* VP, *hǎa* VP *mây*, VP *hǎa mí?* *dây*, *chây* VP, *chây* NP, *chây* (NP) *cà?* VP *hǎa mí?* *dây*, *mí?* *chây* NP, *mây* *chây* NP, *hǎa chây* NP *mây*, and so on are attested (Takahashi 2008: 356-357).

⁵³ The negative morpheme *mii* had several variants such as *mì*, *mí?* and *mí?*.

- (9) a. **ບໍ່** **ເຮັດ** ‘not accustomed’, negative experiential construction
[negation of experience]

> b. **ບໍ່** **ເຮັດ** VP
[realis negative situation (witnessed factual situation)]

> c. **ເຮັດ** VP
[regular negation]

Some twenty years ago, Kullavanijaya (1996 [original ms., 1988]: 89) raised a question about the relationship between *ບໍ່ mii* and *mii*, as follows: “Could it be that the form *ບໍ່ mii* ‘have not’ has gradually developed into *mii* ‘not’?” Since then, however, this question has been left open. This study, therefore, will investigate a plausible development from *ບໍ່ mii* to *mii* / *mī*?, as indicated in (8) above. I will also examine a plausible development of *ບໍ່ ເຮັດ* to *ເຮັດ*, as indicated in (9) above, in order to show a parallel between the two plausible developments (8) and (9).

The purpose of the present study is to motivate a hypothesis on early changes of the two erstwhile verbal negators, *ບໍ່ mii* and *ບໍ່ ເຮັດ*, drawing on Croft’s (1991) account of the typical evolution of negation (see Section 2). The organization of the remainder of this paper is as follows. Section 2 first reviews Croft’s proposal on the course of historical changes in negative existential expressions, which he named ‘negative-existential cycle’. In line with this suggestion, Sections 2.1 to 2.3 present a hypothesis on the early changes of Thai negative expressions based on examination of actual tokens gathered from the inscription data. Essentially, I am suggesting the following. As diagrams (8) and (9) above show, expressions for ‘negation of existence’ (8a) changed into those for ‘irrealis negative situation’ (8b), which is parallel to the change from expressions for ‘negation of experience’ (9a) to those for ‘realis negative situation’ (9b). Furthermore, the two contrastive negative expressions (8b) and (9b) were similarly transformed into expressions for ‘regular negation’ (8c) and (9c). In Section 3, then, I will clarify the types of language change probably involved in the early development of Thai negative system. In particular, I will explain my hypothesis that three well-known types of the diachronic process of language change (i.e. syntactic reanalysis, generalization by analogy and phonological reduction) were involved at different phases of the developments (8) and (9). Section 4 is a brief summary of this study.

2 Negative-existential cycle

Croft (1991) posits a historical linguistic process of ‘negative-existential cycle’, shown graphically in Figure 1 which is adopted from Figure 2 in Croft (ibid.: 6).

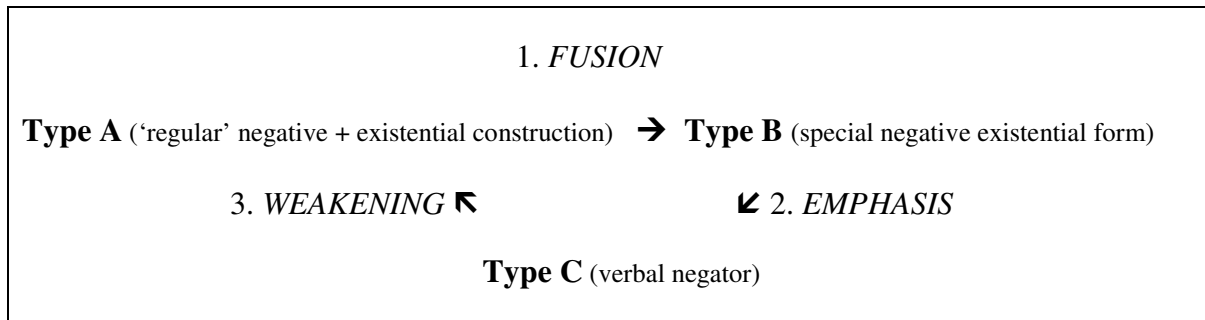


Figure 1: Croft's (1991) proposal for a diachronic 'negative-existential cycle'

In a negative-existential cycle, a special negative existential form (Type B) arises and comes to be used as a verbal negator (Type C) and then is supplemented by the positive existential predicate in its existential function, restoring a 'regular' negative + existential construction (Type A). Croft (1991: 22) states that in this diachronic cycle, fusion of negative and existential occurs first, then the emphatic use of the negative existential as a verbal negator, and finally the analogical use of the positive existential predicate in negative existential constructions being accompanied with attenuation of emphasis. However, he adds that the sequencing is not absolute. For example, in Mandarin Chinese, fusion (Type B) did not occur but the 'regular' negative existential (Type A) jumped straight to the emphatic verbal negator (Type C). Diagram (10) illustrates this.

(10) Development of the negative existential *méi* in Mandarin Chinese:

Type A ('regular' negative existential): *méi* NP 'NP does not exist'

> Type B: none

> Type C (verbal negator): *méi* VP [for negation of complete action]
cf. *bu* VP [for normal declarative negation]

The negative existential *méi* in Mandarin Chinese came to function as verbal negator for negating the completion of an event (Li & Thompson 1981: 421) without any phonological fusion taking place.

Likewise, one of the erstwhile negators in Thai, *bə̀ə mii*, is supposed to have consisted of the negative *bə̀ə* 'not' and the existential verb *mii* 'exist'. A plausible evolution of *bə̀ə mii* is diagrammed in (11).

(11) Development of the negative + existential construction in Thai:

Type A ('regular' negative + existential construction): *bə̀ə mii* NP (e.g. (12))

> Type B: none

> Type C (verbal negator): *bə̀ə mii* VP (e.g. (13))

- (12) **b̀̀̀̀̀** mii ɲuan **b̀̀̀̀̀** mii thoŋ
 NEGATIVE exist silver NEGATIVE exist gold
There was not silver; there was not gold. [1](1292)⁵⁴

- (13) **b̀̀̀̀̀ mii** khananaa th̀̀̀ ləəy
 NEGATIVE recount careful INTENSITIVE
(We) do not recount (what he did) in a careful manner at all (because it was too much to be fully described). [2](1341-1367)

Examination of the inscription corpus data reveals that *b̀̀̀̀̀ mii* NP (Type A) was used until the 16th century and *b̀̀̀̀̀ mii* VP (Type C) until the 18th century (Takahashi 2008).

2.1 From ‘negation of existence vs. of experience’ (Type A) to ‘negation of irrealis vs. realis situation’ (Type C): the process of emphasis

Although in Thai there has been no negative existential with a unique form distinct from other negators (like *méi* ‘not exist’ in Mandarin Chinese), I assume that the development of the Thai negative existential construction *b̀̀̀̀̀ mii* basically corresponds to that of the Mandarin Chinese negative existential *méi* in that both are not fully associated with phonological fusion. The change occurred in the negative existential construction in Thai, however, has something different from its Mandarin Chinese counterpart. Crucially, it is likely that the evolution of *b̀̀̀̀̀ mii* was not an isolated change in the Thai negative system, but another erstwhile negator *b̀̀̀̀̀ h̀̀̀̀̀* appears to have undergone a similar change, as shown in (14) below. *b̀̀̀̀̀ h̀̀̀̀̀* is composed of the negative *b̀̀̀̀̀* ‘not’ and the experiential verb *h̀̀̀̀̀* ‘be accustomed’. Therefore, we may call *b̀̀̀̀̀ h̀̀̀̀̀* a negative experiential construction.

(14) Development of the negative + experiential construction in Thai:

The negative + experiential construction: *b̀̀̀̀̀ h̀̀̀̀̀* ‘not accustomed’

> verbal negator: *b̀̀̀̀̀ h̀̀̀̀̀* VP (e.g. (15))

- (15) **b̀̀̀̀̀ h̀̀̀̀̀** khàat sàk wan sàk khurun
 NEGATIVE lack just day just night
(He did) without missing a single day or a single night. [3](1357)

I hypothesize that the verbal negator *b̀̀̀̀̀ mii*, which likely originates from the negative existential construction, and the verbal negator *b̀̀̀̀̀ h̀̀̀̀̀*, which likely originates from the negative experiential construction, were once contrastive in terms of modal sense (whether the described situation in question is non-factual or factual): irrealis (non-factual) versus realis (factual) negator. The grounds for this hypothesis is my observation of the inscription corpus data that all the 13 tokens of ‘*b̀̀̀̀̀ mii* VP’ could be interpreted as representing negative situation of the irrealis kind (non-factual situation such as non-

⁵⁴ The bracketed number (e.g. [1]) and the parenthesized number (e.g. (1292)) after my free translation of each sample expression from the inscription corpus are, respectively, the ID number and the estimated production year of the source inscription.

realization, impossibility, hypotheticals, dispositional necessity, and generalization) whereas all the 6 tokens of ‘*bə̀ə hə̀ə* VP’ could be interpreted as representing negative situation of the realis kind (factual situation such as experience and perception). Note that unfortunately the number of tokens of negation by the verbal negators *bə̀ə mii* and *bə̀ə hə̀ə* that I have found in the inscriptions is very small: in total, 13 tokens of ‘*bə̀ə mii* VP’ (see Appendix A) and 6 tokens of ‘*bə̀ə hə̀ə* VP’ (see Appendix B).

As exemplified in (16) to (18) below, *bə̀ə mii* is used to express an irrealis situation: such as a non-realized or impossible situation (16), a conditional situation (17), or a habitual or generalized situation (18).

- (16) **bə̀ə mii** sadət loŋ maa
 NEGATIVE proceed descend come
 (The relics) did not come down. [3](1357)

- (17) **phì?** **bə̀ə mii** kaŋwon ...
 if NEGATIVE worry
 If (they) do not worry ... [64](15C)

- (18) maa **bə̀ə mii** khàat
 come NEGATIVE lack
 (Usually they) come without missing an occasion. [78](1796)

In contrast, *bə̀ə hə̀ə* is used to express a witnessed or experienced negative situation, as illustrated in (19) and (20).

- (19) **bə̀ə hə̀ə** khâa fan
 NEGATIVE kill hit
 (He) has not killed or hit (the person quarrelled with him). [5](1361)

- (20) ŋən thə̀əŋ **bə̀ə hə̀ə** mii
 silver gold NEGATIVE exist
 As for silver and gold, they have not existed. [160](1782-1925)

Hence, in past ages there seems to be division of labour between the irrealis negator *bə̀ə mii* and the realis negator *bə̀ə hə̀ə*.

2.2 From ‘irrealis vs. realis negator’ (Type C) to ‘neutral negator’ (Type A): the process of weakening

I further hypothesize that the two negators *bə̀ə mii* and *bə̀ə hə̀ə* both changed into simplified neutral negators *mii* / *mí?* and *hə̀ə*, as respectively diagrammed in (21) and (24) below.

- (21) **bằ** **mii** VP
[irrealis negation]

> **mii / mĩ?** VP
[neutral negation] (e.g. (22), (23))

- (22) nằ lằ **mii** thừan
count and NEGATIVE in full
(They) are countless. [5](1361)

- (23) **mĩ?** rằkỉat
NEGATIVE object to
(He) does not conceive a dislike. [64](15C)

- (24) **bằ** **hằn** VP
[realis negation]

> **hằn** VP
[neutral negation] (e.g. (25))

- (25) cằ? lừum mii sỉ **hằn** dằ
IRREALIS forget exist glory NEGATIVE emerge
It is improbable to forget the existence of the glory. [245](1925-1978)

Typically, when contrasting negators are being neutralized, one of them drops out of use as there is no need for having two generalized negators with the equal status.⁵⁵ However, this story does not go for the contrasting negators *bằ mii* and *bằ hằn*. After shifting to modally neutral negators *mii / mĩ?* and *hằn* which can be used irrespective of the factuality status of the situation described, there existed differences in the distribution of their usages. *mii / mĩ?* served as a neutral negator occurring in non-specific, ordinary discourse; *hằn* became a neutral negator, too, but it appeared only in verses, which is a peculiar type of discourse. Nowadays *mĩ?* is still used as neutral negator but it is rather infrequent since it is confined to formal or literary language. The most frequent negator in present-day Thai is *mằy*, which is supposed to have risen from the fusion of *mĩ?* and a versatile functional morpheme *dằy* or *hằy* (see Section 2.3).

Table 1 below indicates the use time span of each of the above-mentioned old verbal negators, which I have attested in surviving available inscriptions produced during the period from the end of the 13th century to the 20th century.

⁵⁵ This was pointed out by an anonymous reviewer.

Table 1: *Use time span of old verbal negators in Thai*

	13C	14C	15C	16C	17C	18C	19C	20C
<i>bə̀ə</i> VP	-----							
<i>bə̀ə mii</i> ‘exist’ NP	-----							
<i>bə̀ə mii</i> VP	-----							
<i>mii</i> / <i>mí?</i> VP	-----							
<i>bə̀ə hə̀ə̀n</i> VP	-----							
<i>hə̀ə̀n</i> VP								-----

From Table 1, we can see the following:

- (i) the negator *bə̀ə* was used in the inscriptions until the 19th century, and the negative existential construction ‘negative *bə̀ə* + existential verb *mii* + NP’ in particular was used until the 16th century;
- (ii) the negator *bə̀ə mii* lasted until the 18th century, and its probable descendants *mii* / *mí?* until the 20th century;
- (iii) the negator *bə̀ə hə̀ə̀n* persisted until the 20th century, and its probable descendant *hə̀ə̀n* appeared in the 20th century.

In the 20th century, when *bə̀ə hə̀ə̀n* and *hə̀ə̀n* coexisted, *bə̀ə hə̀ə̀n* barely remained as realis negator while *hə̀ə̀n* was given a special function as a literary negator, or so to speak, an elegant negator. Regarding the pair of *bə̀ə mii* and *mii* / *mí?*, *bə̀ə mii* disappeared around the 18th century while *mii* / *mí?* was getting more and more common and became the most common neutral negator by the 19th century. I suppose that the extinction of the irrealis negator *bə̀ə mii* should have some connection with the establishment of the irrealis marker *cà?* before the 18th century (cf. footnote 9), but in this paper I shall make no further inquiry into this point.

Because I could not find a sufficient number of actual tokens of these negators in the inscription corpus data (nor could I utilize documents written in Thai before the end of the 13th century since writing did not exist), these findings concerning the period of use of these negators are clearly speculative, and so they cannot be regarded as decisive evidence for verifying my hypothesis of the evolutions of *bə̀ə mii* to *mii* / *mí?* and of *bə̀ə hə̀ə̀n* to *hə̀ə̀n*. Yet, my findings suggest at least that this hypothesis entails no contradiction. The findings that the surviving time of *mii* / *mí?* is longer than that of *bə̀ə mii* and that *hə̀ə̀n* came to be used later than *bə̀ə hə̀ə̀n*, for example, are not in contradiction to the hypothesis at all, but we could rather say that the findings form circumstantial evidence in favour of the hypothesis.

2.3 Plausible negative-existential cycle in Thai

So far I have analysed the evolution of the negative existential construction (negative *bə̀ə* + existential verb *mii* ‘exist’) and its parallel evolution of the negative experiential construction (negative *bə̀ə* + experiential verb *hə̀ən* ‘be accustomed’) by relying on Croft’s model of negative-existential cycle (Figure 1 above, which is repeated as Figure 3 below). The result of the analysis is that the negative-existential cycle in Thai, as appears to be manifested in inscriptions, is similar but not completely identical to the model postulated by Croft. Figure 2 below graphically depicts a plausible negative-existential cycle in Thai. Compared with the cycle in Figure 3 (= Figure 1), it is notable that the cycle in Figure 2 does not involve phonological fusion in the shift from Type A (regular negative *bə̀ə* plus existential verb *mii*) to Type C (verbal negator *bə̀ə mii*), but instead it has partial phonological loss in the shift from Type C (irrealis negator *bə̀ə mii*) to Type A (regular negative *mii* / *míʔ*).

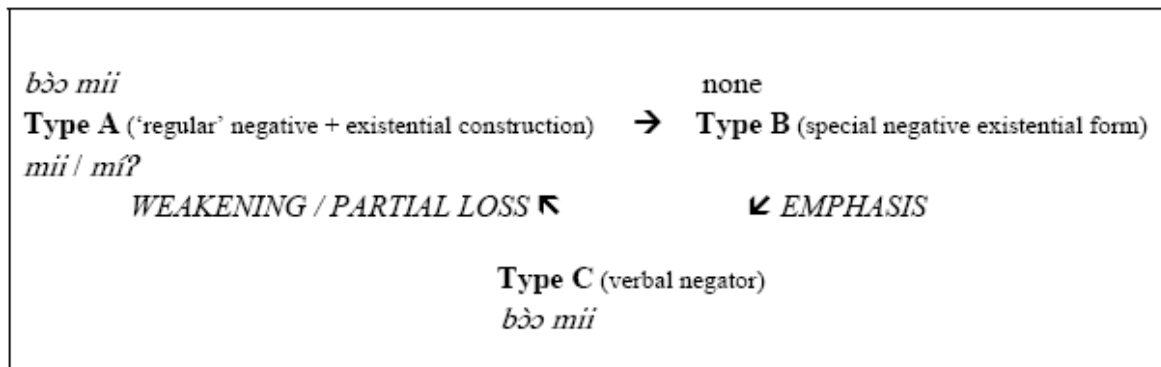


Figure 2: Plausible 'negative-existential cycle' in Thai

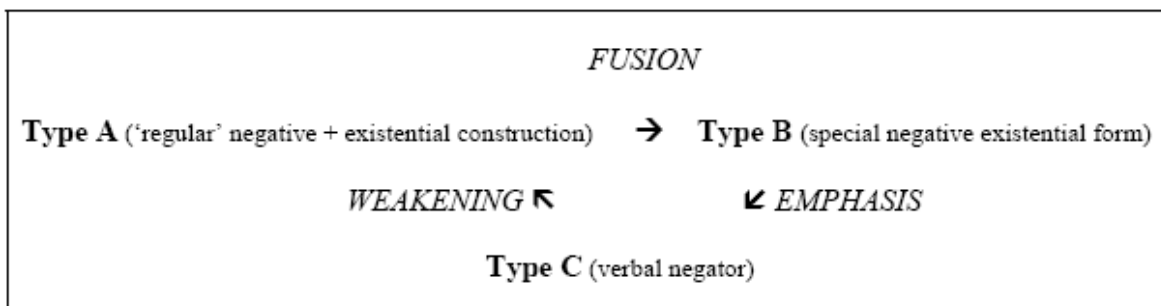


Figure 3: Croft's (1991) proposal for a diachronic 'negative-existential cycle'

It is evident that the plausible evolution of the Thai negative existential construction *bə̀ə mii* (Figure 2) is not in perfect accord with Croft's model of negative-existential cycle (Figure 3 = Figure 1). A salient feature of the plausible evolution of *bə̀ə mii* is that in the shift to the stage of Type C (verbal negator), *bə̀ə mii* gained the function of irrealis negator but did not involve phonological fusion.

The idea that formerly the negator *bə̀ə mii* and its variants *mii* / *míʔ* had a function to mark irrealis negation is supported by Kullavanijaya's (1996: 84) observation on uses of *mii* / *míʔ* in the Sukhothai inscriptions before the middle of the 15th century. She noted that *mii* / *míʔ* usually occur before the modal morpheme *dāy* 'can, capable' and express the

meaning of improbability (viz. something or an event would not happen or could not have happened), and that when they occur without *dây*, there will be a specific lexical item co-occurring such as *càk* ‘intend (not), shall (not)’ or *mûa day lee* ‘(not) any time’ which shows uncertainty in time. This is presumably because *mii* / *míʔ* are a post-irrealis-negator whose original function is to mark irrealis negation. Their affinity with ‘irrealis concept’ (to be specific, the concept of non-factuality pertaining to futurity or uncertainty, including expectation, hope, non-realization, possibility, validity, supposition, generalized situation, etc., which is apt to be marked by the irrealis marker *càʔ* in modern Thai⁵⁶) may remain for some time after developing into a neutral negator, given that preservation of a certain aspect of the former properties of a changing element is by and large seen in language change.⁵⁷ An additional piece of evidence in support of the identity of *míʔ* as post-irrealis-negator is my own observation of the inscription corpus that the irrealis marker *càʔ* did not co-occur with the former negative *bə̀ə*, but around the 18th century at which time *bə̀ə* gradually became less common, *càʔ* came to be used often in combination with *míʔ* (Takahashi 2008). This fact implies that the old negator *bə̀ə* had nothing to do with irrealis concepts, whereas the new negator *míʔ*, which I assume to be a descendant of the irrealis negator *bə̀ə mii*, was closely related with them.

In the cycle in Figure 2, *bə̀ə mii* at the stage of Type C (irrealis negator) then underwent partial phonological loss as well as attenuation of specificity in meaning, which gave rise to neutral verbal negators *mii* / *míʔ* at the stage of Type A (regular negative). In present-day Thai, *míʔ* is still used in formal contexts, but *mii* is no longer used. Ordinarily the latest negative form *mây* is used in both oral and written discourses. Kullavanijaya (1996) also offered a hypothesis on the development of *míʔ* into the currently most common modern negator *mây*. Example (26) is the earliest use of *mây* that I have found in the inscriptions.

- (26) **mây** thuun kèe caw thay ...
 NEGATIVE inform DATIVE the owner
 (If that person) does not inform the owner ... [38](1313-1433)

Her hypothesis on the latest change in the Thai negative system is that *míʔ*, which usually occurred before *dây*, as in (27) below, may have become contaminated with the vowel in *dây* and become *mây* (Kullavanijaya 1996: 87-88).

- (27) **míʔ** dâi rúcàk
 NEGATIVE REALIZATION know
 (They) do not get to know (it). [3](1357)

⁵⁶ According to Diller (1988: 286; 2001: 158), the irrealis marker *càʔ* derives from the verb *càk* meaning ‘desire, intend, consider’. In the middle of the 14th century the verb *càk* began to change its form into *càʔ* by replacing the velar stop with a glottal stop, and the latter phonologically reduced form began to be frequently used in succeeding ages (Diller 1988, Takahashi 2007). The original form *càk* is now confined to a few formal formulaic expressions (e.g. *càk khə̀əp khun yîy* ‘I would like to thank you very much.’).

⁵⁷ Hopper (1991: 22, 28-30) calls this effect occurring particularly in the process of grammaticalization ‘persistence’.

However, it is possible that *mây* has developed from the combination of *mī?* and another versatile functional morpheme *hây*. A serialization of *mī?* and *hây* is exemplified in (28).

- (28) *mī?* *hây* *phón* *ʔàatyaa* *thân* *nán*
 NEGATIVE INDUCEMENT escape crime the authority that
 (They bring him) not to let (him) out of the crime. [38](1313-1433)

To investigate the latest development in the Thai negative system is another intriguing issue, but it is beyond the scope of this study.

3 Three types of changes in the early development of Thai negative system

There are three major types of the historical process of language change that are widely recognized: (i) syntactic reanalysis, (ii) generalization by analogy, and (iii) phonological reduction. ‘Syntactic reanalysis’ is defined as “change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation” (Langacker 1977: 58). It is “a mechanism which changes the underlying structure of a syntactic pattern” (Harris 2003: 532). ‘Generalization by analogy’ here refers to generalization of construction⁵⁸ caused by analogy or “the attraction of extant forms to already existing constructions” (Hopper & Traugott 1993: 56). It is “a mechanism which results in changes in the surface manifestation of a [syntactic] pattern” (Harris 2003: 532). ‘Phonological reduction’ or ‘phonological attrition’ means “the gradual loss of phonological substance” (Lehmann 1995: 126). It brings about the shortening of forms. I suppose that these three types of change were all involved in the plausible early development of the Thai negative system which I have described in the preceding sections. My hypothesis is as follows.

First of all, the negative experiential construction followed by a verb phrase (‘negative *bə̀ə* + experiential verb *hə̀ən* + VP’ meaning that ‘relevant person is not accustomed to situation denoted by VP’) underwent syntactic reanalysis leading to a new (covert) structure consisting of a novel verbal negator *bə̀ə hə̀ən* plus a verb phrase, as depicted in (29). Concurrently, the function of *bə̀ə hə̀ən* as realis negator was established.

(29) Syntactic reanalysis:

negative experiential construction: *bə̀ə* [*hə̀ən* VP]

> realis negator: *bə̀ə hə̀ən* [VP]

Then, the negative existential construction taking a noun phrase (‘negative *bə̀ə* + existential verb *mīi* + NP’ meaning that ‘entity named by NP does not exist’) began to take a verb phrase instead of a noun phrase (*bə̀ə mīi* VP), which can be regarded as generalization of the construction as a result of analogy. In other words, it is a kind of construction-internal generalization resulting from the language users’ recognition of

⁵⁸ The notion ‘generalization of construction’ here roughly corresponds to what Himmelmann (2004: 32) calls ‘host-class expansion’, which is a context-expansion that construction-internally occurs in the process of grammaticalization (viz. the expansion of the class of elements with which a substantive grammaticalizing element is in construction).

structural similarity and semantic contiguity existing between *bə̀ə hə̀ə* and *bə̀ə mii*, or more specifically, their recognition that the form of *bə̀ə hə̀ə* is similar to the form of *bə̀ə mii* and that the meaning of *bə̀ə hə̀ə* is related (speciously contrary) to the meaning of *bə̀ə mii*. This change eventually gave birth to a new verbal negator *bə̀ə mii*, as shown in (30) below. The conventional meaning of irrealis negation associated with *bə̀ə mii* was promoted (i.e. the process of ‘emphasis’ in Croft’s terms) as the two contrastive negative constructions ‘*bə̀ə hə̀ə* VP’ and ‘*bə̀ə mii* VP’ became equally productive and entrenched.

(30) Generalization by analogy:

negative existential construction: *bə̀ə mii* NP

> irrealis negator: *bə̀ə mii* VP
cf. *bə̀ə hə̀ə* VP

Finally, the negators *bə̀ə mii* and *bə̀ə hə̀ə* underwent phonological reduction changing respectively into *mii* and *hə̀ə*, and the former further shortened yielding *míʔ*, as in (31) below. This phonological attrition was accompanied by semantic depletion (i.e. the process of ‘weakening’ in Croft’s terms), namely shift from specific to regular negator (shift from irrealis vs. realis negator to neutral negator).⁵⁹

(31) Phonological reduction:

a. irrealis negator: *bə̀ə mii* VP

> neutral negator: *mii* / *míʔ* VP

b. realis negator: *bə̀ə hə̀ə* VP

> neutral negator: *hə̀ə* VP

4 Conclusion

This study is a preliminary one intended to be an early step for future inquiry into the larger picture of the evolution of the Thai negative system. This work has paid close attention to the two now disused negators *bə̀ə mii* and *bə̀ə hə̀ə* and delved into their functions in the old negative system by analysing actual tokens of negative expressions with these negators occurring in the inscription corpus. It has offered a plausible hypothesis on their contrasting original meanings and correlated historical changes involving three major types of language change, namely, syntactic reanalysis, generalization by analogy, and phonological reduction. I hope that the ideas I have articulated in this paper can make a contribution not only to studies on the Thai

⁵⁹ The semantic depletion or weakening of these negators is equal to their ‘semantic-pragmatic context expansion’, which is the most important context-expansion involved in the process of grammaticalization (viz. the expansion of the semantic and pragmatic contexts in which the construction containing a grammaticalizing element is used) (Himmelman 2004: 33).

grammatical system for negation in particular but also to typological studies regarding historical linguistic change in general.

References

- Bradley, Dan Beach. 1873. *Dictionary of the Siamese Language*. Bangkok. [reprinted in 1971 by Kurusapha Press in Bangkok]
- Bybee, Joan L. and Suzanne Fleischman (eds.) 1995. Modality in grammar and discourse: An introductory essay. In Bybee, Joan L. and Suzanne Fleischman (eds.) *Modality in Grammar and Discourse*, 1-14. Amsterdam, John Benjamins.
- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago, The University of Chicago Press.
- Chung, Sandra and Alan Timberlake. 1985. Tense, aspect, and mood. In Shopen, Timothy (ed.) *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon*, 202-258. Cambridge, Cambridge University Press.
- Diller, Anthony V. N. 1988. Thai syntax and “national grammar”. *Language Sciences* 10.2:273-312.
- Diller, Anthony V. N. 2001. Grammaticalization and Thai syntactic change. In Tingsabadh, Kalaya and Arthur S. Abramson (eds.) *Essays in Thai Linguistics*, 139-175. Bangkok, Chulalongkorn University Press.
- Givón, Talmy. 1982. Evidentiality and epistemic space. *Studies in Language* 6.1:23-49.
- Givón, Talmy. 1994. Irrealis and the subjunctive. *Studies in Language* 18.2:265-337.
- Harris, Alice C. 2003. Cross-linguistic perspectives on syntactic change. In Joseph, Brian D. and Richard D. Janda (eds.) *The Handbook of Historical Linguistics*, 529-551. Malden, Blackwell.
- Himmelman, Nikolaus P. 2004. Lexicalization and grammaticalization: opposite or orthogonal? In Bisang, Walter, Nikolaus P. Himmelman and Björn Wiewer (eds.) *What Makes Grammaticalization?: A Look from its Fringes and its Components*, 21-42. Berlin, Mouton de Gruyter.
- Hooper, Joan B. 1975. On assertive predicates. In Kimball, John P. (ed.) *Syntax and Semantics* 4, 91-124. New York, Academic Press.
- Hopper, Paul J. 1991. On some principles of grammaticalization. In Traugott, Elizabeth C. and Bernd Heine (eds.) *Approaches to Grammaticalization* Vol.1, 17-35. Amsterdam, John Benjamins.
- Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, Cambridge University Press [2nd edition, 2003].
- Kullavanijaya, Pranee. 1996. Undoing homonymy: Cases in Debaio Zhuang and Thai. In Kullavanijaya, Pranee, Amara Prasithrathsint, and Suvanna Kriengkraipetch (eds.) *Collection of Papers on the Relationship between the Zhuang and the Thai*, 78-92. Bangkok, Chulalongkorn University [also included in *Proceedings of the 4th International Conference on Thai Studies, Kunming, Yunnan in May 1988*, 39-47].
- Langacker, Ronald. 1977. Syntactic reanalysis. In Li, Charles N. (ed.) *Mechanisms of Syntactic Change*, 57-139. Austin, University of Texas Press.

- Lehmann, Christian. 1995. *Thoughts on Grammaticalization* (revised and expanded version). München, Lincom Europa.
- Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, University of California Press.
- Lyons, John. 1977. *Semantics*, Vol.2. Cambridge, Cambridge University Press.
- Narrog, Heiko. 2005. On defining modality again. *Language Science* 27:165-192.
- Palmer, F. R. 2001. *Mood and Modality (Second edition)*. Cambridge, Cambridge University Press.
- Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Science*, 65-1. [reprinted in Bright William (ed.) 1992. *The Collected Works of Edward Sapir X: Southern Paiute and Ute linguistics and Ethnography*, 17-314. Berlin, Mouton de Gruyter.]
- Takahashi, Kiyoko. 2007. The irrealis marker in Thai. *Memoirs of Kanda University of International Studies* 19:189-210.
- Takahashi, Kiyoko. 2008. Negative markers in Thai. *Memoirs of Kanda University of International Studies* 20:335-358.
- Whorf, Benjamin L. 1950. An American Indian model of the universe. *International Journal of American Linguistics* 16:67-72.

<Corpus data>

- The Prime Minister's Secretariat, Thailand. 1924/1978. *Corpus of Thai Inscriptions, Installment 1*. Bangkok, The Prime Minister's Secretariat.
- The Prime Minister's Secretariat, Thailand. 1965. *Corpus of Thai Inscriptions, Installment 3*. Bangkok, The Prime Minister's Secretariat.
- The Prime Minister's Secretariat, Thailand. 1970. *Corpus of Thai Inscriptions, Installment 4*. Bangkok, The Prime Minister's Secretariat.
- The Prime Minister's Secretariat, Thailand. 1972. *Corpus of Thai Inscriptions, Installment 5*. Bangkok, The Prime Minister's Secretariat.
- The Prime Minister's Secretariat, Thailand. 1974. *Corpus of Thai Inscriptions, Installment 6.1*. Bangkok, The Prime Minister's Secretariat.
- The Prime Minister's Secretariat, Thailand. 1974. *Corpus of Thai Inscriptions, Installment 6.2*. Bangkok, The Prime Minister's Secretariat.
- Pongsripian, Winai (ed.). 1991. *Corpus of Thai Inscriptions, Installment 7*. Bangkok, The Prime Minister's Secretariat.

Appendix A

13 tokens of ‘*bəw mii* VP’ found in the inscription corpus

1. บ่มีคณนถ์เลย [2](1341-1367)
2. เขาอาราธนาพระบ่มีเสด็จลงมา [2](1341-1367)
3. บุษาทันคน ออกเข้าของทั้งหลายบ่มีว่าถ์เลย [2](1341-1367)
4. ...และบ่มีได้เซ็น [8](1369)
5. คานบ่มีได้มาแลพอย จึงให้ไปบ่าววงสรวงอาราธนาอัญเชิญพระมหาเถรเป็นเจ้าด้วยเคารพหนักหนาคาบหนึ่งโสด [62](1370)
6. พระมหาเถรเป็นเจ้ามาอยู่บ่มีนานเท่าใด ท่านจึงราพึงไปมาพิจารณา... [62](1370)
7. ...ว่าดี และบ่มีพิจารณา หาก... [45](1392)
8. ท่านไปกราบคนก็ดี ไปจอดเรือนก็ดี คล้อยผู้นั้นอุเบกษาและบ่มีเอาและละให้ปล่อยไป [38](1313-1433)
9. ผู้รักษาบ่มีแท้ [306](1466)
10. ผิบบ่มีกั้วลก็ดี ผิจักมีหิณูริพยานไควว่าเรามากท่านให้มาก เรามักน้อยท่านให้น้อย [64](15C)
11. สมบัติคนนองเนื่อง นามานูชาทุกป्ली เดือนมาบ่มีขาด [78](1796)
12. วัดบ่มีได้ออกสักแห่ง เท่าให้ออกเจ้าเมืองงาวผู้เดียวดาข [82](unknown)
13. บ่มีให้คลาดกลางกลางล แท้จริงแล [83](unknown)

Appendix B

6 tokens of ‘*bəw həwən* VP’ found in the inscription corpus

1. นบพ...ส...ร...ในราชมนเทียรบ่ห่อนขาดสักวันสักคืน...จ...เดือนเพ็ญ [3](1357)
2. ...ผู้ลักข้าวลักของบ่ห่อน... [3](1357)
3. ชื่อผู้ใดผิดร้าง . งาม เท่าใดก็ดี บ่ห่อนฆ่าฟัน...สักคาบ [5](1361)
4. และจักให้เถิงที่ลั้มที่ตายดังอันก็ดี บ่ห่อนฆ่าห่อนตีสักคาบ [5](1361)
5. เราบ่อวางไว้ของ เงินทองบ่ห่อนมี [160](1782-1925)
6. สีม่าบ่ห่อนปนเจือ [201](1925-1978)

***klah* IN CONTEMPORARY KHMER: QUANTITATIVE AND QUALITATIVE PLURALITY**

Joseph D. Thach
SEDYL, INALCO – CNRS – IRD
<joseph.thach@inalco.fr>

Denis Paillard
LLF, Paris Diderot University
<denis.paillard@linguist.jussieu.fr>

The present paper deals with *klah* in contemporary Khmer. In the existing grammars and manuals (such as Gorgonief (1966), Huffman (1970), Khin Sok (2000)), *klah* is described as an indefinite pronoun and often considered (at least in translations) as the equivalent of some in English. However, as we will demonstrate below – through a semantic analysis of its uses and values –, *klah* has its own syntactic and semantic values, distinct from those of some, and is a marker of plurality (it should be recalled that in Khmer the number category is not morphologically marked). For this reason, our study of *klah* will come within the framework of a wider discussion about the notion of plurality in Khmer (also linked to nominal and adjectival reduplication⁶⁰).

Our paper endeavors to describe all the uses and values of *klah* without any exclusion. Firstly, it deals with the cases where *klah* is combined with a (count or mass) N. For these two categories of N, we will show that *klah* comes under two interpretations: a. partition of all the instances of the N; b. the construction of a set of singular instances of the N. Secondly, we will study the cases where *klah* relates to a V and constructs various instances of this V. As regards reduplication, we will distinguish between two cases: a. *klah* in case of reduplication of the N; b. and *klah* reduplicated and introducing the notion of vagueness; compare:

- (1) *cam* *klah*
remember *klah*
“I clearly remember some of the things”
- (2) *cam* *klah* *klah*
remember *klah* *klah*
“I faintly remember some of the things/a few things now and then”

We will close our study with the analysis of cases where *klah* is combined with the indefinites/interrogatives *?ey* and *naa*. In this respect, we will point out a difference of behaviour. With *?ey*, *klah* can only stand in postposition and *?ey* is an interrogative⁶¹:

⁶⁰ On reduplication in keeping with plurality, see Paillard (2009)

⁶¹ About *naa* and *?ey* see Thach (2007).

Joseph Thach & Denis Paillard. 2010. *klah* in Contemporary Khmer: Qualitative and Quantitative Plurality. *Journal of the Southeast Asian Linguistics Society* 3.2:93-110

Copyright vested in the authors

Received 31/1/10, revised text accepted 30/9/10

- (3) *praacum prik mɨn kee niʔyiey pii riəŋ ʔey klah*
 meeting morning deict people talk about story ʔey klah
 “Which were the topics broached during this morning meeting?”

With *naa*, *klah* can stand either in ante-position (*naa* is an indefinite: example (4)) or in postposition (*naa* is an interrogative: example (5)):

- (4) *niw knoŋ p^huum nih klah naa ʔat teaŋ ʔaŋkaa*
 be bn village deict klah baa neg. part. rice
hoop p^haaŋ
 eat part.
 “In this village, some inhabitants (whom I cannot or don’t want to point out) don’t even have rice to eat”

- (5) *praacum prik mɨn kee niʔyiey pii riəŋ naa klah*
 meeting morning deict people talk about story naa klah
 “Among the topics on the agenda, which ones have actually been discussed this morning?”

The analysis of these various uses will allow us to verify our hypothesis on *klah* as a marker of quantitative and qualitative plurality.

The examples used in this paper were constructed by us (native speakers) or heard in daily conversations. All these examples were checked by ten native speakers of different social origins (students, teachers, actors, moto-taxi drivers).

In order to discuss the various uses of *klah*, we will use the notion of plurality, but with a meaning different from that usually given to this notion, especially when referring to the number category in languages where the opposition singular vs plural is morphologically marked. In their article entitled The semantics and pragmatics of plurals (2008), D. Farkas and H. de Swart give the following definition of the opposition between singular and plural: “singular nominals take values from the domain of atoms; plural nominals take values from the domain of sum”. In this view, the singular / plural opposition comes down to “atomic vs sum reference”, in other words to a purely quantitative matter. Following M. Jarrega works on the plural in French, a different characterization of plural can be given (Paillard, 2006) associating both quantitative and qualitative information, as shown by the following representation: $X (x_1 \dots x_k \dots x_n \dots)$

Quantitative component (Qnt). X corresponds to a given but not explicit quantity of instances of the category associated to the N under plural. In order to characterize this quantity, we will resort to the expression inclusive plural, adopted by several authors. Inclusive plural is neutral as regards the “atomic vs. sum reference” distinction (see - Do you have children? – Yes, one daughter);

Qualitative component (Qlt). The instances $x_1 \dots x_k \dots x_n$ of the category N are differentiated from each other, the criteria of differentiation being left unspecified (Lasnik, 1995).

This definition for plurality agrees with that given by R. Jackendoff (1991) as “an aggregate of distinguishable individuals” (aggregate: X, distinguishable individuals: xi xk xn).

Our hypothesis on *klah* is the following: *klah* is a marker of inclusive plurality actualizing both components: the quantitative (Qnt) and the qualitative (Qlt).

1. Ncount + *klah*

klah with a count noun as its scope comes under a partition principle: *klah* refers to a subset of instances of the category N, fitting a differential property. *klah* never reduces to refer to a plain sum. This can clearly be shown through a comparison between NPs formed by N + *klah* and those formed by N + *muey camnuən*:

- (6a) *məək* *cuəp* *kɲom* *muəy* *pʰlɛet* *baan* *tee* *kɲom* *mien*
 come meet 1sg. one moment possible part. 1sg. have
riəŋ *klah* *trəw* *niʔyiey*
 story klah must talk

“Could you come and see me for a short while? I have things to talk (to you)”

- (6b) *məək* *cuəp* *kɲom* *muəy* *pʰlɛet* *baan* *tee* *kɲom*
 come meet 1sg. one moment possible part. 1sg.
mien *riəŋ* *muəy-camnuən* *trəw* *niʔyiey*
 have story muəy-camnuən must talk

“Could you come and see me for a short while? I have a number (a series) of points to discuss”

- (6c) *məək* *cuəp* *kɲom* *muəy* *pʰlɛet* *baan* *tee* *kɲom* *mien*
 come meet 1sg. one moment possible part. 1sg. have
riəŋ *Ø* *trəw* *niʔyiey*
 story Ø must talk

“Could you come and see me for a while? I have one / several points to discuss”

In (6a) N + *klah* introduces some instances of the N as they fit a differential property which is not made explicit: the points to be discussed are identified for the speaker.

In (6b) N + *muəy-camnuən* refers to a sum of x (quantificational reference), out of all qualification. In (6c) N + Ø refers to one or several instances of the category N (leaving it to the context to make the difference).

In some cases, *klah* is impossible, and only Ø and *muəy-camnuən* can be used:

- (7a) *siewpʰiw* *muey-camnuən* *nih* *trəw* *yəək* *tʰw* *ʔaoy* *rooŋ-pum* *wɨŋ*
 book muey-camnuən deict. must take go give printing-house part.

“This set of books must be taken back to the printing house !”

- (7b) *siewpʰiw* *Ø* *nih* *trəw* *yəək* *tʰw* *ʔaoy* *rooŋ-pum* *wɨŋ*
 book Ø deict. must take go give printing-house part.

“This / these book(s) books must be taken back to the printing house !”

(7c) **siewp^hiw* *klah* *nih* *trəw* *yɔɔk* *tɪw* *ʔaoy* *rooŋ-pum* *wɪŋ*
 book klah deict. must take go give printing-house part.

In (7) the impossibility of *klah* means that the books in question are copies of one and the same book, which precludes any qualitative differentiation of those books.

Contrary to this case, examples can be met where *klah* only possible:

(8) Context: a police inspector (S1) tries to get the manager (S0) of a restaurant tell him if the wanted person comes to eat daily in the restaurant:

S1: *koat* *mɔɔk* *ɲam* *baay* *nɪŋ* *roal* *tɲay* *rɪi* *yaŋmec*
 3sg come eat rice deict. every day or how
 “Does he come to eat here every day?”

S0: *mɪn* *tiɛŋ* *tee* *tɲay* *klah* *mɔɔk* *tɲay* *klah* *ʔat*
 neg. regular part. day klah come day klah neg.
 “It is not regular, some days he comes, and others he doesn’t”

In (8) *muey-camnuən* and Ø cannot be used. Considering the whole set of the days, *klah1* and *klah2* work out two subsets of days, the first one fitting “come”, and the second one “not come”.

(8bis) - At the market place, a customer (S1) and a seller (S0).

S1: *ʔaoy* *swaay* *kɲom* *muey* *kiiloo* *mɔɔk*
 give mango 1sg. one kilo come
 “Give me one kilo of mangoes!”

S0: *yɔɔk* *swaay* *praap^hɛet* *naa*
 take mango sort naa

pruəh *swaay* *klah* *cuu* *swaay* *klah* *pʔaem* *haey* *swaay*
 because mango klah acid mango klah sweet and mango

klah *tiet* *cuu-ʔaem*
 klah furthermore sweet-and-sour
 “What kind of mangoes do you want? For there are acid mangoes, sweet mangoes and the rest, sweet-and-sour mangoes”.

In (8bis), the mangoes as a whole are divided into three subsets, each subset fitting a differential property: acid, sweet and sweet-and-sour.

(9) - To the question “in which group are there good students?”, S0 answers:

- (9a) *knor* *krom* *nih* *mien* *sih* *Ø* *rien* *puukae* *krom*
in group deict. have student Ø learn gifted group
pseep *tiet* *ʔat* *mien* *tee*
other more neg. have part.

“In this group, there are good students, in others there aren’t any”

- (9b) *knor* *krom* *nih* *mien* *sih* *klah* *rien* *puukae* *krom*
in group deict. have student klah learn gifted group
pseep *tiet* *ʔat* *mien* *tee*
other more neg. have part.

“In this group, there are some good students, in others there aren’t any”

(9a) with Ø plainly states the existence of good students, whereas (9b) with *klah* states that there is a given number of students different from the others as they fit the property “be good” in their studies.

The series (10a-c) confirms that N Ø, contrary to N + *klah*, cannot refer to a partition on a whole set contextually introduced. Example (10c) is possible owing to the fact that a subset is built through an independent source of determination, making it possible to identify a group of NGOs.

- (10a) *niw* *srok* *kmae* *mien* *ʔaŋkaa* *craən* *mɛən* *tae* *ʔaŋkaa*
in country khmer have organisation much be-true only organisation
klah *kɪt* *tae* *pɪi* *rɔək* *luy* *tee*
klah think only of look for money part.

“It’s true that in Cambodia, there are many NGOs, but some (of them) are there only to make money”

- (10b) **niw* *srok* *kmae* *mien* *ʔaŋkaa* *craən* *mɛən* *tae* *ʔaŋkaa*
in country khmer have organisation much be-true only organisation
 \emptyset *kɪt* *tae* *pɪi* *rɔək* *luy* *tee*
 \emptyset think only of look for money part.

- (10c) *niw* *srok* *kmae* *mien* *ʔaŋkaa* *craən* *mɛən* *tae*
in country Khmer have organisation much be-true only
ʔaŋkaa *ʔoostraalii* *kɪt* *tae* *pɪi* *rɔək* *luy* *tee*
organisation Australia think only of look for money part.

“It’s true that in Cambodia, there are many NGOs, but the Australian ones are there only to make money”

In all the above examples, *klah* marks a partition: a differential property is used to identify a group of instances of the N. But the Ncount + *klah* pattern does not necessarily refer to a partition: N+*klah* can refer to a series of instances differentiated from each other. Whereas in the case of the “partition” interpretation, the property is used to distinguish a

subset of instances of the N, in the second case what is emphasized is the fact that each one of the instances in the set has an identity. (6a) is a first example coming under this second interpretation: the different points to be mentioned by the speaker are not presented as a whole and in his view, each one of the points is important

- (6a) *məək* *cuəp* *kɲom* *muəy* *pʰlɛət* *Baan* *tee* *kɲom* *mien*
 come meet 1sg. one moment possible part. 1sg. have
riəŋ *klah* *trəw* *niʔyiey*
 story klah must talk
 “Could you come and see me for a short while? I have things to talk (to you)”

Other examples can be given:

- (11a) *knəŋ* *krom* *nih* *kɲom* *dəŋ* *tʰaa* *mien* *neak* *klah*
 in group deict. 1sg know that have people klah
mɪn *coolcət* *kɲom* *tee*
 neg. like 1sg part.
 “In this group, I know that some people don't like me!”

- (11b) *knəŋ* *krom* *nih* *kɲom* *dəŋ* *tʰaa* *mien* *neak* *Ø*
 in group deict. 1sg know that have people Ø
mɪn *coolcət* *kɲom* *tee*
 neg. like 1sg part.
 “In this group, I know that there are people who don't like me!”

Example (11b) with N + Ø states the existence of one or several unfriendly people; in (11a) the speaker knows who is /are the one (s) in question but does not want to name him / them (let's point out the inclusive interpretation of *klah*: one or several people can be concerned).

- (12) *saalaa* *knəŋ* *kron* *kroan* *mien* *sɪh* *rien* *puukae* *klah*
 school in city enough have student learn be gifted klah
baə *saalaa* *taam* *pʰuum* *wɛŋ* *ʔat* *mien* *sah* *tae*
 if school follow village part. neg. have part. only
mədaaŋ
 once
 “In city schools, it is possible to find some good students, but in rural schools, there are none at all”

In (12) contrary to (9b), *klah* comes after the property “be good”: this property does not work here as the basis for differentiating a subset. In this position, *klah* just means that there are good students, each student fitting specifically the property “be good”. The same is to be found in the series of examples (13a – c) under the form of questions. The context

is the following: an organisation must make an assessment in schools in order to grant scholarships to the students. The assessor asks:

- (13a) *nɨw* *saalaa* *nih* *mien* *sɨh* *puukae* *Ø* *tee*
 at school deict. have student be-gifted Ø part.
baə *saalaa* *taam* *p^huum* *wɛɲ* *ʔat* *mien* *sah*
 if school follow village part. neg. have part.
mədaaŋ
 once
 “Are there good students in this school?”

- (13b) *nɨw* *saalaa* *nih* *mien* *sɨh* *puukae* *klah* *tee*
 at school deict. have student be-gifted klah part.
baə *saalaa* *taam* *p^huum* *wɛɲ* *ʔat* *mien* *sah*
 if school follow village part. neg. have part.
mədaaŋ
 once
 “Are there (even so) good students?”

- (13c) *nɨw* *saalaa* *nih* *mien* *sɨh* *klah* *puukae* *tee*
 at school deict. have student be-gifted klah part.
baə *saalaa* *taam* *p^huum* *wɛɲ* *ʔat* *mien* *sah*
 if school follow village part. neg. have part.
mədaaŋ
 once
 “Are there some students who are good (students)?”

In (13a), the speaker has no prejudice: the question is whether there are (or not) good students. In (13b) with *klah*, the question is not about the existence (or not) of good students but is about the relevance of the property “be good” in order to qualify some of the students, which implies that for the speaker, it is not obvious that there are good students at all. As in (12), *klah* comes after the property ‘be good’. In (13c) we have a different order (N + *klah* + property): the speaker wants to know if there are students who are different from other students by “being good in school”.

In short, when the scope of *klah* is a count noun, it has two interpretations according to the context, corresponding to a ‘weighting’ either on the component X, or on the component (xi xk xn ...): in one case it stands for the partition of a set based on a differential property; in the other, it introduces a series of instances of the category N taken in a qualitative variation.

2. Mass noun as the scope for *klah*

Combined with a mass noun *klah* has two interpretations:

- (14) *tik* *klah* *ɲam* *tɨw* *c^hɨ* *puəh*
 water klah drink go suffer stomach

“Some brands of water cause stomach ache when you drink it”

In (14), *klah* can switch with *muəy-camnuən*. *klah* expresses a partition in the whole of the water brands. Owing to mass nouns properties, the series coming under the partition operated by *klah* is formed by the whole of the brands of bottled water.

- (15) *kom* *teɲ* *tae* *tik-p^hlae-c^həə* *teɲ* *sraa* *klah* *mɔək*
 neg. buy only fruit-juice buy alcohol klah come
 “Don’t buy only fruit juice, buy a certain amount of alcohol (leaving the quantity up to you)!”

In (15) *klah* can switch with *bantec* ‘a little’. *bantec* stands only for a small quantity, whereas *klah* refers to a given quantity, i.e. a quantity which is qualitatively defined: in the case of (15), this corresponds to the quantity the interlocutor will deem suitable as regards the number of people attending the party. This quantity is distinguished in the set of the possible quantities of alcohol.

As R. Jackendoff writes about mass nouns: “With a mass noun like water, one can divide its referent and still get something describable as water. For this second interpretation of *klah*, we put forward the hypothesis that *klah* marks a fragmentation of the mass noun, which means that in X (xi xk xn ...) xi xk xn ... each one of the x corresponds to a specific quantity (or portions) of alcohol. As for X, it refers to a set of *a priori* possible quantities

We now give one more example of *klah* marking a quantitative fragmentation of the mass noun (in (16) *bantec* is possible).

- (16) *baə* *ʔat* *məhoop* *ɲam* *yɔək* *treyɲiet* *klah* *tɨw*
 if neg. side-dish eat take dried-fish klah go
 ɲam *tɨw*
 eat go

“If you have no side dishes, take some fair quantity of dried fish to eat.”

In (16) the quantity of dried fish is not just any quantity: take whatever you need, but leave some for me as well (sharing the quantity)

In short, with the mass nouns the two interpretations already identified for the count nouns are to be met again:

a partition which, owing to the mass nouns properties, distinguishes a subset X (xi xk xn ...) of varieties of N; this subset is based on a differential property. The focus is on X, and the inner subset differences are not taken into account

the fragmentation of the N in a series of distinct quantities X (xi xk xn ...); comparing (17a) with Ø / (17b) with *klah* shows that the fragmentation of

the N results in given specific quantities: the focus is on the series (xi xk xn ...).

3. *klah* as a pronoun

We will speak of *klah* as a pronoun when its scope is not an N, whether it is contextually given (ex. (17)) or *klah* corresponds to the predicate internal object (ex. (18)):

(17) - context: S0 gave some honey to S1, and one week later, S0 asks S1 :

S0 : *tik-kmum* *knom* *ʔaoy* *tɨw* *ʔah* *haəy* *nɨw*
 honey 1sg give go finish already yet

“The honey I gave you, did you finish it or is there some left?”

S1 : *ʔah* *haəy*
 finish already
 “I have already finished it”

(17a) S0 : *caŋ* *baan* *Ø* *tiet* *tee*
 want obtain Ø more part.
 “Do you want some more?”

(17b) S0 : *caŋ* *baan* *klah* *tiet* *tee*
 want obtain klah more part.
 “Do you want some extra?”

In (17), the presence of *tiet* ‘extra, more’ is due to the fact that a first quantity of honey has already been given. In (17a), the question is just about “wanting some more honey”, without the extra quantity being determined. In (17b), the question is prejudiced: S0 is ready to give an extra quantity, but not just any quantity. It depends on what S0 can / or wants to give on a second time (the new quantity being less than the first one).

(18) The topic is about S0 having old recollections that S1 would like to know:

(18a) *cam* *klah* *ʔat* *klah*
 remember klah neg. klah
 “I remember some of the things but not others”

(18b) *cam* *klah*
 remember klah
 “‘There are things I remember quite well I can tell you.”

With the pronoun, we find again the two interpretations previously brought out: partition in (18a), differentiated instances of the category N in (17b) and (18b). Comparing (19a) with (19b) shows that when *klah* refers to a N standing in the left context, *klah* is compulsory to mark a partition.

- (19a) *swaay baan p^hlae craən mēen kɲom som klah mɔɔk*
 mango 2sg fruit much be-true 1sg ask klah come
 “Your mango tree bears actually a lot of fruits, may I ask you for some?”

- (19b) *?swaay baan p^hlae craən mēen kɲom som Ø mɔɔk*
 mango 2sg fruit much be-true 1sg ask Ø come

In (19b) the only possible interpretation is that the request involves the mango tree proper and not part of its fruit.

Directly related to this second interpretation, it must be noted that there are cases when it's the process itself which stands as *klah* scope (in such a case, *bantec* is possible as well):

(20) - In a meeting, the same person (S1) keeps speaking all the time, which irritates S0 who therefore stops him:

- (20a) *sɲat moat Ø tiw*
 quiet mouth Ø go
 “Shut up!”

- (20b) *sɲat moat Ø tiw / tuk ?aoy kee ni?yiey p^haan*
 quiet mouth Ø go let give people speak part.
 “Shut up for a while, let the others speak!”

- (20c) *sɲat moat bantec tiw / tuk ?aoy kee ni?yiey p^haan*
 quiet mouth little go let give people speak part.
 “Just shut up a little [...] !”

In this series (20), *sɲat moat* means ‘shut up’. *moat* which can often be translated by ‘mouth’ is part of the predicate. In (20a), with the form *Ø*, the injunction takes a sharply categorical tone I want you to shut up right now. In (20b), *klah* stands for the construction of one or several instances of the process: all I’m asking you is shut up a little / from time to time. In (20c) with *bantec* the request is limited to one time for a limited period quantitatively defined. The principle here is quite similar to that working with the mass nouns: *klah* can be interpreted as referring to a series of occurrences of the V. In accordance with the notion of inclusive plurality, the actualization may concern one or several occurrences of the V, each occurrence being differentiated.

Here are two other examples:

- (21a) *baə caŋ baan sok-p^hiep l?aa kom cih tae laan*
 if want obtain health good neg. take only car
daə klah tiw
 walk klah go
 “If you want to get in good health, don't take the car, just walk a little (reasonably)”

- (21b) *baə* *caŋ* *baan* *sok-p^hiep* *lʔaa* *kom* *cih* *tae* *laan*
 if want obtain health good neg. take only car
daə *bantec* *tɨw*
 walk little go

“If you want to get in good health, don't just take the car, walk a little / from time to time” (for example when you go at work)

In short, with *klah* as a pronoun, the two interpretations appearing with the count nouns and the mass nouns (partition and series of occurrences qualitatively differentiated) are available. On the other hand, when *klah* has the process for its scope, the second interpretation is the only possible one (with a V there is no available set of Vs)

3. *klah* and the reduplication

As regards reduplication, we will study two different points:

- *klah* in case of reduplication of the N;
- *klah* reduplicated.

In Paillard (2009), we have shown that the reduplication of the N puts N in a qualitative variation out of any quantitative variation: each one of a series of instances (xi xk xn ...) of the category N specifically fits the predicate of which it is an argument (or the qualifying property): NN p means....xi - pi.... xk - pk xn - pn.... .

It appears that the reduplication on the one hand and *klah* on the other have a quite similar semantic value: both involve a series of qualitatively differentiated instances. We hereafter take a series where for a given N, we consider all the possibilities for its determinations: N-O, NN, N-*klah*, NN-*klah*

3.1. *klah* in case of reduplication of the N

(22a) - A popular TV program presenter (S1) goes to the provinces in order to recruit girls for his program. On his arrival in a village, he asks a local inhabitant (S0):

- S1: *nɨw* *p^huum* *nih* *mien* *srey* *sʔaat* *tee*
 in village deict. have girl beautiful part.
 “In this village, are there beautiful girls?”
- S0: *nɨw* *p^huum* *nih* *mien* *srey* \emptyset *sʔaat* *srey* \emptyset
 in village deict. have girl \emptyset beautiful girl \emptyset
ʔat *sʔaat* *ʔat* *dooc* *p^huum* *nuh* *tee* *srey-srey*⁶² *sʔaat*
 neg. beautiful neg. like village deict. part. girl-girl beauti.
taeŋ-ʔah
 all

⁶² In all examples of this 29's series, reduplication of “girl” in second sequence of the sentence is required.

“In this village, some are beautiful ones and some are not, unlike in the village over there where all the girls are beautiful (every girl is beautiful)”

In (22a) the sequence

<i>srey</i>	Ø	<i>sʔaat</i>	<i>srey</i>	Ø	<i>ʔat</i>	<i>sʔaat</i>
girl	Ø	beautiful	girl	Ø	neg.	beautiful

can be replaced by the following sequences:

(b)

<i>srey</i>	<i>srey</i>	Ø	<i>sʔaat</i>	<i>srey</i>	<i>srey</i>	Ø	<i>ʔat</i>	<i>sʔaat</i>
girl	girl	Ø	beautiful	girl	girl	Ø	neg.	beautiful

(c)

<i>srey</i>	<i>klah</i>	<i>sʔaat</i>	<i>srey</i>	<i>klah</i>	<i>ʔat</i>	<i>sʔaat</i>
girl	klah	beautiful	girl	klah	neg.	beautiful

(d)

<i>srey</i>	<i>srey</i>	<i>klah</i>	<i>sʔaat</i>	<i>srey</i>	<i>srey</i>	<i>klah</i>	<i>ʔat</i>	<i>sʔaat</i>
girl	girl	klah	beautiful	girl	girl	klah	neg.	beautiful

(e)

<i>srey</i>	<i>sʔaat</i>	<i>klah</i>	<i>srey</i>	<i>ʔat</i>	<i>sʔaat</i>	<i>klah</i>
girl	beautiful	klah	girl	neg.	beautiful	klah

(f)

<i>srey</i>	<i>srey</i>	<i>sʔaat</i>	<i>klah</i>	<i>srey</i>	<i>srey</i>	<i>ʔat</i>	<i>sʔaat</i>	<i>klah</i>
girl	girl	beautiful	klah	girl	girl	neg.	beautiful	klah

In (a) N Ø property p (‘beautiful’) N Ø property p’ (‘not beautiful’) corresponds to the construction of two subsets of girls in relation with the property p :

The two subsets do not exhaust the whole set, but nothing is said about their number (the "beautiful-girls" subset can count for one single unit).

In (b) N N p (‘beautiful’) N N p’ (‘not beautiful’) according to our hypothesis on the reduplication of the N corresponds to the construction of two series of N in relation with p and p’, each occurrence (individual) of the two series specifically fitting p (or p’). Contrary to (a) the two series are not in a contrastive relation and those two series do not exhaust the whole set of the village girls (some of them are not taken into account as regards the property p).

In (c) N *klah1*-p (‘beautiful’) N *klah2*- p’ (‘not beautiful’) *klah1* and *klah2* correspond to the construction of two subsets of N (in relation with the property p or p’), each instance of the two series are considered as individuals. As in (b) those two subsets don’t exhaust the whole of the village girls.

In (d) N N *klah1*-p (‘beautiful’) N N *klah2*-p’ (‘not beautiful’), two series of differentiated N are at first introduced. In those series, *klah1* et *klah2* distinguish each one for its part a group of girls (previously differentiated), fitting or not the property p. The co-presence of the reduplication of the N and of *klah* reinforces the differentiation between the

individuals in each series. As in (b) and (c) those two subsets don't exhaust the whole of the village girls.

In (c) and (d) *klah* is placed before *p* (and *p'*). In (e) and (f) *klah* is placed after *p* (and *p'*). As was noted before through the examples (12) and (13b), in the second case, the partition is not based on the property *p* (or *p'*).

In (e) *N p* ('beautiful') *klah1 N p* ('not beautiful') *klah2*, *klah1* and *klah2* stand for the construction of two subsets (partition of the set *N*) – this partition is not based on the property *p* (or *p'*).

In (f), *N N p* ('beautiful') *klah1 N N neg p* ('beautiful') *klah2* two series of differentiated *N* are at first introduced and in those series, *klah1* et *klah2* distinguish, each one for its part a group of girls previously differentiated as fitting respectively the properties *p* and *p'*. The distinction between those two groups is not based on the property *p* (or *p'*).

The combinations illustrated by (22d- f) show the modes of interaction between the qualitative plurality (reduplication) and the <quantitative and qualitative> plurality, where both quantitative and qualitative components of the plurality get actualized.

3.2. *klah* reduplicated

As a noun determiner and as an indefinite pronoun as well, *klah* can be reduplicated (but it must be noted that it's impossible to have at the same time reduplication of the *N* and reduplication of *klah*: * *NN klah klah*). As regards the above examples (22 d, f) where the *N klah* has for its scope is reduplicated, we have put forward our hypothesis on the *N* reduplication: each instance of the category *N* specifically fits the predicate. An extension of this hypothesis to the reduplication of *klah* can be validated. In relation to the process, *klah1* and *klah2* construct, each for its part, a qualitatively distinct subset, entailing an effect of vagueness (no way to centre/focus on a given subset). 'Vagueness' is one possible interpretation of reduplicated *N* or Adjective; see:

(23) A client giving indications to the taxi driver:

pɛɛl dal rɔɔbaaŋ kʰiəw kʰiəw nuh som cʰɔp
when arrive fence blue blue deict. please stop

"When you arrive at the blue-like fence, please stop there"

(24) The manager (S1) of a show asks the producer (S0):

rɪəŋ dæl trəw leɛp cap-pʰdaəm hat haəy nɪw
story rel. must play start rehearse already yet

"About the play you are to perform, have you already started the rehearsals?"

(24a)

S0: *cap-pʰdaəm hat klah haəy*
start rehearse klah already

"We have started some sketches"

(24b)

S0: *cap-p^hdaəm hat klah klah haəy*
 start rehearse klah klah already
 “We have started some sketches here and then (nothing serious)”

(25)

cam rɨəŋ klah klah
 remember story klah klah
 “I faintly remember some of the things / some episodes now and then”

(26)

S1: *ʔah baay ɲam haəy nɨw*
 finish rice eat already reater
 “Is there some rice left?”

(26a)

S0: *sal klah dae ʔaac kroan samrap pii neak*
 to be left klah part. can enough for two person
 “There is some rice left (a small amount), it can do for two people”

(26b)

S0: *sal klah klah dae ʔaac kroan samrap pii neak*
 to be left klah klah part. can enough for two pers.
 “There is just a very little left, it’s just enough for two people”

In (24b) and (25) the reduplication of *klah* gives rise to a feeling of vagueness: there is no construction of qualitatively distinct subset. It is therefore impossible to identify the rehearsed sketches (ex. 24b) or the actualized recollections (ex. (25)). In (26b) *klah1* and *klah2* construct two distinct quantities of rice left, which stops from assessing the exact quantity of rice left, therefore interpreted as (very) insufficient.

The working of *klah* when reduplicated is similar to that of *klah* not reduplicated with the construction of two differentiated series (qualitative partition). The criterion liable to make an occurrence belong to such or such series not being specified, this co-presence of two series results in a kind of interference, with a depreciating effect on the occurrences involved, owing to the fact that although differentiated, they are not taken as full individuals.

4. *klah* combining with the interrogatives / indefinites *ʔey* and *naa*.

ʔey is a free choice type of indefinite and an interrogative. As an indefinite–interrogative *naa* means that the items previously made out are related to an undifferentiated set through the introduction of a new property: $x_1 \dots x_k \dots x_n \rightarrow (x(i) \dots x(k) \dots x(n) \dots)$ set (see Thach (2007)).

Whereas with *ʔey*, *klah* can only stand in postposition, with *naa*, *klah* can stand before *naa* (*naa* is an indefinite) or after *naa* (*naa* is an interrogative) (note that *muəy* ‘one’ shows the same distribution as *klah*. On this point, see Thach (2007)).

4.1. *klah* before *naa*

- (27) *kee* *t^haa* *kɔn* *baaraŋ* *lʔaa* *məəl* *mɛɛn* *tee*
 people say picture French good look be-true part.
 “It is said that French films are good, is that true?”

- (27a) *min* *teaŋʔah* *tee* *k^hah* *min* *lʔaa* *dae* *nɨŋ*
 neg. all part. klah neg. good also deict.
 “Not all of them, some (a fixed but not determined quantity) are bad (I can make a list)”

- (27b) *min* *teaŋʔah* *tee* *klah* *naa* *min* *lʔaa* *dae* *nɨŋ*
 neg. all part. klah naa neg. good also deict.
 “Not all of them, some are bad, but I can't tell which ones”.

Note that *naa* alone is not possible in this example.

In (27a) *klah* marks a partition on the set of the French films and constructs the subset of films fitting the property “be bad”. It refers to instances of films which can be identified, but that the speaker cannot list.

In (27b) the presence of *naa* after *klah* means that the films of the subset of distinctive instances constructed by *klah* reduces to an undifferentiated subset: the speaker is unable to identify them; *naa* makes uncertain the previous distinction in the instances of bad films:

Another example:

- (28a) *niw* *knɔŋ* *p^huum* *nih* *klah* *ʔat* *teaŋ* *ʔaŋkaa* *hoop* *p^haaŋ*
 be in village déict. klah neg. part. rice eat part.
 “In this village, some inhabitants (I can identify them), don't even have rice to eat”

- (28b) *niw* *knɔŋ* *p^huum* *nih* *klah* *naa* *ʔat* *teaŋ*
 be in village deict klah naa neg. part.
 ʔaŋkaa *hoop* *p^haaŋ*
 rice eat part.

“In this village, some inhabitants (I can't or don't want to identify them), don't even have rice to eat.”

In (28) as well *naa* neutralizes the differentiation between the occurrences of the N operated by *klah*: *naa* concerns only the qualitative side of *klah*, suspending the prior differentiation. The *klah naa* combination can be represented as follows :

$$\begin{array}{ccc} X \text{ (xi xk xn ...)} & \rightarrow & X \text{ (x(i) x(k) x(n) ...)} \\ \textit{klah} & & \textit{naa} \end{array}$$

4.2 *klah* after *naa* and *?ey*

- (29a) *lən̩jɛc* *n̩ɿŋ* *caŋ* *ɲam* *məhoop* *?ey*
 evening deict. want eat side-dish ?ey
 “Which dishes do you want to eat tonight?” (What do you eat for dinner?)

- (29b) *lən̩jɛc* *n̩ɿŋ* *caŋ* *ɲam* *məhoop* *?ey* *klah*
 evening deict. want eat side-dish ?ey klah
 “Which dishes do you want to eat tonight (there must be some you like best)?”

- (29c) *lən̩jɛc* *n̩ɿŋ* *caŋ* *ɲam* *məhoop* *naa*
 evening deict. want eat side-dish naa
 “Which dishes (out of this list) you want to eat tonight?”

- (29d) *lən̩jɛc* *n̩ɿŋ* *caŋ* *ɲam* *məhoop* *naa* *klah*
 evening deict. want eat side-dish naa klah
 “Among all the dishes on this list, what are those you want to eat tonight?”

(29a-d) are questions (note that in this case *?ey* and *naa* alone are possible). Question (29a), with *?ey* alone, is an open question: *?ey* points the whole of the possible and conceivable dishes. In (29b) *klah* following *?ey* means that for S0 all the dishes of the set are not on the same level for S1 and that the question aims at identifying a subset of dishes fitting the property ‘be preferred by S1’. The difference between (29c) with *naa* and (29a) with *?ey* is due to the fact that at first with *naa* the set is not that of all the possible and conceivable dishes, but a set of dishes on a menu. The question means that S0 doesn’t know which dishes on the list S1 does prefer. We shift from a set of identified dishes to a set of unidentified dishes. In (29d) as well as in (29b), S0 asks S1 to identify a subset of dishes as those S1 wants to eat on the evening.

Another example:

(30) - S0 who didn't attend the meeting asks S1:

- (30a) *pracum* *prɪk* *mɿŋ* *kee* *ni?yɛy* *pɿi* *rɪəŋ* *?ey*
 meeting morning deict. people speak about story ?ey
 “Which were the topics broached in this morning meeting?”

- (30b) *pracum* *prɪk* *mɿŋ* *kee* *ni?yɛy* *pɿi* *rɪəŋ* *?ey* *klah*
 meeting morning deict. people speak about story ?ey klah
 “What were the topics broached in this morning meeting? (I know nothing of the topics liable to be discussed, but since a meeting took place, I take it for granted that some subject(s) were broached)”

- (30c) *pracum* *prɪk* *mɿŋ* *kee* *ni?yɛy* *pɿi* *rɪəŋ* *naa*
 meeting morning deict. people speak about story naa
 “What were the topics on the agenda discussed this morning?”

- (30d) *pracum prik mɨn kee niʔyiey pii riəŋ naa klah*
 meeting morning deict. people speak about story naa klah
 “Among the topics on the agenda, which ones were actually discussed this morning?”

Placed after *ʔey* and *naa*, *klah* introduces a qualitative partition on the set established by *ʔey* and *naa*. This qualitative partition leads to an heterogeneity of the set, which is virtual. Since it comes through a question, *ʔey* and *naa* as question markers are in the locutor's area and *klah* in the interlocutor's. The combination between *klah* with *ʔey* and *naa* can be summed up as follows:

ʔey: (29a)

(xi xk xn ...) set
 S0 (S1 ?)

ʔey klah: (29b)

(xi xk xn ...) set → X (xi ... xn ...) subset
ʔey (S0) klah (S1 ?)

naa: (29c)

xi xk xn → (x(i) x(k) x(n) ...) set
 list S0 (S1 ?)

naa klah: (29d)

xi xk xn → (x(i) x(k) x(n) ...) set → X (xi ... xn ...)subset
 list *naa* (S0) *klah* (S1 ?)

Conclusion

We claim that *klah* associates both a quantitative and a qualitative information, as shown by the following representation:

X (xi xk xn ...)

This characterization is at work in all its various uses and values. If *klah* can be considered as the main marker for plurality in Khmer, it must be pointed out that the qualitative component is nonetheless crucial, as regards the notion of partition as well as the construction of differentiated instances of the N

Abbreviation

deict.:	Deixis, or demonstrative
neg.:	Negation
part.	Particles
1sg.	1st person of singular personal pronoun
2sg.	2nd person of singular personal pronoun
3sg	3rd person of singular personal pronoun
S0 / S1	Speaker/Addressee

References

- Corbett, G. (2000) *Number*, Cambridge University Press, Cambridge
- Gorgonev, Ju. (1966) *Grammatika khmerskogo jazyka*, Moskva.
- Farkas, D. & De Swart, H. (2009) “The semantics and pragmatics of plurals”, ms., Santa Cruz/Utrecht
- Haspelmath M. (1996) *Indefinite pronouns*, Oxford University Press
- Huffman F. (1970) *Modern spoken Cambodian*, Yale University Press, New Haven and London.
- Jackendoff, R. (1991) “Parts and boundaries”, *Cognition*, 41, 9 – 45
- Jarrega, M. (2000) *Le rôle du pluriel dans la construction du sens des syntagmes nominaux en français contemporain*, Thèse de doctorat, Université Paris X – Nanterre.
- Khin Sok (2001) *Grammaire du khmer*, Editions You Feng, Paris.
- Lasersohn, P. (1995) *Plurality, conjunctions and events*, Kluwer, Dordrecht
- Paillard, D. (2006) «Quelque N / quelques N», in: Corblin, Ferrando, Kupfermann (éds) *Indéfinis et prédication*, PU de la Sorbonne, 417 – 428.
- Paillard, D. (2009) «Réduplication du nom et de l’adjectif en khmer», in: S. Osu, G. Col, N. Garric & F. Toupin (éds) *Construction d’identité et processus d’identification*, Peter Lang, Bern, 569 – 586.
- Thach, J. D. (2007) *L’indéfinition en khmer: du groupe nominal au discours*, Thèse de doctorat, INALCO.
- Thach, J. D. (2009) «naa et ʔey: Indéfini – interrogatif en khmer. Deux différentes formes d’indéfinition», *Faits de Langues, Les Cahiers*, n° 1, 2009, Paris, 119-159.
- Vogeleer, S. & Tasmowski, L. (2006) Non definitness and plurality, *Linguistics Today* 95, John Benjamins Publishing Company.

CHILD ACQUISITION OF VIETNAMESE CLASSIFIER PHRASES

Jennie Tran
University of Hawaii
<jennietr@hawaii.edu>

1. Introduction

Vietnamese is an isolating language with lexical tones and monosyllabic word structure. Verbs have no tense or person/number agreement markings. Nouns do not have any morphological inflections for case, gender, person or number. Compared to a language like German, which has tense, person/number agreement markings on verbs, and case, gender, person, and number markings on nouns, Vietnamese seems easier for foreigners to learn and perhaps simpler for children to acquire. However, this is not necessarily true because unlike German and other European languages, Vietnamese is one of several Asian languages with a complex numeral classifier system. This system is not easy for foreigners to grasp and not easy for children to acquire. This is due to the large number of classifiers (over 200), the complex semantic nature of classifiers, and the fixed order of elements in classifier phrases. Classifiers may be particularly difficult to learn in the context of Vietnamese noun phrases, which are not exclusively head-final, but have both postnominal and prenominal modifying elements: Demonstratives, *wh*-words, adjectives, possessive pronouns and relative clauses follow the noun, whereas numerals and classifiers precede the noun. The word order of a full five-element classifier phrase is [numeral/quantifier + classifier + noun + adjective + demonstrative/*wh*-word/ possessive/ relative clause].

From a semantic perspective, classifiers are unbound function words that categorize the head noun based on inherent or salient features of the noun's referent, such as animacy, shape, length, dimension, function, or material. From a syntactic perspective, numeral classifiers can be defined based on three criteria: First, they are obligatory in noun phrases containing a numeral, a demonstrative, an interrogative, or a combination of these elements, with or without an overt head noun, as shown in (1).⁶³ Second, they can be repeated, as is typical in Thai and Burmese (2), or can be reduplicated, as is unique to Cantonese (3).

- (1) hai con (chó) này (Vietnamese)
two CL (dog) this
'these two dogs'

- (2) thaleesàap sǎam thaleesàap (Thai – data from Hundius & Kölver 1983:164)
lake three (CL: lake)
'three lakes'

⁶³ Exceptions exist in Vietnamese and Thai, where the presence of the demonstrative does not require an obligatory classifier (as can be seen in Table 1).

Jennie Tran. 2010. Child Acquisition of Vietnamese Classifier Phrases.

Journal of the Southeast Asian Linguistics Society 3.2:111-137

Copyright vested in the author

Received 15/2/10, revised text accepted 3/12/10

- (3) go3 go3 hok6saang1 (Cantonese – data from Wong 1998:16)
 CL CL student
 ‘every student’

Third, the numeral and the classifier are in all cases adjacent, that is, they are an inseparable pair. An adjective cannot separate the numeral+classifier constituent. The word order of the elements in the numeral classifier phrase can vary depending on the language.

Table 1: *Order of elements in NPs in various numeral classifier languages*

	‘this dog’	‘these two dogs’	‘two black dogs’
Japanese	Dem-N	Dem-Num-CL-N	Adj-N-Num-CL
Cantonese	Dem-CL-N	Dem-Num-CL-N	Num-CL-Adj-N
Mandarin	Dem-CL-N	Dem-Num-CL-N	Num-CL-Adj-N
Thai	N-(CL)-Dem	N-Num-CL-Dem	N-Num-CL-Adj
Vietnamese	(CL)-N-Dem	Num-CL-N-Dem	Num-CL-N-Adj

Since Vietnamese has a different order of classifier phrase elements than that of other languages studied so far, an investigation of the developmental pattern of Vietnamese classifier phrases will add cross-linguistic depth to the classifier acquisition literature. This study represents the first work conducted on the acquisition of the numeral classifier system in Vietnamese. The focus is on investigating how children acquire classifier phrases both in earlier and later stages, as well as on examining the emergence in child speech of the first two-, three- and four-element classifier phrases. Although the study investigates the syntactic as well as the semantic development of classifiers in young Vietnamese-speaking children, the content of the present paper covers only the development of the syntactic aspect of classifiers. The semantic aspect of classifier development will be dealt with in a separate, future paper.

2. Previous research on the syntactic development of classifier phrases

In numeral classifier languages, classifiers play an important role in the development of noun phrase structures. So far, there have been studies on numeral classifier development in six languages (*Cantonese*: Mak 1991; Poon 1981; Szeto 1996; Wong 1998; *Japanese*: Matsumoto 1985a, 1985b, 1987; Muraishi 1983; Sanches 1977; Uchida & Imai 1996, 1999; Yamamoto 2000; *Korean*: Lee 1994, Lee & Lee 2005; *Mandarin*: Erbaugh 1982, 1986; Fang 1985; Hsu 1987; Hu 1993; Loke 1991; Loke & Harrison 1986; Ng 1991; Tse et al. 1991; *Thai*: Carpenter 1987, 1991; Gandour et al. 1984, and *Malay*: Salehuddin & Winkler 2009a.) Among these, the only studies that examined the syntactic properties of classifier development are those by Erbaugh (1982) and Hu (1993) on Mandarin, Carpenter (1987) on Thai, and Wong (1998) on Cantonese. Erbaugh’s and Wong’s studies are the only two longitudinal studies examining naturalistic data from very young children in the age range 1;10 to 3;10 and 1;09 to 2;09 respectively. Carpenter’s is a seminal experimental study on how children between the ages 2;0 to 11;0 acquire the semantic system of Thai classifiers. Erbaugh’s (1982), Hu’s (1993), and Carpenter’s (1987) studies examined the syntactic aspect of classifier phrases rather peripherally. Only Wong’s (1998) study had a primary focus on the syntactic development of noun phrases, including classifier phrases.

With respect to the earliest strategies, previous studies found essentially similar syntagmatic intralinguistic patterns. From very early on, at around age 1;9, children are sensitive to the word order requirements for classifier constructions. They show knowledge of the position of the classifier in the noun phrase. The very first strategy is the 'blank attempt', reported by Gandour et al. (1984), in which children do not yet produce a classifier but make a hesitation after the numeral and a pause to mark the classifier slot. As young as age two, they proceed with the next strategy, filling the classifier slot with the most general classifier, which they use as a default. The general classifier thus serves as a placeholder for the grammatical position of the classifier. Thai acquisition of classifiers manifests two additional strategies: an across-the-board usage of one particular classifier, regardless of the head noun, at around age 3;0, along with an overuse of 'repeaters' across all age groups, where children use the head noun as its own classifier (a strategy referred to as 'overspecification') (Carpenter 1987). These early strategies are important stepping stones into the system as they indicate that children understand the syntactic properties of classifier phrases very early. Children know that (a) the classifier is an obligatory element in the classifier phrase, and (b) classifiers constitute a closed class. Interestingly, the results of Carpenter's (1987) elicitations show that the child subjects never responded with a word that did not conventionally belong in the classifier position, suggesting that children are very aware of the constraints on which words may be classifiers.

The studies by Erbaugh (1982) on Mandarin found that early on, Mandarin-speaking children produce the general classifier together with the demonstrative 'this' and the number 'one', but in the case of specific classifiers, they produce them later, first with the numbers 'two' and above, and then with a demonstrative. Erbaugh further found in her child data that classifiers occur with a noun rather than without. The results in Wong's (1998) study on Cantonese, by contrast, indicate that the grammatical omission of the head noun precedes the mastery of a full classifier phrase. Children tend to combine the classifier with a number before they combine it with a head noun. Their speech in the early multi-word stage exhibits two-element noun phrase structures, namely a very large number of demonstrative+classifier structures at first, then numeral+classifier, followed by classifier+noun. At around age 2;6, their speech exhibits three-element noun phrase structures: demonstrative+classifier+noun first, followed by numeral+classifier+noun. In addition, the classifier phrase occurs earlier and more frequently in object than in subject position. The only instances of ungrammatical patterns are the incorrect use of double classifiers and the omission of an obligatory classifier in the presence of a numeral. These errors are, however, very sporadic (below 0.6%). Similarly, Erbaugh's (1982) investigation of longitudinal naturalistic data comprising over 64 hours of recordings reports a total of only six omissions of obligatory classifiers in the 44,158 utterances produced by a single child.

As can be seen from these findings, children perform better in non-numeral than in numeral constructions. Carpenter's (1987) study on Thai shows that two-year-olds could produce the semantically appropriate classifier in combination with a demonstrative or an adjective, but failed to respond or responded with a single, inappropriate classifier in combination with a numeral. The combination of classifiers with demonstratives is easier because children's understanding of deixis precedes that of counting. This pattern suggests an important relation between cognitive and linguistic development.

3. Classifier phrases in Vietnamese

The three main syntactic properties to be observed when constructing a Vietnamese classifier phrase are the following. First, the classifier is obligatory in the presence of a numeral, both ordinal and cardinal numbers (4a-b), some quantifiers such as *mỗi* ‘every’ (4c), *một vài* ‘a few’ (4d), and the plural markers *những*, *các*, *mấy* (4e). Classifiers are not obligatory with the quantifier *nhều* ‘many, much’ (4f).

(4a) ordinal number + CL	một con (chó)	one CL (dog)
(4b) CL + cardinal number	con (chó) thứ nhất	CL (dog) first
(4c) every + CL	mỗi con (chó)	every CL (dog)
(4d) a few + CL	một vài con (chó)	a few CL (dog)
(4e) plural + CL	những/các con (chó)	PL CL (dog)
(4f) many/much + (CL) + noun	nhều con	many CL
	nhều chó	many dog

The classifier is also obligatory in the presence of demonstratives in direct deixis (4g).

(4g) CL + demonstrative	con (chó) này	CL (dog) this
	con (chó) đó	CL (dog) that
	con (chó) kia	CL (dog) that over there

Classifiers are also required with the *wh*-words *gì* ‘what’ and *nào* ‘which’, when the noun referred to is specific (or definite) (4h), or with the question words *bao nhiêu* or *mấy* ‘how many’ that require a numeral response (4i).

(4h) CL+ <i>wh</i> -word	con (chó) gì?	CL (dog) what?
	con (chó) nào?	CL (dog) which?

(4i) how many + CL	bao nhiêu/mấy ⁶⁴ con (chó)?	how many CL (dog)?
--------------------	--	--------------------

The classifier is not obligatory in direct deixis with demonstratives and *wh*-words when the noun is non-specific. Piriawiboon (2009) notes that in Thai, the noun in a [N-Dem] phrase, where the classifier is absent, can have both an object and a sub-kind reading, whereas the noun in the [N-CL-Dem] phrase, where the classifier is present, has only an object reading. Vietnamese does not pattern like Thai. In (5a), no ambiguity exists in the reading of ‘dog’; it has a sub-kind reading (non-specific). In (5b), the presence of the classifier unambiguously gives ‘dog’ a specific object reading (specific).

(5a) chó này khôn lắm	
dog this smart very	
(i) *‘this particular dog is smart’, (ii) ‘this kind of dog is smart’	

(5b) con chó này khôn lắm	
CL dog this smart very	
(i) ‘this particular dog is smart’, (ii) *‘this kind of dog is smart’	

⁶⁴ Generally, *mấy* is used for numbers below 10 and *bao nhiêu* for numbers above 10.

Another condition for the classifier not being obligatory in direct deixis is the grammaticality and acceptability of the non-occurrence of the classifier with the demonstratives and *wh*-words in less precise speech (Nguyen 1957). The absence of the classifier is more common in spoken discourse.

- (6a) Hai anh muốn ngồi ở bàn nào?
 two you want sit table which
 'At which table do you want to sit?'

- (6b) Anh ngồi xuống ở ghế này.
 you sit down chair this
 'Sit down on this chair.'

(Examples from Emeneau 1951:99)

The CL + Dem sequence in (4g) carries the meaning of singularity and definiteness, whereas in (5a), (6a-b) it is definite, but does not indicate the singular/plural contrast.

Second, the classifier can be used anaphorically, that is, it can co-occur with a numeral (7a), a demonstrative (7b), a *wh*-word (7c), an adjective (7d), a possessive expression (7e), or a relative clause (7f) without the head noun, if the head noun has been sufficiently identified by the previous context. Note that the omitted head noun in (7a)-(e) is, for example, 'dog'.

- | | | | |
|------|----------------------|--|-----------------------------|
| (7a) | Num + CL | hai con | two CL |
| (7b) | CL + Dem | con này | CL this |
| (7c) | CL + <i>wh</i> -word | con nào? | CL which |
| (7d) | CL + Adj | con nhỏ | CL small |
| (7e) | CL + Poss | con của Mi | CL of Mi |
| (7f) | CL+Rel Clause | cái con (voi) | (mà) anh thấy lúc này |
| | | CL CL (elephant) | (which) you see a while ago |
| | | 'the elephant (which) you saw a while ago' | |
- (Nguyen 1957:130)

The classifier + relative clause construction in (7f) is a special characteristic of the Vietnamese classifier phrase as it allows an 'extra classifier' to precede the actual classifier, even though the noun is already classified (Goral 1978; Nguyen 1957). This 'extra classifier' is always the general classifier *cái*, but it is not permitted when the actual classifier is *cái* itself, that is, a sequence of *cái cái* is impossible.

Third, the numeral and the classifier are in all cases adjacent. An adjective, for example, cannot separate the Num-CL constituent, as illustrated (6a) and (6b).

(8a) *Num-CL is inseparable*(8b) *Num-CL cannot be inseparable*Num-CL-Adjmột quả toone CL (fruit;big-round) big

'one big one' (referring to apple)

*Num-Adj-CL*một to quảone big CL (fruit;big-round)

Unlike in Thai and Burmese, in Vietnamese, numeral classifiers cannot be repeated.

There are at least 24 possible two- to four-element classifier phrases in Vietnamese. Table 2 illustrates possible combinations of elements.

Table 2: Possible 2- to 4-element⁶⁵ classifier NPs in Vietnamese with examples

2 elements	Examples	Morphological gloss	Gloss
CL-N	con chó quả trứng	CL (animal) dog CL (fruit/big-round) egg	the dog the egg
CL-Dem ⁶⁶	cái này con này cái đó	CL (general) this CL (animal) this CL (general) that	this this that
CL-wh-word ^o	cái gì? cái nào? con nào?	CL (general) what CL (general) which CL (animal) which	what? which? which?
CL-Adj ^o	con đen	CL (animal) black	the black one
CL-Poss ^o	cái của con	CL (general) of I	mine
Num-CL ^o	hai cái ba con	two CL (general) three CL (animal)	two three

3 elements	Examples	Morphological gloss	Gloss
CL-N-Dem	trái ⁶⁷ banh này cái nồi này	CL (fruit;big-round) ball this CL (general) pot this	this ball this pot
CL-N-wh-word	chiếc xe nào? cái bát nào?	CL (vehicle) car which CL (general) bowl which	which car? which bowl?
CL-N-Adj	cái xe cũ quả bóng vàng	CL (general) car old CL (fruit;big-round) ball yellow	the old car the yellow ball

⁶⁵ By 'element' is meant the grammatical category and not the number of morphemes needed to construct a classifier phrase. For example, the construction CL-Poss consists of two elements, the classifier (CL) and the possessive construction (Poss), but in an actual phrase, CL-Poss consists of three morphemes, e.g. *cái của con*, in which *cái* is CL and *của con* is Poss.

⁶⁶ Both *quả* and *trái* are the classifiers for fruits and big, round shapes. *Quả* is used in the Northern dialect and *trái* in the Southern dialect.

⁶⁷ The phrases marked with the ° symbol are anaphoric uses of the classifier, where the head noun is grammatically omitted. All the other phrases, where the head noun is not omitted, are referred to as full classifier phrases.

CL-N-Poss	cái đàn của con con mèo của em	CL (general) guitar of I CL (animal) cat of I	my guitar my cat
Num-CL-N	hai cây kem ba quả bóng	two CL (long-straight) ice cream three CL (fruit;big- round) ball	two ice creams three balls
Num-CL-Dem ^o	hai cái này mấy cái này	two CL (general) this plural CL (general) this	these two these
Num-CL- <i>wh</i> -wor ^o	hai cái gì đây? hai cái nào?	two CL (general) what here two CL (general) which	what are these two? which two?
Num-CL-Adj ^o	nhiều quả to một ngón nhỏ	many CL (fruit;big- round) big one CL (finger,toe) small	many big ones one small one
Num-CL-Poss ^o	hai cái của con	two CL (general) of I	my two

<i>4 elements</i>	<i>Examples</i>	<i>Morphological gloss</i>	<i>Gloss</i>
CL-N-Adj-Dem	quả trứng bự này	CL (fruit;big-round) egg big this	this big egg
CL-N-Adj- <i>wh</i> - word	chiếc xe nhỏ nào?	CL (vehicle) car small which	which small car?
CL-N-Adj-Poss	quả bóng màu vàng của con	CL (fruit;big-round) ball yellow of I	my yellow ball
Num-CL-N-Dem	hai cục pin này	two CL (small-roundish) battery this	these two batteries
Num-CL-N- <i>wh</i> - word	hai trái banh nào?	two CL (fruit;big-round) ball which	which two balls?
Num-CL-N-Adj	hai chiếc xe nhỏ	two CL (vehicle) car small	two small cars
Num-CL-N-Poss	một cái tất của Linh mấy cái vợt của bố	one CL (one of pair) sock of plural CL (general) racket of Daddy	Linh's one sock Daddy's rackets

It is essential to differentiate between numeral and non-numeral classifier phrases because, as previous research has shown, children perform better in non-numeral than numeral contexts. Table 3 lists this differentiation.

Table 3: *Non-numeral vs. numeral classifier phrases*

<i>Non-numeral</i>	<i>Numeral</i>
CL-N	Num-CL ^o
CL-Dem ^o	Num-CL-N
CL- <i>wh</i> -word ^o	Num-CL-Dem ^o
CL-Adj ^o	Num-CL- <i>wh</i> -word ^o
CL-Poss ^o	Num-CL-Adj ^o
CL-N-Dem	Num-CL-Poss ^o
CL-N- <i>wh</i> -word	Num-CL-N-Dem
CL-N-Adj	Num-CL-N- <i>wh</i> -word
CL-N-Poss	Num-CL-N-Adj
CL-N-Adj-Dem	Num-CL-N-Poss
CL-N-Adj- <i>wh</i> -word	
CL-N-Adj-Poss	

4. Methodology

4.1. Subjects

All child subjects for the current study, a total of 42 participants, were children from monolingual families living in Vietnam. The four subjects for the longitudinal study were from four different families of the same economic and educational level. The other 38 subjects for cross-sectional study were from a daycare center in Ho Chi Minh City, Vietnam.

4.2. Research design

The study employed two different types of data collection: longitudinal naturalistic data and cross-sectional experimental elicited-production data.

4.2.1. Longitudinal

The longitudinal naturalistic data was collected from four children, ages 1;9, 1;11, 2;4, and 2;5, twice a month over a period of six to nine months. Table 4 provides information about these children.

Table 4. *The children in the longitudinal study.*

<i>Name</i>	<i>Gender</i>	<i>Age range</i>	<i>Duration of recording</i>
Minh	boy	1;9 - 2;3	6 months
Ha Mi	girl	1;11 – 2;5	6 months
Liem	boy	2;4 – 2;10	6 months
Giang	girl	2;5 – 3;2	9 months

Each recording session covered about one hour of interaction between the child and his/her parents or caregiver in the child's home. All sessions were both audio-taped and video-recorded. The recording scenarios were playtime, mealtime, bedtime, bath time, and T.V. time. For each session, the investigator brought along various types of toys, props, puppets, pictures and picture books to encourage parents/caregivers and children to use

various types of nouns, since the targeted syntactic constructions are classifier phrases involving nouns. For uncontrolled elicitation, picture book and toys were used.

In addition to the recorded child speech, a questionnaire about the child's noun repertoire was given to the parents to fill out prior to the first recording and then once every month. This questionnaire was designed following the MacArthur Communicative Development Inventory (CDI) for infants and toddlers. It served to reveal any discrepancy between the child's performance during the recorded sessions and his/her actual knowledge of the nouns.

The six to nine-month-stretch of data collection from these four children of different ages is representative of a one and a half-year range of development from 1;9 to 3;2. The age 1;9 is the youngest suitable age for such a study, as it is theoretically two months after the vocabulary spurt and four months before the two-word stage, in which we predict the very first appearance of a classifier, whether in an adult-like form or a phonologically reduced form. On the other hand, the age 3;2 is the oldest appropriate age for such a longitudinal study based on the finding in a pilot study⁶⁸ that children between 3;0 and 3;5 were able to produce numeral+classifier+noun sequences, thus indicating that they had acquired the syntactically adult-like full numeral classifier phrase. This 1.5-year age range is therefore crucial to investigate longitudinally as it covers transitions from bare nouns to two-element and three-element noun phrases. It is during this time that children acquire the grammatical slot of the classifier, the general classifier, and a number of different types of classifier phrases, as reported in previous studies.

4.2.2. Cross-sectional

The cross-sectional experimental elicited-production data was collected every day for a period of six weeks from 38 children between the ages 2;10 and 5;7 in a daycare center in Ho Chi Minh City. The 38 children were divided into three groups. More details can be seen in Table 5.

Table 5. *The children in the cross-sectional study.*

<i>Group #</i>	<i>Group name</i>	<i>Gender</i>	<i>Age range</i>
I	Mầm (youngest)	6 girls, 5 boys	2;10-3;7
II	Chồi (mid)	7 girls, 5 boys	3;8-4;4
III	Lá (oldest)	8 girls, 7 boys	4;7-5;7

The experimental design consisted of two tasks. Each task was conducted with each child individually. The tasks were administered on different days to avoid getting false results due to overtesting and fatigue. The aim of these production tasks was to elicit the obligatoriness of classifiers in constructions with numerals and the use of classifiers in clauses with demonstratives. The children's responses were audio-recorded. Prior to the start of the first elicitation task, each participating child was asked to count a set of eight objects to determine whether the child has mastered the concept of numbers. 114 stimuli

⁶⁸ A pilot study was conducted two years prior to the actual data collection for the present study at a daycare in Ho Chi Minh City, Vietnam with 50 children between the ages 2;6 and 6;5 to find out how many and what types of classifiers Vietnamese children produce, and at what age they can produce which classifiers.

were used, such as pictures, toys, and real objects. Task I elicits the obligatory classifier in numeral constructions: The child is shown pictures of multiple objects and entities on picture cards and asked to say how many of the same objects s/he sees. The question itself does not contain a classifier. The quantity of the objects ranges from 2 to 10. There are 58 different pictures and objects, which attempt to elicit 17 different classifiers. The targeted syntactic structure is numeral+classifier+noun. Task II elicits a classifier phrase with a demonstrative: The child is shown two items that differ only in one feature, for example, two plastic apples - one red and the other green. The child is asked to describe how the two apples are different. There are 43 different pictures and objects. The targeted syntactic structure is classifier+(noun)+demonstrative. The test items used were the same for both tasks, except that task I showed more than two of the same object, whereas task II showed only two of the same object.

For all types of data collection described above, both familiar and unfamiliar objects and entities were used in a 1:1 ratio. For the longitudinal data, the familiar items are those listed by the parents in the noun inventory questionnaire. For the cross-sectional data, the familiar items are those available at the daycare center. The unfamiliar items are those brought along by the investigator.

4.3. Transcription and coding

All children's speech was transcribed and coded following the CHAT format (MacWhinney, 2000), the common transcription format for child language. Each adult and child utterance in all 56 hours of longitudinal data was transcribed with a %mor tier and English gloss. Each single child utterance was coded as spontaneous, imitation, or repetition. Besides the conventional coding for syntactic categories, this study also contains classifier-specific coding and error coding. The analysis was run using the CLAN program (Computerized Language Analysis).

5. Results

Results are primarily derived from the longitudinal data. In the cross-sectional data, only two structures, numeral+classifier+noun and classifier+(noun)+demonstrative, were examined, whereas in the longitudinal data, any classifier phrase produced by any child was extracted.

5.1. General versus specific classifiers

The general classifier in Vietnamese is *cái*. It is used to classify a large group of nouns, covering all kinds of concrete, immobile objects, including furniture, most household items, kitchenware, appliances, tools, clothing items, body parts, very small insects or living things, and abstract nouns. The occurrence of the general classifier is widespread, whereas that of specific classifiers is restricted. The reason for this is that the general classifier also serves as a default classifier. Speakers resort to this classifier when they choose not to use or cannot think of the specific classifier. On the other hand, a (semantically) specific classifier can be applied only with a particular group of nouns whose referents share some specific semantic features and common characteristics associated with the classifier. Using another specific or the general classifier inappropriately with a noun that requires its own semantically appropriate classifier(s) would lead to an unacceptable, ungrammatical overgeneralization on the grounds of semantics and/or pragmatics. Here is a list of the classifiers that occurred in the children's

speech on a frequent basis, in the order of frequency, beginning with those that occurred more often and ending with those that occurred less often.

General classifier:

cái general classifier, inanimate

Specific classifiers:

con	animal, animate
quả/trái	fruit, big-round
chiếc	secondary general classifier, vehicle, one in a pair
cây	long-straight-rigid
củ	root vegetables
cục	small-roundish, undefined
đôi	pair
sợi	long-thin-malleable
viên	small-round
bài	lesson; song; speech
ngôi	house, building; temple; grave
chùm	bunch, bundle
ngón	finger, toe
hột	small-roundish, seed-like

5.2. Findings

Because the focus of the present paper is on the syntactic development of classifier phrases, the analysis of the classifier phrases listed in Table 6 was geared toward only the structure of the classifier phrase, not the semantics of the classifier. This means that the results for each classifier phrase type include both those produced with the general and those produced with the specific classifiers.

Table 6 shows that, in the collected longitudinal data of the four children, classifiers occurred in 19 grammatical and three ungrammatical syntactic positions. Of the 19 grammatical syntactic positions, 11 were produced by all four children; the other eight were produced only by the two older children (2;4-3;2). All four children made ungrammatical syntactic constructions (errors). (See Appendix A for explication of each child's errors.)

The most frequent position in which the classifier occurs is the one preceding the noun, i.e., classifier+noun (CL-N). All four children overwhelmingly prefer this pre-noun position, with usage rates ranging from 59% of all classifier constructions for the oldest child to 82% for the youngest child. The second most frequent position for the classifier is just before the determiner, i.e., classifier+demonstrative (CL-Dem). Compared to the pre-noun position, however, this construction occurs much less frequently, ranging only from 8% to 24% of total usage. The third most frequent position is that preceding the *wh*-word, i.e., classifier+*wh*-word (CL-Wh), which also occurs much less often than the pre-noun position at only 2% to 12%.

Grammatical omission of the head noun precedes the mastery of a full classifier phrase. Compare #1-6 with #7-19 in Table 6. The classifier phrases with determiners and *wh*-words that display head noun omission, CL-Dem and CL-Wh, are more frequent, and also first acquired by all children. The full-fledged three-element classifier phrases with a

determiner or *wh*-word are produced much less frequently than the two-element classifier phrases with the determiner or *wh*-word.

Table 6: *Total number of tokens of classifier phrases produced by each child in all sessions, including impermissible NPs (errors)*

#		CL Phrase Type	Minh 1;09-2;03 (6 months)	Ha Mi 1;11-2;05 (6 months)	Liem 2;04-2;10 (6 months)	Giang 2;05-3;2 (9 months)
1	2-element	CL-N	762 (82%)	591 (77%)	584 (48%)	1638 (59%)
2		CL-Dem	95 (10%)	64 (8%)	285 (24%)	238 (9%)
3		CL-Wh	20 (2%)	41 (5%)	74 (6%)	337 (12%)
4		CL-Adj	2 (0.2%)	4 (1%)	4 (0.3%)	15 (1%)
5		CL-Poss		3 (0.4%)		1 (0.04%)
6		Num-CL	5 (1%)	11 (1%)	32 (3%)	92 (3%)
7	3-element	CL-N-Dem	8 (1%)	6 (1%)	64 (5%)	3 (0.1%)
8		CL-N-Wh			14 (1%)	9 (0.3%)
9		CL-N-Adj	8 (2%)	1 (0.1%)	26 (2%)	31 (1%)
10		CL-N-Poss	8 (1%)	30 (4%)	32 (3%)	43 (2%)
11		Num-CL-N	3 (0.3%)	3 (0.4%)	11 (1%)	287 (10%)
12		Num-CL-Dem		1 (0.1%)	6 (0.5%)	22 (1%)
13		Num-CL-Wh				2 (0.2%)
14		Num-CL-Adj				2 (0.2%)
15		Num-CL-Poss		3 (0.4%)		1 (0.04%)
16	4-element	CL-N-Adj-Poss			1 (0.1%)	1 (0.04%)
17		Num-CL-N-Dem			3 (0.2%)	
18		Num-CL-N-Adj			2 (0.2%)	7 (0.3%)
19		Num-CL-N-Poss				2 (0.2%)
1	Errors	*Num-N	1	3	29	65
		TOTAL	925	770	1206	2799

Vietnamese children combine the classifier with a demonstrative before they combine it with a number. Compare #1 and #6 in Table 6. This pattern also prevails in three-element combinations. The three younger children prefer to combine the demonstrative with CL-N rather than with Num-CL. Compare #7 and #12 in Table 6.

Vietnamese children tend to combine numbers with a classifier alone before they combine them with the classifier+head noun, as can be seen in Figure 1. Numeral+classifier emerges at around 2;01, whereas numeral+classifier+noun emerges more than half a year later, at around 2;09.

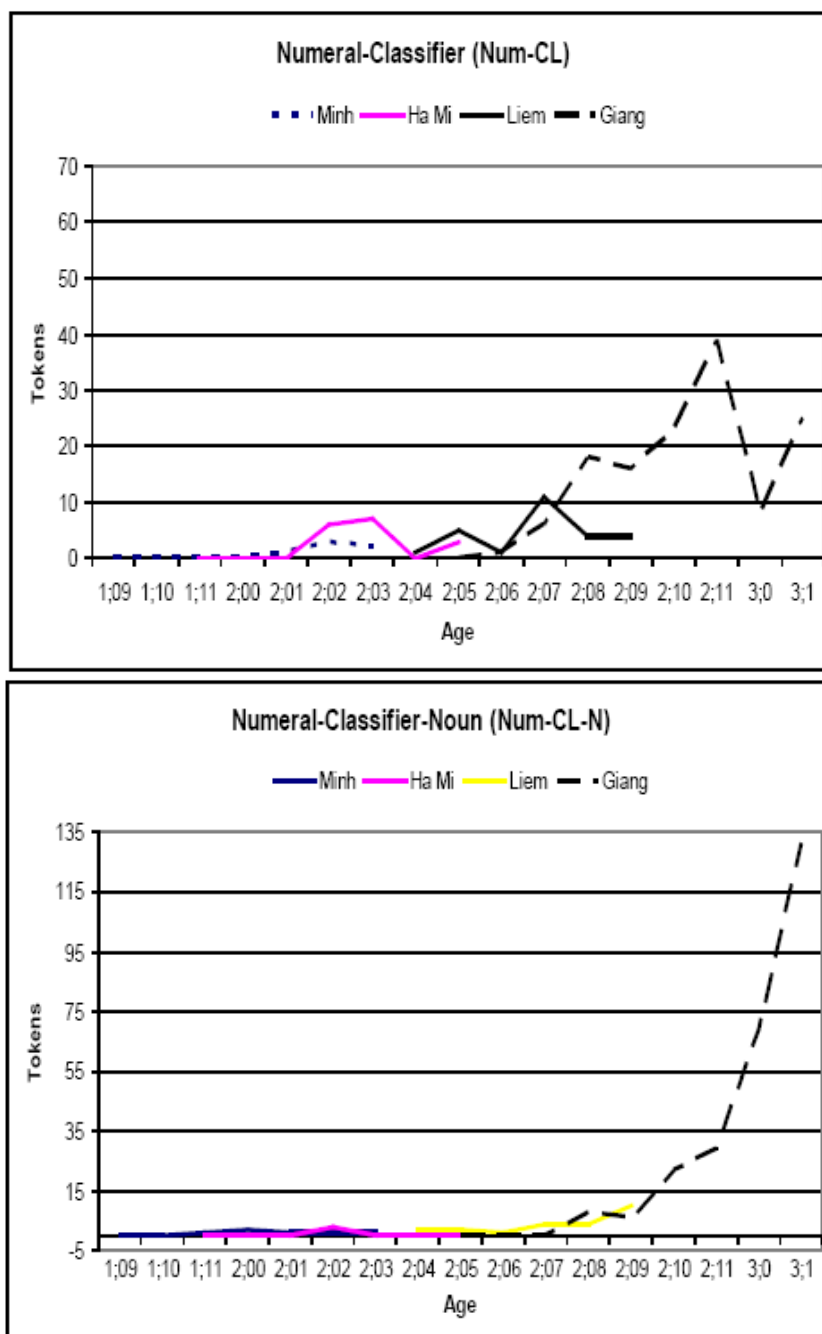


Figure 1: Age of acquisition of Numeral-Classifier and Numeral-Classifier-Noun

An important finding is the order of acquisition of the two-element classifier phrases as it can be compared to previous findings in other numeral classifier languages. This study's findings show that Vietnamese children produce classifier+noun first, classifier+demonstrative second, and numeral+classifier third, as shown in (7a) - (7c).

(7a) cái ly	(7b) cái này	(7c) hai cái
CL glass	CL this	two CL
'the glass'	'this'	'two'

5.3. Number of classifier phrase types

Information on how many classifier phrase types each of the children could produce over the duration of the study is provided in Table 7.

Table 7: *Number of CL phrase types produced by each child*

	Minh 1;09-2;03	Mi 1;11-2;05	Liem 2;04-2;10	Giang 2;05-3;02	Possible number of CLP types
CL phrase types	9	12	16	18	22

The children all display diversity of structure and a neat developmental trajectory. The number of classifier phrase types increases over time: Each older child produces a greater number of classifier phrase types than the next youngest child. Minh, the youngest child, produces nine; Ha Mi, the next youngest, produces 12; Liem, the second oldest child, produces 16; and Giang, the oldest child, produces 18 different types of classifier phrases. More three-element classifier phrases and the more complex four-element ones appear in the speech of the two older children.

The eight most frequently occurring noun phrase and classifier phrase types, arranged in rough order of frequency across all four children, are shown in Table 8.

Table 8: *The eight most frequent NP types in tokens*

NP Type	Minh 1;09-2;03	Ha Mi 1;11-2;05	Liem 2;04-2;10	Giang 2;05-3;02
N	1751	857	1212	2139
CL-N	762	591	584	1638
CL-Dem	95	64	285	238
CL-Wh	20	41	74	337
N-Poss	71	124	103	232
N-Dem	66	19	84	28
N-Adj	10	12	48	49
Num-CL	5	11	32	92

Based on the frequency order presented in Table 8, the most often-used NP types are the one- and two-element ones, whether or not the noun phrase contains classifier.

5.4. Child data examples

Selected actual child utterances from the data illustrating the children's production of some of the classifier phrases listed in Table 6 are presented below. The target utterances are underlined. Note that due to lack of space, the morphological gloss and the English gloss was included only for the target utterances produced by the child; for the other utterances, only one of the glosses was included. (Key of tagging code: sfp = sentence final particle, EMP = emphatic particle, IMP = imperative marker).

CL-N

Minh and auntie Tam were conversing. Suddenly, Minh pointed to an eraser and said:

- *CHI: cái cục cái cục.
CL (general) CL (small roundish) 2x
- *TAM: cục gì?
CL (small roundish) what?
- *CHI: cục gôm. (Minh, Session 2, Age: 1;10)
CL eraser

CL-Adj

Giang and her aunt were looking at picture books. Giang pointed at a picture and asked:

- *CHI: a cái gì vậy Cô_Linh?
'Oh what's that, auntie Linh?'
- *AUN: đồ Giang?
'I ask you.'
- *CHI: cái ... cái a umbrella nè.
CL ... CL oh umbrella sfp
- *AUN: giỏi.
'Good.'
- *CHI: cái dù umbrella nè.
CL umbrella umbrella sfp
- *CHI: cái nhỏ nè. cái to nè. (Giang, Session 16, Age: 3;0)
CL small sfp. CL big sfp.

CL-Poss

Ha Mi and her Mom were looking at pictures on a poster. Her Mom asked her to name the different things she saw on the poster.

- *MOT: cái này?
'This?'
- *CHI: cái của Mi. (Ha Mi, Session 5, Age: 2;1)
CL of Mi
- *MOT: hử?
huh?
- *CHI: cái của Mi nè.
CL of Mi sfp
- *MOT: à, cái của Mi hả?
oh, CL of Mi sfp
'Oh, it's yours?'

CL-N-Dem

Liem and his Mom were playing with cars.

- *CHI: cái xe này bể rồi. (Liem, Session 1, Age: 2;4)
CL car this break PERF
- *MOT: à, xe này bể rồi.
'Oh yes, this car is broken.'
- *CHI: xe mới đi.
car new IMP
- *MOT: xe mới đi hả?
'Oh you want a new car?'

CL-N-Wh

Liem and his Mom were playing with cars.

- *CHI: con chỉ mẹ. con chơi tiếp.
'I show you. I continue playing.'
- *CHI: con chỉ mẹ chiếc xe nào nha. (Liem, Session 7, Age: 2;7)
I show you CL car which sfp
'I show you which car.'
- *MOT: rồi con chỉ mẹ đi.
'Okay show mẹ.'

CL-N-Adj

- *CHI: tập đi ba.
notebook IMP Daddy
- *CHI: cái tập đen. (Minh, Session 6, Age: 2;0)
CL notebook black
- *DAD: rồi ra đây, nè.
'Ok, come out, here.'
- *CHI: đâu rồi ba? tập đâu rồi ba? tập đâu rồi?
'Where, Daddy? Where's the notebook, Daddy? Where's the notebook?'
- *DAD: đây nè.
'Here.'

Num-CL-N

Minh was playing tearing pictures out of a picture book with Dad and Mom. When he found the picture with two balls, he tore it out. He called this game 'cutting'.

- *CHI: ba cắt cho. trái banh đâu mẹ? trái banh đâu rồi?
Daddy cut sfp. CL ball where Mommy? CL ball where EMP?
- *MOT: con kiếm đi.
'Look for it.'
- *CHI: trái banh đâu mẹ?
CL ball where Mommy
- *MOT: hả?
huh?
- *CHI: hai trái banh. (Minh, Session 6, Age: 2;0)
two CL ball
'Two balls.'
- *CHI: cắt. giựt. xong rồi.
'Cut. Throw away. Finished.'

Num-CL-Dem

Liem and his Dad were sitting at the table eating lunch. Liem asked:

- *CHI: mấy cái này gì đây ? (Liem, Session 8, Age: 2;8)
PL CL this what here
'What are these?'
- *DAD: rau.
'Greens'
- *CHI: rau để ăn hả?
'Greens for eating?'
- *DAD: ừm.
'Yes.'
- *CHI: bỏ vô miệng ăn ba nha.
put in mouth eat Daddy sfp
- *DAD: ừm, để đó ba lấy ba ăn.
'Ok, leave it there, I'll take and eat it.'

Num-CL-Wh

Giang and her aunt were looking at picture books. Auntie pointed at a picture.

*AUN: Cô Linh chỉ nè.
'I point here.'

*CHI: hai cái gì vậy cô Linh? (Giang, Session 16, Age: 3;0)
two CL what sfp auntie Linh
'What are those two?'

*AUN: Hư?
'Huh?'

*CHI: chiếc thuyền hả?
CL boat sfp
'Is it a boat?'

*AUN: đây mà cái thuyền a hả?
'What? This is a boat?'

Num-CL-Adj

Auntie asked Giang to tell what she saw in the picture.

*AUN: Cô Linh chỉ cái hình này là Giang nói nè. Gì đây nè? Đây nè.
'I show you this picture then you will tell me. What is this? This.'

*CHI: Cô Linh ơi nhiều quả to Cô Linh ơi nhiều quả .(Giang, Session 15, Age: 3;0)
auntie Linh many CL big auntie Linh many CL
'Auntie Linh, there are many big ones, auntie Linh, many.'
(referring to balloons)

5.5. Emergence order

This study employs Wong's (1998) criteria for emergence order: If the child uses a certain type of structure continuously across three consecutive sessions, the child is deemed able to use that structure productively and the first session of that series is regarded as the onset of this development. Figure 2 below illustrates this emergence order.

Age	Classifier Phrase	Error	numeral/non-numeral, number of elements
1;9 (↑)	CL-N, N-Poss CL-Poss		non-numeral 2-element
1;10	CL-Dem, N-Dem		
1;11	CL-Wh		
2;0	CL-N-Poss, CL-N-Dem		non-numeral 3-element
2;1			
2;2	Num-CL	wrong number indication	numeral 2-element
2;3	N-Wh, N-Adj		
2;4			
2;5			
2;6	MORE DIVERSE		
2;7			
2;8	Num-CL-N, *Num-N Num-CL-Dem, CL-N-Adj	classifier omission	numeral 3-element
2;9			(non-numeral and numeral 4-element)
2;10			
2;11	CL-Adj		
3;0			
3;1			
3;2	fewer *Num-N		
.....			
3;8		fewer classifier omission	
.....			
4;7		fewer wrong # indication	

Figure 2: *Emergence Order*

The classifier phrase that emerges first is CL-N. Evidence for this is provided by Minh, the youngest child, who could produce such phrases already at 1;9; additionally, CL-N phrases appear in the other three children's speech from the time of the first recording. The two-element non-numeral possessive construction, N-Poss, is acquired early as well, before the age of 1;9. Two-element noun phrases also emerge early at around 1;10. Children begin using combinations of the classifier (or the noun) with the determiner and with the interrogative (CL-Dem, N-Dem, CL-Wh) between 1;11 and 2;1.

Children build their first *non-numeral* three-element classifier phrases at around age 2 by adding a third element to the CL-N sequence. These are CL-N-Dem and CL-N-Poss. The general pattern of expanding a two-element to a three-element classifier phrase involves adding an element to the fixed CL-N sequence.

CL-N	+	Dem
		Poss
		Adj

CL-N-Poss is the first three-element NP that emerges, at around 2;0. CL-N-Dem emerges at around the same time. Before 2;0, there is no productive use of three-element phrases. Results from Task II of the cross-sectional data reconfirm this development as they show a 100% correct production of the CL-N-Dem construction in the children aged 2;10 to 5;5.

The first *numeral* classifier phrase that children build consists of two elements, namely Num-CL. This appears later than non-numeral classifier phrases, at around age 2;2. The first numeral three-element classifier phrases are built by adding a third element to the numeral-classifier sequence. These are Num-CL-N, Num-CL-Dem, Num-CL-Poss. They appear later, at around age 2;8.

Num-CL	+	N
		Dem
		Poss

Classifier omission errors (*Num-N) also appear at around age 2;8. CL-N-Adj emerges between 2;7 and 2;8, about the same time as Num-CL-N. CL-N-Adj emerges earlier than CL-Adj. This is because CL-Adj occurs rarely in adult speech. Shortly after the three-element classifier phrases have developed, at around 2;9, the first four-element classifier phrases emerge, including both non-numeral and numeral constructions.

Other constructions that occurred in small numbers are the three-element CL-N-Wh, the four-element CL-N-Adj-Poss, Num-CL-N-Dem, Num-CL-N-Adj, and Num-CL-N-Poss. Only the two older children could produce these more sophisticated four-element classifier phrases; they emerged between ages 2;8 and 3;2.

5.6. Classifier omission errors

Classifier omission errors (*Num-N) involve the ungrammatical omission of the obligatory classifier in a classifier structure containing a numeral (Num-CL-N). Here are some actual examples from the longitudinal data.

(8a) *CHI: ba lấy cho Minh một bóng đi. (Minh 2;3)
 you get for Minh one ball IMP
 ‘Daddy, get a ball for me.’

(8b) *CHI: có hai chổi . (Ha Mi 2;2)
 have two broom
 ‘There are two brooms.’

- (8c) *CHI: hai mèo luôn . (Ha Mi 2;3)
two cat EMP
'There are two cats!'
- (8d) *CHI: hai banh đây . (Liem 2;9)
two ball here
'Here are two balls.'
- (8e) *CHI: một tất nè. (Giang 3;1)
one sock here
'Here is one sock.'
- (8f) *CHI: hai chuối . (Ngoc Tien 3;4, cross-sectional)
two banana
'Two bananas.'
- (8g) *CHI: hai kẹo . (Gia Thinh 3;9, cross-sectional)
two candy
'Two candies.'

In all of these ungrammatical utterances (*Num-N), the obligatory classifier is missing. In (8a), (8d) and (8f), the classifier for fruits and big/round objects, *trái* or *quả*, is missing. In (8b), the classifier *cây* for long, thin objects is missing. In (8c), the classifier for animals, *con*, is missing. In (8e), the classifier *chiếc* for one of a pair is missing. In (8g), the classifier *viên* for small, roundish objects is missing.

The most classifier omission errors occurred in the speech of the two older children, at around 2;8. The two younger children produced very few classifier structures involving a numeral (Num-CL-N) (only four token in the case of Minh and six tokens in the case of Mi) and almost half of them were classifier omission errors (*Num-N). Table 9 shows a comparison among the main numeral classifier phrases: numeral+classifier (Num-CL), numeral+classifier+noun (Num+CL+N), and the erroneous *numeral+noun (*Num+N). This illustrates which of these constructions emerges first and at what ages the children make classifier omission errors.

Table 9: Comparing the main numeral classifier phrases

Minh					Ha Mi				
Session	Age	Num-CL	Num-CL-N	*Num-N	Session	Age	Num-CL	Num-CL-N	*Num-N
1	1;09;20				1	1;11;10			
2	1;10;06				2	1;11;27			
3	1;10;20				3	2;00;11			
4	1;11;04				4	2;00;29			
5	1;11;19		1		5	2;01;27			
6	2;00;03				6	2;02;11	4	3	1
7	2;00;16				7	2;02;27	2		1
8	2;01;01				8	2;03;16	2		
9	2;01;15	1	1		9	2;04;00			1
10	2;01;29	3			10	2;04;14			
11	2;02;13				11	2;04;28			
12	2;02;27				12	2;05;12	3		
13	2;03;11	1	1	1					
TOTAL		5	3	1	TOTAL		11	3	3

Liem					Giang				
Session	Age	Num-CL	Num-CL-N	*Num-N	Session	Age	Num-CL	Num-CL-N	*Num-N
1	2;04;09	1	1		1	2;05;01			
2	2;04;19	5		1	2	2;05;18			
3	2;05;22	4		1	3	2;06;10			1
4	2;06;09				4	2;06;23	1		
5	2;06;20	1			5	2;07;08	6		2
6	2;07;07	4	1		6	2;07;22			
7	2;07;22	5	3	3	7	2;08;01		3	1
8	2;08;11				8	2;08;10	1	4	
9	2;08;25	5	2		9	2;09;01	4	2	
10	2;09;08	6	1	5	10	2;09;18	2	1	
11	2;09;29	1	3	20	11	2;10;00	5		2
TOTAL		32	11	30	12	2;10;14	5	9	7
					13	2;11;00	4	7	6
					14	2;11;17	5	7	3
					15	3;00;03	18	10	2
					16	3;00;17	6	24	5
					17	3;01;00	2	44	7
					18	3;01;14	12	95	15
					19	3;01;27	18	42	10
					20	3;02;13	3	39	6
					TOTAL		92	287	67

It can be inferred from this comparison table that the first *numeral* classifier phrase to emerge in the speech of Vietnamese children is the combination of the numeral+classifier (Num-CL), at around 2;2. The three-element numeral+classifier+noun (Num-CL-N) starts to be used productively at around 2;8. The period before that, between 2;2 and 2;8, is the period of trial and error: Children produce the correct forms while also making classifier omission errors. These errors do not stop when they learn to use the three-element numeral+classifier+noun productively. Rather, such errors continue to occur concurrently with the correct forms, but to a lesser extent, as can be seen in the data from the oldest child, Giang.

The rates of classifier omission errors are shown in Table 10.

Table 10: *Classifier omission errors in longitudinal study*

<i>Child</i>	<i>Age</i>	<i>total CL omission tokens</i>	<i>total expected Num-CL-N tokens</i>	<i>total % classifier omission</i>
Minh	1;9-2;3	1	4	25.0%
Ha Mi	1;11-2;5	3	6	50.0%
Liem	2;4-2;10	30	41	73.0%
Giang	2;5-3;2	67	354	19.0%

The two younger children, Minh and Ha Mi (1;9-2;5) displayed classifier omission error rates of 25% and 50%. Surprisingly, the third child, Liem (2;4-2;10), had a high error rate of 73%. The oldest child, Giang (2;5-3;2) displayed an omission rate of 19%. Despite this inconsistent progression, however, the general trend is for a positive correlation between age and frequency of use of numeral classifier constructions. A child like Liem with a high rate of omission is passing through the period of trial and error.

The results from Task I of the cross-sectional data further show that after around 3;2, classifier omission errors start to decrease. Table 11 illustrates the cross-sectional omission rates.

Table 11: *Classifier omission errors by children in cross-sectional study*

<i>Age</i>	<i># of subjects</i>	<i>total CL omission tokens</i>	<i>total elicited tokens</i>	<i>total % classifier omission</i>
2;10-3;7	n = 11	109	623	17.5%
3;8-4;4	n = 12	89	687	13.0%
4;7-5;7	n = 15	34	847	4.0%

The youngest group (2;10-3;7) had an error rate of 17.5%, the mid group (3;8-4;4), 13%, and the oldest group (4;7-5;7), 4%. The youngest group of the cross-sectional study (2;10-3;7) displayed on average a slightly lower error rate (17.5%) than the oldest child (2;5-3;2) in the longitudinal study (19%). The period in which children's error rates begin to decrease can therefore be roughly estimated to fall between 3;2 and 3;7.

5.7. General classifier as a placeholder

To find out whether young Vietnamese children use the general classifier in the classifier slot as a placeholder, it is necessary to examine data of the two younger children in the

longitudinal study. This is because their speech shows which classifiers children use to build their first classifier phrases. The results reveal that these two youngest Vietnamese children do use the general classifier as a placeholder to build their first two-element classifier phrases, as shown in Table 12. For a detailed chart, refer to Appendix B.

Table 12. *Use of the general classifier by the two youngest children.*

<i>ChildAge</i>	<i>CL phrase</i>	<i>general CL tokens</i>	<i>specific CL tokens</i>
Minh 1;9-2;3	CL-Dem	93	4
	CL-Wh	21	3
	Num-CL	5	1
	Total	119	8

<i>Child Age</i>	<i>CL phrase</i>	<i>general CL tokens</i>	<i>specific CL tokens</i>
Ha Mi 1;11-2;5	CL-Dem	63	4
	CL-Wh	32	11
	Num-CL	12	9
	Total	107	24

As can be seen in this table, the use of the general classifier far exceeds the use of specific classifiers in these two children's early two-element classifier constructions, classifier+demonstrative (CL-Dem), classifier+*wh*-word (CL-Wh), and numeral+classifier (Num-CL). The results in Appendix B also show that very young children before the age of two and a half can mainly produce only four specific classifiers, which are *con*, *trái*, *chiếc*, and *cây*, out of the very large inventory of about 200 Vietnamese classifiers.

6. Discussion

6.1. Observance of syntactic properties

Vietnamese children acquire non-numeral classifier phrases early and make no errors with the structure of these phrases. At the time of their emergence (between 1;9 or earlier and 2;8), all two- to three-element classifier phrases are well-formed syntactically when used with the noun, demonstrative, *wh*-word, possessive, and adjective. When the classifier phrase with a numeral first emerges (at around 2;2), children can produce well-formed two-element numeral+classifier phrases. They observe two of the three syntactic properties in the construction of the Vietnamese classifier phrase: (1) they use the classifier anaphorically and (2) they position the numeral and classifier adjacent to each other. However, children have difficulty observing the third syntactic property, which mandates obligatory classifiers in the presence of numerals. For this reason, they make classifier omission errors when producing the three-element numeral+classifier+noun phrase. Other non-numeral three-element classifier phrases, numeral+classifier+demonstrative and numeral+classifier+possessive, which emerge in limited usage at around the same time, are well-formed. Results from the cross-sectional data show that classifier omission errors decrease only after age 3;2, and begin more rapid decline after 4;7. Young Vietnamese children cannot yet use ordinal numbers and the quantifiers 'every' or 'a few' with the classifier.

6.2. *Order of emergence*

Most non-numeral two-element noun phrases, except for classifier+adjective, are acquired before 2;0. Between 2;0 and 3;2, five of the nine possible three-element classifier phrases are used productively. These results are quite similar to Wong's (1998). Vietnamese children tend to use classifier phrases in more diverse ways after the age of two and a half. The first noun phrase structure involving a numeral that emerges is numeral+classifier, which appears at around age 2;2. The three-element numeral+classifier+noun appears alongside classifier omission errors (*numeral+noun) at age 2;8. Between 2;8 and 3;2, children produce both grammatical and ungrammatical numeral classifier structures. During this same period of time, more three-element and the more complex four-element classifier phrases appear. Not until late in their third year do children produce grammatical numeral classifier phrases.

6.3. *Comparisons with previous studies*

These results show that Vietnamese children demonstrate early knowledge of the classifier slot in noun phrases in non-numeral contexts. They know early on that the slot for the classifier precedes the noun. They do not make any word order errors with classifier phrases. As early as age 1;9, they can produce an obligatory classifier, not only with a noun, but also with a demonstrative and an interrogative. This is consistent with results of previous studies in other Asian languages investigating the development of classifier phrases (Erbaugh 1982, Hu 1993 on Mandarin, Carpenter 1987 on Thai, Wong 1998 on Cantonese). As for Vietnamese children's use of the 'blank attempt' strategy, there is no clear evidence in the current study. A more in-depth phonetic-prosodic examination of the data may reveal that children can rely on phonological or tonal cues to identify the syntactic position of classifiers. As for the 'general classifier as placeholder' strategy, clear supporting evidence is found in the current study. Vietnamese children use the general classifier to fill the grammatical position of the classifier.

A previous classifier acquisition study by Erbaugh (1982) on Mandarin found that classifiers occur with both a demonstrative and a number first. The results of the present study on Vietnamese show that classifiers occur with a demonstrative before they occur with a number. (Refer to Table 6 and Figure 2). Erbaugh further found that classifiers occur with a head noun rather than without. By contrast, a previous classifier acquisition study by Wong (1998) on Cantonese found that grammatical omission of the head noun precedes the mastery of the full classifier phrase. Wong's finding is consistent with the Vietnamese data. When constructing a classifier phrase, Vietnamese children produce the classifier without a head noun rather than with it. The classifier phrases without the head noun, classifier+possessive, classifier+demonstrative, classifier+*wh*-word and numeral+classifier, do precede their full-fledged counterparts in the Vietnamese children's speech. (Refer to Figure 2). Wong's study further found that children tend to combine the classifier with a number before they combine it with a head noun. She explained that, among two-element noun phrase structures, children produce demonstrative+classifier first, then numeral+classifier, then classifier+noun.

The results of the present study on Vietnamese differ from Wong's findings. The emergence order of two-element noun phrase structures for Vietnamese children is classifier+noun first, classifier+demonstrative second, and numeral+classifier third. (Refer to Figure 2 and examples (7a) – (7c).) The present results clearly show that very young Vietnamese children employ a very high number of classifier-noun constructions. The

youngest child alone produced 58 tokens at the first session at age 1;09. (Refer to Appendix A). The children in this study either learned all nouns together with their appropriate classifiers as noun chunks, or they learned only certain nouns with the classifier, and only used these nouns in their speech in their beginning sessions. In order to determine this, an in-depth analysis of the nouns used by each child in each session is necessary.

Wong (1998) further found that at around age 2;6, children's speech begins to exhibit three-element noun phrase structures, starting with demonstrative+classifier+noun and then growing to include numeral+classifier+noun. This is similar to the emergence order in Vietnamese. Classifier+noun+demonstrative is acquired first at around 2;0 and numeral+classifier+ noun develops later at around 2;8. (Refer to Figure 2).

Previous findings have shown that children perform better when using non-numeral constructions than when using numeral constructions. The results of the present study on Vietnamese fully support this finding. However, previous studies did not report any significant rate of classifier omission errors. These previous studies report that by age three, children make very few structural errors such as omission of the classifier from numerical constructions (below 0.6%). On the contrary, the Vietnamese two, three and four year-old children in this study made a higher percentage of omission errors than those found in previous studies in other languages. (Refer to Tables 10 and 11). I speculate that this is because the semantic knowledge that children would need to select the appropriate classifier for each noun is not well developed; thus, young children prefer to leave out the classifier. In other words, Vietnamese children are more conservative in their acquisition of the classifier system than are child speakers of other languages with similar classifier systems. Vietnamese children choose to omit the obligatory classifier rather than deliver the wrong classifier; they prefer to make mistakes in syntax for the sake of avoiding mistakes in semantics.

References

- Carpenter, Kathie Lou. 1987. *How children learn to classify nouns in Thai*. Ph.D. dissertation, Stanford University.
- . 1991. Later rather than sooner: extralinguistic categories in the acquisition of Thai classifiers. *Journal of Child Language* 18: 93-113.
- Emeneau M. B. 1951. *Studies in Vietnamese (Annamese) Grammar*. Berkeley-Los Angeles: University of California Publications in Linguistics 8.
- Erbaugh, Mary. 1982. Coming to order: Natural selection and the origin of syntax in the Mandarin speaking children. Ph.D. dissertation, University of California, Berkeley.
- . 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. In C. G. Craig (ed.), *Noun Classes and Categorization*, pp. 399-436. Amsterdam and Philadelphia: John Benjamins.
- Fang, Fuxi. 1985. An experiment of the use of classifier by 4 to 6 year olds. *Acta Psychologica Sinica* 17,4: 384-392.
- Hsu, Joseph H. 1987. *A study of the various stages of development and acquisition of Mandarin Chinese by children in Chinese milieu*. National Science Council Research Report, College of Foreign Languages, Fu Jen Catholic University.

- Hu, Qian. 1993. The acquisition of classifiers by young Mandarin-speaking children. Ph.D. dissertation, Boston University.
- Gandour, Jack., Soranee Holasuit Petty, Rochana Dardarananda, Sumalee Dechongkit, and Sunee Mukngoan. 1984. The acquisition of numeral classifiers in Thai. *Linguistics* 22: 455-79.
- Goral, Donald R. 1978. Numeral Classifier Systems: A Southeast Asian cross-linguistic analysis. *Linguistics of the Tibeto-Burman Area* 4: 1-72.
- Hundius, Harald, and Kölver, Ulrike. 1983. Syntax and semantics of numeral classifiers in Thai. *Studies in Language* 7: 165-214.
- Lee, Kwee-Ock. 1994. Acquisition of Korean classifiers: Syntactic and semantic factors. Paper presented at the Symposium on Linguistics Theory and the Acquisition of Korean Semantics and Syntax.
- _____, and Sun-Yong Lee. 2005. The acquisition of Korean numeral classifiers. Unpublished manuscript. Kyungsung University & Kyung Hee University.
- Loke, Kit Ken. 1991. A semantic analysis of young children's use of Mandarin shape classifiers. In A. Kwan-Terry (ed.), *Child Language Development in Singapore and Malaysia*, pp. 98-116. Singapore: Singapore University Press.
- _____, and G. Harrison. 1986. Young children's use of Chinese (Cantonese and Mandarin) sortal classifiers. In *Linguistics, Psychology, and the Chinese Language*. Hong Kong: Centre of Asian Studies, University of Hong Kong.
- MacWhinney, B. (2000). *The CHILDES Project : Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mak, David Lai-Woon. 1991. The acquisition of classifiers in Cantonese. Ph.D. dissertation, University of Reading.
- Matsumoto, Yo. 1985a. Japanese numeral classifiers: Their structure and acquisition. M.A. Thesis, Sophia University.
- _____. 1985b. Acquisition of some Japanese numeral classifiers: The search for convention. *Stanford University Papers and Reports in Child Language Development* 24: 89-96.
- _____. 1987. Order of acquisition in the lexicon: Implications from Japanese numeral classifiers. In K. Nelson and A. van Kleeck (eds.), *Children's Language*, Vol. 6. pp. 229-60. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Muraishi, Shozo. 1983. Josuushi tesuto [classifier tests]. In [The National Language Research Institute] (ed.), [*Conceptual Development and Language in Children*]. Tokyo Shoseki: Tokyo.
- Ng, Bee Chin. 1991. Word meaning acquisition and numeral classifiers. *La Trobe University Working Papers in Linguistics* 4: 73-83.
- Nguyen, Dinh Hoa. 1957. Classifiers in Vietnamese. *Word* 13: 124-52.
- Piriyawiboon, Nattaya. 2009. *The role of classifiers in N + Dem in Thai*. Paper presented the 19th annual meeting of the Southeast Asian Linguistics Society (SEALS).
- Poon, Emma Yuen-wai. 1980. Some aspects of the ontological development of nominal classifiers in Cantonese. M.A. Thesis, University of Hong Kong.

- Salehuddin, K., & Winskel, H. 2007. *The role of input and cognitive development in Malay numeral classifier acquisition*. Paper presented at the 15th Biennial Conference of the Australian Human Development Association.
- Sanches, Mary. 1977. Language acquisition and language change: Japanese numeral classifiers. In B. Blount and M. Chanse (eds.), *Sociocultural Dimensions of Language Change*. New York: Academic Press.
- Szeto, Kitty K. 1996. Classifiers in Cantonese-speaking children from 1 year 5 months to 3 years 8 months. In T. Lee, C. Wong, S. Leung, P. Man, A. Cheung, K. Szeto and C. Wong (eds.), *The Development of Grammatical Competence in Cantonese-speaking Children - Report of a project funded by RGC earmarked grant CUHK 1991-94*, pp. 175-202. Hong Kong.
- Tse, John Kwok-ping, Tang Ting-chi, Shie Yu-huei, and Cherry Y. Li. 1991. *Chinese children's language acquisition and development*. National Science Council Research Report, Department of English, National Taiwan Normal University.
- Uchida, Nobuko, and Mutsumi Imai. 1996. A study on the acquisition of numeral classifiers among young children: The development of human-animal categories and generation of the rule of classifiers applying. *Japanese Journal of Educational Psychology* 44, 2: 126-135.
- _____. 1999. Heuristics in learning classifiers: The acquisition of the classifier system and its implications for the nature of lexical acquisition. *Japanese Psychological Research* 41, 1: 50-69.
- Wong, Cathy Sin Ping. 1998. *The acquisition of Cantonese noun phrases*. Ph.D. dissertation, University of Hawaii.
- Yamamoto, Kasumi. 2000. *The acquisition of Japanese numeral classifiers*. Ph.D. dissertation, Cornell University.

PULLING OUT ALL THE STOPS IN VIETNAMESE: A DELINEATION BETWEEN NATIVE AND NON-NATIVE VIETNAMESE SPEECH FOR VOICE ONSET TIME

Alina Twist, Jessica Shamoo Marx, Jessica Bauman, and Allison Blodgett

University of Maryland Center for Advanced Study of Language

<atwist@casl.umd.edu>

<jshamoomarx@casl.umd.edu>

<ablodgett@casl.umd.edu>

<jessicabauman@gmail.com>

0 Abstract

This paper presents results of an acoustic analysis comparing voice onset times (VOTs) of Vietnamese coronal stops produced by adult native speakers of Vietnamese and adult speakers of American English. The native speakers consistently produced three separate, non-overlapping VOT ranges, corresponding to pre-voiced, voiceless unaspirated, and voiceless aspirated stops. Most learners, on the other hand, failed to show a distinction between the VOT ranges for the voiceless stops, collapsing the aspirated and unaspirated sounds into one category. Using the consistent patterns present in the native speaker speech as a model of accuracy, we discuss the predominant learner pattern, as well as the range of learner behavior. Finally, we discuss the implications these comparisons have on the development of methods for improving accurate pronunciation among adult learners of Vietnamese.⁶⁹

1 Introduction

The purpose of this investigation is to determine how successfully English-speaking learners of Vietnamese are able to produce native-like voice onset time (VOT) values for stop consonants, through a quantitative acoustic analysis of native and non-native speech. The study results provide acoustic data that inform future hypothesis testing and may help Vietnamese language instructors better understand the types of articulatory mistakes they encounter among native English-speaking students. Vietnamese has a three-way voicing contrast between coronal stops: pre-voiced *ḍ* [d], voiceless unaspirated *t* [t], and voiceless aspirated *th* [t^h].⁷⁰ English has only a two-way contrast, between voiced or pre-voiced *d* [d] and voiceless *t* /t/. While English contains several allophonic variants of /t/, this study is

⁶⁹ We are grateful to the three anonymous reviewers for their extensive comments, including an interest in biographical characteristics of each speaker. We regret that, in order to protect the privacy of our speakers, we are unable to report on their individual profiles, including age and language learning history. We also thank our speakers for their participation, as well as Anita Bowles, Melissa Fox, Henk Haarmann, Pamela Kling, Sue-Sue Luu, Anton Rytting, and Matt Winn, for their assistance on the project.

⁷⁰ Standard Vietnamese orthography is used throughout this paper.

Alina Twist, Jessica Shamoo Marx, Jessica Bauman, and Allison Blodgett. 2010.

Pulling Out All The Stops In Vietnamese: A Delineation Between Native And Non-Native Vietnamese Speech For Voice Onset Time

Journal of the Southeast Asian Linguistics Society 3.2:138-150

Copyright vested in the authors

Received 1/1/10, revised text accepted 28/10/10

concerned only with word-initial stops, which are all considered to be an aspirated variant, [t^h].

2 Methodology

2.1 Participants

Native speaker participants included four Northern dialect speakers (two female, two male) and two Southern dialect speakers (both female). All were originally from Vietnam and had been living in an English-speaking country for 3 to 26 years. They ranged in age from 32 to 73, and all had experience teaching Vietnamese as a foreign language to adults.

Non-native speaker participants included four Northern dialect learners (one female, three male) and six Southern dialect learners (two female, four male). They ranged in age from 26 to 50. All had been studying Vietnamese intensively (i.e., at least 5 hours a day), but for varying lengths of time. Their weeks of training ranged from 10 to 43. All participants resided in the Washington, DC, area at the time of recording.

2.2 Stimuli

The full set of items was selected for a larger word production study investigating non-native speaker production of tones and vowels. The items comprised 160 real words and included 11 vowels: *i* ([i]), *ê* ([e]), *e* ([ɛ]), *u* ([u]), *ư* ([ɯ]), *ô* ([o]), *ơ* ([ʌ]), *o* ([ɔ]), *a* ([a]), *â* ([ɑ]), *ã* ([ã]). The vowels *i*, *u*, *ư*, *ô*, *ơ*, and *a* appeared with all possible tones for each of three syllable types: open (e.g., [ba:]: *ba*, *bà*, *bạ*, *bá*, *bả*, *bã*), stop-final (e.g., [bat]: *bạt*, *bát*), and nasal-final (e.g., [baŋ]: *bang*, *bàng*, *bạnh*, *báng*, *bảng*; [van]: *văn*). The vowels *ê*, *e*, and *o* appeared only in open syllables. Consistent with Vietnamese phonology, the vowels *â* and *ã* appeared in stop-final and nasal-final syllables only. Each target word was contained in a three-word utterance including a color term, as explained below in the Procedure section.

To the extent possible, targets were matched for initial and final segments within syllable type and within vowel. We attempted to maintain consistent consonant place and manner, but, when necessary, sacrificed one or both in the interest of ensuring that all target stimuli were real words.

The subset of targets that provided the syllables for the VOT analysis conformed to the following four criteria:

- 1) The target syllable began with *đ*, *t*, or *th*.
- 2) The target syllable occurred in a non-utterance-initial position.
- 3) The *đ*, *t*, or *th* of the target syllable was followed by a vowel.
- 4) The target syllable was preceded by an open syllable.

Not all of the items recorded were utilized as targets for VOT analysis. Only those items that began with *đ*, *t*, or *th* and were preceded by open syllables were considered for this study. In addition to the original items, three of the color terms - *tím* (purple), *đen* (black), and *đỏ* (red) were included in the VOT analysis. Only those instances of appropriate color terms following open syllables were considered.

The VOT targets contained a range of tones, vowels, and coda segments, which were not controlled or matched across conditions. Table 1 contains a complete list of the target syllables from the VOT analysis. The diacritics (or lack thereof) in each spelling indicate the particular lexical tone for that word. The six tones of Northern Vietnamese are outlined in Figure 1.

Table 1: Target syllables for VOT analysis

	<i>đ</i> tokens			<i>t</i> tokens	<i>th</i> tokens	
All blocks: color terms <i>All participants</i>	<i>đen</i>	<i>đỗ</i>		<i>tím</i>		
Blocks 1 and 2 only <i>All participants</i>					<i>thình</i> <i>thính</i> <i>thình</i> <i>thỉnh</i>	<i>thịnh</i> <i>thơm</i> <i>thủng</i>
Blocks 3 and 4 only <i>Participants 11-16 only</i>	<i>đâm</i> <i>đan</i> <i>đang</i> <i>đăng</i>	<i>đắt</i> <i>đạt</i> <i>đặt</i> <i>đỏ</i>	<i>đờ</i> <i>đơ</i> <i>đơ</i> <i>đột</i>	<i>tâm</i> <i>tăm</i>		

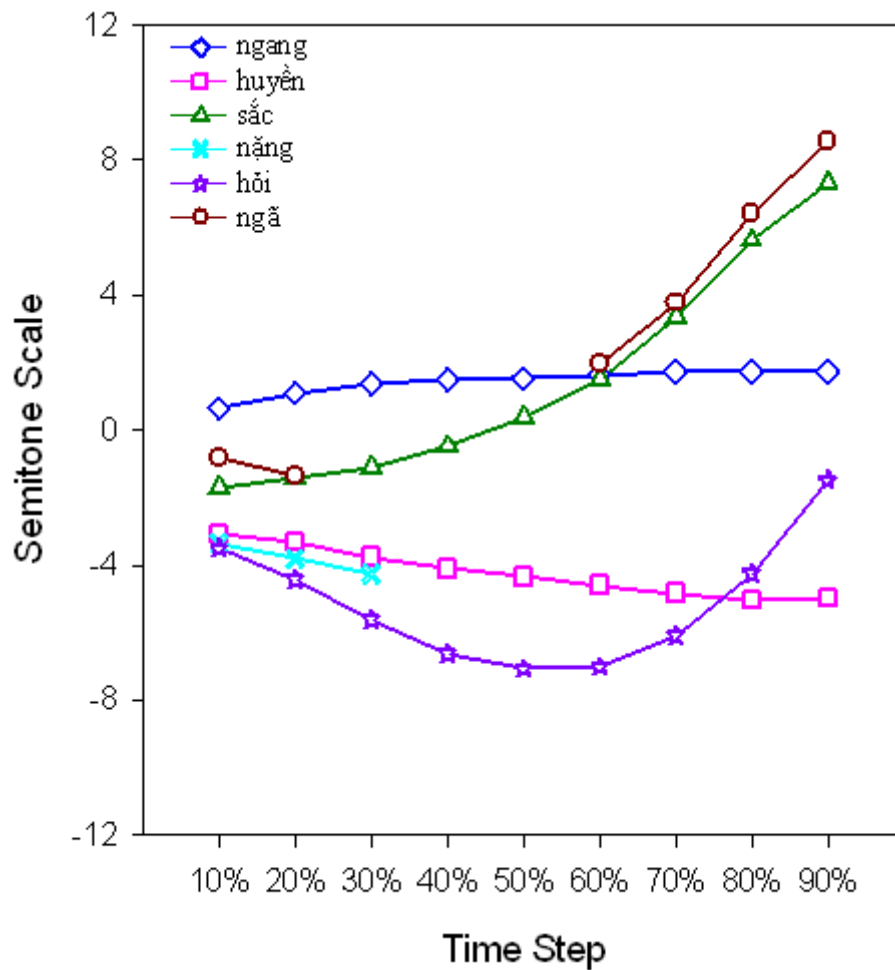


Figure 1: Northern Vietnamese has six lexical tones. Each tone name contains its corresponding diacritic or, in the case of *ngang*, no diacritic. The contours for *ngã* and *nặng* are discontinuous because of glottalization.

2.3 Procedure

Speakers were recorded in a sound-dampened room using Sound Forge 7.0 (22 kHz, 16 bit, mono), a Yamaha 01V96 digital mixing console with no effects settings, and a Neumann TLM 103 microphone.

Participants produced three-word sentences in response to individual words that appeared on a computer screen in red, blue, black, or purple. For example, if the word *bang* appeared in blue, the speaker said "*Từ bang xanh*" ("the word bang is blue"). Participants had access to the written color names as they completed the experimental task. The experiment started with eight practice trials so participants could familiarize themselves with the task. After they completed the practice trials and asked any necessary clarification questions, they proceeded to the main task, which was divided into two or four blocks. Blocks 1 and 2 each contained 102 items with the vowels *i*, *u*, *ɪ*, *ô*, *ơ*, *a*, *â*, and *ã*. Blocks 3 and 4, which were added to the study after data collection had begun, contained 58 items with the vowels *ê*, *ê*, *ơ*, *o*, *a*, *â*, and *ã*. Whereas all 16 participants completed Blocks 1 and 2, only six participants completed Blocks 3 and 4 as well. As a result, all speakers produced 14 utterances containing VOT targets beginning with *th*. Some speakers produced as few as eight *đ* utterances and six *t* utterances. Others produced up to 85 *đ* tokens and 47 *t* tokens. The Analysis section provides additional details regarding the amount of data for each participant.

Items appeared in pseudo-random order such that the vowel, tone, and color of the word always changed from one trial to the next. Items that were paired with *xanh* and *tím* (purple) in Block 1 and Block 3 were paired with *đen* (black) and *đỏ* (red), respectively, in Block 2 and Block 4, and vice versa. Participants thus produced two repetitions of each item, but novel utterances each time. In this self-paced task, participants could repeat any utterance before advancing to the next word. When speakers did repeat, we analyzed only the final repetition.

2.4 Analysis

Target syllables were annotated within their three-word utterances using Praat (Boersma & Weenink, 2008). For each token, the onset of voicing and consonant release were manually marked based on auditory and visual inspection of the waveform, spectrogram, and formant and intensity changes as observed in Praat.

The following criteria were used to determine the onset of voicing:

- 1) The waveform showed a noticeable increase in periodic frequency of energy following a point of very low to no energy corresponding with silence or background noise, or in the absence of such a landmark,
- 2) The spectrogram revealed three consistent vowels formants following a region of aperiodic energy.

Consonant release points corresponded to a sudden burst of noise visible in the spectrogram and waveform. Voice onset and consonant release points were marked at the first positive zero-crossing matching the abovementioned criteria.

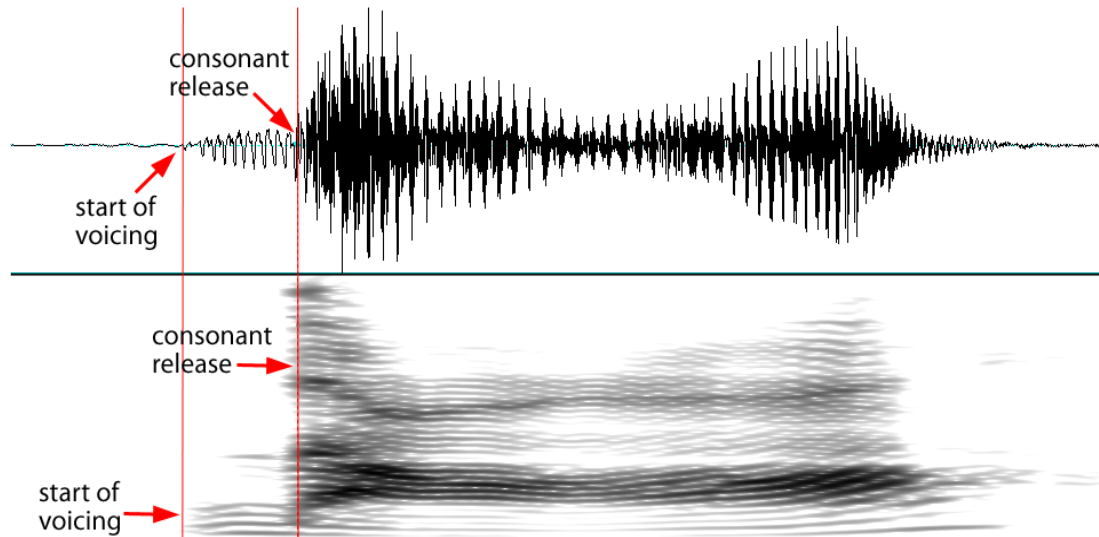


Figure 2: *Demonstration of voice onset time landmarks within the waveform and spectrogram of a native speaker saying đỏ.*

It should be noted that many utterances with target syllables beginning with \bar{d} were removed from the final data set because although they met the characteristics for utterance selection, consonant release and voice onset points could not be identified due to continuous voicing extending from the preceding syllable to the target syllable. The amount of data that was not analyzable varied greatly between speakers. Some speakers produced no continuously voiced tokens, while one speaker produced 80% of her \bar{d} -initial syllables with continuous voicing. Table 2 indicates the amount of data for each speaker.

Table 2: *Amount of data per participant*

Participant	Number of \bar{d} tokens sampled	Number of \bar{d} tokens included in analysis	Number of t tokens sampled and included	Number of th tokens sampled and included
1 Native speaker	8	5	6	14
2 Learner	8	8	6	14
3 Learner	8	8	6	14
4 Learner	8	4	6	14
5 Learner	8	7	6	14
6 Learner	8	7	6	14
7 Learner	8	7	6	14
8 Native speaker	8	5	6	14
9 Native speaker	8	1	6	14
10 Native speaker	8	5	6	14
11 Learner	83	48	44	14
12 Learner	85	72	47	14
13 Learner	83	46	21	14
14 Learner	84	82	43	14
15 Native speaker	80	63	41	14
16 Native speaker	78	20	41	14

3 Results

3.1 Native Vietnamese speakers

Subtracting the time of the consonant release from the time of the onset of voicing yielded a VOT for each target syllable. The VOTs for each stop (\bar{d} , t , th) were then averaged for each speaker. A summary of the averages for the six native Vietnamese speakers is in Figure 3. The native speakers clearly have three distinct categories of VOT. The pre-voiced stop \bar{d} is distinguished by a long negative VOT. The voiceless t is characterized by a slight positive VOT. The aspirated th also has a positive VOT, but it is much longer than that of the plain voiceless t .

For each speaker, the VOTs for target syllables beginning with \bar{d} , t , and th were compared using SPSS to determine whether or not the means of each group were significantly different. The distinctions shown in the figure above are found to be significant in a series of one-way analyses of variance (ANOVAs) assuming unequal variances. For each ANOVA, a Welch statistic was computed along with a typical p value. Welch statistics provide a more conservative measure of differences between means in groups with unequal variances. In no case did the Welch statistic contradict the ANOVA, so the data shown here include only the ANOVA results.

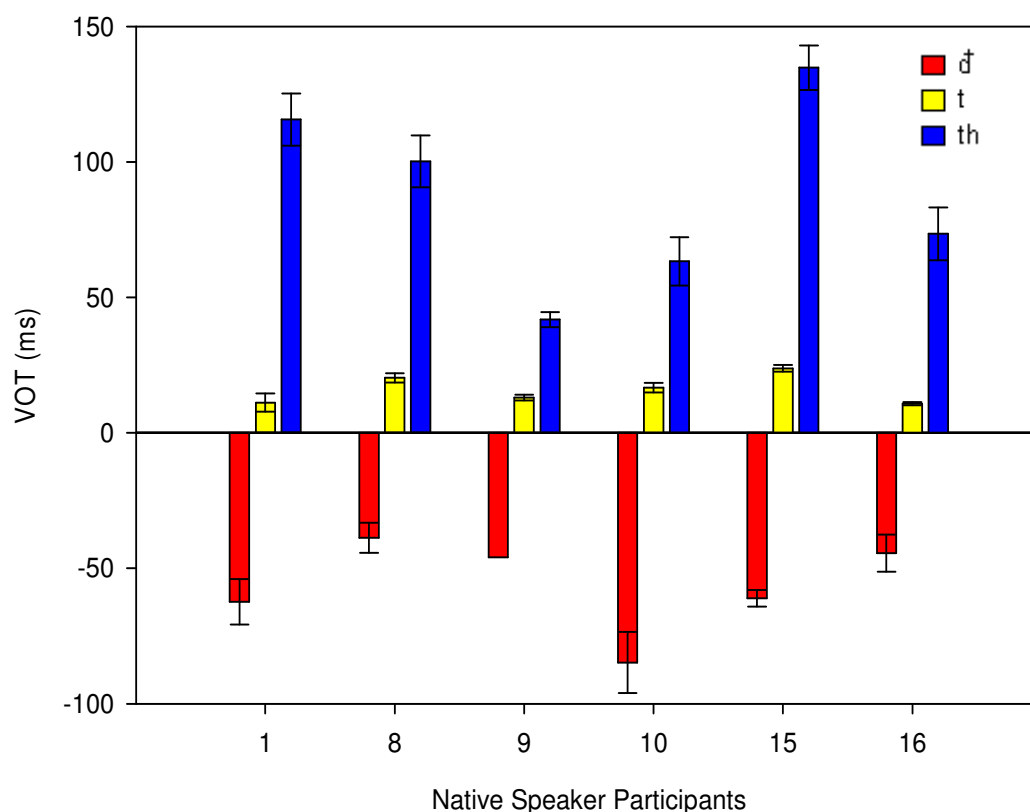


Figure 3: Average VOTs and standard error bars for native speakers of Vietnamese. Speakers are identified by the sequential order in which they were recorded.

In addition, a series of post hoc Tamhane pairwise comparisons, also designed to assess differences between groups with unequal variances, were performed for each speaker. These tests compared VOTs between all three sets of stop pairs in the data: the pre-voiced stop and the voiceless unaspirated; the pre-voiced and the aspirated; and the voiceless unaspirated and the aspirated. For all native speakers, the differences between the three stops were statistically distinct, as shown in Table 3. The table reports the complete results of the ANOVA for each speaker, plus the results of the Tamhane post hoc test for pairs of sounds that are adjacent on the VOT continuum.

Table 3: ANOVA results for native speakers of Vietnamese

Participant	Difference between stops ANOVA	Difference between \bar{d} and t Tamhane's post hoc test	Difference between t and th Tamhane's post hoc test
Speaker 1, female, Northern dialect	$F(2, 27) = 82.1$, $p < .05$	$p < .05$, mean difference = 83.2msec	$p < .05$, mean difference = 104.5msec
Speaker 8, male, Northern dialect	$F(2, 50) = 95.8$, $p < .05$	$p < .05$, mean difference = 63.3msec	$p < .05$, mean difference = 80.0msec
Speaker 9, male, Northern dialect	$F(2, 27) = 118.9$, $p < .05$	$p < .05$, mean difference = 62.4msec	$p < .05$, mean difference = 28.8msec
Speaker 10, female, Southern dialect	$F(2, 27) = 84.5$, $p < .05$	$p < .05$, mean difference = 105.7msec	$p < .05$, mean difference = 46.6msec
Speaker 15, female, Northern dialect	$F(2, 117) = 580.2$, $p < .05$	$p < .05$, mean difference = 86.5msec	$p < .05$, mean difference = 111.0msec
Speaker 16, female, Southern dialect	$F(2, 74) = 112.4$, $p < .05$	$p < .05$, mean difference = 52.6msec	$p < .05$, mean difference = 62.8msec

3.2 English-speaking learners of Vietnamese

English-speaking learners of Vietnamese in this study do not reliably show the same three-way VOT distinction as native speakers. In fact, they generally show only a two-way distinction that collapses two of the Vietnamese segments together. However, not all speakers produce the segments in the same way. Figure 3 shows the average VOTs for non-native Vietnamese learners.

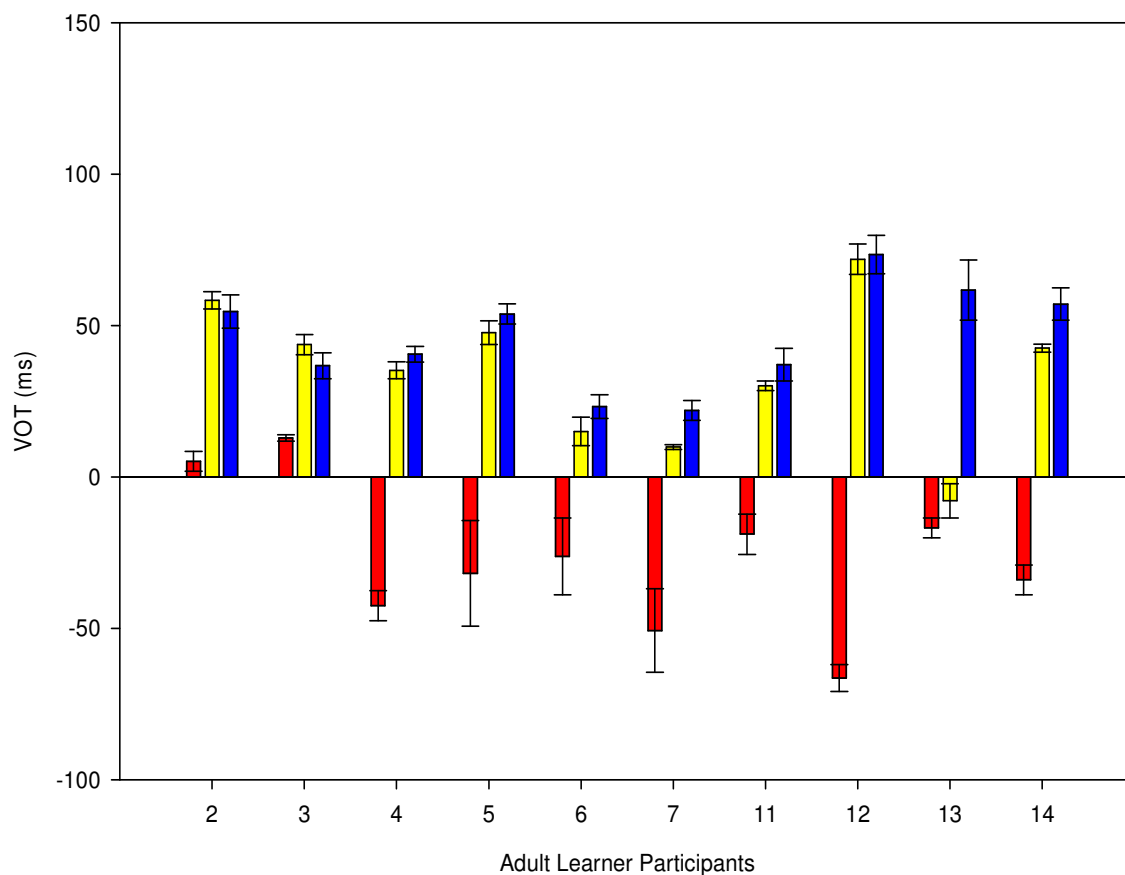


Figure 4: Average VOTs and standard error bars for English-speaking learners of Vietnamese. Speakers are identified by the sequential order in which they were recorded.

Most learners pronounce \tilde{d} with a negative VOT, though the values produced by Learners 6, 11, and 13 do not approach the degree of negativity observed in the native speaker utterances. Learners 2 and 3 consistently fail to prevoice \tilde{d} , resulting in positive, though small, average VOTs.

Utilizing the same univariate ANOVAs and Tamhane pairwise comparisons as for the native speakers, it was determined that most of the Vietnamese learners do not produce a three-way VOT contrast. Most of the students distinguish only the pre-voiced stop from the voiceless stops. The voiceless stops, both aspirated and unaspirated, are collapsed into a single category with overlapping positive VOTs.

A single Vietnamese learner, Learner 13, exhibits the opposite trend, producing a long positive VOT for the aspirated stop, but collapsing the pre-voiced and voiceless stops into a single category with overlapping negative VOT values. Learner 7 is the only non-native learner in the study to produce a three-way VOT contrast between the Vietnamese stops.

Table 4: ANOVA results for English-speaking learners of Vietnamese (*ns* indicates a statistically non-significant result)

Participant	Difference between stops ANOVA	Difference between \bar{d} and t Tamhane's post hoc test	Difference between t and th Tamhane's post hoc test
Learner 2, male, Northern dialect	$F(2, 27) = 28.6$, $p < .05$	$p < .05$, mean difference = 53.2 msec	ns , mean difference = 3.7 msec
Learner 3, male, Northern dialect	$F(2, 50) = 13.8$, $p < .05$	$p < .05$, mean difference = 30.8 msec	ns , mean difference = 7.0 msec
Learner 4, female, Southern dialect	$F(2, 27) = 164.2$, $p < .05$	$p < .05$, mean difference = 86.8 msec	ns , mean difference = 5.3 msec
Learner 5, male, Southern dialect	$F(2, 27) = 25.8$, $p < .05$	$p < .05$, mean difference = 79.5 msec	ns , mean difference = 6.1 msec
Learner 6, female, Southern dialect	$F(2, 27) = 11.2$, $p < .05$	$p < .05$, mean difference = 41.3 msec	ns , mean difference = 5.7 msec
Learner 11, male, Southern dialect	$F(2, 105) = 27.4$, $p < .05$	$p < .05$, mean difference = 47.0 msec	ns , mean difference = 5.9 msec
Learner 12, male, Southern dialect	$F(2, 132) = 260.1$, $p < .05$	$p < .05$, mean difference = 138.6 msec	ns , mean difference = 0.0 msec
Learner 14, female, Northern dialect	$F(2, 138) = 86.1$, $p < .05$	$p < .05$, mean difference = 75.1 msec	ns , mean difference = 14.5 msec
Learner 13, male, Southern dialect	$F(2, 84) = 51.0$, $p < .05$	ns , mean difference = 9.5 msec	$p < .05$, mean difference = 70.0 msec
Learner 7, male, Northern dialect	$F(2, 27) = 27.7$, $p < .05$	$p < .05$, mean difference = 60.6 msec	$p < .05$, mean difference = 12.1 msec

4 Discussion

The results of the analysis show that the English-speaking learners of Vietnamese in this study encounter difficulty producing native-like VOT values. Only one out of the ten learners in the study showed a three-way distinction between pre-voiced, voiceless unaspirated, and voiceless aspirated stops. Even so, the mean difference between his voiceless stops (12.1 msec) was less than half as long as the shortest corresponding difference for native speakers (28.8 msec), and it was much smaller than the difference between his pre-voiced and voiceless stops (60.6 msec). This may indicate that the quality of separation between voiceless unaspirated and voiceless aspirated stops for Learner 13 does not sufficiently mirror the type of distinction in native speaker speech, which may have ramifications for general intelligibility.

Of the learners who produced only a two-way distinction between stops, almost all of them collapsed the VOT ranges of the same two segments: voiceless aspirated and unaspirated stops. One student showed the opposite pattern of clearly distinguishing the voiceless aspirated stop from the voiceless unaspirated and pre-voiced stops. These

opposing patterns may be explained by the prevailing VOT patterns in the learners' native language of English.

Though English speakers make only a two-way distinction between voiced and voiceless stops, they do not all do so in the same manner. VOT values for English /t/ may fall in the range of 40-120 msec, depending on a number of factors including speech rate, following vowel, and stress (Volaitis & Miller, 1992). Establishing a VOT range for English /d/ is not as straightforward. Lisker and Abramson (1964, 1967) maintain that there are two separate possible VOT ranges: that of slightly positive VOTs (under 30msec) and that of negative VOTs as low as -100msec. They argue that VOT ranges for English voiced stops are truly bimodal. Some speakers have a short positive VOT, while others have a consistently negative VOT for these segments. In no case did they observe a speaker producing a mix of positive and negative VOTs. Ryalls, Zipprer, and Baldauff (1997) report effects of gender and ethnicity on VOT values in English. They show that women produced longer positive VOTs for voiceless stops than men. The same study reports a predominance of pre-voicing for English /d/, with women producing shorter negative VOTs than men.⁷¹ However, Ryalls, Simon, and Thomason (2004) and Morris, McCrea, and Herring (2008) conclude that VOT does not vary according to gender. The African-Americans who took part in the 1997 study produced longer pre-voicing for /d/ than the European-Americans, but there was no difference between ethnicities for the voiceless stop. Morris et al. (2008) point out that the many VOT studies that are reported in the literature make use of vastly different experimental utterances, and that factors such as utterance type (e.g., connected speech vs. single words), following vowel context, and tempo may contribute to the conflicting findings.

Whatever the factors that influence the amount and degree of pre-voicing in English, it seems plausible that many English speakers are already familiar with the contrast between negative and positive VOTs. This may explain the predominant pattern of Vietnamese learners who distinguish only between pre-voiced *d̥* and voiceless *t* and *th*. Even the less common pattern exhibited by Learner 13 may be explained in the same way. In this case, the speaker is still making a two-way distinction, but he assigns negative VOTs to both *d̥* and *t*, leaving *th* the only stop in the positive VOT category. Both patterns suggest that speakers are simply extending, albeit imperfectly, their knowledge of categories in their native language to their newly learned language. Several models of speech perception, including the Speech Learning Model (SLM: Flege, 1986, 1990, 1995), the Native Language Magnet model (NLM: Kuhl, 1991, 1992; Kuhl et al., 1992; Grieser and Kuhl, 1989; Iverson and Kuhl, 1996), and the Perceptual Assimilation Model (PAM: Best, 1994, 1995; Best, McRoberts and Goodell, 2001) predict that a speaker's native language shapes that speaker's ability to discriminate contrast in other languages, though the exact nature of the influence is still up for debate. From a practical standpoint, the results also suggest that English-speaking students of Vietnamese do not learn the articulation of coronal stops easily. The native speaker patterns documented here may help instructors explain to students how to modify their productions.

⁷¹ The Ryalls et al. (1997) study differs from Lisker and Abramson's (1964, 1967) work in that positive and negative VOTs are averaged together.

5 New questions to answer

While this study has offered some insight into adult learners' VOT production performance, more exploration of both production and perception processes is needed to provide a thorough understanding of a) why students perform the way they do and b) how their performance impacts their general intelligibility. Only after these questions have been answered will it be possible to determine comprehensively how to improve non-native speaker performance of this parameter.

A general explanation for student performance is that speakers use the tools they already possess, i.e., native phonological knowledge, to process new information from an unfamiliar language. Implicit in this explanation is the assumption that language learners have difficulty comprehending phonological categories that differ from those in their native language. In order to test the relationship between native and non-native categories, data must be collected from both language sources for each speaker. Only then can the native categories be established to compare to the non-native contrasts. In addition, speakers should be probed for their ability to distinguish differences in the language being learned. If they are unable to discern the difference between an aspirated and unaspirated voiceless stop, it is unlikely that they will be able to accurately produce two distinct categories of voiceless stops. Information about native language speech and non-native language comprehension would address the question of why students have difficulty producing Vietnamese stops with the correct range of VOTs.

To investigate how the inaccurate production of stop VOTs impacts intelligibility, native Vietnamese speakers should be polled as to their comprehension of stops produced by Vietnamese learners. If native speakers are able to compensate for substandard VOT values, the effective categorization of stops may not be a priority in language teaching. If, however, VOT values are a primary cue to the quality of a stop that is difficult for native speakers to recover, the issue of how to teach non-native speakers to discriminate and produce native-like stops is an important issue.

6 Conclusions

This study shows that adult second language learners may encounter difficulties with the accurate production of stops if the second language has a different categorization of stops than their native language, based on VOT. The learners in this study apparently drew on the knowledge of their native English in an attempt to capture the contrasts in Vietnamese, resulting in inaccurate production of one or more of the stops. More in-depth analyses of a wider range of data is needed to understand more fully why learners struggle with the Vietnamese VOT distinctions and what pedagogical techniques might be employed to improve performance.

References

- Best, C. (1995). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in Infancy Research* (pp. 217-304). Norwood, NJ: Ablex.
- Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109(2):775-794.
- Boersma, P., & Weenink, D. (2008). Praat: Doing Phonetics by Computer (Version 5.0.06) [Computer Software].

- Flege, J. (1986). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human Communication and Its Disorders*, Volume 2 (pp. 224-401). Norwood, NJ: Ablex.
- Flege, J. (1990). Perception and production: The relevance of phonetic input to L2 language learning. In C. Ferguson & T. Heubner (Eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Philadelphia: J. Benjamin.
- Flege, J. (1995). Second-language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience, Issues in Cross-linguistic Research* (pp. 233-277). Timonium, MD: York Press.
- Grieser, D., & Kuhl, P. (1999). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25: 577-588.
- Iverson, P., & Kuhl, P. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the American Acoustical Society*, 99:1130-1140.
- Kuhl, P. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50(2):93-107.
- Kuhl, P. (1992). Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories. In Y. Tohkure, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 239-264). Tokyo: Ohmsha.
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255:606-608.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops. *Word*, 20:384-422.
- Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10:1-28.
- Morris, R., McCrea, C., and Herring, K. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, 36(2): 308-317.
- Ryalls, J., Simon, M., and Thomason, J. (2004). Voice onset time production in older Caucasian- and African-Americans. *Journal of Multilingual Communication Disorders*, 2: 61-67.
- Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language & Hearing Research*, 40(3): 642-645.
- Volaitis, L., & Miller, J. (1992). Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92:723-735.

BOOK NOTICE

Old Khmer Grammar, Philip N. Jenner & Paul Sidwell, 2010, Canberra: Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University.

I am sorry to admit this, but I have always found Old Khmer a very daunting subject. Despite the availability of excellent dictionaries by Dr. Jenner, Saveros Pou, and Long Seam and many volumes of Old Khmer texts, I have avoided the study of Old Khmer. This volume changes everything. After reading through this Old Khmer Grammar, I'm ready to tackle some texts and see what I've been missing for so many years.

The authors have succeeded in reducing Old Khmer grammar to its basics, covering the phonology, morphology, word classes, and syntax of the phrase, clause, and sentence in 50 pages. The remaining 38 pages contain 50 Old Khmer passages with grammatical analyses, an excerpt from the Vat Samrong Stone with a translation, a bibliography and a glossary (including variant spellings). The section on word classes is the most extensive, containing 30 pages, over a third of the book. Five classes are recognized, verbs, nouns, adverbs, conjunctions and prepositions.

The printing and format are admirably clear, making for great ease of use. The only fault I can find with this book is that, as any good book, it left me with a thirst for more. I hope someday to see a more extensive grammar of Old Khmer with sections on epigraphy and script, history, styles of language and other topics not covered here. This is an excellent book and Doctors Jenner and Sidwell deserve our thanks for making Old Khmer a bit easier to study.

Robert K. Headley
University Park, MD

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. After papers have been reviewed, revised and approved, authors are required to format their papers to style.

Download the template from the homepage at www.jseals.org, and type or paste text directly into this file, and then apply WORD STYLES according to the model.

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, for phonetics use DoulosIPA (or 11 point CharisSIL), for other languages make sure you only use a unicode compliant font.

Line spacing: 14 pt exactly in normal text paragraphs
Set to 'at least' to show larger fonts or graphics

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 table or figure numbers in WORD.

Use footnotes, not endnotes.

If you have any queries, 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 3.2**

December 2010



THE AUSTRALIAN NATIONAL UNIVERSITY



PACIFIC LINGUISTICS

ISSN 1836-6821

