

**JSEALS Special Publication No. 2**

**PAPERS FROM THE  
CHULALONGKORN INTERNATIONAL  
STUDENT SYMPOSIUM ON  
SOUTHEAST ASIAN LINGUISTICS 2017**



**Edited by:**

**Pittayawat Pittayaporn  
Sujinat Jitwiriyanont  
Pavadee Saisuwan  
Bhimbasistha Tejarajanya**

**HAWAII**  
UNIVERSITY of  
HAWAII  
PRESS

© 2018 University of Hawai'i Press  
All rights reserved  
OPEN ACCESS – Semiannual with periodic special publications  
E-ISSN: 1836-6821  
<http://hdl.handle.net/10524/52429>

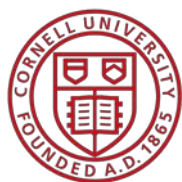


**Creative Commons License**

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

JSEALS publishes fully open access content, which means that all articles are available on the internet to all users immediately upon publication. Non-commercial use and distribution in any medium is permitted, provided the author and the journal are properly credited.

Cover photo courtesy of Pasuree Luesakul.



# JSEALS

Journal of the Southeast Asian Linguistics Society

*Editor-in-Chief*

Mark Alves (Montgomery College, USA)

*Managing Editors*

Nathan Hill (University of London, SOAS, UK)

Sigrid Lew (Payap University, Thailand)

Paul Sidwell (Australia National University, Australia)

*Editorial Advisory Committee*

Marc BRUNELLE (University of Ottawa, Canada)

Kamil DEEN (University of Hawaii, USA)

Gerard DIFFLOTH (Cambodia)

Rikker DOCKUM (Yale University, USA)

San San HNIN TUN (INCALCO, France)

Kitima INDRAMBARYA (Kasetsart University, Thailand)

Peter JENKS (UC Berkeley, USA)

Mathias JENNY (University of Zurich, Switzerland)

Daniel KAUFMAN (Queens College, City University of New York & Endangered Language Alliance, USA)

James KIRBY (University of Edinburgh, Scotland)

Hsiu-chuan LIAO (National Tsing Hua University, Taiwan)

Bradley MCDONNELL (University of Hawai'i at Mānoa, USA)

Alexis MICHAUD (CNRS (Le Centre National de la Recherche Scientifique), France)

Marc MIYAKE (The British Museum)

David MORTENSEN (Carnegie Mellon University, USA)

Peter NORQUEST (University of Arizona, USA)

John D. PHAN (Columbia University, USA)

Pittayawat PITTAYAPORN (Chulalongkorn University, Thailand)

Amara PRASITHRATHSINT (Professor Emeritus of Linguistics, Chulalongkorn University, Thailand)

Alexander D. SMITH (University of North Texas)

Joseph (Deth) THACH (Institut National des Langues et Civilisations Orientales, France)

Alice VITTRANT (Aix-Marseille Université / CNRS-DDL, France)

The Journal of the Southeast Asian Linguistics Society publishes articles on a wide range of linguistic topics of the languages and language families of Southeast Asia and surrounding areas. JSEALS has been hosted by the UH Press since the beginning of 2017.

# CONTENTS

Introduction from the Volume Editors		v
About the Editors and Contributors		vi
Statement from the JSEALS Editor-in-Chief		vii
<b>Papers</b>		
<u>Vectors in Meiteilon Compound Verbs</u>	Potsangbam Chaobimeena	1-9
<u>Origins and Polyfunctionality of Anticipative Expressions: Evidence from Austronesian Taiwan</u>	Yi-Yang Cheng	10-20
<u>Secondary Verbs in Pa-O: A Preliminary Study</u>	Alys Boote Cooper	21-31
<u>The Association of Foreground and Background Information with Verb Stem Alternation in Sizang Chin</u>	Tyler Davis	32-41
<u>The Adjectives in Meiteilon: The Case of So Called Adjective ‘məčə’</u>	Maibam Somobala Devi	42-50
<u>The Polyfunctional Focus Particle Pun in Penang Hokkien: A Contact Perspective</u>	Jia Wen Hing	51-58
<u>A Comparison of Grammaticalization in Shan and Thai</u>	Daniel P. Loss	59-67
<u>Compound Stress in Thai and Effect of Compositionality</u>	Lena Maluleem	68-77
<u>Length Contrast in the Sukhothai Vowel System</u>	Sireemas Maspong	78-89
<u>Serial Verb Constructions in Indonesian: An HPSG Analysis and Its Computational Implementation</u>	David Moeljadi and Viola Ow	90-101

<a href="#"><u>Internally Headed Relative Clauses in Shan</u></a>	Mary Moroney	102-112
<a href="#"><u>Documenting a Fuzzy Tone System: The Case of Muklom Tangsa</u></a>	Mijke Mulder	113-123
<a href="#"><u>The Kachin as participants of an Ethno-Linguistic Area?</u></a>	André Müller	124-135
<a href="#"><u>Unifying Correction and Skepticism in Vietnamese Co</u></a>	Anne Nguyen	136-144
<a href="#"><u>On the Functional Extension of the Thai Form Jan into the Interrogative Marker</u></a>	Noppadon Nuntasukasame, Kachen Tansiri, and Natchanan Natpratan	145-156
<a href="#"><u>Tense, Aspect, and related Markers in Lungmi Rawang</u></a>	Rachel Powelson	157-166
<a href="#"><u>Verbal Agreement in Nocte</u></a>	Syed Iftiqar Rahman	167-176
<a href="#"><u>The Acquisition of Thai Classifiers: Order and Strategies</u></a>	Phetprapa Sangkharam and Kitima Indrambarya	177-185
<a href="#"><u>Stress and Tonal Prominence in Chaoyang</u></a>	Suki S.Y. Yiu	186-194

## INTRODUCTION FROM THE VOLUME EDITORS

Southeast Asian Linguistics is an enchanting field of study due not only to the intriguing diversity of the region but also to the diverse yet congenial network of devoted linguists. For our field to continue growing as an inclusive community with a vibrant academic atmosphere, it is vital to create an environment that promotes academic exchanges among young scholars working on Southeast Asian languages and linguistics. Entitled “Papers from the Chulalongkorn International Student Symposium on Southeast Asian Linguistics,” JSEALS Special Publication No. 2 is a concrete contribution of the first international student conference devoted to the linguistic study of Southeast Asia.

The Chulalongkorn International Student Symposium on Southeast Asian Linguistics (CU-ISSSEAL), which took place during July 8-9, 2017 in Bangkok, Thailand, was organized by the Department of Linguistics, Faculty of Arts, Chulalongkorn University in cooperation with Cornell University, University of Sydney, and Nanyang Technological University. Since its foundation, the linguistics department at Chulalongkorn has been recognized as an important center for Southeast Asian linguistics. Our faculty and students have been actively engaged in research on diverse aspects of Southeast Asian languages. As part of the celebration of Chulalongkorn University’s and the Faculty of Arts’ centenary, the symposium coincided with the “Chulalongkorn Summer School of Southeast Asian Linguistics 2017” and the lecture series “Past, Present, and Future of Southeast Asian Linguistics.” The symposium in particular was designed to serve as a forum for students working on Southeast Asian linguistics from universities around the world. After a screening process by a scientific committee that consists of distinguished faculty members from Chulalongkorn and the three partner universities, 41 out of over 100 abstracts were selected. Attended by over 100 people, the symposium featured 22 oral presentations and 19 poster presentations by 42 students from 25 institutions.

“Papers from the Chulalongkorn International Student Symposium on Southeast Asian Linguistics” is a collection of 19 papers presented at the Chulalongkorn International Student Symposium on Southeast Asian Linguistics. All oral presenters were invited to submit manuscripts for publication in this special volume. All articles were assessed by two anonymous reviewers to ensure high quality. The panel of reviewers consists of a diverse set of linguists working on Southeast Asian languages from many institutions around the world. Although the articles that appear in this volume are student papers, they are all outstanding and reflects the authors’ high potential to become great linguists. Importantly, the number of promising young scholars that contributed to this volume suggests a very bright future for Southeast Asian linguistics as a field.

As organizers of the symposium and editors of this special volume, we would like to thank our partner institutes Cornell University, University of Sydney, and Nanyang Technological University for their academic contribution to the symposium and to this special publication. We are profoundly grateful to all the reviewers whose feedback was all essential in making this volume possible. Furthermore, we would like to express our appreciation to Chula Global Network for their financial support in organizing the symposium and the making of this special volume. Last but not least, we would like to thank Mark Alves, JSEALS editor-in-chief, for his patience and guidance.

### **Editors**

Pittayawat Pittayaporn (Chulalongkorn University)

Sujinat Jitwiriyanont (Chulalongkorn University)

Pavadee Saisuwan (Chulalongkorn University)

Bhimbasistha Tejarajanya (Mahidol University)

## The Editors

PITTAYAWAT PITTAYAPORN (Pittayawat.P@chula.ac.th) is an assistant professor of Linguistics at Chulalongkorn University. He received his PhD in Linguistics from Cornell University in 2009. His research comprises various theoretical issues in historical linguistics and phonology, including tonal change, language contact, and linguistic and population prehistory. He has investigated a variety of Kra-Dai, Austroasiatic, and Mainland Austronesian languages.

SUJINAT JITWIRIYANONT (Sujinat.J@chula.ac.th) is a faculty member in the Department of Linguistics at the Faculty of Arts, Chulalongkorn University, Thailand. He completed his PhD dissertation on tones and tonal coarticulation in Pa-O at the Chulalongkorn University in 2015. His research interests include acoustic phonetics and sociophonetic variation, especially for forensic applications.

PAVADEE SAISUWAN (Pavadee.S@chula.ac.th) is a faculty member in the Department of Linguistics, Faculty of Arts, Chulalongkorn University, Thailand. She received her PhD in Linguistics from Queen Mary University of London in 2016. Her research interests lie in the field of sociolinguistics, focusing on the relationship between language, gender and sexuality.

BHIMBASISTHA TEJARAJANYA (bhimbastha.tej@mahidol.edu) is a researcher at the Faculty of Liberal Arts, Mahidol University, Thailand. He received his PhD in Linguistics from Chulalongkorn University in 2016. His research interests range over morphosyntactic typology and theories as well as diachronic linguistics, especially morphosyntactic change, and grammaticalization.

Authors	Affiliations	Countries
Potsangbam Chaobimeena	Jawaharlal Nehru University	India
Yi-Yang Cheng	University of California, Santa Barbara	USA
Alys Boote Cooper	University of Zurich	Switzerland
Maibam Somobala Devi	University of Hyderabad	India
Tyler Davis	Payap University	Thailand
Jia Wen Hing	National University of Singapore	Singapore
Daniel P. Loss	Payap University, Chiangmai	Thailand
Lena Maluleem	Chulalongkorn University	Thailand
Sireemas Maspong	Chulalongkorn University & Cornell University	Thailand & USA
David Moeljadi, Viola Ow	Nanyang Technological University Singapore, Singapore	Singapore
Mary Moroney	Cornell University	USA
Mijke Mulder	La Trobe University	Australia
André Müller	University of Zurich	Switzerland
Anne Nguyen	National University of Singapore	Singapore
Noppadon Nuntasukasame, Kachen Tansiri, Natchanan Natpratant	Kasetsart University	Thailand
Rachel Powelson	Payap University	Thailand
Syed Iftiqar Rahman	Payap University	Thailand
Phetprapa Sangkharam, Kitima Indrambarya	Kasetsart University	Thailand
Suki S.Y. Yiu	University of Arizona & University of Hong Kong	USA & Hong Kong SAR

## **FROM THE JSEALS EDITOR-IN-CHIEF**

This is the second JSEALS special publication since JSEALS became a University of Hawai'i Press publication in January 2017. The goal of JSEALS special publications is to share collections of linguistics articles, such as select papers from conferences or other special research agendas, as well as to offer a way for linguistic researchers in the greater Southeast Asian region to publish monograph-length works.

In this instance, the volume contains 19 papers primarily by graduate students who had participated in a 2017 symposium of Southeast Asian linguistics at Chulalongkorn University in Bangkok and involving faculty and presenters who are top researchers in the field of Southeast Asian linguistics. The papers cover a full range of linguistic topics, including phonology, morphology, syntax, semantics, ethno/sociolinguistics, historical linguistics, and language acquisition. The languages are from almost all language families in the region: Austroasiatic, Austronesian, Tai-Kadai/Kradai, and Sino-Tibetan/Trans-Himalayan. The languages involved include national languages (e.g. Vietnamese, Thai, and Indonesian), large minority languages (e.g. Shan, Kachin, Chaoyang Chinese, etc.), and small and/or little studied minority languages (e.g. Nocte, Muklom Tangsa, etc.). The papers by these young scholars are relatively short, generally about 10 pages in length, but they are high quality, valuable academic contributions. All have been twice blind-reviewed and thus maintain standards comparable to other research articles in JSEALS. The editors of this volume deserve praise not only for their quality editing but overall support for the authors contributing to this volume.

It is exciting to see such representation by a generation of younger scholars in Southeast Asian linguistics. We are very pleased that JSEALS is able to contribute to the sharing of quality linguistic research in both mainland and insular Southeast Asia.

**Mark J. Alves**

September 30, 2018

Rockville, Maryland



# VECTORS IN MEITEILON COMPOUND VERBS

Potsangbam Chaobimeena  
Jawaharlal Nehru University  
chaochao26@gmail.com

## Abstract

The paper discusses vector verbs in Meiteilon, a Tibeto Burman language. In compound verb constructions (V+V), the verbal roots in the V2 position or the vector give the aspectual meaning of the beginning, continuation, finishing or completion of an action. It may also encode an adverbial meaning of location giving a horizontal or vertical direction of the main verb. Signs of grammaticalization are discernible in Meiteilon vector verbs as delexicalization, or semantic bleaching is involved.

**Keywords:** vector verbs, compound verbs, grammaticalization, semantic bleaching, complex verbs.

**ISO 639-3 codes:** mni

## 1 Introduction

Meiteilon, a Tibeto-Burman language of the north-eastern part of India, is mostly spoken in the state of Manipur and its neighbouring states like Assam, Tripura and also in Myanmar and Bangladesh. Also known as the Manipuri language – the name Meiteilon is a compound – made up of two words: ‘Meitei’ the name of the ethnic group, and ‘lon’ meaning ‘language’, thus together ‘language of the Meitei people’. It is the lingua franca of the state of Manipur. Meiteilon was officially recognized as one of the national languages of India in 1992. A Compound Verb (henceforth CV) is a type of complex predicate where two verbs combine to form a new verbal lexeme consisting of V1 and V2 known as the polar verb and the vector verb respectively. Such a Verb + Verb construction is different from other complex predicate constructions like Noun+ Verb or Adjective+ Verb which are known as Conjoint Verbs. Specifically, a compound verb is a complex verb form comprising the main verb that contains the core meaning while the second component of the compound is semantically delexicalized and bears all the grammatical inflections such as tense, mood, or aspect of the sentence. The vector verb (V2) mostly functions as an auxiliary verb attached to the main verb. It, however, plays a significant role in explicating the meaning of the polar verb. Hence, this type of compound verb is also called an ‘explicator compound verb’. These vector verbs are typically motion-verbs like ‘go’, ‘sit’, ‘come’, ‘give’, ‘take’, etc. in most languages. Evidence showing that a compound verb forms a single lexeme is that all the kinds of grammatical affixes are attached only after the verbs are compounded. The V2 or the vector verb plays a major role in explicating the meaning of the polar or the main verb. According to Abbi and Gopalakrishnan (1991), “[i]n ECV, V2, although homophonous with an independent verb in the language, does not appear with its primary lexical meaning; V2 only occurs in the sequence to mark the main verb V1 for certain ‘grammatical’ features.”

Like in many other languages, the V1 of Meiteilon compound verbs occurs in root/stem form. The name ‘polar’ has been given to it because it contains the core meaning of the complex verb forms. Complex verbs in Meiteilon are formed by derivational elements attached to other existing forms of the verb. The linguistic elements which are used in derivation are *kʰət*, *tʰə*, *sin*, *tʰok*, *pi*, etc.

In Meiteilon, the first component of the verb carries the core meaning of the verb and the second component acts as an explicator in the compound verb. These vector verbs are semantically weak and do not contribute compositionally to the resultant compound verb.

In Meiteilon, there are directional lexemes, namely *kʰət* ‘up’, *tʰə* ‘down’, *tʰok* ‘out’, *sin* ‘in’ and *pi* ‘give’ which can be attached to verbal roots that allow spatial directionality. These lexemes can also be attached to non-directional verbal roots as well as to state verbs. In the latter case, they give certain extended readings which are basically non-directional.

The paper will discuss the various ways in which these directional lexemes are polysemous in Meiteilon and the role of the vector verbs in the compound verb constructions. The five lexemes can have multiple meanings when they occur in the V2 position of the compound verbs. Also, these lexemes can exist independently as lexical items.

### 1.1 Lexical Items

There are certain vector verbs in Meiteilon which are specifically used to code deictic information. These lexemes, nevertheless, can also occur as independent items having their own semantic baggage. They behave like all other independent verbs of the language. They are illustrated in items (1) to (5).

i) *kʰət* ‘lewd’, ‘gesture’ ‘fight’

(1) *əi-nə mit kʰət-li*  
 1.NOM eye gesture/wink-PROG  
 ‘I am winking.’

ii) *tʰa* ‘plant’, ‘slap’

(2) *tomba-nə əŋaŋ-də kʰupak-nə tʰa-i*  
 Tomba-NOM child-ACC palm-INST slap-PRST  
 ‘Tomba slaps the child.’

iii) *tʰok* ‘exit’

(3) *məpan tʰok-u*  
 Out exit-COM  
 ‘Go Out!’

iv) *sin* ‘copy’, ‘change’

(4) *ma-nə fʰurit sil-li*  
 3-NOM shirt change-PROG  
 ‘He is changing his shirt.’

v) *pi* ‘give’

(5) *əi-nə məhak-ki lairik pi*  
 1-NOM 3-HON-GEN book give  
 ‘I give his book.’

### 1.2 Horizontal and Vertical Direction

Most of the lexical items in 1.1 can also encode deictic information when they are the second component of the compound verbs. Information like upward, downward, outward and inward movement is often indicated via the use of these lexemes when they are attached to dynamic verbs unmarked for directionality. Directional meanings are given in capital letters in (6).

i) *kʰət* ‘UP’

(6) *əi-nə mə-bu in-kʰət-li* (Upward Direction)  
 1-NOM 3-ACC push-UP-PRST  
 ‘I push him up.’

Unlike in sentence (1), *kʰət* here behaves differently. *kʰət* in sentence (6) seems to have dellexicalized and encode the directional information of upward movement.

- ii) *tʰə* ‘DOWN’  
 (7) *əi-nə*                      *mə-bu*                      *in-tʰə-i*                      (Downward Direction)  
 1-NOM                      3-ACC                      push-DOWN -PRST  
 ‘I push him down.’

Similarly, *tʰə* when used with a dynamic verb designates the downward movement of the main verb.

- iii) *tʰok* ‘OUT’  
 (8) *əi-nə*                      *mə-bu*                      *in-tʰok-e*                      (Outward Direction)  
 1-NOM                      3-ACC                      push-OUT-PRST  
 ‘I push him out/away.’

Likewise, *tʰok*, which literally means ‘exit’ or ‘out’, when vectorially attached to a dynamic verb gives the meaning of outward direction.

- iv) *sin* ‘IN’  
 (9) *əi-nə*                      *mə-bu*                      *in-sin-li*                      (Inward Direction)  
 1-NOM                      3-ACC                      push-IN-PRST  
 ‘I push him inside.’

Although ‘sin’ in 1.1 (iv) has the meaning of ‘copy’ and ‘change’, here it is functionally different and is used to encode the meaning of inward direction of the main verb. That means these lexemes are similar to adverbs of location. They come in pairs of opposites.

### 1.3 Aspectual Meaning

Some of the lexemes which were used to mark directionality, namely *kʰət* ‘up’, *tʰə* ‘down’, and *tʰok* ‘out’, ‘exit’, are also grammaticalized to give the aspectual meaning of ‘begin’, ‘continue’ and ‘finished’ respectively when attached to non-directional dynamic verbs. However, the other lexeme *sin* is used with verbs to indicate a change in state after a long period of time or gives an inchoative meaning.

#### *kʰət* ‘BEGIN’/’START’

The lexeme *kʰət* when used after a non-directional verb gives the temporal meaning of ‘begin to’ or ‘to start’.

- (10) *ma-nə*    *tʰəbək*    *tou-kʰət-li*  
 3-NOM    work    do-BEGIN-PRF  
 ‘He started to do the work.’

#### ii) *tʰə* ‘CONTINUE’

The lexeme *tʰə* has the temporal sense of continuing an action when used with non-directional verbs.

- (11) *məhak-nə*                      *tomba*    *lak-pə*                      *ŋəi*                      *tʰə-i*  
 3-HON-NOM                      tomba    come-INF                      wait-CONTINUE-PRST  
 ‘He continues to wait for Tomba to come.’

iii) *tʰok* ‘FINISHED’

The lexeme *tʰok* gives the temporal sense of ‘having done something completely or exhaustively’ when used with non-directional dynamic verbs.

- (12) *ram-nə cak ca-tʰok-le*  
 ram-NOM rice eat-FINISH-PRF  
 ‘Ram has eaten rice.’ (Finished)

iv) *sin* ‘COMPLETE’

The lexeme *sin* when attached to some state verbs, provides the meaning of completion.

- (13) *læi-du kən-sil-le*  
 Flower-DET dry-COMplete-PRF  
 ‘The flower has completely become dry.’

### 1.4 Benefaction and Causation

The lexeme *pi* ‘give’ when used vectorially, indicates that an action is performed for the sake of others. This action may be beneficial to the recipient. The lexeme may have undergone the common grammaticalization process and become a benefactive or causative marker when attached to an action verb. In Meiteilon, the vectorial meaning of the verb falls somewhere between a benefactive and a causative.

- (14) *əi nəŋ-gi tʰəbək tou-bi-ge*  
 1SG 2SG-GEN work do-BEN-FUT  
 ‘I will do work for you.’

There is also a permissive function referred to by Matisoff (1991) as a “permissive-causative function”.

- (15) *‘əi yum-də cət-hən-bi-yu*  
 1SG home-LOC go-CAUS-BEN-IMP  
 ‘Please let me go home.’

### 1.5 Attitude

All the following examples fall under the ‘attitudinal category’ discussed in the paper by Abbi and Gopalkrishnan (1991).

- a) The vector verb *pi* ‘give’ in Meiteilon shows disrespect in some cases, as in (16).

- (16) *nəŋ-gi cak-tu əi ca-bi-ge*  
 2SG-GEN rice-DET 1SG eat-GIVE-FUT  
 ‘I shall eat your rice.’ (Disrespectful manner)

The verb *pi* ‘give’ can also be used to mean the act of humility.

- (17) *məhak-nə əŋəŋ-du ləmjiŋ-bi-rəm-mi*  
 3SG.HON-NOM child-DET guide-GIVE-EVID-PRST  
 ‘He is guiding the child.’

b) The lexeme *sin* ‘IN’, ‘copy’, ‘change’ can give the meaning of ‘unexpected surprise’.

- (18) *silā tombā-gi yum-da cəŋ-sin-kʰi*  
 Seela Tomba-GEN house-LOC enter-IN-PRF  
 ‘Seela has entered into Tomba’s house unexpectedly.’

c) *tʰok* ‘OUT’, ‘exit’ can also give the meaning of ‘regret or undesirable situation of the event’.

- (19) *əi-nə kəi hai-dok-kʰre*  
 1SG-NOM what say-OUT-PRF  
 ‘What have I said?’ (Regret)

*əŋaŋ-du-nə əi-gi cak ca-tʰok-kʰre*  
 Child-DET-NOM 1SG-GEN rice take-OUT-PRF  
 ‘The child has eaten my rice.’ (Undesirable)

### 1.6 Adverbial (Manner)

The following are examples of the adverbial (meanings of manner) of the lexemes when they are attached to different forms of verbs:

a) The lexeme *tʰok* ‘OUT’ carries the meaning of ‘carelessness’ when attached to some verbs.

- (21) *tombā-nə tum-tʰok-i*  
 Tomba-NOM sleep-OUT-PRST  
 ‘Tomba sleeps carelessly.’

The same lexeme can also carry the meaning of ‘suddenness’ when used with action verbs.

- (22) *məhak cətnə cətnə lep-tʰok-kʰi*  
 3SG.HON walk-inst walk-inst stop-OUT-PST  
 ‘He stopped suddenly while walking.’

In case of dimensional state verbs which are non-measurable, *tʰok* generally denotes an unnatural or unexpected process whereas *kʰət* denotes a natural process like growth.

- (23) *tombā noi-tʰok-i*  
 Tomba fat-OUT-PRF  
 ‘Tomba has suddenly become fat.’

- (24) *tombā noi-kʰət-li*  
 Tomba fat-UP-PRF  
 ‘Tomba has gradually become fat.’

In case of state verbs of emission, these two lexemes show the corresponding distinction between sudden and gradual processes respectively.

- (25) *tombā-gi ənaba tʰunə pʰə-dok-kʰre*  
 Tomba-GEN sickness soon good-OUT-PRF  
 ‘Tomba has recovered suddenly.’

- (26) *tomba-gi            ənaba    pʰə-gət-lək-li*  
 Tomba-GEN            sickness    good-OUT-FRM-PROG  
 ‘Tomba is recovering gradually.’

b) The lexeme *sin* ‘IN’, ‘copy’, ‘change’ provides ‘volitional meaning’ when it is attached to a static verb. It also gives the meaning of ‘suddenness’ when attached to action verbs.

- (27) *əi-nə    məkʰoi-gi            mərək-tə            kuinə    pʰəm-jil-ləm-i*  
 1SG-NOM 3PL-GEN            middle-LOC            long            sit-IN-EVID-IND  
 ‘I sat for a long time in between them.’

- (28) *əjəj-du    cət-nə            cət-nə    tu-sin-le*  
 Child-DET    walk-INST    walk-INST    fall-IN-PRF  
 ‘The child has fallen suddenly while walking.’

### 1.7 Modality

a) The lexeme *sin* ‘IN’, ‘copy’, ‘change’ when attached to a process verb gives the meaning of ‘emphasis’ or ‘certainty’.

- 29) *məhak-nə            məmə-də            pəisə            pì-sin-kʰi*  
 3SG.HON-NOM    mother-LOC            money            give-IN-PST  
 ‘He gave money to his mother.’ (Certainty)

- (30) *məhak-nə            məmə-də            pəisə            pì-kʰi*  
 3SG.HON-NOM    mother-LOC            money            give-PST  
 ‘He gave money to his mother.’

In the sentence (29), with the use of the lexeme *sin* ‘IN’, the sentence expresses that the money has been definitely given by the boy to his mother. However, in sentence (30), it expresses a simple statement that the boy has given money to his mother. It does not assert certainty on the part of the speaker.

### 1.8 Other Meanings

- i. *lat-tʰok-pə* ‘to worship an evil spirit’
- ii. *tʰə-dok-pə* ‘to offer to an evil spirit’
- iii. *kək-tʰok-pə* ‘to remove an evil spirit from the house’

- (31) *imuŋ-gi    dəsə    fəttə-bə-gi    maiba-nə    lai    lat-tʰok-i*  
 home-GEN    omen    bad-INF-GEN    priest-NOM    God    worship-OUT-PRST  
 ‘The priest worships for the house to protect it from the bad omen.’

The lexeme *tʰok* also provides the meaning ‘doing something openly or publicly’ when used with dynamic verbs.

- (32) *priya-nə            ləmbi-də            kəp-tʰok-i*  
 priya-NOM            road-LOC            cry-OUT-PRST  
 ‘Priya cries on the road.’ (Cry openly in front of people)

b) *sin* has an extended meaning ‘intrusion’ or ‘joining some other persons or objects that are already involved in the event’.

- (33) *soma oja-nə pəra tak-iŋəi-də ŋaŋ-sin-li*  
 Soma teacher-NOM lesson teach-SIM-LOC speak-IN-PRST  
 ‘Soma is speaking while the teacher is teaching the lesson.’

In contrast with the lexeme *tʰok*, *sin* provides the meaning of ‘doing something privately or in a particular place’.

- (34) *mahak-nə cəi-du upu-də ŋa-sin-li*  
 3SG.HON-NOM stick-ACC cupboard-LOC lean-IN-PROG  
 ‘He is leaning the stick against the cupboard.’

c) The lexeme *tʰə* ‘DOWN’, ‘plant’ has the effect of adding intensity to some state verbs to which it is attached.

- (35) *məhak tʰeŋ-tʰə-nə tum-mi*  
 3 SG.HON late-DOWN-INST sleep-PRST  
 ‘He sleeps late.’ (Giving intensity to late)

## 2 Grammaticalization and Polysemy

Some possible scenarios of grammaticalization can be posited in view of the polysemy of the vector verbs in Meiteilon, following the general tendencies proposed by Hopper and Traugott (2003). Their original meanings are most probably the lexical ones described in 1.1.

<i>kʰət</i>	‘lewd’, ‘gesture’, ‘fight’
<i>tʰa</i>	‘plant’, ‘slap’
<i>tʰok</i>	‘exit’
<i>sin</i>	‘copy’, ‘change’
<i>pi</i>	‘give’

A next stage in the process of semantic change could be the extension of these lexical meanings into spatial domains when the vectors are paired with motion verbs used vectorially, as described in 1.2.

<i>kʰət</i>	‘UP’
<i>tʰə</i>	‘DOWN’
<i>tʰok</i>	‘OUT’
<i>sin</i>	‘IN’

These directional meanings are derived both by means of metaphORIZATION and metonymization (Traugott and Dasher 2001). For *kʰət*, the ‘fight’ meaning suggests resistance against antagonistic power, which can be conceptualized as an upward movement. In contrast, an acting force can be metaphorized as a downward, overpowering movement, and thus the ‘DOWN’ meaning of *tʰə* can be accounted for. For *tʰok*, the process of metonymization is involved, as the most salient spatial component of ‘exit’ is to travel ‘OUT’ of somewhere. Similarly, but perhaps less obviously, to ‘change’ implies a transformation ‘IN’ a particular entity’s state, i.e. *sin*. Nevertheless, further research is needed in order to account for the semantic extension of these lexical verbs into manner, attitude, and some other miscellaneous meanings. A mere suggestion to be made here is that the process of (inter)subjectification should be involved (Traugott 2010).

The later stages of the semantic extension of these lexemes hypothetically involve grammaticalization. Their grammatical meanings are as follows.

<i>k<sup>h</sup>ət</i>	‘BEGIN’/‘START’
<i>t<sup>h</sup>ə</i>	‘CONTINUE’
<i>t<sup>h</sup>ok</i>	‘FINISH’
<i>sin</i>	‘COMPLETE’/‘VOLITION’/‘CERTAINTY’
<i>pi</i>	‘BENEFACTION’/‘CAUSATION’

The aspectual meaning ‘BEGIN’/‘START’ of *k<sup>h</sup>ət* can be considered as an extension from the spatial meaning ‘UP’, as an upward movement on the vertical axis can suggest the inceptive phase of an activity. It might be expected that the lexeme with the spatial meaning diametrically opposed to ‘UP,’ i.e. *t<sup>h</sup>a* ‘DOWN’, would also mean ‘FINISH’. This is, however, not the case, since it is *t<sup>h</sup>ok* ‘OUT’ that means ‘FINISH’, and the aspectual meaning of *t<sup>h</sup>a* ‘DOWN’ is in fact ‘CONTINUE’. On second thought, this semantically makes sense, as to move downward can be conceptualized as being engaged in a particular activity, and hence the progressive aspect meaning. Also, if an event is conceptualized as containment, moving out of the containment can be compared to the completion of the event. The opposite of *t<sup>h</sup>ok* ‘OUT’ is *sin* ‘IN’, and it might be expected that *sin* should mean inception or initiation. On the contrary, it in fact means ‘COMPLETE’, which is similar to ‘FINISH’, but entails a change of state as a result of an event. A semantic link can be proposed, connecting ‘IN’ to ‘COMPLETE’ by means of metaphorization. That is, if a state is conceptualized as containment, a movement completely into that containment can mean the resultant change subsequent to an event.

Furthermore, apart from its aspectual meaning, *sin* can also mean ‘VOLITION’ and ‘CERTAINTY’. These two extended modal meanings are different yet related. They are different in that while the former is a dynamic modal, the latter is epistemic in nature (Palmer 2001). That is, ‘VOLITION’ involves a participant-internal force of desire acted out in language, whereas ‘CERTAINTY’ is participant-external and in fact involves the speaker’s own attitude towards the likeliness of the situation. These two modal meanings, however, are unified in that they both involve a high level of modal force (compared to ‘ability’ and ‘probability,’ which are weak modality). This strong nature of ‘VOLITION’ and ‘CERTAINTY’ could explain their extension from ‘COMPLETE’, which suggests an absolute state of a particular event. Last, the grammaticalization pathway of a verb meaning ‘GIVE’, as in this case *pi*, into a benefactive and/or causative marker is well documented in Heine and Kuteva’s (2002) World Lexicon of Grammaticalization. Conceptually, it is probably not difficult to discern the relatedness between these meanings. An event of giving prototypically involves three participants, i.e. the giver, the thing given, and the recipient of the thing. Therefore, it is not surprising that a verb of giving gives rise to ‘BENEFACTION’, marking the recipient of the event. Moreover, when the verb is used causatively, the sense of allowance is involved. That is, allowing somebody to do something involves CAUSATION.

### 3 Conclusion

The paper gives an analysis of how Meiteilon undergoes the process of grammaticalization. This language uses the same lexemes for representing the same lexical forms, directional and the aspectual meaning as well as the other extended usages. The vector verbs in the V2 position may designate the horizontal and the vertical movement of the main verb, or an aspectual meaning, benefaction, causativity, or an adverbial meaning of location. The vector verbs may have undergone the process of grammaticalization in such a way that they have become distinct entities from their lexical counterparts. These vector verbs can occur with most verbs in Meiteilon, with different meanings in different environments.



## References

- Abbi, Anvita. 1991. Semantics of explicator compound verbs, South Asian Languages. *Language Sciences* 13(2):161-180.
- Abbi, Anvita. 1999. What is Indian languages: A case of grammaticalization of explicator verb 'GO'. *PILC, Journal of Dravidic Studies* 9(2):105-120.
- Bhat, D.N.S. and M.S. Ningomba. 1995. *Manipuri grammar*. Mysore: Central Institute of Indian Languages.
- Chaobimeena, Potsangbam. 2013. *A process of grammaticalization of directional lexemes: a case study of Meiteilon*. (Unpublished M. Phil. Dissertation, New Delhi: Jawaharlal Nehru University.
- Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. New York: Cambridge University Press.
- Hopper, Paul J. and Elizabeth Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Palmer, Robert L. 2001. *Mood and modality*. Cambridge University Press, Cambridge.
- Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In *Subjectification, intersubjectification and grammaticalization*, ed. by Davidse, Kristin, 39. Berlin/New York: Walter de Gruyter.
- Traugott, Elizabeth C., & Dasher, Richard B. 2001. *Regularity in semantic change*. Cambridge: Cambridge University Press.

# ORIGINS AND POLYFUNCTIONALITY OF ANTICIPATIVE EXPRESSIONS: EVIDENCE FROM AUSTRONESIAN TAIWAN<sup>1</sup>

Yi-Yang Cheng  
University of California, Santa Barbara  
ycheng@ucsb.edu

## Abstract

Anticipative expressions are verb-modifying expressions that encode anticipative necessity—a type of necessity meaning in which the state of affairs expressed in the sentence is anticipated by the speaker due to his or her knowledge of some past temporal property of the state of affairs concerned. Two types of anticipative expressions have been identified across five Austronesian languages of Taiwan. One type refers to states of affairs that had been planned in the past but are somehow never realized; the other refers to states of affairs that had been regularly realized in the past and whose realization in the present or future is highly anticipated. Anticipative expressions represent a typologically unique set of expressions, as the necessity meaning they express cannot be assigned to established modal categories. More comparative studies are required to determine how widespread anticipative expressions are both within and outside the Austronesian language family.

**Keywords:** modality, typology, language contact, Formosan languages, anticipative necessity

**ISO 639-3 codes:** tay, xnb, trv, tsu, sxr

## 1 Introduction

Areality in modality has emerged as a prominent topic in the literature on language contact, whereby modal expressions may show shared properties that characterize linguistic areas in which one single genealogical origin may not be established to account for the shared properties. Polyfunctionality in modal expressions, for example, may be a widespread property found in a linguistic area. This has been reported in Europe, where polyfunctionality in encoding both epistemic modality and non-epistemic modality is found in the modal expressions of nearly all European languages (Hansen and de Haan 2009, van der Auwera and Ammann 2011). The phenomenon can be exemplified by the English possibility modal *may* and necessity modal *must*: in (1a) and (2a), the modals serve non-epistemic functions as they modify *states of affairs* (henceforth SoA(s)<sup>2</sup>) as enabled or compelled, whereas in (1b) and (2b), the modals serve epistemic functions as they are concerned not with the properties of SoAs,

---

<sup>1</sup> This study was supported by two research projects granted to Dr. Li-May Sung at National Taiwan University (Formation and Fluidity of the Island World, National Taiwan University [NTU 102R3108] and A Typological Study of Austronesian Languages in Taiwan and Their Revitalization, National Science Council [NSC 100-2420-H-002-035-MY3]), which funded fieldwork sessions where primary data on Mayrinax Atayal (Matu'uwal), Kanakanavu, Seediq, and Tsou were collected. I would like to thank Dr. Li-May Sung for introducing me to the study of modality in Formosan languages. I would also like to extend thanks to Dr. Marianne Mithun for all the intellectual resources she offered me while I developed ideas for this paper. Special thanks are given to Dr. Daniel Kaufman and Lilian Li-ying Li for inspirations and fruitful discussions on modality in the languages they specialize in. I am grateful to the language consultants I worked with from Tai'an Township, Namasia, Puli, and Alishan, whose efforts in explaining to me the subtle semantic features of anticipative expressions have taught me so much about their languages and their cultures. My final thanks go to two anonymous reviewers, whose comments have reminded me to attend to the finest details of my own words and to have all my wild ideas in check. All the remaining errors in this paper are, of course, mine.

<sup>2</sup> The term “state of affairs” is used in this paper to refer to both events and states.

but “[the speaker’s] estimation ... of the chances or the likelihood that the SoA expressed in the clause applies in the world” (Nuyts 2016:38).

(1a) *You **may** go home now.*

(1b) *Bob **may** be mistaken about the cause of the accident.* (van der Auwera and Ammann 2011)

(2a) *You **must** go home now.*

(2b) *Terry **must** be from Northumberland.* (van der Auwera and Ammann 2011)

A widely shared grammaticalization pathway for the development of modal expressions is another property of modality that may be characteristic of a linguistic area. Modal expressions of possibility in Mainland Southeast Asia (MSEA), for instance, commonly develop from verbs of acquiring, which may be of common or unrelated lexical sources (Enfield 2003). This can be illustrated by the verb *daj4* ‘to acquire’ in Lao, which has developed functions encoding *ability*, as in (3a), *attainment of a state that involves ability*, as in (3b), and *epistemic modality*, as in (3c).

(3a)<sup>3</sup> *Khòj5                      vaw phaa2                      saa3                      laaw2      **daj4***  
 I                                  speak                      language Lao                      can/ok  
 ‘I can speak Lao.’ (Enfield 2003:78)

(3b) *Khòj5                      **daj4**      pajò      lòj2      nam4*  
 I                                  able      go      swim      water  
 ‘I got to go swimming.’ (Enfield 2003:102)

(3c) *Qaat5      cao      thùj5      cap2      kao      pên3      **daj4***  
 might      IRR      suffer      catch      FOC.PL      be      can  
 ‘It’s possible that you would even be arrested.’ (Enfield 2003:78)

Modal categories that are commonly or rarely found to be encoded by dedicated forms can also distinguish one region from another. English modal auxiliaries, such as those exemplified in (1) and (2), include modal expressions that encode both *possibility* and *necessity*, which is the epitome of modal expressions of Europe, where existence of expressions of both modal categories are commonly found in modal systems (Hansen and de Haan 2009; van der Auwera and Ammann 2011). Offering an overview of modality in the languages of Oceania, Lichtenberk (2016) reports on the prominence of a less commonly reported type of modality—*timitive modality*, which expresses “[a] possible state of affairs as undesirable or even as something (to be) feared” (ibid. 346). The Toqabaqita example shown in (4) involves the clause-initial particle *ada* as the form for expressing timitive modality, but the same concept may be expressed through post-verbal particles or complementation with verbs of fearing in other Oceanic languages (Lichtenberk 2016:346-347).

<sup>3</sup> Glossing abbreviations used in this paper follow the Leipzig Glossing Rules. Additional abbreviations are as follows (note that data cited from other studies retain the glossing abbreviations used by the original authors, which appear in parentheses here): (AD)=adnominal (suffix); ANTICIP=anticipative; AV=agent voice; (CLF)=classifier; CONJ=conjunction; COS=change of state; (FOC)=focus; LNK=linker; LV=locative voice; PV=patient voice; (RDP)=reduplication; (REAL)=realis; RED=reduplication; (SEQ)=sequential; (TIM)=timitive.

- (4) *Ada wane qeri ka riki nau*  
 TIM man that 3SG.SEQ see 1SG  
 ‘[I fear] the man might see me.’ (Lichtenberk 2016:346)

Furthermore, Oceanic languages are also commonly found to lack dedicated forms for the expression of *obligation*, or deontic modality. As can be seen in Manam and Lolovoli in (5) and (6), respectively, the deontic concept of obligation is expressed through the irrealis, whose function is generally to encode future states of affairs.

- (5) *ʔúsi ne-m g-iásaʔ-i*  
 loincloth POSS.CLF-2SG.AD 2SG.IRR-wash-3SG.OBJ  
 ‘Wash your loincloth!/You must/should wash your loincloth.’  
 (Lichtenberk 1983, cited from Lichtenberk 2016:334)

- (6) *Go=ni quqagi vunu go=mo ga-gani*  
 2SG.SBJ=IRR wash.hands then 2SG.SBJ=REAL RDP-eat  
 ‘You (must/will) wash your hands before you eat.’<sup>4</sup>  
 (Hyslop 2001:256, cited from Lichtenberk 2016:334)

Building on findings from studies on areality in modality, the main goal of the present study is to introduce the reader to a modal category—*anticipative necessity*—that has not been widely discussed in the literature, but can be identified across at least five Austronesian languages of Taiwan (also known as the Formosan languages). Anticipative necessity forms a unique category of modality, as the modal meaning ‘should’ it represents does not fit well with established modal categories of necessity (including dynamic and deontic necessity), but is strongly tied to the speaker’s anticipation arising from knowledge of one of the following two past temporal properties in the modified SoA:

- (7a) The past SoA had been planned but did not end up realized.  
 (7b) The past SoA had been regularly realized but ended up unrealized.

The expressions that encode anticipative necessity are referred to as *anticipative expressions* in the discussions that follow, and can be categorized into two types according to their semantic properties: the *planned-in-the-past* type (7a) is found in Mayrinax Atayal (henceforth Matu’uwal<sup>5</sup>), Seediq, and Tsou, and the *regularly-realized-in-the-past* type (7b) in Kananavu and Saaroa. It will be demonstrated that anticipative expressions identified so far unanimously show both a modal function of expressing anticipative necessity, and a non-modal function that indicates some past temporal property of the SoA in question associated with either (7a) or (7b).

The findings from this study will serve as a starting point for a more comprehensive investigation of the properties and distribution of anticipative expressions. More modality-oriented comparative studies conducted within and outside the Austronesian world should determine how widespread anticipative necessity is, and how it may interact with other (grammatical or semantic) categories across the world’s languages.

This paper is organized as follows. In Section 2, the modal and non-modal functions of anticipative expressions are discussed with data from two Formosan languages: Matu’uwal and Kananavu. In Section 3, the modal meaning of anticipative necessity as illustrated by the two languages is

<sup>4</sup> Interestingly, Lolovoli does have a form *mas* for encoding deontic modality, but it was borrowed from Bislama (Melanesian Pidgin spoken in Vanuatu), which borrowed the form from English *must*. Before contact with English, Lolovoli did not have any dedicated forms for expressing obligation or necessity. (Hyslop 2001:256, cited from Lichtenberk 2016:334)

<sup>5</sup> As one of the Atayal dialects commonly known in Mandarin as 汶水泰雅語, Matu’uwal is conventionally referred to in the Formosan literature as Mayrinax Atayal. However, speakers of the language have shown a strong preference for Matu’uwal over Mayrinax, the latter never really used by the community members.

demonstrated to be distinguished from the necessity meanings of two established modal categories—dynamic modality and deontic modality. In Section 4, data from three more Formosan languages (Seediq, Tsou and Saaroa) are included to show that there are at least two types of anticipative expressions that can be observed in Austronesian Taiwan. The summary and conclusion of the present study are provided in Section 5.

## 2 Anticipative expressions: non-modal and modal functions

In this section, two functions that are typically served by anticipative expressions—a non-modal function of temporal modification, and a modal function of expressing anticipative necessity, are discussed with data from two Formosan languages—Matu’wual and Kanakanavu.

In Matu’wual (Atayalic), the anticipative expression *naki* serves as a verb-modifying expression that occurs in clause-initial position. In its non-modal function, *naki* serves to modify the SoA expressed in the sentence as having been planned by the agent participant in the past, but somehow never realized. In other words, the SoA expressed in the sentence and modified by *naki* (*Kaynu’ shooting boars* in (8a) and *Limuy buying salt* in (8b)) are presented as non-occurring at the time when the sentence is uttered but known by the speaker to have been planned in the past.

- (8a) *Naki*            ‘i’    *s<um>bu’*    *cu’*    *bawwak*            ‘i’    *Kaynu’*  
 ANTICIP            LNK    <AV>shoot    ACC    boar                NOM    Kaynu’  
 ‘Kaynu’ was going to shoot boars (but somehow didn’t).’

- (8b) *Naki*            ‘i’    *m-<in>baynay* *cu’*    *timu’*    ‘i’    *Limuy*  
 ANTICIP            LNK    <PFV>AV-buy    ACC    salt                NOM    Limuy  
 ‘Limuy was going to buy salt (but somehow didn’t).’

In (8), *naki* serves a specific past-temporal modification function, as the speaker objectively describes the non-occurrence of some SoA that had been planned in the past, and no modal meaning is available in the interpretation. However, it can also be used as a modal expression whereby the SoA in the sentence is specified as non-occurring, but nonetheless strongly anticipated by the speaker due to his or her knowledge that it had been planned in the past. This is shown in (9), where the sentences containing *naki* in (8) are situated in a context where the above-mentioned anticipation is foregrounded (by a follow-up sentence headed by the adverb *aqih gi’* ‘unfortunately’). Here, *naki* is interpreted modally as ‘should’ and serves to mark the modified SoA as non-occurring, but it is anticipated to occur due to knowledge of the planned-in-the-past nature of the SoA in question.<sup>6</sup>

- (9a) *Naki*            ‘i’    *s<um>bu’*    *cu’*    *bawwak*            ‘i’    *Kaynu’*  
 ANTICIP            LNK    <AV>shoot    ACC    boar                NOM    Kaynu’

*Aqih*    *gi’*    *ma-tubah*    *Kiya’*    *a*    *waw=niya’*  
 bad    CONJ    AV-return    EXIST    NOM    matter=3SG.GEN

‘Kaynu’ should shoot boars (because he was going to). Unfortunately, he returned. He had something else to do.’

<sup>6</sup> In other words, in (9a), the speaker anticipates Kaynu’ to shoot boars because he or she knows that Kaynu’ had planned to do so, but didn’t, whereas in (9b), the speaker anticipates Limuy to have bought salt because he or she knows it to have been the plan in the past, although it never got carried out.

(9b) *Naki*                    ‘i’        *m-<in>baynay*                    *cu’*    *timu’*    ‘i’        *Limuy*  
 ANTICIP                    LNK        <PFV>AV-buy                    ACC        salt        NOM        Limuy

*Aqih*    *gi’*            *ungi’-an=niya’*                    *la*  
 bad        CONJ        forget-LV=3SG.GEN                    COS

‘Limuy should have bought salt (because she was going to). Unfortunately, she had forgotten about it.’

Much like *naki* in Matu’uwal, the anticipative expression *masiraru* in Kananavu (Saarua-Kananavu<sup>7</sup>) also serves as a verb-modifying expression occurring in clause-initial position and has both non-modal and modal functions. In its non-modal function, *masiraru* indicates that the modified SoA has been regularly realized by the agent participant in the past, to the extent that its realization at any time period would be strongly anticipated. The past regular realization of the SoA may be attributed to some inherent property of the participant involved, as in (10a), or to some routine that has been followed by the participant, as in (10b).

(10a) *Masiraru*                    *putukikio*                    *Pani*  
 ANTICIP                    work.AV                    Pani  
 ‘Pani works (not surprisingly, since he has always been a diligent worker).’

(10b) *Masiraru*                    *tia*            *tarasangai*                    *nguain*  
 ANTICIP                    FUT        rest.AV                    3SG  
 ‘They will (definitely) take some rest (as it is time that they rest according to their schedule).’

Used non-modally as in (10) above, *masiraru* is used to both objectively describe some SoA as actually occurring, or certain-to-occur, and at the same time highlight its status as having been regularly realized in the past. Again, no modal meaning is available here. However, *masiraru* can also be used to indicate that the SoA in the sentence is surprisingly non-occurring despite its regular realization in the past, and that it is anticipated by the speaker to occur due to his or her knowledge of its past regular realization<sup>8</sup>. This may involve unexpected violation of the agent’s behaviour pattern, as in (11a), or regularly followed routines that were discontinued, as in (11b). In either case, there is a mismatch between the reality and the speaker’s anticipation out of knowledge of past regularity, whereby *masiraru* is interpreted as expressing the necessity meaning ‘should’.

(11a) *Masiraru*                    *po-kari-kari*                    *Kanakanavu*                    *nguain*  
 ANTICIP                    utter-speech-RED                    Kananavu                    3SG  
 ‘S/he should speak Kananavu (since s/he has always spoken it).’

(11b) *Masiraru*                    *kaisisi*                    *sua*                    *Kanakanavu*  
 ANTICIP                    river.ceremony                    NOM                    Kananavu  
 ‘The Kananavu people should hold the river ceremony (as it’s an old tradition).’

Comparing the functions of *naki* and *masiraru* from (8) through (11), the reader may find that there are two important differences distinguishing the two expressions, which motivate the categorization as mentioned in (7). First, the two expressions differ in the past temporal property of the SoA specified (planned-in-the-past for *naki* vs. regularly-realized-in-the-past for *masiraru*). Second, the reality status

<sup>7</sup> The two languages have been shown to be closely related, but different subgroupings have been proposed. See, for example, Chang (2006), Ross (2009) and Zeitoun and Teng (2016) for relevant discussions.

<sup>8</sup> In other words, the speaker thinks it’s necessary for the third person singular referent in (11a) to speak Kananavu because the referent is known by him or her to have always been speaking the language, but somehow stopped doing so. Similarly, in (11b), the speaker knows the Kananavu people to have always held the river ceremony. Since they somehow stopped doing so, the speaker now anticipates them to continue holding the ceremony.

of the SoA they modify is also different. In Matu’uwal, the SoA remains non-occurring regardless of which function is served by *naki*. That is, the difference between (8) and (9) lies only in whether the speaker’s anticipation is additionally expressed. Whereas in (8), the use of *naki* is strictly descriptive in temporally characterizing the SoA as planned in the past but non-occurring, the use of *naki* in the contexts of (9) involves an additional modal element of the speaker expressing anticipation that the planned-in-the-past SoA “should” occur despite its non-occurring status. In contrast, the two functions of *masiraru* in Kanakanavu are accompanied by SoAs whose reality statuses are different, in addition to the modal meaning ‘should’ available in (11). In (10), the SoAs involved are presented as actually occurring, or certain to occur, regardless of their past regular realization. In (11), on the other hand, the SoAs involved are non-occurring, therefore anticipated by the speaker to occur due to their past regular realization.

### 3 Anticipative necessity vs. dynamic/deontic necessity

As can be seen in the previous section, *naki* in Mayrinax Atayal and *masiraru* in Kanakanavu are not expressions that exhibit multiple, unrelated functions. Instead, in each expression, the modal function of expressing anticipation as necessity is strongly tied to the non-modal function of specifying some past temporal property of the SoA in the sentence. Nevertheless, the type of necessity expressed by anticipative expressions—referred to as *anticipative necessity* in this paper—does not appear to be assignable to existing modal categories of necessity established in the typological literature on linguistic modality. In this section, this is demonstrated by comparing anticipative necessity to necessity meanings belonging to the established categories of dynamic modality and deontic modality, and by further examining the expression of necessity meanings other than anticipative necessity in Matu’uwal and Kanakanavu.

To begin with, anticipative necessity is clearly distinguished from necessity notions belonging to the category of *dynamic modality*, in which the expression of necessity is grounded in the notions of *need* or *root necessity*. The notion of *need* represents a specific type of necessity that is internal to a participant engaged in the SoA (hence its other commonly seen labels *participant-internal necessity* (van der Auwera and Plungian 1998), *participant-inherent necessity* (Nuyts 2016), etc.) which does not have anything to do with the speaker’s anticipation:

- (12a) *I need to hear a good loud alarm in the mornings to wake up.* (Bybee et al. 1994:177)  
 (12b) *Boris needs to sleep ten hours every night for him to function properly.* (van der Auwera and Plungian 1998:80)

The notion of *root necessity* refers to the existence of restraining circumstances that necessitate the occurrence of the SoA. This type of necessity has also been called *participant-external necessity* (van der Auwera and Plungian 1998) as it involves circumstances external to the participants engaged in the SoA. Again, the speaker’s anticipation arising from past temporal knowledge of the SoA does not come into play in the necessity expressed.

- (13a) *To get to the station, you have to take bus 66.* (van der Auwera and Plungian 1998:80)  
 (13b) *Clay pots must have some protection from severe weather.* (Coates 1983:35)

Another modal category that is commonly associated with the expression of necessity is *deontic modality*, which involves indication of the existence of *moral desirability* of the SoA expressed in the sentence (Nuyts 2016:36). Typically, this type of necessity refers to *obligations* or *ethical norms* that “oblige” the participant to engage in the SoA (van der Auwera and Plungian 1998:81). Deontic necessity has been widely exemplified by English modal auxiliaries *must* and *should* in the literature:

- (14a) *All students must obtain the consent of the Dean of the faculty concerned before entering for examination.* (Coates 1983:35)  
 (14b) *I just insisted very firmly on calling her Miss Tillman, but one should really call her President.* (Coates 1983:59)

While obligations and ethical norms may also be analyzed as a type of speaker anticipation<sup>9</sup>, they typically refer to much more generalized authorities or rules that do not necessarily have to coincide with the speaker’s knowledge of the past temporal property of the specific SoA concerned. For example, in (14a) the modal meaning of *must* does not stipulate that the SoA in question (*all students obtaining the consent*) has to have been planned or regularly realized by the participants (i.e. *students*) in the past in order for the necessity to hold. Instead, it is the established norms concerning the SoA (that students obtain consent from some authority for the process of examination) that makes its occurrence necessary. In the same vein, the modal meaning of *should* in (14b) does not indicate the SoA concerned (*calling her President*) to have been specifically planned or regularly realized in the past. In fact, it is even made clear in this sentence that the speaker him/herself is a violator of the deontic rule.

In Matu’uwal and Kananavu, necessity that can be considered as belonging to dynamic or deontic modality is expressed by forms different from the anticipative expressions discussed in Section 2. Unlike English, Matu’uwal does not have a specific modal expression for *need*. Instead, *need* is implied through the desiderative expression *sumi’uwa* ‘desire’ when the discourse context implies the existence of physical needs inherent in the participant involved.

- (15) *Ma-’uway*      *ku*      *hihihi’=mu*      *ru’*  
 AV-sick      NOM      body=1SG.GEN      CONJ
- s<um>i’uwa’*      *i’*      *ma-nabuwaw*      *cu’*      *qusiya’*  
 <AV>desire      LNK      AV-drink      ACC      water  
 ‘I am sick today, and I want/need to drink some water.’

There is, however, an expression *asi ki* in Matu’uwal, whose meaning ‘must’ may cover both root necessity (16a) and deontic necessity (16b), which refer to restraining circumstances and social norms/obligations, respectively, as the necessity expressed. As the speaker’s knowledge of the planned-in-the-past nature of the SoA does not come into play, *asi ki* ‘must’ does not express anticipative necessity, and is therefore distinguished from *naki*<sup>10</sup>.

- (16a) *Asi*      *ki*      *qal’-an=su’*      *ku*      *xuil=su’*  
 must      LNK      keep.in.door-LV=2SG.GEN      NOM      dog=2SG.GEN
- mutux*      *k<um>atcu’*      *cuquliq*  
 tend      <AV>biteACC      person  
 ‘You must keep your dog inside the house. It bites people.’

<sup>9</sup> For example, in the following sentences containing the English auxiliary *must*, the states of affairs may also be considered as “anticipated” by the speaker.

(a) *You must have respect for other people’s property.* (Coates 1983:34)

(b) *Clay pots must have some protection from severe weather.* (Coates 1983:35)

<sup>10</sup> If the necessity meaning in (15a) was anticipative necessity, there would be a presupposition that the speaker knew the dog to have specifically planned to bite people, but never really succeeded (or that there had been a plan for the dog to bite people, but the plan never really worked). In the same vein, in (15b), a similar presupposition would be available (the women in the past had planned to weave, but never really did) if *naki* was used.



- (16b) *Asi ki t<um>inun ku ka-kanayril cu' cunga' gi'*  
 must LNK <AV>weave NOM RED-woman ACC past CONJ
- yani ku gaga' cunga'*  
 that NOM Atayal.convention past
- ‘Women in the past had to weave because that was an ancient Atayal tradition.’

In contrast, Kanakanavu does not appear to show any dedicated modal expressions for encoding dynamic or deontic necessity meanings. Instead, the future marker *tia* and the adverbial expression of certainty *cucuru* ‘really’ may be used together in a sentence to imply need or root necessity when the agent of the sentence is the speaker.<sup>11</sup>

- (17) *Cucuru tia m-o-kusa Taibei sua iiku*  
 really FUT AV-move-toward Taipei NOM 1SG
- ‘I am really going to Taipei.’ (I need to/must go to Taipei.)

In this regard, anticipative necessity can be considered a unique type of necessity not only crosslinguistically but within the two Formosan languages that show anticipative expressions. On the one hand, anticipative necessity cannot be readily categorized as either dynamic necessity or deontic necessity, as the former refers to more specific notions of *need* and *root necessity* as the compelling circumstances, whereas the latter refers to more generalized notions of *obligations* or *ethical norms*. On the other hand, anticipative expressions in Matu’uwal and Kanakanavu are also clearly distinguished from other types of expressions/ constructions conveying necessity. In Matu’uwal, *need*, *root/deontic necessity*, and *anticipative necessity* each have a dedicated expression/construction. The anticipative expression *masiraru* in Kanakanavu, however, appears to be the only modal expression dedicated to expressing necessity.

#### 4 The distribution of anticipative expressions across Austronesian Taiwan

So far, examples of anticipative expressions have been examined with data from two languages—Mayrinax Atayal and Kanakanavu. Nevertheless, the distribution of anticipative expressions does not appear to be limited to just the two Formosan languages discussed above, but has a much wider distribution within Austronesian Taiwan. Specifically, the planned-in-the-past (7a) type of anticipative

<sup>11</sup> Our Kanakanavu language consultants have consistently indicated that

- (i) the necessity meaning in this construction cannot be translated as *yinggai* ‘should’ in Mandarin, or refer to social norms or obligations;
- (ii) the expression closest to Mandarin *yinggai* ‘should’ in meaning is the anticipative expression *masiraru*, which, however, has to presuppose that the SoA (e.g. *the speaker going to Taipei* in (17)) had been regularly realized in the past; and that
- (iii) such a presupposition is not available in the meaning of (17).

They have also intuited that there are no ways to specifically express the fact that a SoA is deemed necessary due to social norms or obligations. Whether deontic modality indeed does not have any dedicated expressions in Kanakanavu requires more systematic investigation into the expression of related meanings in the language. This does not mean, however, that the concept of obligation does not exist or could not be conveyed in the language. Verbs in the imperative, for example, can be readily used to imply that the hearer is demanded to perform an action out of obligations, as exemplified below:

- (iv) *M-oran-a kangvang inia.*  
 AV-help-IMP all 3SG.OBL  
 ‘Help him/her, everyone!/everyone should help him/her!’

Both the imperative and future marking appear to be among the common strategies for expressing obligation in Oceanic languages (Lichtenberk 2016:334).

expressions is further observed in Seediq and Tsou, and the regularly-realized-in-the-past (7b) type, in Saaroa<sup>12</sup>.

In Seediq, a language closely related to Matu’uwal, both classified as Atayalic languages, anticipative necessity is expressed by the anticipative expression *naka*. Similarly, *naka* may function as a non-modal modifier indicating the SoA in the sentence as having been planned in the past but never ending up being realized, as in (18a). It may also be used to express anticipative necessity, as in (18b).

(18a) *Ma=su*                      *ini*    *durun*    *han?*    *Naka=su*                      *krngiy-an*  
 how.come=2SG.NOM    NEG    tell    first    ANTICIP=2SG.NOM    reserve-LV  
 ‘Why didn’t you tell (me) in advance? (I) would have reserved (one) for you/was going to reserve (one) for you (but didn’t).’

(18b) *Naka=ku*                      *naq*    *m-eyah*                      *ciga*  
 ANTICIP=1SG.NOM    only    AV-come                      yesterday  
  
*Yah-un=ku=na*                      *s<m>gila*                      *de-Pawan*  
 come-PV=1SG.NOM=3SG.GEN    <AV>tarry                      PL-PN  
 ‘I should have come yesterday. (I wanted to, but) I was held up by Pawan and them.’

A similar pattern is found in Tsou (Tsouic), which is distantly related to the Atayalic languages. Here, two forms *nte* and *nto* are identified as anticipative expressions, in which there is a semantic distinction based on the difference in temporal distance between the planned-in-the-past SoA and the utterance time, as can be seen in (19a) and (19a’). Again, there is a modal function served in expressing anticipative necessity, as in (19b).

(19a) *Nte-’o*                      *hioa.*    *Mi-ta-cu*                      *hioa*  
 ANTICIP-1SG.NOM    work    AUX.AV-3SG.NOM-PFV    work  
 ‘I was going to do (the work). (But) he did it (instead).’

(19a’) *Nto-’u*                      *hioa.*    *Mi-ta-cu*                      *hioa*  
 ANTICIP-1SG.NOM    work    AUX.AV-3SG.NOM-PFV    work  
 ‘(Long ago,) I was going to do (the work). (But) he did it (instead).’

(19b) *Nte-’o/nto-’u*                      *bonu*    *to*    *s’os’o*  
 ANTICIP-1SG.NOM    eat    OBL    medicine  
 ‘I should take some medicine (since I was going to but didn’t).’

In Saaroa, a language that has been shown to be closely related to Kanakanavu (see Chang 2006; Ross 2009; Zeitoun and Teng 2016), the anticipative expression *mairahlu*<sup>13</sup> may serve as a non-modal modifier in characterizing a SoA as having been regularly realized in the past (20a), or as a modal modifier in expressing anticipative necessity (20b).

(20a) *Mairahlu*                      *ka*    *ma-maini*                      *makahlumuhlumu*  
 ANTICIP                      LNK    RED-small                      wet.bed  
 ‘Kids just wet the bed. (That’s just what they have been doing.)’

<sup>12</sup> Another Formosan language, Bunun, was included in an earlier version of this paper (presented at Chula-ISSSEAL2017) as also exemplifying the (7b) type. Although it does involve an anticipative expression *nau*, the specifics of its non-modal and modal functions turned out to be more complicated than previously thought, and so is now excluded in this paper. I thank Lilian Li-ying Li for discussing her Bunun data with me in considerable detail.

<sup>13</sup> The Saaroa data are retrieved from the *Indigenous Languages Online Dictionary* compiled by Taiwanese indigenous activists and/or scholars and funded by the Council of Indigenous Peoples (CIP), Taiwan <<http://e-dictionary.apc.gov.tw/Index.htm>>. Glossing is based on the author’s analysis of the data.

- (20b) *Mairahlu*=*mu*            *mauhlauhlangu*            *salia*  
 ANTICIP=2PL.NOM        remain                            home  
 ‘You should stay home (since you always have).’

Generalizing from the data across the five Formosan languages discussed above, it can be observed that at least two types of anticipative expressions can be identified. Both types of expressions exhibit the modal function of expressing necessity from the speaker’s anticipation, which arises from some temporal property of the SoA indicated by the no-modal function. The anticipative expressions in Matu’uwal, Seediq, and Tsou involve the past temporal property of the modified SoAs as “planned in the past but never realized”. In contrast, Kanakanavu and Saaroa show anticipative expressions based on temporal modification of the SoA as “regularly realized in the past”.

## 5 Summary and concluding remarks

To summarize, the present paper has discussed the properties of anticipative expressions as evidenced in five Formosan languages. It has been shown that all the anticipative expressions identified so far share both a non-modal function of expressing temporal modification and a modal function of expressing necessity. Anticipative expressions are distinguished from modal expressions that encode dynamic and/or deontic necessity. The type of necessity they express is referred to as anticipative necessity, which is strongly tied to the speaker’s anticipation arising from his or her knowledge of some past temporal property of the modified SoA. The past temporal property specified by anticipative expressions can be categorized into two types. The first type involves the SoA modified as having been planned in the past, but never carried out, which is found in Matu’uwal, Seediq, and Tsou. The second type modifies the SoA as having been regularly realized in the past, to the extent that its realization in the present or future is strongly expected. Kanakanavu and Saaroa have been shown to exhibit this latter type of anticipative expression.

While the present study has argued for the unique status of anticipative necessity in the typology of modal meanings and its distribution in five Formosan languages, more descriptive and comparative studies are still required to determine its status both within and outside the Austronesian world. On the one hand, there might be additional types or patterns of polyfunctionality<sup>14</sup> observed in languages with anticipative expressions, which may inform historical linguistics or areal linguistics<sup>15</sup>. On the other hand, there might be other (grammatical or semantic) categories interacting, or associated with anticipative necessity in interesting ways<sup>16</sup>. In any case, this paper has served as a starting point for a more comprehensive investigation on anticipative necessity, both typologically and areally.

## References

- Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Chang, Henry Y. 2006. Rethinking the Tsouic subgroup hypothesis: A morphosyntactic perspective. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-Kuei Li on his*

<sup>14</sup> Cheng (2013) has described other functions of *naki* in Matu’uwal. Cheng and Sung (2015) discussed how anticipative expressions in Seediq and Tsou may also express dynamic/deontic necessity.

<sup>15</sup> The reader may have observed that there appears to be formal similarity across the anticipative expressions in each type proposed in this paper: the planned-in-the-past type expressions all begin with the form *n(a)-*, whereas the regularly-realized-in-the-past type expressions seem to show formal similarity in the final syllables. However, as pointed out by an anonymous reviewer, the evidence is still very thin as to whether the expressions in each type of anticipative expressions are historically related to each other. It is also pointed out that whether areal diffusion can account for the distribution still requires more comparative research both within and outside Taiwan.

<sup>16</sup> A category somewhat similar to anticipative is *frustrative*, which expresses “the non-realization of some expected outcome implied by the proposition expressed in the marked clause” (Overall 2017:479) and may be considered an areal feature characterizing Amazonian languages (ibid. 501-504).

- 70th birthday, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 565-83. Taipei: Academia Sinica.
- Cheng, Yi-Yang. 2013. *Modality in Mayrinax Atayal*. Taipei: National Taiwan University. (MA Thesis.)
- Cheng, Yi-Yang and Li-May Sung. 2015. The expression of modality in Kanakanavu. *Oceanic Linguistics* 54(1):17-53.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Enfield, Nick J. 2003. *Linguistic epidemiology: semantics and grammar of language contact in Mainland Southeast Asia*. London and New York: Routledge and Curzon.
- Hansen, Björn and Ferdinand de Haan. 2009. Concluding chapter: modal constructions in the languages of Europe. In *Mood in the languages of Europe: a reference work*, ed. by Hansen, Björn and Ferdinand de Haan, 511–559. Berlin: Mouton de Gruyter.
- Lichtenberk, Frantisek. 2016. Modality and mood in Oceanic. In *The Oxford hand book of modality and mood*, ed. by Nuyts, Jan and Johan van der Auwera, 330-365. Oxford: Oxford University Press.
- Nuyts, Jan. 2016. Analyses of the modal meanings. In *The Oxford hand book of modality and mood*, ed. by Nuyts, Jan and Johan van der Auwera, 31-49. Oxford: Oxford University Press.
- Overall, Simon E. 2017. A typology of frustrative marking in Amazonian languages. In *The Cambridge handbook of linguistic typology*, ed. by Aikhenvald, Alexandra and R.M.W. Dixon. Cambridge: Cambridge University Press.
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. by Adelaar, Alexander and Andrew Pawley, 295-326. Canberra: Pacific Linguistics.
- van der Auwera, Johan and Andreas Ammann. 2011. Overlap between situational and epistemic modal marking. In *The world atlas of language structures online*, ed. by M. S. Dryer and M. Haspelmath Munich: Max Plank Digital Library, ch. 76.
- van der Auwera, Johan, and Vladimir Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2:79-124.
- Zeitoun, Elizabeth and Stacy F. Teng. 2016. Reassessing the position of Kanakanavu and Saaroa among the Formosan languages. *Oceanic Linguistics* 55(1):162-198.

# SECONDARY VERBS IN PA-O: A PRELIMINARY STUDY

Alys Boote Cooper  
*University of Zurich*  
*alys.bootecoop@uzh.ch*

## Abstract

Serial verbs and secondary verbs are a widely described feature of languages of Southeast Asia. The Karenic languages of Myanmar and Thailand also make extensive use of such constructions. This paper aims to examine how several verbs including ‘come’, ‘move away’, ‘give’ and ‘hit’ are used as secondary verbs to express various grammatical functions and modify the main verb of a clause in Pa-O, an underdescribed Karenic language from Myanmar. This paper aims to be a preliminary description of several secondary verbs in this language.

**Keywords:** serial verbs, secondary verbs, Karenic  
**ISO 639-3 codes:** blk, eky, kjp, ksw, pdu

## 1 Introduction

One typical feature of the languages of Southeast Asia is the frequent use of serial verb constructions (Enfield & Comrie 2015:8). Broadly speaking, serial verb constructions are made up of at least two verbs without any overt marker of dependency. As in Southeast Asia, these complex verbal predicates are common in Pa-O, a Karenic (Tibeto-Burman) language from Eastern Myanmar (glottocode: paok1235). Verbs may take on a grammatical function as ‘secondary verbs’ for different purposes, for example, for valency-increasing or tense, aspect and modality marking. In many languages, a verb may have different grammatical functions depending on its position in relation to the main verb(s). This is also the case in Pa-O and other Karenic languages. This preliminary study aims to introduce the reader to some secondary verbs and examine their degree of integration in the verb complex in Pa-O, formerly known by the exonym Taungthu, an underdescribed language.

The Karenic languages are spoken mostly in Myanmar and Thailand along with a significant diaspora of speakers overseas. They are predominantly spoken in Eastern Myanmar and Western Thailand near the Myanmar-Thailand border. The Karenic languages have long been regarded as a distinct group of languages in the Sino-Tibetan family, but despite this, Karenic’s exact place in the Sino-Tibetan family has not always been clear. One of the most striking features of Karenic languages is their verb-medial word order. Since almost all languages classified as Tibeto-Burman are consistently verb-final, it was unclear whether this word order was inherited or arose due to language contact with neighbouring SVO languages although the extent of Karenic’s contact with Tai-Kadai and Austroasiatic languages is still unclear. It was previously suggested that Karenic was a sister branch to Tibeto-Burman (for example Benedict 1972), but today Karenic is largely considered to be a daughter branch of Tibeto-Burman instead (Manson 2009)<sup>1</sup>. Despite the fact that Karenic has been long regarded as one group, the internal classification of its languages is not yet entirely clear.

Bradley (1997) divides Karenic into three branches: Northern, Central and Southern. Although Pa-O is spoken in both the northern and southern parts of the area covered by Karenic languages, it is usually classified as a ‘Northern’ Karenic language since the largest population is located to the north of the territory inhabited by Karenic speakers. Also classified as Northern is Kayan. Among Central Karenic, the most well-known languages are Bwe and Kayah whereas Sgaw and Pwo both belong to the Southern branch. Another grouping is by Manson (2011).

Although Pa-O is a reasonably large group (with over 600,000 speakers in Myanmar), it has been little researched until now. Pa-O has been included in some works on reconstruction of Proto-Karenic

---

<sup>1</sup> The existence of a Tibeto-Burman subgroup is not entirely uncontroversial, however, but this is beyond the scope of this paper.

(for example Jones (1961), Manson (2009)), but research on the grammar is lacking apart from one short grammatical sketch (Hackett 1954).

There are two main dialects of Pa-O: ‘Highland Pa-O’ (Northern Pa-O) and ‘Lowland Pa-O’ (Southern Pa-O). The majority of Pa-O speakers are Northern Pa-O and live in southern Shan State around Taunggyi and surrounding areas in Kayah state, although Northern Pa-O is far from being a homogenous group itself. The centre of Southern Pa-O is Thaton in Mon state, but Pa-O may also be found nearby in Kayin State. Between these two main centres there exists a dialect chain in Karen state and Bago Region. It is assumed that Pa-O speakers migrated northwards at some stage during the last millennium (Christensen & Kyaw 2006). There is some contact between speakers of these different dialects, particularly now due to increasing Internet usage, but they have some difficulty communicating in Pa-O due to dialectal differences and therefore sometimes switch to Burmese to ease communication.

Pa-O speakers today are subject to strong pressure from the official language of Myanmar, Burmese, which is dominant in schooling and media. Most young Pa-O speakers are therefore also proficient in Burmese and language shift is occurring in larger towns like Taunggyi and Thaton. There are two separate writing systems: Northern Pa-O and Southern Pa-O. Both writing systems are based on the Burmese script and have some extra symbols for tone markings and letter combinations to account for differences in the language. The Pa-O support their language by organising numerous summer school classes to develop speakers’ proficiency in reading and writing their own language. There are some regular publications written in Pa-O, but due to the only rare usage of Unicode encoding in Myanmar, most speakers communicate online in Burmese rather than Pa-O.

The Northern Pa-O have had extensive language contact with Shan (Tai-Kadai) as can be seen from its many Shan borrowings. In some areas there is also contact with smaller groups like Danu and Intha (Burmese varieties). In the south, there has been contact with Mon (Austroasiatic) and other Karenic languages like Sgaw and Pwo. The Northern Pa-O remained more isolated from the influence of Burmese for longer than the Southern Pa-Os due to their geographical location on the Shan plateau rather than the plains.

The majority of Pa-O are, in contrast to many other Karenic speakers, overwhelmingly Buddhists and easily recognisable in Myanmar due to their traditional clothing made up of all black or navy-blue clothing and brightly-coloured turbans for both men and women. This clothing is still worn by older Northern Pa-O speakers in the countryside (particularly women) and by speakers of all ages at any kind of festival. It has been claimed that the Southern Pa-O have been strongly Burmanised and the North Pa-O Shanised, but there is no in-depth study on the Pa-O dialects to measure whether this outsider’s impression is also reflected in the language.

## 2 Serial verbs

There is debate over exactly what defines a serial verb construction since different authors have used various definitions. Indeed, Lord (1993:1) calls it “a sticky business”. In this paper, the author will use Aikhenvald’s typology (2006) of these constructions. According to Aikhenvald (2006:4-21), the main criteria for serial verb constructions are: that the verbs form a single predicate without any markers of dependency between the verbs in a single clause; prosody is as in a clause with a single verb; same tense, aspect, modality (TAM) and polarity; the verbs describe one single event; the verbs share arguments and verbs must be independent lexical verbs. Several of these criteria may be problematic. It can be notoriously difficult to define what an ‘event’ is, for example, and the intonation of a clause is a phonological rather than syntactic criterion and is beyond the scope of this paper. Even the terms predicate and clause can be controversial. This analysis will stick to morpho-syntactic criteria. Example (1) is considered by the author to fulfil the aforementioned criteria. Three verbs ‘stay’, ‘hide’ and ‘be quiet’ are adjacent without any marker of coordination or subordination. The future marker and negators apply to the whole complex. All verbs are lexical verbs.

- (1) *tə-ʔí*                      *ʔòŋ*      *pəǹ̀?*      *té?*      *tâw*  
 NEG-FUT                  stay      hide      be.quiet      NEG  
 ‘I won’t stay here hiding quietly.’

## 2.1 Symmetry

Aikhenvald (2006) makes a distinction between symmetrical and asymmetrical serial verb constructions. In asymmetrical constructions, one of the verbs comes from a class which is somehow restricted. The verbs from a restricted class modify the event described by the verb or verbs from the open (or large) class. In this paper, the grammaticalised verbs in asymmetrical constructions are referred to as secondary verbs. Aikhenvald (2006:28-30) defines several main semantic types of symmetrical constructions: sequence of actions or concomitant actions related together; cause-effect serial verb constructions; manner serial verb constructions; synonymous verb serialisation. In addition, eight asymmetrical types are defined by Aikhenvald (2006:22-28): direction and orientation; aspect, extent and change of state; secondary concept serialisation; serialisation of complement-clause-taking verbs; increasing valency and specifying arguments; reducing valency; comparatives and superlatives. Relevant to this paper are direction and orientation, increasing valency, reducing valency.

## 2.2 Contiguity and core and nuclear serialisation

Another distinction has been made between nuclear and core serialisation (Foley & Van Valin 1984) which is called contiguous and non-contiguous by Aikhenvald (2006). These terms are part of Role and Reference Grammar and are a useful tool to look at multiverbal complexes although this paper does not follow the overall theory here. Foley & Van Valin (1984) distinguish between three levels of structure: nuclear (predicate), core (predicate + core arguments) and periphery (core + all other constituents). This pattern is displayed in Fig. 1. In nuclear serialisation, all arguments are arguments of the complex nucleus. No elements may be inserted between verbs in the nucleus. However, serialisation on the core level does allow insertion since the clause consists of two nuclear elements with separate core arguments.

*Figure 1: (Van Valin & Lapolla 1997:442)*

Nuclear juncture	[CORE ... [NUCL PRED] ... + ... [NUCL PRED] ...]
Core juncture	[CLAUSE ... [CORE ... ] ... + ... [CORE ... ] ... ]
Clausal juncture	[SENTENCE ... [CLAUSE... ] ... + ... [CLAUSE ... ] ... ]

In many verb-medial languages of Southeast Asia, verb serialisation may frequently occur on both the nuclear and core levels (Bisang 1992). In Karenic, however, a different pattern may be observed. It has been noted for Eastern Kayah Li (Solnit 2006), Geba (Swanson 2011) and Pwo (Kato 2003) that these languages are mostly restricted to nuclear serialisation. This is also the case for Pa-O. In fact, almost all multiverb complexes analysed thus far have consisted of nuclear serialisation. When both verbs share an object, the object appears after both of them, as in the purposive SVC (2). In (2) it would be grammatical to place the object between the verbs in a language like Thai, for example, but this is ungrammatical in Pa-O (2a). Example (3) shows that there is no transitivity harmony. A transitive verb like ‘strike’ may perfectly well occur with an intransitive verb ‘die’ in a multiverb complex.

(2) *k<sup>h</sup>wè p<sup>h</sup>re ʔám mók.cók*  
 1SG buy eat orange  
 ‘I buy oranges to eat.’

(2a) \**k<sup>h</sup>wè p<sup>h</sup>rè mók.cók ʔám*

(3) *k<sup>h</sup>wè twé sì t<sup>h</sup>wí*  
 1SG strike die dog  
 ‘I struck the dog so that it died.’

Objects can rarely intervene between verbs. Example (4) is from Pwo Karen which is possible only when the second verb is an unvolitional verb (see Kato forthcoming). The author has not come across any similar examples yet in her texts, but according to language consultants this structure is also possible in Pa-O.

- (4) *ʔəwé ʔáN mì blè*  
 3SG eat rice be.full  
 ‘He ate rice and got full.’  
 Pwo (Kato forthcoming)

A difference between the non-contiguous constructions of Thai and contiguous constructions of Pa-O may be seen in examples (5) and (6). Example (5) shows core serialisation in Thai where both verbs may take their own object arguments and the object of the first verb appears between the two verbs. This exact pattern is impossible in Pa-O since there is one valency for the entire verb complex, which shows that the verbs make up one unit. The valency is decided by the most transitive verb or can be increased by one through the use of a secondary verb like ‘give’ (§ 3.2). Since the O argument of the first verb cannot be inserted into the verbal complex, instead of the object of ‘carry’, this cannot be expressed by a multiverb construction. Example (6) does not fulfil the criteria of an SVC. For example, the verbs can have different polarity or TAM marking here.

- (5) *mê: p<sup>h</sup>a: lû:k-lû:k k<sup>h</sup>ûn bandaŋ*  
 mother escort offspring-PL ascend stairs  
 ‘Mother took the children up the stairs.’  
 Thai (Adapted from Diller 2006)

- (6) *ʔəmɿ jəʔ pa tóm ʔəpò ʔáw t<sup>h</sup>aŋ tədəŋ*  
 mother carry include COM child be.located climb stairs  
 ‘Mother climbed the stairs carrying children.’

In (7), the goal of the movement, *c<sup>h</sup>e* ‘market’, follows the verb *lwe* ‘go’. In (8), however, when *lwe* occurs with another verb, it is ungrammatical for the goal to appear between the two verbs. Instead, in (9) the object of the second verb directly follows the verb complex and the goal of the verb ‘go’ must occur in a postpositional phrase. This demonstrates a preference for nuclear over core serialisation.

- (7) *k<sup>h</sup>wè lwe c<sup>h</sup>e*  
 1SG go market  
 ‘I go to the market.’

- (8) *\*k<sup>h</sup>wè lwe c<sup>h</sup>e p<sup>h</sup>rè t<sup>h</sup>áʔ.pé*  
 1SG go market buy fish

- (9) *k<sup>h</sup>wè lwe p<sup>h</sup>re t<sup>h</sup>áʔ.pé c<sup>h</sup>e kò*  
 1SG go buy fish market inside  
 ‘I go buy fish at the market.’

Solnit (2006) suggests that this restriction could be a reflex of its original Tibeto-Burman verb-final word order, but this remains to be shown since there has not been any research done on the historical syntax of Karenic. Further research aims to discover and describe in which limited contexts core serialisation may occur.



### 3 Secondary verbs in Pa-O

In Pa-O, there appear to be several verbs which have a specific conventionalised meaning when used as a ‘secondary verb’. These verbs may all be used independently as the single verb in a monoverbal clause and also as secondary verbs, or V2s. Morphosyntactically, secondary verbs, unlike main verbs, are not syntactically independent. Rather, they select a verb as their syntactic host. V2s have a strict order within the verb complex (shown in Figure 2). A different order may result in an ungrammatical utterance or completely change the meaning. Meanwhile, the order of main verbs is often iconic i.e. relating to the temporal order in which actions are performed, rather than strictly grammatical. V2s are more bound than main verbs: they may not be individually questioned or individually be repeated in an answer, which is the case for main verbs. This paper will discuss secondary verbs ‘leave’, ‘come’; ‘give’ and ‘hit’ which mark direction, permissive, benefactive, ‘affectedness’ and obligation/necessity, respectively. This paper will also show that although functions are similar in Karenic languages, they are not identical.

#### 3.1 Directionals

Some verbs denoting directional movement may be used both as main verbs or in a secondary function. In (10) and (11), *lən* ‘come’ and *t<sup>h</sup>o* ‘go away’ act as main verbs in monoverbal clauses. Semantically, physical movement is performed by the agent.

- (10) *nì m̂ lən lén*  
 1PL mother come PERF  
 ‘(our) mother has come to meet me.’

- (11) *pò-mú t<sup>h</sup>o t<sup>h</sup>a*  
 offspring-FEM go.away thing  
 ‘The daughter moves away (from home).’

‘Come’ and ‘go away’ as secondary verbs may be considered as ‘movement toward or away from the centre of interest’ and typologically, direction verbs are very commonly grammaticalised in languages with serial verbs (Aikhenvald 2006:22). When used with a direction verb, ‘come’ and ‘go away’ specify the direction of the movement (12) as ‘towards’ or ‘away’.

- (12) *ʔòŋ t<sup>h</sup>o p<sup>h</sup>é təcíʔ làʔ nè*  
 stay go.away give a.little get Q  
 ‘Could you move aside a little, please?’

As a secondary verb, ‘come’ and ‘go away’ appear to be commonly used with other physical movement verbs like ‘run’ or ‘go’, but this is not a constraint. So far, the type of main verb they may appear with appears to be unrestricted: for example in (13) with a transitive volitional verb and in (14) an intransitive unvolitional verb. It is possible that these verbs express aspect as well as direction, but this is not yet clear and is the subject of continuing research. There is, however a semantic connection between the verbs. In (13) and (14), ‘go away’ is combined with ‘sell’ and ‘die’, whereas with the verbs ‘buy’ and ‘be born’, ‘come’ is preferred.

- (13) *k<sup>h</sup>wè tə-cà t<sup>h</sup>o t̂w d̂n k<sup>h</sup>wè ka*  
 1SG NEG-sell go.away NEG yet 1SG car  
 ‘I haven’t sold my car yet’

- (14) *we sì t<sup>h</sup>ò lén nì néŋ*  
 3SG die go.away PERF two year  
 ‘He’s been dead for two years.’

(15) and (16) clearly show the difference in position of V1 and V2 with ‘come’. (15) may be read as purposive. The mother comes in order to meet someone, whereas in (16), the mother calls the other, but no movement is specified in the process. Here ‘come’ rather refers to the direction of the call in relation to the speaker. Typically, in Pa-O, the main verb appears before the V2 (head-final order), which supports the theory that the original word order of Karenic was verb-final.

(15) *nì m̂ l̂n ĥ? kʰwè*  
 1PL mother come call 1SG  
 ‘(our) mother has come to meet me.’

(16) *m̂ ĥ? l̂n pʰôŋ*  
 mother call come 1SG  
 ‘Mother just called me.’

### 3.2 ‘Give’

In Pa-O, the verb *pʰé* ‘give’ has several functions. Firstly, the lexical verb may be used as the sole verb in a clause as in (17) where an object, *lítú?* ‘book’ is transferred. Additionally, ‘give’ may encode permissive or benefactive depending on its position relative to the main verb(s). Pre-verbal ‘give’ encodes a permissive meaning whereby a second participant is permitted to perform an action (18). The causer comes first, followed by ‘give’, V and then causee. When negated, the reading is also permissive as in (19).

(17) *cʰərá-mú pʰé t̂pɛ lítú?*  
 teacher-fem give pupil book  
 ‘The teacher gave the book to the student.’

(18) *kʰwè pʰé lwe pʰre na njáw cʰe kò*  
 1SG give go buy 2SG cat market LOC  
 ‘I let you go buy a cat at the market.’

(19) *kʰwè pʰé lwe t̂w na*  
 1SG give go NEG 2SG  
 ‘I won’t let you go.’

A similar pattern may be observed in Kayan, where *pʰè*, presumed by Manson (2010) to also be related to the main verb *pʰé* ‘give’, occurs before the main verb as a permissive marker as in (20).

(20) *θ̂n sá pʰè-l̂n zūθakén là-m̂n*  
 tree NEG allow-descend squirrel one-CLFs  
 ‘[But] no tree would let Squirrel descend.’  
 Kayan (Manson 2010:142)

Cross-linguistically, there is a connection between ‘give’ and causatives (Jenny 2015). This connection between ‘give’ and causation/permissive is more overt in Pa-O than in some other Karenic languages. In Sgaw, for example, the verb *ýé?* ‘give’ is not used as a secondary verb, but rather *du?*, a cognate root meaning ‘give’. This form is shown in (21). This form appears to be related to Kayah Li *dá*. In Kayah Li, *dá* is used both as a main verb ‘give’ and secondary verb preverbally (22).

- (21) *tàs<sup>h</sup>á lokò p<sup>h</sup>aʔdóʔ hò yə-duʔ-ʔó kɔʔ lɔ*  
 sickness fever large also 1SG-CAUS-drink all CFP  
 ‘Also, I let people who have a high fever drink it.’  
 Sgaw (Jones 1961)

- (22) *vē dɿ cwá ne to*  
 1SG give go 2SG NEG  
 ‘I won’t let you go.’  
 Kayah Li (Solnit 1997:158)

Postverbally, ‘give’ encodes benefactive in Pa-O where the recipient of the benefit occurs after the verb complex (23). Benefactives are not necessarily expressed by the verb ‘give’ in Karenic, for example, in Kayah Li (24) the marker *pè* corresponds rather with the verb ‘to win’ according to Solnit (1997). In Sgaw, this function is expressed by the verb *né* ‘get/take’ (see Chappell 1991) which occurs postverbally as in (25) and in Kayan by the verb *ba* ‘hit’ as in (26):

- (23) *dɔʔ p<sup>h</sup>é bá we néʔ*  
 say give hit 3SG Q  
 ‘Do I have to talk to him for you?’

- (24) *ʔa vē pè lū heso*  
 3 dig BEN OBV earth  
 ‘He dug out earth for her [i.e. doing her job].’  
 Kayah Li (Solnit 1997:157)

- (25) *θəráʔmúʔ pye né pya tàkíʔnàʔ*  
 teacher.female buy take person cloth  
 ‘The teacher bought us the cloth.’  
 Sgaw (adapted from Jones 1961)

- (26) *ŋá prà prì-bâ há písāp<sup>h</sup>ò tá*  
 that CLF buy-hit COMPL child fish  
 ‘That man bought the child a fish.’  
 Kayan (Manson 2010:146)

There is an interesting parallel observed in Burmese, Pa-O speakers’ current main contact language. In Burmese, the verb *pè* ‘give’ is similar in shape to Pa-O *p<sup>h</sup>é*. After the main verb, it gives a benefactive reading, like Pa-O. Before a main verb, *pè*, may express permissive like in Pa-O, but only for some speakers. This function of *pè* appears to be a recent phenomenon possibly due to influence from Mon State Burmese which has in turn been influenced by Mon and that has not (yet?) spread to all Burmese speakers (Okano 2005).

### 3.3 ‘Hit’

As a main verb, *bá* means ‘to hit’ (a target) or ‘to effect/come into contact with’. The same verb appears to have been grammaticalised in other Karenic languages and other unrelated languages in Southeast Asia. As a main verb in (27).

- (27) *tənát ʔáw lʃ-lʃ ká-ká tɛʔ bá k<sup>h</sup>wè*  
 gun exist exceed-RED cross-RED FOC hit 1SG  
 ‘A gun hit me accidentally.’

However, *bá* may also appear as a secondary verb. As a secondary verb, *bá* often encodes necessity or obligation as in (28). When negated, the negative marker *tâw* appears after *bá* and the gives the reading ‘I do not have to watch my daughter’.

- (28) *pò-mú*                      *thýŋ*                      *té?*                      *t<sup>h</sup>wâ*                      *bá*  
 child-FEM                      stand                      FOC                      look                      hit  
 ‘When my daughter wakes up I have to watch her/look after her.’

The meaning may be extended to that of ‘affectedness’. In (29) the translation could be more literally rendered as “have you been affected by receiving your salary?”

- (29) *ráp*                      *bá*                      *lên*                      *ròn*                      *là*                      *né?*  
 receive hit                      PERF                      money                      month                      Q  
 ‘Have you received your salary?’

*Bá* occurs very frequently in negative sentences with the particle ‘yet’ as in (30). Here, the sentence does not mean “did not have to yet”, but rather “did not get to yet/was not able to yet”. The subject has not yet come into contact with the action of reading this book.

- (30) *k<sup>h</sup>wè* *ʔəkhij* *ʔáw*                      *tâw*                      *tə*                      *lítú?*                      *jò*                      *p<sup>h</sup>át*                      *bá*                      *tâw*                      *dîn*  
 1SG time exist NEG because book PROX read hit NEG yet  
 ‘I have not read this book because I did not yet have the time.’

In Sgaw, *bá?* has a similar function, but occurs preverbally, as shown in (31). According to Rattanakul (2004) *ba?* is also present in a focus construction (32) where *ba?* appears cognate to Pa-O *bá*. In this example, the door is fronted as the focus in what has been called a passive-like structure. Manson (2010) describes an ‘experiential’ marker -*bà* which occurs in ‘passive’ constructions and does not consider it related to the verb ‘hit’ (however the benefactive marker in (26), *bâ*, is).

- (31) *kə-bá?*                      *yɔ?*                      *t<sup>h</sup>ɔ?*                      *t<sup>h</sup>i*  
 FUT-hit carry up water  
 ‘(she) has to carry up water on her head.’  
 Sgaw (adapted from Jones 1961)

- (32) *krɛ*                      *ba?*                      *mò*                      *ká?ti*                      *ʔɔ*  
 door touch mother shut 3SG  
 ‘It was the door that was shut by the mother.’  
 Sgaw (Rattanakul 2004:127)

This exact construction appears to be ungrammatical in Pa-O, but *bá* may occur following *k<sup>h</sup>ám* ‘to endure/bear’. The verb *k<sup>h</sup>ám* has been borrowed from Burmese. The shape of Pa-O *k<sup>h</sup>ám* indicates that it was borrowed from Burmese before Burmese nasal consonants became realised as nasalisation of the preceding vowel. Here, the subject (and semantic object of the actual action) is fronted and the agent occurs after ‘receive touch’. Here ‘blow’ is the noun, not the verb ‘to blow’. *Bá* reduces the volitionality of the action from the subject’s perspective. This construction may be used in situations with physical or non-physical contact (33).

- (33) *k<sup>h</sup>wè*                      *k<sup>h</sup>ám*                      *bá*                      *c<sup>h</sup>ərá*                      *k<sup>h</sup>ûn*  
 1SG bear hit teacher scold  
 ‘I had to endure being scolded by my teacher’

This structure appears to be borrowed from the Burmese construction shown in (34). Whereas in Burmese *kʰan* is used together with *yá* ‘get’, ‘get’ also has the secondary function of ‘must’, which corresponds with the function of ‘touch’ in Pa-O. Unlike in Burmese, most Pa-O speakers rejected sentences without an agent.

- (34) *θu-dó ʔəmé ʔə-pyò kʰan-yá-de*  
 3-ASS.Plmother.DEP NML-speak accept-get-NFUT  
 They were reprimanded by the mother.’  
 Burmese (Jenny & Hnin Tun 2016)

Rattanakul (2004) presents other examples where inanimate objects or animals were acted upon, but all these examples were not acceptable to Pa-O speakers. Any use of ‘hit’ in this kind of construction appears limited to animate undergoers. Since ‘hit’ may be used in this way in other Karenic languages, it seems that the pattern *kʰám bá* could be the combination of an indigenous structure and a structure obtained through contact. Grammaticalisation of ‘hit’ is not unusual in Southeast Asia. A similar structure also exists in neighbouring languages, for example Thai, Khmer and Mon. In Thai, *tʰù:k* encodes adversative along with *tèh* in Mon as in (35).

- (35) *ɲə yèh tɿʔ pèh tèh ʔəpa tək raʔ*  
 day morning that 2 hit father beat foc  
 ‘Tomorrow you’ll be beaten up by your father.’  
 Mon (Jenny 2005:108)

In summary, the order of items mentioned in this paper are represented in Figure 2. The current analyses show negation occurring at the beginning (optionally) and end of this domain followed by some TAM markings whose origin is not yet clear. Several items may occur in this slot, but since this is not the focus of this paper, the slot has been simplified as ‘TAM’ here. Before and after the main verb(s) (MV) there are slots for V2s. These V2s themselves have a fixed order. This fixed order within the verbal domain suggests that they are more grammaticalised. ‘Come’ and ‘go away’ appear to share the same slot, likely due to semantic reasons. V2s may be freely combined with each other (36).

- (36) *kʰwè pʰré làn pʰé bá tâw na mók.cók*  
 1SG buy come give hit NEG 2SG orange  
 ‘I didn’t buy you oranges.’

Figure 2: Order in the Pa-O verb complex

(NEG)	TAM	V2 <i>give</i>	MVs	V2 <i>come/go away</i>	V2 <i>give</i>	V2 <i>hit</i>	NEG
-------	-----	-------------------	-----	---------------------------	-------------------	------------------	-----

#### 4 Conclusion

As we have seen, some verbs have been grammaticalised multiple times with secondary functions, and the function of a secondary verb may be different according to its position in relation to the main verb in Pa-O and other Karenic languages. Although the functions expressed by secondary verbs are similar in these languages, there is no one-to-one matching of secondary verb to function. The function of directional verbs like ‘go away’ and ‘come’ appears linked to their semantics as full verbs, but they may now be used with a wide variety of verbs including non-movement verbs. ‘Give’ to encode permissive seems to be relatively stable among the languages mentioned, although the form of the secondary verb is not always the same. Pa-O uses the same verb to encode permissive and benefactive, but this is less common in other Karenic languages. Perhaps this difference is due to the influence of Burmese where a similarly sounding verb for give *pè* encodes both permissive and benefactive in some dialects. ‘Hit’

appears to encode the necessity of an event, but also the unvolitional necessity to experience an event. This still remains to be more thoroughly examined by the author, but there are interesting parallels in neighbouring languages like Thai and Mon. The differences between Thailand Sgaw and Pa-O are striking in relation to this construction and will be explored further. Overall, the verb complex in Pa-O provides support for an original verb-final word order in Karenic by demonstrating untypical structures for verb-medial serialising languages from Southeast Asia like a predisposition for nuclear or contiguous serialisation over core or non-contiguous.

## References

- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In *Serial verb constructions: a cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 1-68. Oxford: Oxford University Press.
- Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.
- Bisang, Walter. 1992. *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer: vergleichende grammatik im rahmen der verbserialisierung, der grammatikalisierung und der attraktorpositionen*. Tübingen: Gunter Narr Verlag.
- Bradley, David. 1997. Tibeto-Burman languages and classification. In *Papers in Southeast Asian linguistics 14*, ed. by David Bradley, 1-71. Canberra: Pacific Linguistics.
- Chappell, Hilary. 1991. *Syntax and semantics of the benefactive construction in Moulmein Sgaw Karen*. LaTrobe University Working Papers in Linguistics 4:37-52.
- Christensen, Russ., & Sann Kyaw. 2006. *The Pa-O: Rebels and refugees*. Chiang Mai, Thailand: Silkworm books.
- Diller, A. V. N. 2006. Thai serial verbs: cohesion and culture. In *Serial verb constructions: a cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 161-177. Oxford: Oxford University Press.
- Enfield, Nick J., & Bernard Comrie. 2015. Languages of Mainland Southeast Asia: The state of the art. In Nick J. Enfield & Bernard Comrie (eds.), *Languages of Mainland Southeast Asia. The state of the art*, 155–208. Berlin: De Gruyter Mouton.
- Foley, W. A., & R. D. Jr Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge Studies in Linguistics London (38).
- Hackett, Marion. 1954. *Sketch of Pa-O grammar*. ms.
- Jenny, Mathias. 2015. The far West of Southeast Asia: ‘Give’ and ‘get’ in the languages of Myanmar. In *Languages of Mainland Southeast Asia. The state of the art*, ed. by Nick J. Enfield & Bernard Comrie, 155-208. Berlin: De Gruyter Mouton.
- Jenny, Mathias. 2005. *The verb system of Mon*. Zurich: ASAS.
- Jenny, Mathias., & San San Hnin Tun. 2016. *Burmese: a comprehensive grammar*. London and New York: Routledge.
- Jones, Robert B., Jr. 1961. *Karen linguistic studies*. Berkeley & Los Angeles: University of California Press.
- Kato, Atsuhiko. In Press. Pwo Karen. In *The Mainland Southeast Asia Linguistic Area*, ed. by Alice Vittrant & Justin Watkins, 131-175. Berlin, Boston: De Gruyter Mouton.
- Kato, Atsuhiko. 2003. Pwo Karen. In *The Sino-Tibetan languages*, ed. by Graham Thurgood & Randy J. LaPolla, 632-650. London: Routledge.
- Lord, Carol. 1993. *Historical change in serial verb constructions*. Amsterdam/Philadelphia: John Benjamins.
- Manson, Ken. 2011. *The subgrouping of Karen*. Paper presented at the 21st Annual meeting of the Southeast Asian Linguistics Society (SEALS), Kasetsart University, Bangkok.
- Manson, Ken. 2010. *A grammar of Kayan, a Tibeto-Burman language*. Bundoora, Victoria: La Trobe University dissertation.

- Manson, Ken. 2009. Prolegomena to reconstructing Proto-Karen. *LaTrobe Working Papers in Linguistics* 12.
- Okano, Kenji. 2005. The verb ‘give’ as a causativiser in colloquial Burmese. In *Studies in Burmese Linguistics*, ed. by Justin Watkins, 97-104. Pacific Linguistics: Canberra.
- Rattanakul, Suriya. 2004. Karen. *Collected papers Vol. 1*. Nakhorn Pathom: Mahidol University.
- Solnit, David B. 2006. Verb serialization in Eastern Kayah Li. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 144-159. Oxford: Oxford University Press.
- Solnit, David B. 1997. *Eastern Kayah Li: Grammar, Texts, Glossary*. Hawai’i: University of Hawaii Press.
- Swanson, Kirstie. 2011. *Serial verb constructions in Bwe Karen*. Chiang Mai: Payap University MA thesis.
- Van Valin, Robert D. Jr. & Randy J. Lapolla. 1997. *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press.

# THE ASSOCIATION OF FOREGROUND AND BACKGROUND INFORMATION WITH VERB STEM ALTERNATION IN SIZANG CHIN

Tyler Davis  
*Payap University*  
*tdavlinguist@gmail.com*

## Abstract

Sizang Chin exhibits verb stem alternation in which one form of a verb (Stem I) occurs in certain environments and a secondary form (Stem II) occurs in other environments. Henderson (1965) correlates Stem I with final clauses and Stem II with non-final clauses. King (2009) correlates Stem II with subordinate clauses. She also correlates Stem I with the agentive voice and Stem II with the non-agentive voice. Building on the work of Henderson (1965) and King (2009), this study examines the correlation of verb stem alternation and foreground and background information in Sizang third-person narrative discourse. The study demonstrates that foreground information clauses mostly correlate with Stem I and setting (background) information clauses always correlate with Stem II. However, most background clauses conveying possible events or events that did not happen correlate with Stem I, not Stem II as expected. Therefore, verb stem alternation in Sizang Chin cannot be accounted for by foreground and background information alone.

**Keywords:** Tibeto-Burman, Kuki-Chin, verb stem alternation, narrative discourse

**ISO 639-3 codes:** cnh, csy, ctd

## 1 Introduction

Verb stem alternation is a phenomenon in Kuki-Chin languages in which a verb may surface in one of two stems, which are often phonologically distinct. The primary and secondary stems are often referred to as “Stem I” and “Stem II”, respectively. This paper examines the correlation between verb stem alternation and correlation with foreground and background information in Sizang Chin (csy), one of the Northern Kuki-Chin languages, which is spoken in the south of Tedim Township, Chin State, Myanmar (Burma).

## 2 Literature review

Henderson (1965) is one of the first researchers to examine verb stem alternation in a Northern-Kuki-Chin language, analyzing two Tedim Chin (ctd) texts at the word, phrasal, and sentence level. Stem I, per Henderson, is found in “conclusive sentences” and is associated with the expression of indicative mood, and Stem II is found in “inconclusive sentences” and is associated with the expression of subjunctive mood (1965:84). Henderson further comments that Stem I verbs “are greatly in the majority [within a narrative]” (Henderson 1965:86).

Kathol (2003) and Kee Shein Mang (2006) investigated verb stem alternation in terms of its syntactic, semantic, and pragmatic functions. Kathol (2003:219), in his study of verb stem alternation in Hakha Lai, concludes that there is no single form-function relationship between stems and their functions. However, his findings suggest that valency and polarity are the two strongest factors in determining the verb stem of the clause. Kee Shein Mang (2006:107–108) examined factors such as valency, clause subordination, clause dependency, and deontic modality in K’Cho. He also argues that Stem II is used in valence increasing operations.

King (2009), in a typological examination of verb stem alternation across different Kuki-Chin subgroups, claims that verb stem alternation is “fundamentally the morphosyntactic manifestation of the agentive voice [Stem I] and its logical counterpart, the non-agentive voice [Stem II]” (King 2009:142). She determines this by first considering the argument proposed by Chhange (1993), that Stem



II originally encoded nominalization. Stem II was then further grammaticalized to being used in subordinate clauses, and then Stem I and Stem II respectively acquired subject focus and oblique/object focus. King admits that the stems do not always correlate with each function, due to the development of causative, benefactive, and antipassive morphemes. Therefore, King proposes that Stem I correlates to the agentive voice – meaning the focus of the clause is on the agent – and Stem II correlates to the non-agentive voice – meaning the focus is not on the agent. Her conclusions are summarized in Table 1.

**Table 1:** Functions of Kuki-Chin verb stems according to King (adapted from King 2009: fig. 7)

Function	Stem I	Stem II
nominalization	agentive	nonagentive/verbal nouns
subordination		adverbial/non-finite
relative clauses	subject relativized	object/oblique relativized
content questions	subject questioned	object/oblique questioned
valence-changing operations	causatives <sup>1</sup>	causatives/benefactives
	antipassive	ergative

The issue with King’s methodology is that she does not make a clear distinction between the function of verb stem alternation in clausal-level phenomena (causatives/benefactives, adverbials, antipassive) and its function in argument-level phenomena (agentive, nonagentive, ergative, subject/object, etc.). Furthermore, most of the previous studies of verb stem alternation in Kuki-Chin languages focus only on the syntactic environments in which the verbs alternate. DeLancey (1987:54) suggests that “[the] interpretation of any utterance is in fact based at least in considerable part on real-world context” and therefore, analyzing sentences in a discourse allows that discourse to be a “substitute for the real-world context”. Therefore, this paper focuses on how each verb stem surfaces in foreground and background information in third-person narrative discourse. Li (2014:101) calls grounding analysis the “functional study of the form–function relationships between structural features and their grounding functions”.

Regarding the morphophonemic properties of verb stem alternation in Sizang, Button (2011a) observes that most Stem II verbs in Sizang have a falling tone, but the tone of the proto-Stem I verb often determines the coda of the Stem II verb. Not all stems may be accurately predicted this way, as demonstrated, for example, by the alternation of *t<sup>h</sup>εŋ* (level) → *t<sup>h</sup>êt* (falling) ‘arrive’ in both the author’s data and in Button’s (cf. Button 2011b:30). Button’s reconstruction contradicts these data by predicting *t<sup>h</sup>ên* as the Stem II form. However, Button’s reconstructions provide a stable framework for determining the stem of a verb in Sizang discourse, if both stems are similar in every way, except for the tone (for example, *tăm* → *tâm* ‘many’).

### 3 Foreground and background information

Regarding narratives, Labov and Waletzky (1967) were the first to make a distinction between clauses which demonstrate “something that the speaker [or narrator] presents as having happened, or that a particular participant did” and those that provide additional information off the timeline. Hopper (1979:213) terms the first type of clause “foreground information” and all other types of clauses “background information”.

The foreground information within narrative discourse parallels Grimes’ (1975) idea of events in narratives, and the background information likewise parallels Grimes’ idea of “non-events”. There are six types of non-events that Grimes defines, which have been further defined by Dooley and Levinsohn (2001). *Participant identification* introduces, reintroduces, or describes a participant/character in a narrative. *Setting* provides the locational and temporal information about an event in the discourse. *Explanation* is information through which the narrator (or the speaker in a conversation) clarifies what

<sup>1</sup> In Falam, Tedim, and Sizang.

is happening in the narrative. *Collateral* information explains what did not happen in relation to what did happen and is realized as negatives, clauses with irrealis and deontic modality, interrogatives, and conditionals. *Evaluations* are instances in which the narrator voices their personal feelings or perspective about the present situation. *Performative* information usually contains statements from the narrator addressed to the audience, specifically about the narrative (e.g. morals and conclusions) (Dooley & Levinsohn 2001:83). No examples of evaluation or performative information have been found in the corpus used in this study (see below). Furthermore, due to space constraints, only the expression of foreground information, along with the expression of setting and collateral background information, is examined in this paper.

#### 4 Data and methodology

For this paper, a total of nine third-person traditional narratives were examined. Three folktales come from Stern (1984). A further six narratives were collected by the author in March 2017, in Kalaymyo, Myanmar. Within the examples, Stern’s data will be cited with his name, and my data will be cited with the name of the narrative from whence it came.

For this study, the narratives were examined clause-by-clause in a spreadsheet. Comparing the verb of each clause to the verb forms independently elicited from a native speaker, each verb was labeled as “Stem I”, “Stem II”, “Identical”, or “Unknown”. The label “Identical” was used when a particular verb appears to only have one form and the label “Unknown” was used if the speaker, from whom I elicited the verbs, could not give both stems. In this paper, only the verbs whose stems could be verified as either Stem I or Stem II are examined, but the number of occurrences for all stems are still included in the tables below. To determine foreground and background information, each clause was given a grounding value, based on the definitions of foreground and background information, which were given above. After independently determining the grounding value of each clause and the verb stem of each clause, the data was sorted in order to find any semantic correlations between the verb stems and grounding values.

The remainder of the paper is structured as follows: The expression of foreground information is examined in Section 5, followed by the expression of setting information in Section 6 and collateral information in Section 7. The conclusions are then given in Section 8.

#### 5 Foreground information

Foreground information contains the temporally-sequenced main events of the narrative, making up the central storyline. Out of 409 total clauses examined from nine narratives, 148 clauses contain foreground information. Of these 148 clauses, 129 of those foreground information clauses in the data have the head verb surfacing as Stem I, as shown in (1)<sup>2</sup>.

- (1) *á-há:u pǎ: ín tʰíŋ.kú:ŋ túŋ á=sî:a hí:*  
 ATTR-rich man ERG tree top 3S=set.snare.I be  
 ‘The rich man set his snare atop a tree.’ (Stern 1984:45)

In this story, a rich man and a poor man set out to hunt barking deer. The rich man sets his trap on top of a tree (where a barking deer cannot go) and the poor man sets his trap in a field. The sentence in (1) contains only one clause and one event: the setting of the snare. The end of the clause is also marked with the verb *hí:* which indicates a realized event.

<sup>2</sup> In this paper, I use the following abbreviations: 1=1st person, 2=2nd person, 3=3rd person, I=Primary Stem, II=Secondary Stem, AUG=augmentative, CAUS=causative, CIS=cislocative, CL=clause, COM=comitative, COND=conditional, DIR=directional, E=(either) identical verb stem, HORT=hortative, IMP=imperative, IRR=irrealis, MAIN=main clause, NEG=negative, NF=non-final, NMLZ=nominalizer, PL=plural, Q=interrogative, SC=subordinate clause, SUB=subordinator

By contrast, two clauses and two events are shown in (2).

- (2) [áma: sî:a-na t<sup>h</sup>íj.kú:ŋ túj a: š:k-sâk a:]<sub>MAIN</sub>  
 3S trap.II-NMLZ tree top at snare.I-CAUS NF

[tîa: kîk hî:]<sub>MAIN</sub>

**return.I again be**

‘[He] caused [the deer] to be caught in his own trap on top of the tree and went back home.’  
 (Stern 1984:45).

In the coordinate clause in (2), the main verb *š:k-sâk* ‘caused to be caught’ surfaces as Stem I, as does the second event *tîa:* ‘return’. It is demonstrable that the first event is indeed Stem I and not Stem II, because according to Stern (1984:49 note 10), if the event were Stem II, the verb would mean ‘allow to be caught’ because causative *sâk* would inherit its secondary meaning, the benefactive.

The ten instances of Stem II in foreground information clauses were all verbs containing the comitative enclitic =*pûi*, as demonstrated in (3).

- (3) [a=ɔai liăt pîa a:]<sub>CL1</sub> [a=ɔm-nũ: t ên=pûi: hî:]<sub>CL2</sub>  
 3s.POSS=buffalo eight give.I NF 3s.POSS=friend-FEM **establish.II=COM** be  
 ‘[The village head] gave his eight buffalo and married with [the man’s] girlfriend.’ (*Tei le Zuang Thu*)

In this story, a boyfriend and girlfriend meet a rich man, who asks the boyfriend if the woman with him is his girlfriend or wife. The boyfriend lies, saying she is his wife. Therefore, the rich man declares he will take the “wife” and gives eight buffalo as payment. Like other instances of foreground information in these data, this sentence ends in the verb *hî:* ‘be’ and is the final event in a sequence. However, the head verb of this clause ‘establish’ is *teaj* in its Stem I form. The Stem II form exhibits a change in tone and monophthongization, which is a clear indication of stem alternation. However, as this verb is a main event, the only evident explanation as to why it is surfacing as Stem II is the enclitic attached to the verb stem. Peterson (2007:18) explains the use of the *-pii* in Hakha Chin, which is cognate to Sizang’s =*pûi*, saying that it “signals the addition of an object argument which is understood to be a co-participant with the subject in the performance of the action described by the verb”. Thus, in this example, the subject who is performing the action is the rich man, and the oblique object argument is the *ɔm-nũ:* ‘girlfriend’.

A summary of the instances of verb stem alternation in foreground information is given in Table 2.

**Table 2: Summary of foreground information by stem type**

Stem Type	Instances
Unknown	3 (2.0%)
Identical	5 (3.4%)
Stem II	10 (6.8%)
<b>Stem I</b>	<b>129 (87.2%)</b>
Total Foreground	148 (100%)

In foreground information, the majority of head verbs within clauses surface as Stem I, with 129 instances. Every instance of Stem II surfacing within the clauses contains the enclitic =*pûi*. Therefore, for foreground information clauses, the default stem is Stem I, as hypothesized.

## 6 Setting information

Setting information provides the temporal and locational context of the event. Out of 83 total setting information clauses, the head verb of setting information clauses surfaces as Stem II 51 times. 21 instances are verbs whose stems do not change, and 12 instances are of verbs whose stem forms are unknown. There are no instances where Stem I surfaces in setting information clauses. An example of a typical setting clause is given in (4).

- (4) [*ĩn*     *á=t<sup>h</sup>ét*             *tíay*     *in*]<sub>SUB</sub>    [... *mi:*     *zɔ:ŋ-pǎ:*     *mún*     *a:*  
**house**   **3SG=arrive.II**     **when**     **SUB**     ... person   poor-MASC place   at
- [*ĩ:*=*sǎ:*     *sî:a*             *víl*             *ká=tɔ:ŋ*]<sub>QUOTE</sub>    [*tí:*     *a:*]<sub>MAIN</sub>    [*pá:hî:*]<sub>MAIN</sub>  
 1P=animal    set.snare.II    watch.I    DIR=HORT    say.I    NF            invite.I be
- Lit: house he arrive when, person poor man place at “our traps check upwards let’s” say and invite.  
 ‘Upon arriving at home, the rich man invited the poor man and said “let’s go check our traps”.’  
 (Stern 1984:45)

The sentence in (4) is comprised of three separate clauses. The first is a temporal subordinate clause, distinguished by the word *tíay* ‘when’ and the subordinating marker *in*. The main verb within the temporal clause, surfacing as *t<sup>h</sup>ét*, is the secondary stem of *t<sup>h</sup>eŋ* ‘arrive’. Within the quote, which is a hortative imperative, the head verb *víl* ‘watch’ is in Stem I. Finally, the head verb within the main clause *pá:* ‘invite’ also surfaces as Stem I. Therefore, within this sentence, *pá:* is the main verb, providing the foreground information for this particular event on the timeline, whereas *t<sup>h</sup>ét* is providing the background information regarding when the main event took place.

In (5), the event setting clause expresses a completed event, subordinate to the main event of the sentence.

- (5) [*tu:a*     *t<sup>h</sup>iy-pǐ:*     *ŋau*             *a=sêp*                     *zô:k*                     *tíay=in.nǎ*]<sub>SC</sub>  
 that     Thingpi    sacrifice   **3=work.II**             **finish.II**             **when=SUB**
- [[*ŋǔa:*     *hoŋ*             *zu*             *kî:k*             *hî:*]<sub>QUOTE</sub>    *tí:*             *hî:*]<sub>MC</sub>  
 rain    CIS    rain.I    again    be    say    be
- ‘When they had finished working the Thingpi sacrifice, they said “it’s raining on us again!”’  
 (*Thingpi Taangthu*)

The subordinate clause, in (5), expresses the completed event of working the *Thingpi* sacrifice by combining the Stem II form of *seam* ‘work’ with the Stem II form of *zô:* ‘finish’, followed by the temporal phrase *tíay in.nǎ*. ‘when’, giving an ‘after’ temporal sense. This is followed by the main event *tí:* ‘say’ in the following main clause. In (4), the events occur in a sequence with the two Stem I-headed clauses separated by the NON-FINAL *a:*. However, because the “working” of the sacrifice occurs in a background subordinate clause, the head verb surfaces as Stem II in (5).

**Table 3:** Summary of setting information by stem type

Stem Type	Instances
Unknown	13 (15.3%)
Identical	21 (24.7%)
Stem I	0 (0.0%)
<b>Stem II</b>	<b>51 (60.0%)</b>
Total Setting	85 (100%)

The results for setting clauses are summarized in Table 3. As demonstrated in Table 2, all 51 setting information clauses within the data, whose stem type could be determined, occur with Stem II verbs. Unlike setting clauses, collateral information clauses show a mix of Stem I and Stem II head verbs.

## 7 Collateral information clauses

Collateral information relates what did not happen to what did happen in the main events of the narrative. In the corpus, collateral information clauses surface as conditionals, imperatives, interrogatives, negatives, purposives, and other irrealis clauses, either denoting probability or future events. Out of the 79 collateral clauses in the data, 58 have a Stem I head verb, six have a Stem II head verb. In another four clauses, there is no verb stem alternation. In the last eight clauses, the head verbs cannot be verified as either Stem I or Stem II. In the discussion of collateral information, conditionals are discussed first, followed by imperatives, interrogatives, negatives, purposives, and finally other irrealis clauses, either denoting probability or future events.

The sentence in (6) contains two types of collateral information clauses: a conditional information clause followed by an imperative clause.

- (6) [na=tí:a núam lé:]<sub>COND</sub> [ná:=da:k-bǔ: tǒ: hoŋ kí:.tǎn tân]<sub>IMP</sub>  
 2=return.E want.I COND 2=gong.set with CIS relinquish.I IMP  
 ‘If you want to go back home, give me your gong set.’ (*Pu ZaDo Daak Ngakna Thu*)

The first clause in (6) is marked as a complex conditional clause by the conditional marker *lé:*. The complex predicate consists of two verbs *tí:a* ‘return.home’, which has only one form, and *núam* ‘want’, which is *núam* in Stem I and *nôp* in Stem II. Within the conditional clause in (6), Stem I surfaces. The second clause is overtly marked as an imperative by the imperative marker *tân*. The predicate consists of one verb *kí:.tǎn* ‘relinquish’, which is in Stem I. There are 14 instances of imperative clauses in the corpus, with 13 instances of a Stem I verb surfacing and one instance of an unknown stem.

Within the corpus, there are 13 instances of clauses containing the marker *lé:*. Six of those instances occur with a Stem I verb, as in (6), four instances occur with a verb whose stem does not have a secondary form (“Identical”), and three instances occur with a verb whose stem is unknown. While Stern (1963) and King (2007) posit that Stem II conditionals exist, I argue that the marker *lé:* has two functions. The first is as a conditional marker, which is demonstrated in (6) and the second is as a temporal marker, which is demonstrated in (7).

- (7) [tu:a t<sup>h</sup>íŋkú:ŋ-tě: ní: p<sup>h</sup>ǔ:k a:]<sub>C1</sub>  
 that tree-PL two chop.I NF  
 [[á=p<sup>h</sup>ǔ: lé:]<sub>SC</sub> pǔ:k t<sup>h</sup>ěi-ŋǒl hêt]<sub>C2</sub>  
 3=chop.II when fell.I able.I-NEG probably.not  
 ‘They chop at those trees and **when** they chop, they probably can’t fell them.’ (*Tei le Zuang Thu*)

(7) contains three clauses. The first clause is a foreground clause with the Stem I form *p<sup>h</sup>ǔ:k* ‘chop’. This is followed by a temporal subordinate clause which repeats the information in the preceding clause in a tail-head linkage relationship. This clause’s head verb is the Stem II form *p<sup>h</sup>ǔ:* ‘chop’. After glossing this story in the field and discussing with native speakers, the use of Stem II here appears to be signifying a *temporal setting* clause, rather than a conditional. King (2009:152) briefly addressed the issue of conditionals sometimes surfacing as Stem I, where, according to her hypotheses, Stem II would be expected. King states that the reason for the alternation may be due to some conditional clauses being contrafactuals, thus making it a realis-irrealis issue. However, in the Sizang corpus, all conditionals appear in situations that are assumed to be true. As for Stern (1963:276), he analyzes one form of conditional to be “polite” and the other “colloquial”, and states that the stems change accordingly. His hypothesis is also not supported by the data in the corpus.

A verbal interrogative collateral information clause is shown in (8).

- (8) [[ká=le:i zôn le:i]<sub>NP</sub> á=tij zí:am tí:=in]<sub>QUOTE</sub> dôn hí:  
 1S.POSS=tribute search.II tribute **3=be.worthy.I Q** say=SUB ask.I be  
 ‘[They] asked, “is the tribute [animal] that we searched for worthy?”’ (Thingpi Taangthu)

The head verb of the interrogative collateral information clause is *tij* ‘be.worthy’, whose Stem II form is *tín*. Out of eight interrogative collateral clauses in the corpus, six of them contain Stem I verbs. However, there is one instance where the interrogative clause contains a Stem II verb, as demonstrated in (9).

- (9) [sáha:ŋ=nă: baŋ.bâŋ ŋun kân=pûi:zí:am]<sub>QUOTE</sub> tí: lê:  
 tiger=ERG how river **cross.II=COM Q** say COND  
 ‘If you ask, “how did the tiger cross the river [with him]?”’ (Pu ZaDo Daak Ngakna Thu)

As in example (3), the head verb of the interrogative collateral information clause *kân* ‘cross’ is suffixed by the comitative enclitic and surfaces in Stem II. Because the verbs within foreground and interrogative clauses all surface as Stem I, except for instances with the comitative, I posit that verbs in Sizang suffixed by the comitative enclitic always surface as Stem II.

A negative collateral clause is shown in (10). Out of 15 instances of negative collateral information clauses, 14 of the instances contain Stem I verbs. The stem of the verb in the remaining instance is unknown. A negated clause with a Stem I verb is illustrated in (10).

- (10) mi: há:u=pă: sî:a-na: [t<sup>h</sup>ij túŋ=a:]<sub>LOC</sub> bâŋmâ: ǰ:k-ŋǰl hí:  
 person rich=MASC set.snare.II-NMLZ tree top=at **nothing snare.I-NEG be**  
 ‘The rich man’s trap at the top of the tree snared absolutely nothing.’ (Stern 1984:45)

In (10), just like in (1), the head verb of the clause surfaces as Stem I. Therefore, both events, the archetype of foreground information, and negatives, the archetype of collateral information, appear with Stem I verbs.

An example of a purposive collateral information clause is demonstrated in (11).

- (11) [í:=kî:=zâ:ktâ:-na:=tû:=in]<sub>SC</sub>  
 1PL.INCL.AGR=MID=**respect.II-NMLZ**=PUR=SUB  
 [akûa u ha:m=bêl hí: zí:am]<sub>MAIN</sub>  
 who older.sibling aged=SUPL be Q  
 ‘In order for us to respect each other, who is the eldest?’ (Stern 1984:50).

In Sizang, purposive constructions are formed by suffixing (-na:)=tû:=in to the verb. The purposive =tû: is grammaticalized from the postposition *a:tû:* ‘for’. In all three instances of a purposive collateral information clause, the verb surfaces as Stem II.

Finally, an irrealis collateral information clause is demonstrated in (12).

- (12) î:=kî:-k<sup>h</sup>ên t<sup>h</sup>âp kul tú:  
**1P.INC=MID-divide.II** individually **must** IRR  
 ‘We will have to split up individually’ (Stern 1984:50).

In (12), the deontic modal verb *kul* ‘must’ occurs with the Stem II form of *k<sup>h</sup>ên* ‘divide’. Other researchers have mentioned the affinity of Stem II with deontic modality (Mang 2006). When discussing this phenomenon with native speakers, I learned that *kul* may also appear as a head verb, as shown in (13).

- (13)  $\hat{a}:ma:-t\acute{e}:$   $\emptyset$   $kul-buan-t\acute{e}:$   
 3S-PL 3s need.I-NEG-belief  
 ‘They don’t need it.’ (Sing Za Nang 2010:8)

In (13), *kul* is a transitive verb with the meaning ‘need’. The predicate has two arguments, the third-person plural pronoun  $\hat{a}:ma:-t\acute{e}:$  and an absent object argument. The English free translation clarifies that this verb is transitive, as it is not saying ‘they are not needed’, but rather ‘they do not need it’, denoting the existence of an object argument. This sheds light on the interpretation of (12), in which the subject argument is indicated by the pre-verbal first-person plural inclusive agreement marker  $\hat{i}:=$ . Therefore, just as there is nominalization occurring in setting information clauses, the dynamic verb *k<sup>h</sup>en* ‘divide’ is nominalized to become the object argument of *kul*.

The association of different collateral information types with verb stem types is presented in Table 4.

**Table 4:** Occurrences of stems within collateral information clauses

	Stem Type	Stem I	Stem II	Identical	Unknown
Collateral Clause Type	Conditionals	8	0	4	3
	Imperatives	13	0	0	1
	Interrogatives	6	1	1	0
	Negatives	14	0	0	1
	Purposives	0	3	0	2
	Irrealis	14	1	0	1
	TOTAL	58 (76.3%)	5 (6.6%)	5 (6.6%)	8 (10.5%)

As demonstrated in Table 3, the majority of collateral information clauses, with the exception of Purposives, occur with Stem I verbs, contrary to the expectation that background information correlates with Stem II verbs. The two instances of Stem II occur in collateral information clauses, as explained in the discussions of (9) and (12). The one instance of Stem II which occurs in an interrogative clause is explained by the occurrence of the comitative enclitic  $=pu\acute{i}$ : *on the head verb*, which coincidentally was the marker of all 10 occurrences of Stem II in foreground information (see (3)). The one instance of Stem II occurring within an irrealis collateral information clause, as explained in the discussion of (10), is due to the nominalization of *k<sup>h</sup>en* ‘divide’. Therefore, the explanation of verb stem alternation in collateral information clauses is not solely due to the fact that collateral information is a type of background information.

## 8 Conclusion

Explaining verb stem alternation on the basis of foreground information versus background information introduces another approach for predictability regarding the use of each verb stem in discourse. Looking at verb stem alternation from a discourse perspective could help linguists to properly separate clausal-level phenomena from argument level-phenomena, if applied to other languages. The hypothesis was, that because both Henderson (1965) and King (2009) associated Stem II with subordinate clauses, that Stem II would correlate with background information in narrative discourse. Stem I, likewise, would correlate with foreground information. This paper demonstrates that foreground clauses contain verbs in Stem I except when the clause contains the comitative enclitic. As described by Peterson (2007), because the comitative is an applicative which adds a co-participant to the clause, the valency of the clause is then raised. Because this pattern occurs every time the comitative enclitic is suffixed to the verb stem, Peterson’s argumentation explains every instance of Stem II surfacing in foreground information in the data (see (3)), and the one instance of Stem II surfacing in collateral information (see (9)). Kathol (2003), Kee Shein Mang (2006), and King (2009) agree that raised valency is a function

that denotes the use of Stem II verbs. Indeed, Chhangte (1993:86) argues that verb stem alternation in Mizo, is “the product” of nominalization and valence change. Therefore, I posit that the Stem II occurrences in foreground information in these data are the result of valence raising, due to the COMITATIVE enclitic. As for setting clauses, the head verbs always surface as Stem II, as hypothesized. However, although collateral information clauses are also background information, the majority of verbs in its clauses surface as Stem I. Therefore, it is clear that verb stem alternation in Sizang Chin does not directly correlate to foreground information and background information. Further investigation is thus needed, to see why background information always correlates with Stem II, but collateral information does not.

This paper contains several limitations. Although Button (2011a) provides a good description of how verbs alternate in Sizang Chin, the patterns are not always predictable. Therefore, it is necessary to review every stem in a set of data with a native speaker, so no “Unknown” tokens exist in the analysis. Also, due to length constraints, only three types of information are examined in this paper. Looking back at King’s initial classifications, this paper addresses everything on the clausal level for Sizang Chin that King examined, except for relative clauses, which are treated as arguments in Sizang Chin, and benefactives, which do not appear in the corpus. So, a broader corpus with natural examples of benefactives from a narrative would help to fully address King’s hypotheses of valency. Regarding King’s interpretation of conditionals, this paper briefly demonstrates that the conditional marker with a Stem II verb denotes setting information, rather than conditional collateral information. However, further investigation into the nature of contrafactuals within Sizang is needed, as neither this corpus nor King provides adequate examples of them. Also, a further look into the nature of nominalization in Sizang is needed, as results slightly suggest that it is a large factor of Stem II surfacing at the clausal level for setting information clauses and collateral information clauses.

## References

- Button, Christopher. 2011a. *Proto Northern Chin: Vol I.* (STEDT Monograph 10). Berkeley, CA: STEDT.
- Button, Christopher. 2011b. *Proto Northern Chin: Vol II.* (STEDT Monograph 10). Berkeley, CA: STEDT.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.
- Chhangte, Lalnunthangi. 1993. *Mizo syntax.* Eugene: University of Oregon.
- Chung, Sandra & Alan Timberlake. 1985. Tense, aspect, and mood. In *Language typology and syntactic description*, ed. by Timothy Shopen, 202-258.
- DeLancey, Scott. 1987. Transitivity in grammar and cognition. In *Coherence and grounding in discourse: Outcome of a symposium, Eugene, Oregon, June 1984*, ed. by Russell S. Tomlin, 53-68. (Typological Studies in Language 11).
- Dooley, Robert A. & Stephen H. Levinsohn. 2001. *Analyzing discourse: A manual of basic concepts.* Dallas, Texas: SIL International.
- Givón, T. 2001. *Syntax: An introduction, Vol II.* Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Grimes, Joseph E. 1975. *The thread of discourse.* (Janua Linguarum Series Minor, 207). The Hague & Paris: Mouton.
- Henderson, Eugénie J. A. 1965. *Tiddim Chin: A descriptive analysis of two texts.* (London Oriental Series 15). Oxford: Oxford University Press.
- Hopper, Paul J. 1979. Aspect and foregrounding in discourse. In *Syntax and Semantics Vol. 12*, ed. by Talmy Givón, 213–241. New York: Academic Press.
- Kathol, Andreas. 2003. Cooperating constructions in Lai “Lexical insertion”. In *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State*



- University, ed. by Stefan Müller, 203-221. Stanford, CA: CSLI Publications.  
<http://csli-publications.stanford.edu/HPSG/2003/kathol.pdf>.
- Kee Shein Mang. 2006. *A syntactic and pragmatic description of verb stem alternation in K'chò, a Chin language*. Chiang Mai, Thailand: Payap University.
- King, Deborah. 2009. Structural and pragmatic functions of Kuki-Chin verbal stem alternations. *Journal of the Southeast Asian Linguistics Society* 1:141-157.
- Labov, William & Joshua Waletzky. 1967. Narrative analysis: Oral versions of personal experience. In *Essays on the verbal and visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*, ed. by June Helm. Seattle: University of Washington Press.
- Li, Wendan. 2014. Clause structure and grounding in Chinese written narrative discourse. *Chinese Language and Discourse* 5(2):99-145.
- Payne, Doris L. 1992. Narrative discontinuity versus continuity in Yagua. *Discourse Processes* 15(3), 375-394.
- Peterson, David A. 2007. *Applicative constructions*. (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
- Stern, Theodore. 1984. Sizang (Siyin) Chin texts. *Linguistics of the Tibeto-Burman Area* 8(1):43-58.

# THE ADJECTIVES IN MEITEILON: THE CASE OF SO CALLED ADJECTIVE *məč̣a*

Maibam Somobala Devi  
University of Hyderabad  
somobala@gmail.com

## Abstract

The paper examines the morpho-syntactic property of Meiteilon adjectives with special emphasis on the so-called adjective *məč̣a* ‘small’. Meiteilon does not have a distinct word-class of adjectives. Most of the attributive adjectives are derived from stative verb roots and occurs prenominal or postnominal. However, the exceptional case of the adjective *məč̣a* ‘small’ does not apply derivational rules and it obligatorily occurs postnominal. The question here is that if *məč̣a* ‘small’ is an adjective, why it is different from the other adjectives in terms of formation and syntactic occurrence? Is this exceptional word an adjective or some other noun-modifying word which has fixed ordering that is postnominal? This paper tries to propose that analyzing *məč̣a* as an adjective poses a problem for morpho-syntactic analysis of adjectives in general as it does not follow the derivational rules to form adjectives and it has a fixed word order (i.e. postnominal unlike other adjectives). Again, it tries to propose that *məč̣a* ‘small’ does not come under the adjective word class morphologically, syntactically and semantically. Moreover, *məč̣a* ‘small’ can be analyzed as a diminutive word as it expresses both the concrete meaning of physical size and the abstract meaning of endearment.

**Keywords:** adjective, prenominal adjective, postnominal adjective, morpho-syntax, gradability

**ISO 639-3 codes:** mni

## 1 Introduction

### 1.1 The Language

Meiteilon (the Meitei + lon ‘language’) or Manipuri is spoken basically in the state of Manipur which is in North-eastern India. It is also spoken in the neighbouring states namely Assam, Tripura, Mizoram and neighbouring countries namely Myanmar and Bangladesh by the Meitei inhabitants of these places. Meiteilon ~ Meiteiron or Manipuri, a Tibeto-Burman language, belongs to the language family of the Kuki-Chin sub-family (Grierson 1903 vol. III part III). Meiteilon, a Tibeto-Burman language is a SOV, agglutinative language. Structurally, it is a head-final language which shows left branching nodes in syntactic trees. It is also a tonal language. It is the most advanced Tibeto-Burman language spoken in India having its own developed literary language and script (archaic script). Meiteilon is the lingua franca of Manipur. Since 20 August 1992, Meiteilon became the first Tibeto-Burman language to receive recognition as a VIII Schedule language of India<sup>1</sup>.

### 1.2 About the Topic

According to the traditional grammars, an adjective is a word that describes or modifies a noun. The main syntactic role is to qualify a noun or noun phrase, giving more information about the quality of the object signified. Earlier research on the adjectives were on the notion of adjectives being a separate

---

<sup>1</sup> The VIII Schedule to the Constitution Article 344 (1) and 351 of India is the list of the official languages of the Republic of India and this list meant that the languages which are listed are entitled to representation on the Official Languages Commission.

word class but the aspects of research has been changed to look at the adjective word class as to whether it is the same for every language or not. Since every language is not the same, the composition and the structure of the language vary substantially, and it is crucial to understand whether there is a separate word class for adjectives in every language or not (Dixon, et al. 2004).

This paper aims to discuss the nature of adjectives in Meiteilon. It examines the morphological structure and the syntactic function of adjectives in Meiteilon in general with special focus on the so-called adjective *məča*.

## 2 Adjectives in Meiteilon

From Dixon, et al. (2004), we can see that in the field of studies on adjectives there is the controversy on whether all the languages have adjectives or not and the structure of adjectives is the same or not. Thus, let us discuss the nature of Meiteilon adjectives.

### 2.1 Formation of Meiteilon adjectives

Meiteilon does not have a distinct word-class of adjectives. Meiteilon has no adjective form of a word such as ‘bad’ or ‘small’ in English. The lexical items which are used as adjectives in this language are derived from stative verb roots. The independent adjectives are formed through two respective rules; (i) prefixation of marker ‘ə-’ and suffixation of NZR suffix ‘-bə ~ -pə’ to the monosyllabic stative verb root and (ii) suffixation of NZR suffix ‘-bə ~ -pə’ to the polysyllabic (mostly disyllabic) state verb root.

i) ə- + monosyllabic stative verb root+ -bə

- |     |                 |               |
|-----|-----------------|---------------|
| (1) | <i>ə-cəw-bə</i> | <i>jum</i>    |
|     | PRF-big-NZR     | house         |
|     |                 | ‘big house’   |
| (2) | <i>ə-pik-pə</i> | <i>jum</i>    |
|     | PRF-small-NZR   | house         |
|     |                 | ‘small house’ |
| (3) | <i>ə-son-bə</i> | <i>huy</i>    |
|     | PRF-weak-NZR    | dog           |
|     |                 | ‘Weak dog’    |

ii) Polysyllabic root stative verb+ -bə

- |     |                            |                    |
|-----|----------------------------|--------------------|
| (4) | <i>p<sup>h</sup>əjə-bə</i> | <i>ləy</i>         |
|     | beautiful-NZR              | flower             |
|     |                            | ‘beautiful flower’ |
| (5) | <i>nuŋay-bə</i>            | <i>pao</i>         |
|     | happy-NZR                  | news               |
|     |                            | ‘happy/good news’  |
| (6) | <i>həraw-bə</i>            | <i>mikup</i>       |
|     | joy- NZR                   | moment             |
|     |                            | ‘Joyful moment’    |

Other verbal roots with the prefix ‘ə-’ nominalizer ‘-bə ~ -pə’ can also be used as adjectives in the language.

- |     |                 |                   |
|-----|-----------------|-------------------|
| (7) | <i>ə-pai-bə</i> | <i>uček</i>       |
|     | PRF-fly-NZR     | bird              |
|     |                 | ‘The flying bird’ |
| (8) | <i>ə-cen-bə</i> | <i>huy</i>        |
|     | PRF-run-NZR     | dog               |
|     |                 | ‘The running dog’ |

- (9) *ə-tum-bə*            *huy*  
 PRF-sleep-NZR    dog  
 ‘The sleeping dog’

Again, the verbs in progressive and perfective with nominalizer ‘-bə’ can be used as adjectives in Meiteilon (Singh 2000).

- (10) *sat-li-bə*            *ləy*  
 bloom-PROG-NZR    flower  
 ‘A blooming flower’
- (11) *ca-ri-bə*            *əŋaŋ*  
 eat-PROG-NZR      child  
 ‘The eating child’
- (12) *tum-li-bə*            *əŋaŋ*  
 sleep- PROG-NZR    child  
 ‘The sleeping child’
- (13) *thoŋ-lə-bə*          *cinjak*  
 cook-PERF-NZR      food  
 ‘The cooked food’
- (14) *pət-lə-bə*            *həy*  
 rot- PERF-NZR        fruit  
 ‘The rotten fruit’
- (15) *pok- lə-bə*            *əŋaŋ*  
 birth-PERF-NZR      child  
 ‘The born child’

However, there is the exceptional case of the adjective *məčə* ‘small’, where the above-mentioned derivational rules cannot be applied, and it obligatorily occurs postnominally. We can see the detailed discussion on the so-called adjective *məčə* ‘small’ in the next section.

### 2.1.1 Exceptional case of adjective *məčə* ‘small’

The so-called Meiteilon adjective *məčə* ‘small’ (considered as an adjective due to its meaning) differs from the above-discussed adjectives which are derived from the two respective rules. It does not follow any of the formation rules and it is an independent adjective. The modifying word *məčə* ‘small’ is traditionally considered as adjectives (because of its meaning) in Meiteilon and this modifying word is different from the other type of adjectives which are derived from the stative verb roots as the above-mentioned adjective formation rule cannot apply (Devi 2017).

- (16) *huy*            *məčə*  
 dog            small  
 ‘small dog’
- (17) \* *huy*            *məčə-bə*  
 dog            small-NZR  
 ‘small dog’

From the example (6), we can see that the adjective formation rule for polysyllabic rules cannot be applied to this particular modifier which is polysyllabic in nature. Therefore, it is morphologically different from the other derived adjectives.

## 2.2 Syntactic positions of Meiteilon adjectives

According to Hofherr (2010), adjectives can appear in two main types of syntactic contexts i.e., a) as attributive adjectives which directly modifies a noun and b) as predicative adjectives as the complement of a copula. Let us see the syntactic nature of Meiteilon adjectives in the next section.

### 2.2.1 Attributive Adjectives in Meiteilon

Meiteilon adjectives formed through the application of the two derivational rules can be used as attributive adjectives which modifies the head noun. Attributive adjectives can occur prenominally and postnominally.

a) Attributive adjectives in prenominal position

- (18) a) *acawbə* *layrik*  
big book  
'big book'
- b) *phəjəbə* *ləy*  
beautiful flower  
'beautiful flower'

b) Attributive adjectives in postnominal position

- (19) a) *layrik* *acawbə*  
book big  
'big book'
- b) *ləy* *phəjə-bə*  
flower beautiful  
'beautiful flower'

However, in the exceptional case of modifier *məčə* 'small', neither of the above-mentioned derivational rules can be applied and it obligatorily occurs postnominally.

- (20) a) *layrik* *məčə*  
book small  
'small book'
- b) \**məčə* *layrik*  
small book  
'small book'

We can see from (20)-(a) and (b) that the modifier *məčə* obligatorily occurs postnominally, otherwise it is ungrammatical.

### 2.2.2 Predicative adjectives

The adjectives which are derived by the two rules and are used as attributive adjectives and the so-called adjective *məčə* can appear in the predicate position with copula '*ni*'.

- (21) a) *layrik əsi əhənbə-ni*  
book DEM new-COP  
'This book is new.'
- b) *layrik əsi ətaybə-ni*  
book DEM expensive-COP  
'This book is expensive.'
- c) *layrik əsi phəjəbə-ni*  
book DEM beautiful-COP  
'This book is beautiful.'

- d) *layrik əsi əcəwbə -ni*  
 book DEM big-COP  
 ‘This book is big.’
- e) *layrik əsi məča -ni*  
 book DEM small-COP  
 ‘This book is small.’

### 2.2.3 Distribution of more than one attributive adjective

As we have seen earlier, Meiteilon derived adjectives can occur prenominally or postnominally. They can be moved around the noun without causing any ungrammaticality (Devi 2017) as we have seen in (18) and (19). Now we will examine the distribution of more than one adjective inside the noun phrases.

- (22) a) *nuysi-bə əjoubə əpikbə huy əni si*  
 cute white small dog two DEM  
 ‘These two small white cute dogs’
- b) *əjoubə əpikbə nuysibə huy əni si*  
 white small cute dog two DEM  
 ‘These two small white cute dogs’
- c) *əpikbə əjoubə huy nuysibə əni si*  
 small white dog cute two DEM  
 ‘These two small white cute dogs’
- d) *huy nuysibə əpikbə əjoubə əni si*  
 dog cute small white two DEM  
 ‘These two small white cute dogs’

In the above examples (22a-b) all the derived adjectives are prenominally modified the head noun, and can be freely moved in the prenominal position. In (22c) we can see that adjectives can occur prenominally and postnominally and modify the head noun. Again, all the adjectives postnominally modified the head noun in (22d). Therefore, derived adjectives can be moved around the noun as well as the other adjectives. They do not have any restrictions in movements.

However, the adjective *məča* cannot move around the noun, it has a fixed postnominal position, however the noun and adjective *məča* together (N+ *məča*) can move around the adjectives. Let us consider the examples (23a-d) given in (23).

- (23) a) *nuysibə əjoubə huy məča əni si*  
 cute white dog small two DEM  
 ‘These two small white cute dogs’
- b) *huy məča nuysibə əjoubə əni si*  
 dog small cute white two DEM  
 ‘These two small white cute dogs’
- c) *nuysibə huy məča əjoubə əni si*  
 cute dog small white two DEM  
 ‘These two small white cute dogs’
- d) *\*nuysibə huy əjoubə məča əni si*  
 cute dog white small two DEM  
 ‘These two small white cute dogs’
- e) *\*huy nuysibə məča əjoubə əni si*  
 dog cute small white two DEM  
 ‘These two small white cute dogs’

As we have seen in example (23 a-c) *məča* postnominally modifies the noun *huy* ‘dog’ and always occurs together, and it is ungrammatical if they do not occur together as we can see in example (23d-e).

Thus, it is evident that adjective *məča* is different from the other adjectives morphologically and syntactically. Meiteilon has another adjective *əpikpə* which has the same meaning of *məča* ‘small’ and which is derived by the rule (1) and can occur prenominally or postnominally. The question here is that if *məča* ‘small’ is an adjective, why is it different from the other adjectives in terms of formation and syntactic occurrence? Is this exceptional word an adjective or some other noun-modifying word?

### 3 Modifying nature of Meiteilon adjectives

As far as the nature of the modifiers goes, Cinque (2010) suggests two sources of modification i.e., a) Direct Modification (DM) and b) Indirect Modification (IM). According to him, the characteristics of both types of modifiers are given below:

#### *Direct modification (DM):*

- Syntactically, adjectives which are of direct modification source involves merger of different classes of APs in the specifiers of various dedicated functional heads.
- These types of adjectives are subject to ordering restrictions. i.e. they cannot be moved around.
- Most DM are closer to the head noun.
- E.g. *electrical engineer*.

#### *Indirect modification (IM):*

- They are reduced relative clauses that are generated in the specifier of a higher functional projection.
- The ordering between the reduced relative clauses is free. i.e., movement of such type of adjective is allowed.
- IM can stay further away from the head noun (e.g. *old beautiful house*).

We have seen in Section 2.2.3 that adjectives derived through the two adjective-formation rules can move around the noun and they can be further away from the head noun. It looks like an indirect modifier which is adjoined to the noun.

- (24) *pʰəjəbə*    *əŋəŋbə*    *ləy*    *əsi*  
 beautiful    red    flower    DEM  
 ‘This beautiful red flower’
- (25) *əpikpə*    *nuŋsibə*    *yum*    *əsi*  
 small    cute    house    DEM  
 ‘This small cute house’

However, adjective *məča* ‘small’ has a fixed order which is postnominal (restriction in the movement) and it is adjacent to the head noun. It looks like a direct modifier which is head of the higher functional projection. Consider the examples (26) and (27).

- (26) *huy*    *məča*    *nuŋsibə*    *əsi*  
 dog    small    cute    DEM  
 ‘This small cute dog’
- (27) \**huy*    *nuŋsibə*    *məča*    *əsi*  
 dog    cute    small    DEM  
 ‘This small cute dog’

#### 4 Semantic property of adjectives

Hoefherr (2010) suggests *gradability* as the prototypical property of adjectives which cannot apply to other categories.

- **Gradability** = degree expressions (very or too)+adjective  
 Ex: too big  
 very good

Meiteilon derived adjectives can combine with degree words and be validated as an adjective class but the modifying word *məča* ‘small’ cannot be combined with degree expressions. Consider the examples in (28).

- (28) a) *məsi*      *yamnə*      *pʰəjəbə*      *ləy*      *ni*  
 DEM      very      beautiful      flower      COP  
 ‘This is a very beautiful flower’
- b) *məsi*      *yamnə*      *(ə)pikpə*      *yum*      *ni*  
 DEM      very      small      house      COP  
 ‘This is a very small house’
- \**məsi*      *yamnə*      *məča*      *huy*      *ni*  
 DEM      very      small      dog      COP  
 ‘This is a very small dog.’

We can see from the above example (28 a, b) that Meiteilon derived adjectives *pʰəjəbə* and *(ə)pikpə* occur with the degree word *yamnə*, however example (28c) shows that the modifying word *məča* cannot occur with the degree word.

Therefore, considering *məča* as an adjective creates a problem in analyzing Meiteilon adjectives morphologically, syntactically and semantically as it gives a different direction of analysis from the other type of adjectives. As we have seen above morphologically, it is different from the derived adjectives as it does not apply the derivative rules. Again, syntactically it is different as it has a fixed word order (postnominal) whereas the derived adjectives can occur pronominally and postnominally. Finally, semantically *məča* is different from the derived adjectives as *məča* cannot occur with degree words but the derived adjectives can occur with degree words.

From the above discussions we can assume that the word *məča* ‘small’ does not fall under the adjective word class morphologically, syntactically and semantically. In order to find out the exact word class or category, this paper tries to propose the analysis of *məča* as follows.

*məča* ‘small’ = diminutive word

According to Dahl (2005), diminutives are the words formed by derivational processes that add a semantic element having to do with size to the meaning of the word. As in English, adding the suffix ‘-ette’ to the noun *kitchen* and *cigar* yields the diminutive meaning *kitchenette* ‘small kitchen’ *cigarette* ‘small cigar’; ‘-let’ to the noun *book* and *pig* gives the diminutive forms *booklet* ‘small book’ and *piglet* ‘a young pig’. However, diminutives tend to be used in a multitude of ways, involving semantic and pragmatic elements that go far beyond the simple notion of size (Dahl 2005).

We assume that Meiteilon word *məča* is a diminutive word as it gives a diminutive meaning when we use it with both animate and inanimate nouns and kinship terms in this language. Consider the examples in (29).



- (29) a) *həinəw* ***məča***  
 mango small  
 ‘tiny mango’
- b) *əŋaŋ* ***məča***  
 kid small  
 ‘young/tiny kid’
- c) *indon* ***məča*** = (*indomča*)  
 aunty small  
 ‘Mother’s younger sister’
- d) *ice* ***məča*** = (*icemča*)  
 elder sister small  
 ‘elder sister who is younger to the eldest one’
- e) *tamo* ***məča*** = (*tamoča*)  
 elder brother small  
 ‘elder brother who is younger to the eldest one’

We can see from examples (29 a-b) that both the animate and inanimate nouns occur with *məča* which gives the meaning of ‘smaller in size’ as it deals with the physical feature of the object both animate and inanimate. However, the kinship terms with diminutive word *məča* (28c-e) expresses an endearment meaning other than the simple notion of size. As mentioned earlier, diminutives tend to be used in a multitude of ways, involving semantic and pragmatic elements that go far beyond the simple notion of size (Dahl 2005). The Meiteilon diminutive word *məča* expresses both the concrete meaning of physical size and the abstract meaning of endearment.

Thus, it is evident that *məča* gives the meaning of smaller in size or endearment to the noun or kinship terms. Therefore, this paper proposes that this particular word *məča* is a diminutive word which expresses both the concrete meaning of physical size and the abstract meaning of endearment, and like other noun-modifying words, it has the fixed postnominal position.

## 5 Summary

Meiteilon does not have a distinct word class of adjectives. They are derived from state verb roots by two respective rules, polysyllabic verb root + *-bə* and *ə-* + monosyllabic verb root+ *-bə*. Another independent word *məča* ‘small’, which is considered as an adjective due to its meaning, is available in the language which differs from the derived adjectives morphologically and syntactically. As the word *məča* ‘small’ is an independent word, the derivative rules to form adjectives cannot be applied. Meiteilon has two alternate word order for DP elements; A> N> NUM> DEM and N> A> NUM >DEM as Meiteilon adjectives can occur pronominally and postnominally, which is possible according to Cinque (2005) by movement of head noun to the higher functional heads. The derived adjectives can occur pronominally or postnominally, but the word *məča* ‘small’ obligatorily occurs postnominally. Therefore, the modifying word *məča* ‘small’ fails to fall under the adjective class morphologically and syntactically.

Again, it creates problems in identifying Meiteilon adjective as direct or indirect modifiers as proposed by Cinque (2010). According to Cinque (2010), direct modifiers are restricted in movement and occur near the head noun, whereas indirect modifiers are not restricted in movement and can occur far from the head noun. In the case of Meiteilon adjectives, derived adjectives bear the nature of indirect modifiers as they can move around the nouns, they can occur pronominally or postnominally, and they can occur far from the nouns in a noun phrase. However, the word *məča* ‘small’ seems like a direct modifier as it has a fixed position (postnominal) and always occurs next to the head noun. Otherwise, it is ungrammatical. Therefore, analyzing the word *məča* ‘small’ as an adjective creates problems in confirming the nature of Meiteilon adjectives in general.

Semantically, *məča* ‘small’ is different from the other derived adjectives as derived adjectives can combine with degree words and be validated as an adjective class, but the modifying word *məča* ‘small’ cannot be combined with degree expressions.

Finally, this paper proposes that *măča* ‘small’ can be analysed as a diminutive word form which expresses both the concrete meaning of physical size and the abstract meaning of endearment, and like other noun modifying words (determiner, classifiers, quantifiers), it has a fixed ordering which is postnominal.

### Abbreviations

A	adjective
COP	copula
DEM	demonstrative
DP	determiner phrase
N	noun
NUM	number
NZR	nominalizer
PERF	perfective
PRF	prefix
PROG	progressive

### References

- Bhat, D.N.S and M.S. Ningomba. 1997. *Manipuri grammar*. Munchen, Newcastle:Lincom Europa.
- Cinque, Guglielmo. 2005. Deriving Greenberg’s Universal 20 and Its Exceptions. *Linguistics Inquiry* 36(3):315-332.
- Cinque, Guglielmo. 2010. *The syntax of adjectives: a comparative study*. Linguistic Inquiry. Monographs. Massachusetts: MIT Press.
- Dahl, Ö. 2005. Diminutives and Augmentatives. In *Encyclopaedia of language and linguistics*, ed. by Keith Brown, 594-595. 2<sup>nd</sup> Edition. Oxford, UK: Elsevier Science and Technology.
- Devi, Maibam Somobala. 2017. Distribution of adjectives in Meiteilon, *IOSR Journal of Humanities and Social Science (IOSR-JHSS)* 22(5):18-22.
- Dixon, R. M.W. and Alexandra Y. Aikhenvald (eds.). 2004. *Adjective classes: a cross-linguistic typology*, 370. Oxford: Oxford University Press.
- Grierson, G. A. 1903. *Linguistics survey of India*, Vol. III, Part III, 1<sup>st</sup> Edition, Motilal Banarsidas. Delhi.
- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language*, ed. by J.H Greenberg, 73- 113. Cambridge, Massachusetts, and London, England: MIT Press.
- Hofherr, Patrica Cabredo. 2010. Adjectives- an Introduction. In *Adjectives- Formal analyses in syntax and semantics*, ed. by Cabredo Hofherr, Patrica & Ora Matusshanky. Amsterdam: John Benjamin.
- Singh, Chungkham Yashwanta. 2000. *Manipuri grammar*. New Delhi: Rajesh publication.

# THE POLYFUNCTIONAL FOCUS PARTICLE *PUN*<sup>53</sup> IN PENANG HOKKIEN: A CONTACT PERSPECTIVE

Jia Wen Hing  
National University of Singapore  
jwhing@u.nus.edu

## Abstract

This paper studies the focus particle *pun*<sup>53</sup> in Penang Hokkien, which has developed into a degree adverb that can occur in post-predicate position and proposes that its polyfunctionality arises from relabeling Southern Min *ma*<sup>21</sup> with the phonetic string of Malay *pun*, and semantic change.

**Keywords:** focus, degree adverb, contact, relabeling, semantic change  
**ISO 639-3 codes:** msa, sit, mis

## 1 Introduction

Penang Hokkien is spoken mostly among the Chinese population in Penang, the only Malaysian state with a Chinese majority (42%), where a myriad of languages are spoken, including English, Malay, Tamil, and Chinese dialects such as Mandarin, Cantonese, and Teochew.<sup>1</sup> Penang Hokkien is considered an overseas Southern Min dialect of Zhangzhou origin (Tan 1993; Lim 2010), with many words adapted from Malay and other local languages (Yeoh 2013; Soon 2014). Some of these words take up important grammatical meanings. This paper focuses on *pun*<sup>53</sup> -- a focus particle claimed as a loanword from colloquial Malay (Lim & Teoh 2007) that eventually reconditions the adverbial system of Penang Hokkien.<sup>2</sup>

A focus-structure is a discourse structure that partitions a clause into two parts: one that is focused or highlighted (indicated by [ ]<sub>F</sub> in this paper), and one that is backgrounded or assumed. Focus is expressed in many ways: intonational prominence (cf. Jackendoff 1972; Rooth 1996, *inter alia*); word order (cf. Baker 1995; Birner & Ward 1998), and focus particles (cf. König 1991; Hartmann 1994).<sup>3</sup> In many languages, focus particles can be associated with several readings. For example, English *even* can be analyzed as additive with an additional scalar component (see König 1991:66-73; Traugott 2006; Gast & van der Auwera 2011 *inter alia*). The focus of this paper, *pun*<sup>53</sup>, has several core functions as well, some of which are shared with Malay *pun*.

## 2 The distribution of *pun* and *pun*<sup>53</sup>

Focus particle *pun* is highly versatile and very common in colloquial Malay (Goddard 2001). Its functions have been categorized differently under different approaches. In her grammar on contemporary Malay, Haji Omar (1980) first identified five functions of *pun*: *persamaan* ‘parallelism’, *waktu* ‘temporal’, *tolak ansur* ‘concessive’, *tokokan* ‘additive’, and *kepastian* ‘determinative’ [my translation]. These functions are later regrouped by Goddard (2001), based on the Natural Semantic

---

<sup>1</sup> See *2010 Population and Housing Census of Malaysia*. Department of Statistics, Malaysia, pp.16-61. Archived on 5 February 2013.

<sup>2</sup> Notation follows the convention in Chinese tone literature (Chao 1930) where 1 signifies the lowest relative pitch and 5 highest so that 53 tone value would denote a high-to-mid falling tone. Tone IIa in Proto-Fujian Southern Min is a high-falling tone that has an allotone of high-rising 445 in Penang Hokkien. To avoid complication, tone IIa is marked as 53 in this paper. Wherever sandhi is performed, sandhi tone is indicated with -N after the citation tone, so that 53-44 would denote a high flat sandhi tone for tone IIa.

<sup>3</sup> Focus particles are sometimes called “focus adverbs” (cf. Nevalainen 1991) or “focus modifiers” (cf. Huddleston & Pullum 2002).

Metalanguage (NSM) framework, into three broad categories: topic-focus, event-focus and predicate-focus.

Based on data from naturally occurring speech collected during participant observation, I identified five core functions of *pun*, which are compared to the findings of existing literature, and validated by my consultants.<sup>4</sup>

- (1) [Dia]<sub>F</sub> *pun* mau pergi  
3SG ADD want go  
'He also wants to go.'
- (2) [Dia]<sub>F</sub> *pun* tak larat angkat meja itu  
3SG ADD NEG capable lift table that  
'Even he is not capable of lifting that table.'
- (3) Dia [dah makan tiga mangkuk nasi]<sub>F</sub> *pun* tak kenyang  
3SG already eat three bowl rice CONC NEG full  
'Although he ate three bowls of rice, he is not full.'
- (4) Dia [apa]<sub>F</sub> *pun* sanggup lakukan  
3SG what INDF will do-CAU  
'He would do anything.'
- (5) Setelah dia mandi, dia *pun* makan nasi  
After 3SG bathe 3SG CONJ eat rice  
'After he bathed, he ate rice.'

*Pun* is most frequently used to indicate focus on the constituent to its left, which is usually the subject or topic in kernel clauses (cf. Goddard 2001). *Pun* in (1) asserts that the constituent in focus (i.e. 'he') is not the only person who performs the predicate (i.e. 'want to go'), but another person must have performed it before (henceforth, ADDITIVE *pun*). *Pun* in (2) implies that someone else other than the constituent in focus (i.e. 'he') met the standard of the property denoted by the predicate (i.e. 'incapable of lifting that table'), and that 'he' was the least likely person to meet the standard (henceforth, SCALAR ADDITIVE *pun*, cf. König 1991:66-73). *Pun* in (3) denotes that the situation described in the constituent in focus (i.e. 'ate three bowls of rice') is considered an unfavourable condition for the situation described in the following clause (i.e. 'not feeling full') (henceforth, CONCESSIVE *pun*, cf. Forker 2016). *Pun* in (4) forms indefinite pronoun with interrogative word 'what' (henceforth, INDEFINITE PRONOUN *pun*). Indefinite pronoun *pun* can also be added to other items (e.g. nouns, adjectives and verbs) to express indefiniteness (cf. Weber 1989:375-376). *Pun* in (5) indicates that the meaning of the clause in which it occurs (i.e. 'ate rice') is linked to the meaning of the preceding clause (i.e. 'after he bathed') in a way that the second clause is a continuation or a consequence of the situation referred in the first clause (henceforth, CONJUNCTIONAL ADVERB *pun*, cf. Goddard, 2001:38-40).<sup>5</sup> Conjunctional adverb *pun* gives instructions to process the containing sentence in a certain context (cf. König 1991:63). Its main function is to conjoin two clauses, rather than indicate focus.

Penang Hokkien focus particle *pun*<sup>53</sup> is also versatile, but less prevalently used due to competition with similar focus particles like *ia*<sup>21</sup>/*ma*<sup>21</sup> (similar to Mandarin Chinese *yě*, cf. Hole 2004), and *to*<sup>21</sup> (similar to Mandarin Chinese *dōu*, cf. Lin 1998). There are five core functions:

<sup>4</sup> Participant observation was carried out among three native speakers of colloquial Malay, and six native speakers of Penang Hokkien who learnt standard Malay in school, and speak colloquial Malay occasionally in informal domains, e.g. among friends. Kernel clauses containing *pun* were collected and analysed, and the analysis was validated by the same pool of speakers.

<sup>5</sup> Goddard (2001) has identified several other functions of *pun*, including utterance-final *pun* and equi-complement *pun*, which give the sentence a "somewhat rhetorical effect". These functions are not considered in this paper due to its pragmatic nature.

- (6a) [i<sup>44</sup>]<sub>F</sub> **pun**<sup>53-44</sup> ai<sup>21-44</sup> k<sup>h</sup>i<sup>21</sup>  
 3SG ADD want go
- (6b) [i<sup>44</sup>]<sub>F</sub> ai<sup>21-44</sup> k<sup>h</sup>i<sup>21</sup> **pun**<sup>53</sup>  
 3SG want go ADD  
 ‘He also wants to go.’
- (7) [i<sup>44</sup>]<sub>F</sub> **pun**<sup>53-44</sup> kəŋ<sup>53-44</sup>-be<sup>21-21</sup>-k<sup>h</sup>i<sup>53</sup> hi<sup>53-44</sup> tĩũũ<sup>44</sup> tok<sup>3-4</sup>-teŋ<sup>13</sup>  
 3SG ADD lift-NEG-up that CLF table-top’  
 ‘Even he is not capable of lifting the table.’
- (8) i<sup>44</sup> [tsiaŋ<sup>4-3</sup>-liau<sup>53</sup> sã<sup>44-21</sup>-uã<sup>53-44</sup> pũũ<sup>21-21</sup>]<sub>F</sub> **pun**<sup>53-44</sup> be<sup>21-21</sup>-pa<sup>53</sup>  
 3SG eat-PF three-CLF rice CONC NEG-full  
 ‘Although he ate three bowls of rice, he is not full.’
- (9) i<sup>44</sup> [ha<sup>53-44</sup>-miŋ<sup>4</sup>]<sub>F</sub> **pun**<sup>53-44</sup> tso<sup>21-44</sup>-e<sup>21-21</sup>-ts<sup>h</sup>ut<sup>3</sup>  
 3SG what INDF do-POT-out  
 ‘He would do anything.’
- (10) tsit<sup>3-4</sup> tiau<sup>13-21</sup> sã<sup>44</sup> sui<sup>53</sup> **pun**<sup>53</sup>  
 this CLF shirt pretty DEG  
 ‘This shirt is quite pretty.’

Penang Hokkien *pun*<sup>53</sup> indicates focus on the constituent to its left. Some of its core functions are shared with Malay *pun*: ADDITION (6), SCALAR ADDITION (7), CONCESSION (8), and INDEFINITE PRONOUN (9). In (10), *pun*<sup>53</sup> functions as a compromiser or downtoner on the gradable adjective ‘pretty’, reducing the degree of ‘pretty’ to ‘less than pretty’ (henceforth, DEGREE ADVERB *pun*<sup>53</sup>, cf. Martínez, 2009).

Considering the similarities between Malay *pun* and Penang Hokkien *pun*<sup>53</sup>, existing literature considers *pun*<sup>53</sup> as a loanword of *pun* (Lim & Teoh 2007; Churchman, 2017). Lim & Teoh (2007:152) claims that most loanwords in Penang Hokkien are “retained in their entirety in terms of semantic content, phonology (excepting the superimposition of tones to the words pronounced) and formal classes”. Some loanwords experience semantic broadening (e.g. sən<sup>21</sup>.dok<sup>4</sup> ‘ladle > spoon/ ladle’; ləŋ<sup>21</sup>.kaŋ<sup>44</sup> ‘drain > drain/ trench’), while some experience semantic narrowing (e.g. ba<sup>21</sup>.lai<sup>53</sup> ‘station/ hall > police station’; bit<sup>3</sup>.tsa<sup>44</sup>.la<sup>21</sup> ‘to speak > a trial or court case’). Lim & Teoh (2007) also observes that Penang Hokkien speakers sometimes innovate on the meaning of certain Malay lexical items that they have borrowed (e.g. bu<sup>44</sup>.ta<sup>13</sup> ‘blind > all for nothing’; lu<sup>21</sup>.mak<sup>4</sup> ‘fat > flirtatious’). While these descriptions seem to explain the polyfunctionality of Penang Hokkien *pun*<sup>53</sup>, they do not explain: (i) why certain functions are shared and others not (i.e. CONJUNCTIONAL ADVERB as in 5); (ii) why *pun*<sup>53</sup> occurs after the predicate (in 6b & 10), when additive focus particles and degree adverbs typically occur before the predicate in Southern Min dialects, or (iii) how a focus particle developed into a degree adverb (as in 10).

To answer these questions, I propose relabeling as the process that introduced the phonetic string of *pun* into Penang Hokkien.

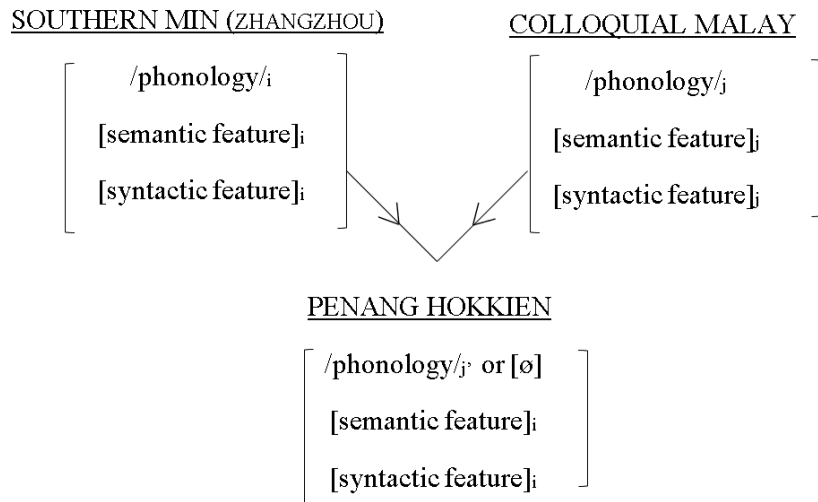
### 3 Relabeling

Relabeling involves assigning a lexical entry of a given language, a new label taken from another language. Relabeling happens in contact language of different nature, and the extent of relabeling across these languages is determined by factors including the number of languages involved, the amount of access to these languages, and whether code switching or language shift took place (Lefebvre, 2008). It is a major process in language contact, which has been greatly underestimated due to the fact that this

process is referred to in the literature by a number of terms (e.g. “relexification” in Muysken 1981; “transfer” in Siegel 2001 *inter alia*) that hide its unity and prominence in language contact situations.

The relabeling process that I propose for *pun*<sup>53</sup> involves assigning a lexical entry in the substrate language, a new phonological representation drawn from the superstrate language (cf. Muysken 1981; Lefebvre & Lumsden 1994). In the case of Penang Hokkien, the substrate language is Southern Min dialect and the superstrate language is colloquial Malay.<sup>6</sup>

**Figure 1:** Proposed relabeling process for Penang Hokkien



Relabeling is semantically driven, i.e. the semantic representations of the substrate and superstrate language entries must be partially overlapped (Muysken 1981:62).<sup>7</sup> Hence, *pun*<sup>53</sup> is likely the result of relabeling *pun* onto Southern Min *ia*<sup>21</sup>/*ma*<sup>21</sup>, which shares four core functions with Malay *pun*: ADDITIVE (11), SCALAR ADDITIVE (12), CONCESSIVE (13), and INDEFINITE PRONOUN (14).<sup>8</sup> Since *ia*<sup>21</sup>/*ma*<sup>21</sup> does not function as conjunctive adverb (15), *pun*<sup>53</sup> that follows its semantic and syntactic features does not function as conjunctive adverb as well.

- (11a)  $[i^{44}]_F$      $ma^{21-21}$      $ai^{21-44}$      $k^h i^{21}$   
 3SG    ADD    want    go
- (11b)  $*[i^{44}]_F$      $ai^{21-44}$      $k^h i^{21-21}$      $ma^{21}$   
 3SG    want    go  
 ‘He also wants to go’
- (12)  $[i^{44}]_F$      $ma^{21-21}$      $kəŋ^{53-44}$ - $be^{21-21}$ - $k^h i^{53}$      $hi^{53-44}$      $tĩũ^{44}$      $tok^{3-4}$ - $teŋ^{13}$   
 3SG    ADD    lift-NEG-up    that    CLF    table-top  
 ‘Even he is not capable of lifting the table’
- (13)  $i^{44}$      $[tsiaŋ^{4-3}$ - $liau^{53}$      $sã^{44-21}$ - $uã^{53-44}$      $pũ^{21-21}]_F$      $ma^{21-21}$      $be^{21-21}$ - $pa^{53}$   
 3SG    eat-PF    three-CLF    rice    CONC    NEG-full  
 ‘Although he ate three bowls of rice, he is not full’

<sup>6</sup> This paper adopts the position that defines ‘substrate language’ as an earlier language in a contact situation and ‘superstrate language’ as a succeeding language.

<sup>7</sup> The phonological representation derived from the superstrate language is identified as *j*’ instead of *j*, because the new phonological form associated with the lexical entry is not that of language *j*, but rather a phonological form established on the basis of a phonetic form in language *j* interpreted by the phonological principles of the language *i*.

<sup>8</sup> *ia*<sup>21</sup> is the literary counterpart of *ma*<sup>21</sup>. *ma*<sup>21</sup> is preferred in everyday conversation due to the language’s nature as a spoken language.

- (14)  $i^{44}$   $[ha^{53-44}-mi^{21}]_F$   $ma^{21-21}$   $tso^{21-44}-e^{21-21}-ts^h ut^3$   
 3SG what INDF do-POT-out  
 ‘He would do anything.’
- (15)  $*tsang^{13-21}-ho^{53-44}-ek^4$   $i^{44}$   $ma^{21-21}$   $tsiah^{4-3}-pũ^{21}$   
 wash-complete-bath 3SG eat-rice

While relabeling answers to question (i) on why certain functions are shared between *pun* and *pun*<sup>53</sup>, and others not, it does not answer to question (ii) on why *pun*<sup>53</sup> occurs after the predicate, and (iii) on how a focus particle developed into a degree adverb. As shown in Figure 1, the syntax of *pun*<sup>53</sup> must conform to that of *ia*<sup>21</sup>/*ma*<sup>21</sup>, i.e. *pun*<sup>53</sup> has to occur before the predicate to express addition, scalar addition, concession, and indefiniteness. However, additive *pun*<sup>53</sup> may occur after the predicate (as in 6b). This is unlikely induced by the syntax of Malay *pun*, but another additive focus particle might be of influence. Goddard (2001:34) observes that additive *juga* can be used interchangeably with additive *pun*. For example, the sentence ‘he also wants to go’ can be translated as *dia pun mau pergi* (as in 1), and also as in (16).

- (16a)  $[Dia]_F$  ***juga*** *mau* *pergi*  
 3SG ADD want go
- (16b)  $[Dia]_F$  *mau* *pergi* ***juga***  
 3SG want go ADD

Unlike *pun*, *juga* occurs before or after the predicate. *Juga* also occurs in conjunction with *pun*. When a sentence contains both *pun* and *juga*, *pun* always occur before the predicate, while *juga* after. In post-predicate position, *juga* functions like additive *too* in English (cf. Goddard 1986:636-637):<sup>9</sup>

- (17)  $[Dia]_F$  *pun* *mau* *pergi* ***juga***  
 3SG ADD want go ADD

Since both *pun* and *juga* are used to express additive, and additive particle occurs both before and after the predicate, it is not impossible for Penang Hokkien speakers to assume a post-predicate position for *pun* in Malay, and subsequently *pun*<sup>53</sup> in Penang Hokkien.<sup>10</sup> In other words, the use of *pun*<sup>53</sup> in Penang Hokkien involves relabeling *ia*<sup>21</sup>/*ma*<sup>21</sup> with the phonetic string of *pun*, and syntactic feature of additive particle *pun* and *juga* in Malay. The result is an additive *pun*<sup>53</sup> that occurs after verbal predicates (as in 6b), and adjectival predicates.

- (18)  $tsit^{3-4}$   $tiau^{13-21}$   $sã^{44}$   $sui^{53}$  ***pun*<sup>53</sup>**  
 this CLF shirt pretty ADD  
 ‘This shirt is also pretty’

Post-predicative additive *pun*<sup>53</sup> always occur at sentence-final position. This position is commonly filled by sentence-final particle in Penang Hokkien, as well as other Southern Min dialects. In line with this observation, the post-predicative positioning of *pun*<sup>53</sup> does not seem revolutionary. However, as will be shown in (20), *pun*<sup>53</sup> occurs after the head adjective it modifies as a degree adverb in a prenominal attributive adjective phrase.

<sup>9</sup> *Juga* is itself a polyfunctional particle, and in certain contexts convey meanings translatable into English as ‘still’, ‘all the same’, e.g., *Mau tak mau mereka naik juga* ‘Like it or not they went up all the same’ (Goddard 2001:57).

<sup>10</sup> Penang Hokkien speakers during the late 19<sup>th</sup> to 20<sup>th</sup> century are believed to speak, or to at least understand Baba Malay (Churchman, 2017), a creolized language that uses *pun* and *juga* prevalently (cf. Lim 1981; Lee 2014). Mid 20<sup>th</sup> century onwards, Penangites have to learn standard Malay as a subject in school. Hence, it is safe to assume that Penang Hokkien speakers of different times have sufficient understanding of basic Malay grammar.

#### 4 Semantic change

One might have noticed by now that although the surface structure of (18) is exactly the same with that of (10), the meanings are different. Post-adjectival *pun*<sup>53</sup> has two functions: *pun*<sup>53</sup> in (18) expresses ADDITION by denoting that *tsit*<sup>3-4</sup> *tiau*<sup>13-21</sup> *sã*<sup>44</sup> ‘this shirt’ is not the only *sã*<sup>44</sup> ‘shirt’ that is *sui*<sup>53</sup> ‘pretty’, but another *sã*<sup>44</sup> that is previously mentioned (linguistically or non-linguistically) must have also been *sui*<sup>53</sup>; *pun*<sup>53</sup> in (10) indicates DOWNTONING by reducing the degree of ‘pretty’ to ‘less than pretty’. Linguistic context plays a crucial role in the interpretation of meaning. The additive meaning surfaces only in comparative structures (19), while the downtoning meaning applies elsewhere.

- (19) *hi*<sup>53-44</sup> *tiau*<sup>13-21</sup> *sã*<sup>44</sup> *sui*<sup>53</sup>, *tsit*<sup>3-4</sup> *tiau*<sup>13-21</sup> *sã*<sup>44</sup> *sui*<sup>53</sup> ***pun***<sup>53</sup>  
 that CLF shirt pretty this CLF shirt pretty ADD  
 ‘That shirt is pretty, this shirt is also pretty’

While post-predicative additive *pun*<sup>53</sup> occurs only at the sentence-final position, downtoner *pun*<sup>53</sup> occurs at the sentence-final position, and also in the adjective phrase that occurs preminally to modify a noun.

- (20) *tsit*<sup>3-4</sup> *tiau*<sup>13-21</sup> *sui*<sup>53-445</sup> ***pun***<sup>53-44</sup> *e*<sup>13-21</sup> *sã*<sup>44</sup>  
 this CLF pretty DEG LP shirt  
 ‘This quite pretty shirt’

In Southern Min dialects, downtoning is indicated with a degree adverb that is placed before the adjective.<sup>11</sup>

- (21a) *tsit*<sup>3-4</sup> *tiau*<sup>13-21</sup> *sã*<sup>44</sup> *kai*<sup>53-44</sup> *sui*<sup>53</sup>  
 this CLF shirt DEG pretty  
 ‘This shirt is quite pretty’
- (21b) *tsit*<sup>3-4</sup> *tiau*<sup>13-21</sup> *kai*<sup>53-44</sup> *sui*<sup>53-445</sup> *e*<sup>13-21</sup> *sã*<sup>44</sup>  
 this CLF DEG pretty NOM shirt  
 ‘This quite pretty shirt’

Semantically, downtoner *pun*<sup>53</sup> resembles *kai*<sup>53</sup>. It does not, however, conform to the syntax of *kai*<sup>53</sup> or degree adverbs in general, as predicted by the relabeling hypothesis. Hence, on top of relabeling, I propose that *pun*<sup>53</sup> has undergone semantic change, and that degree adverb *pun*<sup>53</sup> stems out of post-predicative additive *pun*<sup>53</sup> via semantic change.

Both degree adverbs and focus particles are subjective in that they involve speaker’s assessment and evaluation of intensity, position on a scale, ordering of alternatives, etc. (Traugott 2006). Among the many paths of semantic changes identified in Traugott & Dasher (2002), degree adverb *pun*<sup>53</sup> exemplifies the changes as:

- i) Concrete > abstract. In general, modifiers that can be placed on a scale are abstract. In the case of *pun*<sup>53</sup>, the additive meaning is relatively objectively definable and concrete, as compared to the downtoning degree that can vary by context;
- ii) Less > more subjective. The semantic change from less > more subjective is commonly observed in the development of “intensives” and “downtoners” (Stoffel, 1901). The scalar nature of degree adverb *pun*<sup>53</sup> makes it relatively more subjective than its additive counterpart, as it is construed in the mind of the individual speaker who establishes the scale and signals how he or she conceives the degree to be ordered with respect to the scale.

<sup>11</sup> In Penang Hokkien, an adjective that is modified by intensifiers is linked to the head noun it modifies by the linking particle (LP) *e*<sup>13</sup> that behaves functionally similar to Mandarin Chinese *de* and Cantonese *ge*.



It should be noted that not all Penang Hokkien speakers use degree adverb *pun*<sup>53</sup> in non-sentence-final position. Two out of six of my Penang Hokkien consultants, aged above 50 years old, expressed doubt on sentences like (20) though both of them understood the function of *pun*<sup>53</sup> in these sentences. It is more appropriate, then, to infer that degree adverb *pun*<sup>53</sup> is a developing feature.

## 5 Summary

Many words in Penang Hokkien whose phonetic string comes from Malay have always been regarded and studied as loanwords. In fact, it is a common practice to classify non-sinitic words in Hokkien as loanwords from languages in contact, e.g. Japanese loanwords in Taiwan Hokkien, Spanish loanwords in Philippine Hokkien, English loanwords in Singapore Hokkien, etc.. Through detailed study on *pun*<sup>53</sup>, this paper proposed alternative ways to analyse these words, which better explains their polyfunctionality. The results of this study emphasize the importance of analysing a language, without being bound to its name. It is not the objective of this paper to classify Penang Hokkien, but if Penang Hokkien were to be treated as a localized version of Southern Min dialect that is loaded with loanwords, much of its unique linguistic features, including but not limited to the post-predicative position of focus particles, would be left unexplained.

## References

- 2010 *population and housing census of Malaysia*. Department of Statistics, Malaysia.
- Baker, Carl L. 1995. Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English. *Language* 71:63-101.
- Birner, Betty J., & Gregory L. Ward. 1998. *Information status and non-canonical word order in English* (Studies in language companion series 40). Amsterdam: Benjamins.
- Chao, Yuen.-Ren. 1930. A system of “tone-letters”. *Le Maître Phonétique* 45:24-27.
- Churchman, Catherine (2017). The eclectic nature of Penang Hokkien vocabulary, its historical background and implications for character writing. *Journal of Taiwanese Vernacular* 9(1):86-125.
- Forker, Diana. 2016. Toward a typology of additive markers. *Lingua* 180:69-100.
- Gast, Volker, & Johan van der Auwera. 2011. Scalar additive operators in the languages of Europe. *Language* 87 (1):2-54.
- Goddard, Cliff. 1986. The natural semantics of too. *Journal of Pragmatics* 10:635-644.
- Goddard, Cliff. 2001. The polyfunctional Malay focus particle *pun*. *Multilingua* 20:27-59.
- Haji Omar, Asmah. 1980. *Nahu Melayu Mutakhir*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Hartmann, Dietrich. 1994. Particles. In *The encyclopedia of language and linguistics*, ed. by Ronald E. Asher, & James M.Y. Simpson, 2953-2958. Oxford: Pergamon Press.
- Hole, Daniel. 2004. *Focus and background marking in Mandarin Chinese: system and theory behind cai, jiu, dou and ye*. London: Routledge.
- Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London: Routledge.
- Lee, Nala H. 2014. *A grammar of Baba Malay with sociophonetic considerations*. University of Hawai'i at Manoa: Ph.D. dissertation.
- Lefebvre, Claire. 2008. Relabelling: A major process in language contact. *Journal of Language Contact* 2(1):91-111.

- Lefebvre, Claire, & John S. Lumsden. 1994. Relexification in creole genesis. In *The central role of relexification in creole genesis: The case of Haitian Creole*. Research report prepared for SSHRCC on the project La genèse du créole haïtien: un cas particulier d'investigation sur la forme de la grammaire universelle at Université du Québec à Montreal.
- Lim, Beng S. 2010. Penang Peranakan Hokkien: A sociological reflection of a hybrid tongue. In *Penang Peranakan Chinese and Chinese Muslims: A historical and cultural journey*, ed. by Sohaimi Abdul Aziz. Penang: Persatuan Karyawan Pulau Pinang.
- Lim, Beng S., & Boon S. Teoh. 2007. Malay lexicalized items in Penang Peranakan Hokkien. *SEALS VIII: Papers from the 8th meeting of the Southeast Asian Linguistics Society*, ed. by Mark Alves, Pauls Sidwell, and David Gil, 148-165. Canberra: Pacific Linguistics.
- Lim, Sonny. 1981. *Baba Malay: The language of the "Straits-born" Chinese*. Department of Linguistics, Monash University, Melbourne: Master of Arts Thesis.
- Lin, Jo-wang. 1998. Distributivity in Chinese and its complications. *Natural Language Semantics* 6 (2):201-243.
- Muysken, Pieter C. 1981. Halfway between Quechua and Spanish: the case for relexification. In *Historicity and variation in creole studies*, ed. by Albert Valdman & Arnold Highfield, 52-78. Ann Arbor: Karoma Publishers.
- Nevalainen, Terttu. 1991. *But, Only, Just: Focusing adverbial change in Modern English 1500-1900*. Helsinki: Société Néophilologique.
- Rooth, Mats. 1996. Focus. In *The handbook of contemporary semantic theory*, ed. by Shallom Lappin, 271-297. Malden, MA: Blackwell.
- Siegel, Jeff. 2001. Koine formation and creole genesis. In *Creolization and contact*, ed. by Norval Smith & Tonjes Veenstra, 175-197. Amsterdam/ Philadelphia: John Benjamins.
- Soon, Samantha S. M. 2014. *Tonal assignment of Malay loanwords in Penang Hokkien*. Masters thesis: Nanyang Technological University.
- Stoffel, Cornelis. 1901. *Intensives and downtoners: A study in English adverbs*. Heidelberg: Carl Winter.
- Tan, Chee B. 1993. *Chinese peranakan heritage in Malaysia and Singapore*. Kuala Lumpur: Fajar Bakti.
- Traugott, Elizabeth C. 2006. The semantic development of scalar focus modifiers. In *The handbook of the history of English*, 335-359. Blackwell.
- Traugott, Elizabeth C., & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Weber, David J. 1989. *A grammar of Huallaga (Huanuco) Quechua*. University of California Publications in Linguistics.
- Yeoh, Yinyin. 2013. *Phonetics and phonology of Min dialect in Penang*. Masters thesis: Peking University.

# A COMPARISON OF GRAMMATICALIZATION IN SHAN AND THAI<sup>1</sup>

Daniel P. Loss  
Payap University  
lossdaniel224@gmail.com

## Abstract

Inspired by previous research into Tai syntactic change and grammaticalization (Diller 2001), this study looked at two closely related Tai languages – Shan and Thai – with the aim of better understanding grammaticalization in each language. The starting point for this research was the extensive literature on Standard Thai grammaticalization from which the researcher compiled several grammaticalized morphemes (grams), for the purpose of comparison with Shan. This study also adopted an approach based on Post (2007), which entailed the use of two genre-specific corpora of each language to compare the relative frequencies of analogous grams. Despite many similarities, the comparison revealed some differences between Shan and Thai grams in areas of verb-phrase and noun-phrase syntax. This study lays out the findings of this comparison by showing the areas of the comparison and findings on Shan, where it differed from Standard Thai. It ends with some brief ideas and suggestions for more comparative grammatical research.

**Keywords:** grammaticalization, Shan, Thai  
**ISO 639-3 codes:** shn, tha

## 1 Introduction

Grammaticalization, the process by which lexemes come to serve grammatical functions, provides many insights into the historical origins and processes of language. Among Southeast Asian languages, Standard Thai has been the focus of much grammaticalization research. Yet, to better frame historical developments of grammar within Tai languages and of the nature of Proto-Tai, more information about grammaticalization in languages other than Standard Thai should also be considered. Inspired from insights gained into the evolution of Thai grammatical markers, this current work seeks to extend the explanatory power of previous Thai grammaticalization research by bringing data from a related language – Shan<sup>2</sup> – into the discussion. The first obvious benefit of this type of research is it brings more attention to the Shan language – a lesser known Tai language in regards to grammaticalization, but additionally it aims to contribute to cross-linguistic research and discussion on Tai languages, by answering calls for more research in this area by Diller (2001), and studies since its publication.

The Thai language is the most studied Southeast Asian language in regard to grammaticalization – due in part to the existence of written texts and inscriptions dating back over 700 years. This historical record of language provides researchers a diachronic resource from which to create corpora and investigate grammaticalization in the Thai language. The prevalence of grammaticalization studies for Thai can be seen in Diller et al. (2008) which features an entire chapter composed of articles focused on areas of grammaticalization in Thai, but there are also several other papers and theses on the topic, many of which are written in Thai. Some of these works are directly cited within this study, but it by no means presents the entire scope of these works. From the extensive body of literature on Thai grammaticalization, dozens of grammatical morphemes or “grams” were compiled for use in data

---

<sup>1</sup> This article has been adapted from research in my Master’s Thesis of the same title from Payap University.

<sup>2</sup> The term Shan is problematic in that it’s an exonym under which several dialects or varieties are included. A discussion on issues of ethnic and linguistic classification are outside of the scope of this current study. The use of Shan throughout this paper refers to the consultants and participants of this study all who were “Tai-Yai/Tai-Long” speakers.

collection through direct elicitations, a review of published Shan texts, and a corpus-based approach. The underlying logic and rationale for this study is that evidence of diachronic change (in Thai) can serve as the context of a synchronic comparison with a related language (Shan), an inference supported by general factors of language typology as well as concepts within grammaticalization theory.

## **2 Literature Review**

Before considering the findings presented within this study some theoretical concerns should be kept in mind. Knowledge of only grammaticalization theory may not be sufficient for interpretation of the differences between Shan and Thai. This brief literature review points to important ideas worked on by Walter Bisang and Nick J. Enfield dealing with matters of Southeast Asian language change.

### ***2.1 Grammaticalization and Southeast Asian languages***

In classic models of grammaticalization, it was thought that as grams developed from their lexical origins, they would become more abstract in meaning, have less syntactical autonomy, and undergo phonetic reduction. But, as pointed out in several works of Bisang (2004, 2008), when it comes to languages in East and Mainland Southeast Asia (EMSEA), there are unique issues to consider, namely “the lack of the coevolution of meaning and form”. Essentially, research on grammaticalization in EMSEA languages must take into account the multi-functional nature of grams in these languages. One of the main typological features that explains the absence of the coevolution of meaning and form in these languages is that languages here display a lack of obligatory grammatical categories, thereby making it difficult to categorize many morphemes used in these languages. In other words, the indeterminateness of grams refers to the fact that many grams also still serve largely lexical functions. Besides posing some challenges for general grammatical analysis it also presents challenges for grammaticalization theory as well. To address these issues Bisang recommends that researchers look beyond just individual grams and include constructions —form and meaning pairing of linguistic information into their investigations. By looking to constructions, a better understanding of grams that still maintain their lexical roles can be investigated at the level of multi-word constructions.

### ***2.2 Areal versus genetic inheritance***

Typically, grammaticalization refers to a language internal process of semantic shifting and functional expansion. While grammaticalization can account for some language change, when researching any languages of Mainland Southeast Asia (MSEA) it is important to recognize the sociolinguistic factors that have shaped the languages currently inhabiting this area. In MSEA prolonged language contact between languages may be seen as responsible for structural reanalysis and other processes of language shift. These factors and other issues surrounding the topic of MSEA languages have been covered extensively in the work of Enfield (2002, 2003, 2005). Typologically languages in MSEA not only share many broad linguistic features but they also share some quite specific grammatical items and constructions (e.g. the syntax of classifier constructions). Finding grammatical similarities across the boundaries of unrelated languages makes the job of historical reconstruction more complicated than in more homogeneous language areas and reinforces the need for awareness of areal factors. Enfield points out that for linguists working with these languages “the puzzle in MSEA, like in every region, is often seen to be that of distinguishing between two kinds of cause for the existence of common structure in languages: internal vs. external change, vertical vs. horizontal transmission, descent vs. diffusion” (Enfield 2011:74). Essentially these questions are ones of determining whether some linguistic (grammatical or lexical) item is the result of areal or genetic inheritance.

The background to this issue is the unique social, linguistic, and cultural histories that have brought languages in the Austronesian, Hmong-Mien, Austroasiatic, Sino-Tibetan, and Tai-Kadai families together in close proximity and created a unique linguistic environment. Researchers must keep in mind how the history and movement of people who speak these languages may have shaped various aspects of each language’s current form. Along with how current sociolinguistic factors like diglossia and bilingual speech communities also bring more variables into play. One factor that allows for a type of diffusion of linguistic units is “typological poise” —the system internal aspects of a language that allow

it to accept new innovations. This means that for closely related languages like Thai and Shan the lines between processes of language change are likely to be blurred if not entirely illusory. These concerns should then leave researchers cautious on the ultimate status of any supposed Shan grammaticalizations. Ultimately, historical claims of language change—especially grammaticalization—requires evidence that shows bridging contexts, like has been done for cases of grammaticalization in Thai through historical-corpus data.

### 3 Verb-phrase grams

Verb-phrase grams are grammatical morphemes used in verb-phrase syntax. In Table 1, findings from the comparison of Shan and Thai verb-phrase grams are shown. Brief discussion of these grams and examples are presented in the following sub-sections.

**Table 1:** Comparison of Shan and Thai verb-phrase grams.

Shan	Thai	Function
<i>pǎn</i>	<i>hâj</i>	CAUS/BEN <sup>3</sup>
<i>tě</i>	<i>cà</i>	IRR
<i>tûk</i>	<i>kamlaŋ</i>	CONT
<i>lò (laj)</i>	<i>tôŋ</i>	MOD
<i>lǎ / se</i>	<i>kwàa</i>	COMPAR
<i>ŋáa</i>	<i>thùuk</i>	PSSV

#### 3.1 Preverbal TAM markers

Pre-verbal auxiliaries for both Shan and Thai include several possible markers of aspect and modality. Research into many Thai auxiliaries has revealed that they are often the result of grammaticalization from full verbs. For example, the Thai irrealis marker *cà* has grammaticalized from a verb meaning *càk* ‘to intend, consider’ (Diller 2001). In Shan the gram *tě* serves as an irrealis marker, and though it is not cognate with Thai *cà*, it is analogous in its grammatical functions. One general finding from the comparison of Shan and Thai showed that although there are many analogous pre-verbal markers, Shan has not been subject to the same areal influences. The Thai continuous aspectual marker *kamlaŋ* is grammaticalized from a Khmer noun ‘strength’. A similarity between Shan and Thai pre-verbal markers at the level of constructions is shown in the (1), which features the Shan ‘inceptive construction’ modifying the main verb *kĩn* ‘to eat’. This construction would be accomplished in Thai with the same structure of [CONT+IRR+V] but with each slot filled by Thai’s respective grams (i.e. *kamlaŋ cà + V*).

- (1) *háw*    *tûk*    *tě*    *kĩn*    *koj*  
 1SG    CONT    IRR    eat    banana  
 ‘I am about to eat a banana.’

The Thai modal *khuan* ‘should’ also a Khmer loan does not appear to be used in Shan, which instead uses *thùk* a verb meaning ‘to be suitable’ as a pre-verbal marker of deontic modality ‘should’. Both the Thai and Shan forms often appear along with each language’s respective irrealis marker. Shan also differs in deontic modality marking signifying a strong obligation ‘must’. Thai uses *tôŋ* a pre-verbal auxiliary, grammaticalized from a verb ‘to touch’ (Chancharu 2009; Meesat 1997). Shan speakers instead use *lò* originally a Burmese verb ‘to need’, a usage also found in Dai Lue. Another interesting case of Burmese influence is the effect Burmese may have had on the meaning/function of Shan *laj* ‘to

<sup>3</sup> Abbreviated grammatical functions in this paper refer to the following; ADJVZR=Adjectivizer, ASP=Aspect, BEN=Benefactive, CAUS=Causative, CLF=Classifier, COM=Complementtizer, COMPAR=Comparative, CONT= Continuous, CONN=Connective, IRR=Irrealis, MAL=Malefactive, MOD=Modal, NEG=Negative, NOM=Nominalizer, NUM=Numerical, POSS=Possession, PSSV=Passive, PREP=Preposition, REL=Relativizer, V=Verb

acquire’. Jenny (2009) holds that Burmese *yá* ‘to acquire’ has gone through grammaticalization to mark obligative modality. Jenny hypothesizes this development in Burmese may have influenced Shan *laj* in modal constructions indicating a sense of ‘must’, a usage not entirely analogous with the Thai cognate *dâaj*.

### 3.2 Verbs ‘to give’

The grammaticalization of *hâj* ‘to give’ is one of the most written-about cases of Thai grams. Grammaticalized uses of Thai *hâj* include a marker of benefactive, causative or malefactive clauses. Grammaticalized uses of *hâj* ‘to give’ for causative and benefactive clauses occur within certain constructions which still reveal the position of clausal reanalysis (Iwasaki & Yap 1998). The grammaticalization of Thai *hâj* ‘to give’ parallels the functional expansion of the Shan verb *pǎn* ‘to give’. Jenny (2012) also noted the use of *pǎn* in some Shan varieties for causative and benefactive clauses. Additionally, the usage in example (2) adds a syntactically similar but semantically differentiated use of *pǎn* for malefactive clauses.

- (2) *mũ kɔ má khóp pǎn kô nân kámlěw*  
 pig CONN come bite MAL CLF that immediately  
 ‘The wild pig then came to bite that person immediately.’

This case here points toward a key concept in grammaticalization theory, that of ‘clines’ – pathways of universal grammatical developments that occur cross-linguistically. Clines are the common movement of source lexical material into grammatical roles. Shan and Thai verbs ‘to give’ developing into similar grammatical markers can serve as evidence in positing clines spoken of within grammaticalization literature. A further factor to consider here though would be ‘typological poise’, a language’s readiness to calque functors from proximate languages (See Enfield 2003), the implication being that one language uses its own lexical source material to mirror a grammaticalization in another language, as opposed to having it develop naturally or genetically from within.

### 3.3 Directional Verbs

Within Tai languages the class of lexical verbs known as “directional verbs” has been noted to also serve grammatical functions. Typical directional verbs are those that refer to movements *go*, *come*, *ascend*, *descend*, *enter*, and *exit*. While all of these verbs can act as main predicates, their grammaticalized form can be seen as resultatives or success markers. Thepkanjana & Uehara (2008) see this grammaticalized usage as normally occurring in the construction [NP1 V NP2 DV].

While in general Shan and Thai are mostly similar here, a difference with respect to directional verbs is that Thai uses *khâw* ‘to enter’ and *ʔə̀k* ‘to exit’ to describe some cognitive states such as ‘understanding, thinking, legibility’ where Shan uses a verb *pə̀j* ‘to pierce’, as shown in (3).

- (3) *háw ʔàan lajmu su ʔàm pə̀j*  
 1SG read handwriting you NEG ASP  
 ‘I can’t read your handwriting.’

Though Shan does not use the same directional verbs for these constructions, the underlying metaphor seems to be similar with Thai with cognitive states being compared to an enclosed space, and understanding or comprehension is achieved by movement through or “piercing” of the barrier. The analogous Thai form of (3) would be accomplished with a verb ‘to exit’. Usage of *pə̀j* in this manner has also become lexicalized in *pə̀j cǎu* ‘pierce-heart’ or ‘understand’, again mirroring a usage of the Thai directional verb *khâw* ‘to enter’.

Besides the overall similarity in general directional verb usage, one interesting thing to point out here is that the Proto-Tai *kwà* ‘to go’ is grammaticalized as a comparative marker and temporal conjunction in Thai (Diller 2001) yet remains completely lexical in Shan. Shan like many other Tai dialects uses other words used for comparative marking (Morev 1998:97). Two identified Shan

comparative markers are *lǎ* a verb ‘to exceed’ and *se* a verb ‘to separate’, have both lexical and grammatical uses, and there also seems a functional expansion of *se* into a clause-final particle.

### 3.4 Passives

Prasithratsint (2006) outlines the gradual development of the Thai passive marker *thùuk*. Originally *thùuk* was a lexical transitive verb meaning ‘to touch’, ‘to hit off the point’, but it has since grammaticalized from a lexical verb to a full neutral passive marker. Prasithratsint (2006) holds that a stage in the development of the Thai neutral passive was that of *thùuk* occurring within serial verb constructions where the verb that followed negatively affected an animate agent. Prasithratsint believes that passive markers in other Southeast Asian languages may have developed in same way. An important takeaway from Prasithratsint’s analysis here is the role that serial-verb constructions play, as well as the distinction between adversative and neutral passives. The environment of their development is chains of serial verbs which provide a context for reanalysis and decategorialization of lexical verbs. Prasithratsint (2004) holds that for many Southeast Asian languages, verbs meaning ‘to incur’ or a related meaning are likely to act as adversative passives before grammaticalizing into full-neutral passives. Shan passive constructions are possible with the verb *náa* ‘to meet, encounter’ such as in (4), but this seems to be mostly used as an adversative passive.

- (4) *kón waan kó náa cáaη hét háaj tǎ wàat*  
 people village CONN PSSV elephant do very wound wound  
 ‘People of the village were gravely injured by a stampeding elephant.’

Even in its active verbal sense *náa* was often found as a verb ‘to meet, encounter’ within the context of an adversarial situation. Compared to the Thai *thùuk*, Shan *náa* has not achieved the same sort of semantic bleaching which characterizes later stages of grammaticalization.

## 4 Noun-phrase grams

Noun-phrase grams are grammatical morphemes used in noun-phrase syntax. In Table 2 findings from the comparison of Shan and Thai grams in noun-phrase syntax are shown. Brief discussion of these grams and some examples are presented in the following sub-sections.

**Table 2:** Comparison of Shan and Thai noun-phrase grams

Shan	Thai	Function
---	<i>nák</i>	NOM (Agent)
<i>táaη</i>	<i>khwaam</i>	NOM (Abstract)
<i>lǎη</i>	<i>kaan</i>	NOM (Gerundive)
<i>lǐ</i>	<i>nâa</i>	ADJVZR
<i>tǐ</i>	<i>thǐi</i>	REL
---	<i>khǎoη</i>	POSS

### 4.1 Nominalizers

The Thai nominal prefix *nák-* originates from the Old Khmer *anak* ‘person’ (Diller 2001:142) but is now a classifying prefix transforming verbs into agents of the action that their verb signifies. Through comparison of Thai words which are produced through the use of *nák-*, Shan was found to accomplish some similar nominalizations with the use of the word *mǎ* ‘expert’. Despite being commonly used in compounds, the Shan *mǎ* and the Thai cognate *mǎo* are still semantically transparent and can function as a free morpheme in both languages. Uses of Shan *mǎ* also a verb/pre-verbal marker have also been found, with the sense of “be able”, such as *su mǎ lat kwam sǎη pǎη* (‘you able speak language which languages’) or “Which languages can you speak?”. Grammatically the basic process of nominalizing compounds referring to agents or actors in Shan and Thai are similar. But the morpheme of *nák-* was most likely borrowed as a bound-prefix from Khmer and thus explains its productivity.

Another Thai nominalizer *khwaam* has grammaticalized from a noun meaning ‘sense or substance of a matter’ (Prasithrathsint 1997). It is found mostly on verbs dealing with emotive or mental processes forming abstract nominalizations. Prasithrathsint (1997) found *khwaam* nominalizations going as far back as the 13th century. In Shan the lexical noun *táaŋ* ‘path/road’ functions as a nominal prefix, cover similar nominalizations as *khwaam* does in Thai. The noun *táaŋ* still holds its lexical sense in Shan as does the Thai cognate *thaaŋ* ‘path’, but the Shan form has grammaticalized further into a productive prefix for abstract nominalizations. There are also examples like *táaŋ-kín* ‘food’ which is a nominalization of a verb ‘to eat’, there is no corresponding nominalized Thai form like *\*khwaam-kin*. Prasithrathsint (1997) also studied the historical development of *kaan* nominalizations (denoting an ‘action’ or ‘activity’) and though they are first found in the Ayutthaya Period, it is not until the Bangkok Period that *kaan* nominalizations regularly occur, sometime after abstract *khwaam* nominalizations developed. Shan does not use *kaan* yet similar nominalizations are done with a productive prefix *lôŋ* which denotes a topic or case, such as *lôŋ-tem* ‘writing’, or *lôŋ-jù-lôŋ-kín* ‘livelihood, way of living’.

The grammaticalization of nominalization markers in Thai are hypothesized by Prasithrathsint (1997) to be the consequence of periods of modernization in Thai society, and contact with English. With Prasithrathsint’s (1997) hypothesis on the role sociolinguistics factors have played here, we could perhaps understand some of the differences, but more research is needed on the differences between Shan and Thai in regards to frequency of overall nominalization as well as the grammaticalization of nominalizers.

#### 4.2 Adjectivizers

Adjective-forming prefixes or “adjectivizers” modify verbs and adjectives to produce adjectival words which are then used to modify nouns. Meesat (1997) holds that Thai *nâa* grammaticalized from an intransitive verb meaning “to be good, to be special” into an adjectivizer. Meesat hypothesizes that this development might have begun before the Sukhothai period. In Shan the morpheme *lǐ* ‘good’ is used to accomplish the same function of creating adjectives from verbs. Shan examples like *lǐ-kín* ‘edible/tasty-looking’, *lǐ-hak* ‘lovely, cute’ correspond to Thai constructions of [*nâa*+V]. Later on around the time of King Rama V *nâa* grammaticalized further to also have a function as an epistemic modal marker. Shan *lǐ* does not have the modal function.

#### 4.3 Corpus approach to Relativization and Possession

This study also adopted an approach based on Post (2007) which entailed the use of two corpora, for the purpose of comparing the relative frequencies of grams in each language. Use of this approach required the creation of a small topic-specific Shan corpus, made up of texts obtained through a video elicitation method. These texts provided data and several contexts for comparison of the two languages, first in generating many comparable examples and uses of language and second in also allowing frequency measurements of analogous grams to be done in each language’s corpus. A main idea for the use of frequency measurements is the idea that high token frequency can serve as an indicator of a grammatical morpheme (See Bybee, 2003). High-frequency grams are likely to be semantically bleached and have additional meanings and functions not present in their lexical forms.

##### 4.3.1 Relativization

Kullavanijaya’s (2008) investigation into the grammaticalization of *thîi* identified several functions beyond that of just a noun meaning ‘place’. For this study the functions identified in Kullavanijaya (2008) were used to classify uses of Thai *thîi* and the Shan cognate *ti* in the corpus. The overall frequency in the corpora for *thîi* and *ti* were both 1.3%, but as shown in Table 3, uses of Thai *thîi* varied more so than Shan *ti*.



**Table 3:** Comparison of functions of *thûi/ti*

	N	NUM	NOM	PREP	COM	REL	CLF
<b>Shan - ti</b>	---	---	---	31%	13%	55%	---
<b>Thai - thûi</b>	3%	2%	7%	14%	13%	54%	1%

Besides the similar uses as a relativizer and complementizer, speakers within the Thai corpus used *thûi* for functions in which the Shan speakers did not. In her study Kullavanijaya (2008) compared findings on Thai *thûi* with other Tai dialects (Mueang Hiam dialect of Lao, and Shangsi Zhuang) and concluded that Thai *thûi* has had a unique development with Thai sentences becoming “more complex with embedding and modifying constructions” (p.466). The findings here also seem to support a hypothesis of the Thai *thûi* being more ‘deeply grammaticalized’. The results of this study should not be taken as proof that Shan *ti* does not have the other functions (for example uses as an ordinal number marker have been found in other sources), but instead point out how questions on particular grams could be explored with more advanced corpus-based approaches.

#### 4.3.2 Possession

In Thai the noun *khǎwng* ‘thing’ has been grammaticalized as a genitive marker indicating possession, though a grammaticalized marker *khǎwng* is still lexical in both Thai and Shan. The Thai genitive *khǎwng* construction is [N<sub>POSSESSED</sub> [(*khǎwng*) N<sub>POSSESSOR</sub>]]<sub>NP</sub> like in (5) were a hat being possessed by the child is explicitly marked. However, the *khǎwng* possessive construction is largely optional in Thai like in example from Shan (6) which shows possession just based contiguous word order [N<sub>POSSESSED</sub> [N<sub>POSSESSOR</sub>]]<sub>NP</sub>.

(5) *mùak khǎwng dèk*                      **Thai**  
 hat      POSS      child  
 ‘The child’s hat’

(6) *mòk.hǎ luk?ǎn*                      **Shan**  
 1SG      CONT  
 ‘The child’s hat’

Though the use of *khǎwng* for possession in Thai is optional, a look at its use in both the Thai and Shan corpora shows a marked difference in how often it is used. There were a total of 20 Thai possessional NPs, 12 of which were marked with *khǎwng*. The overall frequency of *khǎwng* in Thai was 0.22% which includes 2 lexical ‘thing’ usages. In Shan *khǎwng* only appeared 3 times in the corpus with only one use marking possession. Of the 24 possessional NPs in Shan, only one was marked with *khǎwng* and the other two appearances of *khǎwng* were lexical uses. These differences in the marking of possession are shown in Table 4.

**Table 4:** Comparison of noun-phases of possession

	Shan	Thai
Genitive <i>khǎwng</i> construction	4%	60%
Contiguous word order	96%	40%

The findings here suggests that possession phrases are more likely to be marked by *khǎwng* in Thai, while if Shan *khǎwng* has grammaticalized the possessive marking function, it is not used as often and its use could possibly be the result of influence from Thai.

## 5 Conclusion

As a basis of comparison, grammaticalization proved to be useful, at least identifying several Shan grammars. But these findings can also lead to other interesting questions, concerning a broader Pan-Tai grammatical comparison which focuses on grammaticalization. It seems that with more comparative data, Tai languages have something to add to general grammaticalization theory, such as adding to the inventory of ‘clines’, a key concept in the literature. This line of thought is based on findings here which showed similar developments from similar lexical content (e.g. V ‘to be good’ > ADJVZR, V ‘to give’ > CAUS, V ‘to touch’, > PSSV etc.). Thus, while identifying differences between Shan and Thai is interesting for its own sake, it also presents areas for further exploration. With consideration of more general sociolinguistic factors, it also seems likely that with any influence on Shan from Thai, it could have the potential of one day subsuming unique Shan linguistic features. This possibility places a higher value on identifying unique features throughout Shan and its varieties, which could in turn present a glimpse at broader Tai grammatical phenomenon. In doing this type of research, the work of historical Thai researchers, and their work on grammaticalization, may give an interesting and valuable context from which to begin some discussions on these matters. Despite the largely exploratory nature of this study, it hopes to point to areas for further research.

## References

- Bisang, W. 2004. Grammaticalization without coevolution of form and meaning: The case of tense-aspect-modality in east and mainland Southeast Asia. In *What makes grammaticalization? - A look from its fringes and its components*, ed. by W. Bisang, N. P. Himmelmann & B. Wiemer, 109-138. Berlin: Mouton de Gruyter.
- Bisang, W. 2008. Grammaticalization and the areal factor: The perspective of east and mainland Southeast Asian languages. In *Rethinking grammaticalization: New perspectives*, ed. by M. J. López-Couso, & E. Seoane, 15-35. Amsterdam & Philadelphia: John Benjamins.
- Bybee, J. L. 2003. Mechanisms of change in grammaticalization: the role of frequency. In *The handbook of historical linguistics*, ed. by B. D. Joseph & R. D. Janda, 602-623. Malden, MD: Blackwell Publishing.
- Chancharu, N. 2009. *The polyfunctionality of the word form / tɔŋ / in Thai: A cognitive linguistic study*. (Unpublished Master’s Thesis, Chulalongkorn University).
- Diller, A. V. N. 2001. Grammaticalization and Tai linguistic change. In *Essays in Tai linguistics*, ed. by M.R. Kalaya Tingsabadh, & A. S. Abramson, 139-176. Bangkok: Chulalongkorn University Press.
- Diller, A. V. N., Edmondson, J. A. & Luo, Y. 2008. *The Tai-Kadai languages*. New York: Routledge.
- Enfield, N. J. 2002. Parallel innovation and ‘coincidence’ in linguistic areas: On a bi-clausal extent/result constructions of mainland Southeast Asia. In *Proceedings of the 28th meeting of the Berkeley Linguistics Society: Special session on Tibeto-Burman and Southeast Asian linguistics*, ed. by P. Chew, 121-128. Berkeley: Berkeley Linguistics Society.
- Enfield, N. J. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge Curzon.
- Enfield, N. J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34, 181-206.
- Enfield, N. J. 2011. Linguistic diversity in mainland Southeast Asia. In *Dynamics of human diversity: The case of mainland Southeast Asia*, ed. by N. J. Enfield, 63-80. Canberra: Pacific Linguistics.
- Iwasaki, S., & Yap, F. H. 1998b. ‘Give’ constructions in Thai and beyond: A cognitive and grammaticalization perspective. *Proceedings of the 1st International Conference on Tai studies*, 371-382.
- Jenny, M. 2009. Modality in Burmese: ‘may’ or ‘must’ – grammatical uses of yá ‘get’. *Journal of the Southeast Asian Linguistics Society* 1:111-126.

- Jenny, M. 2012. *The far west of Southeast Asia – ‘give’ and ‘get’ in the languages of Myanmar*. Paper presented at Mainland Southeast Asia - State of the art, MPI Leipzig.
- Kullavanijaya, P. 2008. A historical study of /thûi/ in Thai. In *The Tai-Kadai languages*, ed. by Diller, A. V. N. et al., 445-467. London: Routledge Press.
- Meesat, P. 1997. *A study of auxiliary verbs developed from verbs in Thai*. Unpublished Master's Thesis, Chulalongkorn University.
- Morev, L. N. 1998. Expressing comparison in the Tai languages. *Mon-Khmer Studies* 28:91-100.
- Post, M. W. 2007. Grammaticalization and compounding in Thai and Chinese: A text-frequency approach. *Studies in Language* 31(1):117-175.
- Prasithrathsint, A. 1997. Emergence and development of abstract nominalization in Standard Thai. In *Southeast Asian linguistics studies in honor of Vichin Panupong*, ed. by Arthur S. Abramson. 179-190. Chulalongkorn University Press.
- Prasithrathsint, A. 2004. The adversative passive marker as a prominent areal feature of Southeast Asian languages. In *Papers from the 11th annual meeting of Southeast Asian Linguistics Society*, ed. by S. Burusphat, Tempe, Arizona: Arizona State University, Program for Southeast Asian Studies, Monograph Series Press.
- Prasithrathsint, A. 2006. Development of the thùuk passive marker in Thai. Passivization and typology: Form and function. In *Proceedings of the workshop on Passives at the 20th Scandinavian Conference of Linguistics*, ed. by W. Abraham & L. Leisio [Typological Studies in Language]. Amsterdam: John Benjamins.
- Thepkanjana, K., & Uehara, S. 2008 Directional verbs as success markers in Thai: Another grammaticalization path. In *The Tai-Kadai languages*, ed. by Diller, A. V. N. et al., 484-506. London: Routledge Press.

# COMPOUND STRESS IN THAI AND EFFECT OF COMPOSITIONALITY

Lena Maluleem  
*Chulalongkorn University*  
*lynamaluleem@gmail.com*

## Abstract

The present study provides empirical acoustic evidence and phonological analysis of compound stress in Thai by comparing high and low compositionality compounds with disyllabic phrases and monomorphemic words. Duration and intensity were measured as the acoustic correlates of stress. The results indicate differential stress patterns between compounds with different degrees of compositionality. Low compositionality compounds have secondary-primary stress pattern, identical to monomorphemic words. High compositionality compounds on the other hand show primary-primary stress pattern identical to disyllabic phrases despite acoustic differences due to effects of phrasal prominence.

**Keywords:** phonology, acoustic phonetics, compound, stress, compositionality

**ISO 639-3 codes:** tha

## 1 Introduction

Stress has been suggested as one of the characteristics which differentiate compound words from phrases in Thai (Iangubol 1982; Injan 2015). Disyllabic phrases bear primary stress on both syllables (Iangubol 1982; Injan 2015) while compounds have been claimed to either bear unstressed-primary stress (Naksakul 1981) or secondary-primary stress pattern (Peyasantiwong 1986). The analysis of compound stress is controversial because different stress patterns exist among compounds (Injan 2015; Naksakul 1981; Peyasantiwong 1986).

The present study compared high and low compositionality compounds with disyllabic phrases and monomorphemic words to examine whether compounds with different degrees of compositionality demonstrate distinct patterns of stress. Taking duration and intensity to be the cues to rhythmic stress, high compositionality compounds are predicted to have roughly equal duration and intensity on both syllables, indicating a primary-primary stress pattern. Low compositionality compounds, in contrast, are hypothesized to show shorter duration and lesser intensity on the first syllable, showing a secondary-primary stress pattern. Moreover, the stress pattern of high compositionality compounds is expected to be identical with those of disyllabic phrases while low compositionality compounds have the same pattern as disyllabic monomorphemic words.

## 2 Literature Review

### 2.1 Stress

According to metrical and prosodic phonological frameworks (Hayes 1995; Kager 1995; Nespor 2007; Nespor & Vogel 1986), stress is the manifestation of prominence applied to the syllables in each level of the prosodic domain and can be acoustically realized in longer duration, greater intensity, and higher fundamental frequency ( $f_0$ ). The metrical grid in Figure 1 marks the prominence and represents the levels of stress. The syllable showing the highest grid marks is the most prominent and considered to carry the primary stress. The lower ones show secondary stress and unstressed.

**Figure 1:** Metrical representation of word stress (Goldsmith, 1976; Liberman & Prince, 1977)

(x) (x)	(. x)	(. x)	prosodic word
(x) (x)	(x) (x)	(. x)	foot
(x) (x)	(x) (x)	(x) (x)	syllable
primary stress	secondary stress	unstressed	

Acoustically, stress is realized as a cluster of duration, intensity, and f<sub>0</sub>. Some languages consider duration as the most salient cue e.g. Catalan (Ortega-Llebaria & Prieto 2011), Uyghur (Yakup & Sereno 2016), and Dutch (Sluijter & Van Heuven 1996) while in some other languages, stress is realized only through the fundamental frequency e.g. Japanese (Beckman 1986).

## 2.2 Stress in Thai

From a phonological point of view, Thai is an iambic language in which the highest prominence in the phonological word is always on the right-most syllable (Bennett, 1994, 1995). Phonetically speaking, among potential cues of stress reported in the literature, duration is proposed to be the most salient (L-Thongkum, 2011; Nitisaroj, 2004; Potisuk, Gandour, & Harper, 1996). In addition, Potisuk et. al (1995) finds that intensity, though not very consistent, seems to relate to stress in some contexts. As for f<sub>0</sub>, average f<sub>0</sub> does not indicate stress, but each lexical tone in stressed and unstressed syllables are realized with different f<sub>0</sub> contours. Based on these findings, duration and intensity are measured in this paper. The f<sub>0</sub> contours, although considered another indicator of stress in Thai, are not included in this analysis because the lexical tone of target words could not be controlled.

As for compound stress, one controversial issue is how many levels of stress exist in compounds. While some authors (Naksakul, 1981; Peyasantiwong, 1986) describe the first syllable of compounds as having secondary stress, others do not posit intermediate levels, thus labeling the syllable as unstressed (Injan, 2015). Another issue is whether all compounds show the same stress pattern. Peyasantiwong (1986) has categorized compounds into two types based on their stress patterns. Subordinate compounds are constructed from words with different meanings and bear a weaker stress on the first syllable, e.g. /tôn. 'má:j/ “tree.” Coordinate compounds, on the other hand, are built from words with the same meaning and show strong stress on both syllables, e.g. /'nê:. 'chát/ “clear, accurate.” In contrast, Injan (2015) observes that this type of compound usually bears primary stress only on the second syllable. Only a few cases carry primary stress on both syllables, e.g. /'kāk. 'kān/ “detain.” I observe that all the words listed as exceptions are those whose meaning is related to both of their components, e.g. /'dī. 'ŋā:m/ “good + beautiful” = “good” or “fine.” On the other hand, the other stress pattern also includes the words which have different meaning from its parts, e.g. /tôn. 'tō:/ “(tree) trunk + stump” = “source.” This seems to indicate that stress patterns are not based on the type of compounds but vary according to their degrees of compositionality, the degree in which a complex expression’s meaning is determined by the meanings of its parts (Pelletier, 1994). Therefore, the hypotheses of this study are based on the idea of compositionality.

## 3 Methods

### 3.1 Materials

The word list includes 20 high compositionality compounds, 20 low compositionality compounds, 7 disyllabic monomorphemic words, and 7 disyllabic phrases. Because syntactic and semantic heads may affect stress assignment (Samek-Lodovici, 2005), only exocentric compounds, i.e. ones whose heads could not be identified, are included to make sure the difference comes from the degree of compositionality. Monomorphemic words and disyllabic phrases were selected based on 7 syllable structures including CVV, CVO, CVG, CVN, CVVO, CVVG, and CVVN. Disyllabic phrases include four verb phrases with noun phrase modifiers and three noun phrases with adjective phrase modifiers. Syllable structures were kept the same for all word types. All words had high-mid, or low-mid vowels to control for the possible intrinsic duration of vowels. Initial consonants could not be controlled.

However, as initial consonants may have effects on syllable duration (L-Thongkum, 2011), place, manner, and voicing of the initial consonants were included as random effects in linear mixed effects regression model.

The compositionality rating is based on 3 Bangkok Thai native speakers' judgement (which is the different group from those participated in recording session). This pilot study aims primarily to study compound only at the two ends of the compositionality continuum which has been suggested to be gradient with no actual boundary between high and low degree (Bybee, 2014). Participants were presented the word lists of known compounds and were asked to judge whether the compound as a whole is similar in meaning to their constituents. This would categorize the words into two groups of high and low compositionality. Only words with unanimous judgement from all participants were chosen.

### 3.2 Procedure

In this study, the control variables were the native dialect and age of the participants. Four female Thai undergraduate students at Kasetsart University, aged between 19-22, born and living in Bangkok, participated in this study. The target words were embedded in a carrier sentence /phû:t wâ: \_\_\_ ɿ:k thī:/ or /phû:t wâ: \_\_\_ ná?/ 'Say \_\_\_ again,' to avoid final lengthening at intonational phase boundaries (Dechapiatskul, 2014; L-Thongkum, 2011). The participants were presented with the target sentences in randomized order. They were requested to recite the sentences at an average speed as speaking rate can affect syllable duration (Gros, Pavesic, & Mihelic, 1997; Kessinger & Blumstein, 1998). There are totally 648 tokens recorded (4 participants, 54 words, 3 times). The recordings were conducted in a silent office room. The sampling frequency was at 44.1 kHz.

### 3.3 Measurements

Syllable duration and intensity were measured using Prosodic Praat Script (Elvira-Garcãa, 2014). Each token was segmented after the offset of the preceding vowel. The end of nasal consonants was at the offset of the nasal murmur. The end of obstruents was marked at the beginning of formant structure for sonorants or at the beginning of fricative noise. To normalize duration and intensity, the two values were transformed into duration ratio and intensity ratio of the first to the second syllable.

$$\left(\text{duration ratio} = \frac{\text{1st syllable duration}}{\text{2nd syllable duration}}\right), \left(\text{intensity ratio} = \frac{\text{1st syllable intensity}}{\text{2nd syllable intensity}}\right)$$

## 4 Results

To test whether compositionality and word types could be predicted by duration and intensity ratio, they were analyzed using a Linear Mixed Effects Regression Model (LMER) using the lmerTest package in R software (Kuznetsova, Brockhoff, & Christensen, 2015). The indicators are duration ratio and the intensity ratio of the first syllable. For each indicator, fixed effects are *compositionality* (high or low), and *word types* (compound, monomorphemic word, and phrase). The random intercepts are *places*, *manners*, *voicing* of the initial consonants, and *speaker*.

**Table 1:** Compounds' compositionality predicted by duration ratio

	Estimate	Std. error	df	t-value	p-value
(Intercept)	0.99574	0.06050	16.458	16.458	< .001
CompositionLow	-0.06999	0.0218	474.3	-3.288	<b>.001</b>

**Table 2:** Word types predicted by duration ratio

	Estimate	Std. error	df	t-value	p-value
(Intercept)	0.88776	0.03401	197.2	26.104	< .001
WordtypeHigh	-0.05166	0.02192	578	-2.357	<b>.01</b>
WordtypeLow	-0.06430	0.02201	579	-2.922	< <b>.01</b>
WordtypeMono	-0.23519	0.04100	579	-5.737	< <b>.001</b>

The LMER results reveal the duration is a significant predictor of a compound's compositionality (Table 1) as well as word types (Table 2). In contrast, the intensity ratio shows no significance (see Tables 3 and 4).

**Table 3:** Compounds' compositionality predicted by intensity ratio

	Estimate	Std. error	df	t-value	p-value
(Intercept)	<.001	<.001	109.5	N/A	< .001
CompositionLow	<.001	<.001	<.001	-0.472	0.63694

**Table 4:** Word types predicted by intensity ratio

	Estimate	Std. error	df	t-value	p-value
(Intercept)	<.001	<.001	<.001	N/A	< .001
WordtypeHigh	<.001	<.001	578	-1.607	0.10869
WordtypeLow	<.001	<.001	579	-1.465	0.14354
WordtypeMono	<.001	<.001	579	1.084	0.27873

#### 4.1 Duration

To identify the stress pattern, the duration and duration ratio of each syllable of all word types were compared using Tukey's HSD Pairwise Post-hoc Test. Means and standard deviations for all duration variables are shown in Table 5. The first syllable of compounds and monomorphemic words are shorter than its second syllable. A huge difference is found in light initial monomorphemic word. The first syllable is 195 ms shorter than its second syllable ( $F(1, 22) = 963.54, p < .001$ ). The first syllable of heavy initial monomorphemic word and low compositionality compounds also show a shorter duration compared the second syllable ( $F(1, 252) = 1.86, p < .001$ ). However, the differences are only 26 ms and 27 ms respectively. On the other hand, both syllables of high compositionality compounds are about the same duration (316 vs 317 ms,  $F(1, 472) = 0.08, p = 0.7$ ). Moreover, disyllabic phrases show a significantly longer duration of the first syllable ( $F(1, 166) = 4.36, p = .03$ ).

**Table 5:** Mean durations (ms) by word types and compositionality, averaged across speaker. Standard deviation in parentheses.

Word types	Syllable weight/ Compositionality	1 <sup>st</sup> syllable	2 <sup>nd</sup> syllable
Monomorphemic word	Light initial	187.25 (19)	382.35(16)
	Heavy initial	283.34 (19)	309.11 (39)
Compound	Low compositionality	302.31 (45)	329.49 (52)
	High compositionality	316.10 (49)	317.37 (51)
Phrase		325.81 (39)	311.15 (50)

Focusing on the initial syllable, their average duration ratio varies according to word types and syllable structures. The initial syllables in monomorphemic words with light initial syllables e.g. [thā.nǒn] 'road' has a significantly smaller duration ratio than the heavy initial monomorphemic words ( $p < .001$ ). Crucially, it is the shortest and the least prominent compared to initial syllables in all other word types (duration ratio = 0.49).

**Table 6:** Mean duration ratio of the first to the second syllable

Word types	duration ratio	duration ratio
Mono (light initial)	0.49	1
Mono (heavy initial)	0.93	1
Low compositionality	0.92	1
High compositionality	0.99	1
Phrase	1.06	1

Interestingly, the average duration ratio of the first syllable of heavy initial monomorphemic words e.g. [sām.li:] ‘cotton’ is almost equal to low compositionality compounds ( $F(1, 310) = 1.01, p = 0.3$ ). The duration ratios in both word types are significantly different from the first syllable of monomorphemic word with light initial ( $F(1, 249) = 63.31, p < .001$  for low compositionality compounds and  $F(1, 27) = 70.94, p < .001$  for monomorphemic words with heavy initial). In other words, they are more prominent than the first syllable of monomorphemic word with light initial. In the case of compounds, the two types of compounds display a significantly different duration of the first syllable. The mean duration ratio of low compositionality compounds is significantly shorter than another type ( $F(1, 476) = 18.82, p < .01$ ). As for disyllabic phrases, a different pattern is observed. Their first syllable is more prominent as it shows significantly greater duration than the second syllable.

To summarize, the pairwise test of duration ratio across word types revealed three different levels of prominence. The strongest prominence found in the second syllable of all word types and the first syllable of high compositionality compounds. The first syllable of low compositionality compounds and heavy initial monomorphemic words shows an intermediate prominence. Finally, the least prominence is in the first syllable of light initial monomorphemic words. The acoustic results support the hypotheses that compounds with different degree of compositionality have different stress patterns and that low compositionality compounds have similar prominent patterns as heavy initial monomorphemic word. However, high compositionality compounds and disyllabic phrases reveal the acoustically different patterns contrast to what has been proposed in the hypothesis. A detailed discussion of stress patterns is provided in Section 5.

#### 4.2 Intensity

The LMER results show the significance of neither compositionality nor word types on intensity ratio (see Tables 3 and 4) and there is no correlation between duration and intensity either. The light initial syllable of monomorphemic words which is hypothesized to be an unstressed syllable, according to their syllable structures, also show no significantly different intensity ratio. This probably suggests that intensity is not a cue to stress unless the difference between unstressed and stressed syllables should be observed.

**Table 7:** Mean intensity (dB) by word types and compositionality, averaged across speakers. Standard deviation in parentheses.

Word types	Syllable weight/ Compositionality	1 <sup>st</sup> syllable	2 <sup>nd</sup> syllable	p-value
Mono morphemic word	Light initial	73.66 (1)	70.66 (4)	.002
	Heavy initial	76.82 (2)	73.52 (3)	.002
Compound	Low compositionality	80.16 (3)	79.17 (3)	.004
	High compositionality	79.65 (3)	78.46 (4)	.001
Phrase		81.73 (3)	78.65 (4)	<.001



**Table 8:** Mean intensity ratio of the first to the second syllable

Word types	1 <sup>st</sup> syllable	2 <sup>nd</sup> syllable
Mono (light initial)	1.04	1
Mono (heavy initial)	1.04	1
Low compositionality	1.01	1
High compositionality	1.01	1
Phrase	1.04	1

The pairwise test shows that the mean intensity of the first syllable is significantly greater than the second syllable in all word types (Table 7). The greatest significance was found between the first and the second syllable of disyllabic phrases ( $F(1, 166) = 30.16, p < .001$ ). The significantly lower intensity of the second syllable might be affected by the decreasing energy at the end of the sentence. The decreasing intensity at the boundary of prosodic domain is also reported by Lesure & Clements (2015) in Ch'ol language.

## 5 Discussion

The acoustic results suggest three different patterns of prominence (Table 9), which can be analyzed as different stress profiles. High compositionality compounds show equal prominence on both syllable (2 2) and are thus analyzed as primary–primary word stress pattern. Light initial monomorphemic words bear the least prominence on the first syllable and the strongest one on the second syllable (0 2). It is thus an unstressed–primary stress pattern. Monomorphemic words with heavy initial and low compositionality compounds manifests in an intermediate prominence on the first syllable and the strongest prominence on their second syllable (1 2), which is secondary–primary stress pattern. The metrical structures of all word types are shown in Figure 4.

**Table 9:** The levels of prominence in all word types  
(0=the least prominence, 1= intermediate prominence and 2=the most prominence).

Word types	Syllable weight/ Compositionality	Levels of prominence	
		1 <sup>st</sup> syllable	2 <sup>nd</sup> syllable
Monomorphemic word	Light initial	0	2
	Heavy initial	1	2
Compound	Low compositionality	1	2
	High compositionality	2	2
Phrase		2	2

In the Metrical and Prosodic Phonological frameworks (Hayes, 1995; Kager, 1995; Nespor, 2007; Nespor & Vogel, 1986), stress is the manifestation of rhythmic structure. The acoustically different stress patterns appear to reflect a different metrical structure between two types of compounds. The primary-primary stress pattern of high compositionality compounds indicates that the highest degree of metrical prominence is marked on both syllables (Figure 2). Assuming obligatoriness and culminativity, two generally accepted properties of word stress (Hyman, 2006, p. 239; 2009, p. 217; 2014, p. 60), it is possible that each member of high compositionality compounds forms separate prosodic words manifested with primary stress on both syllables. In contrast, both syllables of low compositionality compounds and heavy initial disyllabic morphemes belong to the same prosodic word. The highest degree of prominence is thus applied only to the rightmost syllable following the iambic law and manifested in a primary stress. The other syllable is marked with less prominence and realized as a secondary stress. The acoustic results further suggest differential prosodic structure between high compositionality compounds and disyllabic phrases. The first syllable of disyllabic phrase is significantly longer than their second syllable which is probably the effects of phrasal prominence. In

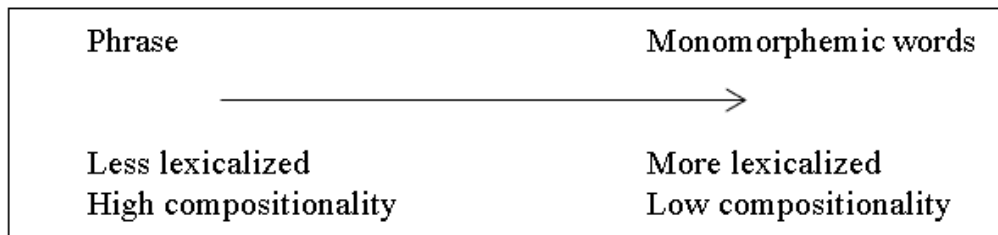
the prosodic phrase domain, the prosodic head appears to be on the initial syllable aligned with a syntactic head and manifested in the significantly longer duration of the first syllable.

**Figure 2:** Proposed metrical structure of all word types

			(x .)	prosodic phrase
(. x)	(. x)	(x) (x)	(x) (x)	prosodic word
(. x)	(x) (x)	(x) (x)	(x) (x)	foot
(x) (x)	(x) (x)	(x) (x)	(x) (x)	syllable
Mono	Mono/Low	High	Phrase	
(light initial)	(heavy initial)			

The different compound stress patterns can be explained in terms of lexicalization. Following Brinton and Traugott’s “lexicalization as fusion” approach (Brinton & Traugott, 2005), lexicalization refers to a set of changes in which a complex construction becomes simpler, i.e. syntagm (i.e. syntactic phrase) > lexeme or compound/complex words > simplex word. The processes relate to parallel changes in several levels of grammar, including the decrease in compositionality, loss of syntactic variability, development of morphological fixedness, and phonetic erosion, typically in a less prominent syllable. Compounds are known with the universal tendency to go through lexicalization processes, most of the time resulting in a monomorphemic-like structure, e.g. gospel < OE god “good” + spell “tidings,” (Brinton & Traugott, 2005:50). High and low compositionality compounds represent the point near the beginning and the end of the lexicalization pathway in which high compositionality words have undergone only some degree of lexicalization and low compositionality reflects highly lexicalized construction, as schematized in Figure 3.

**Figure 3:** Lexicalization and compositionality (adapted from Brinton & Traugott, 2005)



In the first stage, as represented by high compositionality compounds, both elements form their own prosodic word. After going through lexicalization, the prosodic word boundary between each member seems to become less transparent or, in some extreme cases, erased so that the two constituents of low compositionality compounds come to belong to the same prosodic word. The explanation is supported by the similar stress pattern between low compositionality compounds and heavy initial monomorphemic words. It is likely that the morphological and prosodic structure of low compositionality compounds may have been lexicalized and thus reanalyzed as one disyllabic morpheme.

The acoustic characteristics reflect at least three distinct levels of stress in Thai; unstressed, secondary stress, and primary stress. The second syllable of disyllabic words is always primary stress, as predicted by iambic law, consistent with all previous studies (Bennett, 1994, 1995; Naksakul, 1981; Peyasantiwong, 1986). In contrast, the first syllable varies according to word type, syllable structure and compositionality. This study argues that the compound stress patterns significantly relate to the words’ compositionality. However, this paper is based only on very small number of speakers (n = 4). The results should be considered preliminary. For future works, types of compounds (i.e. coordinate/subordinate, exocentric/endocentric), may also have to be investigated together with compositionality and further studies that take intermediate stages of lexicalization into account are needed to fully understand compound stress in Thai.

## 6 Conclusion

This paper shows that there is a significant effect of compound compositionality on their stress patterns. The acoustic results also confirm the hypotheses that low compositionality compounds bear secondary-primary stress and that high compositionality compounds bear primary stress on both syllables. There is also similarity between low compositionality compounds and heavy initial monomorphemic word which might reflect a similar morphological structure between these two types. For disyllabic phrases, their word stress pattern is identical to high compositionality compounds. However, the phrasal stress seems to be on the first syllable which makes disyllabic phrases acoustically different from compounds with high compositionality.

## References

- Beckman, M. E. 1986. *Stress and non-stress accent* (Vol. 7). Walter de Gruyter.
- Bennett, J. 1994. Iambicity in Thai. *Studies in the linguistic sciences* 24(1/2):39-57.
- Bennett, J. 1995. *Metrical foot structure in Thai and Kayah Li: Optimality-theoretic studies in the prosody of two southeast Asian languages*. (Ph.D.), University of Illinois at Urbana-Champaign, Urbana, IL. Retrieved from Full text -- IDEALS <http://hdl.handle.net/2142/22242>
- Brinton, L. J., & Traugott, E. C. 2005. *Lexicalization and Language Change*: Cambridge University Press.
- Bybee, J. 2014. Analytic and holistic processing in the development of constructions. *Language in Interaction: Studies in honor of Eve V. Clark* 12.
- Dechapiatskul, M. 2014. *Prosodic Constituents in Thai: Syntax-Prosody Mapping and Boundary Strength*. (Doctor of Philosophy program in Linguistics), Chulalongkorn University, Bangkok.
- Elvira-Garcãa, W. 2014. *Prosodic-data-extraction v2.1*. [Praat script].
- Goldsmith, J. A. 1976. *Autosegmental phonology* (Vol. 159). Citeseer.
- Gros, J., Pavesic, N., & Mihelic, F. 1997. *Syllable and Segment Duration at Different Speaking Rates in the Slovenian Language*. Paper presented at the Fifth European Conference on Speech Communication and Technology.
- Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*: University of Chicago Press.
- Hyman, L. M. 2006. Word-prosodic typology. *Phonology*:225-257.
- Hyman, L. M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31(2):213-238.
- Hyman, L. M. 2014. Do all languages have word accent. *Word accent: Theoretical and typological issues*:239.
- Iangubol, A. 1982. *An Analytical study of Compound Words in Thai*. (Master of Arts), Chulalongkorn University, Bangkok.
- Injan, A. 2015. Coordinate Compounds in Thai.
- Kager, R. W. J. 1995. The metrical theory of word stress. *Blackwell handbooks in linguistics* 1:367-402.
- Kessinger, R. H., & Blumstein, S. E. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26(2):117-128.
- L-Thongkum, T. 2011. Thai Sounds: An Acoustic Study. In (pp. 330). Bangkok: Chulalongkorn University.
- Lesure, C., & Clemens, L. 2015. *Prosodic boundary marking in ch'ol: Acoustic indicators and their applications*.
- Lieberman, M., & Prince, A. 1977. On stress and linguistic rhythm. *Linguistic inquiry* 8(2):249-336.
- Naksakul, K. 1981. *Thai phonological systems*. Bangkok: Chulalongkorn University.
- Nespor, M. 2007. *Prosodic phonology: with a new foreword*. New York; Berlin: Mouton de Gruyter.

- Nespor, M., & Vogel, I. 1986. *Prosodic phonology*: Dordrecht, Holland: Foris Publications; Riverton, N.J.: Foris Publications.
- Nitisaroj, R. 2004. Perception of stress in Thai. *The Journal of the Acoustical Society of America* 116(4):2645-2645.
- Ortega-Llebaria, M., & Prieto, P. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, 0023830910388014.
- Pelletier, F. J. 1994. The principle of semantic compositionality. *Topoi* 13(1):11-24.
- Peyasantiwong, P. 1986. *Stress in Thai*. Paper presented at the Papers from a Conference on Thai Studies in Honor of William J. Gedney. Michigan Papers on South and Southeast Asia, Center for South and Southeast Asian Studies, University of Michigan, Ann Arbor.
- Potisuk, S., Gandour, J., & Harper, M. P. 1996. Acoustic correlates of stress in Thai. *Phonetica* 53(4):200-220.
- Samek-Lodovici, V. 2005. Prosody–syntax interaction in the expression of focus. *Natural Language & Linguistic Theory* 23(3):687-755.
- Sluijter, A. M., & Van Heuven, V. J. 1996. *Acoustic correlates of linguistic stress and accent in Dutch and American English*. Paper presented at the Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on.
- Yakup, M., & Sereno, J. A. 2016. Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association* 46(01):61-77.

### Appendix: Word list

No.	Word	Meaning	Word types
1	dèn.c <sup>h</sup> át	‘obvious’	high
2	p <sup>h</sup> óp.hě̀n	‘see’	high
3	p <sup>h</sup> ı̄w.nǎŋ	‘skin’	high
4	rēw.wāj	‘quick’	high
5	k <sup>h</sup> ̄j.j.c <sup>h</sup> ̄ɪn	‘familiar’	high
6	t <sup>h</sup> ̄áj.t <sup>h</sup> ̄õ:n	‘redeem’	high
7	bèŋ.pān	‘share’	high
8	k <sup>h</sup> ón.hǎ:	‘find’	high
9	sō:n.rén	‘hide’	high
10	nê:.c <sup>h</sup> át	‘clear’	high
11	nâ:.tā:	‘visage’	high
12	k <sup>h</sup> á:.k <sup>h</sup> ǎ:j	‘trade’	high
13	dī:.ŋā:m	‘good’	high
14	lì:k.nǐ:	‘escape’	high
15	suǎj.ŋā:m	‘beautiful’	high
16	k <sup>h</sup> wâ:ŋ.pā:	‘throw’	high
17	bâ:n.ruān	‘house’	high
18	lē:w.sā:m	‘wicked’	high
19	suām.sàj	‘wear’	high
20	khát.luāk	‘select’	high
21	lín.c <sup>h</sup> ák	‘drawer’	low
22	kêʔ.dām	‘black sheep’	low
23	p <sup>h</sup> ǎw.k <sup>h</sup> õn	‘at close range’	low

No.	Word	Meaning	Word types
24	wāj.fāj	‘flammable (person)’	low
25	tā.j.t <sup>h</sup> ŭn	‘basement’	low
26	sāj.duān	‘earth worm’	low
27	kūŋ.jīŋ	‘stye’	low
28	năŋ.sŭ:	‘book’	low
29	ná:m.ŋŭn	‘blue’	low
30	p <sup>h</sup> ô:.mót	‘wizard’	low
31	p <sup>h</sup> ĩ:.suâ	‘butterfly’	low
32	tā:.k <sup>h</sup> ă:w	‘timid’	low
33	nâ:.tâ:ŋ	‘window’	low
34	lû:k.suă	‘boy scout’	low
35	mē:w.mō:ŋ	‘scout’	low
36	p <sup>h</sup> ě:n.t <sup>h</sup> ĩ:	‘map’	low
37	wā:ŋ.kâ:m	‘swagger’	low
38	dā:w.t <sup>h</sup> iam	‘satellite’	low
39	ná:m.tâw	‘calabash’	low
40	nók.wì:t	‘whistle’	low
41	p <sup>h</sup> ā:.să:	‘language’	mono
42	kā:ŋ.kē:ŋ	‘pants’	mono
43	mâ:t.trā:	‘section’	mono
44	săm.lī:	‘cotton’	mono
45	kāw.lăw	‘a kind of soup’	mono
46	pó:j.siān	‘a kind of plant’	mono
47	t <sup>h</sup> a.nôn	‘road’	mono
48	p <sup>h</sup> â:.nîm	‘soft fabric’	phrase
49	p <sup>h</sup> ôm.jă:w	‘long hair’	phrase
50	tùk.sŭ:ŋ	‘high building’	phrase
51	bâ:n.măj	‘new house’	phrase
52	pāj.wát	‘go (to the) temple’	phrase
53	k <sup>h</sup> ă:j.ná:m	‘sell water’	phrase
54	tâ:k.suâ	‘dry clothes’	phrase

# LENGTH CONTRAST IN THE SUKHOTHAI VOWEL SYSTEM

Sireemas Maspong  
*Chulalongkorn University & Cornell University*  
*sm2627@cornell.edu*

## Abstract

This paper analyzes vowel length contrast in the Thai language of the Sukhothai period, as a descendant of Proto-Southwestern Tai (PSWT). The analysis is based on the Sukhothai graphemic system, correspondence of graphemes and PSWT phonemes and Old Khmer, the donor language of many loanwords in Sukhothai. The study reveals that the length contrast in PSWT is retained in high vowels and /a, a:/, while upper mid vowels /e, e:/, /ɤ, ɤ:/ and /o, o:/ develop length contrasts due to lexical borrowing from OK. The remaining vowels, lower mid vowels /ɛ:/ and /ɔ:/ and diphthongs, do not have length contrast.

**Keywords:** Thai language, Sukhothai inscriptions, vowel length contrast, graphemic analysis, contact induced change

**ISO 639-2 codes:** tai, tha

## 1 Introduction

The Thai language of the Sukhothai period (henceforth the Sukhothai language), as recorded in inscriptions dating back to the 13<sup>th</sup>-16<sup>th</sup> centuries, is one of the few attested medieval languages in the Southwestern branch of the Tai language family. Although its exact relationship to modern Tai varieties has never been clarified, many scholars, including Ittaratana (1975), Jansomwong (1987), Dhananjayananda (1993), consider the language to be a direct ancestor of Thai, which is currently spoken as the official national language of Thailand. Because Sukhothai inscriptions are the oldest original documents of a Tai language that survive today, studying the Sukhothai language may reveal unique information about the history of Thai and the Tai language family in general.

One puzzle of crucial relevance involves the length contrast. Even though there is a literature about the Sukhothai vowel system, there is no consensus on vowel inventories. Mostly, the studies of the vowel system agree on vowel quality distinctions, but this is not the case for vowel length. Building on my earlier work (Maspong 2016a; Maspong 2016b; Maspong & Pittayaporn to appear), this paper, therefore, studies the length contrast in the Sukhothai vowel system based on the Sukhothai inscriptions, methodologically combining graphemic analysis and Comparative Tai, following the Old Mon analysis of Shorto (1965). Moreover, this paper studies the development of length contrast in the Sukhothai language from Proto-Southwestern Tai (henceforth PSWT) and a possible process(es) involved.

The results show that vowel graphemes mostly reflect length distinction accurately, except for the case of upper mid vowels /e, e:/ and /ɤ, ɤ:/, which I argue to be underrepresented in the Sukhothai writing systems. Moreover, the results reveal that all monophthongs are contrastive for length, except lower mid vowels /ɛ:/ and /ɔ:/, for which vowel length is allophonic. The length contrast in PSWT is retained in high vowels and /a, a:/, while upper mid vowels /e, e:/, /ɤ, ɤ:/ and /o, o:/ develop length contrasts. The remaining vowels, lower mid vowels /ɛ:/ and /ɔ:/ and diphthongs, do not have length contrast. The innovated vowel length contrast is the result of lexical borrowing from Old Khmer (henceforth OK).

## 2 Previous Studies

Previous studies on Sukhothai vowel length can be mainly divided into two groups according to their methodologies and results. The first group of studies proposed that Sukhothai retained vowel length contrast for high vowels and /a, a:/ from PSWT and developed the contrast for upper mid vowels /e, e:/, /ɤ, ɤ:/ and /o, o:/. Both Brown (1965) and Rodin (1991) reached this conclusion by applying the

Comparative Method to data from Southern Thai. Their position lies on the assumption that the Sukhothai language is the direct ancestor of modern Southern Thai varieties. Nevertheless, no convincing evidence for the genetic relationship of Southern Thai dialects and the Sukhothai language is provided.

Another view on the status of vowel length in the Sukhothai vowel system is that the language lost a vowel length contrast in high vowels, but developed the contrast for /o, o:/, /ɛ, ɛ:/ and /ɔ, ɔ:/. Focusing on the writing system of Sukhothai inscriptions, Danvivathana (1981) and Jansomwong (1987) attempted to posit the sound represented by each graph. These studies suggest that the length distinction in high vowels is not phonemic in the Sukhothai language, as the symbols representing long and short high vowels appear interchangeably. Although they rightly used epigraphic data from Sukhothai inscriptions to arrive at their suggestion, the main shortcoming is that they did not consider the possibility that graphic variation may have been due to factors, such as origin of words, syllable structure, or function of words.

More seriously, when a script developed for writing one language is applied to the writing of another, such as scripts in Sukhothai inscriptions, which is developed from Old Mon (henceforth OM) and OK scripts, the resulting adaptation is likely to be neither systematic nor consistent in the matching of graphemes to phonemes (Shorto 1965:89-90). In other words, the assumption that a grapheme represents one phoneme is not tenable, since there could be cases of allography or homography. A graph could represent more than one speech sound, e.g. Mon graphs *ṣ* and *s* spellings as *kṣīw*, *pṣuk*, *kusīw*, *suk* infer one sound /s/. On the other hand, a speech sound could also be represented by more than one graph, e.g. Mon *cap* as an alternant of *cup*, *cip*, ‘to arrive as’, is a homograph of *cap* ‘to adhere to.’

Although, those using data from the Sukhothai inscriptions rightly take spelling variation as significant but fail to assess its nature systematically. Because inscriptions are the only primary source for the Sukhothai language, this paper follows the second group in taking spelling variation found in the Sukhothai inscriptions as crucial data. However, the importance of careful graphemic analysis as well as knowledge of PSWT and OK are also stressed.

### 3 Data and Methodology

The inscriptions used in this study are the inscriptions collected and transliterated in *Prajum Jārik Bhāḥ VII: Jārik Sukhothai [Collection of inscriptions, part VIII: Sukhothai inscriptions]* (2005) published by the Fine Arts Department. Of the 63 inscriptions in the volume, only those with more than one hundred legible words were used. Each of the thirty inscriptions selected were analyzed separately to allow for both potential variation due to the possibility of being different dialects and other extra-linguistic factors yet to be detected. The selected inscriptions included those written with Sukhothai Thai script and Sukhothai Khmer script, the two scripts that were used to record the Sukhothai language.

Following Shorto’s (1965) methodology, the graphemes in Sukhothai inscriptions were extracted. For each inscription, the vowel symbols were first analyzed using graphemic analysis. To discover their pronunciation, each word was first matched with a corresponding PSWT reconstruction. The correspondence between graphemes and PSWT phonemes were then set up for further analysis of the sound represented by each grapheme. With respect to the PSWT vowel system, the majority view is that PSWT has a vowel length distinction not only in \*a, \*a: but also in high vowels (Li 1977; Pittayaporn 2009; Sarawit 1973). This paper, therefore, adopted the view that PSWT has length contrast in high vowels and \*a, \*a:.

**Table 1:** Proto-Southwestern Tai vowel system (Pittayaporn 2009)

	Front		Mid		Back	
	Short	Long	Short	Long	Short	Long
<b>High</b>	i	i:	ɯ	ɯ:	u	u:
<b>Mid</b>	e	-	ɤ	-	o	-
<b>Low</b>	-	ɛ:	a	a:	-	ɔ:
<b>Diphthong</b>	iə		ɯə		uə	

Many words cannot be reconstructed in PSWT but are loanwords from Indic or Khmer. In such cases, the original forms in the donor languages were used instead. The original forms analyzed here were extracted from Jenner (2009a) for Khmer loanwords and Indic ones that are considered to have been borrowed through the Khmer language (Nacaskul 1962; Varasarin 1984).






After the correspondence of graphemes and proto-phonemes had been set up, the Sukhothai sound represented by each grapheme was extracted. Following Shorto (1965), discrepancies between the graphemic and phonological systems were also taken into account. If a grapheme corresponded with only one proto-phoneme, it was assumed to represent that particular PSWT phoneme retained in Sukhothai. Cases that did not show one-to-one correspondence were analyzed either as allography and homography of the grapheme or as evidence for a change in phonemic distinction depending on how they fit within the general picture.

#### 4 Vowel Length Contrast in the Sukhothai Language

As mention earlier, the Sukhothai writing system developed from the OM and OK systems. In OM and OK scripts, only graphemes representing high vowels and /a. a:/ show length distinction, as in Table 2.

**Table 2:** Old Khmer vocalic writing system (Jenner 2009a; Jenner 2009b)

	Front		Mid		Back	
	Short	Long	Short	Long	Short	Long
<b>High</b>	<i>	<ī> <sup>1</sup>	-		<u>	<ū>
<b>Mid</b>	<e>		-		<o>	
<b>Low</b>	-		<a>	<ā>	-	
<b>Diphthong</b>	<y>		-		<v>	

Sukhothai, however, invented new characters to represent its vowel system, including high back unrounded vowels <ī / ->, upper mid back unrounded vowel <ō / >, lower mid vowels <è / >, <a / -> and back unrounded diphthong <ōə / >. Interestingly, all innovated graphs do not show length distinctions. The Sukhothai vocalic writing system is summarized in Table 3.

<sup>1</sup> Transliteration used in this paper follows the system used by Na Nagara and Griswold in Epigraphic and Historical Studies (1967-1979) modified from Graphic System used by Coedès (1924).



**Table 3: Sukhothai vocalic writing system**

	Front		Mid		Back	
	short	long	short	long	short	long
<b>High</b>	<i / -ᵿᵿᵿᵿ >	<ī / -ᵿᵿᵿᵿ >	<ī̄ / -ᵿᵿᵿᵿ >		<u / -ᵿᵿᵿᵿ >	<ū / -ᵿᵿᵿᵿ >
<b>Mid</b>	<e / ɛ- >		<ö / ɨ-ᵿᵿᵿᵿ >		<- - >	<o / ɔ- >
<b>Low</b>	<è / ɯ- >		<a / -ᵿᵿᵿᵿ >	<ā / -ᵿ >	<a / -ə >	
<b>Diphthong</b>	<y / -ʊ >		<öa / ɨ-ᵿᵿᵿᵿ >		<v / -ɹ >	

In this section, I summarize the results from my studies on vowel length of high vowels, non-high vowels, and diphthongs in the Sukhothai language (cf. Maspong 2016a; Maspong 2016b; Maspong & Pittayaporn to appear).

#### 4.1 High vowels and (a)<sup>2</sup>

High vowels and /a, a:/ are grouped together as they are vowels that have length distinctions in PSWT. In previous works using graphemic analysis, short and long high vowels are not phonemic, while /a, a:/ remain contrastive for length in the Sukhothai language. Jansomwong (1987) used the variation of graphs representing high vowels as the main evidence against length contrast of high vowels. For example, the word ‘ride’ can be written as *khi* and *khī*, ‘night’ as *gin*, *gīn* and *gīn̄*, and ‘plant’ as *pluk* and *plūk*. However, Maspong and Pittayaporn (to appear) proposed that variations involving the high vowels are surprisingly few and explainable. The high vowels and /a, a:/ are contrastive for length in the Sukhothai language as in PSWT.

The analysis reveals that most of the variations are due to differences in the number of graphemes that exist in the graphemic system of the texts, which are variations between inscriptions. There are three system of graphemes used in Sukhothai inscriptions according to number of graphemes representing high vowels. Note that even though Sukhothai inscriptions vary with respect to the graphemic systems used, there is no obvious effect from any specific extra-linguistic factors, e.g. age, region, and content.

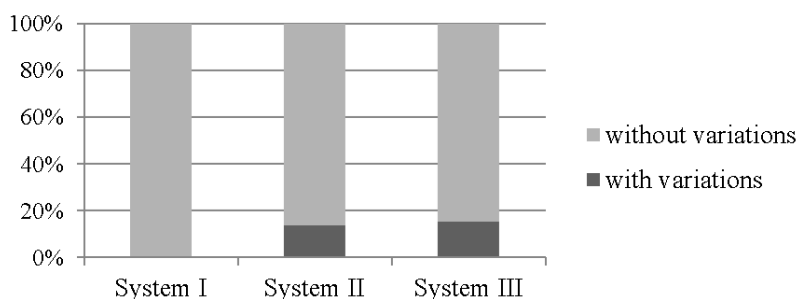
**Table 4: Groups of inscriptions having different numbers of high vowel graphemes**

	<i>	<ɨ>	<ī̄>	<u>	<ū>	Numbers of Inscriptions
<b>System I</b>	✓	-	-	✓	✓	1
<b>System II</b>	✓	✓	-	✓	✓	21
<b>System III</b>	✓	✓	✓	✓	✓	8

Variations are also found within each system. Figure 1 shows the type frequency of variant and invariant words. Counting only words that appear more than once in an inscription, the result shows that only 15 and 18 percent of words in inscriptions using System II and System III respectively appear in more than one form.

<sup>2</sup> (V) is used to represent a vowel phoneme without length indication.

**Figure 1:** Percentage of words with a grapheme representing high vowels that show variations compared to those that do not



Moreover, this type of variations is limited to specific groups of words which intrinsically allow variations: open-syllable words, loanwords, and function words (See Maspong 2016a and Maspong & Pittayaporn to appear for more details). In order to analyze length contrast in the Sukhothai language, variations need to be filtered out. I use System II as the representative of high vowel graphemic systems since it is the most frequent system. After filtering out all variations, there are four graphemes representing high vowels, namely <i>, <ī>, <u> and <ū>.

In terms of correspondence with PSWT and OK, graphemes representing high vowels behave similarly to graphemes representing (a). Short and long vowels in PSWT and OK correspond to different sets of graphemes, as illustrated in Table 5.

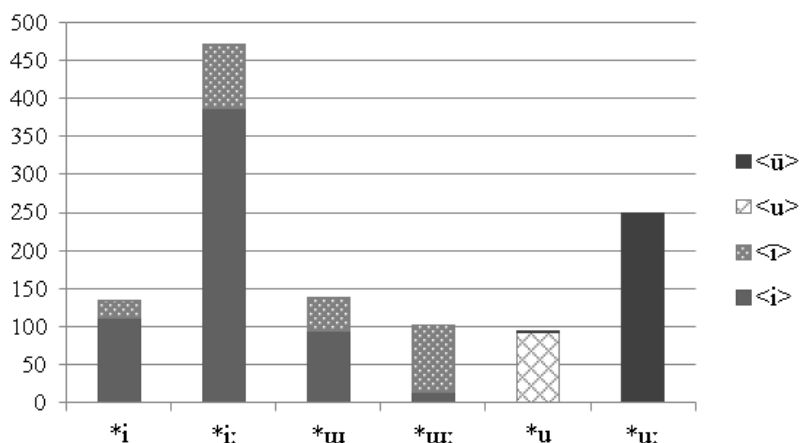
**Table 5:** Correspondence of graphemes representing high vowel and (a) with PSWT and OK

Correspondence with proto-short vowel			Correspondence with proto-long vowel		
Grapheme	Proto-phoneme		Grapheme	Proto-phoneme	
	PSWT	OK		PSWT	OK
<i>	*i, *u	/i/, /i:/	<ī>	*i:, *u:	/i/, /i:/
<u>	*u	/u/, /u:/	<ū>	*u:	/u/, /u:/
<ă>	*a	/a/, /ɔ/	<ā>	*a:	/a:/

According to Table 5, a grapheme corresponds to a PSWT phoneme in most cases, except for <i> and <ī>. Each of them corresponds to two proto-vowels. The short high unrounded vowels PSWT \*i, \*u correspond to <i>, while the long counterparts correspond to <ī>. The general pattern of correspondence still shows that the short proto-vowels correspond with one set of graphemes, <i>, <u> and <ă>, and the long proto-vowels with the other set, <ī>, <ū> and <ā>. This analysis does not apply to the case of loanwords containing high vowels in Sukhothai inscriptions.

Not every item found corresponds with PSWT as straightforwardly as mentioned earlier. Figure 2 illustrates less than 35 per cent of words with \*i, \*i:, \*u, and \*u: nucleus and approximately 2 per cent of words with \*u, \*u: nucleus in PSWT deviate from the normal cases. All of those words are considered to vary both between different inscriptions and within the same inscriptions if they appear more than once, as mentioned earlier.

**Figure 2:** Type frequency of correspondences of graphemes representing (i), (u), and (u) and PSWT phonemes



To recapitulate, the short and long proto-vowels generally correspond with two sets of graphemes with few overlap or variation. These patterns support the existence of the length distinction. If length had not been phonemic in the Sukhothai language, long and short high vowels in PSWT should not have been represented differently in Sukhothai inscriptions.

#### 4.2 Non-high vowels

Unlike high vowels and \*a, \*a:, the other non-high vowels do not have a length contrast in PSWT. Moreover, the length distinction is also absent in OK writing system (Table 2) and in the Sukhothai writing system (Table 3). Although there is no graphemic variation as high vowels, the patterns of correspondence between non-high proto-phonemes and Sukhothai graphemes are not as straightforward. Their correspondences and analyses are summarized in this section. See Maspong (2016a) and Maspong (2016b) for more details on length distinction of non-high monophthongs.

##### 4.2.1 Upper mid vowels

The upper mid vowels are represented by two types of graphemes: those showing a length contrast and those not showing a length contrast. The graphemes in the first group are the graphemes representing (o), namely <--> and <o>. They are graphically distinguished and correspond to different proto-vowels, as in Table 6.

**Table 6:** Correspondence of graphemes representing (o) with PSWT and OK

Correspondence with proto-short vowel			Correspondence with proto-long vowel		
Grapheme	Proto-phoneme		Grapheme	Proto-phoneme	
	PSWT	OK		PSWT	OK
<-->	*o	/ɔ/	<o>	-	/o:/

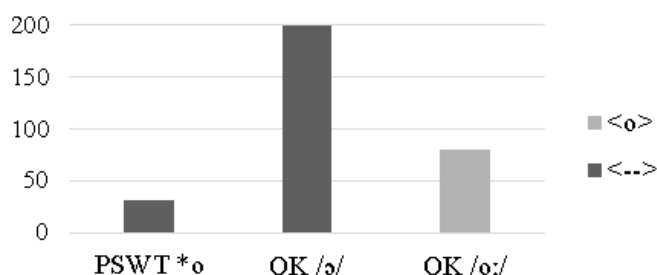
**Figure 3:** Type frequency of correspondences of graphemes representing (o) with PSWT and OK

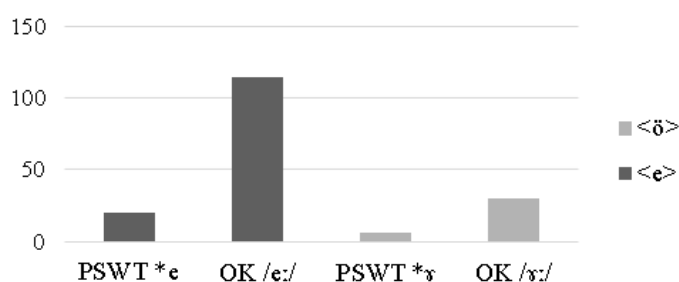
Table 6 and Figure 3 shows that the grapheme <--> corresponds with short vowels \*o in PSWT and /ɔ/ in OK, while <o> corresponds with the long vowel /o:/ in OK. As <--> only correspond to short proto-vowels, it is plausible that the grapheme represents a short vowel. I analyze it to be short /o/ and PSWT \*o and OK /ɔ/ are realized as a same vowel in the Sukhothai language.

The case of <o> is more complicated, as it only corresponds to an OK vowel. There are two possible analyses. Firstly, the foreign vowel /o:/ was imported together with OK loanwords into Sukhothai vowel system and gave rise to vowel length distinction. The alternative analysis is that it retains spelling from OK, but it is pronounced as /o/ or /ɔ:/, which are PSWT vowels, and the Sukhothai language does not have length distinction for /o/ and /o:/. If the latter analysis is true, there should be spelling variation between <o> and other graphemes representing /o/ or /ɔ:/. The variation however cannot be found. Therefore, it is more plausible that <--> represents a short vowel and <o> a long vowel and vowel length is contrastive.

The graphemes representing (e) and (ɤ) behave differently from those representing (o), as they do not show length distinction. They correspond to short vowels in PSWT and long vowels in OK. The pattern of correspondence is summarized in Table 7 and Figure 4.

**Table 7:** Correspondence of (e) and (ɤ) graphemes with PSWT and OK

Grapheme	Proto-short vowel		Proto-long vowel	
	PSWT	OK	PSWT	OK
<e>	*e	-	-	/e:/
<ö>	*ɤ	-	-	/ɤ:/

**Figure 4:** Type frequency of correspondences of graphemes representing (e) and (ɤ) with PSWT and OK

We cannot conclude if length is contrastive or not only from the correspondence pattern of upper mid vowel graphemes and proto-phonemes. The distribution of graphemes in different syllable structures is, therefore, needed. In PSWT and Tai languages in which vowel length is phonemic, length is not distinctive in open syllables. To be more specific, only long vowels occur in open syllables (Pittayaporn 2009).

Unlike PSWT phonemes, upper mid vowel graphemes are found in open syllables. If the constraint holds in Sukhothai, it can be concluded that there are long [e:] and [ɤ:]. They are all in loanwords, as PSWT does not have long \*e: and \*ɤ:. The question is whether they are phonemic or not. One group of

explanations suggests that vowel length is not contrastive. It can be either that (i) the vowel graphemes might represent only long vowels and short native vowels are merged with long vowels, or (ii) vowel length is predictable with short vowels in closed syllables and long vowels in open syllables. The alternative explanation is that vowel length is contrastive, and each grapheme represents both short and long vowels. The short vowels are in native words and long vowels are in loanwords.

I analyze the Sukhothai language to have length contrast for upper mid vowels following the general trend of Southwestern Tai languages, e.g. Thai and some dialects of Lue. They reflect short vowels for native words with short nuclei and long counterparts are from different sources, which conforms to the situation in Sukhothai. This analysis is more elegant in that the vowel system preserves symmetry of length distinction for vowels with the same height, as /o, o:/ are contrastive for length.

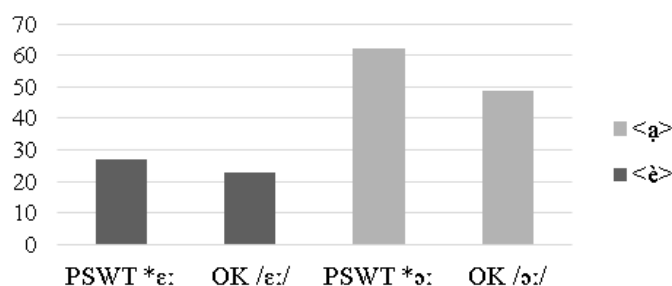
#### 4.2.2 Lower mid vowels

The graphemes representing lower mid vowels do not show vowel length contrast. They are different from <e> and <ö>, as they correspond to only long vowels in PSWT and OK, as shown in Table 9 and Figure 5.

**Table 8:** Correspondence of lower mid vowel graphemes with PSWT and OK

Grapheme	Proto-short vowel		Proto-long vowel	
	PSWT	OK	PSWT	OK
<è>	-	-	*ɛ:	/ɛ:/
<a>	-	-	*ɔ:	/ɔ:/

**Figure 5:** Type frequency of correspondences of graphemes representing (ɛ) and (ɔ) with PSWT and OK



Since they do not correspond to short vowels, it is more plausible that there is no short lower mid vowel in the Sukhothai language and vowel length is not contrastive for lower mid vowels. However, the distribution of the graphs in Sukhothai inscriptions suggests the existence of short vowel forms.

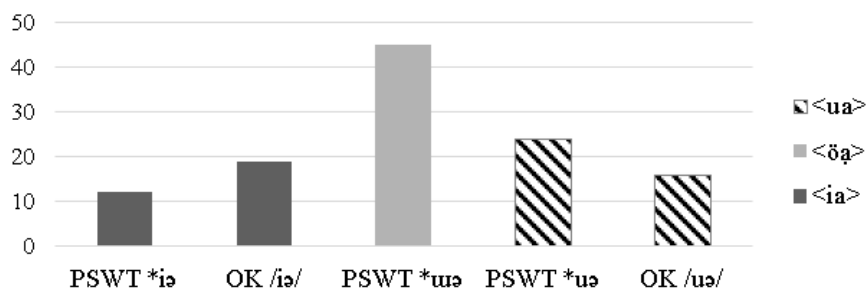
The graphemes <è> and <a> co-occur with <h> in two and five words respectively. The <h> coda is only found in non-native words with graphs representing short vowels. Previous studies used this distribution as an evidence supporting length distinction of lower mid vowel (*cf.* Danvivathana 1981; Jansomwong 1987). However, their alternations are predictable. Therefore, I analyze the short vowels to be allophonic and lower mid vowels are not contrastive for length.

#### 4.3 Diphthongs

The graphemes representing diphthongs behave similarly to the lower mid vowels. They correspond to only diphthongs, as in Table 10.

**Table 9:** Correspondence of diphthong graphemes with PSWT and OK

Grapheme	Proto-short vowel		Proto-long vowel	
	PSWT	OK	PSWT	OK
<ia>			*iə	/iə/
<öa>			*uə	-
<ua>			*uə	/uə/

**Figure 6:** Type frequency of correspondences of graphemes representing diphthongs with PSWT and OK

Diphthongs are not contrastive for length in the Sukhothai vowel system. Moreover, as they do not co-occur with <h>, they cannot be analyzed to have allophonic short forms like in the case of lower mid vowels.

In conclusion, vowel length is contrastive in the Sukhothai language. However, not all vowels in the system are contrastive for length. High, upper mid vowels, and /a - a:/ are contrastive for length, while lower mid vowels and diphthongs are not. The vowel system can be illustrated as in Table 11.

**Table 10:** Sukhothai vowel system

	Front		Mid		Back	
	short	long	short	long	short	long
<b>High</b>	i	i:	ɯ	ɯ:	u	u:
<b>Mid</b>	e	e:	ɤ	ɤ:	o	o:
<b>Low</b>	[ɛ]	ɛ:	a	a:	[ɔ]	ɔ:
<b>Diphthong</b>	iə		uə		uə	

## 5 From Proto-Southwestern Tai to the Thai language of Sukhothai and Thai

Literature on diachronic Thai mostly assumes that the Sukhothai language is the direct ancestor of Thai. However, there is no study that focuses on their genetic relationship, nor any study that provides clear support for the genealogy. In this section, I therefore compare the Sukhothai vowel system with PSWT, its direct ancestor, and Thai, its *possible* direct descendent, to prove the genetic relationship.

There are only high vowels \*i, \*i:, \*ɯ, \*ɯ:, \*u, \*u: and low vowels \*a, \*a: that are contrastive for length in PSWT. The Sukhothai language retains length contrast for those vowels from PSWT and also develops length distinction for upper mid vowels /e - e:/, /ɤ - ɤ:/, /o - o:/. The vowel system in Thai is similar to the Sukhothai system, except that length is contrastive for all monophthongs. Vowel inventories of PSWT, Sukhothai and Thai are illustrated in Table 12.

**Table 11:** Comparison of PSWT, Sukhothai and Thai vowel inventories

PSWT	Sukhothai	Thai
*i	/i/	/i/
*i:	/i:/	/i:/
*u	/u/	/u/
*u:	/u:/	/u:/
*e	/e/	/e/
-	/e:/	/e:/
*ɤ	/ɤ/	/ɤ/
-	/ɤ:/	/ɤ:/
*o	/o/	/o/
-	/o:/	/o:/
-	-	/ɛ/
*ɛ:	/ɛ:/	/ɛ:/
-	-	/ɔ/
*ɔ:	/ɔ:/	/ɔ:/
*a	/a/	/a/
*a:	/a:/	/a:/
*iə	/iə/	/iə/
*uə	/uə/	/uə/
*uə	/uə/	/uə/

The data in Table 12 show that the Sukhothai vowel system fit nicely as an intermediate stage between PSWT and Thai. Based on this assumption, we can propose that length contrast in the Thai vowel system developed in steps. Length contrast derived for upper mid vowels first in the Sukhothai language and later for lower mid vowels in Thai.

This proposal is supported by studies on the development of vowel length in Thai (*cf.* Brown 1979:14-15). They found that all monophthongs are contrastive for length, except the lower mid vowels /ɛ:/ and /ɔ:/, in a dictionary published in 1854 by Pallegoix. The vowel system in 1854 conforms with the Sukhothai vowel system. It also supports the claim that vowel length contrast of lower mid vowels developed after the Sukhothai period, based on the assumption that Sukhothai is the direct ancestor of Thai.

A more elaborate study is, however, needed to prove that the Sukhothai language is or is not the direct ancestor of Thai. This study only suggests that Sukhothai vowel system is identical to the intermediate stage between PSWT and Thai and is a possible direct ancestor of Thai.

## 6 Length Contrast as a Result of Contact-Induced Change

In this study, I found that the Sukhothai language develops length distinction for upper mid vowels /e - e:/, /ɤ - ɤ:/, /o - o:/ that is absent in PSWT. One possible explanation for the emergence of length distinction is that length contrast for upper mid vowels is a result of language contact with OK. This proposal is not entirely new, as Dhananjayananda (1993) in her study on the emergence of length distinction of mid vowels in diachronic Thai also proposed a similar idea. Note that she suggested the length distinction to emerge much later than the Sukhothai period.

Even though it is clear that the language contact plays an important role in the emergence of the length distinction, the contact situation is still puzzling. There are two competing proposals on contact situation between the Sukhothai language and OK that induced change. I would use the terminology proposed by Thomason and Kaufman (1988) to capture the two basic types of language contact or *interference*: (i) borrowing and (ii) interference through shift. The first approach, borrowing, was proposed by Li (1977) and Dhananjayananda (1993). They claimed that the vowels /e:/, /ɤ:/, /o:/ might

have been introduced into the vowel system through borrowing. It is assumed that these vowels are borrowed through loanwords into the Sukhothai language by Sukhothai speakers.

On the other hand, Diller (1988) suggested the other situation in which numbers of OK speakers coming to speak the Sukhothai language (or Tai varieties) in the fourteenth century failed to learn the target language perfectly and imported features from their source language to the target language. This situation is the case of interference through shift. Note that Diller did not propose this to capture the emergence of the length contrast, but the merger of velar consonants.

The length contrast in the Sukhothai language, as a case of contact-induced change, supports the first approach, in which the contact situation between Sukhothai and OK is borrowing. Since vowel length distinction is a structural element, if it emerged from interference through shift, we would expect the source language to have length distinction. However, upper mid vowels are not contrastive for length in OK in contrast to the Sukhothai language. It is therefore unlikely for length distinction to be directly transferred to the language through shift. The emergence of the length distinction seems to be the result of lexical borrowing instead. To be more specific, borrowing of OK loanwords into the Sukhothai language filled in the gap in the system, since long upper mid vowels are absent in PSWT. This led to the emergence of the vowel length contrast of upper mid vowels.

In conclusion, the emergence of the length distinction from PSWT in the Sukhothai language gives more insights on the contact situation of OK and the Sukhothai language or Tai varieties in general that it is borrowing instead of interference through shift that led to vowel length contrast. However, more studies are needed to reach a stronger conclusion.

## 7 Conclusion

In an attempt to study the vowel length distinction in the Thai language of the Sukhothai period, this paper analyzed the graphemic system and spelling variations found in the Sukhothai inscriptions. In addition, Sukhothai graphemes were also compared with vowel phonemes in PSWT, the direct ancestor, and OK, the donor language of loanwords. Length distinctions were analyzed from the correspondence of graphemes and proto-phonemes with additional evidence from the distribution of the vowel graphemes.

The results show that length contrast is retained in the Sukhothai vowel system. However, not all vowels in the system are contrastive for length. High, upper mid vowels, and /a - a:/ have length distinction, while lower mid vowels and diphthongs do not have length distinction. All PSWT vowels are retained in the Sukhothai language. The significant innovation is the length distinction in upper mid vowels, which is the result of borrowing. The Sukhothai language was also proposed to be a possible direct ancestor of Thai.

## References

- Brown, J. Marvin. 1965. *From ancient Thai to modern dialects*. Bangkok: Social Science Association Press.
- . 1966. The language of Sukhothai: Where did it come from? And where did it go? *Social Science Review* 3:40-42.
- . 1979. Vowel length in Thai. *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honor of Eugenie J. A. Henderson*, 10-25.
- Coedès, George. 1924. *Recueil des inscriptions du Siam*. Bangkok: Bangkok Times Press.
- Danvivathana, Nantana. 1981. *The Thai writing system*. Ph.D. dissertation: University of Edinburgh.
- Dhananjayananda, Puttachart. 1993. *The emergence of length distinction of mid and low vowels in Thai*. Ph.D. dissertation: Chulalongkorn University.
- Diller, Anthony V. N. 1988. Consonant mergers and Inscription One. *The Journal of the Siam Society* 76:46-63.
- Gedney, William J. 1947. *Indic loanwords in spoken Thai*. Ph.D. dissertation: Yale University.



- Ittaratana, Chuanpis. 1975. *Kān Chai Kham Lae Samnuen Nai Samai Sukhothai* [A study of Thai words and idiomatic expressions in the Sukhothai period]. M.A. thesis: Chulalongkorn University. (in Thai)
- Jacob, Judith M. 1960. The Structure of the Word in Old Khmer. *Bulletin of the School of Oriental and African Studies, University of London* 23:351-68.
- Jansomwong, Kruawan. 1987. *Nuey Sieng Phāsā Sukhothai Jāk Jāreuk* [The phonemes of the language used in the Sukhothai Inscriptions]. M.A. thesis: Silpakorn University. (in Thai)
- Jenner, Philip N. 2009a. *A dictionary of Angkorian Khmer*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- . 2009b. *A dictionary of pre-Angkorian Khmer*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Li, Fang Kuei. 1977. *A handbook of comparative Tai*. Honolulu: University of Hawai'i Press.
- Maspong, Sireemas. 2016a. *Vowel Length Contrast in The Thai Language of Sukhothai Period*: Chulalongkorn University.
- . 2016b. Kwām Tāng Thī Mī Naj Samkhan Dān Kwām San Yāw Khōng Nuaj Sieng Sara (e) (ε) (o) Lae (o) Naj Phāsā Thai Samaj Sukhothai [Vowel length contrast of (e) (ε) (o) and (o) in the Thai language of Sukhothai period]. *Journal of Letters* 45(2):59-92. (in Thai)
- Maspong, Sireemas & Pittayawat Pittayaporn. to appear. Length contrast of high vowels in the Thai language of the Sukhothai period: What do the inscriptions say? *Cahiers de Linguistique Asie Orientale*.
- Na Nagara, Presert & Alexander B. Griswold. 1992. *Epigraphic and Historical Studies*. Bangkok: Historical Society Under the Royal Patronage of H.R.H. Princess Maha Chakri Sirindhorn.
- Nacaskul, Karnchana. 1962. *A study of cognate words in Thai and Cambodian*. M.A. thesis: University of London.
- Pittayaporn, Pittayawat. 2009. Proto-Southwestern-Tai revised: a new reconstruction. *Journal of Southeast Asian Linguistics Society* 5:119-143.
- Prajum Jāreuk Bhāk VII: Jāreuk Sukhothai [Collection of inscriptions, part VIII: Sukhothai inscriptions]. 2005. Bangkok: Fine Arts Department. (in Thai)
- Rod-in, Chanida. 1991. *A phonological reconstruction of the Sukhothai initial consonants in 3 inscriptions of the 14th century A.D., with a comparison to Inscription 1*. M.A. thesis: Mahidol University.
- Sarawit, Mary E. 1973. *The Proto-Tai vowel system*. Ph.D. dissertation: University of Michigan.
- Shorto, Harry L. 1965. The interpretation of archaic writing system. *Lingua* 14:88-97.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Varasarin, U. 1984. *Les éléments khmers dans la formation de la langue siamoise*. Paris: Société d'Etudes Linguistiques et Anthropologiques de France.
- Wimonkasem, Kannika. 2009. *Tamrā Rian Aksorn Thai Borān: Aksorn Khōm Thai, Aksorn Dharmma Lanna, Aksorn Dharmma Isan* [Textbook of ancient Thai scripts: Thai Khmer scripts, Lanna Dharmma scripts, Isan Dharmma script]. Bangkok: Department of Eastern Language, Faculty of Archeology, Silpakorn University. (in Thai).

# SERIAL VERB CONSTRUCTIONS IN INDONESIAN: AN HPSG ANALYSIS AND ITS COMPUTATIONAL IMPLEMENTATION

David Moeljadi  
Nanyang Technological University Singapore  
D001@e.ntu.edu.sg

Viola Ow  
Nanyang Technological University Singapore  
VOW001@e.ntu.edu.sg

## Abstract

This paper discusses syntactic and semantic properties of Serial Verb Constructions (SVCs) in standard Indonesian. Analyses of Indonesian SVCs can be found in Englebretson (2003), as well as in reference grammars such as Sneddon et al. (2010). A syntactic analysis in Lexical Functional Grammar (LFG) (Kaplan and Bresnan 1982; Dalrymple 2001) was done by Arka (2000). However, no work has been done on modeling Indonesian SVCs within the Head Driven Phrase Structure Grammar (HPSG) framework (Pollard and Sag 1994) and Minimal Recursion Semantics (MRS) (Copestake et al. 2005). This paper aims to fill this gap. As for our data source, we employ the Indonesian section of the Nanyang Technological University — Multilingual Corpus (NTU-MC) (Tan and Bond 2012). We wrote a Python script to extract the Indonesian SVCs. Our HPSG analysis is implemented and tested in the Indonesian Resource Grammar (INDRA), a computational grammar for Indonesian (Moeljadi et al. 2015).

**Keywords:** serial verbs, Indonesian, HPSG, MRS, computational grammar  
**ISO 639-3 codes:** ind

## 1 Introduction

The Serial Verb Constructions (SVCs) are a type of syntactic feature with a sequence of two or more juxtaposed verbs in a single monoclausal structure in which neither is an auxiliary; they refer to a single (possibly complex) event and are not separated by any conjunctions; they share at least one semantic argument and together have a single intonation contour (Kroeger 2004, Aikhenvald and Dixon 2006). Sneddon et al. (2010:279) note that some Indonesian verbs can have a verbal clause as a complement. The main verb and the complement verb have the same subject. Alwi et al. (2014:169-170) mention that some verbs and verb phrases can function as complements (*pelengkap*) or adjuncts (*keterangan*) in verbal predicate sentences. Mintz (2002:381-383) notes that some verb sequences in Indonesian lack conjunctions or prepositions meaning “to” or “for” in between, thus appearing side by side, while in English prepositions are required.

Englebretson (2003:128-133) defines serial verbs in Indonesian as ‘at least two adjacent verbs without intervening material, which refer to closely related events, occur in the same intonation unit, and share at least one argument’. A single intonation unit refers to ‘the intonation properties of a single verb clause, and not a sequence of clauses’ (Aikhenvald 1999:470). Three criteria are established by Englebretson (2003) as a characterization of prototypical serial verb constructions in Indonesian, namely occurring contiguously, encompassing a single intonation unit, and sharing at least one argument. SVCs encompass a broad range of semantic relationships which is strictly inferential, based on verb meaning and context. There is no overt syntactic marking, conjunction, or other morpheme to indicate the semantic relation between the verbs. The meaning of SVCs is determined by both semantic compositionality and extra-lexical meaning components. Kroeger (2004) states that different languages

impose different restrictions as to which specific combinations of verbs are permissible, and that these restrictions are sometimes due to cultural factors.

It may be more proper and less misleading not to use the term SVC, but to use a more inclusive term such as ‘verb series’. However, this paper follows the term ‘serial verb constructions’ in Englebretson (2003), focuses on the surface form of the constructions, and considers constructions with juxtaposed verb predicates as SVCs. We describe the classification of Indonesian SVCs done by Englebretson (2003) in Section 2, with additional data from other previous works. We explain our work in extracting Indonesian SVCs from the Nanyang Technological University Multilingual Corpus (NTU-MC) (Tan and Bond 2012) and show some extracted data in Section 3. We give a brief explanation of HPSG and MRS in Section 4 and propose our analyses in Section 5. This paper analyzes SVCs in the standard “High” variety of Indonesian. The previous work by Englebretson (2003) uses a corpus of a colloquial “Low” variety of Indonesian.

## 2 Previous works

Analyses of Indonesian SVCs can be found in Englebretson (2003), Polinsky and Potsdam (2008), as well as in reference grammars such as Sneddon et al. (2010), as mentioned in Section 1. A syntactic analysis in LFG (Kaplan and Bresnan 1982, Dalrymple 2001) was done by Arka (2000). In this section, we summarize Englebretson (2003)’s analysis based on semantic relations and give our syntactic analysis in terms of headedness and types of verbs or transitivity. Englebretson (2003) classifies SVCs in Indonesian into two big groups, i.e. ‘serial verb as putative complements’ and ‘serial verbs with other semantic relationships’. We add one more group (i.e. ‘other semantic relationships not mentioned in Englebretson (2003)’).

1. **Serial verbs as putative complements.** This is the most common type of Indonesian SVCs. The second verbs in the SVCs (V2s) are reduced complements and the first verbs (V1s) are complement-taking predicates (CTP) which semantically can be classified as modality verbs in Englebretson (2003) or commitment verbs and orientation verbs in Arka (2000).<sup>1</sup> Arka (2000) notes that the commitment verbs are those in which the committer commits himself or herself to bring about some state of affairs. For example, *mencoba* “try”, *menolak* “refuse”, *berusaha* “attempt”, and *mulai* “begin”. The orientation verbs are typically experiencer verbs. Some examples are *mau* “want”, *ingin* “desire”, *berhak* “to have rights”, *perlu* “need”, *suka* “like (to do something)”, and *bisa/tahu* “know how to”. Arka (2000)’s commitment verbs and orientation verbs are classified as modality verbs in Englebretson (2003).<sup>2</sup> These are at the top or close to the top of the binding hierarchy defined by Givón (1980) and tend to be tightly integrated, reflecting a close conceptual bond between events. According to Givón (1980), the binding hierarchy refers to the strength of the semantic bond between two events, which contributes to the syntactic integration of the two events into a single event. Modality and manipulative verbs are at the top of this binding hierarchy, and perception-cognition-utterance verbs are lower in the hierarchy. Sentence (1) is one example. In this and all subsequent examples, serial verbs are underlined.<sup>3</sup>

---

<sup>1</sup> Englebretson (2003) also includes manipulative verbs like *suruh* “order” and *paksa* “force” as CTP which correspond to the influence verbs in Arka (2000). Arka (2000) states that they are characterized by having the influenced argument, i.e. the person asked, as controller. Sneddon et al. (2010:281) notes that the object of the main verb is simultaneously subject of the complement verb. Thus, the verbs are not juxtaposed in active voice because the object of V1 (or the subject of V2) appears between V1 and V2. We will not discuss this in this paper since we limit our discussion to SVCs with juxtaposed non-passive verbs.

<sup>2</sup> Englebretson (2003) includes *malas* “lazy” as a modality verb. On the other hand, Sneddon et al. (2010:285-286) regard this as one of a limited number of adjectives which can take an active verb as a complement. We follow Sneddon et al. (2010) and regard this as an adjective. This will not be dealt with in this paper.

<sup>3</sup> In this and all subsequent examples, the English translation uses simple present tense in default because Indonesian verbs do not inflect for tense.

- (1) *Budi*            *mencoba*        *mengejar*            *Adi.*  
 Budi                meN-try            meN-chaseAdi  
 ‘Budi tries to chase Adi.’

The construction in Example (1) is regarded as a control construction in Arka (2000) which can be defined as a relation of referential dependence between an unexpressed argument in an embedded clause (controlled argument) and an expressed or unexpressed argument (the controller) in a matrix clause. In (1), the unexpressed subject of *mencoba* “try” and *mengejar* “chase”, i.e. Budi, is said to be controlled (or shared). The head of this construction is V1 which belongs to a group called ‘control verbs’. V2 can be intransitive, e.g. *Budi mencoba tidur* “Budi tries to sleep”.

In addition, Arka (2000) mentions another complement clause called ‘raising’, in which the argument that is ‘thematically’ associated with an embedded clause is syntactically expressed as the argument of the matrix verb, where the matrix verb does not assign any thematic role to the ‘raised’ argument. The following is an example of raising. V1 is the head and belongs to ‘raising verbs’. V2 can be intransitive, for example, *Budi tampak tidur* “It appears that Budi sleeps”.

- (2) *Budi*            *tampak*            *mengejar*            *Adi*  
 Budi                appear            meN-chase            Adi  
 ‘It appears that Budi chases Adi.’

2. **Serial verbs with other semantic relationships.** Englebretson (2003) states that verb serialization also encodes four other relationships, i.e. manner, purpose, causation, and coordinated action. The relationships between the verbs must be inferred on the basis of verb semantics and context.

(a) **Manner serialization.** V2 expresses how V1 is done (V1 is the head). Example (3) shows that V1 is intransitive and V2 is transitive. V2 can be intransitive, e.g. *Budi pergi bersepeda* “Budi goes by riding a bicycle”. However, V1 cannot be transitive because the object of V1 may appear before V2, e.g. *Budi mengejar Adi menggunakan mobil* “Budi chases Adi by using a car”.

- (3) *Budi*            *berjalan*            *menggunakan*        *tongkat.*  
 Budi                walk                meN-use                stick  
 ‘Budi walks using a stick.’

(b) **Purpose serialization.** This type denotes mandatory temporal sequence. V1 enables V2 to happen, and V1 must be done first (V1 is the head). In (4), V1 is intransitive and V2 is transitive. Intransitive verbs can fill in the position of V2, e.g. *Budi pergi tidur* “Budi goes (away) to sleep”. Transitive verbs cannot fill in the position of V1 for the same reason mentioned in ‘manner serialization’ above.

- (4) *Budi*            *pulang*            *mengambil*            *uang*  
 Budi                go.home            meN-take                money  
 ‘Budi goes home to get money.’

(c) **Periphrastic causative construction.** V1 constructs an individual complex predicate with V2, as illustrated in (5). This is usually non-standard in Indonesian and does not occur in formal Indonesian, except for the lexicalized *beri tahu* “inform” (*beri* “give”, *tahu* “know”). Since this type is periphrastic (non-standard) and this paper only deals with the standard Indonesian, we will not discuss this type further. In addition, we did not find any example of this type in our data in Section 4.

- (5) *Budi*            *kasih*            *mati*                    *mereka*  
 Budi                give                die                        3PL  
 ‘Budi kills them.’

- (d) **Coordinated actions.** V1 and V2 occur rapidly and repetitively, appearing to be simultaneous, as shown in (6). The syntactic head is the V2. The order of V1 and V2 is interchangeable, e.g. *Mereka mengeroyok memukul Budi* “they gang up and beat up Budi”. In (6), both V1 and V2 are transitive verbs. V1 and V2 can be intransitive, e.g. *Budi tertawa bertepuk tangan* “Budi laughs and claps hands”, also *Budi bertepuk tangan tertawa* “Budi claps hands and laughs” (*bertepuk tangan* “clap hands” is an intransitive verb in Indonesian).

(6) *Mereka memukul mengeroyok Budi*  
 3PL meN-beat meN-gang.up Budi  
 ‘They beat up and gang up Budi.’

3. **Other semantic relationships not mentioned in Englebretson (2003).** In addition to the above semantic relationships, Alwi et al. (2014:170) notes Example (7), which has an ‘origin’ or ‘source’ meaning.

(7) *Budi pulang bertamasya*  
 Budi go.home have.a.picnic  
 ‘Budi goes home from picnic.’

We think this is ambiguous. Depending on the context, (7) can be interpreted as ‘purpose serialization’ and thus the meaning is “Budi goes home to have a picnic”. Regarding the headedness and transitivity of V1 and V2, the behavior is the same as the one in ‘purpose serialization’.

There is also another construction which has a ‘resultative’ meaning. Kridalaksana (1989:139) mentions that this type is a combination of an action (predicated by V1, the head) and a result (predicated by V2) which is caused by that action. In this construction, V1 is transitive and V2 is intransitive, as shown in (8), which consists of an action *Budi membunuh Adi* “Budi kills Adi” and a result *Adi mati* “Adi dies”.

(8) *Budi membunuh mati Adi.*  
 Budi meN-kill die Adi  
 ‘Budi kills Adi until Adi dies.’

Englebretson (2003) mentions that SVCs in Indonesian may appear with connectors between the verbs, e.g. *untuk* “in order to” which may appear after a manipulative or modality verb in the first type above. However, he notes that while *untuk* potentially indicates some sort of relationship between a series of verbs, how to characterize the relationship is not clear and he found only three examples of *untuk*, compared with 517 SVCs in his corpus with no connective. Englebretson (2003) also states that there is apparent grammaticization of some of the Indonesian SVCs into auxiliaries marking aspect and modality, e.g. *mau* “want” as a modality verb in SVC type one and as a temporal auxiliary marking future. Polinsky and Potsdam (2008) analyze *mau* followed by a passive predicate which generates an unexpected, ambiguous interpretation. Sneddon et al. (2010:280-281) also notes that *mau* with a passive complement raises ambiguity as to whether it is the subject or the agent of the complement who wants the action to occur. This is an interesting phenomenon. However, in this paper we limit our discussion to non-passive predicates only.

### 3 Indonesian data

We wrote a Python script to extract SVCs from a Sherlock Holmes short story and a Japanese short story in the Indonesian database in the Nanyang Technological University Multilingual Corpus (NTUMC) (Tan and Bond 2012), a parallel English-Chinese-Japanese-Indonesian corpus containing 2,975 Indonesian sentences from three sources: Singapore Tourism Board website ([www.yoursingapore.com](http://www.yoursingapore.com)), a Sherlock Holmes short story “The Adventure of the Speckled Band”, and a Japanese short story written by Akutagawa Ryunosuke: “The Spiders Thread”. The Sherlock Holmes short story and the Japanese short story are originally written in English and Japanese, respectively, and

translated into standard Indonesian. Both the original texts and the translations are part-of-speech (POS) tagged.

The NTU-MC data is organized according to sentence IDs (SID), word IDs (WID) for V1, WID for V2, V1, V2, parts-of-speech (POS), and sentence. After extracting all possible Indonesian SVCs, we used suggested PARSEME annotation guidelines to determine if an extracted SVC is an SVC. After that, SVCs following the patterns we observed were automatically tagged and the remaining complex verbs were manually tagged. Automatic tags were then hand checked for errors claims.

### 3.1 Extracting SVCs

The PARSEME annotation guidelines are state-of-the-art for annotation of verbal multi-word expressions (Candito et al. 2016). They are written with the assumption that a person, not a computer, is doing the annotation. The annotators can use the guidelines to write scripts to automate the annotation. Since PARSEME does not have specific guidelines for Indonesian, we modified the guidelines for English and wrote suggested PARSEME guidelines for the extraction and identification of Indonesian SVCs based on the findings in Section 3. The guidelines describe a series of tests to identify and classify Indonesian SVCs (see Table 1).

1. If V1 is a control verb like *ingin* “desire”, *mau* “want”, *mencoba* “try”, *tahu* “know”, and *mampu* “be able to”, and the meaning of V1+V2 is compositional, V1+V2 is a control SVC.
2. If V1 is a raising verb like *terlihat* “seems”, *terasa* “feels like”, and *tampak* “looks”, and the meaning of V1+V2 is compositional, V1+V2 is a raising SVC.
3. If V2 expresses the manner of V1, and V1 is the head of the SVC, V1+V2 is a manner SVC.
4. If V2 indicates the purpose of V1, V1 is the head of the SVC, and V1 and V2 are bound by a temporal sequence (V1 happens before V2), V1+V2 is a purpose SVC.
5. If V1 and V2 occur rapidly and seem to be simultaneous actions, and the meaning of V1+V2 is compositional, V1+V2 is a coordinated action SVC.
6. If V2 happens before V1 and V1 is the head of the SVC, V1+V2 is a source SVC.
7. If V1 indicates an action, V2 is the result of V1 (V1 happens before V2), and V1 is the head of the SVC, V1+V2 is a resultative SVC.
8. If *untuk* “for/to” can be inserted between V1 and V2, V1+V2 is a control or purpose SVC, where control SVCs are not temporally constrained, but purpose SVCs require the V1 to occur before V2.
9. If a pronoun like *dia* “s/he” can be inserted between V1 and V2, as in Example (9), V1+V2 is NOT an SVC because it does not fulfill two of the three criteria for prototypical SVCs, i.e. occurring contiguously and encompassing a single intonation unit, as mentioned in Section 1.

(9) *Dia*        *mengatakan*        (*dia*)        *mau*        *ke*        *kota...*  
 3SG        meN-say        (3SG)        want        to        city  
 ‘He says he wants to go to the city...’ (SID:10255)

10. If the complementiser *bahwa* “that” or *apakah* “whether” can be inserted between V1 and V2, where V1 is a saying verb like *mengatakan* “say” or an asking verb like *menanyakan* “ask”, V1+V2 is NOT an SVC, for the same reason mentioned above. (10) is an example.

(10) *Dia*        *mengatakan*        (*bahwa*)        (*dia*)        *mau*        *ke*        *kota...*  
 3SG        meN-say        (that)        (3SG)        want        to        city  
 ‘He says that he wants to go to the city...’ (SID:10255)

All possible Indonesian SVCs are two words tagged as “verb” appearing beside each other in the corpus. Following the suggested PARSEME guidelines for Indonesian SVCs, we wrote a script to semi-automatically assign the tags. First, we checked for each sentence in the corpus whether the POS tag V and V occur side-by-side, afterwards we checked whether V1 is a raising or a control verb. If V1 is neither a raising nor a control verb, we checked and assigned the tags manually. Instances where V1s or V2s turn out to be segmentation errors, nouns, adverbs, or prepositions that were wrongly tagged in

the corpus were then marked and tagged as “NOT VV”. Semi-automatically tagged SVCs were then checked for errors.

**Table 1: Indonesian SVCs and the corresponding tags**

Type	Tag	V1	V2
Control SVC	control	control verb	(in)transitive
Raising SVC	raising	raising verb	(in)transitive
Manner SVC	manner	intransitive	(in)transitive
Purpose SVC	purpose	intransitive	(in)transitive
Coord.action SVC	coord.action	(in)transitive	(in)transitive
Source SVC	source	intransitive	(in)transitive
Resultative SVC	resultative	transitive	intransitive

**Table 2: Distribution of Indonesian SVCs in the corpus**

Type	Number	Percentage
control	21	72.4%
raising	3	10.3%
manner	2	6.9%
purpose	2	6.9%
coord.action	1	3.4%
source	0	0.0%
resultative	0	0.0%
Total	29	100.0%

### 3.2 Result

Out of 45 candidates which have the POS tags V-V side-by-side and can be automatically extracted, only 29 are SVCs. Out of the 29, 21 are control SVCs, three are manner SVCs, two are purpose SVCs, two are raising SVCs, and one is a coordinated action SVC. Regarding the remaining 16 candidates, 13 of them have incorrect POS tags (e.g. the word *masalah* “problem” was tagged V, but it should have been tagged N), two of them are incorrectly segmented (have segmentation errors), and the rest (one candidate) is the one in (10) where its V1 is a *saying* verb. Table 2 shows the distribution of Indonesian SVCs from the extracted data. All the extracted SVCs in the data comprise two juxtaposed verbs. SVCs having more than two juxtaposed verbs were not found in the corpus. A test-suite (a sample of text illustrating a particular language phenomenon or construction, formatted in interlinearized glossed text according to Leipzig glossing rules) containing the 29 SVC sentences was made.<sup>4</sup> The sentences were slightly edited to accommodate INDRA, focusing on the SVCs only.<sup>5</sup>

For control SVCs, V1s are the head of the SVCs, and are control verbs like *jadi* “manage (to)”, *ingin* “wish”, and *bermaksud* “intend”. The V2s attached to these V1s are complements, and the meaning of V1+V2 is compositional for all control SVCs in the data, as in (11).

- (11) ...*Holmes*    ...*mencoba*    *membuka*    *palang*    *itu*...  
 Holmes    meN-try    meN-open    shutter    that  
 ‘...Holmes ...tries to open that shutter...’ (SID:10417)

<sup>4</sup> The test-suite can be accessed in INDRA repository in GitHub: <https://github.com/davidmoeljadi/INDRA/blob/master/testsuites/SVC.txt>

<sup>5</sup> The computational grammar for Indonesian (INDRA) is still being developed and it does not cover phenomena such as subordination at present. The 29 SVC sentences were edited so that they do not contain subordinate clauses.

For Indonesian manner SVCs, V2s indicate the manner and direction of V1. For example, in *pulang* “return home” + *melalui* “pass through”, the act of returning home is done by passing through a place, as illustrated in Example (12). The V2 *melalui* “pass through” is a verb in (12) but can also function as a preposition meaning “through”.<sup>6</sup> Payne (2008:312) notes that serial verbs can also be a source for adpositions.

- (12) ...*saya*      *pulang*      *melalui*      *halaman...*      *itu...*  
 1SG      go.home      pass.through      yard      that  
 ‘...I go home by passing through that... yard.’ (SID:10500)

For the raising SVC category, all V1s are raising verbs and are the heads of each SVC, and V2s are complements. The meaning of V1+V2 is compositional, as in (13).

- (13) *Waktu*      *terasa*      *berlalu*      *dengan*      *lambat*      *sekali.*  
 time      feel      pass      with      slow      very  
 ‘It seems that time passes very slowly.’ (SID:10585)

For purpose SVCs, V1 is the head of the SVC and V2 indicates the reason for doing V1, and V1 has to happen before V2. The meaning of every V1+V2 is compositional. For example, in *bersiap* “prepare” + *pergi* “go”, the act of preparing was done for the purpose of going somewhere, as shown in (14). The other purpose SVC we found is *menggapai-gapai* “reach out” + *mencari* “search”. However, one can argue that *bersiap pergi* “prepare to go” is a control SVC and *menggapai-gapai mencari* “reach out and search” is a coordinated action SVC. Some SVCs have ambiguous semantic relations (see also (7)).

- (14) ...*Holmes*      ...*bersiap*      *pergi.*  
 Holmes      prepare      go  
 ‘...Holmes ...is prepared to go.’ (SID:10587)

The only coordinated action SVC in the extracted data is *berlari* “run” + *menuju* “head towards”, meaning ‘run and go towards’. V1 and V2 happen rapidly and repetitively to describe the seemingly simultaneous action of running and going towards somewhere.

- (15) *Saya*      *segera*      *berlari*      *menuju*      *kamar*      *ayah*      *tiri*      *kami...*  
 1SG      soon      run      head.toward      room      father      step-      1PL.EXCL  
 ‘I soon run towards our stepfather’s room...’ (SID:10193)

#### 4 HPSG and MRS

We use the theoretical framework of HPSG (Pollard and Sag 1994). HPSG is monostratal, handling orthography, syntax (SYN), and semantics (SEM) in a single structure (sign), modeled through typed feature structures. Signs in HPSG include words, phrases, sentences and utterances (Sag et al. 2003). Types are classes of linguistic entities. Each type is associated with a particular feature structure. Feature structures are sets of features or attributes and value pairs which represent objects. Features or attributes are unanalyzable atomic symbols from some finite set and values are either atomic symbols or feature structures themselves. Feature structures are usually illustrated with an Attribute-Value Matrix (AVM). HPSG is unification- and constraint-based. The words and phrases are combined according to constraints of the lexical entries based on the type hierarchy. We use MRS (Copestake et al. 2005) as the semantic framework because it is adaptable for HPSG typed-feature structure and suitable for parsing and generation. The semantic structures in MRS are underspecified for scope and thus suitable for representing ambiguous scoping.

<sup>6</sup> If *melalui* can be passivized, it is a verb, not a preposition. The V2 *melalui* in Example (12) is a verb because we can change the order of the subject *saya* “1SG” and the object *halaman itu* “that yard” and change the verb form to passive: *halaman itu saya lalui* “that yard is passed through by me”.



## 5 Analysis and computational implementation

### 5.1 Serial verbs as putative complements

This group contains control and raising constructions, following Arka (2000). Control verbs assign a semantic role to their subject, while raising verbs do not. Sag et al. (2003:376) also state that subject control verbs express a relation between an individual and a situation, while subject raising verbs express properties of situations.

**Control SVC.** The head daughter of this construction is V1. V1’s predications involve two things: the first one is the subject (SUBJ) which is the shared argument (a noun) of V1 and V2 and the second one is some situation predicated by V2 (COMPS). V2 is a semantic argument of V1’s predication. Its index is linked to both V1 and V2. V2’s semantic (SEM) external argument (XARG) is identified with its subject’s semantic index. That index is identified with the first argument (ARG1) in the lexical relation (KEYREL) introduced by the verb. This is illustrated in 16. See also Figure 1 on the next page for the Dependency Minimal Recursion Semantics (DMRS) representation of the control construction. The head-complement rule unifies V1’s constraints on its complement with those of V2 which results in the identification of the XARG value of V2 with the index of the subject. The head-complement rule also propagates up the constraints on the subject from the head daughter *mencoba* “tried” to the verb phrase *mencoba membuka palang itu* “tried to open that shutter”. The semantics of this verb phrase preserve the semantic properties of its daughters, including the desired reentrancies with the subject index. The subject-head rule combines *Holmes* with *mencoba membuka palang itu* and identifies the ARG0 value of *Holmes* with the ARG1 values in both V1 and V2. This is illustrated in Figure 1.

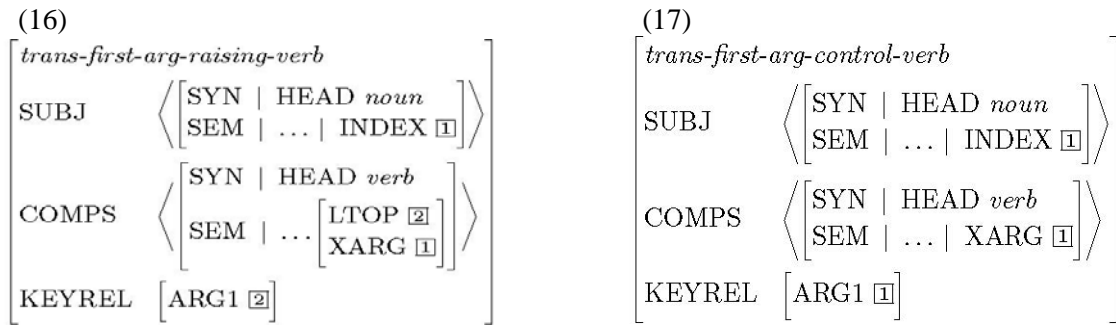


Figure 1: Parse tree and DMRS representation of a control SVC

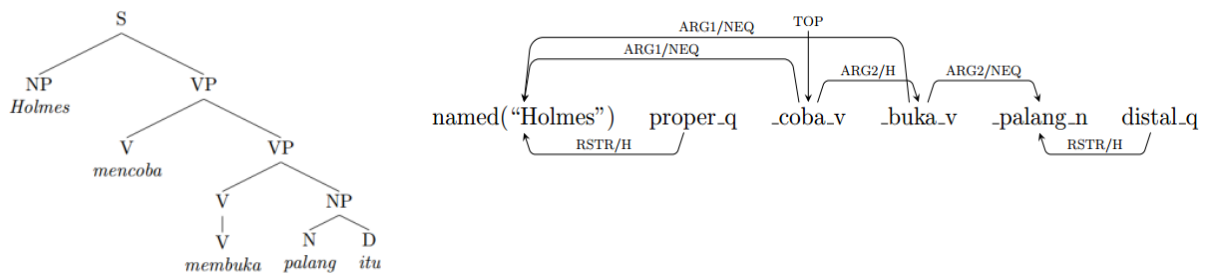
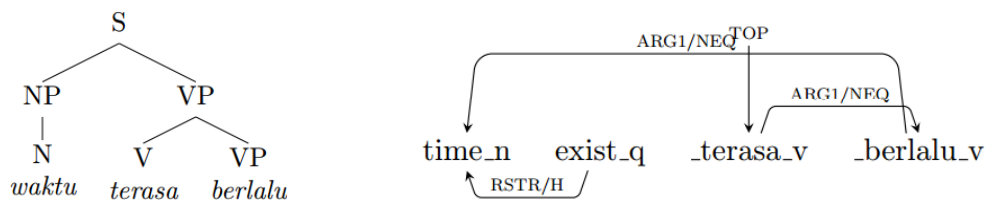


Figure 2: Parse tree and DMRS representation of a raising SVC

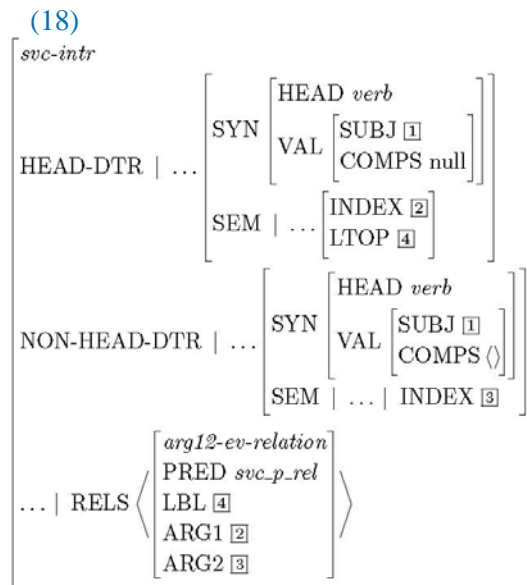


**Raising SVC.** The raising phenomenon is where a syntactic argument’s semantic index is not linked to any semantic argument position in a given lexical entry’s semantic relation, but is instead assigned to a role in the semantics of another syntactic argument of the lexical entry. In a raising construction, the head daughter (V1) can have two syntactic arguments, an NP (as SUBJ) and a VP (as COMPS), as in (13). V1 does not do anything semantically with its subject. Sag et al. (2003:366) use ‘active-passive pairs’ which have essentially the same meaning. V1 identifies the external argument (XARG) of V2 with the subject’s index. However, it takes the proposition introduced by V2 as its only semantic argument, rather than assigning a second semantic role to the subject’s index. The local top (LTOP) is identified with the first argument (ARG1) in the lexical relation (KEYREL) introduced by the verb. This is illustrated in (17). The head-complement rule unifies V1’s constraints on its complement with those of V2 and identifies the subject’s index with the XARG value of V2. The subject-head rule combines the subject *waktu* “time” with the VP *terasa berlalu* “seems passing” and identifies the ARG0 value of the subject with the ARG1 in V2.

### 5.2 Serial verbs with other semantic relationships

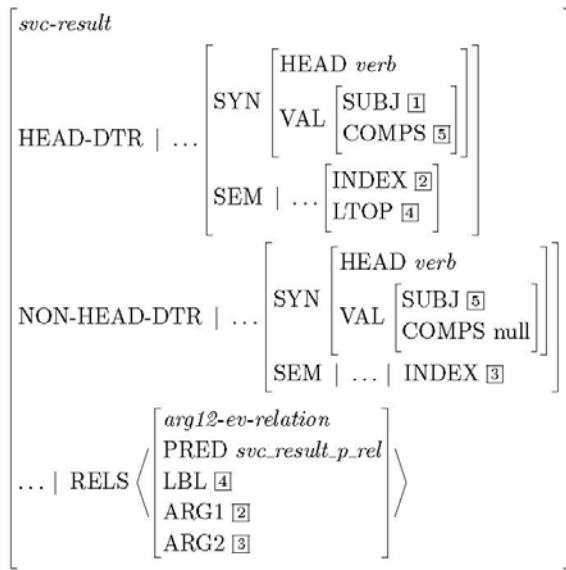
The result in Section 4.2 shows us that V1 in the manner, purpose, and coordinated action SVCs can be any verbs other than control and raising verbs and there are ambiguities in the semantic relationships. Because of this, we made rules based on the transitivity and we introduce semantic relations *svc\_p\_rel*, *svc\_coord\_p\_rel*, and *svc\_result\_p\_rel*.

**V1 is intransitive.** In an SVC where the V1 is intransitive, V1 (the head daughter, HEAD-DTR) and V2 (NON-HEAD-DTR) share the same subject (SUBJ) but not the same object. The V1 is assumed not having an object and thus in the valence (VAL), the COMPS has the value *null*. The object, if there is any, is the argument of V2 (the COMPS is not *null*). Here we introduce a new relation (RELS) *svc\_p\_rel* having two arguments. Its first argument (ARG1) is identified with the index of the head daughter (V1), while its second argument (ARG2) is identified with the index of V2. The label (LBL) is identified with the local top (LTOP) of V1. This is illustrated in (18). The head-complement rule unifies V2 with its object. The SVC rule unifies V1 and V2 (with its object argument) and propagates up the constraints on the subject from the head daughter *pulang* “return home” to the verb phrase *pulang melalui halaman itu* “go home by passing through that yard”. The semantics of this verb phrase retain the semantic properties of its daughters and the *svc\_p\_rel* predicate relation. The subject-head rule combines *saya* “I” with *pulang melalui halaman itu* and identifies the ARG0 of the subject with the ARG1 of V1 and V2. This is illustrated in Figure 3.

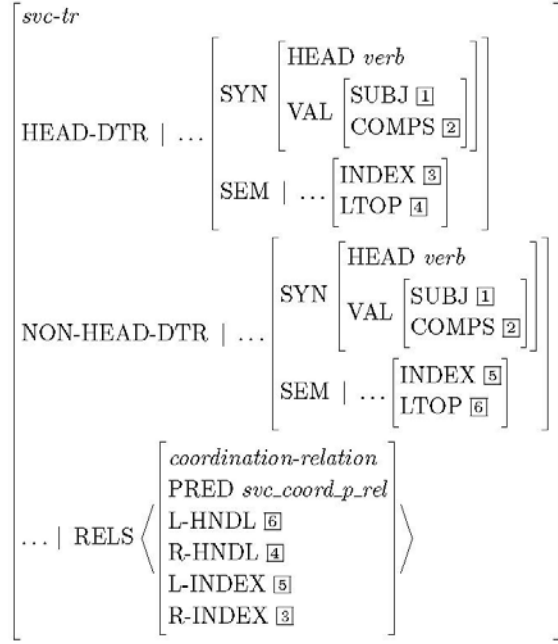


**V1 and V2 are transitive with a shared object.** In this SVC, both V1 and V2 share the same subject (SUBJ) and the same object (COMPS). V2 is the syntactic head (HEAD-DTR), but semantically the *svc\_coord\_p\_rel* is the head. The left index (L-INDEX) and the right index (R-INDEX) of *svc\_coord\_p\_rel* are identified with the index of V1 and V2, treating the semantics the same as “V1 and V2”. This is illustrated in (19). The SVC rule unifies V1 *memukul* “beat up” and V2 *mengeroyok* “gang up” and propagates up the constraints on the object *Budi* and the subject *mereka* “they”. The head-complement rule unifies V1+V2 with their shared object and identifies the ARG0 of the object with the ARG2 of V1 and V2. The subject-head rule combines the shared subject with the VP and identifies the ARG0 of the subject with the ARG1 of V1 and V2 (see Figure 4).

(19)

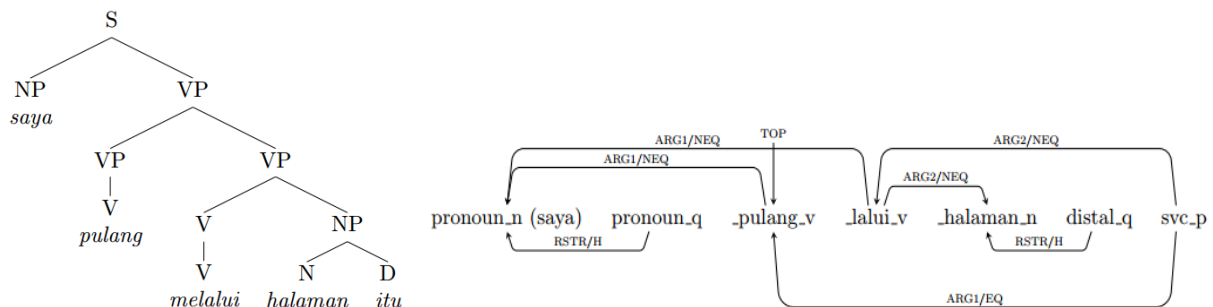


(20)

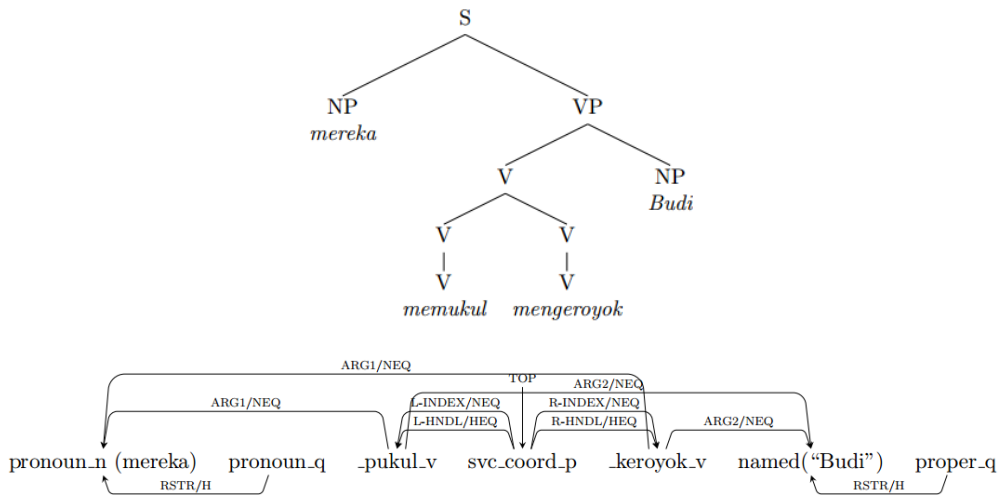


**V1 is transitive and V2 is intransitive, having a resultative meaning.** This SVC is a particular construction with regard to the transitivity of V1 and V2 and the semantic relationship. V1 is a transitive verb its subject (SUBJ) and object (COMPS), being the same as the subject and the object of the SVC respectively. V2 is an intransitive verb, its subject (SUBJ), corresponding to the object of the SVC. V2 does not have an object and thus the value of its COMPS is *null*. In terms of semantic relation (RELS), it is resultative and thus we posit a new relation *svc\_result\_p\_rel*. Its ARG1 and ARG2 are identified with the indices of V1 and V2 respectively. Its label (LBL) is identified with the local top (LTOP) of V1. This is illustrated in (20). The SVC rule unifies V1 *membunuh* “kill” and V2 *mati* “die”. The head-complement rule unifies V1 with its object (also V2 with its subject argument) and identifies the ARG0 of the SVC’s object with the ARG1 of V2 and the ARG2 of V1. The subject-head rule combines the SVC’s subject with the VP and identifies the ARG0 of the subject with the ARG1 of V1 (see Figure 5 on the next page).

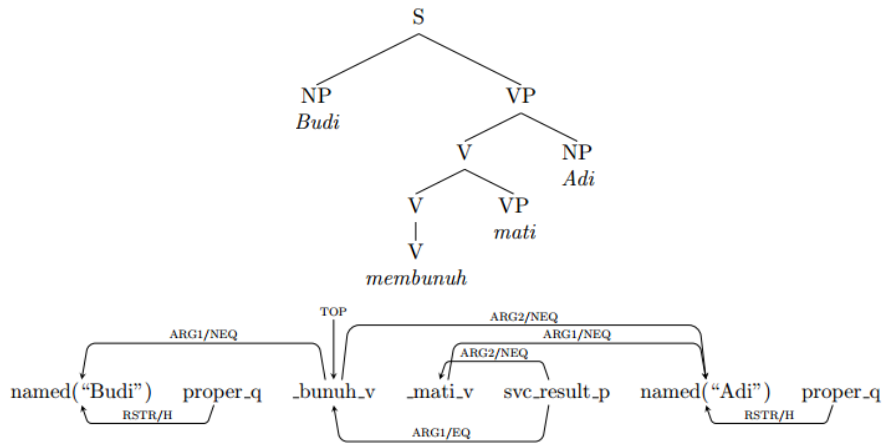
Figure 3: Parse tree and DMRS representation of Example (12)



**Figure 4:** Parse tree and DMRS representation of Example (6)



**Figure 5:** Parse tree and DMRS representation of Example (8)



## 6 Evaluation and conclusion

Our analyses of Indonesian SVCs depart from Englebretson (2003) with references from Arka (2000) and other reference grammars. We wrote a Python script to extract SVCs from a Sherlock Holmes short story and a Japanese short story, both translated into standard Indonesian, in NTU-MC (Tan and Bond 2012). A test-suite containing a representative set of sentences designed to show SVC phenomena for Indonesian was created based on the 29 sentences extracted from NTU-MC (all positive/grammatical test items). We analyzed and modeled Indonesian SVCs using HPSG and MRS and implemented our analyses in INDRA (Moeljadi et al. 2015). We posit lexical types for control and raising verbs. For verbs with other semantic relationships, we made rules based on the transitivity of V1 and V2 and the shared arguments, as well as the semantic relationships. After the lexical acquisition and the addition of lexical types and SVC rules, we did the evaluation by checking the coverage of parsed sentences in the test-suite using gTest, a DELPH-IN Grammar Testing Tool (<https://github.com/goodmami/gtest>). The result was 29 out of 29 sentences in the test-suite can be parsed (overall coverage 100%). There may be a closed class of verbs that occurs as V1 or V2 in each semantic relationship and thus, in the future, we will get more data from different corpora and do more work on verb subcategorization, and we will analyze which verbs in particular SVCs are likely to appear in which semantic relations (manner, purpose, coordinated action, source, and resultative).

## 7 Acknowledgements

Thanks to Sanghoun Song for his help in implementing the control and raising SVCs in the early stage. Thanks to Francis Bond, František Kratochvíl, and Hiroki Nomoto for the precious comments and advice.

## References

- Aikhenvald, Alexandra Y. and R. M. W. Dixon. 2006. *Serial verb constructions: a cross-linguistic typology*. Oxford University Press.
- Aikhenvald, Alexandra Y. 1999. Serial constructions and verb compounding: evidence from Tariana (North Arawak). *Studies in Language* 23(3):469-498.
- Alwi, Hasan, Soenjono Dardjowidjojo, Hans Lapoliwa, and Anton M. Moeliono. 2014. *Tata bahasa baku bahasa Indonesia [The standard grammar of Indonesian]*. Jakarta: Balai Pustaka.
- Arka, I Wayan. 2000. Control and argument structure: explaining control into subject in Indonesian. *Papers presented at Fourth International Symposium on Malay/Indonesian Linguistics, Jakarta*.
- Candito, Marie, Fabienne Cap, Silvio Cordeiro, Vassiliki Foufi, Polona Gantar, Voula Giouli, Carlos Herrero, Mihaela Ionescu, Verginica Mititelu, Johanna Monti, Joakim Nivre, Mihaela Onofrei, Carla Parra Escartn, Manfred Sailer, Carlos Ramisch, Monica-Mihaela Rizea, Agata Savary, Ivelina Stonayova, Sara Stymne, and Veronika Vincze. 2016. *PARSEME shared task - annotation guidelines*. [accessed on January 11, 2017] (ver 1.6b).
- Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: an introduction. *Research on Language and Computation* 3(4):281-332.
- Dalrymple, Mary. 2001. Lexical-Functional Grammar. *Syntax and Semantics*, 34. Academic Press.
- Englebretson, Robert. 2003. *Searching for structure: the problem of complementation in colloquial Indonesian conversation*. John Benjamins Publishing.
- Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4(3):33-77.
- Kaplan, Ronald and Joan Bresnan. 1982. Lexical Functional Grammar: a formal system for grammatical representation. In *The mental representation of grammatical relations*, 173-281. Cambridge: the MIT Press.
- Kridalaksana, Harimurti. 1989. *Pembentukan kata dalam bahasa Indonesia [Word formation in Indonesian]*. Jakarta: Gramedia.
- Kroeger, Paul. 2004. *Analyzing syntax: a lexical-functional approach*. Cambridge University Press.
- Mintz, Malcolm W. 2002. *An Indonesian and Malay grammar for students*. 2nd edition. Perth: Indonesian / Malay Texts and Resources.
- Moeljadi, David, Francis Bond, and Sanghoun Song. 2015. Building an HPSG-based Indonesian Resource Grammar (INDRA). In *Proceedings of the GEAF Workshop, ACL 2015*, 9–16.
- Polinsky, Maria and Eric Potsdam. 2008. The syntax and semantics of wanting in Indonesian. *Lingua* 118(10):1617-1639.
- Pollard, Carl and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press.
- Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic theory: a formal introduction*. 2nd edition. Stanford: CSLI Publications.
- Sneddon, James Neil, Alexander Adelaar, Dwi Noverini Djenar, and Michael C. Ewing. 2010. *Indonesian reference grammar*. 2nd edition. New South Wales: Allen & Unwin.
- Tan, Liling and Francis Bond. 2012. Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). *International Journal of Asian Language Processing* 22(4):161-174.

# INTERNALLY HEADED RELATIVE CLAUSES IN SHAN

Mary Moroney  
Cornell University  
mrm366@cornell.edu

## Abstract

This paper demonstrates that internally headed relative clauses (IHRCs) are found in Yǔn Shan, a variety of Shan, a Southwestern Tai language. I discuss the Shan nominal structure and propose to give a syntactic analysis of Shan relative clauses that uses head raising to account for both its post-head and internally headed relative clauses.

**Keywords:** relative clauses, internally headed relative clauses, head raising, Shan, Tai  
**ISO 639-3 codes:** shn, tha

## 1 Introduction

Shan, a Southwestern Tai language, has externally headed, post-head relative clauses, as are found in many languages, including English and Thai, another Southwestern Tai language. In this sort of relative clause, shown in (1),<sup>1,2,3</sup> the head *lik* ('book') appears before the clause that modifies it. This is the same as the structure that appears in Thai, shown in (2), where the head *naŋsǎ̌* ('book') again precedes the clause.

- (1) [*Lik ʔǎn háw hǎn nâj*] *mán lěŋ*  
book COMP 1 see this 3 red  
'The book that I am seeing is red.' (Shan; 2016.05.10.MM.P.07b)
- (2) [*Naŋsǎ̌ thîi raw hě̌n nîi*] *mán dɛɛŋ.*  
book COMP 1.PL see this 3.SG red  
'The book that I see is red.' (Thai; 2016.09.28.06)

In addition to the post-head relative clauses, examples like (3) are also found in the language. This construction is called an internally headed relative clause (henceforth, IHRC). The Thai equivalent of the Shan IHRC in (3) is ungrammatical, as is demonstrated in (4). The defining characteristic of the internally headed relative clause is that the head appears inside the clause itself, often in the position it would be found in an independent, non-relativized clause. Shan is a language with SVO word order, so the head *lik* ('book') appears after the verb in (3).

- (3) [*ʔǎn háw hǎn lik ʔǎn nâj*] *mán lěŋ.*  
COMP 1 see book AN this 3.SG red  
'This book that I see is red.' (Shan; 2016.05.10.P.MM.07a)

---

<sup>1</sup> Thanks to Aye Twei Soe, Sireemas Maspong, and Ngampit Jagacinski who provided the Shan and Thai data. Thanks also to those who have given me comments on this paper, especially John Whitman. All errors are mine.

<sup>2</sup> A note on the glosses: 1: first person; 3: third person; AN: pronominal use of *ʔǎn*; CLF: classifier; COMP: complementizer; IMPF: imperfect; NEG: negation; PL: plural; PRF: perfect; SG: singular; TOP: topic marker

<sup>3</sup> The data on Shan comes primarily from my fieldwork on Shan working with a Shan speaker in Ithaca, NY from January 2016 to present. My Shan consultant is from a village near Papun in Kayin State, Myanmar and speaks the Yǔn Shan dialect — which is very different from the Taunggyi dialect. She also speaks Karen and received her education in Burmese. She has been in the United States for 5 years and speaks English as well. Data was collected using a variety of elicitation methods: direct translation, grammaticality judgments, telling short stories, felicity judgments on grammatical sentences in specific contexts.

- (4) \**[ʔan/thîi raw hěn najsěw ʔan níi] mán dεεŋ.*  
 COMP 1.PL see book AN this 3.SG red  
*Intended:* ‘This book that I see is red.’ (Thai; 2016.09.28.02-03)

It is somewhat surprising that internally headed relative clauses are used in addition to post-nominal externally headed relative clauses because IHRCs have not been reported in Thai or Lao, closely related languages (Warotamasikkhadit 1972; Morev & Moskalev 1979; Prasithratsint & Yaowapat 2009). In fact, they have not been reported in any Tai languages that I have found.

This finding is also surprising because VO languages are less likely to have internally headed relative clauses than OV languages. At one point in time it was assumed that a head-initial, VO language like Shan could not have internally headed relative clauses because only OV languages were thought to allow IHRCs (Cole 1987). This assumption has been disproven in verb initial languages like Seediq and Tagalog (Aldridge 2004) and SVO languages like Buli (Hiraiwa 2003) and American Sign Language (Liddell 1980). According to the World Atlas of Language Structures (WALS), 58 OV languages have IHRCs and only 5 VO languages do (Dryer & Haspelmath 2013); however, none of the cases of SVO languages with IHRCs are identified in WALS, so this count may not represent the real distribution. Shan’s use of IHRCs appears to be more evidence that VO languages can have this construction, but this leaves open the question of why head initial languages are less likely to have internally headed relative clauses. This is a question I hope to address in future research.

The goal of this paper is to identify and examine the structure of the Shan relative clause and give a syntactic analysis of both the externally and internally headed relative clauses. Section 2 discusses the relative clauses found in Shan. Section 3 examines the nominal structure of Shan. Section 4 offers an analysis of post-head and internally headed relative clauses. Section 5 concludes.

## 2 Relative clauses in Yǔn Shan: Internally headed and post-head externally headed

When a relative clause has a single overt head, there are three places that the head can appear: to the left of the clause, to the right of the clause, or inside the clause. When the head appears to the left of the clause, it is called a post-head relative. If it is to the right, it is called a pre-head relative. Finally, if the head is inside the clause itself, it is an internally headed relative clause.

### 2.1 Assumptions

In order to distinguish between internally and externally headed relative clauses, I use particular words that should appear between the relative clause head and the right or left edge of the clause if the head is still inside the clause. Since Shan has SVO word order, the object of an internally headed relative clause will follow the verb. In pre-head relative clauses, the extracted object will follow the verb because it appears to the right of the relative clause. Thus, it is important to be able to show whether the relative clause head is inside the clause. To mark the right edge of the CP, I use *wánnâj* (‘today’). A similar issue would make it difficult to distinguish post-head relatives with an extracted subject from IHRCs with the subject as the head of the clause. The complementizer *ʔăn*—which can be distinguished from other uses of *ʔăn*, as will be discussed in a later section—can indicate the left edge of the clause by appearing between an external head and its modifying clause.

As shown in (5-6), *wánnâj* (‘today’) can appear at the beginning or end of a clause. As (7) shows, it cannot appear anywhere else within the clause. Thus, when an object is to the left of *wánnâj* or a subject is to the right, we can assume that it is inside the clause.

- (5) *Nan Li sù kàj wánnâj.*  
 Nan Li buy chicken today  
 ‘Nan Li bought a chicken today.’ (Shan; 2017.03.21.29)

- (6) *Wánnâj* *Nan* *Li* *sù* *kàj*.  
 today Nan Li buy chicken  
 ‘Today, Nan Li bought a chicken.’ (Shan; 2017.03.21.32)
- (7) *Nan* *Li* (*\*wánnâj*) *sù* (*\*wánnâj*) *kàj*.  
 Nan Li today buy today chicken  
 Intended: ‘Nan Li bought a chicken today.’ (Shan; 2017.03.21.30-31)

By utilizing *wánnâj* it is possible to differentiate pre-head and internally headed relative clauses when the object is extracted. The following sections show what types of relative clauses can be used in a few different syntactic environments.

### 2.2 Relative clause head as matrix object

When I asked for a translation of a relative clause with the head as the matrix object, the first version offered was a post-head relative, as in (8). Here, *kàj* (‘chicken’) is the head of the relative clause, and it is the object of the matrix clause. The internally headed relative clause, as in (9), was also grammatical. For this sentence, the pre-head relative clause in (10) was not accepted. The bracketing notation in these examples is outer brackets around the full nominal structure and inner brackets around what seems to be inside the sentential clause portion of the relative clause. The relative head is in bold.

- (8) *Saj* *K<sup>h</sup>am* *ǰǎw* [***kàj*** [*ǰǎn* *Nan* *Li* *sù* *máa*  
 Saj Kham take chicken COMP Nan Li buy  
*máa wánnâj*]].  
 come today  
 ‘Saj Kham took a chicken that Nan Li bought today.’ (Shan Post; 2017.03.07.MM.21a)
- (9) *Saj* *K<sup>h</sup>am* *ǰǎw* [[*ǰǎn* *Nan* *Li* *sù* *máa*  
 Saj Kham take COMP Nan Li buy come  
***kàj*** *wánnâj*]].  
 chicken today  
 ‘Saj Kham took a chicken that Nan Li bought today.’ (Shan IH; 2017.03.07.MM.21c)
- (10) *\*Saj* *K<sup>h</sup>am* *ǰǎw* [[*ǰǎn* *Nan* *Li* *sù* *máa*  
 Saj Kham take COMP Nan Li buy come  
*wánnâj*] ***kàj*** *nâj*].  
 today chicken this  
 Intended: ‘Saj Kham took a that chicken Nan Li bought today.’  
 (Shan Pre; 2017.03.07.MM.21b)

### 2.3 Relative clause head as matrix subject

When I asked for a translation of a relative clause with the head as the matrix subject, the first version offered was an internally headed relative, as in (11). Here, *kàj* (‘chicken’), the head of the relative clause and the subject of the matrix clause, is clearly inside the relative clause. The post-head relative clause, as in (12), was also acceptable. For this sentence, the pre-head relative clause in (13) was not accepted.



- (11) [[ʔǎn Nan Li sù kàj wánnâj]] mán pěn  
 COMP Nan Li buy chicken today 3 be  
 sǐi kʰǎaw.  
 color white  
 ‘A chicken that Nan Li bought today was white.’ (Shan IH; 2017.03.07.MM.04a)
- (12) [Kàj [ʔǎn Nan Li sù wánnâj]] mán pěn  
 chicken COMP Nan Li buy today 3 be  
 sǐi kʰǎaw.  
 color white  
 ‘A chicken that Nan Li bought today was white.’ (Shan Post; 2017.03.07.MM.04b)
- (13) \*[[ʔǎn Nan Li sù wánnâj] kàj (nâj)] mán pěn  
 COMP Nan Li buy today chicken this 3 be  
 sǐi kʰǎaw.  
 color white  
*Intended:* ‘A chicken that Nan Li bought today was white.’ (Shan Pre; 2017.03.07.MM.04c,d)

Demonstrated here are only a few types of syntactic constructions that use relative clauses. More were tested but are not included due to space limitations. Post-head and internally headed relative clauses were accepted as grammatical almost consistently. Pre-head relatives were marginally accepted in a few cases, but upon retesting, they were rejected.

Most of the data in this paper was elicited; however, in a recording of a parable that my consultant told, there was an example of what is either an internally headed relative clause or a pre-head relative. This can be seen in (14). Here the head *sɿkʰó* (‘clothes’) follows the verb inside the relative clause. Given that pre-head relatives have been found to be less acceptable than IHRCs, it is likely that this is an internally headed relative clause.

- (14) Luk kɔ̄ kǎaŋ nâjhan̄kɔ̄ mán sù sàw  
 child CLF.PERSON middle TOP 3 buy put  
 [ʔǎn mán cɿk sɿ-kʰó hâw].  
 COMP 3 like clothes-clothes PL  
 ‘The middle child, he bought and put the clothes that he liked,’ (Shan; 2016.04.26.C.05)

From the evidence presented above, I conclude that this dialect of Shan uses post-head and internally headed relative clauses, but at this point there is not consistent evidence to suggest that Shan has pre-head relatives as well.

### 3 Shan nominal structure

Before giving an analysis of the Shan relative clause, it is first necessary to give an analysis of the Shan nominal structure. The basic order of nominal elements is NP Adj Num Clf Dem, as in (15).

- (15) Mǎa jàw sǎam tǔ nâj mán kʰǎaw  
 dog big three CLF.ANIMAL this 3 white  
 ‘These three dogs are white.’ (Shan; 2016.11.28.MM.27a)

#### 3.1 The functions of ʔǎn

Cognates of ʔǎn are found in several Southwestern Thai languages (Gedney 1994). Typically, they function as a general classifier for inanimate things and/or as a pronominal element like English ‘one’. Since it is relevant to the structure of the Shan nominal, I discuss the uses of ʔǎn in the language.

## 3.1.1 Classifier/Pronominal use

Classifiers appear after a noun when the noun is modified by a numeral, adjective, or demonstrative, as in (16-19). Each category of nouns (e.g., people, animals, round things, etc.) uses a different classifier. In (16), *má* is the classifier that goes with the word meaning ‘spoon’. For a small number of nouns, *ʔǎn* can be used with numerals in place of the noun-specific classifier, as (16) shows. The same can be found in Thai for an apparently broader class of nouns (Iwasaki and Ingkaphirom 2005:77). Some nouns can optionally have the noun-specific classifier appear before adjectives, as in (17), but this has only been found for nouns referring to people.

- (16) *cʰɔ̃ sǎam ʔǎn/má*  
 spoon three CLF.GENERAL/CLF.LONG  
 ‘three spoons’ (Shan; 2017.03.28.MM.10a,b)

- (17) *mɛjít kɔ̃ sǔŋ*  
 woman CLF.PERSON tall  
 ‘tall woman’ (Shan; 2016.11.28.MM.22)

More commonly, *ʔǎn* appears with the demonstrative *nâj*, as in (18). Appearing alone without a noun, *ʔǎn nâj* means ‘this’. Note that the noun-specific classifier *mík* can be used in addition to *ʔǎn nâj* in (18), which adds the meaning of ‘one’. This suggests that there are potentially two separate projections in the nominal structure: one for numeral classifiers and one associated with the demonstrative. The word *ʔǎn* can act as either a classifier or a pronominal type element associated with the demonstrative—like ‘this one’ in English. There are, however, restrictions on which nouns *ʔǎn* can appear with. For example, nouns referring to people are not compatible with *ʔǎn*, as shown in (19). Instead, the noun-specific classifier *kɔ̃* should be used.

- (18) *tʰò mík (ʔǎn) nâj*  
 bean CLF.BEAN AN this  
 ‘this one bean’ (Shan; 2016.11.28.MM.47)

- (19) *\*mɛjít (ʔǎn) nâj*  
 woman AN this  
 Intended: ‘this woman’ (Shan; 2016.11.28.MM.59,61)

## 3.1.2 Complementizer

The *ʔǎn* used in the circumstances mentioned in the previous subsection is homophonous with what I have been glossing as COMP (complementizer) in the relative clause. There are several reasons to think that the classifier and pronominal uses of *ʔǎn* are different from the *ʔǎn* that appears before relative clauses.

First, *ʔǎn* is not available as a numeral classifier for most nouns. For example, it would not be able to appear with nouns like *mɛjít* (‘woman’) in (19) since it is incompatible with that noun when *ʔǎn* is acting as a classifier or pronominal. Thus, it cannot simply be a classifier preceding a relative clause in the same way that the classifier precedes an adjective in (17). Only nouns referring to people have been found to allow classifiers in this position. Additionally, there are examples where the classifier *kɔ̃* appears in addition to *ʔǎn* before a relative clause, as in (20), so the noun-specific classifier used in this way is different from the *ʔǎn* in front of the relative clause.

- (20) *mɛjít kɔ̃ ʔǎn háw hǎn mɔ̃wáa nâj mán khèn*  
 woman CLF.PERSON COMP 1 see yesterday this 3 clever  
 ‘The woman that I saw yesterday is clever.’ (Shan; 2016.10.07.MM.13)

- (21) *mejítj kô nâj ʔăn mán ʔàan lik nâj mán khèn*  
 woman CLF.PERSON this COMP 3 read book this 3 clever  
 ‘This woman that read the book is clever.’ (Shan; 2016.04.11.C.MM.10c)

When it is used as a pronominal, it appears in conjunction with *nâj*, so the *ʔăn* that appears before relative clauses is not likely to be of this use. Also, the use of a classifier in conjunction with a demonstrative can co-occur in with the *ʔăn* preceding a relative clause, as in (21). In (3) above, *ʔăn* in this pre-relative clause position co-occurs with *ʔăn* as the demonstrative.

Another reason to think that *ʔăn* acts as a complementizer is that it appears in the same position as the Thai complementizer, *thîi*, in relative clauses (Jenks 2011). In addition, it is possible, but less common, to form a relative clause using the Shan cognate of *thîi*, *ti* ‘place’ (Gedney 1994). An example of this can be seen in (22). I would predict that *ʔăn* is not allowed to co-occur with *ti* in this position.

- (22) *lik ti háw hăn nâj mán lěj*  
 book COMP 1 see this 3 red  
 ‘The book that I see is red.’ (Shan; 2016.11.15.MM.01)

For these reasons, I analyse the *ʔăn* preceding relative clauses as a complementizer.

### 3.2 Demonstratives

Shan does not have articles like *a* and *the* in English, but it does have proximate and distal demonstratives. Here, I discuss the proximate demonstrative *nâj* (‘this’). When used in conjunction with a classifier, which might be *ʔăn* or a noun-specific classifier, *nâj* has the meaning ‘this’, as discussed in the previous section. The AN/CLF+Dem structure can appear alone or modifying a noun.

In addition to its use as a demonstrative, *nâj* also frequently appears at the right edge of a relative clause, as in (23). Here, the interpretation of *nâj* is not the same as ‘this’. Compare (23) with (24) which has the same surface structure except that the addition of *ʔăn* give the meaning of ‘this’.

- (23) [*ʔăn háw hăn lik jù nâj*] *mán lěj*  
 COMP 1 see book IMPF this 3 red  
 ‘The/A book that I am seeing is red.’ (Shan; 2017.02.28.MM.59)
- (24) [*ʔăn háw hăn lik jù ʔăn nâj*] *mán lěj*  
 COMP 1 see book IMPF AN this 3 red  
 ‘This book that I am seeing is red.’ (Shan; 2017.02.28.MM.60)

In (25), *nâj* appears to modify a headless relative clause, acting in much the same way as *nâj* in (23).

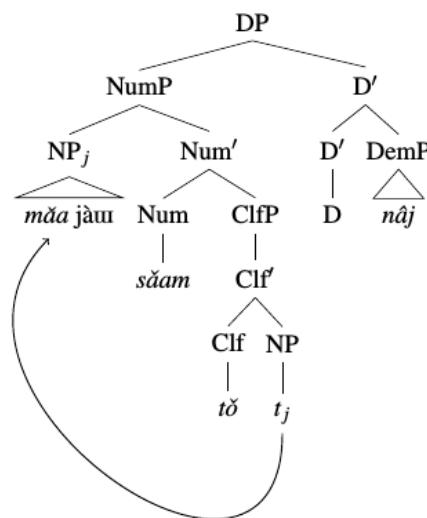
- (25) [*ʔăn máu jùm nâj*] *mán màw cau*  
 COMP 2 believe this 3 NEG true  
 ‘What you believe is not true.’ (Shan; 2016.04.26.P.MM.04)
- (26) *Măa tǒ nâj mán wáaŋŋaj Cu mán nâj Spot*  
 dog CLF.ANIMAL this 3 clever name 3 this Spot  
 ‘This dog is clever. His name is Spot.’ (Shan; 2017.02.07.MM.36)

In (26), the *nâj* in the second sentence is used in a way different from English ‘this’. In Thai, the demonstratives or particles derived from them mark subject topics (Iwasaki and Ingkaphirom 2005:361). At this point, I will only claim that *nâj* has the demonstrative meaning when it precedes *ʔăn* or a noun-specific classifier and leave the meaning of *nâj* in other circumstances to future work.

### 3.3 Proposed structure

I follow Simpson’s (2005) proposal for the structure of the Thai DP in having the numeral and classifier in separate projections and the ClfP selecting the NP. Then, the NP moves to a position to the left of the numeral to get the word order NP Num Clf. The demonstrative is right-adjoined to the DP. Since a classifier is necessary to get the demonstrative interpretation, I propose that the pronominal *ʔǎn* appears in D. For nouns that are not compatible with *ʔǎn* in this position, they would require the noun-specific classifier. Since this classifier is never doubled, it would need to be deleted to avoid repetition. This would be consistent with other Shan data that shows that string-adjacent *nâj*’s are not allowed. Another possibility would be for NumP to be the complement of D and then move to a higher position. Then, the noun-specific classifier would move out of the Clf to D before NumP moves to a higher position. The full proposed structure can be seen in Figure 1.

Figure 1: Proposed structure for Shan DP



## 4 Analysis

### 4.1 Position of nominal structures in relation to relative clause

In (27), the head of the relative clause *mǎa* (‘dog’) is modified by the noun-specific classifier *tǎ* and *nâj*. Since, in the interpretation of the sentence, the seeing took place yesterday, *mɔwáa* (‘yesterday’) is interpreted inside the relative clause. Thus, *mǎa tǎ nâj* is inside the relative clause. Given that ‘yesterday’ can be translated as *mɔwáa* or *mɔwáanâj*, I am not making any claims about the position of *nâj* in this sentence.

- (27) [*ʔǎn háw hǎn mǎa tǎ nâj mɔwáa nâj*] *mán lěj*  
 COMP 1 see dog CLF.ANIMAL this yesterday this 3 red  
 ‘This dog that I saw yesterday is red.’ (Shan; 2016.05.17.P.MM.5)

(28) shows coordination of two nouns modified by classifiers within the relative clause. This indicates that when the Clf+Dem is adjacent to the head, also in (27) above, the head and Clf+Dem can form a constituent.

- (28) [ʔǎn háw hǎn mǎa tǒ nâj táŋhêŋ mjaw tǒ nâj]  
 AN 1 see dog CLF.ANIMAL this and cat CLF.animal this  
 mán kʰǎaw.  
 3 white  
 ‘The dog and the cat that I see are white.’ (Shan; 2016.10.18.MM.42)
- (29) [ʔǎn háw hǎn lik jù ʔǎn nâj] mán lěŋ  
 COMP 1 see book IMPF AN this 3 red  
 ‘This book that I am seeing is red.’ (Shan; 2017.02.28.MM.60)

Examples like (24), repeated in (29), show that in some cases there can be a word intervening between the head and the right edge of the clause. Thus, the Clf+Dem, ʔǎn nâj appears to be outside the relative clause. I would argue that this position is higher than the relative clause and part of the DP structure that contains the relative clause. (30) shows that when two relative clauses are modifying the same noun, there is only one nâj adjacent to the relative clause, and it appears to the right of the final relative clause.

- (30) [màmɔ̂ [ʔǎn mí nǚ lik (\*nâj)] [ʔǎn Saj Kʰam láaŋ jâw]  
 apple COMP exist on book this COMP Saj Kham wash PRF  
 =nâj] mán lěŋ  
 this 3 red  
 ‘The apple, which is on the book, which Saj Kham washed already, is red.’  
 (Shan; 2017.05.30.MM.43,44)

Comparing data in (27) and (29) and taking into consideration that nâj can appear more than once in a nominal structure containing a relative clause, it seems like there are two possible positions for a DP to be projected: at the highest layer of the nominal structure and at the base-generated position of the head inside a relative clause.

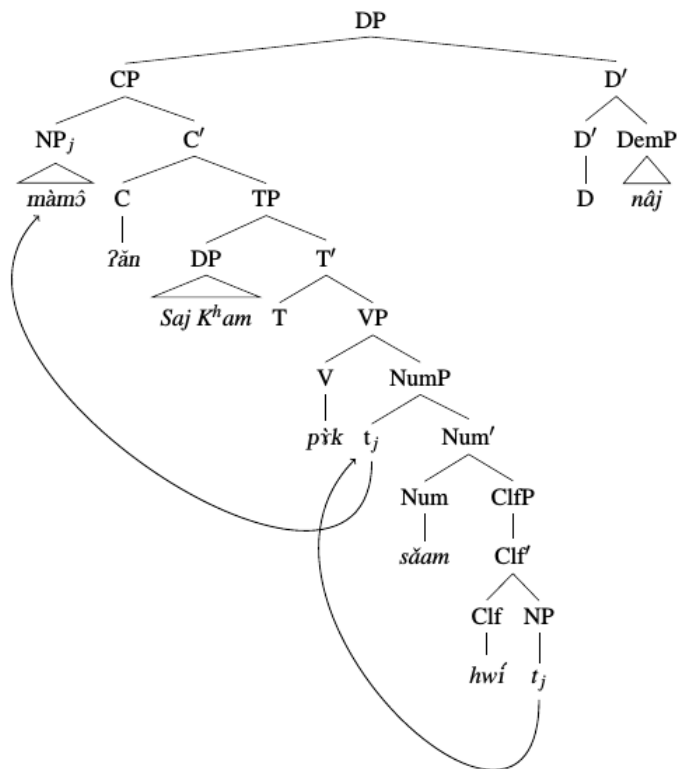
#### 4.2 A raising account of Shan relative clauses

I am proposing a raising analysis as has been used for Choctaw by Broadwell (1985) and for Ancash and Imbabura Quechua by Cole (1987), among others. This type analysis assumes that the LF representation of internally and externally headed relative clauses is the same.

For examples like (31), where there is a quantifier inside the relative clause and the head is either inside the clause (an IHRC) or outside the clause (a post-head RC), the proposed structure would be as in Figure 2.

- (31) Nan Li cǐn pěn [(màmɔ̂) [ʔǎn Saj Kʰam pǐk]  
 Nan Li eat be apple COMP Saj Kham peel  
 (màmɔ̂) sǎam hwí] nâj]  
 apple three CLF.ROUND this  
 ‘Nan Li ate apples that Saj Kham peeled three of.’ (Shan IH/Post; 2017.02.28.MM.13,17)

Figure 2: Proposed structure for numerals inside relative clause



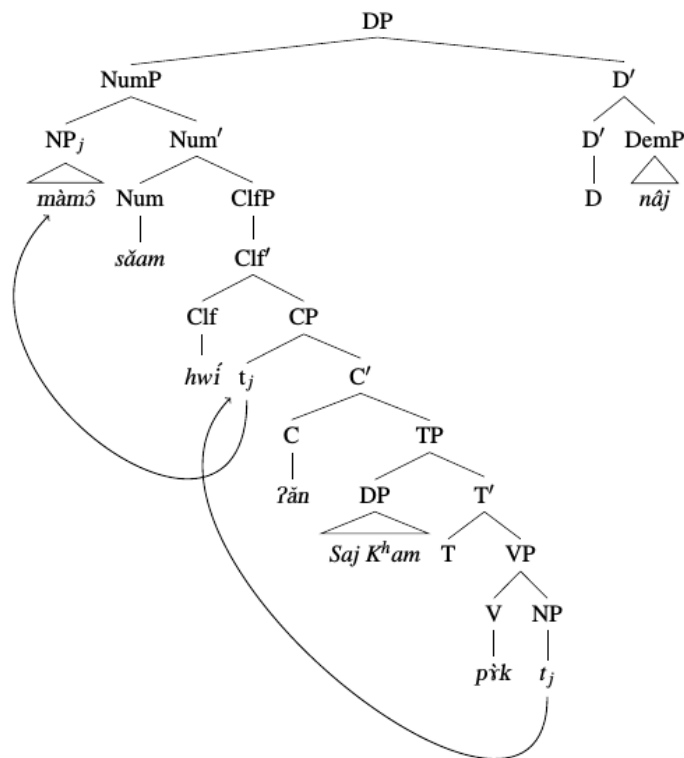
Here, the head is generated as the complement of the classifier phrase and moves through the SpecNumP into the SpecCP of the relative clause. For internally headed relative clauses, the movement out of the NumP would happen at LF.

For examples like (32), where the quantifier is outside the relative clause, there are two options: either the whole DP/NumP moves from inside the clause to SpecCP or just the head NP is generated inside the clause and that moves through the SpecCP of the relative clause into a higher position in the highest DP. Given that we find externally headed examples like (21) with NP Clf Dem appearing immediately to the left of the relative clause and internally headed examples like (27) with NP Clf Dem inside the relative clause, we might want to allow for full DP movement.

- (32) *Nan Li cĭn pĕn [mâmô sǎam hwí [ʔǎn Saj*  
 Nan Li eat be apple three CLF.ROUND COMP Saj  
*K<sup>h</sup>am pøk] nâj]*  
 Kham peel this  
 ‘Nan Li ate three apples that Saj Kham peeled.’ (Shan Post; 2017.02.28.MM.14)

There are a few reasons to think that the numeral/quantifier material is base-generated in the surface position. First, the quantificational material is interpreted at its surface scope, as can be seen in the difference in interpretation between (31) and (32). Second, in none of the data does NP Num Clf Dem appear immediately to the left of the relative clause. When a relative clause has the interpretation ‘these # NPs that...’ the structure is NP Num Clf Rel Clf Dem. This suggests that the numeral is a part of the nominal structure above the relative clause. Thus, the proposed structure for (32) is as in Figure 3.

Figure 3: Proposed structure for (32) with numerals outside relative clause



Note that this analysis predicts that it should be possible to get a Num Clf in front of an internally headed relative clause. This is something that still needs to be tested.

### 4.3 Wh island effects

A raising account predicts that extraction from a relative clause embedded within another relative clause would be bad. This appears to be the case in Shan. In (33) is an internally headed relative clause embedded in another internally headed relative clause. The externally headed version is similarly bad.

- (33) \**[ʔăn Nan Li waa kǎn táŋhêŋ [ʔăn kón ʔàan lik nâj]] mán lěŋ*  
 COMP Nan Li spoke together with COMP person read book  
 this 3 red

Intended: ‘The book that Nan Li spoke with the people who read (it) is red.’

(Shan; 2017.05.30.MM.38)

- (34) \**[lik [ʔăn Nan Li waa kǎn táŋhêŋ [ʔăn kón ʔàan nâj]] mán lěŋ*  
 book COMP Nan Li spoke together with COMP person read  
 this 3 red

Intended: ‘The book that Nan Li spoke with the people who read (it) is red.’

(Shan; 2017.05.30.MM.39)

## 5 Conclusion

In this paper, I have demonstrated that Yün Shan, a Tai language, does have internally headed relative clauses in addition to post-head relative clauses. Further, I have proposed an analysis of the Shan DP and a simple head raising analysis of Shan post-head and internally headed relative clauses. Given that a VO word order language like Shan is less likely to have this construction than an OV language and that closely related languages like Thai and Lao do not allow IHRCs, it would be interesting to investigate what features are common in languages that have IHRCs, particularly VO languages. This will be pursued in future work.

## References

- Aldridge, Edith. 2004. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 5:99-129.
- Cole, Peter. 1987. The structure of internally headed relative clauses. *Natural Language & Linguistic Theory* 5:277-302.
- Broadwell, G. A. 1985. Internally headed relative clauses in Choctaw. *Kansas Working Papers in Linguistics* 10.
- Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info> [accessed on May 19, 2017].
- Gedney, William. 1994. *J. William J. Gedney's Southwestern Tai Dialects: Glossaries, Texts, and Translations*. Edited by Thomas J. Hudak. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- Hiraiwa, Ken. 2003. Relativization in Buli. *Studies in Buli Grammar. Working Papers on Endangered and Less Familiar Languages* 4:45–84.
- Iwasaki, Shoichi, and Ingkaphirom, Preeya. 2005. A reference grammar of Thai. Cambridge University Press.
- Jenks, Peter. 2011. *The hidden structure of Thai noun phrases*. Harvard University Cambridge, Massachusetts dissertation.
- Liddell, Scott K. 1980. *American Sign Language Syntax* Vol. 52. Mouton De Gruyter.
- Morev, Lev Nikolaevic, and Alekseĭ Alekseevich Moskalev. 1979. *The Lao language*.
- Prasithrathsint, Amara and Natchanan Yaowapat. 2009. A typology of relative clauses in mainland Southeast Asian languages. *Mon-Khmer Studies* 38:1-23.
- Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. *The Oxford handbook of comparative syntax*, 806-838.
- Warotamasikkhadit, Udom. 1972. *Thai syntax: An outline* Vol. 68. Mouton De Gruyter.



# DOCUMENTING A FUZZY TONE SYSTEM: THE CASE OF MUKLOM TANGSA<sup>1</sup>

Mijke Mulder  
*La Trobe University*  
*mijke.mulder@gmail.com*

## Abstract

This paper provides a preliminary description, based on primary data, of the tone system of Muklom Tangsa. This is an endangered and understudied Tibeto-Burman language variety that is spoken in Northeast India. A striking feature of Muklom tones is their variable or fuzzy realisation. It is argued that tonal fuzziness is allowed by (a) the complex nature of Muklom tone categories, i.e. hearers rely on multiple cues in tone perception, and (b) low functional load of the tones. Fuzzy realisation, in combination with low metalinguistic awareness, makes Muklom Tangsa tones particularly hard to determine. Fortunately, cooperation between researcher and consultant can tackle tone determination in a fuzzy tone language.

**Keywords:** tones, Muklom, Tangsa  
**ISO 639-3 codes:** nst

## 1 Introduction

When we take up the task of analysing a previously undescribed tone language, we are likely to come across some challenges in determining the character and number of tones. In a special series titled ‘How to Study a Tone Language’, in the journal of Language Documentation and Conservation, edited by Steven Bird and Larry Hyman (2014), scholars developed different strategies of tackling the description of tone systems, while providing examples from various languages. This paper aims to make a small contribution to this endeavour by exploring the tone system of Muklom Tangsa and by reflecting on tone data elicitation methods for languages that have similar tone systems. The discussion of Muklom tones is based on a database of over 2000 lexemes, which has been collected during two recent field trips by the author, and on metalinguistic discussions with Muklom consultants.

Muklom is an endangered and understudied Tibeto-Burman language spoken in the Changlang district of the state Arunachal Pradesh in Northeast India. The speakers, who refer to themselves as [muklɔm] and to their language as [muklɔm raʔ] ‘Muklom speech’, dwell in two clusters of villages: the oldest settlements near Changlang Town in the hills, and the more recent settlements near Kharsang Town in the plains. The Mukloms are part of a larger ethno-cultural group called Tangsa, their so-called ‘tribe’. The Tangsa region spans the India-Myanmar border, but the Muklom people only reside on the Indian side. In attempts to classify the Tibeto-Burman languages of Northeast India, scholars have assigned the language varieties of Tangsa subtribes to various higher and lower level groups. The proposed branches Sal, Konyak, and Northern Naga subsume Tangsa (e.g. Shafer 1955; Benedict 1972; French 1983; Bradley 1997; Burling 2003; Blench & Post 2015).

---

<sup>1</sup> The research on which this paper is based and the presentation of this paper have been made possible by funds from La Trobe University, Melbourne, Australia (scholarships LTUPRS and LTUFFRS, grants from HuSS IRGS and Linguistics DRP). This study would not have been possible without the help and hospitality of many Muklom consultants, mainly from Kuttom, New Chingsa and Kharsang Town, but also from New Janman and New Khimiyong. For tone elicitation and determination in particular, I would like to thank Ms Tchithan Techhi and Mr Nongmai Changmi who contributed many hours of hard work in sessions on lexicon, but also ‘i<sup>3</sup> βΛŋ<sup>2</sup>’ Mr Ngulang Changmi, Ms Sunumi Changmi, Mr Phanglong Sajung, Mr Chupan Changmi, Mr Munkap Khimhun, and Mr Munson Khimhun. I am grateful to Stephen Morey and two anonymous reviewers for their comments on the first version of this paper and to Tamir Dingjan for proofreading the same version.

Three different names are used to refer to this larger language group that Muklom belongs to: Tangsa, Tangshang, and Tase (ISO 639-3: nst). Prior to the coining of these umbrella terms, the subgroups were referred to by their subtribe names, with the possible addition of ‘Naga’. The first term, ‘Tangsa’, was coined by Indian official Bipin Borgohain (foreword in Barua 1991:vii), who states that ‘Tang’ stands for ‘mountain’ and ‘sa’ means ‘person’. ‘Tangsa’ remains to be used in India. In Myanmar, however, the label ‘Tangshang’ has been in use since its inauguration by the Tangshang Central Culture and Literature Committee in Nanyun in 2003. This term is not directly cognate with ‘Tangsa’ but derives from the names of two ancestors in local oral history, the brothers Tangnyu Wang and Shangnyu Wang (Statezni 2013:5). ‘Wang’ means king or chief. It should be noted that the scope of Tangshang differs from that of Tangsa. The label Tangshang in Myanmar subsumes Nocte, Wancho and Tutsa, while in India these are considered separate groups (Statezni 2013:7). The third language label, ‘Tase’, or fully ‘Naga Tase’, is employed by Ethnologue (Simons and Fennig 2017) and represents the pronunciation of the word Tangsa in the Chamchang<sup>2</sup> variety. For the sake of simplicity, I will use the term Tangsa throughout this paper, as this is the standard inside India.

The Tangsa group is linguistically quite diverse: Morey (2014b: 63; 2017) estimates there are approximately 70-80 Tangsa subtribes on both sides of the Indian-Burman border. Each subtribe speaks their own language variety, which are sometimes mutually intelligible but often are not. The internal classification of Tangsa is still a work in progress, but our knowledge of Tangsa varieties has been steadily growing since the start of the DoBeS project on Tangsa, Tai and Singpho in Northeast India in 2005, directed by Stephen Morey, and the commencement of language surveys around the same time in Myanmar by the Linguistic Society Naga Survey Team in cooperation with SIL International (Statezni 2013:25). While the label Tangsa serves its function as a modern ethno-cultural group term, it has yet to be determined whether the Tangsa dialects constitute a valid linguistic subgroup. I will not go deeper into the topic of sub-grouping here, although it may be worth noting that Nocte has been suggested as the closest linguistic relative of Tangsa (e.g. Shafer 1953:228; French 1983:726; Burling 1983:17; Bradley 1997:21; Burling 2003:175). Indeed, some language varieties in the Tangsa and Nocte territories are sub-grouped based on their geographical location rather than on linguistic grounds. For example, the speakers of the Hakhun and Phong/Ponthai language varieties are referred to as Tangsa when they live within Tangsa territory in the Changlang district of Arunachal Pradesh or in Assam, but Nocte when they dwell within Nocte territory in the Tirap district of Arunachal (Morey 2017).

While research into Tangsa is fortunately on the rise, published original literature on Muklom Tangsa specifically remains extremely limited. The only published modern linguistic research incorporating field work has been carried out by Morey (2014a:667; 2015: passim; 2017), who has also reinterpreted data from earlier sources (Morey 2011). In addition to this, there is work underway: the author of this paper is carrying out a PhD project on the Muklom variety and Kellen van Dam is carrying out a PhD project on tone systems across Tangsa varieties, both under supervision of Stephen Morey.

Earlier sources on Muklom Tangsa are: 1) the work of the Language Officer of Changlang in Arunachal Pradesh, Ngemu (1977), a sixty-six-page description of Muklom including remarks on grammar, phrases with English translation and a Muklom-English wordlist; 2) the work of the Director of Research in Arunachal Pradesh, Das Gupta (1980), a six-page language sketch; 3) the work of Bandyopadhyay (1989), a thirteen-page wordlist. It should be pointed out that the language variety under discussion here is usually referred to in previous sources as ‘Moklum’ (e.g. Das Gupta 1980), though ‘Muklom’ is also found. Since the native pronunciation of the subtribe and its language is [muklɔm], I consider the latter spelling preferable.

This paper starts off by providing a preliminary description of the Muklom Tangsa tone system (§ 2). Muklom exhibits a typical Tangsa tone system, though at least one deviant characteristic may exist. The language has a relatively simple three-way tone system, but tone determination is impeded by fuzzy realisation and low metalinguistic awareness (§ 3). Three factors that may explain fuzziness and low awareness are (1) the complex nature of tone categories, (2) low functional load of tones, (3) tone Sandhi. To determine tone on lexical and grammatical items in Muklom, I have applied a strategy that combines researcher metalinguistic knowledge with native speaker intuition of consultants (§ 4).

---

<sup>2</sup> An alternative spelling is ‘Kimsing’.

## 2 A typical Tangsa tone system

The Muklom tone system consists of three tone categories, as is typical for Tangsa dialects, which tend to have a three-way tone contrast on open syllables and syllables with a non-plosive coda (together ‘non-stopped syllables’), and no contrastive tone on syllables ending in a plosive (‘stopped syllables’) (Morey 2014a:641). For Tangsa languages that follow this tendency, we have to decide how to analyse the tonal status of stopped syllables. In my understanding, we can either propose they do not carry tone, or we can propose that syllables with a stopped coda constitute a fourth tone. Morey chooses the latter option in a discussion of tone systems in seven different Tangsa dialects and treats ‘stop final words as a fourth category’ (Morey 2014a:658).

Although the majority of Tangsa varieties exhibit a three-way tone system, the realisation of the tone categories varies considerably between the language varieties. For this reason, we need to keep tone category and tone realisation strictly separated. Let me exemplify the distinction by looking at the three tone categories that occur on non-stopped syllables. When considering groups of cognates in a comparative Tangsa wordlists (Morey, p.c.), we find patterns of correspondences in tone realisation. For example, we see that many words with mid-to-low falling tone in Muklom Tangsa, carry high, level creaky tone in Cholim Tangsa, and high, falling tone in Mungray Tangsa (Morey 2014a:658). This category will be referred to as tone 1. In another group of cognates, we find that mid-high-mid rising-falling tone in Muklom corresponds to high falling tone in Mueshaungx Tangsa<sup>3</sup>, and mid, level tone in Chamchang Tangsa. This category will be referred to as tone 2. In a third group of cognates, we find that mid, level tone in Muklom corresponds to high, rising in Mueshaungx Tangsa, and low, falling in Mungray Tangsa. This category will be referred to as tone 3. The Tangsa varieties that are mentioned here have been selected randomly; the same applies to the lexicons of other varieties. The point that I want to make by providing these examples is that while the abstract categories are the same across dialects (at least for these three tones), the phonetic realization of the categories may differ considerably across dialects. Referring to tones by category number is therefore preferred to labels based on pitch features. This is not a new convention in the context of Tangsa languages. For example, Morey (2011; 2014a; 2014b; 2015; 2017 to name a few) employs superscript numbers to indicate the tone of the preceding syllable, and Weidert (1979; 1987) uses superscript numbers to indicate the tone of the following syllable.

Table 1 illustrates the tone system of Muklom with three example words. It is important to point out that in this dialect, tone consists of more than pitch height or contour. As is not uncommon for tone systems, additional cues play a role. For Muklom, I would like to propose that phonation and duration are relevant cues, in addition to pitch height and contour. The data seem to point towards a three-way distinction for phonation, namely breathy, modal, and creaky. In order to describe pitch height and contours on a phonemic level, three values seem to suffice: high, mid, low. More research is required to determine whether there is a statistically significant difference in duration between tokens of the three tone categories, but the impression is that tone 1 is shorter, and tone 2 is longer. The suggestion that duration is an important cue in tone perception is supported by the fact that the presence of an inherently short vowel in a word negatively affects the ability of speakers to name the tone of that word in a metalinguistic conversation. For example, words carrying tone 2 which seems to be characterised by long duration, are less quickly recognised when the vowel is short. This topic will be explored further in Section 4.

In Table 1, the tone-carrying syllable has the nasal /ŋ/ in coda position. Other codas on tone-carrying syllables are: nasals /m/ and /n/, and lateral /l/. Coda-less syllables also carry tone. In other words, all tone-carrying syllables are of the type (C)V(N/L), where N stands for nasal and L for lateral. See Table 2 for an example of a CV minimal tone triplet.

---

<sup>3</sup> Alternatively, spellings include Mueshaung, Mossang and Moshang.

**Table 1:** *Tone categories in Muklom Tangsa*

Tone	Example	Gloss	Pitch	Phonation	Duration
1	<i>k<sup>h</sup>uŋ<sup>1</sup></i>	nose	mid-low falling	creaky	short
2	<i>k<sup>h</sup>uŋ<sup>2</sup></i>	plate	mid-high-mid rising-falling	modal	long
3	<i>k<sup>h</sup>uŋ<sup>3</sup></i>	path	mid level	breathy	medium

**Table 2:** *Coda-less tone triplet*

Example	Gloss
<i>kɔ<sup>1</sup></i>	threshold
<i>kɔ<sup>2</sup></i>	few
<i>kɔ<sup>3</sup></i>	genitals

The remaining consonants that Muklom allows in coda position do not to combine with tone, though there may be a few exceptions. The remaining codas are /p/, /t/, /k/, and /ʔ/, all non-aspirated voiceless plosives. The working model adopted here is that Muklom syllables of the type (C)VP, where P stands for plosive, do not carry tone. However, potential counterexamples exist. Several consultants identify a tone contrast on stopped syllables. Muklom has an inventory of ten vowels: /i/, /ɪ/, /u/, /ɔ/, /ʌ/, /a/, /ɛ/, and /e/, plus two long vowels /u:/ and /ɔ:/. The trend is that consultants identify stopped syllables with nuclei /ɪ/ and /ʌ/ as tone 1. Tone 1 is low and short in duration, at least compared to tone 2. It seems that these two vowels /ɪ/ and /ʌ/ are also short in duration, hence the confusion with tone 1. Never are stopped syllables containing /ɪ/ or /ʌ/ associated with long duration tone 2 by speakers. Similarly, stopped syllables with /i/ are never identified as carrying tone 1, unless the coda is a glottal stop, creating easy confusion with the creakiness of tone 1. It appears that /i/ is longer in duration than the two previously mentioned vowels, hence the confusion with tone 2. In this manner we can explain consultants' tone identification of syllables that we did not expect to be tone-bearing as deriving from interference by vowel duration. Table 3 presents a few examples of near (not full) minimal pairs that consultants often perceive as carrying tone.

**Table 3:** *Stopped syllables (incorrectly?) identified as carrying tone*

Word (tone 1)	Gloss	Word (tone 2)	Gloss
<i>tʌp</i>	fireplace	<i>tap</i>	shed, tent
<i>p<sup>h</sup>ʌt</i>	wash hair	<i>p<sup>h</sup>at</i>	vomit
<i>pit</i>	ask	<i>pit</i>	soft
<i>k<sup>h</sup>it</i>	dig out	<i>k<sup>h</sup>it</i>	silver

As mentioned above, potential counterexamples do exist. Stopped syllables containing /u/ or /ɔ/ are not identified by speakers as belonging to one singular tone category but fall into two groups: syllables that consultants assume to carry tone 1 and syllables that they take to carry tone 2. I have found that vowel duration plays a role: there is a phonemic distinction between short /u/ and long /u:/, and between short /ɔ/ and long /ɔ:/. A few minimal pairs are shown in Table 4.

**Table 4:** *Stopped syllables (incorrectly?) identified as carrying tone*

Short vowel	Gloss	Long vowel	Gloss
<i>puk</i>	stomach (organ)	<i>pu:k</i>	elephant
<i>luk</i>	frog	<i>lu:k</i>	tall
<i>tɔk</i>	maternal aunt	<i>tɔ:k</i>	winnow
<i>p<sup>h</sup>ɔt</i>	boil	<i>p<sup>h</sup>ɔ:t</i>	smoke

Although I propose to interpret these and other minimal pairs of the same type as phonemic vowel distinctions rather than tonal distinctions, it should be noted that consultants have so far been unwilling to accept the idea that stopped syllables cannot carry contrastive tone. Consultants do, however, accept unstressed, (C)V coda-less syllables as being toneless. Consultants often indicate they do not perceive

tone on these syllables, and comment that they are ‘too short’ to catch a tone or that their tone is ‘in between’ categories. When consultants do try to determine the tone category of this type of syllable, they will assign it to tone 3 or tone 1. Note that these syllables are never thought of by consultants as carrying tone 2, a tone which is characterised by long duration.

A few examples of such unstressed (C)V syllables are provided in Table 5. Toneless open syllables are always grammatical elements, like the four morphemes presented in Table 5. A short note on the nominalisers: /ʌ/ derives a noun from an adjective, and /u/ derives a noun from a verb.

**Table 5: Toneless (C)V syllables**

Morpheme	Gloss	Example word	Gloss
ʌ	NMLZ	ʌ-xʌl <sup>2</sup>	(the) good (one)
u	NMLZ	u-wat	hitting
t <sup>h</sup> ʌ	RECIP	t <sup>h</sup> ʌ-mɔl <sup>3</sup>	fight one another
ta	CAUS	ta-k <sup>h</sup> am <sup>2</sup>	warm up (trans)

If consultants are willing to accept toneless open syllables, but less keen to accept the idea of toneless stopped syllables, we should maintain the possibility that tone is playing a (limited) role in the latter syllable type. Perhaps the distinction between short and long vowels /u/ and /u:/, and /ɔ/ and /ɔ:/ in stopped syllables is a ‘side-effect’ of tone, just as it is a property of tone in non-stopped syllables.

When investigating lexical elements, we find out that if two mono-syllabic lexical morphemes combine to form a compound, the first element almost always retains its tone, as illustrated by Examples 1-3.

- (1) wɔ<sup>1</sup> → wɔ<sup>1</sup>raŋ<sup>2</sup>  
bird hornbill (bird)
- (2) k<sup>h</sup>ɔ<sup>2</sup> → k<sup>h</sup>ɔ<sup>2</sup>tɛip  
hair hair clip
- (3) na<sup>3</sup> → na<sup>3</sup>tsɔŋ<sup>3</sup>  
ear earring

Although the general pattern seems to be that these first elements of compounds retain their tone, I have found one exception in the data so far collected. Example 2 shows how the element k<sup>h</sup>ɔ<sup>2</sup> ‘head’ can retain its tone after compounding, but it does not always do so. In fact, in most other compounds with ‘head’ encountered so far, the first element changes from tone 2 into tone 3. Tone changes for the word ‘head’, as found in Muklom, appear to be quite common across Tangsa varieties (Morey, p.c.). Examples 4-6 illustrate this process. Note that k<sup>h</sup>ɔ<sup>2</sup> can mean both hair and head. To disambiguate, speakers can use k<sup>h</sup>ɔ<sup>3</sup>run<sup>2</sup>, literally ‘head hair’.

- (4) k<sup>h</sup>ɔ<sup>2</sup> + run<sup>2</sup> → k<sup>h</sup>ɔ<sup>3</sup>run<sup>2</sup>  
head hair head hair
- (5) k<sup>h</sup>ɔ<sup>2</sup> + tsip → k<sup>h</sup>ɔ<sup>3</sup>tsip  
hair/head scales dandruff
- (6) k<sup>h</sup>ɔ<sup>2</sup> + fit → k<sup>h</sup>ɔ<sup>3</sup>fit  
hair/head brush hair brush

From the available examples, we cannot deduce an underlying mechanism based on meaning or synchronic phonological context to explain this pattern. The meanings of the words do not point us in the direction of an answer. On the contrary, k<sup>h</sup>ɔ<sup>2</sup>tɛip ‘hair clip’ and k<sup>h</sup>ɔ<sup>3</sup>fit ‘hair brush’ are quite similar in meaning, both being items that are used on the hair, yet the first does not show tone change while the second does. Similarly, synchronic phonological context does not provide an answer. The second

elements of the compounds *kʰɔʔtɛip* ‘hair clip’ and *kʰɔʔtsip* ‘dandruff’ differ only by place of articulation of the onset affricate, yet in the first compound no tone change takes place while it does in the second. As one of the anonymous reviewers of this paper suggested, there may be a diachronic explanation. We could propose a tone Sandhi rule that changed the tone of first elements in compounds into third tone, which at some point of time stopped being productive. Compounds with third tone like *kʰɔʔrun²* ‘head hair’, *kʰɔʔtsip* ‘dandruff’, and *kʰɔʔʃit* ‘hair brush’ would have been formed before or while the tone Sandhi rule was still active. Contrastively, the compound *kʰɔʔtɛip* ‘hair clip’ would have been introduced into the language after the rule had stopped being active. Tone change in compounding remains a topic for further research, and the hypothesis discussed here can be tested by creating a task in which consultants coin new compounds with ‘head’. If they maintain the second tone, it suggests the third tone compounds are fossilized expressions that reflect an earlier tone Sandhi rule.

The Muklom tone system is here described as a three-way system, but it is worth noting that in some Tangsa orthographies, syllables with a glottal stop coda are treated as a fourth tone category (e.g. Chamchang orthography, presented by Bynn Kham Lann 2015; Hakhun orthography, as reported by Morey 2014a:656; Mueshaungx orthography, as reported by Morey 2014a:657). Muklom also exhibits glottal stop codas, but these are treated here as belonging to the same group as other syllables with plosive codas: -ʔ, -p, -t, -k. The glottal stop coda is without doubt distinct from tone 1. While in casual speech the two may be realised in quite similar manner, in careful speech the glottal stop coda syllables are shorter in duration and do not necessarily require falling pitch, as opposed to tone 1 syllables. A few minimal pairs are provided in Table 6. For /ɛ/ I have not found an exact minimal pair. A near minimal pair would be the second syllable of /rʌŋ²nɛʔ/ ‘afternoon’ and the word /mɛ¹/ ‘tail’.

**Table 6:** *Minimal pairs: tone 1 versus glottal stop coda*

Word	Gloss	Word	Gloss
<i>wa¹</i>	father	<i>waʔ</i>	bamboo
<i>xi¹</i>	sacred	<i>xiʔ</i>	faeces
<i>kɔ¹</i>	threshold	<i>kɔʔ</i>	give
<i>su¹</i>	grandchild	<i>suʔ</i>	thorn

### 3 Fuzzy realisation and low metalinguistic awareness

Despite the apparent simplicity of a three-way tone system, the tones of Muklom do present obstacles to analysis. I have identified the following two main challenges: 1) fuzzy realisation, 2) low metalinguistic awareness. This section will discuss what these challenges entail and proposes that the cause of these issues lies in 1) the complex nature of Muklom tones, 2) the low functional load of tone in Muklom, 3) presence of tone Sandhi.

When I arrived at the field site near Kharsang for the first time, I expected to be confronted with a tonal language. Stephen Morey had visited the Mukloms a few years earlier and succeeded to record a few tone minimal pairs<sup>4</sup>. Bandyopadhyay (1989) incorporates tone marking in his wordlist of Muklom, using grave and acute accent. This publication does not mention which variety is discussed, but the vast overlap in lexicon beyond doubt confirms it is Muklom. It is therefore surprising to observe that his tone transcription deviates considerably from my findings. For example, on the first page he uses an acute accent for both a tone 1 word and a tone 2 word: {aahúŋ} which I recorded as *ʌhup¹* ‘there is not,’ and {áang} which I recorded as *ʌŋ²* ‘to be’. Several factors may be responsible for this discrepancy: 1) the variety of Muklom that Bandyopadhyay reports on may be different from the one reported on here, 2) tone changes may have taken place, 3) some errors may have crept into his transcription. Even though the tone indication is different, and perhaps unreliable, it shows one thing: the writer was aware of the presence of a tone system in Muklom. Further reason for me to expect a tonal language was the fact that in almost all the Tangsa dialects for which scholars had collected a wordlist, tone was found to play a role. However, when I started learning the language, the tones did not stand out to me. They

<sup>4</sup> “Muncha – Minimal Pairs” from The Language Archive: <https://hdl.handle.net/1839/00-0000-0000-0017-C4EA-F@view>.

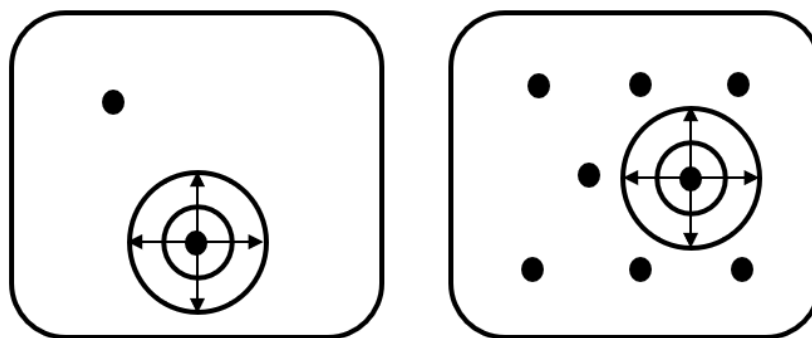
were certainly present, but they did not seem very pronounced. Interestingly, a native speaker of Mueshaungx Tangsa who is in contact with speakers of Muklom, can understand at least some words, and has a high level of metalinguistic awareness of his own variety due to involvement in language documentation, pointed out to me that ‘Muklom has no tones’. Clearly, even to a speaker of a tonal, related Tangsa variety, the Muklom tones may not seem very pronounced. The reason why Muklom tones are not very salient in the perception of (some) outsiders is that their realisation is quite variable.

What makes Muklom tones variable, or fuzzy? As described above, Muklom tones consist of a bundle of features. Pitch height, pitch contour, phonation, and duration are all relevant cues in tone perception. Even in careful speech such as wordlist recordings, on which the current analysis of tones is based, not all cues match the prototype. For example, speakers differ greatly as to how creaky their first tone is. In some speakers, tone 1 tokens in wordlist recordings are more often pronounced with modal rather than creaky phonation. Similarly, for tone 3, breathiness can often not be perceived. Interestingly, tone 2 tokens sometimes occur with breathy voice, while typically they have modal phonation. The same fuzziness is observed in pitch contours. For example, in some of the wordlist recordings, one older speaker exhibits typical phonation but deviant pitch contour. For example, he frequently does not show level contour for tone 3 tokens. Instead, the speaker exhibits a falling pitch unaccompanied by creaky voice (creaky would be typical for tone 1). The same speaker also pronounces various tokens with tone 1 as delayed falling or even with a slight rise in pitch before the fall instead of the direct falling pitch associated with prototypical tone 1.

The complex nature of the Muklom tone system, in which tones consist of a bundle of features rather than one main cue such as pitch, is one factor that allows fuzzy realisation. The hearer has multiple cues to rely on, so there is room for articulation to be sloppy. For example, if phonation type is ‘off’, the hearer can still rely on pitch and duration. Similarly, if pitch deviates from the prototype, the hearer can still rely on phonation and duration. If duration happens to be not prototypical of the tone category, the hearer can still rely on pitch and phonation to decode the signal. The idea that tone consists of a bundle of features in certain languages is not new. For another account of a fuzzy TB tone system, see for example Mazaudon’s detailed description of tone in Tamang (Mazaudon 2014).

Imagine possible tone space as a square like in Figure 1 (similar to plotting vowels into possible vowel space). With a three-way tone system, the small number of tone categories allows greater variation in articulation. However, with an eight-way tone system, the chance of misperception by the hearer increases. In other words, tone categories in a low-number tone system could occupy a bigger territory than tone categories in a high-number tone system.

*Figure 1: Tone targets in imaginary tone space for a three- and eight-way system*



Another factor that can help explain the flexibility in tone realisation in Muklom Tangsa is functional load. Tone languages vary as to the importance of tone for disambiguation of word meaning or word recognition in general. Speakers of certain tone languages rely more on context and less on tone than those of other tone languages (Konnerth 2014:78). We say that in the former type of language, tone has a low functional load (Konnerth 2014:76-83). Across the Tangsa language varieties we find tone systems that I suspect carry low functional load (based on impressions and not yet quantified by experimental data; compare Morey 2014a:640 on Singpho).

While fuzzy tones do not pose a problem to native hearers for perception, the outside researcher must spend more effort to identify the number of categories, their characteristics, and to determine the tones on lexical and grammatical items, a key task when transcribing recorded texts or composing a lexicon.

The second challenge for the researcher in the case of Muklom (besides the first challenge which is sloppy realization) is low metalinguistic awareness of tones. The Muklom speakers that I have been in contact with were not metalinguistically aware of the existence of three tone categories in their language. While speakers were aware that certain word pairs differed in what they referred to as ‘melody’ or ‘tune’, none of the consultants independently suggested three such melodies existed, or that two different words could share the same melody. Once the concept of three ‘tunes’ was explained to consultants, many were able to determine at least some of the tones on new lexemes when asked to do so. Accuracy varied greatly between consultants, and even after repeated practice none of them, including the most experienced consultant, were able to independently determine tones reliable enough for lexicon composition. Muklom is not an isolated case in this. Mazaudon (2014) reports similar challenges in the determination of tone for Tamang, which like Muklom has tones that consist of a bundle of features. The speakers of the TB language Karbi, spoken in Northeast India, also ‘experience a lot of difficulty in identifying the tone category of a morpheme’ (Konnerth 2014:77).

Low metalinguistic awareness combined with fuzzy realisation poses a challenge to the researcher, but low awareness also hinders the language community should they wish to write their language in a standardised orthography that incorporates tone marking. There is currently no standard orthography in use for Muklom. Spelling in written or typed texts, such as song lyrics, text messages, or posts on social media, show considerable inter and intra writer variation. People mostly or even exclusively use Roman script. Tone is usually not expressed in the writings. Although, it seems syllable-final {h} is sometimes used to express sound features that are associated with tone. Note that /h/ as a phoneme only occurs in onset position and is not allowed in coda position in Muklom. The syllable final character {h} that writers use, can express both phonation type, and a phonemic glottal stop. Interestingly, with respect to phonation the character is clearly used to express two opposites, namely both breathy voice and creaky voice. The three examples (7) to (9) show how in non-standardised orthography writers sometimes use final {h} to express breathy phonation (Example 7), creaky phonation (Example 8) and final glottal stop (Example 9). Ideally, a standard orthography will provide a means of distinguishing between these. For clarity, creakiness and breathiness that are associated with tone are marked in the IPA lines in Examples 7-9.

(7) {enai    akhal                          beh?}  
 i<sup>3</sup>    ŋa<sup>1</sup>                          ai<sup>2</sup>                          ʌ-xa<sup>1</sup>                          bɛ<sup>3</sup>?  
 1SG.POSS                                  sister                          VOC                          NMLZ-good                          Q  
 ‘Sister, how are you?’ (social media message)

(8) {thumal                                  hun                          angta<sup>h</sup>}  
 e<sup>2</sup>t<sup>h</sup>i<sup>3</sup>                          ma<sup>1</sup>                          hu<sup>3</sup>                          ʌŋ<sup>2</sup>t-g<sup>1</sup>.  
 DEM                          like                          really                          be-PST-3  
 ‘It really was like that.’ (short story draft)

(9) {thumoh                                  angta}  
 t<sup>h</sup>u<sup>3</sup>-m-ɔ?                          ʌŋ<sup>2</sup>-t-g<sup>1</sup>.  
 arrive-NEG-3                          be-PST-3  
 ‘It did not arrive.’ (social media message)

#### 4 Cooperation tackles fuzziness

In this section, I present examples from my own field work on Muklom to show how tone determination was, eventually, tackled. The strategy that is described may very well prove useful in the description of other fuzzy tone languages.



As mentioned above, fuzzy realisation of tones frustrates the determination of tones on lexical and grammatical items. Before visiting the field site, I found that while trying to determine tone in previously recorded wordlist elicitations by Stephen Morey and Jürgen Schöpf<sup>5</sup>, I would often change my estimation of tone category for a given token. Very few words were confidently assigned to tone categories by me. Tone 1 was the easiest to recognise, but tone 2 and 3 were much harder to detect. It was evident that I needed more exposure to the language to be able to determine tones more easily. After working with consultants for several months, my tone judgement had improved but was not yet satisfactorily reliable yet. I often changed my analysis based on new input from consultants.

Should I then fully rely on consultants' judgement? This turned out not to be an option either because of the following. In the interest of time, a relatively accurate consultant proposed to independently mark tones on a word list. I soon discovered, however, that they lacked the desired reliability, often changing their judgement when asked to repeat the word or compare it to another word from the assigned tone category.

Alone, neither the researcher, nor the consultants could complete the task satisfactorily. The strategy that worked best for Muklom was to cooperate, and we did this in the following manner. I would prepare a wordlist for elicitation based on the recordings of stories and on field notes from spoken conversation that were written down in a notebook. Entry by entry I would ask the consultant to pronounce the word, and request them to identify the tone, using the minimal tonal triplet *k<sup>h</sup>uŋ<sup>1</sup>* 'nose', *k<sup>h</sup>uŋ<sup>2</sup>* 'plate', and *k<sup>h</sup>uŋ<sup>3</sup>* 'small path' as reference words. I would compare their judgement with my own, and if there were any discrepancies between their and my judgement or the judgement of another consultant, I would ask more questions. Based on the replies I would most of the time be able to confirm that my own judgement or another consultant's first judgement was incorrect, or the replies would allow the consultant to readjust their assessment.

Through hours of lexicon elicitation sessions, it became clear that certain phonological contexts were easier, and that certain parts of speech were easier to judge than others. With respect to phonological context, consultants struggled more with syllables containing short vowels /ʌ/ and /ɪ/. It was particularly difficult for consultants to identify tone 2 on these vowels, a tone category that is normally associated with long duration. Another phonological context that proved more difficult, was syllable type (C)V. Consultants sometimes found it difficult to compare an open syllable word to the three reference words which do have a coda, namely *-ŋ*. To bypass this issue, I would instead ask them to compare the (C)V element under discussion to other coda-less words, such as *wa<sup>1</sup>* 'father' for tone 1, and *na<sup>3</sup>* 'ear' for tone 3. The issue with vowels, and also the issue of conflicting judgements, was frequently resolved by repeating the word with slightly exaggerated tone pronunciation. For example, if the word under discussion was *pu* 'snake' and two consultants disagreed, one claiming it carried tone 2, and the other claiming it carried tone 3, then I would repeat the word back to the consultant, asking if it was more correct to say *pu* (with exaggerated rising-falling contour) or *pu* (with exaggerated level contour). However, this example uses a noun, while the most challenging category of words are the verbs, since 1) they sometimes show tone alternation, and 2) person markers carry different tones depending on tense. These properties seem to negatively affect the ability of consultants to correctly identify the tone of a verb root when inflected. The described strategy of pronouncing the words with exaggerated tone and phonation almost always led to a more confident judgement.

## 5 Conclusion

A fuzzy tone system has various implications for descriptive linguistic research. It is likely to take longer to a) figure out whether such a language has tone at all, b) identify the number of categories, c) determine what the relevant cues are, and d) determine the tones on individual lexical and grammatical items. For all languages, but particularly in the case of a fuzzy tone language like Muklom Tangsa, it is necessary that recordings are collected from more than one or two speakers. Also, it is advisable to avoid relying exclusively on analysis by a non-native linguist; we should consider and report on native speaker intuitions with respect to tone categories and tone determination.

<sup>5</sup> "Muncha and Thangseng – CALMSEA word list" from The Language Archive: <https://hdl.handle.net/1839/00-0000-0000-0017-C4EB-9@view>.

**Abbreviations**

1	first person	P	plosive
3	third person	POSS	possessive
C	consonant	PST	past tense
CAUS	causative	Q	question
DEM	demonstrative	RECIP	reciprocal
L	lateral	SG	singular
N	nasal	V	vowel
NMLZ	nominaliser	VOC	vocative

**References**

- Bandyopadhyay, Swapon Kumar. 1989. A Tangsa wordlist. *Linguistics of the Tibeto-Burman Area* 12(2):79-91.
- Barua, Surendra Nath. 1991. *Tribes of the Indo-Burma border. A socio-cultural history of the inhabitants of the Patkai Range*. New Delhi: Mittal Publications.
- Benedict, Paul K. 1972. *Sino Tibetan. A conspectus*. Cambridge: Cambridge University Press.
- Bird, Stephen and Hyman, Larry. (eds.). 2014. How to study a tone language. *Themed series in Language Documentation and Conservation* 8:453-777.
- Blench, Roger, and Post, Mark. 2015. *(De)classifying Arunachal languages. Reconsidering the evidence*. Manuscript. [Accessed through [www.rogerblench.info](http://www.rogerblench.info).]
- Bradley, David. 1997. Tibeto-Burman languages and classification. In Bradley, David (ed), *Papers in Southeast Asian Linguistics No. 14: Tibeto-Burman languages of the Himalayas*:1-72.
- Burling, Robbins. 1983. The Sal languages. *Linguistics of the Tibeto-Burman Area* 7(2):1-31.
- Burling, Robbins. 2003. The Tibeto-Burman languages of Northeastern India. In *The Sino-Tibetan Languages*, eds. By Thurgood, G., and LaPolla, R., 169-191. London: Routledge.
- Bynn Kham Lann. 2015. *Tangshang Naga orthography report*. Unpublished presentation.
- Das Gupta, K. 1980. *The Tangsa language. A synopsis*. Shillong: North East Indian Frontier Agency.
- French, Walter T. 1983. *Northern Naga. A Tibeto-Burman mesolanguage*. Dissertation. City University of New York.
- Konnerth, Linda A. 2014. *A grammar of Karbi*. Dissertation, University of Oregon.
- Mazaudon, Martine. 2014. Studying emergent tone-systems in Nepal: Pitch, phonation and word-tone in Tamang. *Language Documentation and Conservation* 8:587-612.
- Morey, Stephen. 2011. Tangsa agreement markers. In *North East Indian Linguistics*, ed. by Hyslop, G., Morey, S., and Post, M. 3:76-10.
- Morey, Stephen. 2014a. Studying tones in North East India. Tai, Singpho and Tangsa. *Language Documentation and Conservation* 8:637-671.
- Morey, Stephen. 2014b. Ahom and Tangsa: Case studies of language maintenance and loss in North East India. *Language Documentation & Conservation, Special Publication* 7:46-77.
- Morey, Stephen. 2015. The internal diversity of Tangsa: vocabulary and morphosyntax. In *Language and culture in Northeast India and beyond. In Honor of Robbins Burling*, ed. by Post, Mark W., Morey, Stephen D., and DeLancey, Scott. Canberra: Asia-Pacific Linguistics.
- Morey, Stephen. 2017. Tangsa. To appear in *The Sino-Tibetan Languages*, ed. by Graham Thurgood and Randy LaPolla. London and New York: Routledge (2nd Edition).
- Ngemu, T. 1977. *Moklum language guide*. Shillong: Government of Arunachal Pradesh.
- Shafer, Robert. 1953. Classification of the northern-most Naga languages. *Journal of the Bihar Research Society* 39(3):225-264.
- Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word: Journal of the Linguistic Circle of New York* 11:94-111.

- Simons, Gary F. and Fennig, Charles D. (eds). 2017. *Ethnologue: Languages of the world*. Twentieth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Statezni, Nathan. 2013. *Fifty-five dialects and growing. Literacy and comprehension of vernacular literature among the Tangshang Naga in Myanmar*. MA Thesis, Graduate Institute of Applied Linguistics.
- Weidert, Alfons. 1979. The Sino-Tibetan tonogenetic laryngeal reconstruction theory. *Linguistics of the Tibeto-Burman Area* 5(1):49-127.
- Weidert, Alfons. 1987. *Tibeto-Burman tonology*. Amsterdam: John Benjamins.

# THE KACHIN AS PARTICIPANTS OF AN ETHNO-LINGUISTIC AREA?

André Müller  
University of Zurich  
andre.mueller3@uzh.ch

## Abstract

The Kachin are a multilingual ethnic group, inhabiting the northern part of Myanmar and parts of Yunnan Province in China, where they are subsumed under the name Jingpo. In this paper, I argue that they do not only share many cultural peculiarities due to intense contact between the subgroups, but also a number of linguistic features, indicating that the Kachin/Jingpo form an ‘ethno-linguistic area’. Languages within this area participate in it in varying degrees, while related languages outside of the Kachin umbrella or unrelated languages in the same geographical context do not share these features. Thus, long-standing cultural and linguistic contact led to strong convergence between the ethnic subgroups and their languages.<sup>1</sup>

**Keywords:** Kachin, Sino-Tibetan languages, areal linguistics, Jinghpaw, language contact  
**ISO 639-3 codes:** acn, atb, bxd, ckh, duu, kac, kht, lhu, lis, lsi, mhx, mya, nun, raw, sgp, shn, zkd

## 1 Introduction

The north of Myanmar is home to several ethnic groups that are generally referred to as ‘Kachin’ in English and in the national language Burmese. From a linguistic perspective, the Kachin can be described as a multilingual group, speaking several languages from different branches of the Tibeto-Burman family, with Jinghpaw<sup>2</sup> being the *lingua franca* in the area. Historically, the speakers of these languages have long been living as a loose federation in the same region: the mountainous highlands of what is now the Kachin State of northern Myanmar, a part of ‘Zomia’, the vast upland massif of Mainland Southeast Asia (van Schendel 2002; Scott 2009). Centuries of intense contact between these ethnic subgroups have led to a strong sociocultural convergence, and the rise of an overarching social space with the exonym ‘Kachin’.

There is an ongoing debate about the Kachin as an ethnic group or ‘transethnic category’ (Robinne & Sadan 2007:300) since at least the mid 20<sup>th</sup> century with Leach’s (1954) seminal work on the political and societal system of the highland cultures of the Kachin and the Shan and their interactions. Most of this literature (*inter alia* Leach 1954, Robinne & Sadan 2007, Gilhodes 1996, Chit Hlaing 2007, Maran 2007, Sadan 2013) is concerned with the cultural or anthropological similarity of the Kachin, and relatively little has so far been published on the linguistic evidence for this long-lasting contact scenario, though cases of intensive lexical borrowing have already been mentioned by Enriquez (1933:46-47). The present study strives to present more evidence for this contact situation, which is the basis of my claim that the Kachin social complex deserves to be named an ‘ethno-linguistic area’, a social space in which the participating languages show certain linguistic features that were not inherited from a common source, but that have spread as a result of diffusion under long and intense language contact. The ethno-linguistic area is thus not based on geographic, but rather cultural and social boundaries delimiting the distribution of the areal features. In this specific case, the boundaries of this area are those between the ethnic umbrella of the Kachin to other groups such as the Burmese, Shan, Khamti,

---

<sup>1</sup> The project is sponsored mainly by the Swiss National Foundation (Grant No. 100012\_150136).

<sup>2</sup> In this paper, I use the spelling *Jingpo* to refer to the ethnic group in China, while I use *Jinghpaw* for the name of the Jinghpaw language in both countries. *Kachin* stands for the ethnic group in Myanmar. The term *Kachin languages* is here used for the languages pertaining to the Kachin/Jingpo people in the wider sense: Jinghpaw, Lacid, Lashi, Lisu, Ngochang, Rawang, and Zaiwa.

etc. The expectation is that the languages considered as part of the Kachin share certain areal features that non-Kachin contact languages within the geographical area, like Burmese, Shan, Khamti, and Kadu, but also closely related non-Kachin languages such as Turung or Achang outside of it lack. The present study proposes a number of such features (lexical, phonological, and morphosyntactic) that can in this case be used as evidence for the suggested ethno-linguistic area.

The data used in this study is taken from published literature, but also draws heavily from the fieldwork I have conducted over 5 field trips to Myanmar and Yunnan between 2014 and 2017. The work presented here is still in progress and further fieldwork and research is necessary to complete the picture.

## 2 Ethno-linguistic situation of the Kachin area

The people who are commonly called ‘Kachin’ and who understand themselves as belonging to this ethnic unit historically inhabit the north of the Greater Burma Zone, a region describing the modern country of Myanmar including adjacent regions. Many Kachin also live in the Northern Shan State or have migrated to other places within the country. They are one of the eight larger ethnic groups officially recognized by the government of Myanmar. However, they are not a homogenous ethnic group, but rather a loose federation of several subgroups, each with their own ethnic identity in addition to the umbrella category ‘Kachin’ — an exonym used mainly in English and Burmese, but rarely as an autonym in the respective languages themselves. Traditionally, the main subgroups of the Kachin are the Jinghpaw, Zaiwa, Lhaovo, Lacid, and Lisu, with the Rawang and Ngochang<sup>3</sup> being included but often not mentioned explicitly. Each of these groups speak a different language, usually in addition to the *lingua francae* Jinghpaw and Burmese, showing the societal as well as individual multilingualism in this group (Deumert 2011:272).

The question whether the Kachin themselves have always formed a larger unit, ethnically or socio-culturally, or whether this had only come about in recent centuries, is currently very difficult to answer, as written sources before the time of British colonialization mentioning the Kachin or their subgroups are sparse. Jinghpaw and most of the other Kachin languages were not written before 19<sup>th</sup> century Baptist missionaries developed a writing system for Jinghpaw and later also the other languages. Also Burmese sources mentioning the Kachin people or the people inhabiting the Kachin mountains do not go very far back in time, as the area was generally difficult to access. The etymology of the term ‘Kachin’ itself is unclear and much has been speculated on it; one possible origin could be the Jinghpaw toponym *ga hkyeng* ‘red earth’ describing the Jinghpaw’s homeland between the Nmai and Mali rivers (Wang 1997:4, Sadan 2007:41-45). The term, Leach (1954:41-44) explains, was first spelled *Kakhyen* in early British sources and derived from the Burmese term <*kakhyān*> /kət<sup>h</sup>in/ and denoted a group of multilingual hill tribes near Bhamo and Hsenwi at the modern border between Kachin and Shan State. Since 1886, after the British established their Christian missions and recruited many men from different ethnic groups, the term ‘Kachin’ was formalized to refer to all non-Shan and non-Burmese people in the Kachin hills; linguistic and cultural diversity was downplayed by colonial officers and later also by Kachin nationalist politic leaders who sought to unify the different groups (Sadan 2007a:47f, Sadan 2007b:68).

But not all of the groups mentioned above identify equally as Kachin. Among the Lisu, only a small group living in close vicinity with the Kachin actually see themselves as members of this group (Bradley 1996:749, Müller 2016:34). Leach (1954:45) describes them as ‘marginal to the Kachin Hills’. The Lisu also do not participate in the *mayu-dama* system that regulates the exogamous intermarriage between the different Kachin clans and forms one of the cornerstones of the Kachin society (cf. e.g. Leach 1954, Robinne 2007:290-295, Sadan 2007a:60).

Other languages spoken in and around the Kachin State not participating in the Kachin social complex are Burmese, Shan, Khamti, and the Kadu language further in the south, which is, however,

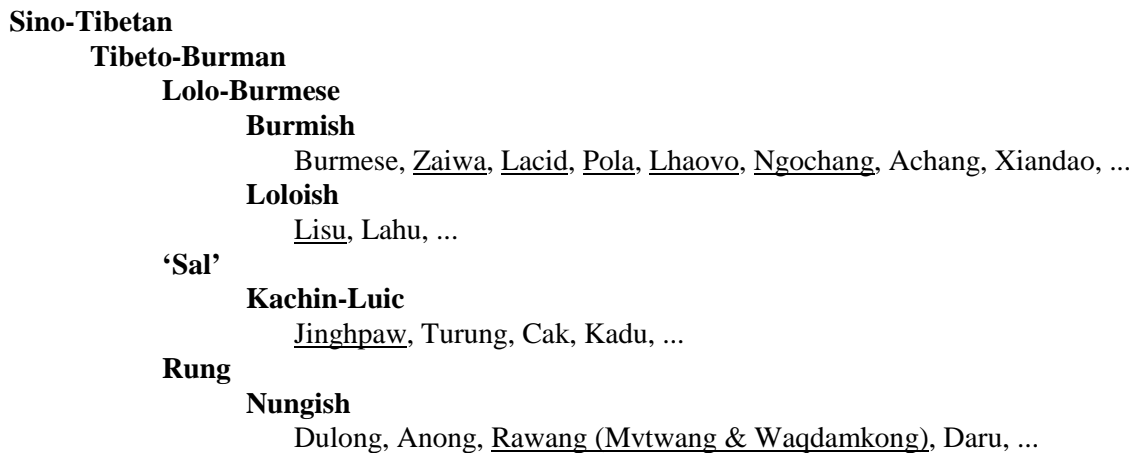
---

<sup>3</sup> The names of these subgroups and their languages vary. For the Jinghpaw there are multiple spellings such as Jingpo, Jinghpo, Chingpaw, and their language is often simply called ‘Kachin’. The Lhaovo are also known as Maru or Langsu, the Zaiwa as Azi or Atsi, the Lacid as Lashi, Leqi, or Lacheik. The Ngochang are commonly known as Achang in China.

one of the closest relatives of the Jinghpaw language (Sangdong 2012:29). Close relatives and varieties of the Kachin languages inside Myanmar are also found outside the country, across the border in China (e.g. Anong/Nung, Dulong/ Drung, Pela/Bola) and India (e.g. the Turung variety of Singpho described in Morey 2010). The Kachin varieties of Jinghpaw, Lhaovo, Lacid, and Zaiwa that are spoken in the Chinese province of Yunnan, have not been Christianized, did not fall under British rule, and are subsumed under the ethnic label *Jǐngpō zú* (景颇族). They exclude the speakers of Achang/Ngochang, as well as those of Anong/Nung and Dulong/Drung, which all belong to other ethnic groups (Mullaney 2011). However, Wang (1997:35) argues that the ‘Jingpo’ have only arrived in the Dehong region of Yunnan in the 18<sup>th</sup> century, before the ethnic labels were imprinted on the whole groups. It thus makes sense to treat China’s ‘Jingpo’ and Myanmar’s ‘Kachin’ as one and the same group for the present study.

Despite the Kachin narrative of a common origin of all the Kachin tribes and their subsequent migration from the Tibetan highlands into today’s Kachin State (cf. Enriquez 1933:24-28; Leach 1954:230; Dai 2012:8-9), the Kachin languages do not belong to a single branch of the Sino-Tibetan family. Thurgood’s (2003) approach of a subgrouping of the Sino-Tibetan tree should suffice here as one example of many attempts to classify the languages of this family; naming and spelling have been adapted:

**Figure 1:** *Affiliation of the Kachin languages, based on Thurgood (2003)*



Only relevant branches are shown. The underlined languages represent the ones treated as part of Kachin/Jingpo. It is evident from the classification that they belong to at least four different branches, each with closely related languages that are not counted as Kachin, such as Burmese, Lahu, Kadu, or Dulong.

### 3 (Ethno-)linguistic areas

According to Thomason (2001:99), a linguistic area or Sprachbund is ‘a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor’. In this definition, the notion of a geographical region is a bit problematic, as already noted by Muysken (2008). In the modern study of linguistic areas, the boundaries of such areas should not be understood as purely geographic, but as defined by social ties. A useful notion is the concept of the ‘social space’: ethnic groups like the Burmese, Shan, and Khamti, living in the same geographic region, are excluded, while Kachin communities living outside of the geographical region, are still included.

For a community such as the Kachin, for which exists strong convergence not only of linguistic patterns, but also of socio-cultural properties, and for which the self-proclaimed ethnic affiliation is the base of the social space, I use the term ‘ethno-linguistic area’.<sup>4</sup> These traits have already been described

<sup>4</sup> The term is sometimes used as a vague notion but has never been clearly defined before, to my knowledge.

in detail and from various points of view in the literature (Wehrli 1904; Carrapiett 1929; Leach 1954; Gilhodes 1996; Wang 1997; Maran 2007; Robinne 2007; Sadan 2007a; Sadan 2013) and include marriage and burial rituals, various festivals, an overarching clan system with exogamous marriage restrictions which cross-cuts the language groups, common traditions and cultural values, Christianity as the prevailing religion,<sup>5</sup> a rather egalitarian (non-hierarchical) societal structure, the same kinship system<sup>6</sup> along with heavy borrowing from Jinghpaw (Maran 2007:56-58, own fieldwork notes), and common origin and descent myths. In this light, an ethno-linguistic area can be defined as a social space in which several ethnic groups or subgroups share linguistic features as well as cultural traits as a direct result of long-lasting contact as opposed to coincidence, inheritance from a common ancestor, or common tendencies of language or culture.

If the Kachin/Jingpo social complex is indeed an example of a smaller ethno-linguistic area, embedded in the larger linguistic macro-area of Mainland Southeast Asia (described in Enfield 2001; Enfield 2005; Jenny 2015), what shared linguistic features do we expect, and what is expected of their spread within and outside of this area? The most salient effects of language contact are of course loanwords, but as Thomason (2001:100) points out, loanwords occur even in casual contact scenarios and are therefore much less interesting than structural interference, that is, the distribution of grammatical features. We thus expect non-inherited, non-random inter-language similarities not only in the lexicon, but especially in the realms of morphology, syntax, but also phonology, which is often the earliest to be affected by convergence processes (cf. Aikhenvald & Dixon 2001:17). We furthermore expect that there is a group of ‘core’ Kachin languages with more shared features, and several more ‘peripheral’ languages with less (Muysken 2008, cf. also the *Standard Average European* linguistic area elaborated in Haspelmath 2001 and Heine & Kuteva 2006). Additionally, non-Kachin languages spoken in close geographic vicinity of the area – such as Burmese, Kadu, Shan, Lahu – should not share these features, and neither should geographically more remote languages that are closely related to the Kachin languages in question, like Singpho or Xiandao.

#### 4 Linguistic evidence

The following sections contain the linguistic evidence pointing to the convergence as the outcome of the contact situation of the Kachin/Jingpo languages.

##### 4.1 Lexicon

As a prestige language and the *lingua franca* of the area, Jinghpaw unsurprisingly is the source language for many loanwords in most of the Kachin languages. This ‘matter replication’ (cf. Matras & Sakel 2007) is especially noticeable in the Northern Burmish languages of China and Myanmar, Zaiwa, Lhaovo, Lacid, and Ngochang, and also in Rawang (evidenced by the many Jinghpaw loanwords marked in LaPolla & Sangdong 2015). Kurabe (2016:458) remarks that Zaiwa is the language most influenced by Jinghpaw, as even its kinship terminology is largely adopted from Jinghpaw (this is also true for Lhaovo and Rawang, cf. Müller 2016:34). Many loanwords from Burmese and Shan have also entered these languages via Jinghpaw. The long-lasting contact situation is evidenced further by the abundance of loans of function words from Jinghpaw, especially in Zaiwa, cf. Zaiwa *ʔmá*<sup>7</sup> ‘also’ < Jinghpaw *má*, Lacid *ləma* < Jinghpaw *ləma* ‘if’, or the word ‘very’, borrowed from Jinghpaw *gràj* into Zaiwa *gʲaj*, Lhaovo *gʲàj*, Lacid *gʲâj*, and even Rawang *gàj* ~ *gráj*.

Lisu, as one of the culturally more peripheral, hardly shows any loanwords from Jinghpaw, suggesting that the contact there must have been more recent and less intense. There are hardly any direct

<sup>5</sup> This applies more to the Kachin in Myanmar. The Jingpo in China rather adhere to traditional animist belief (cf. Wang 1997:169-219).

<sup>6</sup> There is further evidence for the peripheral Kachin languages adopting Jinghpaw’s kinship system: only the middle and southern groups of Rawang have a system very close to that of Jinghpaw (Omaha type), while northern Rawang groups have the Iroquois type (Sarep 1996:95; Straub 2015).

<sup>7</sup> Words are transcribed phonetically using the IPA in this work, regardless of the orthography or transcription used in the sources.

Jinghpaw loans in Northern or Southern Lisu (Bradley 1994 and Bradley 2006, respectively), or in the Lisu variety spoken near Myitkyina (own fieldwork).<sup>8</sup> Jinghpaw loanwords are virtually absent in Lahu, Burmese, Kadu, Shan, and Khamti.

Figure 2: Demonstrative morphemes with altitude distinction

Family/ Branch	Language	Affil.	‘that up there’	‘that over there’	‘that down there’	Source
Sal	Jinghpaw	CK	tʰɔ̄-	wɔ̄-	lé-	Kurabe (2016)
	Turung	—	ho daj			Morey (2010)
	Cak	—	húŋmá			Huziwara (2016)
	Kadu	—	ʔúŋ			Sangdong (2012)
Burmish	Zaiwa	CK	hú	éê	ʔmɔ̄	Lustig (2010)
	Lhaovo	CK	xù	tʰò	ʔmɔ̄	Dai (2005)
	Lacid	CK	xu	tʰə	ʔmɔ	Dai & Li (2007)
	Pola	CK	xú	tʰɔ̄j	ʔmá	Dai et al. (2007)
	Ngochang	PK	how	tʰji	ʔmó	Inglis et al. (2005)
	Achang	—	xáw			Dai (2010)
	Xiandao	—	ʂùzɔ̄	tʰété	ŋòŋó	Wang (2005)
	Burmese	—	ho			Jenny & Hnin Tun (2016)
Loloish	Lisu	PK	nɸ	go	ɬə	Bradley (2003)
	Lahu Shi	—	nā	ʔutɛ̄i-	mā-	Straub (2014)
Nungish	Rawang	PK	kʰū			Straub (2014)
	Waqdamkong	PK	lɔ̄	kʰū	ʔu	Barnard (1934)
	Daru	PK	kʰū / lū			Straub (2016)
	Anong	—	ná	ʔó / kʰú	ʔí	Sun & Liu (2009)
Tai-Kadai	Shan	—	pún			own fieldwork
	Khamti	—	pǔn (nèj)			own fieldwork

An interesting phenomenon concerning lexical semantics is found in the demonstratives of these languages, which exhibit an altitude distinction in their remote sets of demonstratives, encoding the relative height of the referent in relation to the speaker in monomorphemic stems. These can usually be translated as ‘that up there’, ‘that over there’, and ‘that down there’. This particular set of demonstratives is found in all of the Northern Burmish languages, including Pola on the Chinese side, in the Waqdamkong dialect of Rawang (Barnard 1934), as well as in Anong (Sun & Liu 2009:73). However, it is not found in Daru or the Mvtwang varieties of Rawang. Lisu (Straub 2014) and Lahu (Matisoff 1973:51-52) both have a similar altitude distinction. Lianghe Achang (Dai 2010:348), spoken in China, does not have this distinction, while the related Xiandao (Wang 2005:71), also from China, does. For Ngochang, my consultants used only circumlocutions with locative expressions, like tʰoʔ-mà-dá ʔjen (top-LOC-LNK house) ‘the house on top’, and used the general distal demonstrative háw. The forms how ‘that up there’, tʰji ‘that over there’, and ʔmó ‘that down there’ are listed in a provisional dictionary (Inglis et al. 2005) only, perhaps indicating that the altitude distinction is not made by all speakers anymore, something also found in younger and more urban speakers of Jinghpaw. Although all varieties of Jinghpaw spoken in Myanmar and China have the distinction, it is not mentioned for the Turung variety of Singpho from Assam (Morey 2010). The distinction is equally absent from Burmese, Kadu, Cak, Shan, and Khamti. Figure 2 compares the phonetic forms of these demonstratives. It is worth noting that the forms are not cognate between all the languages showing the distinction, which indicates that the words were not merely borrowed, but that the distinctions were made using language-internal

<sup>8</sup> There are a few lexemes shared with Jinghpaw, such as Northern Lisu pàèʔ / Southern Lisu pàj ~ pøj ‘festival’ (cf. Jinghpaw páj ‘festival, event’), but they are likely borrowed from a common source, like Burmese pwe ‘festival, event’, perhaps via Shan páj. Others are shared inheritances from Proto-Tibeto-Burman.



material. The column “Affiliation” shows which languages belong to the Core Kachin group (CK), to the Peripheral Kachin group (PK), and which are not part of the Kachin area (—).

#### 4.2 Phonology

Phonological data also present evidence of conversion especially between the Northern Burmish languages and Jinghpaw (both core Kachin). Yabu (1988) points out that Zaiwa’s phonological system is closer to that of Jinghpaw, despite its much closer relationship to Lhaovo and Lacid. The other Burmish languages Lhaovo, Lacid, Pola, and even Burmese itself have diphthongs occurring in closed syllables, followed by a consonant, e.g. *-auk*, *-auŋ*, *-auj* for Lhaovo (Sawada 2004); *-auʔ*, *-eit*, *-ein*, *-ouʔ*, *-ouk*, *-ouŋ*, *-uəŋ*, *-uəŋk* for Lacid (Wannemacher 2011); *-auʔ*, *-auŋ* for Pola (Dai et al. 2007). In both Jinghpaw and Zaiwa, diphthongs cannot occur before final consonants and are better analyzed as a combination of vowel and off-glide. Additionally, both Jinghpaw and Zaiwa have a very regular set of rhymes almost without gaps (cf. Kurabe 2016:459), while Lhaovo, Lacid, Pola, and also written Burmese have many more restrictions. Yabu (1988) explains this as the result of intensive language contact between Jinghpaw and Zaiwa. Not mentioned by Kurabe is Ngochang, which allows four phonemic diphthongs according to Inglis & Inglis’ (2003) analysis, but whose diphthong *-ai-* always occurs in the rhymes *-aiʔ* and *-ain*; thus, Ngochang behaves similarly to the other Burmish languages.

As an argument for an areal phenomenon among the Kachin/Jingpo, however, the lack of pre-coda diphthongs in Zaiwa and Jinghpaw is relatively weak, as the phenomenon is only remarkable in Zaiwa, which behaves differently from its sister languages. Generally speaking, pre-coda diphthongs with high vowels as their second segment are rare in the languages of Southeast Asia and hardly occur outside of the Burmish languages. Among the languages looked at for this study, only Cak and Kadu have a diphthong *-ai-* that can occur in front of *-k* (in Kadu), *-ʔ* (in Cak), or *-ŋ* (in both) in a restricted set of words (Huziwara 2016:483-486, and Sangdong 2012:72-73, respectively).

#### 4.3 Morphosyntax

Three morphosyntactic features are relevant for the discussion of areal features in the Kachin/Jingpo ethno-linguistic area, and all of them concern variability.

The most striking one is perhaps the variable positioning of demonstratives relative to the head noun, a phenomenon that is otherwise very uncommon outside the area and rare from a global perspective (Dryer 2013<sup>9</sup>). In Jinghpaw from both China and Myanmar it is possible and common to use either DemN or NDem word order without any change in meaning, allowing both *ndaj ləʔnaw* (DEM cat) and *ləʔnaw ndaj* (cat DEM) to express ‘that cat’. For Jinghpaw, this free distribution has been mentioned before in the literature numerous times (e.g. Dai & Xu 1992:368; Cheung 2007:34-37; Kurabe 2016:167). Cheung (2007:37) shows that the variability in word order between demonstrative and noun ‘can neither be due to topicalization of the demonstrative or the head noun [...] nor can it be attributed to movement of either element outside of the noun phrase’. Most of the core Kachin languages, including Pola, have a variable word order between demonstrative and noun, as I confirmed during my fieldwork.<sup>10</sup> My Ngochang consultants did only accept and produce a DemN word order, and judged \*NDEM as ungrammatical; a corpus analysis of the Ngochang New Testament also shows consistent DemN patterns. Rawang has DemN word order, but the demonstrative can occur after the noun if at the same time it precedes a classifier. I interpret this NDEMClf as a structure similar to DemN, as in both cases the demonstrative acts as a determiner to a nominal head.<sup>11</sup> Most languages outside of the Kachin area have a fixed word order, so we find the order NDEM in Lisu, Shan, and Khamti, and DemN in Daru, Anong, Lianghe Achang, Burmese, Cak, and Kadu. Interestingly, Turung has a

<sup>9</sup> I do not agree with some of the feature values presented in the WALS data, as several of the Kachin languages are glossed as having just one particular order instead of two.

<sup>10</sup> A reviewer suggested that this could be due to Tai-Kadai influence, as Jinghpaw and the Northern Burmish languages have undergone heavy influence from Shan. However, Singpho was and still is in much closer contact with surrounding Tai languages, yet it does not allow \*NDEM word order.

<sup>11</sup> Unlike some generative theories, I take the demonstrative to be a dependent instead of the clausal head of a DP.

consistent DemN word order as well (Morey 2010:306), in contrast to its relatives in Yunnan and the Kachin State. Xiandao is a special case with NDem and DemClf word order (Wang 2005:71) and requires further analysis. There is one outlier in the sample: Lahu, although not a member of the Kachin/Jingpo ethnic group(s) by any stretch, also allows flexible order in this case (Matisoff 1973:52-53).

The second case of variability can be found in the verbal complex of the languages. Verb chains or multi-verb predicates, including serial verb constructions and auxiliary verb constructions, are ubiquitous across and beyond Southeast Asia. Usually, auxiliary verbs come in a fixed position in relation to the main verb in the languages of Southeast Asia, depending on their semantics (Lord 1993; Enfield 2003; Anderson 2006). In Lao, a verb with the lexical meaning of ‘get’ or ‘acquire’ can serve as a secondary verb expressing possibility (e.g. ‘can go’) when used after the main verb, or attainment (e.g. ‘manage to go’) when used before it; relative position and verbal semantics are strongly correlated in this case (Enfield 2001). This is also true for Shan. In the core Kachin languages, however, a group of secondary verbs can appear preverbally or postverbally without any semantic difference, as shown in examples (1a) and (1b).

- (1a) *εi*        *daj*    *mətî*                    *εá*    **lù**    *ʔaj*  
 3SG        MEDL mushroom        eat    **get**    DECL  
 ‘He **can** eat that mushroom.’
- (1b) *εi*        *daj*    *mətî*                    **lù**    *εá*    *ʔaj*  
 3SG        MEDL mushroom        **get**    eat    DECL  
 ‘He **can** eat that mushroom.’

The order appears to be in free variation, even within the same speaker in the same discourse. While in Jinghpaw there are eight such flexible verbs (Kurabe 2016:272-274, 311-319), in Zaiwa, Lhaovo, Lacid, and Ngochang, only two have the same flexibility: ‘know’ and ‘want’.<sup>12</sup> This interesting and still understudied phenomenon is neither found in Singpho, which only has postverbal secondary verbs, nor in any of the other languages in the scope of this study. In Pola, *náw* ‘want’ is strictly postverbal, but Dai et al. (2007:120) mention *kàn* ‘to volunteer’ as being able to precede and follow the main verb, expressing ‘casually, arbitrarily, whatever’ and ‘for fun’, respectively, which does imply a semantic difference. It appears to be restricted to the core languages of the area.

The third variability occurs with singular possessive personal pronouns in first, second, and often also third person. Again, Jinghpaw serves as an example. Jinghpaw has two sets of pronouns for the singular pronouns: an unmarked set *ηaj* ‘1SG’, *naη* ‘2SG’, *εi* ‘3SG’, and a possessive set *ηéʔ* ‘1SG.POSS’, *náʔ* ‘2SG.POSS’, and *εiʔ* ‘3SG.POSS’. To express possession, several strategies are possible with varying degree of grammaticality, shown in examples (2) to (7), partly involving an intervening genitive marker *ʔàʔ* or the more colloquial ablative marker *ná*. Examples (3) and (7) were partly rejected by my informants, perhaps on the grounds that *ηéʔ* ‘1SG.POSS’ and *ʔàʔ* ‘GEN’ are both rather formal, while *ná* ‘ABL’ is more colloquial and preferred by younger speakers. A bare unmarked personal pronoun as in example (2) is ungrammatical as a productive marker of possession and only occurs in a few fixed compounds together with kinship terms, e.g. *ηaj-nù* ‘my mother’.

<sup>12</sup> Note that *know* is one of the flexible verbs in both Jinghpaw and the Northern Burmish languages, but *want* is only flexible in Northern Burmish languages. Also note that these verbs are not necessarily cognates in the Kachin languages (e.g. Lhaovo *bà* vs. Jinghpaw *tεε* ‘know’, or Ngochang *me* vs. Lhaovo *nùk* ‘want’).

(2)	*	<i>ηaj</i>		<i>ńtâ</i>		(5)		<i>ńéʔ</i>		<i>ńtâ</i>
		1SG		house				1SG.POSS		house
(3)	ʔ	<i>ηaj</i>	<i>ʔàʔ</i>	<i>ńtâ</i>		(6)		<i>ńéʔ</i>	<i>ʔàʔ</i>	<i>ńtâ</i>
		1SG	GEN	house				1SG.POSS	GEN	house
(4)		<i>ηaj</i>	<i>ná</i>	<i>ńtâ</i>		(7)	ʔ	<i>ńéʔ</i>	<i>ná</i>	<i>ńtâ</i>
		1SG	ABL	house				1SG.POSS	ABL	house

This flexibility is (in part) mentioned in Dai (2012:69), Kurabe (2016), and existed already a century ago (cf. Hanson 1896:40, Hertz 1902:9) – so this is not a new phenomenon. A similar variability exists in other Kachin/Jingpo languages in a similar form, such as Zaiwa (Lustig 2010:227-230; Wannemacher 2017), Lhaovo (Dai 2005:38; own fieldwork confirms this for Lhaovo in Myanmar), Lacid (Dai & Li 2007:78-79; Luk 2017:41), Pola (Dai et al. 2007:62-63). Outside of the Kachin/Jingpo ethno-linguistic area, this variability is only found in Lianghe Achang, spoken in China (Dai 2010:347), but not in its relative Xiandao (Wang 2005:90). It seems to be absent in Dulong (LaPolla 2017), Anong, Lisu, and Burmese. Also, the closer and less close relatives of Jinghpaw – Turung, Cak, and Kadu – do not show this variability in pronominal possession. Shan and Khamti (Inglis 2007:16-17) have an optional overt marker (*k<sup>h</sup>ʔη* and *k<sup>h</sup>ʔη*, respectively) between noun and pronoun.

## 5 Conclusion

As I have tried to demonstrate in this paper, the ethnic group referred to as Kachin or Jingpo with its subgroups forms a large social complex in the region of Kachin State and Northern Shan State in Northern Myanmar and Yunnan in Southern China. The people of this group see themselves and each other as belonging to a greater whole, a ‘super-ethnic category’ (Chit Hlaing 2007:116), even though some under this umbrella do not see themselves as such and reject the Kachin dominance (Sadan 2007a:61). A common history and long-standing contact, but also more recent ambitions from within and outside to unify them, have shaped a common cultural frame, shared traditions, and a sense of community in those who think of themselves as Kachin. This unity has had influence on the linguistic makeup of the languages spoken by the Kachin, I claim. The evidence leading to this conclusion are summarized in figure 3, with the Kachin features marked in a darker gray, the non-Kachin features in a lighter shade of gray. Missing (‘?’) and non-applicable data (‘n/a’) is colored in very bright gray.

The initial expectations were fulfilled: indeed, we find evidence of linguistic convergence in all parts of the language. In the lexica of the languages involved we find lots of shared vocabulary, mostly spread from or through the *lingua franca* Jinghpaw. We also find shared semantic distinctions in the demonstratives which mark differences in relative altitude; here it is unclear where these distinctions spread from and if they are parallel innovations or shared retentions due to contact. In the phonology, there are tendencies of convergence towards Jinghpaw. The most striking convergent patterns are found in the morphosyntax of the languages: Variability in word order: demonstratives can appear pronominally or postnominally. Variability in grammatical marking: possession can be expressed with possessive personal pronouns with or without a dependency marker, or with the nominative pronoun followed by such a marker. Most of these features are shared among the languages most typically associated with the Kachin (Jinghpaw and the Northern Burmish languages), and to a lesser degree also with the languages less strictly associated with the Kachin (like Rawang and Lisu). This is where the socio-culturally based notions of ‘core Kachin’ and ‘peripheral Kachin’ are reflected in the linguistic structure. Furthermore, closely related languages clearly outside of the area, or non-Kachin languages in the same geographic region like Burmese, Shan, Khamti, or Lahu, do not usually share these features (with some exceptions that require further studies).

Figure 3: Overview of the areal features among the languages in the study

Family/ Branch	Language	Affil.	Jinghpaw loanwords	Altitude Dem	*-aiC, *-auC	Order of Dem & N	Flexible Aux.	Flexible Poss.
Sal	Jinghpaw	CK	n/a	yes	absent	both	yes	yes
	Turung	—	n/a	no	n/a	DemN	no	no
	Cak	—	none	no	present	DemN	no	no
	Kadu	—	none	no	present	DemN	no	no
Burmish	Zaiwa	CK	many	yes	absent	both	yes	yes
	Lhaovo	CK	many	yes	present	both	yes	yes
	Lacid	CK	many	yes	present	both	yes	yes
	Pola	CK	many	yes	present	both	yes?	yes
	Ngochang	PK	many	yes	partly	DemN	yes	yes
	Achang	—	none	no	partly	DemN	?	yes
	Xiandao	—	some	yes	n/a	NDem?	?	no
Burmese	—	none	no	present	DemN	no	no	
Loloish	Lisu	PK	none	yes	n/a	NDem	no	no
	Lahu	—	none	yes	n/a	both?	?	no
Nungish	Rawang	PK	many	no	n/a	DemN	no	no
	Waqdamkong	PK	many	yes	n/a	DemN	?	?
	Daru	PK	?	no	n/a	DemN	?	?
	Anong	—	none	yes	n/a	DemN	?	no
Tai- Kadai	Shan	—	none	no	n/a	NDem	no	no
	Khamti	—	none	no	n/a	NDem	?	no

Looking at the evidence, it seems useful to speak of an ethno-linguistic area with Jinghpaw, Zaiwa, Lhaovo, Lacid and arguably also Pola and Ngochang as a core, and Rawang and Lisu at the periphery. Yet, several data points and relevant languages are still missing to complete the picture. I am positive that further research will still reveal other shared features and also determine the inclusion and further affiliation of some of the speech varieties to the core or peripheral languages, as well as an estimated time-depth for the contact. More data from Achang and Rawang varieties, the aberrant Gauri dialect of Jinghpaw, or the moribund Northern Burmish language Hpon is needed. This kind of research is not only necessary to further our understanding of the history and interaction of the people in Southeast Asia, but is also important for contact linguistics, areal linguistics in general and in mainland Southeast Asia in particular.

## References

- Aikhenvald, Aleksandra & Dixon, R.M.W. 2001. Introduction. In *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, ed. by Aleksandra Aikhenvald & R.M.W. Dixon, 1-26. Oxford: Oxford University Press.
- Anderson, Gregory D.S. 2006. *Auxiliary Verb Constructions*. Oxford/New York: Oxford University Press.
- Barnard, J.T.O. 1934. *A handbook of the Rāwang dialect of the Nung language*. Rangoon: Superintendent of Gov't Printing and Stationery.
- Bradley, David. 1994. *A Dictionary of the Northern Dialect of Lisu (China and Southeast Asia)*. Pacific Linguistics C-126. Canberra: Australian National University.
- Bradley, David. 1996. Kachin. In *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas, Volume 1: Maps*, ed. by Stephen Wurm, 749-751. Berlin/New York: Mouton de Gruyter.

- Bradley, David. 2003. Lisu. In *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla, 222-235. London: Routledge.
- Bradley, David. 2006. *Southern Lisu Dictionary*. Berkeley: University of California Press.
- Carrapiett, W.J.S. 1929. *The Kachin Tribes of Burma*. Rangoon: Superintendent of Gov't Printing and Stationery.
- Cheung Candice Chi-Hang. 2007. On the Noun Phrase Structure of Jingpo. *USC Working Papers in Linguistics* 3:32-56.
- Chit Hlaing, F.K.L. 2007. Some Remarks upon Ethnicity Theory and Southeast Asia, with Special Reference to the Kayah and the Kachin. In *Exploring Ethnic Diversity in Burma*, ed. by Mikael Gravers, 107-122. Copenhagen: Nordic Institute of Asian Studies.
- Dai Qingxia. 2005. 浪速语研究 [A Study of the Langsu Language]. Beijing: Publishing House of Minority Languages.
- Dai Qingxia. 2010. 梁河阿昌语概况 [A Brief Introduction of the Lianghe Achang Language]. In: 藏缅语系语言研究 (5) [Research on Tibeto-Burman Languages 5], 336-383. Kunming: Yunnan Nationality Publishing House.
- Dai Qingxia. 2012. 景颇语参考语法 [Jingpo Reference Grammar]. Beijing: China Social Science Press.
- Dai Qingxia & Li Jie. 2007. 勒期语研究 [A Study of the Leqi Language]. Beijing: Central University of Nationalities Publication House.
- Dai Qingxia & Xu Xijian. 1992. *The Grammar of Kachin*. Beijing: Zhongyang Minzu Xueyuan Chubanshe.
- Dai Qingxia, Jiang Ying & Kong Zhi'en. 2007. 波拉语研究 [A Study of the Bola Language]. Beijing: Publishing House of Minority Languages.
- Deumert, Ana. 2011. Multilingualism. In *The Cambridge Handbook of Social Linguistics*, ed. by Rajend Mesthrie, 261-282. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2013. Order of Demonstrative and Noun. In: Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/88> [accessed on October 8, 2017].
- Enfield, Nick J. 2001. On Genetic and Areal Linguistics in Mainland South-East Asia: Parallel Polyfunctionality of 'acquire'. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 255-290. Oxford/New York: Oxford University Press.
- Enfield, Nick J. 2003. *Linguistic Epidemiology: Semantics and Grammar of Language Contact in Mainland Southeast Asia*. London: RoutledgeCurzon.
- Enfield, Nick J. 2005. Areal Linguistics and Mainland Southeast Asia. In: *Annual Review of Anthropology* 2005, 181-206.
- Enriquez, Colin M. 1933. *Races of Burma*. Delhi: Manager of Publications.
- Gilhodes, Charles. 1996. *The Kachins: religions and customs*. Bangkok: White Lotus.
- Hanson, Ola. 1896. *A Grammar of the Kachin Language*. Rangoon: American Baptist Mission Press.
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In *Language Typology and Language Universals: An International Handbook*, ed. by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible, 1492-1510. Berlin/New York: Walter de Gruyter.
- Heine, Bernd & Tania Kuteva. 2006. *The Changing Languages of Europe*. Oxford/New York: Oxford University Press.
- Hertz, H.F. 1902. *A Practical Handbook of the Kachin or Chingpaw Language*. Rangoon: Government Printing.

- Huziwara Keisuke. 2016. *Cak-English-Bangla Dictionary*. Dhaka: A H Development Publishing House.
- Inglis, Douglas. 2007. *Nominal Structure in Tai Khamti*. Research Paper #312. Chiang Mai: Payap University.
- Inglis, Douglas & Connie Inglis. 2003. *A Preliminary Phonology of Ngochang*. Chiang Mai: Payap University.
- Inglis, Douglas, Nasaw Sampu, Wilai Jaseng & Thocha Jana. 2005. *A preliminary Ngochang-Kachin-English Lexicon*. Chiang Mai: Payap University.
- Jenny, Mathias. 2015. The far West of Southeast Asia: ‘Give’ and ‘get’ in the languages of Myanmar. In *Languages of Mainland Southeast Asia*, ed. by Nick J. Enfield & Bernard Comrie, 155-208. Berlin: De Gruyter.
- Jenny, Mathias & San San Hnin Tun. 2016. *Burmese: a comprehensive grammar*. New York: Routledge.
- Kurabe Keita. 2016. *A Grammar of Jinghpaw, from Northern Burma*. PhD dissertation. Tokyo: Tokyo University of Foreign Studies.
- LaPolla, Randy J. 2017. Dulong. In *The Sino-Tibetan Languages, 2<sup>nd</sup> edition*, ed. by Graham Thurgood & Randy J. LaPolla. London/New York: Routledge.
- LaPolla, Randy J. & David Sangdong. 2015. *Rawang-English-Burmese Dictionary*. Privately Published.
- Leach, Edmund R. 1954. *Political Systems of Highland Burma: A Study of Kachin Social Structure*. London: The Athlone Press.
- Lord, Carol. 1993. *Historical Change in Serial Verb Constructions*. Amsterdam: John Benjamins.
- Luk, Hkaw. *A Grammatical Sketch of Lacid*. (MA thesis) Chiang Mai: Payap University.
- Lustig, Anton. 2010. *A Grammar and Dictionary of Zaiwa*. Leiden/Boston: Brill.
- Maran, La Raw. 2007. On the continuing relevance of E. R. Leach’s “Political Systems of Highland Burma” to Kachin Studies. In *Social Dynamics in the Highlands of Southeast Asia: Reconsidering Political Systems of Highland Burma by E. R. Leach*, ed. by François Robinne & Mandy Sadan, 31-66. Leiden: Brill.
- Matisoff, James A. 1973. *The Grammar of Lahu*. Berkeley: University of California Press.
- Matras, Yaron & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4):829-865.
- Morey, Stephen. 2010. *Turung: a variety of Singpho language spoken in Assam*. Canberra: Pacific Linguistics.
- Mullaney, Thomas S. 2011. *Coming to terms with the nation: ethnic classification in modern China*. Berkeley: University of California Press.
- Müller, André. 2016. Linguistic convergence within the ‘Kachin’ languages. In: *International Institute for Asian Studies: The Newsletter* 75:34-35.
- Muysken, Peter. 2008. *From Linguistic Areas to Areal Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Robinne, François. 2007. Transethnic social space of clans and lineages: A discussion of Leach’s concept of common ritual language. In *Social Dynamics in the Highlands of Southeast Asia: Reconsidering Political Systems of Highland Burma by E. R. Leach*, ed. by François Robinne & Mandy Sadan, 283-297. Leiden: Brill.
- Robinne, François & Mandy Sadan. 2007. Reconsidering the Dynamics of Ethnicity Through Foucault’s Concept of ‘Spaces of Dispersion’. In *Social Dynamics in the Highlands of Southeast Asia: Reconsidering Political Systems of Highland Burma by E. R. Leach*, ed. by François Robinne & Mandy Sadan, 299-308. Leiden: Brill.

- Sadan, Mandy. 2007a. Constructing and Contesting the Category ‘Kachin’ in the Colonial and post-Colonial Burmese State. In *Exploring Ethnic Diversity in Burma*, ed. by Mikael Gravers, 34-76. Copenhagen: Nordic Institute of Asian Studies.
- Sadan, Mandy. 2007b. Translating Gumlau: History, the ‘Kachin’ and Edmund Leach. In *Social Dynamics in the Highlands of Southeast Asia: Reconsidering Political Systems of Highland Burma* by E. R. Leach, ed. by François Robinne & Mandy Sadan, 67-87. Leiden: Brill.
- Sadan, Mandy. 2013. *Being and becoming Kachin: histories beyond the state in the borderworlds of Burma*. Oxford: Oxford University Press.
- Sangdong, David. 2012. *A Grammar of the Kadu (Asak) Language*. PhD dissertation. Bundoora: La Trobe University.
- Sarep Hpung. 1996. A study of the morphology of verbs and nouns in the Sinwal dialect of the Rawang language. In: *Linguistics of the Tibeto-Burman Area* 19(2):93-184.
- Sawada Hideo. 2004. A tentative etymological word-list of Lhaovo (Maru) language. In *Approaches to Eurasian Linguistic Areas*, ed. by Fujishiro Setsu, 61-122. Kobe: Kobe City College of Nursing.
- Scott, James C. 2009. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven/London: Yale University Press.
- Straub, Nathan. 2014. *THIS and THAT in an undocumented dialect of Rawang*. Conference paper. 47<sup>th</sup> International Conference on Sino-Tibetan Languages and Linguistics. Kunming: Yunnan Normal University, October 7-19, 2014.
- Straub, Nathan. 2015. *The Logic of Kinship Terms in Rawang*. Presentation held on the 25<sup>th</sup> Annual Meeting of the Southeast Asian Linguistic Society. Chiang Mai: Payap University, May 27-29, 2015.
- Straub, Nathan. 2016. *Direction and Time Reference in the Rvmøɽ (Dvru) Dialect of Rawang, from Northern Myanmar*. MA thesis. Chiang Mai: Payap University.
- Sun Hongkai & Liu Guangkun. 2009. *A Grammar of Anong: Language Death Under Intense Contact*. Leiden/Boston: Brill.
- Thomason, Sarah Grey. 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
- Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: the interaction between language contact, change, and inheritance. In *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla, 3-21. London/New York: Brill.
- van Schendel, Willem. 2002. Geographies of knowing, geographies of ignorance: jumping scale in Southeast Asia. In: *Environment and Planning D: Society and Space* 20(6):647-668.
- Wang Zhaohui. 2005. 仙岛人及其语言 [*The Xiandao People and their Language*]. Beijing: Publishing House of Minority Languages.
- Wang Zhusheng. 1997. *The Jingpo: Kachin of the Yunnan Plateau*. Tempe: Arizona State University.
- Wannemacher, Mark. 2011. *A Phonological Overview of the Lacid Language*. Chiang Mai: Payap University.
- Wannemacher, Mark. 2017. *Zaiwa to English Dictionary – preliminary*. Unpublished manuscript.
- Wehrli, Hans J. 1904. *Beitrag zur Ethnologie der Chingpaw (Kachin) von Ober-Burma*. Leiden: Brill.
- Yabu Shiro. 1988. A Preliminary Report on the Study of the Maru, Lashi, and Atsi Languages of Burma. In *Historical and Cultural Studies in Burma*, ed. by Ishizawa Y., 65-132. Tokyo: Institute of Asian Cultures, Sophia University.

# UNIFYING CORRECTION AND SKEPTICISM IN VIETNAMESE *CƠ*

Anne Nguyen<sup>1</sup>  
National University of Singapore  
ntn@u.nus.edu

## Abstract

This paper aims to present the first unified formal semantics for the Vietnamese sentence-final particle *cơ*, which appears in corrective focus declaratives and in *wh*-interrogatives conveying skepticism. *Cơ* expresses the speaker's belief that a contrasting, previously mentioned proposition is false. My proposal uses Rooth's (1992) Alternative Semantics framework and Roberts' (1996) organization of discourse structure into a set of *moves*.

**Keywords:** sentence-final particles, semantics, discourse, focus, interrogatives

## 1 Introduction

190-6393 codes: vie

Vietnamese is a language with a rich system of sentence-final particles (SFPs), many of which have not been thoroughly described (Thompson 1965, Nguyen 1997, Le 2015). In this paper, I describe the semantics of Vietnamese sentence-final *cơ*. At first glance, the particle appears to have different functions in declaratives and in *wh*-interrogatives, conveying correction and skepticism respectively. A more careful investigation, however, reveals that the two uses share many similarities. I show that the two uses are subject to the same distribution constraints regarding some contrasting statement in the prior discourse. In addition, each use exhibits sensitivity to the position of focus or *wh*-words. To my knowledge, these points have not been discussed in previous work. Nguyen (1997) only describes its use in declaratives expressing preference. Though Le (2015) mentions that the particle might appear in multiple sentence types, its function is informally described as increasing the sentences' force. My aim thus is to present a formal semantics for the particle *cơ*, which can account for its properties in different configurations and thus unify the two uses.

## 2 Properties of *Cơ*

The Vietnamese sentence-final particle *cơ* can be used in both declaratives and *wh*-interrogatives, but with seemingly different functions. In declaratives, *cơ* appears in the situations in which the speaker attempts to correct what has been stated. This corrective use of *cơ* is illustrated in example (1).

- (1) A: 'What does Minh like?'  
B: 'Minh likes rabbits.'  
C: *Không phải. Minh thích [mèo]<sub>F</sub> cơ*  
not correct M. like cat *CƠ*  
'That's not right. He likes [cats]<sub>F</sub>.'

The people here are discussing what animals Minh likes. When speaker B claims that Minh likes rabbits, speaker C can follow up first by denying B's claim and then rectifying the false claim by adding

---

<sup>1</sup> I would like to thank my supervisor Michael Yoshitaka Erlewine for his suggestions and advice, and my Vietnamese friends, Ngo Van Binh, Nguyen Dinh An, Nguyen Thi Kieu Thanh, Nguyen Thanh Huong, and Nguyen Ha Khoa for discussion of the data and comments. Thanks also to the audience at the 27<sup>th</sup> meeting of the Southeast Asian Linguistics Society in Padang, Indonesia.



that Minh likes cats instead. In contrast, the use of *cơ* in *wh*-interrogatives expresses the speaker's skepticism about what has been said. An example of this skeptical use is in (2).<sup>2</sup>

- (2) A: 'Minh likes rabbits.'  
 B: *Minh thích gì cơ*      *Tớ tưởng Minh thích mèo chứ*  
 M. like what CỎ      I think M. like cat PRT  
 'Minh likes WHAT? I thought Minh likes cats.'

Speaker B in this example assumes that Minh likes cats, later explicitly indicated in their utterance. As this assumption is incompatible with A's claim that *Minh likes rabbits*, it is reasonable for B to challenge this statement with the *cơ wh*-question conveying skepticism.

Despite the fact that the two uses yield different pragmatic effects, further examination shows that both uses are subject to the same distributional constraints, opening up the possibility of a unified analysis. In the following subsections, I describe the constraints on the distribution of the particle in both declaratives and *wh*-interrogatives. In addition, I will discuss the sensitivity of the particle to focus and *wh*-constituent.

### 2.1 Distributional constraints

The presence of the sentence-final particle *cơ* is syntactically optional and does not affect the utterance's at-issue content (for declaratives, its truth condition). Despite this, the use of the particle is not always felicitous, but instead restricted by certain constraints. First, consider examples (3) and (4):

- (3) A: *Minh thích mèo (#cơ)*  
 M. like cat CỎ  
 'Minh likes cats.'
- (4) A: 'What animal does Minh like?'  
 B: *Minh thích mèo (#cơ)*.  
 M. like cat CỎ  
 'Minh likes cats.'

The statement containing *cơ* in (3) is discourse-initial, uttered at the beginning of the conversation. In (4), it is an answer to the immediately previous *wh*-question. The presence of *cơ* in both utterances is inappropriate. A comparison between these cases with example (1) in which the use of *cơ* is well-formed, shows that in (1), the *cơ* utterance is preceded by a contrasting claim by speaker B. This suggests that the reason for the inappropriateness of *cơ* in (3–4) is due to the absence of this contrasting claim in the prior discourse.

The use of skeptical *cơ* is restricted in the same way.

- (5) A: *Minh thích gì (#cơ)?*  
 M. like what CỎ  
 'What does Minh like?'  
 B: 'Minh likes cats.'

The *cơ* question in (5) is used at the beginning of the conversation and thus is not preceded by a contrasting statement. The infelicity of *cơ* here, therefore, is as expected. In short, the use of the particle requires a contrasting statement in prior discourse.

<sup>2</sup> *Wh*-interrogatives with *cơ* can also be used when the speaker has not heard parts of a prior utterance, i.e. as an echo question. Note that Vietnamese is a *wh*-in-situ language.

However, this condition is not enough to license the use of *cơ*, as shown in (6–7).

- (6) A: ‘What does Minh like?’  
 B: ‘(I know) Minh likes rabbits.’  
 C: (*Tớ biết*) *Minh thích [mèo]<sub>F</sub> (#cơ)*  
 I know M. like cat CỒ  
 ‘I know Minh likes cats.’  
 A: ‘So he likes rabbits and cats.’
- (7) A: ‘Minh likes rabbits.’  
 B: # *Minh thích gì cơ? Tớ cũng nghĩ thế*  
 M. like what CỒ I also think that  
 ‘Minh likes WHAT? I also think so.’

It is obvious that both the *cơ* statement of speaker C in (6) and the *cơ* question of speaker B in (7) are preceded by the contrasting claim *Minh likes rabbits*. However, the use of the particle in both examples is inappropriate, suggesting the involvement of some other requirement. To see what that requirement is, let’s compare (6) and (7) with minimally contrasting discourses in which the particle is licensed such as examples (1) and (2), repeated here as in (8) and (9).

- (8) A: ‘What does Minh like?’  
 B: ‘Minh likes rabbits.’  
 C: *Không phải M. thích [mèo]<sub>F</sub> cơ*  
*c cơ.*  
 not correct M. like cat CỒ  
 ‘That’s not right. He likes [cats]<sub>F</sub>.’
- (9) A: ‘Minh likes rabbits’  
 B: *M. thích gì cơ? Tớ tưởng M. thích mèo*  
 M. like what CỒ I think M. like cat  
 ‘Minh likes WHAT? I thought M. likes cats.’

Take a look at (6) and (8) first. In (8), speaker C thinks B’s claim is false and issues a correction in C. Speaker C in (6) above, however, does not share the belief that B’s utterance was false, leading A to conclude that Minh likes both. B’s utterance acts as a correction in (8) but not (6), corresponding to *cơ*’s felicity in (8) but not (6). Similarly, the contrast between (7) and (9) is also related to the speaker’s belief. In (9), by uttering the *cơ* question, B expresses his skepticism toward A’s claim that Minh likes rabbits. This is made clear by B’s continuation indicating that he thinks Minh likes another kind of animal, namely, cats. In contrast, speaker B continues in (7) with ‘I also think so,’ clearly indicating that B is in agreement with A. The particle *cơ* is infelicitous in (7B) because the speaker believes A’s claim.

In short, the felicity of the particle *cơ* in both its corrective and skeptical uses requires the existence of a contrasting statement in the previous discourse which the speaker believes to be false.

## 2.2 Sensitivity to focus and *wh*-constituent

I have shown that the use of *cơ* marks the existence of a contrasting proposition earlier in the discourse. This explains why we interpret a sentence containing *cơ* as introducing an inference regarding some contrasting proposition. Consider example (10). The utterance in (10) asserts that Minh bought a certain cat at a certain store. At the same time, the use of *cơ* with focus marking on the demonstrative *này* ‘this’ of the object ‘cat’ introduces an inference about some other cat.<sup>3</sup>

<sup>3</sup> On the phonetic realization of focus in Vietnamese, see Jannedy (2007).

- (10) *Minh mua con mèo [này]<sub>F</sub> ở cửa hàng này cơ*  
 M. buy CL cat [this]<sub>F</sub> at store this CỎ  
 ‘Minh bought this cat at this store.’  
 ↗ (The speaker thinks) Minh did not buy some other cat at this store.’

When stress is placed on another constituent, the demonstrative *này* ‘this’ of the locative as in (11), a different inference pertaining to some other store is drawn.

- (11) *Minh mua con mèo này ở cửa hàng [này]<sub>F</sub> cơ*  
 M. buy CL cat this at store [this]<sub>F</sub> CỎ  
 ‘Minh bought this cat at this store.’  
 ↗ (The speaker thinks) Minh did not buy the cat at some other store.’

This variation in interpretation of *cơ* in (10) and (11) correlating with the placement of focus suggests that the corrective *cơ* is focus sensitive. Skeptical *cơ* has a very similar behavior as illustrated in items (12) and (13).

- (12) *Minh mua [con mèo nào] ở cửa hàng này cơ*  
 Minh buy CL cat which at store this CỎ  
 ‘Minh bought WHICH cat at this store?’  
 ↗ The speaker doesn’t believe that Minh bought a particular cat at the store.
- (13) *Minh mua con mèo này ở [cửa hàng nào] cơ*  
 Minh buy CL cat this at store which CỎ  
 ‘Minh bought this cat at WHICH store?’  
 ↗ The speaker doesn’t believe that Minh bought the cat at a particular store.

The presence of skeptical *cơ* allows the speaker to convey his skepticism about a contrasting proposition in the previous discourse. In (12), when the *wh*-constituent is the DP *con mèo nào* ‘which cat,’ the speaker’s skepticism relates to some particular cat mentioned previously. In contrast, in (13) when the place where Minh bought the cat is in question, the utterance expresses skepticism about the cat being purchased at some particular location. In other words, skeptical *cơ* is sensitive to the position of *wh*-constituents.

To sum up, I have presented the main properties of *cơ*, through which we can see how closely related its two uses are. Both uses are subject to certain distributional constraints: the prior discourse must include a contrasting proposition which the speaker believes to be false. I also presented another parallel between corrective *cơ* and skeptical *cơ*, namely, their interpretations are sensitive to the position of focus and *wh*-constituents, respectively. An adequate formalization for *cơ* therefore must encode these various requirements.

### 3 Proposal

In this section, I will provide a formalization for the particle based on the generalization in the previous section using the Mats Rooth’s Alternative Semantics framework for the interpretation of focus and Craige Roberts’ theory of discourse as organized into a set of moves.

First, consider the sensitivity of the particle to the position of focus and *wh*-constituents. The focus sensitivity of corrective *cơ* suggests that the particle, like others that associate with focus, should be theoretically treated as quantifying over a set of alternatives determined by the position of focus (Rooth 1985, 1992). This approach can also be naturally extended to skeptical *cơ* in *wh*-interrogatives, which descriptively reflects a sensitivity to the position of *wh*-constituents. The semantics of questions is also

standardly treated as denoting a set of alternatives (Hamblin 1958; Karttunen 1977; von Stechow 1991; Beck 2006; Kotek 2014, 2016). Therefore, the particle’s two uses can be unified as referencing a contrasting proposition that belongs to the alternative set determined by focus-marking or a *wh*-constituent.

The felicity conditions of *cơ* also require that the contrasting proposition is previously mentioned in the discourse. This part of meaning “previously mentioned” can be encoded using Roberts’ (1996) organization of discourse structure into a set of *moves*. The contrasting proposition must be a discourse move in the set of moves prior to the point that *cơ* is uttered.

I propose the semantics for the sentence-final particle *cơ* as follows.  $[[\alpha \text{ cơ}]]$  has the at-issue content of  $[[\alpha]]$  and introduces the not-at-issue meaning in (14):<sup>4</sup>

- (14)  $[\alpha \text{ cơ}] \rightsquigarrow \exists p . ( p \in M \quad \text{and} \quad \text{(i)}$   
 $\quad \quad \quad p \in [[\alpha]]^f \quad \text{and} \quad \text{(ii)}$   
 $\quad \quad \quad \text{the speaker believes that } p \text{ is false } ) \quad \text{(iii)}$

Here, *M* is the set of previous discourse *moves* (Roberts, 1996) prior to *cơ*’s utterance. The focus alternative set  $[[\alpha]]^f$  is the set of all interpretations obtained by replacing all the F-marked constituents in  $\alpha$ , which can be a focus (Rooth 1992) or a *wh* constituent (Roberts 1996), with contextually-determined alternatives.

#### 4 Analysis

The main focus of this section is to re-examine the examples in Section 2 with respect to the proposed semantics for the particle *cơ* above. I will first demonstrate that the proposal can account for the felicity of the particle. Then, I will discuss how this proposal accurately rules out the use of *cơ* in the infelicitous conversations, motivating each clause of the not-at-issue content in the proposal above. Lastly, I will show that the proposal accounts for the correlation between the interpretation of the particle with the placement of focus and *wh*-constituents.

##### 4.1. The felicity of corrective *cơ* and skeptical *cơ*

I will begin the discussion with conversation (1) repeated here as (15). I argue that the felicity of *cơ* in (15) is predictable under the proposed semantics. The analysis for the sentence with *cơ* (15C) under the proposed semantics is as in (16).

- (15) A: ‘What does Minh like?’  
 B: ‘Minh likes rabbits.’  
 C: *Không phải. Minh thích [mèo]<sub>F</sub> cơ*  
 not correct M. like cat CO  
 ‘That’s not right. He likes [cats]<sub>F</sub>.’

- (16) a. The scope of *cơ*:  $\alpha = [{}_S \text{ Minh likes } [{}_{\text{cat}}]_F]$   
 b. Previous moves:  $M = \{ m_1 - \text{‘What animal does Minh like?’},$   
 $\quad \quad \quad m_2 - \text{‘Minh likes rabbits’},$   
 $\quad \quad \quad m_3 - \text{‘That’s not right’ (= Minh does not like rabbits) } \}$   
 c.  $[[\alpha]]^f = \{ \text{that Minh likes cats, that Minh likes rabbits, ... } \}$   
 d. Not-at-issue content:  $\exists p . ( p \in M \quad \text{and} \quad \text{(i)}$   
 $\quad \quad \quad p \in [[\alpha]]^f \quad \text{and} \quad \text{(ii)}$   
 $\quad \quad \quad \text{the speaker believes that } p \text{ is false } ) \quad \text{(iii)}$   
 e. At-issue content:  $[[\alpha]] = \text{that Minh likes cats}$

<sup>4</sup> The not-at-issue meaning in (15) may be a presupposition or conventional implicature. I do not distinguish between these options here.

We first check that the felicity condition on the use of *cơ* in (16d) is met in (15). The requirement in (16d) will be met by the proposition  $p = \textit{that Minh likes rabbits}$ . The proposition  $p$  belongs both to the set of prior discourse moves  $M$  in (16b) ( $p = m_2 \in M$ ) and to the focus alternatives of  $\alpha$ ,  $[\alpha]^f$ , in (16c). Moreover, by making move  $m_3$  in (16b), the speaker  $C$  explicitly indicates that he thinks the proposition  $p$  is false. Therefore, the proposition  $p = \textit{that Minh likes rabbits}$  satisfies all three requirements in (16d), predicting the felicity of *cơ* in (15). The at-issue content *that Minh likes cats* (16e) is then asserted.

The felicity of skeptical *cơ* in example (2), repeated here as (17), follows the same logic.

- (17) A: ‘Minh likes rabbits.’  
 B: *Minh thích gì cơ? Tôi tưởng Minh thích mèo chứ*  
 M. like what CỐ I think M. like cat PRT  
 ‘Minh likes WHAT? I thought Minh likes cats.’
- (18) a. The scope of *cơ*:  $\alpha = [{}_S \textit{Minh likes what}]$   
 b. Previous moves:  $M = \{ m_1 - \textit{‘Minh likes rabbits’} \}$   
 c.  $[\alpha]^f = \{ \textit{that Minh likes cats, that Minh likes rabbits, ...} \}$   
 d. Not-at-issue content:  $\exists p . ( p \in M \quad \text{and} \quad \text{(i)}$   
 $p \in [\alpha]^f \quad \text{and} \quad \text{(ii)}$   
 $\text{the speaker believes that } p \text{ is false} ) \quad \text{(iii)}$   
 e. At-issue content: the question ‘What does Minh like?’<sup>5</sup>

The requirements of *cơ* in (18d) are met by the proposition  $p = \textit{that Minh likes rabbits}$ . First, we note that  $p = m_1$  in the set of prior discourse moves  $M$  (18b). For the alternative set of  $\alpha$  in (18c), I follow Beck (2006) and Kotek (2014, 2016), who compute  $[\alpha]^f$  for  $\alpha$  containing the *wh*-word ‘what’ to be the set of possible answers. The proposition  $p$  is in this set as well. Finally, the speaker of (17B) must believe that  $p$  is false. That speaker B does not believe  $p$  is reflected in (17) above by B’s continuation, that he thinks Minh likes cats instead.

Notice that the last requirement on  $p$  in (18diii) — independently motivated through the study of corrective *cơ* — explains the sense of skepticism felt in such examples. Even in a discourse without B’s explicit continuation as in (17B) above, the addition of *cơ* to a *wh*-interrogative will express the speaker’s skepticism towards a previously mentioned proposition.

#### 4.2 Infelicitous examples motivate clause (14i) regarding the set of previous discourse moves $M$

The felicity of the particle requires that there be a proposition  $p$  which simultaneously satisfies the three clauses in the not-at-issue content of *cơ* (14). This implies that the inappropriate use of the particle can be seen in the situations in which the requirement of one of the clauses is not met. Consider conversations (3) to (5), repeated here as (19) to (21).

- (19) A: *Minh thích [mèo]<sub>F</sub> (#cơ) (Câu ấy không thích thỏ)*  
 M. like cat CỐ he not like rabbit  
 ‘Minh likes [cats]<sub>F</sub>. (He doesn’t like rabbits.)’
- (20) A: ‘What (animal) does Minh like?’  
 B: *Câu ấy thích mèo (#cơ) (Câu ấy không thích thỏ)*  
 he like cat CỐ he not like rabbit  
 ‘He likes [cats]<sub>F</sub>. (He doesn’t like rabbits.)’

<sup>5</sup> The question here may be considered an echo question. See e.g. Ginzburg and Sag (2001), Artstein (2002), Sudo (007) on the semantics of echo questions.

- (21) A: *Minh thích gì (#cơ)? (Tớ tưởng cậu ấy ghét thỏ.)*  
 M. like what CỐ I think he hate rabbit  
 ‘Minh likes WHAT? (I think he hates rabbits.)’

The uses of corrective *cơ* in (19) and (20) and skeptical *cơ* in (21) are unacceptable because the *cơ* utterances are not preceded by a contrasting proposition. Under my proposal, they fail to meet clause (14i). For example, the most appropriate proposition  $p = \textit{that Minh likes rabbits}$  can only meet (14ii–iii) and fails to meet (14i). In (19) and (21),  $[\alpha \textit{ cơ}]$  is a discourse-initiation move in the strict sense, which means that the set of discourse moves  $M$  at this point is totally empty. There is not a single move, either explicit or implicit, in  $M$  at this point. In (20),  $M$  consists of a single *wh*-question move. Although the *wh*-question in (20A) and the proposition  $p$  are related —  $p$  is a possible answer to A’s question — the proposition  $p$  itself is not in the previous discourse moves  $M$ .

These examples together motivate clause (14i), requiring the proposition  $p$  to be in the set of previous discourse moves  $M$ .

#### 4.3. Infelicitous examples motivate clause (14iii) regarding the speaker’s belief

Now I will briefly discuss conversations (6–7), repeated here as (22) and (23).

- (22) A: ‘What does Minh like?’  
 B: ‘(I know) Minh likes rabbits.’  
 C: *(Tớ biết) Minh thích [mèo]<sub>F</sub> (#cơ)*  
 I know M. like cat CỐ  
 ‘I know Minh likes cats.’  
 A: ‘So he likes rabbits and cats.’
- (23) A: ‘Minh likes rabbits.’  
 B: *# Minh thích gì cơ? Tớ cũng nghĩ thế*  
 M. like what CỐ I also think that  
 ‘Minh likes WHAT? I also think so.’

As discussed in the previous section, the use of the particle here as corrective *cơ* or skeptical *cơ* is ruled out because the speaker does not hold the belief that some previous, contrasting claim is false. Consider the proposition  $p = \textit{that Minh likes rabbits}$ . This belongs both to the set of previous discourse moves  $M$  and the alternative set  $[[\alpha]]^f$  which is determined by the focus or *wh*-constituent. Accordingly, the infelicity of the particle in both conversations is clearly because the requirement of (14iii) is not satisfied, that is, the speaker does not think  $p$  is false. The infelicitous conversations, therefore, directly motivate clause (14iii) pertaining to the speaker’s belief about the contrasting previously mentioned proposition.

#### 4.4 Sensitivity to focus and *wh*-constituent and clause (14ii) involving the alternative set $[[\alpha]]^f$

The proposal can also provide an explanation for the sensitivity of the particle to the placement of focus or *wh*-constituents. Take the contrast between (12–13), repeated here as (24–25), as an example.

- (24) *Minh mua [con mèo nào] ở cửa hàng này cơ*  
 Minh buy [CL cat which] at store this CỐ  
 ↗ The speaker doesn’t believe that Minh bought a particular cat at the store.
- (25) *Minh mua con mèo này [ở cửa hàng nào] cơ*  
 Minh buy CL cat this [at store which] CỐ  
 ↗ The speaker doesn’t believe that Minh bought the cat at a particular store.

I argue that the conditions of the use of *cơ* (14) make (24) and (25) appropriate in different discourse contexts, raising the different inferences observed. The not-at-issue content of the particle requires some proposition *p* to belong to the alternative set  $[[\alpha]]^f$ . This set differs between the *wh*-question in (24) and (25), varying in the choice of *cat* in (24) and the choice of *store* in (25), resulting in the different inferences introduced by the particle.

In short I have demonstrated that the proposed semantics for the particle can account for the contrasts documented in Section 2. In addition, the analysis clearly indicates that the corrective use and skeptical use of *cơ* can be unified under the proposed semantics. Lastly, the elaboration in this section allows us to see how the parts of the proposal are motivated based on the collected data.

## 5 Enriching previous discourse moves

The discussion so far indicates that the particle *cơ* reflects the existence of a contrasting proposition which has been previously mentioned, which is formalized as being in the set of a previous discourse moves. The presence of *cơ* hence allows the hearer to track the structure of the previous discourse. In some cases, the use of *cơ* can in fact shed light on the organization of the previous discourse, contributing to more effective communication. In this section, I will show that the felicity of *cơ* provides evidence for the set of discourse moves *M* containing moves strengthened with conversational implicatures by cooperative principles. Consider the following examples:

- (26) A: ‘Which book did Minh buy?’  
 B: ‘Minh bought this book.’  
 C: *Minh mua cả quyển kia nữa cơ*  
 M. buy also CL that PRT CƠ  
 ‘Minh also bought that book.’  
 a. ✘ ↗ The speaker believes that Minh didn’t buy this book.  
 b. ✔ ↗ The speaker believes that Minh didn’t ONLY buy this book.

In (26), under my proposal, the felicity of the particle in C’s utterance requires the existence of a proposition *p* which C thinks is false. A candidate for *p* is the content of B’s claim, *that Minh bought this book*. But notice that speaker C in (26) is not correcting B’s claim and in fact seems to believe *p*. Instead, we interpret (26C) as indicating that the speaker disbelieves a strengthened version of this claim, *p’ = that Minh only bought this book*.

We can verify that the speaker in C does not disbelieve the proposition *p = that Minh bought this book* by explicitly denying it, as in (27). Here, the *cơ* utterance is infelicitous.

- (27) A: ‘Which book did Minh buy?’  
 B: ‘Minh bought this book.’  
 C: # *Không phải. Minh mua cả quyển kia nữa cơ*  
 Not true M. buy also CL that PRT CƠ  
 ‘That’s not true. Minh also bought that book.’

We conclude that *cơ* in (26) is licensed by the strengthened move *p’ = that Minh only bought this book*. However, notice that no speakers in (26) explicitly mentioned this strengthened statement *p’*. I claim that this move arises implicitly by default because B’s answer ‘Minh bought this book’ is congruent to the Question Under Discussion ‘Which book did Minh buy?’ (26A). The contrast between (26–27), then, indicates that the felicity of *cơ* marks the presence of this implicit, strengthened move in the representation of previous discourse moves. Thus, the particle provides evidence of the formal enrichment of discourse moves through conversational implicature.

## 6 Conclusion

The paper presented a formal semantics for the Vietnamese sentence-final particle *cơ*, which can unify its corrective and skeptical uses. The particle reflects the existence of a contrasting proposition in the previous discourse which the speaker believes to be false. This meaning is formalized by adopting the Roothian Alternative Semantics framework and Roberts' formalization of discourse into a set of discourse moves. I showed that SFP *cơ* can manifest the implicit structure of discourse. The detailed study of SFP *cơ*, therefore, can refine our theory of speakers' mental representations of discourse.

## References

- Artstein, Ron. 2002. *Parts of Words*. Ph.D. dissertation, Rutgers University.
- Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1-56.
- Ginzburg, Jonathan Ginzburg, and Ivan Sag. 2001. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford: CSLI.
- Hamblin, Charles. 1973. Questions in Montague English. *Foundations of Language* 10:41-53.
- Jannedy, Stephanie. 2007. Prosodic focus in Vietnamese. In *Interdisciplinary Studies on Information Structure* 8, ed. by Ishihara, S., S. Jannedy, and A. Schwarz, 209-230.
- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3-44.
- Kotek, Hadas. 2014. *Composing questions*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Kotek, Hadas. 2016. On the semantics of wh-questions. In *Proceedings of Sinn und Bedeutung 20*, ed. by Nadine Bade, Polina Berezovskaya, and Anthea Schöller.
- Le, Ha Giang. 2015. *Vietnamese sentence final particles*. MA thesis, University of Southern California.
- Nguyen, Dinh Hoa. 1997. *Tieng Viet khong son phan* [Vietnamese without veneer]. London Oriental and African Language Library.
- Roberts, Craige. 1996. Information Structure: Towards an integrated formal theory of pragmatics. In *OSUWPL Volume 49: Papers in Semantics*, ed. by Jae Hak Yoon and Andreas Kathol.
- Rooth, Mats. 1985. *Association with Focus*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75-116.
- Sudo, Yasutada (2007) A metalinguistic semantics for echo questions. In *Proceedings of the 16th Amsterdam Colloquium*, 205-211.
- Thompson, Laurence C. 1965. *A Vietnamese reference grammar*. Seattle: University of Washington Press.
- von Stechow, Arnim. 1991. Focusing and backgrounding operators. In *Discourse particles: Descriptive and Theoretical Investigations on the Logical, Syntactic and Pragmatic Properties of Discourse Particles in German*, ed. by Werner Abraham, 37-84. John Benjamins.



# ON THE FUNCTIONAL EXTENSION OF THE THAI FORM *jaŋ* INTO THE INTERROGATIVE MARKER<sup>1</sup>

Noppadon Nuntasukasame  
*Kasetsart University*  
*ujingshian@gmail.com*

Kachen Tansiri  
*Kasetsart University*  
*tansiri.kc@gmail.com*

Natchanan Natpratan  
*Kasetsart University*  
*y.natchanan@gmail.com*

## Abstract

Generally, the form *jaŋ* in Modern Standard Thai, similar to the counterpart *noch* in German (van Der Auwera 1993), *hai* in Mandarin (Yeh 1998), and *still* in English (Michaelis 1993), serves as a continuative aspect marker expressing the persistence of a state or an activity (Jarajjarungkiat, 2008 and Tansiri 2011). In addition to being a continuative aspect marker, *jaŋ* serves as an interrogative marker appearing at the end of an affirmative sentence to inquire as to whether the state has already taken place or whether the action has already been done. Based on 1,500 sampled sentences in which *jaŋ* appeared in the Thai National Corpus (TNC), it is hypothesized that the continuative aspect marker *jaŋ* has extended its grammatical function as the interrogative marker *jaŋ*. The main purpose of this study was to investigate the motivation and the mechanism of the functional extension of the continuative aspect marker *jaŋ* as the interrogative marker *jaŋ*, which has never been thoroughly studied in previous research concerning *jaŋ*. The findings revealed that the motivation of this development could be attributed to inference in accordance with Bybee et al.'s (1994:281-301) framework; furthermore, the mechanism in the initial stage of the change proposed by Traugott (2011) might have been repetition leading to frequency. Effects of high token frequency led to the erasure of word boundaries, i.e., [*rű:*] 'or' + [*jaŋ*] > [*rű: jaŋ*], and the phonological reduction in turn came about at the end of the process—i.e. [*rű: jaŋ*] > [*rű jaŋ*] > [*jaŋ*].

**Keywords:** grammaticalization, lexicalization, functional extension, continuative aspect marker, interrogative marker

**ISO 639-3 codes:** tha

## 1 Introduction

The form *jaŋ* in Modern Standard Thai has received a great deal of attention among Thai and foreign linguists because of its syntactic and semantic properties. Boonyapatipark (1983) proposed that *jaŋ* serves as an imperfective aspect marker indicating the persistence of a state or an activity. For instance, *jaŋ* in (1) indicates the persistence of Somchai's studying at his old school.

---

<sup>1</sup> This article is a part of the first author's dissertation titled "The Semantico-syntactic Properties of the Word /*jaŋ*/ in Thai".

- (1) *Somchai*      *jaŋ*      *rian*      *jù:*      *tʰi:*      *ro:ŋrian də:m*  
 Somchai      still      study      be      at      school      old  
 ‘Somchai has still studied at his old school.’

Jaratjarungkiat (2008) suggested four uses of *jaŋ* in Modern Standard Thai as follows:

- a) *jaŋ* as a verb conveys the causative meaning, which is used in only some specific contexts, i.e., the subject of a sentence has to be *abstract noun*. One such example is given in (2).

- (2) *ka:n*      *banlú?*      *pʰõnsǎmrèt*      *kʰráŋ*      *ní:*      *nâ: teà?*      *jaŋ*      *kʰuam*      *jindi:*  
 NOM      achieve      success      time      this      should      cause      NOM      pleased  
*kæ:*      *tʰúk*      *kʰon*  
 to      every      man  
 ‘This success will likely please everyone.’

Note that the example in (2) shows that the subject *this success* is an abstract noun, and the verb *jaŋ* means *to cause, to make*.

- b) *jaŋ* as a pre-verbal auxiliary indicates continuance aspect meaning. In (3), *jaŋ* expresses *my existence at the office*, which has continued up to the time when the speech was uttered.

- (3) *tɛʰǎn*      *jaŋ*      *jù:*      *tʰi:*      *tʰam*      *ŋa:n*  
 I      still      be      place      do      work  
 ‘I am still at the office.’

- c) *jaŋ* as a preposition must occur with motion verbs, especially *paj* ‘go’ and *ma:* ‘come’; hence, the word has only one meaning that indicates the case of ‘goal’ as in (4).

- (4) *pâ:*      *Somjai*      *dən*      *tʰa:ŋ*      *paj/ma:*      *jaŋ*      *Kanchanaburi*      *ta:m*      *lampʰaŋ*  
 aunt      Somjai      walk      way      go/come      to      Kanchanaburi      by      alone  
 ‘Auntie Somjai went/came to Kanchanaburi (province) alone.’

- d) *jaŋ* as a conjunction has two meanings, viz.: additive relation in (5) and adversative relation found in the context of the comparative construction in (6).

- (5) *nô:k teàk*      *Suda*      *teà?*      *mâj*      *pʰû:t*      *læ:w*      *jaŋ*      *tʰam*      *nâ:*      *tɛʰǎ:j mə:j*  
 besides      Suda      will      NEG      speak      PFV      also      do      face      indifferent  
*sàj*      *tɛʰǎn*      *ʔik*  
 on      I      too  
 ‘Not only did Suda keep silent, but she was also indifferent to me.’

- (6) *dían*      *kʰít*      *wâ:*      *ka:n*      *sú:*      *hǔaj*      *tâj*      *din*      *jaŋ*      *dí:*  
 I      think      COMP      NOM      buy      lottery      under      earth      still      good  
*kùa*      *ka:n*      *kʰâw*      *bò:n*      *pen*      *nǎj*      *nǎj*  
 more      NOM      enter      casino      be      where      where  
 ‘I think buying an illegal lottery is still much better than getting into the casino.’

In (5), *jaŋ* provides additional evidence to confirm the text theme, i.e., the theme being Suda’s keeping silent while the clause marked by *jaŋ* provides another situation of her indifference to me, which involves the theme. In (6), *jaŋ* signifies adversative relation which implies a comparison between buying an illegal lottery and getting into the casino.

Besides the aforementioned uses, Tansiri (2011:160) added the interrogative marker, which is used for asking whether a state has already occurred or whether an action has already been accomplished, in its syntactic properties, as in (7).

- (7)    *mâj*    *sâ:p*    *wâ:*    *tæ:ɲtua*    *sèt*    *jaŋ*    *kʰráp*  
 NEG    know    COMP    dress    finish    INT    PT  
 ‘Have you dressed up yet?’

In (7), the speaker would like to know whether you have already dressed up by means of adding the interrogative marker *jaŋ* at the end of the sentence.

As for the grammaticalization of *jaŋ*, previous studies mentioned only the development from verb to other functions. Jaratjarungkiat (2008) proposed that the historical development of *jaŋ* has four pathways, as follows: (1) the development from the verb meaning to exist to a pre-verbal auxiliary (2) the development from the verb meaning to exist to a conjunction (3) the development from the verb meaning to exist to a causative verb, and (4) the development from the verb meaning to stay or to be in/at to a preposition. Pattaranavik (2008) suggested that grammaticalization of *jaŋ* in the Sukhothai period (1238-1438) is also the development from verb to preposition and pre-verbal auxiliary. However, these studies did not examine how *jaŋ* as an interrogative marker correlated with other functions.

The objective of the present study is to argue that the interrogative marker *jaŋ* is derived from the aspectual marker *jaŋ*. Furthermore, this paper also examines mechanisms and motivations of this development—how the change occurs, and why there is a change.

The organization of this paper is as follows: Section 2 reviews previous work on the mechanism and motivation of grammaticalization proposed in Traugott (2011) and Bybee et al. (1994) and the notion of the *layering principle* suggested by Hopper (1991). Section 3 focuses on linguistic context having an effect on the functional extension of the form *jaŋ*. Section 4 investigates the motivation behind the functional extension of the form *jaŋ* according to Bybee et al.’s (1994) framework. Section 5 is concerned with the mechanism of the development of the form *jaŋ* from the aspectual marker to the interrogative marker. Section 6 subsequently discusses two changes—erasure of word boundaries and phonological reduction—which arise from high token frequency. Section 7 verifies that the development of form *jaŋ* from an aspectual marker to an interrogative marker is taking place at the present time. The last section summarizes the present findings.

## 2 Literature Review

### 2.1 An overview of grammaticalization

Grammaticalization in Hopper and Traugott ‘s point of view (Hopper and Traugott 2008:2) has been defined in two perspectives. The first, a historical perspective, investigates the source of grammatical forms and the process of development which occurred over time. From this perspective, grammaticalization is considered as the subset of linguistic changes in which a lexical item or construction in certain linguistic contexts acquires grammatical function or in which an existing grammatical unit acquires a new grammatical function. The second is a more synchronic perspective studied from the viewpoint of fluid, dynamic patterns of language used at a moment in time. The present study essentially focuses on the process whereby grammatical function develops to further grammatical function synchronically; as a result, the terms, functional extension and grammaticalization, are used interchangeably throughout the paper.

Hopper and Traugott (2008:2) also emphasized the importance of a *specific context* as to whether a change occurred or not; that is to say, the lexical item or construction has to be in a very local context; thus, development into a new grammatical function will take place, such as the example *going to* in (8) and (9).

- (8)    I *am going to* London.  
 (9)    I *am going to* marry Bill.

In English, *going to* becomes the future marker *gonna* only when it is in the specific context where a verb follows *to* as shown in (9); otherwise, the change does not occur, such as the sentence in (8).

## 2.2 Mechanism and motivation of functional extension

Traugott (2011) proposed three mechanisms of change, i.e. how a change occurred, which are reanalysis, analogy, and repetition.

As for the first mechanism, reanalysis was defined by Langacker (1977:58 as cited in Hopper and Traugott 2008:51) as “change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic change of its surface manifestation”. Furthermore, he continued by subdividing it into two subtypes. One is *resegmentation*, i.e., boundary loss, boundary creation, and boundary shift. The other is *syntactic/semantic reformulation*, i.e., ranging from semantic change to change in agreement patterns. Crucially, both subtypes of reanalysis need *ambiguity* (also called *opacity*) in the surface structure; in other words, there are at least two interpretations for the structure to be analyzed (Hopper and Traugott, 2008:52).

Secondly, analogy is the overt mechanism which modifies surface manifestations, but in itself does not affect rule change (Hopper and Traugott 2008:39). Having compared reanalysis and analogy, Hopper and Traugott (2008:64) concluded that the former causes rule change and modifies deep representations while operating on a syntagmatic axis; on the contrary, the latter operates on a paradigmatic axis. In morphology, there are two types of analogy which are often mentioned. One is *leveling*, i.e., a reduction of stem allomorphs. The other is *extension or generalization*, i.e., changing the form which derives from a less predominant way to the form which derives from a more predominant way (Traugott 2011:4).

The last mechanism is repetition leading to frequency. Bybee (2003:603) proposed a new definition of grammaticalization—the importance of repetition—as “the process by which a frequently-used sequence of words or morphemes becomes automated as a single processing unit”. Moreover, the frequency was classified as *token frequency*, i.e., the number of times a particular form occurs in text and *type frequency*, i.e., the number of items that are available to a particular class of forms. She argued that the high token frequency of grammaticalizing phrases provides the trigger device for many changes, such as semantic fading, phonological reduction, positional fixing, and erasure of word boundaries, and also explained that the reason why high token frequency leads to many changes was that “it affects the nature of the cognitive representation” (Bybee 2003:604).

Nicolle (1998:15) claimed that the motivation—the cause of any case of grammaticalization—is likely to be *semantic change* rather than other formal change, which is in accord with Bybee et al.’s framework. They identified five mechanisms of semantic change that were involved in the development of grams, namely *metaphorical extension*, *inference*, *generalization*, *harmony*, and *absorption of contextual meaning* (Bybee, Perkins and Pagliuca 1994:282-301).

The first mechanism is metaphorical extension, “a semantic shift from one domain (which is more concrete) to another one (which is more abstract) with preservation of some of the relational structure originally expressed” (Bybee et al. 1994:283). Heine, Claudi and Hünemeyer (1991:48) proposed the metaphorical extension in the form of metaphorical chains: PERSON > OBJECT > ACTIVITY > SPACE > TIME > QUALITY. They illustrated the effect of metaphorical extension on the use of the Ewe word *megbé* ‘back—a body part’ (OBJECT) towards a location either as an adverb or as a postposition ‘behind’ (SPACE), temporal marker ‘late’ (TIME), and finally denoting “dull or mentally retarded” (QUALITY).

As for the second mechanism of semantic change, inference is the process by which the listener uses additional knowledge to make sense of what is not explicit in an utterance from the speaker (Yule 2014:131). According to Bybee et al. (1994), inference is the force behind the grammaticalization process if, and only if, the inference suggested by the context, invited inference or pragmatic meaning, has been conventionalized to be a coded meaning (Nicolle 1998:18; Evans and Green 2006:721); in other words, the situations in which the inference is present must occur frequently enough for language-users to assume that inference is a necessary part of the meaning of form (Bybee et al.1994:197). Consequently, inference, which is the force behind the grammaticalization process, is also called *conventionalization of implicature* (Bybee et al. 1994:282).

Thirdly, generalization can be described in terms of “the loss of specific features of meaning with the consequent expansion of appropriate contexts of use for a gram” (Bybee et al. 1994:289). An example of generalization is illustrated in the development of *can* from signifying a sense of mental

ability to general ability and finally to root possibility. Nevertheless, it is not clear whether generalization in itself is a mechanism of change or whether generalization is just the outcome of change caused by some other mechanism (Bybee et al. 1994:290).

As for harmony, it is referred to as using form which semantically harmonizes with context, subsequently spreading to similar contexts. Bybee et al. (1994:239-240) argued that “harmony would be restricted to cases in which the grammaticizing element had already lost most of its semantic content”; the reason is that the meaning of a gram does not contribute any new information to a sentence since it merely echoes the temporal reference, modality, etc. expressed in the main clause (Nicolle 1998:19). For instance, in British English, *should* is used in a subordinate clause where the meaning of *should* harmonizes with the modal force of the main clause (Bybee et al. 1994:293). However, Thepkanjana (2016:59-60) commented that it is not clear how *harmony* was the motivation behind the grammaticalization process.

The final mechanism of semantic change, absorption of contextual meaning, is a process in which a form absorbs parts of its contextual meaning (Lessau 1994:4). Lessau (1994:4) views it as a semantic consequence of the pragmatic process of inference, but Bybee et al. (1994:296) emphasized that contextual absorption differs from pragmatic inference in that its meaning is absorbed from linguistic contexts rather than being conventionalized from pragmatic contexts.

### 2.3 The layering principle

Layering is one of the five principles proposed by Hopper (1991). It refers to the phenomenon where a language can develop many forms for the same function at any one synchronic moment in time. Layering may be considered as *variability* (Hopper and Traugott 2008:124).

Hopper (1991:22) proposed that “[w]ithin a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers”. Nevertheless, the diverse forms serving the same function may have slightly different meaning or may simply be recognized as stylistic alternatives or sociolinguistic registers (Hopper 1991:23).

Bybee et al. (1994:21) exemplified the layering of futural expressions in English: *will*, *shall*, and *be going to*, as in examples (12)-(14).

- (10) I *am going to* study English.
- (11) I *will* study English.
- (12) I *shall* study English.

In view of futural expression, all still appear in Modern English; however, there is a little difference with regard to these three futural expressions; namely, *will* and *shall* are used for predicting future events which are not already decided or obviously *on the way*; whereas, *be going to* is used for future events which are already decided ahead of time (Swan 1995:134-135). As to the difference between *will* and *shall*, the Oxford Dictionary mentioned that “the two words are used more or less interchangeably, and this is now an acceptable part of standard British and American English; however, the word *shall* is now seldom used in any normal context in American English.” (<https://en.oxforddictionaries.com>)

### 3 Specific linguistic contexts of functional extension from the aspectual marker *jaŋ* to the interrogative marker *jaŋ*

Because specific linguistic contexts have had an influence on grammaticalization, it is hypothesized that specific linguistic contexts causing functional extension from the aspectual marker *jaŋ* to the interrogative marker *jaŋ* are as follows.

Firstly, the aspectual marker *jaŋ* must occur in the context of asking someone to choose between two alternatives—alternative interrogative sentences; furthermore, the verb which appears in both alternatives must be the same. If there is a noun phrase following the verb, the first noun phrase has to be similar to the second one. In addition, the aspectual marker *jaŋ* must co-occur with the negative

marker (NEG) *mâj*; moreover, both the aspectual marker *jaŋ* and the negative marker *mâj* must appear only in the second alternative. Lastly, when the verbs appearing in both alternatives are action verbs, the perfective marker (PFV) *lê:w* must appear in the first alternative. Examples include (13) to (17).

- (13) *Suda kin kʰâ:w-tɛʰáw lê:w rǔ: jaŋ mâj dâj kin kʰâ:w-tɛʰáw*  
 Suda eat breakfast PFV or still NEG get eat breakfast  
 ‘Has Suda had breakfast yet?’  
 (Literal translation: ‘Has Suda had breakfast, or not?’)

- (14) *tʰə: mâj tɛʰɔ:p faŋ ple:ŋ rók lê:w rǔ: jaŋ tɛʰɔ:p faŋ*  
 you NEG like listen song rock PFV or still like listen  
*ple:ŋ rók jù*  
 song rock PROG  
 ‘Didn’t she like listening to rock music?’, or ‘Did she still like listening to rock music?’

- (15) *tʰə: kliat kʰăw rǔ: jaŋ rák kʰăw jù*  
 you hate him or still love him PROG  
 ‘Do you hate him, or still love him?’

- (16) *núua-kʰû: tɛʰăn taj paj lê:w rǔ: jaŋ mâj kət*  
 soulmate I die go PFV or still NEG be born  
 ‘Has my soulmate already died, or hasn’t he been born yet?’

- (17) *tʰə: mâj jà:k mi: kraj rǔ: jaŋ mâj jà:k mi: raw*  
 you NEG want have who or still NEG want have I  
 ‘Don’t you want anyone else?’, or ‘Don’t you want me?’

In (13), the contexts in which the aspectual marker *jaŋ* occurs conform with all of the conditions mentioned above; therefore, the sentence could be changed to *Suda kin kʰâ:w tɛʰáw lê:w jaŋ* ‘Has Suda had breakfast yet?’ with the interrogative marker *jaŋ* at the end of the sentence. Because of not conforming to the second condition—the negative marker *mâj* appears in the first alternative; the sentence in (14) is not developed into *\*tʰə: mâj tɛʰɔ:p faŋ ple:ŋ rók lê:w jaŋ*. In (15), it is not changed to *\*tʰə: kliat kʰăw jaŋ* because verbs in the first and the second alternatives are different; moreover, the negative marker *mâj* does not appear in the second alternative. In (16), it is impossible for the sentence to change to *núua kʰû: tɛʰăn taj paj lê:w jaŋ* because verbs in both alternatives are different. However, *núua kʰû: tɛʰăn taj paj lê:w jaŋ* is grammatical and meaningful; it does not refer to ‘Has my soulmate already died, or has not been born yet?’, but it means ‘Has my soulmate died yet?’. In (17), it cannot be changed to *\*tʰə: mâj jà:k mi: kraj jaŋ* because the negative marker *mâj* appears in the first alternative, and noun phrases following the verbs in both alternatives are different.

#### 4 Motivation behind functional extension

Consider the construction behind an alternative marker *rǔ: ‘or’*. To distinguish *jaŋ* before and after change, let *jaŋ*<sub>1</sub> ‘still’ be *jaŋ* prior to the change, and let *jaŋ*<sub>2</sub> be *jaŋ* after the change, as in (18).

- (18) a. VP + [*rǔ:*] + [*jaŋ*] + [*mâj*] + VP > VP + [*rǔ:*] + [*jaŋ*]  
 b. [*jaŋ*<sub>1</sub>] + [*mâj*] + VP > [*jaŋ*<sub>2</sub>]

*Remark:* VP stands for verb phrase; the square bracket [...] stands for word boundary; *mâj* is a negative marker in Thai.

Based on Bybee et al. (1994:285-289), it might be argued that the motivation behind the development from [*jaŋ*<sub>1</sub>] + [*mâj*] + VP > [*jaŋ*<sub>2</sub>] is inference because the meaning of [*jaŋ*<sub>1</sub>] + [*mâj*] + VP has already been conventionalized to the coded meaning of *jaŋ*<sub>2</sub>.

- (19) *kʰǎw ta:j paj lé:w rǔ: jaŋ₂ tʰâ: taj paj lé:w kô?*  
 he die go PFV or jaŋ if die go PFV then  
*mòd-rú:ŋ-mòd-ra:w tɛ: tʰâ: jaŋ₁ māj ta:j hāj du: wâ:*  
 end a story but if still NEG die let see  
 COMP  
*kʰǎw mi: mia mi: lû:k rǔ: plàw*  
 he have wife have kid or NEG  
 ‘Has he died yet? If he has already died, you don’t have to do anything; conversely, if he hasn’t died yet, please check whether he has a family or not.’

In (19), *jaŋ₂* refers to *jaŋ₁ māj ta:j* ‘(he) hasn’t died yet’ (*[jaŋ₁] + [māj] + VP*) because the speaker requires the listener to select between *ta:j paj lé:w* ‘(he) has already died’ and *jaŋ₂*. Subsequently, the speaker gives the listener more information so that if the listener selects the first choice *ta:j paj lé:w* ‘(he) has already died’, the listener does not have to do anything. However, if the listener selects the second choice *jaŋ₂* in which the meaning is given in the latter sentence—*jaŋ₁ māj ta:j* ‘(he) hasn’t died yet’, the listener should verify whether he has a family.

- (20) *tʰə: tʰam ka:n-bâ:n sèt rǔ: jaŋ tʰâ: jaŋ kô? rî:p*  
 you do homework finish or jaŋ if jaŋ then immediately  
*tʰam jà: mua tɛ: ?ò: ?è:*  
 do NEG only for dawdle  
 ‘Have you finished your homework yet? If not, do it immediately. Don’t dawdle.’

- (21) Q: *Suda klàp ma: rǔ: jaŋ*  
 Suda back come or jaŋ  
 ‘Has Suda come back yet?’

A: jaŋ  
jaŋ  
 ‘Not yet.’

Over time, *jaŋ₂* indicating *jaŋ₁ + māj + VP* ‘still not + VP’ has been used so frequently that ‘still not + VP’ is conventionalized to be the coded meaning of *jaŋ₂*, such as in examples (20) and (21). In (20) and (21), the listener can infer that *jaŋ* means ‘(you) have not finished your homework’ and ‘(Suda) has not come back yet’, respectively, even though the speaker does not need to offer further information to the listener about what *jaŋ* signifies.

## 5 Mechanisms of functional extension

According to Traugott’s (2011) framework, the mechanisms of change includes reanalysis, analogy, and repetition. It is very likely that the mechanism in the initial stage of functional extension from the aspectual marker *jaŋ* to the interrogative marker *jaŋ* is repetition. One thousand five hundred sampled sentences in which *jaŋ* appeared from the Thai National Corpus (TNC) were collected and examined. It should be noted that there were four types of construction used for asking whether a state has already occurred or whether an action has already been accomplished in Thai, as given in Table 1.

Table 1 shows statistically significant differences in token frequency for each construction. Findings indicate that *VP + rǔ: jaŋ* has the highest token frequency (52.83%) when compared with other constructions. Bybee (2003:604) and Bybee and Thomson (1997, as cited in Hopper and Traugott, 2008:127-128) argued that high token frequency triggers some changes that take place in grammaticalization, i.e., semantic fading, erasure of word boundaries, and phonological reduction; furthermore, such effects reveal the fact that frequently used forms are eroded at a faster rate than less

frequently used forms. For this reason, the high token frequency of construction VP + *rǔ: jaŋ* plays an important role in the occurrence of the erasure of word boundaries and phonological reduction.

**Table 1:** Token frequency of each construction used for asking whether a state has already occurred or whether an action has already been accomplished in Thai.

Constructions	Token Frequencies	Percentage (%)
VP + <i>rǔ: jaŋ mâj</i> + VP	2	3.78
VP + <i>rǔ: jaŋ</i>	28	52.83
VP + <i>rǔ jaŋ</i>	12	22.64
VP + <i>jaŋ</i>	11	20.75
<b>Total</b>	<b>53</b>	<b>100</b>

## 6 Effects of high token frequency

While the aspectual marker *jaŋ* extended its function, token frequency of the grammaticalizing construction VP + *rǔ: jaŋ* appeared at a much higher rate than the construction VP + *rǔ: jaŋ mâj* + VP, which led to erasure of word boundaries and phonological reduction.

### 6.1 Erasure of word boundaries

The frequent use of the alternative marker *rǔ:* ‘or’ with the form *jaŋ* ‘still not VP’ in the construction VP + [*rǔ:*] + [*jaŋ*]<sup>2</sup> caused the word boundary between *rǔ:* and *jaŋ* to disappear; in other words, these two words—[*rǔ:*] and [*jaŋ*—have fused to become a single word [*rǔ: jaŋ*]. This is consistent with Brinton and Traugott’s concept of “the unification, or univerbation of a syntactic phrase or construction into a single word”—lexicalization as fusion (Brinton and Traugott 2005:48).

There is an important piece of evidence that [*rǔ: jaŋ*] has become a full-fledged single unit in itself; namely, it was fossilized into an inseparable unit.

- (22) *tʰə: tʰam ka:n-bâ:n lé:w [rǔ: jaŋ]*  
 you do homework PFV or not yet  
 ‘Have you finished your homework yet?’

- (23) *tʰə: tʰam ka:n-bâ:n lé:w [rǔ:] [jaŋ] mâj tʰam ka:n-bâ:n*  
 you do homework PFV or still NEG do homework  
 ‘Have you finished your homework yet?’

Examples (22) and (23) are identical in meaning; nevertheless, the complementizer *wâ:* cannot be inserted between *rǔ:* and *jaŋ* in (22), whereas it can be inserted between *rǔ:* and *jaŋ* in (23), as shown in (24) and (25), respectively.

- (24) \* *tʰə: tʰam ka:n-bâ:n lé:w [rǔ: wâ: jaŋ]*

- (25) *tʰə: tʰam ka:n-bâ:n lé:w [rǔ:] wâ: [jaŋ] mâj tʰam ka:n-bâ:n*  
 you do homework PFV or COMP still NEG do homework  
 ‘Have you finished your homework yet?’

Moreover, [*rǔ: jaŋ*] is regarded as lexicalization by uniting [*rǔ:*] ‘or’ and [*jaŋ*] ‘still not VP’, which correlates well with Phanthumetha’s notion that [*rǔ: jaŋ*] is a single word belonging to *supplementary word*, used for inquiring whether an action has occurred. for example, *kʰǎw paj rǔ: jaŋ* ‘Has he gone yet?’ (Phanthumetha 2006).

<sup>2</sup> The square bracket [...] stands for word boundary.



## 6.2 Phonological reduction

The developments, both from *rű: jaŋ* to *rű jaŋ* and from *rű jaŋ* to *jaŋ*, are phonological reduction, having an order of change as follows.

Luksaneeyanawin (1993 as cited in Tumtavitikul 1998:63) argued that a 2-syllable-word in Thai has an *unstressed-stressed* pattern on the basis of Metrical Phonology such as *sàʔtìʔ* ‘mental, mind’ (the first syllable *sàʔ* is unstressed, and the second syllable *tìʔ* is stressed). Tumtavitikul (2004:110) observed that vowel reduction, in Thai, mainly occurs in the unstressed syllable. A final glottal deletion, together with tone neutralization to mid tone, always precedes vowel reduction, such as *sàʔtìʔ* > *sàtìʔ* > *sàtìʔ* > *sətìʔ*. Similarly, it is likely that a 2-syllable-word *rű: jaŋ* undergoes vowel reduction to become *rű jaŋ*, i.e., *rű: jaŋ* > *rű jaŋ*.

The latter step, which occurred after vowel reduction, is weak syllable deletion—omission of a weak (unstressed) syllable preceding a stressed syllable. In Thai, there are many examples in line with these two steps, i.e., first, there was vowel reduction, then weak syllable deletion, e.g., *ʔan raj* > *ʔaʔ raj* > *ʔa raj* > *raj* ‘what’ (Warotamasikkhadit 1999 as cited in Tikhachunhatian 2002), *bàt dīaw* > *pràʔ dīaw* > *pra dīaw* > *dīaw* ‘for a moment, for a little while’ (Kullavanijaya, 2002), and *[pen] + [jà.ŋ raj]* > *pen jaŋ ŋaj* > *pen ŋaj* > *ŋaj* ‘how are you?’ (Pothipath 2014). Accordingly, *rű jaŋ*, which originated from vowel reduction of *rű: jaŋ*, underwent weak syllable deletion to become *jaŋ* towards the end. The final outcomes, *raj*, *dīaw*, *ŋaj*, and *jaŋ*, are new words or new functions stored in the user’s lexicon. However, these new words or new functions are mostly used in casual or informal speech; for example, (26) is more formal than (27).

(26) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *lɛ́:w*      [*rű: jaŋ*]  
 you      eat      breakfast      PFV      or not yet

(27) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *jaŋ*  
 you      eat      breakfast      INT  
 ‘Have you had breakfast yet?’

## 7 How to verify occurrence of functional extension

According to Hopper’s layering principle (Hopper 1991), there were four forms found in sample sentences in which *jaŋ* appeared from the TNC. They have been used for inquiring as to whether a state has already occurred or whether an action has already been accomplished, i.e., ‘VP + *rű: jaŋ mâj* + VP’, ‘VP + *rű: jaŋ*’, ‘VP + *rű jaŋ*’, and ‘VP + *jaŋ*’, as shown in (28).

(28a) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *lɛ́:w*      *rű: jaŋ mâj dâj kin kʰâ:w-tɛʰáw*  
 (28b) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *lɛ́:w*      *rű: jaŋ*  
 (28c) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *lɛ́:w*      *rű jaŋ*  
 (28d) *tʰə:*      *kin*      *kʰâ:w-tɛʰáw*      *lɛ́:w*      *jaŋ*

All of these four forms in (28a)-(28d) convey the analogous meaning ‘Have you had breakfast yet?’, but the distinction among them is stylistic alternative; namely, it is intuitively obvious for a speaker whose first language is Thai to perceive the contrast between the form ‘VP + *rű: jaŋ mâj* + VP’ as having the most formal style (as in 28a.) and the form ‘VP + *jaŋ*’ as having the least formal style (as in 28d.). Because of the occurrence of the layering phenomenon (all of these four forms coexist in the modern Thai language), it is feasible to confirm that functional extension from the aspectual marker *jaŋ* to the interrogative marker *jaŋ* is taking place.

## 8 Discussion

When a continuative aspect marker *jaŋ* becomes an interrogative marker *jaŋ* through the process known as functional extension, it undergoes syntactic and semantic change. Syntactically, interrogative marker appears only at the end of the sentence, even though it has to lose its previous position being before a

verb phrase. Semantically, the meaning relating to the persistence of a state or an activity develops into the function asking about whether something has already occurred.

During the occurrence of these changes, there were some analogous phenomena which might be found not only in Thai, but also in other Southeast Asian languages. For instance, *ñang* in Lao is used as an aspectual-modal, it means ‘still’ or ‘(not) yet’ such as the example in (29). In addition, when *ñang* is used as a stand-alone answer, it conveys the meaning of ‘not yet’ despite not taking negative marker *bòø*, as is exemplified in (30) A1. On the other hand, *ñang* in (30) A2 is used in combination with a verb phrase, it must co-occur with the negative marker *bòø* so as to convey the meaning of ‘not yet’ (Enfield 2007:207-208).

(29) *fèèn*                      *laaw*    *kaø*                      *ñang*    *maa*    *juu*    *vaa*  
 sweetheart              she              T.LNK                      still              come              PROG              INT  
 ‘So does her sweetheart still come?’

(30) Q:              *khaw*    *suk*              *lèèw*    *vaa*  
 rice              cooked              PFV              INT  
 ‘Is the rice cooked already?’

A1:              *ñang*  
 not yet  
 ‘Not yet’

A2:              *ñang*    *bòø*              *suk*  
 still              NEG              cooked  
 ‘It is still not cooked.’ (i.e. ‘it is not cooked yet.’)

*ñang* in Lao is similar to *jaŋ* in Thai in that both of them are able to signify ‘not yet’ in spite of not taking the negative maker when they are used in isolation. It may be assumed that such a phenomenon, perhaps, arose from inference; in other words, *ñang* + *bòø* + VP has occurred together so often that it has been conventionalized to be the coded meaning of *ñang*. Nevertheless, *ñang* in Lao does not carry on developing into an interrogative marker like *jaŋ* in Thai.

## 9 Conclusion

In summary, this paper has investigated the motivations and linguistic mechanisms of functional extension from the continuative aspect marker *jaŋ* to the interrogative marker *jaŋ*. The results of this study suggest that the motivation for functional extension was inference and the mechanism in the beginning stage of the development was repetition. Based on Bybee’s notion (Bybee, 2003), the effects of high token frequency are conducive to the disappearance of word boundaries, i.e., [*rú:*] + [*jaŋ*] > [*rú: jaŋ*], and phonological reduction, i.e., [*rú: jaŋ*] > [*rú jaŋ*] > [*jaŋ*].

However, only if the continuative aspect marker *jaŋ* appears in the following specific contexts will this change to the interrogative marker *jaŋ* be allowed to occur.

As for the first specific context, the aspectual marker *jaŋ* must occur in the second selection of alternative interrogative sentences; besides this, the two verbs that appear in both selections must be identical. Secondly, the negative marker *mâj* must appear only in the second selection. More crucially, it must not appear in the first selection. Lastly, when the verbs appearing in both selections are action verbs, the perfective marker *lè:w* must appear in the first selection. Moreover, the evidence from the layering phenomenon supports the notion that functional extension from the aspectual marker *jaŋ* to the interrogative marker *jaŋ* currently occurs.

### Capitalized abbreviations

COMP	complementizer
INT	interrogative marker
NEG	negative
NOM	nominalizer
PFV	perfective aspect marker
PROG	progressive marker
PT	particle
T.LNK	topic linker
VP	verb phrase

### References

- Boonyapatipark, Tasanalai. 1983. *A study of aspect in Thai*. Doctoral dissertation, University of London.
- Brinton, Laurel J., and Elizabeth C. Traugott. 2005. *Lexicalization and Language Change*. Cambridge, UK: Cambridge University Press.
- Bybee, Joan L. and 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.
- Bybee, Joan. 2003. Mechanisms of change in grammaticalization: The role of frequency. In *Handbook of Historical Linguistics*, ed. by Brian D. Joseph and Richard Janda, 602-623. Oxford: Blackwell Publishers.
- Evans, Vyvyan and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Enfield, Nick J. 2007. *A Grammar of Lao*. Berlin, Germany: Walter de Gruyter & Co Press.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
- Hopper, Paul J. On some principles of grammaticalization. In *Approaches to Grammaticalization* Vol. I, ed. by Elizabeth Closs Traugott and Bernd Heine, 17-36. Amsterdam: John Benjamins.
- Hopper, Paul J., and Elizabeth C Traugott. 2008. *Grammaticalization*. Cambridge, England: Cambridge University Press.
- Jaratjarungkiat, Sureenat. 2008. */jay/: A diachronic study*. Master's Thesis, Chulalongkorn University. (In Thai)
- Kullavanijaya, Pranee. 2002. *Save little by little*. Nakhon Pathom: Silpakorn University Press. (In Thai)
- Lessau, Donald A. 1994. *A Dictionary of Grammaticalization*. Bochum: Universitätsverlg Dr. N. Brockmeyer.
- Michaelis, Laura A. 1993. 'Continuity' within three scalar models: the polysemy of adverbial still. *Journal of Semantics* 10:193-237.
- Nicolle, Steve. 1998. A relevance theory perspective on grammaticalization. *Cognitive Linguistics* 9(1):1-35.
- Oxforddictionaries.com [homepage on the Internet] Oxford University Press.  
<https://en.oxforddictionaries.com/usage/shall-or-will> [accessed on April 13, 2017].
- Pattaranavik, Prapatsorn. 2008. Grammaticalization of *jay* in the Sukhothai period. *Parichart Journal* 20(2):36-53. (In Thai)
- Phanthumetha, Nawawan. 2006. *Thai Grammar*. Bangkok: Chulalongkorn University Press. (In Thai)

- Pothipath, Vipas. 2014. Lexicalization in Thai. *Journal of Language and Linguistics* 32(2):1-24. (In Thai)
- Swan, Michael. 1995. *Basic English Usage*. Oxford & New York: Oxford University Press.
- Tansiri, Kachen. 2011. *Interactions between grammatical aspect and lexical aspect in Thai: a corpus-based study*. Doctoral dissertation, Chulalongkorn University. (In Thai)
- Thepkanjana, Kingkarn. 2016. *Grammaticalization*. Bangkok: Chulalongkorn University Press. (In Thai)
- Tikhachunhatian, Rungarun. 2002. *Indefinite pronoun and interrogative pronoun: A diachronic study*. Doctoral dissertation, Chulalongkorn University. (In Thai)
- Traugott, Elizabeth C. 2011. Grammaticalization and mechanisms of change. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbooks of Grammaticalization*. Oxford handbooks online. doi:10.1093/oxfordhb/9780199586783.013.0002
- Tumtavitikul, Apiluck. 1998. The Metrical Structure of Thai in a Non-Linear Perspective. *Papers presented to the Fourth Annual Meeting of the Southeast Asian Linguistics Society 1994*, ed. by Udom Warotamasikkhadit and Thanyarat Panakul, 53–71. Tempe, Arizona: Program for Southeast Asian Studies, Arizona State University.
- Tumtavitikul, Apiluck. 2004. *Phonology: Systematic Organization of Sounds in Languages*. Bangkok: Kasetsart University Press. (In Thai)
- Van der Auwera, J. 1993. Already and still: beyond duality. *Linguistics and Philosophy* 16:613-653.
- Yeh, Meng. 1998. On *hai* in Mandarin. *Journal of Chinese Linguistics* 26(2):236-280.
- Yule, George. 1996. *Pragmatics*. Oxford: Oxford University Press.

# TENSE, ASPECT, AND RELATED MARKERS IN LUNGMI RAWANG<sup>1</sup>

Rachel Powelson  
Payap University  
*rachel\_powelson@sil.org*

## Abstract

This paper describes how tense and aspect are marked in Dinglak, a variety of Lungmi Rawang, of the Tibeto-Burman language family. It also touches on some related concepts, including modality and directional marking. These are marked in two ways in Lungmi: first, with a group of post-verbal particles called “orientation markers”, and second, with clause-level structures such as cleft constructions containing nominal clauses. The temporal orientation markers in Lungmi are similar in function to those found in other Rawang varieties but have different forms. Cleft constructions, consisting of a nominal clause and a copula, can have tense, aspect, and modality implications in Lungmi that are similar to those in other Rawang varieties.

**Keywords:** tense, aspect, modality, directionality, language comparison, Nungish  
**ISO 639-3 codes:** raw

## 1 Introduction

Lungmi (also called Longmi or Nung Lungmi) is spoken by approximately 10,000 people<sup>2</sup> in Kachin State, Myanmar. It is a Tibeto-Burman language of the Nungish cluster, considered to be a variety of Rawang [ISO: raw]. There are significant lexical and phonological features that set Lungmi apart from other Rawang varieties. Based on my observations and others', intelligibility between Lungmi and other Rawang varieties seems low.

This paper will describe how tense and aspect are marked in Dinglak, a variety of Lungmi Rawang. It will also touch on some related concepts, including modality and directional marking. These are marked in two ways in Lungmi: first, with a group of post-verbal particles called “orientation markers”, and second, with clause-level structures such as cleft constructions with nominal clauses. This paper will also compare the structures in Lungmi with what these linguists have found by previous researchers in Matwang, Sinwal, and Rvmøl Rawang. There is a good deal of grammatical similarity among these varieties, despite their lexical and phonological differences.

The data used in this paper comes from Dinglak Lungmi (also written as Dingra or Dingraq), a variety spoken in the villages of Dinghket and Namhkam in the Machanbaw Township of Kachin State. The estimated population of Dinglak speakers is 1,000. The data was collected from eight different Dinglak speakers in Myitkyina, Myanmar and Chiang Mai, Thailand. In Myitkyina, approximately 10 minutes of spoken expository and narrative text were collected for use in this study, in addition to 250 sentences elicited directly through translation in Burmese. Later, in Chiang Mai approximately 25 hours were spent interviewing and collecting additional data from the speakers in Chiang Mai.

---

<sup>1</sup> More information about the verb phrase structures described in this paper can be found in my thesis, which will be published through Payap University in 2017 or 2018.

<sup>2</sup> This is my own estimation based on previous estimates of Lungmi and Rawang population. In 1997, Bradley estimated the Lungmi population as 30,000, approximately 17% of the total Rawang population, 172,000. In 2017 the Ethnologue (Simons 2017) estimates for the total Rawang population as 63,000. If the proportion of Lungmi population within the Rawang group has stayed consistent, we can estimate the modern Lungmi population to be 10,000 based on the Ethnologue's estimate. However, many factors, including migration and ongoing conflict, may have affected the Lungmi population in the two decades since Bradley's estimate, so this should be treated as a very rough estimation.

Lungmi is part of the Rawang language group, which consists of between 70 and 100 varieties (Sarep 1995). These dialects are further categorized into 4 groups: Matwang, Daru-Jerwang, Tangsar, and Lungmi. Morse described the Rawang group as “a gold mine yet to be tapped by comparative linguists” (1988:239). The Matwang variety of Rawang has been well documented by scholars such as Morse (1965), LaPolla (2006, 2010)<sup>3</sup>, and Sarep (1995)<sup>4</sup>, but there has been very little documentation on the other varieties, especially Lungmi. The most extensive description of a non-Matwang variety to date has been written by Straub (2016), who studied Rvmø1, a variety of Daru.

## 2 Order of verb phrase constituents

Tense and aspect in Lungmi are marked primarily by clitic-like particles in the verb phrase. In Sarep’s description of Sinwal Rawang (1995), he calls these “orientation” markers, dividing them into two groups: “temporal orientation” and “spatial orientation” markers (p. 57). Orientation particles can indicate either tense or directional meanings, sometimes both. Temporal orientation refers to tense and aspect markers. Spatial orientation refers to location and directional markers.

Lungmi clauses are verb final. The verb phrase can be followed by a sentence-final particle or subordinator. These are not considered a part of the verb phrase because they operate on a clause level, but they play important roles in the semantics of the clause and verb phrase. The following structure outlines show the internal structure of the Lungmi verb phrase and verb stem.

**VP** → [Adverb **V<sub>STEM</sub>** =benefactive =3rd.intransitive =**orientation** =person/number]

**V<sub>STEM</sub>** → [non1st.subject– negative– causative– <imperative>**V<sub>ROOT</sub>** –1st.singular ]

The verb phrase consists of a verb stem, preceded by an optional adverb and followed by several post-verbal particles, which are marked as clitics in this paper. The verb stem consists of a verb root with affixes. This article focuses mostly on orientation markers. Information about other particles and affixes in the Lungmi verb phrase can be found in Powelson (2018).

In this description of Lungmi, the distinction between affixes and clitics or particles is based on the degree of phonological influence between the marker and the verb root. Benefactive, 3rd person intransitive, orientation, and person/number markers may undergo minor phonological changes based on their environment, but they are largely phonologically independent from the verb root.

Example (1) shows a complete clause in Dinglak Lungmi, with a subject, two objects, a verb phrase, and a sentence-final question particle. The verb phrase is marked in bold.

- (1) nâ-ní-ē      âŋ-ū      dāŋ      gāmbāŋ      ə-dzǐ      =əgiʔ      =sā      nâ      ʔ<sup>5</sup>  
 1pers-dual-AG 3pers-PT    DAT    money    **n1-give**    =**PF.rec**    =**D/P**    **Y/N**  
 ‘Did you two already give him money?’

## 3 Orientation markers in Dinglak Lungmi

Five orientation markers have been identified in Lungmi. These are shown in table 1. A Lungmi verb phrase may have either a temporal or a spatial orientation marker, but not both. The same rule applies in other Rawang varieties (Sarep 1995; Straub 2016).

<sup>3</sup> LaPolla has written many other papers about Matwang. The bibliography section of this paper only contains a few that are related to this study.

<sup>4</sup> Sarep documented the Sinwal variety of Rawang, which is closely related to Matwang.

<sup>5</sup> The Lungmi data in this paper is presented in an IPA orthography with some modifications for readability. Because there is only a two-way contrast between aspirated and unaspirated stops and affricates, their voiced counterparts are used to represent unaspirated voiceless stops. *y* is used to represent the palatal approximant /j/, and the vowel /ɐ/ is represented by *ɹ* to avoid inconsistencies when switching between fonts.

**Table 1: Tense, Aspect, and Direction Particles**

Particle	Meaning	Gloss
(ə)giʔ	Perfective, Recent Past	PF.rec
(ə)lɔk	Perfective, Middle Past	PF.mid
(ə)yāŋ	Perfective, Remote Past	PF.rem
(ə)lɔŋ	Cislocative directional	DIR
(ə)lak	Cislocative directional	DIR

Each of these particles has an allomorph that omits the initial weak vowel when preceded by a word ending in a mid-tone vowel. Example (2) shows the allomorph of *əgiʔ* [giʔ] that occurs after the intransitive marker *-ī*.

- (2)    tʃén    =ī        =giʔ  
        go     =3:I     =PF.rec  
        ‘(He/she) went.’

A verb phrase with no orientation marker can usually be understood as imperfective, as shown in example (2).

- (2)    ŋâ     dzén     =l̄  
        1pers go     =SG  
        ‘I am going.’

### 3.1 Perfective temporal orientation markers

There are three perfective markers with semantic distinctions based on how far in the past the event occurred: *əgiʔ* ‘recent perfective’, *əlɔŋ* ‘mid-distant perfective’, and *əyāŋ* ‘remote past’.

The recent past perfective marker *əgiʔ* can mark events that happened recently or several days ago. It is not used for events in the remote past. Example (3) shows this marker in use. In this example, the 1st person singular affix *ŋ*<sup>6</sup> as been added to the perfective marker because the subject of the verb phrase is 1st person singular.

- (3)    ŋâ-i     gəlóm    ām        =əgi<ŋ>ʔ  
        I-AG    all        eat        =PF.rec<1SG>  
        ‘I ate all (the food).’

The origin of this morpheme is unknown. It shares a similar vowel with the perfective marker [bu] found in other Rawang varieties, as described in the next section, but the rest of the word seems dissimilar to the perfective markers found in any other Rawang variety.

The next temporal orientation marker in Lungmi is *əlɔk*, which marks events that took place one or more days ago, but not more than a few months ago. Example (4) shows this marker in use.

- (4)    âŋ        nā        gat        dāŋ     dzén        =əlɔk .  
        3pers    TP        market    DAT     go         =PF.mid  
        ‘He went to the market (yesterday or before).’

The ranges of use of *əlɔk* and *əgiʔ* overlap slightly. They can both be used to describe events that happened just a few days ago, but *əlɔk* seems to mark events that are less relevant to the current situation. The idea of a distinction based on current relevancy is based on Straub’s description of Rvmøl Rawang, where he described the perfective marker *bu*, which is similar in distribution to *əgiʔ* in Lungmi, as marking both completion and current relevance (Straub 2016:111). Straub found that the Rvmøl

<sup>6</sup> *-ŋ* is usually a suffix, but when occurring on a glottalized syllable, it is pronounced before the glottal stop. For that reason, it has been marked here as an infix.

perfective marker *bu*, like *əgiʔ* in Lungmi, can be used to mark events that happened a few days or a week ago, as shown in example (5).

- (5)    *sá:t-bú*        =wà  
           kill:3-PFV    =hearsay  
           ‘(he) killed, it is said (not long ago, possibly a day or a week ago)’ -Rvmø1 Rawang (Straub 2016:110)

The third temporal orientation marker in Lungmi is the remote past marker *əyāŋ*. It can mark events that took place from several months ago to many years ago. This particle is generally used in storytelling, both in traditional stories and in personal stories that happened in the remote past. Example (6) is from a story about an experience the speaker had several years ago.

- (6)    *ənôŋ*            *p<sup>h</sup>ly- ē*            *sīlot*        =*əsā*        =*əyāŋ*  
           Lisu            male-AG        rescue       =D/P        =PF.rem  
           ‘A Lisu man rescued us.’

The verb in (6) is followed by two particles. The inverse dual/plural particle *əsā* is used because the syntactic patient of the verb is 1st person plural. More information about person and number marking in Lungmi can be found in Powelson (2018).

### 3.2 Comparison with past-tense marking in other Rawang varieties

This section compares the temporal orientation markers from Lungmi with similar markers from three other varieties of Rawang: Matwang, as documented by LaPolla (2015) and LaPolla and Poa (2001); Sinwal, a variety closely related to Matwang, as described by Sarep (1995); and Rvmø1, a variety related to Daru, as described by Straub (2016). Straub’s Rvmø1 data, like the Lungmi data in this paper, is presented in a modified IPA orthography. All Matwang and Sinwal data is presented in standard Rawang orthography, followed by the IPA transcription for examples where the phonological form is important.

**Table 2:** *Nungish Tense Particles*

Particle	Variety	Meaning	Source
<i>əgiʔ</i>	Lungmi	perfective, recent	collected data
<i>bu</i>	Rvmø1	perfective	Straub 2016
<i>bó</i> [bú]	Matwang, Sinwal	perfective	LaPolla and Poa 2001
	Sinwal (Matwang)	toward-remote	Sarep 1995
<i>zó</i> [zú]	Sinwal (Matwang)	unknown-remote	Sarep 1995
<i>dír</i> [dér]	Matwang	hours past	LaPolla 2015
(ə)lónŋ	Lungmi	perfective, mid-distant	collected data
<i>ri</i>	Rvmø1	time-days	Straub 2016
<i>ap</i>	Matwang	time-days	LaPolla 2015
	Sinwal (Matwang)	near-remote	Sarep 1995
(ə)yāŋ	Lungmi	perfective, remote	collected data
<i>yáŋ</i>	Rvmø1	time-years	Straub 2016
<i>yàng</i>	Matwang	time-years	LaPolla 2015
	Sinwal (Matwang)	far-remote	LaPolla 2015

These three varieties of Rawang have documented a three-way or four-way distinction between recent and distant events. Table 2 summarizes the perfective markers found in Lungmi along with the tense markers found in Matwang and Rvmø1, along with their meanings as described by previous researchers. As Table 2 shows, the remote marker is the most phonologically consistent across varieties. The mid-distant perfective marker *əlónŋ* in Lungmi and the time-days marker *ri* in Rvmø1 also show



possible phonological relation, because the shift from [r] in other Rawang varieties to [l] in Lungmi is very consistent.

Straub (2016) describes three past tense particles in Rvmø1 Rawang: perfective *bu*, days-past *ri*, and years-past *yāη*. While the forms of these particles are different in Rvmø1, they seem to have the same ranges of meaning and distribution as the three perfective markers in Lungmi. *bu* is equivalent in meaning and distribution to *əgiʔ* in Lungmi, *ri* is equivalent to *ələk*, and *yāη* is equivalent to *əyāη*.

LaPolla describes Matwang Rawang as having a four-way distinction in past-tense marking. Like Rvmø1, Matwang has a perfective morpheme *bó* or *bō* [bú or bū], a days-past marker *ap* or *vp* [əp], and a years-past marker *yàng* (LaPolla and Poa 2001, p. ix). In addition, LaPolla also describes an hours-past marker *dár* or *dvr* [dár] (LaPolla 2015:41). Example (7) shows the Matwang hours-past marker *dvr* in use. In this example the speaker is saying that he dug up a bamboo rat earlier, and now he is giving the listeners (supposedly) the bamboo rat to eat.

- (7) kām-bōng svrò-cè dòng **dvr**-à  
 kind.of.bamboo-section rat-DIM dig **TMhrs-TNP**  
 ‘I dug up a bamboo rat.’ -Matwang Rawang (LaPolla and Poa 2001:67)

Sarep’s description of Sinwal Rawang is very similar to LaPolla’s description of Matwang, although Sarep does not include *dvr*. Also, Sarep says these particles mark only distance from the present, and that distance may be either in the past or the future. He gives one example of the particle *bō* [bū] being used in a future-tense verb phrase, in example (8).

- (8) kū kaq dī -ng -bō -ng -ē  
 there to go 1s t-rmt 1s imprf  
 ‘I will go there.’ -Sinwal Rawang (Sarep 1995, p. 59)

Sarep also identifies a second particle *zó*, which has similar meaning to *bó* but marks events that the speaker was unaware of until recently (1995:59). No such particle has been identified in Lungmi, so far.

While the Lungmi particles discussed in the previous section are very close in meaning to the other Rawang particles, there are some differences in analysis. First, all three morphemes in Lungmi have been labeled as perfective. While Sarep’s research indicated that orientation particles in Sinwal may mark actions that have not been completed yet, every example of these particles in Lungmi, found thus far, has been in a perfective verb phrase. Until evidence is found that the Lungmi particles *əgiʔ*, *ələk*, and *əyāη* can be used in non-perfective verb phrases, it seems most appropriate to label all three of them as perfective.

Secondly, the analysis presented in this paper has omitted the use of words like *day* and *year* in the labeling of these morphemes, in favor of vaguer words like *recent* and *remote*. This is because the precise meaning of each of these morphemes can change based on context. For example, the morphemes *əgiʔ* in Lungmi and *bu* in Rvmø1 usually mark events that took place within the past day, but they can sometimes mark events that took place a few days ago, as long as those events are still relevant in the present. Straub also found that the morpheme *ri* in Rvmø1, which usually marks events that took place at least 1 day ago, can be used for more recent events if that event is being contrasted with an event, marked with *bu*, which took place later (2016:116). Therefore, the labels of recent past, mid-distant past, and remote past more accurately capture these meanings than labels based on hours, days, and years.

### 3.3 Spatial orientation markers

In addition to temporal orientation markers, Lungmi also has two spatial orientation markers, also known as directional markers. The particles *ələk* and *əlóη* indicate direction towards a point of reference. These two morphemes are similar in meaning, and it has been difficult to determine their semantic differences. Further semantic analysis of these morphemes using a larger corpus of text would be useful.

Both particles can be used to mark both perfective and imperfective verb phrases. Examples (10) and (11) show these two particles in use.

- (10) ηâ dzén =əłŋ  
 1pers go =DIR  
 ‘I am coming / almost there.’

In example (10), the particle *əłŋ* marks an imperfective verb phrase. In example (11), the particle *əlak* marks a perfective verb phrase.

- (11) ηâ dzén =əlak  
 1pers go =DIR  
 ‘I have come / already arrived.’

However, *əlak* can also be used to mark imperfective events, as shown in example (12).

- (12) ηâ-ndŋy-ē nâ-u dāŋ sədŋy gŋmbŋŋ dzí =əlak =nŋŋ dī  
 1pers-PL-AG 2pers-PT DAT tomorrow money give =DIR =2PL  
 affirmation  
 ‘We-pl will give you-sg money tomorrow.’

Then, in example (13), (*ə*)*łŋ* is used to mark a completed event.

- (13) âŋ-ē ləŋuʔ sí lû =łŋ  
 3pers-AG banana bring =DIR  
 ‘She brought bananas.’

Based on the previous four examples, it cannot be said that *əlak* and *əłŋ* consistently mark tense or aspect in verb phrases. Rather, the meaning seems to be based on directionality and motion.

The middle-distance perfective marker *əlak* is interesting because of its phonological similarity to directionality markers, which reflects Straub’s research on tense and directionality markers in Rvmøł (2016).

Straub’s work focused on the historical relation between directionality markers and past tense markers in Rvmøł and notes the phonological similarity between the mid-distance past marker *ri* and cislocative markers *ra*, *re*, and *rvt*. The phonological similarity between Lungmi’s two cislocative directional markers *əlak* and *əłŋ* and the perfective marker *əlak* directly reflects Straub’s observations in Rvmøł and confirms his assertion of a historical semantic connection between tense markers and cislocative markers.

Straub (2016) and Sarep (1995) have also documented more specific directional markers in Rvmøł and Sinwal, such as the downward directional markers *zaq* [zaʔ] or *zak* in Sinwal and *zaʔ* or *zək* in Rvmøł and the upward directional markers *nəŋg* in Sinwal and *luŋ* in Rvmøł. No such directional markers have been found in Lungmi.

#### 4 Intransitive 3rd-person marker *ī*

While the intransitive 3rd person marker *ī* is not a tense or aspect marker, its use is closely tied to that of perfective and directionality markers. When past-tense markers and directionality markers are used in an intransitive verb phrase with a 3rd person subject, they are usually accompanied by the 3rd person intransitive marker *ī*. When this marker is used, the optional /ə/ at the beginning of the orientation marker is usually dropped, as it typical when those particles follow an open syllable.

- (17) âŋ nā gat dāŋ dzén =ī =giʔ  
 3pers TP market DAT go =3:I =PF.rec  
 ‘He went to the market.’

This marker is only used if all three of the following circumstances are met: 1) there is a time or directional orientation marker present; 2) the clause is intransitive; and 3) the subject of the clause is 3rd person.

In example (18), the use of  $\bar{t}$  is ungrammatical because there is no perfective or directional marker.

- (18)     $\hat{a}\eta$          $n\bar{a}$          $gat$          $d\bar{a}\eta$          $d\bar{z}\acute{e}n$      $=*i$          $=\bar{a}$         .  
          3pers    TP        market    DAT        go         $=3:I$          $=SG$   
          ‘He goes to the market.’ (ungrammatical)

In example (19), the use of  $\bar{t}$  is ungrammatical because the subject is not 3rd person.

- (19)     $\eta\hat{a}$          $gat$          $d\bar{a}\eta$          $d\bar{z}\acute{e}n$      $=*i$          $=g\acute{i}?$   
          1pers    market    DAT        go         $=3:I$          $=PF.REC$   
          ‘I went to the market.’ (ungrammatical)

Other Rawang varieties also have an intransitive marker  $-i$ , but in other varieties, this marker occurs after the orientation marker. Example (20) shows the use of this marker in Matwang.

- (20)     $m\grave{o}ng-k\grave{o}m$   $vd\acute{v}m$      $taq$          $dv-$          $k\acute{o}m$      $-sh\grave{i}$          $d\acute{a}r$          $-i$   
          all-meet    plain    LOC        CAUS-    meet    -R/M        TMhrs    -IP  
          ‘At Mongkom (All Meet) Plain they gathered everyone together.’ - Matwang Rawang (LaPolla and Poa 2001:74)

So far, this paper has focused on tense and aspect marking using orientation markers. The following section will look at other methods of marking tense, aspect, and modality.

## 5 Other tense-aspect related constructions

This section looks at constructions besides orientation markers that have tense and aspect implications. These include three types of constructions: reduplication, auxiliary verbs, and copula cleft constructions.

Reduplication of the verb root can be used to indicate repetition or continuation. In example (21), the verb  $y\acute{i}$  ‘go down’ is reduplicated to show ongoing motion.

- (21)     $n\bar{a}ms\acute{i}$          $y\acute{i}t$          $p^h\acute{\lambda}y$          $n\bar{a}$          $th\acute{i}?$   $k^h\acute{a}t$      $y\acute{i}$          $y\acute{i}$          $=s\bar{i}$          $d\bar{a}...$   
          fruit        pick        male        TP        again        go.down    go.down    =R/M        then...  
          ‘The man who was picking fruit came and came down (and then...).’

The auxiliary verbs  $d\bar{a}\eta$  ‘finish’ and  $d\bar{v}$ , glossed here as ‘FUT’ (future) can be used to mark completive aspect and future tense, respectively. The examples here are bracketed to show the proposed clause embedding structure.

The auxiliary verb  $d\bar{v}$  forms a future tense construction, as shown in example (22). LaPolla records a similar auxiliary verb  $d\acute{v}n$  in Matwang (2010:3).

- (22)    [ [  $\hat{a}\eta$          $n\bar{a}$          $gat$          $d\bar{a}\eta$         [  $d\bar{z}\acute{e}n$  ]<sub>VP</sub> ]<sub>S</sub>    [  $d\bar{v}$          $=\bar{a}$  ]<sub>VP</sub> ]  
          [ [ 3pers    TP        market    DAT        [ go ] ]        [ FUT        =SG ] ]  
          ‘He will go to the market.’

While a perfective marker can be used to mark a verb phrase as completed, the speaker can put more emphasis on the completion by embedding the clause within a clause with the main verb  $d\bar{a}\eta$  ‘finish’, as shown in (23).

- (23)    [ [  $t^h\bar{a}ml\acute{a}$      $n\acute{i}$                     [  $\bar{a}m$  ]<sub>VP</sub> ]<sub>S</sub>    [  $d\bar{a}\eta$          $=s\bar{i}$          $=g\acute{i}?$  ]<sub>VP</sub> ]  
          [ [ child    rice.cooked    [ eat ] ]        [ finish    =R/M        =PF.rec ] ]  
          ‘The child ate the rice already.’

The cleft construction [ [S]<sub>NOM</sub> *í*  $\bar{\lambda}$  ], in which a nominal clause is followed by a copula, can have various aspectual or modal meanings, depending on which nominalizer is used. The nominalizer *nā* marks probability or possibility. A cleft construction with this nominalizer is shown in example (24).

- (24) [ [ [  $\hat{a}\eta$       *ní*              *ām* ]<sub>S</sub> ·      **nā** ]<sub>NP</sub>    [ *í*      = $\bar{\lambda}$  ]<sub>VP</sub> ]  
 [ [ [ 3pers    rice.cooked    eat ]      **PSB** ]    [ COP =SG ] ]  
 ‘He will (probably) eat the rice.’

The nominalizer *lám* is used to mark intention. This is a very common construction in Lungmi conversation, used to express future plans and goals.

- (25) [ [ [  $\eta\hat{a}$ -i    *pān*      *sān* ]<sub>S</sub> ·      **lám** ]<sub>NP</sub> [ *í*      = $\bar{\lambda}$  ]<sub>VP</sub> ]  
 [ [ [ I-AG    plate    wash ]      **PUR** ]    [ COP =SG ] ]  
 ‘I will wash the dishes.’

The construction in (26) is interesting because, as the parenthesis around the copula *í* indicate, the copula is optional. Without the copula, the sentence would need to be analyzed as a single clause with a single verb phrase, making *lám* a verb phrase particle. This indicates that the nominalizer *lám* may be grammaticalizing in to a post-verbal particle. The variation in this construction warrants further investigation from a diachronic viewpoint.

- (26) [  $\eta\hat{a}$ -i      *pān*    [ *sān*    *lám*      = $\bar{\lambda}$  ]<sub>VP</sub> ]  
 [ I-AG      plate    [ wash    PUR      =SG ] ]  
 ‘I will wash the dishes.’

The nominalizer *-sak* is used to create a past-tense participle, giving the clause a completive sense. In example (27), the construction emphasizes the finished state of the event. The subject has already been to the market, finished his shopping, and returned.

- (27) [ [ [  $\hat{a}\eta$       *nā*      *gat*      *dāŋ*      *tʃén* ]<sub>S</sub> ·      **sak** ]<sub>NP</sub>    [ *í*      = $\bar{\lambda}$  ]<sub>VP</sub> ]  
 [ [ [ 3pers    TP      market    DAT    go ]      **PART** ]    [ COP =SG ] ]  
 ‘He already went to the market (finished, returned).’

All three of these constructions are similar to constructions found in Matwang, described by LaPolla (2006).

## 6 Conclusion

In summary, tense and aspect in Lungmi Rawang are marked primarily through temporal orientation markers, which mark past perfective events in three degrees of remoteness. In comparison with other Rawang varieties, we can see that Lungmi has had significant phonological and lexical changes to the tense and aspect marking morphemes. However, many aspects of the grammatical structure have stayed consistent.

Tense, aspect, and modality can also be marked by structures outside of the verb phrase. The use of cleft constructions with nominal clauses can have various tense, aspect, and modal meanings based on the nominalizer used. Verbs such as *dā* ‘future’ and *dāŋ* ‘finish’ can also have tense and aspect meanings. Finally, reduplication of the verb stem can be used to indicate the ongoing or repetitive nature of the event.

This paper is based on a longer thesis, Powelson (2018), which gives an overview of the entire verb phrase in Lungmi.

### Abbreviations

1/2/3pers	1st/ 2nd/ 3rd person
1PL	1st person plural
2PL	2nd person plural
AG	agent
BEN	benefactive
CLF	classifier
COP	copula
D/P	dual/plural
DAT	dative
DIR	directional
DL	dual
3:I	3rd person, intransitive
LOC	locative
OB	object
n1	non-1st person subject marking
PART	participle
PF.rec	perfective, recent
PF.mid	perfective, mid-distant
PF.rem	perfective, remote
PSB	possibility
PT	patient
PUR	purpose
R/M	reflexive/middle marker
SB	subject
SFP	sentence final particle
SG	singular marker <sup>7</sup>
SG:3PT	3rd person patient with singular subject
TP	topic
Y/N	yes/no question marker

### *Additional abbreviations from Matwang examples from LaPolla and Poa (2001)*

CAUS	Causative
DIM	post-nominal Diminutive marker
IP	Third person intransitive past marker
NP	Non-Past sentence final marker
PREF	Prefix (noun prefix, intransitivizing verb prefix)
TMhrs	Past tense markers, within last 24 hrs
TMdys	Past tense marker, 2-3 days up to a year ago
TMyrs	Past tense marker, years ago
TNP	Transitive Non-Past (3rd person object marker)

### *Additional abbreviations from Rvmøl examples from Straub (2016)*

CIS	cislocative direction marker
PFV	perfective

### *Additional abbreviations from Sinwal examples from Sarep (1995)*

t-rmt	time- remote
1s	1st person singular
imprf	imperfect

---

<sup>7</sup> In Lungmi, any 3rd person is treated as singular for the purposes of person/number marking in the verb phrase.

## References

- LaPolla, Randy J. 2006. Copula Constructions in Rawang. In *Linguistic Studies in Chinese and Neighboring Languages*, ed. by Dah-an, et. al., 1059-1076. Taipei: Academia Sinica.
- LaPolla, Randy J. 2010. Transitivity Harmony in the Rawang Language of Northern Myanmar. In *The southern hemisphere: parameters of language variation*, ed. by de Beuzeville, L. and Peters, P. Sydney, Australia: University of Sydney.
- LaPolla, Randy J. 2015. On the logical necessity of a cultural and cognitive connection for the origin of all aspects of linguistic structure. In *Language structure and environment: social, cultural, and natural factors*, ed. by Rik De Busser & Randy J. LaPolla, 31–44. Amsterdam & Philadelphia: Benjamins.  
[https://www.academia.edu/13461775/On\\_the\\_logical\\_necessity\\_of\\_a\\_cultural\\_and\\_cognitive\\_connection\\_for\\_the\\_origin\\_of\\_all\\_aspects\\_of\\_linguistic\\_structure](https://www.academia.edu/13461775/On_the_logical_necessity_of_a_cultural_and_cognitive_connection_for_the_origin_of_all_aspects_of_linguistic_structure) [accessed on July 6, 2015].
- LaPolla, Randy J. and Poa, Dory. 2001. *Rawang Texts*. München, Germany: Lincom Europa.
- Morse, Robert. 1965. Syntactic Frames for the Rvwàng (Rawang) Verb. *Lingua* 15:338-369.
- Morse, Stephen A. 1988. Five Rawang Dialects Compared Plus More. In *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*, ed. by D. Bradley, E.J.A. Henderson, & Mazaudon, M., 237-250. Canberra, Australia: Pacific Linguistics.
- Powelson, Rachel. 2018. *The Verb Phrase in Lungmi Rawang*. Manuscript in preparation.
- Straub, Nathan. 2016. *Direction and Time Reference in the Rvmøl (Dvru) Dialect of Rawang, from Northern Myanmar*. Chiang Mai, Thailand: Payap University.
- Sarep, Hpung. 1995. *A Study of the Morphology of Verbs and Nouns of the Sinwal Dialect of the Rawang Language*. Chiang Mai, Thailand: Payap University.

# VERBAL AGREEMENT IN NOCTE

Syed Iftiqar Rahman  
Payap University  
iftiqar.rahman@gmail.com

## Abstract

Nocte belongs to the Northern Naga subgroup within the Tibeto-Burman language family. Ethnologue records the population of the Nocte community to be 33,000 (2001 census) and the language status as 6a (vigorous). Burling (2003) categorizes Nocte together with Bodo-Garo, Koch, Konyak and Jingphaw languages into the Sal subgroup of Tibeto-Burman languages. The Nocte verb is always followed by an uninflected or inflected operator or an agreement word. Verbal operators give additional information to the main verb like tense, aspect, mood, negation etc. Besides these, verbal operators also include cislocative and inverse marking morphemes. An unusual aspect of Nocte is having a hierarchical agreement system. Hierarchical agreement is a kind of special agreement system, where the transitive verb agrees with a person that is higher in the hierarchy irrespective of its semantic and grammatical role. In Nocte person hierarchy, first person is higher than second person and both first person and second person are higher than third person i.e. 1>2>3. Apart from hierarchy in the agreement system Nocte also has inverse marker */-h-/* that functions similar to the person indexation system as it also concentrates on marking the hierarchy of a person in a speech event.

**Keywords:** Nocte, Tibeto-Burman, hierarchical agreement, inverse, cislocative.

**ISO 639-3 codes:** njb

## 1 Introduction

This paper is a description of the verbal agreement system of Nocte, a language variety spoken by the Noctes settled in the Tirap district in Arunachal Pradesh, India. An unusual aspect of Nocte is having a hierarchical agreement system. Hierarchical agreement is a kind of special agreement system, where the transitive verb agrees with a person that is higher in the hierarchy irrespective of its semantic and grammatical role. In Nocte person hierarchy, first person is higher than second person and both first person and second person are higher than third person i.e. 1>2>3. Similar hierarchical systems are also found in Rawang (LaPolla 2010), Jingphaw (DeLancey 2011) and some other Northern Naga varieties like Muklom Tangsa (Morey 2011).

Section 1.1 presents the Nocte literature review, and Section 2 talks about the structure of the verbal system. Section 3 explains hierarchical agreement and inverse marking whereas Section 4 discusses agreement in focus. Section 5 discusses cislocative marking and Section 6 elaborates on agreement in extended constructions like negatives and serial verb construction.

Tirap district is the home to the greater Nocte community. Tirap shares a district border with Changlang and Longding, a state border with Nagaland and Assam, and an international border with Myanmar. There is one Nocte village in the Tinsukia district of Assam called the Paltan Basti or Dihing Kinar Nocte Gaon. Recently, I learned of a few more Nocte villages in the Changlang district as well.

Nocte belongs to the Northern Naga subgroup within the Tibeto-Burman family with the code ISO 639:3 njb Naga Nocte. Ethnologue records the population of the Nocte community to be 33,000 (2001 census) and the language status as 6a (vigorous). Burling (2003) categorizes Nocte together with Bodo-Garo, Koch, Konyak and Jingphaw languages into the Sal subgroup of Tibeto-Burman languages. Similar subgrouping of Tibeto-Burman languages is also proposed by linguists such as Bradely (1997) and Thurgood (2003).

My paper is mainly based on the data collected from Borduria village. The neighbouring Nocte villages are Kaimai, Paniduria, Khonsa, Polung, Laptang, and Pansumthong. There are also two Wancho Naga villages in the area, namely Lapnan and Lokthong. Wancho is a closely related language

variety of the Nocte and this community resides in the Longding district. In Nocte, *nok* means ‘village’ and *te* stands for the people, so Nocte refers to the people living in the village. According to a 1971 census, there were 58 Nocte villages with a population of 21,853. And according to a 2001 census the population is about 33,000. There are approximately 29 Nocte villages under the Khonsa division, the district headquarters of Tirap district. The Nocte villages I have visited so far are, Borduria, Kheti, Paltan, Dadam, Thinsa, Polung, New Tupi, Deomali, Hoka and Khonsa.

The Borduria people speak the Ha?wa variety of Nocte. Hindi is their second language. However, the old generation can speak Assamese very fluently as Assamese was once the medium of instruction in schools. Now, Hindi and English are widely used in schools.

### ***1.1 Nocte Literature review***

Very little linguistic description is available on Nocte. However, a wider research allowed me to explore a few linguistic descriptions available on Nocte. One such description is the ‘Tibeto-Burman family’ by Grierson (1903). This is one of the earliest documentation of the language and one of the first writings on Nocte available to me. It can be safely argued that Grierson was inspired by Robinson (1849), a predecessor who wrote a grammar sketch of the Namsangia variety of Nocte. These works briefly describe the lexical categories like nouns, adjectives, pronouns, verbs, adverbs and conjunctions. There are no descriptions about the verbal agreement system which is the main focus of my paper in these writings. ‘An Introduction to the Nocte language’ by K. Das Gupta (1971) came after a long gap. Although this volume introduces the verbal agreement marking in Nocte, it is still far from describing any of the post-verbal complexities like the hierarchy in the agreement marking or the inverse marking present in the language. Linguistically, a lesser relevant work ‘The Noctes’ by Parul Dutta (1978) enabled me to have a glance at the socio-political and religious life of the Noctes. Although these books do not meet modern linguistic standards, they are nonetheless a reliable source of the socio-economical background of the Noctes and also provide geographical information about the Nocte inhabited areas.

The second phase of linguistic research on Nocte began with the works of linguists like Alfons Weidert and Scott DeLancey. Alfons Weidert gave his unpublished notes on Nocte to Scott DeLancey and through him I had the opportunity to study them. Weidert can be said to be the first linguist who studied Nocte in more depth in comparison to his predecessors. Weidert’s notes gave an outline of the agreement paradigm present in Nocte but they seemed to be based on limited data as he mostly collected sentences presented as examples in K. Das Gupta’s book ‘An Introduction to the Nocte language’. Moreover, the source of his data is not known. Then came a more concrete work on Nocte that is ‘Nocte and Jingphaw: Morphological correspondences’ by Scott DeLancey (2011). This paper is an attempt by DeLancey to do a comparative study of Jingphaw and Nocte morphology to show the close relationship between the two languages and thus to confirm Burling’s (2003) hypothesis which suggests a special relationship between Bodo-Garo, the Konyak Naga languages and Jingphaw.

The approach of linguists like Alfons Weidert and Scott DeLancey is more relatable to the topic of this paper than any other works available. They set the outline for my research. The understanding of the complex concepts like the hierarchical agreement and inverse marking system opened new horizons. DeLancey’s paper was immensely helpful in understanding the complex post-verbal auxiliaries present in Nocte. I started my research with a basic understanding of the agreement system in Nocte. It cannot be denied that DeLancey’s paper threw light on the agreement particles and complexities of the hierarchical agreement system present in Nocte. In addition to this, he also discussed the grammatical inverse system present in the language.

Before moving on to describing the complex post verbal morphology of Nocte, it is worth describing the grammatical aspects of the language such as the hierarchical agreement system and inverse marking.

Unlike K. Das Gupta’s (1971) analysis of Nocte tense and agreement markers as verbal suffixes, this paper works in alignment with DeLancey (2011, 2014) which considers auxiliaries carrying information about TAM, negation and agreement particles as separate words from the main verb, more specifically as sentence final words (SFWs).

The Nocte agreement system is based firstly on person hierarchy and secondly on the grammatical roles of the agent and patient. In other words, in a transitive verb construction in Nocte, the verb agrees



with a person that is higher in the person hierarchy. In Nocte person hierarchy, first person is higher than second person and both first person and second person are higher than third person i.e. 1>2>3. On the other hand, inverse marking shows whether an object (O) argument outranks the agent (A) argument or not. According to DeLancey (2011, 2014), with hierarchical agreement marking, the presence of the inverse marker confirms that the agreement is with the object and not with the agent.

At this stage there is a lot more work to do on collecting conversational data and more narratives or natural texts which might produce examples contrary to my statement. In other words, after detailed study there may arise some examples defying my analysis of the hierarchical agreement. For now, I can say that Nocte does have some kind of person hierarchy in the agreement marking and potentially it also has deviations in the agreement hierarchy.

## 2 Structure of the verbal system

According to DeLancey (2014), in languages in northern Myanmar and Northeast India, agreement markers do not attach to the verb stem but occur as phonologically independent words which he refers to as the agreement words. These agreement markers can also get attached to certain morphemes carrying information about tense, aspect and other verbal categories. It is in this sense as suggested by DeLancey that these agreement words can be referred to as verbal operators. This section of the paper gives an overview of how the verbal operators work in Nocte.

Morey (2011) recorded some of the Tangsa varieties having agreement as verbal suffixes. Das Gupta (1971) records Nocte as having verbal suffixes to mark tense and aspect. However in Nocte I have recorded agreement words together with verbal operators as separate phonological words. A phonological word is defined in terms of a variety of phenomena, namely, prosodic specifications like stress, segmental specifications like assimilation sandhi, phonotactics etc. and one grammatical word may consist of more than one phonological word (Post 2007). It is also evident that consonants and vowel harmony occurs within a word and not across word boundaries whereas in Nocte, there seems to be no assimilation between the verb root and the agreement word or particle that follows. For example, *say-tak* ‘sell-PAST-1SG’ ‘sold’ is realized as *say tak* as two separate words. Also Alfons Weidert in his unpublished notes on Nocte records agreement, TAM, and negation as separate words. However, deeper studies need to be done to understand the agreement words in Nocte.

Before discussing the verbal agreement in Nocte, a brief introduction to the Nocte verb morphology is provided in this section.

Nocte has a very complicated post-verbal morphology. DeLancey (2011) discusses the complex SFWs (Sentence Final Word, coined by Dai and Diehl, 2003) in Jingphaw and Nocte. Another important paper by DeLancey (2014) talks about the presence of similar SFW in archaic Kuki-Chin language varieties which he quotes as ‘transparently grammaticalized auxiliary verbs’.

The agreement markers in Nocte can also get attached to certain morphemes carrying information about tense, aspect and other verbal categories. It is in this sense as suggested by DeLancey that these agreement words can be referred to as verbal operators. In Nocte verbal operators give additional information to the main verb like tense, aspect, mood, negation etc..

This section of the paper gives an overview of how the verbal operators work in Nocte. Verbal operators can be of two types: inflected and uninflected. For example, Nocte tense-aspect markers can be both inflected and uninflected. Nocte future tense can be expressed in two different ways: one with the agreement words that inflect for person and the other with an uninflected verbal operator that does not change according to person.

The following examples show verbal operators and agreement word with no operator. In (1), the verb is followed by the operator *t<sup>hu</sup>* and it can be considered as a verbal operator as it gives information about the tense and aspect. Since, *t<sup>hu</sup>* is invariant for person i.e. there is no agreement particle in it, it is just an uninflected operator.

- a. V+ uninflected operator  
 (1) *John hu-kɔ vɛt t<sup>hu</sup>*  
 John dog-ERGhit PRS-PROG  
 ‘John is hitting the dog.’

In (2), *k-a* is a verbal operator that follows the verb stem *vət* ‘hit’ where *-a* the third person agreement marker gets attached with the present progressive aspect marker *-k-*. Here, in *k-a*, *k-* is the operator and *-a* is the agreement word.

b. V+ inflected operator

- (2) *John hu-kə vət k-a*  
 John dog-ERGhit PRS-PROG-3  
 ‘John is hitting the dog.’

In (3), the third person agreement marker *a* follows the verb stem *ka* ‘go’. Although we cannot see any overt verbal operator attached with the agreement particle, the future tense information is already incorporated into the agreement word.

c. V+ agreement word with no operator

- (3) *ate pit-naŋ ka a*  
 3SG field-LOC go FUT-3  
 ‘he will go to the field.’

Besides tense, aspect and negation, verbal operators also include inverse marking and cislocative morphemes which are discussed in Section 3 and Section 5 respectively.

### 3 Hierarchical agreement and inverse marking

Hierarchical agreement is a kind of special agreement system, where the transitive verb agrees with a person that is higher in the hierarchy irrespective of its semantic and grammatical role. In Nocte person hierarchy, first person is higher than second person and both first person and second person are higher than third person i.e. 1>2>3.

Apart from hierarchy in the agreement system, Nocte also has inverse marking that functions similar to the person indexation system as it also concentrates on marking the hierarchy of a person in a speech event. However, its main function is to show whether the object argument outranks the agent or not. Nocte has an inverse marker */-h-/* that occurs after the main verb. But since Nocte already has a hierarchical agreement system to depict the person hierarchy, the overt inverse marking merely serves to make the understanding of the agreement unambiguous.

**Table 1:** Hierarchical agreement and inverse marking

A>P	Agreement			A>P	Agreement
1SG>2SG	aŋ	<==		2PL>1SG	h-aŋ
1SG>2PL	aŋ			2PL>1PL	h-i
1SG>3	aŋ	=		2PL>3SG	ɛn
1PL>2	ɛ	<==		2PL>3PL	ɛn
1PL>3	ɛ	<==		3>1SG	h-aŋ
2SG>1SG	h-aŋ	<=====		3>1PL	h-i
2SG>1PL	h-i	<=====		3>2SG	h-ə
2SG>3SG	ə	<==		3>2PL	h-ɛn
2SG>3PL	ə	<==		3>3	a

Table 1 shows the person hierarchy and inverse marking in Nocte verbal agreement. The first and the third columns in the table with the heading A>P read as the agent (A) working (>) on the patient (P) whereas, the second and the fourth columns present the agreement markers. Similar agreement markers that reappear throughout the table are linked with equal signs and vertical lines to show the evidence of the hierarchy in the Nocte agreement system. Some-examples showing hierarchical agreement are as in (4).

- (4) *e-ma*                      *dihjaʔnja-naŋ*                      *ko*                      *a*  
 3SG-ERG                      girl-ABS                      give                      FUT-3  
 ‘He will give to the girl.’

Example (4) does not show any hierarchy in the agreement marking as both the agent and the recipient are 3SG, whereas in example (5) below the recipient 1SG is higher in the hierarchy than 3SG, therefore the verb agrees with the 1SG.

- (5) *ate-ma*                      *ŋa-naŋ*                      *ko*                      *h-aŋ*  
 3SG-ERG                      1SG-ABS                      give                      INV-FUT-1SG  
 ‘He will give to me.’

- (6) *ŋa-mε*                      *ate-naŋ*                      *ko*                      *aŋ*  
 1SG-ERG                      3SG-ABS                      give                      FUT-1SG  
 ‘I will give to him.’

Comparing examples (5) and (6), it is clear that the inverse marker appears when the hierarchy forces the agreement with the P argument instead of the A argument. In (5), the agreement is with the P argument as 1P>3A, therefore we also see the inverse marker /h-/ attached to the agreement word. However, example (6) does not show the inverse marking as the agent itself (1A) is higher in the hierarchy and therefore the verb also agrees with it and there is no need to have the inverse marker.

#### 4 Agreement in focus

In elicitation, we always get the hierarchy in the agreement system where the verb agrees with the argument that is higher in the hierarchy. However, a transitive sentence can be marked for either of the arguments to mark focus. In other words, it is also possible that apart from tense, person and aspectual information, the agreement markings can also carry some pragmatic and contextual information. This inference is required for the understanding of the apparently irregular agreement marking of the subject and objects in Nocte. Two such situations are described below.

Firstly, in Nocte there is a special agreement marking between 1SG subject and 2SG object, where the verb instead of agreeing with the 1SG subject and 2SG object rather is marked with 1PL agreement. One example is in (7).

- (7) *ŋa-ma*                      *naŋ-naŋ*                      *ko*                      *εʔ*  
 1SG-ERG                      2SG-ABS                      give                      FUT-1PL  
 ‘I will give to you.’

Here in (7), the verb instead of being marked for 1SG or 2SG is marked for 1PL which suggests that the action of giving is seen as a process that involves both the giver *ŋa* (1SG) and the receiver *naŋ* (2SG) achieving the meaning of we *ni* (1PL).

Secondly, the agreement marking can also suggest whether the subject or the object is marked. For example, in the following two sentences the pragmatics and context will help the listener to understand the speaker’s selective use of 3<sup>rd</sup> person agreement marking over the 1<sup>st</sup> person agreement or vice versa.

- (8) *ate-ma*            *ŋa-naŋ*            *ko*            *t-a*  
 3SG-ERG            1SG-ABS            give            PAST-3SG  
 ‘He gave it to me.’
- (9) *ate-ma*            *ŋa-naŋ*            *ko*            *t-h-aŋ*  
 3SG-ERG            1SG-ABS            give            PAST-INV-1SG  
 ‘He gave it to me.’

In example (8), the verb agrees with the 3SG and thus emphasizes the 3<sup>rd</sup> person. Pragmatically this sentence answers the question ‘Who gave it to me?’, whereas in sentence (9) the verb agrees with the 1SG object and thus emphasizes 1SG, answering the question, ‘Whom did he give it to?’. Presumably, the presence of the inverse marker *-h-* helps the listener or the addressee to confirm that the emphasis is on the 1SG recipient. Apparently, the presence of the inverse marker looks redundant as the verb already agrees to the 1SG which is higher in the person hierarchy and thus is already emphasizing the 1SG recipient, but the inverse marker still co-occurs with the agreement marker in Nocte. One more example showing the presence of both the inverse marker and the hierarchical agreement marking is given in (10).

- (10) *Peter-ma*            *ŋa-naŋ*            *vet*            *t-h-aŋ*  
 3SG-ERG            1SG-ABS            hit            PAST-INV-1SG  
 ‘Peter hit me.’

## 5 Cislocative marking

Nocte has a cislocative marker /-i-/ that gets prefixed to the agreement morpheme. The cislocative marker indicates ‘going’ or ‘coming’ motion in reference to a particular location or person. DeLancey (2011) describes the cislocative marker and gives a comparison between Nocte and Jingphaw. DeLancey also refers to Weidert’s claim that the cislocative marker is possible only in the present tense. Table 2 presents the Nocte cislocative paradigm from DeLancey (2011).

**Table 2:** Cislocative paradigm (DeLancey, 2011)

<i>Person</i>	<i>Singular</i>	<i>Plural</i>
<i>1</i>	<i>-i-ŋ</i>	<i>-i-ŋ?</i>
<i>2</i>	<i>-i-ɔ?</i>	<i>-i-ŋŋ</i>
<i>3</i>	<i>-i-a</i> (non-past) <i>-i-a?</i> (past)	

Some examples are provided in (11) and (12).

- (11) *naŋ*            *ka*            *ɔ*  
 2SG            go            IMP-2SG  
 ‘You go.’
- (12) *naŋ*            *ka*            *-i-ɔ*  
 2SG            go            CIS-go  
 ‘You come.’

The only difference between examples (11) and (12) is the cislocative marker /-i-/ which occurs with motion verbs in Nocte to show that the movement is toward the deictic centre. In (12), the cislocative marker /-i-/ occurs with the verb *ka* ‘go’ to mean come instead. Another example is (13).

- (13) *naŋ-me                    ŋa-laŋ                    kɔləm    wən            ɿ-ɔ                    a*  
 2SG-ERG                    1SG-for   pen            take            CIS-FUT-2SG                    QP  
 ‘Will you bring a pen for me?’

In (13), the usual meaning for the verb *wən* is ‘take’ but when cislocative /ɿ-/ occurs, the meaning changes to ‘bring’, that is, the motion of the action is towards the deictic centre or the speaker.

## 6 Agreement in extended constructions

This section describes verbal agreement in extended constructions like negatives and serial verb constructions. Section 6.1 reviews the different ways negatives can be constructed in Haɔwa Nocte. Section 6.2 focuses on prohibitive constructions and 6.3 presents a comparative view of the negative constructions found in Nocte and other Tangsa/Tangshang varieties. Section 6.4 discusses agreement in Nocte serial verb construction.

### 6.1 Negative constructions

The negative morphemes are the uninflected operator *ho*, the verb prefix *la-*, and the negative particle *ma-*. Negative *ma-* is the basic negator. It can be prefixed either to the verb, as in example (14), or to the agreement word, as in (15) and (16). In the first type of negative construction, negator *ma-* occurs as a prefix to the main verb.

- (14) *nime                    tʃam matʃaʔ                    ki*  
 ni    -mɛ    tʃam    ma-    tʃaʔ    k                    -i  
 1PL    ERG    rice    NEG    eat    PROG    1PL  
 ‘We are not eating rice.’

In (14), *ma-* occurs as a prefix to the verb *tʃaʔ* ‘eat’.

- (15) *ətʃinme                    ninanŋ                    k<sup>h</sup>e mat<sup>h</sup>i*  
 ətʃin    -mɛ    ni-    nanŋ    k<sup>h</sup>e    ma-    t                    -<sup>h</sup>                    -i  
 they    ERG    we    ABS    see    NEG    PAST    INV    1PL  
 ‘They did not see us.’

In (15), the negator *ma-* occurs as a prefix to the agreement word. Similarly, in (16), the negator prefix *m-* is prefixed to the agreement particle.

- (16) *naŋ tʃam tʃaʔ mɔʔ*  
 naŋ    tʃam    tʃaʔ    m-                    ɔʔ  
 you    rice    eat    NEG    2SG  
 ‘You do not eat rice’.

The following examples illustrate the use of uninflected negative operator *ho* and uninflected verb prefix *la*. Here in (17), uninflected negative operator *ho* ‘have not’ occurs sentence-finally after the verb *ket* ‘go’.

- (17) *naŋ ket ho*  
 naŋ    ket    ho  
 you    go    have not  
 ‘You do not have to go.’

Similarly, in (18), uninflected verb prefix *la-* occurs as a prefix to the verb *tʃam* ‘eat’ and the future tense is marked by the invariant future tense marker *min*.

- (18) *ate tfam latfa? min*  
 ate tfam la- tfa? min  
 3SG rice NEG eat FUT  
 ‘He will not eat rice.’

Negation with *ma-* may be combined with other operators, as in examples (19) and (20). In example (19), the verb *luam* ‘play’ is nominalized by *-te* and the negator *-m* occurs as an infix between the auxiliary verb *daŋ* ‘do’ and the agreement particle.

- (19) *ŋa football luamte daŋmak*  
 ŋa football luam -te daŋ -m -ak  
 I football play NOM do NEG 1SG  
 ‘I don’t do football playing’

In the ‘not-yet’ constructions, negator *ma-* occurs between *t<sup>h</sup>uhe* ‘yet’ and the main verb. (20) is an example.

- (20) *atea t<sup>h</sup>uhe ma hu k<sup>h</sup>uəm ke*  
 ate -a t<sup>h</sup>uhe ma hu k<sup>h</sup>uəm ke  
 3SG ABS yet NEG reach walk PERF  
 ‘She has not reached it yet.’

Table 3 summarizes the possible negative constructions in Nocte.

**Table 3: Schemas of negatives in Nocte**

<b>Schema</b>	<b>comments</b>
<i>1a. ma-‘NEG’ + V + agreement</i>	<i>Shows speaker’s intention.</i>
<i>1b. ma- ‘NEG’ + V+daŋ _’do’+ -wa ‘invariant past tense marker’</i>	<i>This construction is used to negate actions in simple past tense.</i>
<i>1c. t<sup>h</sup>uhe ‘yet’+ ma- (NEG) + V + (-k- (PROG)) + agreement</i>	<i>Used to show that the action is not completed yet.</i>
<i>2a. V + ma-‘NEG’ + agreement</i>	<i>It shows a kind of certainty in speaker’s speech.</i>
<i>2b. V+m-‘NEG’+agreement</i>	<i>It reflects speakers comment on some habitual present events.</i>
<i>2c. V+NZ -te+ daŋ _’do’+ -m- ‘NEG’+agreement</i>	<i>This type of construction reflects speaker’s dislike or disability or attitude toward certain kind of activity.</i>
<i>3. V (stem 2)+ho ‘invariant negative copula’</i>	<i>This type of construction portrays speaker’s suggestion or conception about some event.</i>
<i>4. la-‘NEG’+V+min ‘invariant future marker’</i>	<i>It used to negate some action in the future.</i>

## 6.2 Prohibitive *nak*

*nak* ‘prohibitive’+V+agreement

In Nocte, there is a prohibitive particle *nak* that occurs before the verb. The addressee is mostly covered. The following examples illustrate this. In example (21), *nak* occurs before the verb *ka* ‘go’ and in (22) occurs before the verb *t<sup>h</sup>o* ‘tell’.

- (21) *ɿenkɔ nak ka ɔʔ*  
 ɿen -kɔ nak ka ɔʔ  
 there LOC PRH go 2SG  
 ‘Don’t go there!’

- (22) *mot<sup>h</sup>o*                      *nak*      *t<sup>h</sup>o*      *ɛn*  
 mo-      t<sup>h</sup>o      nak      t<sup>h</sup>o      ɛn  
 lie      tell      PRH      tell      FUT-2PL  
 ‘Don’t tell a lie.’

### 6.3 Serial verb construction

In Haʔwa Nocte serial verb constructions, a non-finite linker *le* occurs between the verb phrases and the final verb is marked for agreement. One good example is (23). In (23), *le* connects the verb phrases *tʃaʔ* ‘eat’ and *tʃuen* ‘run’ and the final verb *tʃuen* is followed by the agreement particle.

- (23) *ate*      *tʃam*      *tʃaʔ*      *le*      *mɛ*      *tʃuen*      *taʔ*  
 3SG      rice      eat      NF      ERG      run      PAST.3  
 ‘He ate rice and ran.’

Causative constructions in Nocte are also a kind of serial verb construction where the same non-finite linker *le* connects the cause and result noun phrases. In addition, an optional, possible verbalizer *daŋ* occurs after the NPCAUSE. The schema for causative NPs is given below.

Reason/Cause: [[NPi.....*le*...V(no agreement)..]SCause -mɛ ɔi V (agreement)]

Some good examples are (24 and (25). In examples (24) and (25), non-finite particles *le* connects the NP cause and NP experiencer.

- (21) *ate* *mele.ia* *daŋ* *le* *mɛ* *ɪ* *taʔ*  
 ate *mele.ia* *daŋ* *le* -mɛ *ɪ* *taʔ*  
 3SG malaria AUX NF ERG die PAST.3  
 ‘Having malaria, he died.’

- (22) *ate* *pu* *k<sup>h</sup>e* *le* *mɛ* *tʃuen* *taʔ*  
 ate *pu* *k<sup>h</sup>e* *le* -mɛ *tʃuen* *taʔ*  
 3SG snake see NF ERG run PAST.3  
 ‘On seeing a snake, she ran away.’

## 7 Conclusion

Nocte post-verbal morphology is very complex with the presence of phonologically independent agreement words that can also get attached with other verbal categories like tense, aspect, mood and negation. In addition, Nocte has other fascinating features like the hierarchy in the agreement marking system or the presence of the inverse marker etc.. Further studies will allow the complexity of the Nocte verbal morphology to be explained with greater details.

### Abbreviations

SG	Singular
PL	Plural
ERG	Ergative
ABS	Absolutive
CIS	Cislocative
INV	Inverse
QP	Question particle
LOC	Locative
PROG	Progressive
FUT	Future

HAB	Habitual
1P	1 <sup>st</sup> person
2P	2 <sup>nd</sup> person
3P	3 <sup>rd</sup> person
PRS	Present

## References

- Burling, R. 2003. The Sal Languages. *Linguistics of the Tibeto-Burman Area* 7(2).
- Das Gupta, K. 1971. *An Introduction to the Nocte Language*. Shillong: North-East Frontier Agency.
- DeLancey, Scott. 2004. The blue bird of ergativity. In *Ergativity in Amazonia III*, ed. by Francesc Queixalós, 1–37. Paris: CELIA-CNRS.
- DeLancey, Scott. 2011. Nocte and Jingphaw: Morphological correspondences. In *Papers from the Northeast Indian Linguistic Society conference III*, ed. by G. Hyslop, S. Morey and M. Post, 61-75. New Delhi, Cambridge University Press India.
- Dutta, P. 1978. *The Noctes*. Shillong: Directorate of Research, Government of Arunachal Pradesh.
- Endangered Languages* (n.d.). <http://www.unesco.org>
- Grierson, G. A. 1903. *Tibeto-Burman Family Vol. 3 part ii*. Delhi: India
- LaPolla, R. 2010. *Hierarchical Person Marking in the Rawang Language*.  
<http://victoria.linguistlist.org>
- Morey, Stephen. 2011. Tangsa agreement markers. In *Papers from the Northeast Indian Linguistic Society conference III*, ed. by G. Hyslop, S. Morey and M. Post, 76-104. New Delhi: Cambridge University Press India.
- Morey, Stephen. 2013. *Tone in Tangsa Languages- a comparative study*. Paper presented at the Himalayan Languages symposium. Canberra, Australia.
- Post, M. 2007. *A Grammar of Galo*. PhD Dissertation. Research Center for Linguistic Typology. Australia: La Trobe University.
- Robinson, G. 1894. Namsangia. *Asiatic Society of Bengal*, XVHI.
- Weidert, Alfons. n.d. Unpublished notes on Nocte.



# THE ACQUISITION OF THAI CLASSIFIERS: ORDER AND STRATEGIES<sup>1</sup>

Phetprapa Sangkharam  
*Kasetsart University*  
*phetprapa34033@gmail.com*

Kitima Indrambarya  
*Kasetsart University*  
*kitima.i@ku.th*

## Abstract

The present study aimed to analyze the acquisition of classifiers in Thai children from 3-6 years of age, focusing on strategies and their order. The cross-sectional data was collected by one-on-one interview with 37 Thai children in Bangkok who speak only standard Thai. The selected entities in 5 categories were given to them. The findings revealed that the children's developmental stages can be roughly divided into two stages, each of which exhibits rather different strategies in using classifiers: 1) non-classifier stage, using noun reduplicating, number counting, coding and message strategies and 2) classifier stage, using noun repetition, general classifier, general noun, shape, specific part, specific classifier and specific reason strategies.

**Keywords:** classifiers, language acquisition, Thai  
**ISO 639-3 codes:** tha

## 1 Introduction

Tuaycharoen (1984) investigated the acquisition of classifiers among Thai children and found that they used six developmental strategies at the age of 2 to 5 years: the early attempt (2;0-2;6), noun repetition (2;0-2;6), identical noun deletion (2;6-3;0), over-extension (3;0-4;6), trial and error (4;6-5;0) and dodging (4;6-5;0) (Rungrojsuwan 2003:26) shown in Table 1.

The researchers did a cross-sectional study with 9 children between 2 and 5 years of age. These children use standard Thai and live in Bangkok, Thailand. Differences in terms of strategies in using classifiers are found among children in the study. The findings revealed that the children between 3 - 3.6 years of age use number counting more often than using classifiers. Instead of using the number of objects followed by classifiers, the children further counted the number. For example, when the researchers pointed at the picture of two dogs, the children replied by counting the number "one two..." without any classifier. The number counting of the children may be "one two three...", "three two one..." or "one three..." etc.. Moreover, noun repetition strategy was not found at all at this age. Tuaycharoen (1984) proposed that the children at the age 3-3.6 years old use the noun repetition strategy most and they will go on to the next steps, namely the identical noun deletion and the over-extension strategies. Children at the age of 5 use fewer varieties of classifiers than those at the age of 4.

The cross-sectional data of this study are different from Tuaycharoen's (1984). The data of this study were then analyzed differently. It is found that strategies used by children were interesting because their concepts and reasoning behind their use of classifiers differed, to a certain extent, from the adults' and those strategies were in a certain order. The researchers then recollected the data from 37 children and analyzed the order and strategies again.

---

<sup>1</sup> This article is part of the thesis entitled "The Acquisition of Classifiers in Thai Children". Special thanks to Kachen Tansiri, Nattawut Chaicharoen and Satit Leelathawornchai for invaluable comments.

**Table 1:** Developmental strategies in the acquisition of classifiers in Thai (Tuaychareon, 1984)

Strategies	Age	Descriptions	Examples
The Early Attempt	2;0-2;6	Realization of classifiers in the language	nók sǎam nók bird three bird 'three birds'
Noun Repetition	2;0-2;6	Noun repetition after number	mǎa sǒɯŋ mǎa → mǎa sǒɯŋ dog two dog dog two 'two dogs'
Identical Noun Deletion	2;6-3;0	Noun deletion in classifier position	kradaat nùŋ lēm (nùŋ pʰèn) paper one CL (one CL) 'one paper'  mǒɔtʰɯsai nùŋ kʰrúaŋ (nùŋ kʰan) motorcycle one CL (1 motorcycle) 'one motorcycle'
Over-Extension	3;0-4;6	- from generic to specific - from major to component	nùŋ sôm → nùŋ plùək one orange one CL (shell) 'an orange(CL)' → 'a shell(CL)'
Trial and Error	4;6-5;0	A trial	Show a picture of one shoe kʰrúŋnùŋ 'half'
Dodging	4;6-5;0	avoidance of unfamiliar classifiers	nók sǎam nók bird three bird 'three birds'

The objective of the present study is to analyze the acquisition of classifiers in Thai children from 3-6 years of age, focusing on strategies and their order. This study ignored the standard criteria of Thai classifiers for adults.

## 2 Methodology

For this study, 43 nouns from 5 categories were selected as experimental stimuli: human/human-like, animals, vegetables/fruits, vehicles, and others. The nouns were selected from a first-word book for children aged 2 -5 years old, *Language in the two-year-old* (Susan Goldin-Meadow, Martin E.P. Seligman Rochel Gelman, 1976) and *First Words: Communicative Development of 9- to 24-month-old Thai Children* (Rungrajsuan 2003). All of the selected nouns were represented by pictures to be shown to the children. Each noun was identified with semantic features. Only nouns with different semantic features were selected (e.g. dog vs. fish but not dog vs. cat). The researchers expected that the children acquire the selected nouns already. There were no more than 5 items in the picture to be shown. A one-on-one interview was conducted to ask the children to identify the classifiers of the selected entities. The children tried to create some stories from the picture and the researchers tried to ask questions or play games about classifiers. The children produced 44 classifiers, all of which were subjected to a closer analysis. All selected nouns are given in table 2.

Table 2: Selected nouns

Human/ human-like	Animals	Vegetables/Fruits	Vehicles	Others
man	chicken	durian	car	spoon
woman	fish	tomato	ship	fork
doll/robot	bird	carrot	train	book/notebook
angel	dog	mushroom	plane	chair
	crab	corn	bicycle	table
	elephant	banana	motorcycle	sun
	snake	watermelon		paper
	giraffe	orange		cloud
	octopus	rose apple		pencil
	turtle	mango		ball
		apple		clock
				house

Figure 1: Examples of pictures shown to the children



### 3 Participants

37 preschool children participated in the study:30 from the same school and seven from another school. The children were 3.0 to 6.0 years old. They speak Thai as their first language.

### 4 Findings and discussions

The findings revealed that the children’s developmental stages can be roughly divided into two sub-stages: 1) the non-classifier stage, using noun reduplicating, number counting, coding and message strategies and 2) the classifier stage, using noun repetition, general classifier, general noun, shape, specific part, specific classifier and specific reason strategies.

#### 4.1 Strategies

##### 4.1.1 Non-classifier stage

##### A. Noun reduplicating strategy

After counting the entities in the picture, the children answered by reduplicating the name of the entity, as in the following example.

- (1) *ʔéppân ʔéppân ʔéppân*  
 apple apple apple  
 ‘3 apples’ (3.0 yrs.)

According to the above example, instead of using *ʔéppân sāam lûuk* ‘apple three CL’, the child reduplicated the noun, *ʔéppân* ‘apple’ three times to mean “three apples”. Therefore, it can be said that

classifiers do not appear yet at this stage. Since nouns and verbs are acquired earlier than numbers and classifiers (Clark 2009), the children at this age normally reduplicate nouns to represent the number of entities.

#### B. Counting number strategy

The children acquire numbers at the age of 3. They learn to count the number in order first. Basically, the series of one-two-three is more frequently found. This strategy refers to the use of another series of numbers in the position of classifier which follows the number of the entities.

(2)    *nùŋ*     *sǎŋ*  
          one    two  
          ‘1 snake’ (3.3 yrs.)

(3)    *sǎŋ*     *lé’*     *sǎam*  
          two    and    three  
          ‘2 pencils’ (3.3 yrs.)

In example (2), after being shown the pictures of one snake, the child replied *nùŋ sǎŋ* ‘one two’ instead of using *nùŋ tua* ‘one CL’. The child used the consecutive number *sǎŋ* ‘two’ in the position of the classifier. When the researchers asked them “how many snakes in this picture?” they replied *nùŋ* ‘one’. After the researchers asked “one what?”, they replied again with *nùŋ sǎŋ* ‘one two’.

(4)    *sǎŋ*     *nùŋ*     *sǎŋ*     *sǎŋ*  
          two    one    two    two  
          ‘2 balls’ (3.6 yrs.)

This strategy can be seen as the early stage of Tuaychareon’s (1984) trial and error strategy at 4.6 - 5.0 years. At this stage, children played with numbers. They tended to pair up the numbers in each set and start counting them. For example, in (4) when two balls were shown, they count *sǎŋ nùŋ* ‘two one’ and *sǎŋ sǎŋ* ‘two two’. The first number in each set designates the number while the last number refers to a classifier.

#### C. Coding strategy

For this strategy, the children use their own specific code in the position of a classifier.

(5)    *sìi*     *te’et*     *kâo*     *sǎŋ*  
          four    seven    nine    two  
          ‘4 bananas’ (3.4 yrs.)

For this step, the children learn more numbers and are able to create the group of number by themselves. When they were asked about number and classifier, they knew that the number must be followed by something. In example (5), a child counts *sìi* ‘four’ for four bananas. When asked “how many bananas?” he replied *sìi* ‘four’. Then, the researchers asked *sìi ?àraj* ‘four what’ to which he replied *sìi te’et kâo sǎŋ* ‘four seven nine two’. The series of number following the first number may reveal the realization of classifier position in children.

#### D. Message strategy

This strategy refers to the use of a short message.

(6)    *hùnjon*   *nùŋ*     *te’hǎp*   *lên*     *dâaj*  
          Robot   one    like    play    can  
          ‘One robot that I can play with’ (3.11 yrs.)

It is the step after the counting number strategy. For this step, they counted the number followed by a short message. The children already started acquiring classifiers but they might not know how to use them. They knew that the number had to be followed by something. It can be seen that children realize the classifier position after the number.

In example (6), after seeing a picture of one robot, the child replied *nùŋ* ‘one’. Then he was asked *nùŋ ʔàraj* ‘one what’, he replied *hùnjon nùŋ tɛʰɔ̌p lɛ̀n dâaj* ‘robot one can play’ which means ‘a robot that I can play with’. In this case, children chose to add a post-nominal modifier in the position of classifiers.

#### 4.1.2 Classifier stage

At this stage, the children use classifiers by means of various strategies.

##### A. Noun repetition strategy

The noun repetition strategy is where the head noun is copied to be the classifier.

(7) *rua      nùŋ      rua*  
 ship      one      ship  
 ‘one ship’ (3.7 yrs.)

(8) *pʰûən    sɿi      pʰûən    pʰûən*  
 friend    four      friend    friend  
 ‘4 friends’ (4.10 yrs.)

This strategy is similar to Tuaycharoen’s (1984) where the noun repetition strategy was found to be at the age of 2.0 -2.6 years. However, the present research found the noun repetition stage to be at the age of 3. The children used noun repetition when they did not know the specific classifiers of the nouns. This is the simplest strategy. Example (7) shows the noun repetition strategy of *rua* ‘ship’. In Example (8), a child used noun repetition with the human noun *pʰûən* ‘friend’. The child reduplicated the noun *pʰûən* ‘friend’ to identify plurality of the head noun as *pʰûən sɿi pʰûən pʰûən* ‘friends’. The child didn’t use the general classifier *kʰon* ‘person’.

##### B. General classifier strategy

The use of general classifiers is a major strategy of the children (Carpenter 1987) because it might be the most basic pattern and the children can acquire it quite early. General classifiers in Thai are *ʔan* for things, *tua* for animals and *kʰon* for humans and human-like figures. However, the children may use these 3 general classifiers in their own way.

(9) *puu      sɔ̌ŋ      ʔan*  
 crab      two      CL  
 ‘two crabs’ (3.3 yrs.)

(10) *mǎa      sɔ̌ŋ      tua*  
 dog      two      CL  
 ‘two dogs’ (3.0 yrs.)

(11) *naaŋfáa sɔ̌ŋ      kʰon*  
 angel      two      CL  
 ‘2 angels’ (4.6 yrs.)

These general classifiers were used by the children 3-6 years of age. Particularly, the classifier *ʔan* was most frequently found. The children acquired these general classifiers (*ʔan/tua/kʰon*) before employing other strategies.

In example (9), the classifier *ʔan* was used with animals for a child at the age of 3.3 years. This child used *ʔan* for almost all categories (things, animals, vehicles, etc.) In example (10), the classifier *tua* was used with two dogs and in example (11), the classifier *kʰon* was used with angels because they look like humans.

### C. General noun strategy

In this strategy, a noun referring to the basic category of the given noun is used as a classifier, for example the basic noun *pʰàk* ‘vegetable’ is used as a classifier for the nouns *kʰɛròt* ‘carrot’ and *kʰâopʰôt* ‘corn’.

(12) *kʰɛròt* *nùŋ* *pʰàk*  
 carrot one CL  
 ‘one carrot’ (5.4 yrs.)

(13) *kʰâopʰôt* *sǝŋ* *pʰàk*  
 corn two CL  
 ‘two corns’ (5.4 yrs.)

This strategy represents the child’s categorization of the given noun, for example in examples (12) and (13), a child categorizes carrot and corn as vegetables, using *pʰàk* ‘vegetable’ as the classifier for carrot and corn. Tuaycharoen (1984) called this type of strategy over-extension. The present research found this strategy to be used at the age between 4.0 and 4.6 years.

### D. Shape strategy

The shape strategy refers to the use of classifiers representing the shape of the entity referred to by the given noun, for example, the classifier *kǝŋ* which is normally used with a sphere-shaped entity is used with the noun *tʰúrian* ‘durian’ rather than the more appropriate classifier *lûuk/pʰôn*.

(14) *tʰúrian* *nùŋ* *kǝŋ*  
 durian 1 CL  
 ‘one durian’ (5.3 yrs.)

(15) *lûukbǝŋ* *sǝŋ* *wonklom*  
 ball 2 CL  
 ‘2 balls’ (5.4 yrs.)

In (14), a child used the classifier *kǝŋ* ‘lump’ with *tʰúrian* ‘durian’ rather than the more appropriate classifier, *lûuk /pʰôn* which is generally used for a sphere-shaped fruit .

The present study found this strategy to be used at the age 4.5-5.5 years. In Thai, there are several classifiers used with rounded entities, each of which has to be used with a certain set of nouns, for example, the classifier *lûuk* is normally used with fruits and ball-like entities, while the classifier *kǝŋ* ‘lump’ is normally used with stones, ice cubes, ice-cream scoops and clouds. Children at an early age may not use them appropriately. In example (15), a child used the classifier *wonklom* ‘circle’ with *lûuk bǝŋ* ‘ball’. Again, the specific classifier for ball is *lûuk /bai*.

### E. Specific part strategy

This strategy refers to the use of a noun referring to a specific part of the given object as a classifier, for example, in (16) and (17) the child used the noun *kʰăa* ‘leg’ and *kʰɛn* ‘arm’ which are observable parts of the objects *kâoʔi* ‘chair’ and *hùnjon* ‘robot’ as their classifiers respectively.

- (16) *káoʔi sǝŋ kʰǎa*  
 chair two CL  
 ‘two chairs’ (two legs) (3.7 yrs.)
- (17) *hùnjon nùŋ kʰǎen*  
 robot one CL  
 ‘one robot’ (one arm) (4.6 yrs.)
- (18) *mǝtʰrsai nùŋ kʰrúaŋ*  
 motorcycle one CL  
 ‘one motorcycle’ (one engine) (4.5 yrs.)
- (19) *naalíkaa nùŋ kʰrúaŋ*  
 clock one CL  
 ‘one clock’ (one engine) (4.6 yrs.)

Besides the observable parts, children may use nouns referring to salient parts of the given object as its classifier, for example, in (18-19), the child used the noun *kʰrúaŋ* ‘engine’ which is the salient part which may not be clearly observable from the outside of the objects *mǝtʰrsai* ‘motorcycle’ and *naalíkaa* ‘clock’ as their classifier. This strategy reveals that children’s selection of a specific classifier might be influenced by the most salient parts of the given object.

#### F. Specific classifier strategy

This strategy refers to the use of certain specific classifiers which are not general classifiers, repeated head nouns or general nouns and do not represent shape and specific parts of the given objects. This group of classifiers has to be used with a specific set of objects, for example (20), the classifier *ruən* is specifically used with the noun *naalíkaa* ‘clock’.

- (20) *naalíkaa nùŋ ruən*  
 clock one CL  
 ‘one clock’ (4.5 yrs.)

Because of its specific use which cannot be deduced from the physical feature of the given noun, it is hypothesized that the children might be trained to use them appropriately. This strategy is found rather late, at the age between 4 and 5 years.

#### G. Specific reason strategy

By this strategy, children may use their own specific reasons in selecting a classifier for the given object. The chosen form may not conform to the aforementioned strategies as in the following examples.

- (21) *kʰrúaŋbin nùŋ fáa*  
 plane one CL  
 ‘one plane’ (one sky) (4.10 yrs.)
- (22) *káoʔi sǝŋ máai*  
 chair two CL  
 ‘two chairs’ (two wood) (5.2 yrs.)

In example (21), a child used the noun *fáa* ‘sky’ as the classifier of the object *kʰrúaŋbin* ‘plane’. When asked, the child explained that he used the classifier *fáa* ‘sky’ because the plane flies in the sky. In example (22), the child used the noun *máai* ‘wood’ for the object *káoʔi* ‘chair’. Even though the

picture shown to him was not of a wooden chair, the child explained that the chair that he knows of is made of wood, so he would like to use the classifier *máai* ‘wood’ with the chair. In other words, children may use their own specific reasons when using classifiers.

#### 4.2 Order

The non-classifier stage and classifier stage begin at the age of three but the non-classifier stage ends roughly at the age of four. In the non-classifier stage at the age of three, the most frequently found strategy is the coding strategy, found in 7 children out of 12 (58.33%), followed by the number counting strategy, found in 5 children out of 12 (41.67%), the message strategy, found in 3 children out of 12 (25%), and the noun reduplication strategy, found in 2 children out of 12 (16.67%). These strategies disappeared at the age of 4 except for the coding and message strategies.

*Table 3: Stages of the acquisition of classifiers and strategies*

Strategies	3.0-3.11 yrs. (12 children) (100%)	4.1-4.10 yrs. (12 children) (100%)	5.1-6.0 yrs. (13 children) (100%)
<b>Non-classifier Stage</b>			
Noun Reduplication Strategy	2 (16.67%)	0 (0%)	0 (0%)
Number Counting Strategy	5 (41.67%)	0 (0%)	0 (0%)
Coding Strategy	7 (58.33%)	1 (8.33%)	0 (0%)
Message Strategy	3 (25%)	1 (8.33%)	0 (0%)
<b>Classifier Stage</b>			
Noun repetition Strategy	4 (33.33%)	5 (41.67%)	7 (53.85%)
General Classifiers Strategy	12 (100%)	12 (100%)	13 (100%)
Shape Strategy	2 (16.67%)	10 (83.33%)	13 (100%)
General Noun Strategy	0 (0%)	0 (0%)	2 (15.38%)
Specific Part Strategy	2 (16.67%)	2(16.67%)	3 (23.08%)
Specific Reason Strategy	1 (8.33%)	10 (83.33%)	2 (15.38%)
Specific Classifier Strategy	2 (16.67%)	8 (66.67%)	13 (100%)

At the classifier stage, six strategies were found at the age of three. Table 3 shows that the most frequent strategy used was the general classifier strategy. A hundred percent of the children at all ages preferred to use general classifiers. For children aged 4 and 5, the shape strategy is the second most used strategy while the specific classifier strategy is the third most used strategy. This shows that children acquire classifiers better as they grow older. It is interesting to note that the specific classifiers are those classifiers children learned and were taught by adults. It is obvious that they understand these classifiers and make better use of this strategy when they become older. The noun repetition strategy increases in usage following the children’s age. However, the general noun strategy, specific part strategy and specific reason strategy are strategies used by a small number of children and seem to depend on the individual. What is interesting is ten children (83.33%) at the age of 4 use the specific reason strategy and this strategy is dropped by the age of five. It is noted that children aged 4 tend to



use a variety of classifiers. It is possible to compare the specific reason strategy for children at the age of four with the trial and error stage as stated by Tuaycharoen (1984).

## **5 Conclusion**

In conclusion, the acquisition of classifiers can be divided into two stages: the non-classifier stage and the classifier stage. All strategies found in the non-classifier stage end by the age of four. In the classifier stage, children at all ages prefer to use general classifiers. Moreover, all children at each age use the general classifiers the most. The use of the noun repetition strategy, shape strategy and specific classifier strategy increases as children grow older. The specific strategy using specific classifiers needed to be taught to children such as the classifiers for houses and ships. Some strategies, however, seem to be idiosyncratic. These strategies are the general noun strategy, specific part strategy and specific reason strategy. Thus, the order of classifier acquisition is not clear-cut. More detailed investigation may shed light on this matter.

## **References**

- Balvin, Edith. 2009. *Cambridge Handbook of Child Language*. New York: Cambridge University Press.
- Carpenter, K. 1987. *How Children Learn to Classify Nouns in Thai*. Ph.D. Dissertation, Stanford University.
- Clark, Eve V. 2009. *First Language Acquisition*. Cambridge: Cambridge University Press.
- Rungrojsuwan, Sorabud. 2003. *First Words: Communicative Development of 9- to 24-month-old Thai Children*. Ph.D. Dissertation, Department of Linguistics, Graduate School. Bangkok: Chulalongkorn University.
- Susan Goldin-Meadow, Martin E.P. Seligman Rochel Gelman. 1976. *Language in the two year old*. University of Pennsylvania.
- Tuaycharoen, P. 1984. Developmental strategies in the acquisition of classifiers in Thai. In *Selected Papers from the international Symposium on Language and Linguistics*, ed. by P. Tuaycharoen et al., 203-211. Chiang Mai University.

# STRESS AND TONAL PROMINENCE IN CHAOYANG<sup>1</sup>

Suki S.Y. Yiu

*University of Arizona & University of Hong Kong*  
*syutji@connect.hku.hk*

## **Abstract**

This paper argues that the durational asymmetry predicted by the Iambic-Trochaic Law (Hayes 1985, 1995) for languages typically considered as stress languages is also found in complex tone languages such as the Chaoyang variety of Teochew (Southern Min, Hong Kong/Chaozhou). In Chaoyang, there are six phonemic tones, and tone sandhi can apply from both left and right. Given the bidirectional tone sandhi, duration and intensity measurements were obtained from ten disyllabic sandhi minimal pairs of two speakers to compare the relative prominence of different sandhi positions, citation positions, and sandhi-citation positions. Results show that a duration contrast holds for anterior type sandhi and an intensity contrast for posterior type sandhi, despite a tendency for higher intensity on the first syllable and longer duration on the second syllable across sandhi types. A metrical interpretation of prominence in complex tone languages is largely supported by the results, in addition to tonal prominence.

**Keywords:** tonal prominence, metrical prominence, tone and stress, bidirectional tone sandhi, Iambic-Trochaic Law, Chaoyang, Teochew

**ISO 639-3 codes:** nan

## **1 Introduction**

One fundamental insight in the literature of tone studies is that tonal behaviour is diagnostic of metrical prominence. When two or more tones are in contact, the ones which stay intact are considered to be carrying accentual prominence and have the power to spread their influence to adjacent tones or even tones in a wider scope, whereas the others are in metrically recessive positions which tend to modify, neutralise, or even completely drop the tones (Chen 2000:2 94). Maddieson (1978:341) uses the term ‘stress-related tonal stability’ to refer to assimilatory processes driven by tones which are associated to stressed syllables.

Even though the above is generally assumed to be true, the possibility of stress interacting with tonal dominance motivates examining the role of phonetic correlates of stress other than pitch, notably duration and intensity. A clearer understanding of the role of duration and intensity in signalling prominence in tone languages can help us evaluate the legitimacy of the adoption of stress, and hence feet, as an organizing constituent for tonal distribution. If the duration and intensity measurements match the behaviour of pitch (from tone sandhi) in terms of prominence, we would like to see how closely they observe the predictions of the Iambic-Trochaic Law, which forms the core of Metrical Stress Theory (Hayes 1995). Otherwise, we may conclude that pitch (or related sandhi phenomena) signals prominence in tonal languages independent of other stress correlates.

In this paper, I seek phonetic evidence for a metrical representation of tone in complex tone languages, particularly in terms of duration and intensity, which are especially relevant in defining the asymmetric foot inventory for iambs and trochees. The metrical view hypothesizes that duration and intensity cues are present on top of tone sandhi to determine whether a language has iambic or trochaic-

---

<sup>1</sup> I sincerely thank Diana Archangeli and Stephen Matthews for their invaluable comments and suggestions. Many thanks to Elaine Lau and Katie Lu for offering their knowledge on Teochew and identifying subjects, and to the Chaoyang subjects for providing the recordings. This paper has benefited from constructive feedback by the audience and two anonymous reviewers of Chula-ISSSEAL2017, and the audience of Manchester Phonology Meeting 25. I am very grateful to Fulbright and Universitas 21 for their generous financial support. All errors are my own.

like rhythmic structure. I present first-hand phonetic data from Chaoyang (Southern Min, Hong Kong/Chaozhou) to explore whether duration and intensity cues match bidirectional tone sandhi in signalling relative prominence of different positions within disyllabic units.

This paper gives a brief overview of the interaction between tone and stress in Chaoyang from a metrical perspective. In Section 2, I introduce the tone system of Chaoyang and the bidirectional tone sandhi. Section 3 lays out the basic expectation correlating sandhi location with duration/intensity. Section 4 details the hypotheses, predictions and results to see whether the phonetic measurements align with the tone sandhi behaviours. Section 5 discusses the results regarding the Iambic-Trochaic Law (Hayes 1985, 1995). Section 6 concludes the paper.

## 2 Tone system of Teochew (Chaoyang dialect)

There are six lexical tones in the Chaoyang variety of Teochew (Zhang 1979, 1981, 1982, 2016:18-19), as listed in the first column of Table 1. The six lexical tones shown here occur in sonorant-final syllables. There are also two checked tones, occurring in syllables that end with an unreleased final stop, which are shorter versions of the high tone and low tone shown in Table 1. The second column is a tone-letter representation of the tone numbers adopted from Yip (2001). Among the six citation tones, three of them are level tones: H(igh), M(id) and L(ow), and three of them are contour tones: HM, ML and LM.

*Table 1: Tones in Chaoyang*

Citation tone (Zhang 2016)	Yip (2001)	Anterior ( <i>T<sub>x</sub>T<sub>y</sub></i> )	Posterior ( <i>T<sub>x</sub>T<sub>y</sub></i> )
33	M	M	L
53	HM	ML	ML
31	ML	H	L
55	H	L	L
313	LM	M	L
11	L	M	L

Given any disyllabic unit, either the tone on the first or second syllable undergoes sandhi to another tone, regardless of the tonal content of the trigger. Italicised *T<sub>x</sub>* and *T<sub>y</sub>* in the headings of Table 1 indicate the sandhi tone in anterior type and posterior type sandhi respectively.<sup>2</sup>

Anterior type tone sandhi changes the tone on the first syllable. For example, ML changes to H when it is on the first syllable of a disyllabic unit, regardless of the tone of the second syllable. Posterior type tone sandhi, on the other hand, refers to the type of sandhi happening on the second syllable with the tone on the first syllable remaining unchanged. For example, when ML is on the second syllable of a disyllabic unit, it undergoes sandhi as L, no matter which tone docks on the first syllable.

Note also in Table 1 that the variety of tone in the posterior position is more limited than the anterior position since it only holds two different tones (L and ML), rather than four as in the anterior position (H, M, L and ML) or even six as in the citation tone inventory (H, M, L, HM, ML and LM). I will come back to the discussion of this in Section 5. The bidirectional sandhi phenomenon makes it possible to examine the relative prominence of syllables in different positions in a disyllabic unit across sandhi types, and how well (or not) the relative prominence matches the metrical view.

With this background, we are in position to understand the hypotheses and predictions for different comparisons based on the sandhi types, the methods for testing those predictions, and the data itself, acoustic measurements and statistical tests of ten minimal pairs.

<sup>2</sup> This kind of bidirectional tone sandhi is widely reported in Chinese languages, especially in the Wu dialects, to a lesser extent in Southern Min. Chen's (2000) book on tone sandhi patterns across Chinese dialects is an invaluable reference. This paper does not deal with the morphosyntactic side of the tone sandhi phenomenon; therefore, it cannot speak for its influence on arriving at such tonal outputs given the same tonal inputs. Instead, this paper focuses on exploring a metrical alternative to diagnose prominence by examining relevant phonetic measurements in terms of duration and intensity.

### **3 Hypotheses and predictions**

This section lays out the basic expectation correlating sandhi location with duration/intensity. I will fine-tune the basics for the specific tests being discussed in Section 4.

According to the IT-Law (Hayes 1985, 1995), iambic and trochaic rhythmic types tend to adopt different acoustic cues to signal prominence (or stress): Iambs use length while trochees use intensity. This predicts that the second syllable is longer than the first syllable in iambic units, whereas the first syllable is louder than the second syllable in trochaic units. In the case of tone sandhi, based on the correlation between tonal stability and prosodic prominence, syllables with citation tones are considered to be more prominent than those with sandhi tones. Since anterior sandhi changes the tone on the left, it is expected to pattern with the iambic rhythm. As the posterior sandhi changes the tone on the right, it is expected to behave like trochees.

- Hypothesis 1 tests whether anterior sandhi units adopt duration and posterior sandhi units adopt intensity to signal prominence.
- Hypothesis 2 tests whether syllables carrying sandhi tones are less prominent than syllables carrying citation tones.
- Hypothesis 3 tests whether there are differences in terms of prominence between the sandhi syllables in both sandhi types depending on the rhythmic type that they pattern with.

### **4 Phonetic evidence for a metrical representation of tone**

Ten disyllabic sandhi minimal pairs are selected for both sandhi types. They are identical in their segmental and tonal content before tone sandhi but differ in their tonal make-up depending on the type of sandhi taking place.

Two native speakers of Chaoyang were invited to make audio-recordings in a quiet room, reading 72 disyllabic units, among which there were ten sandhi minimal pairs (or 20 units). The disyllabic units were randomised and presented on paper for the speakers to read in a natural and casual way. Segmentation of the recordings, and extraction of syllable duration and maximal intensity were performed in Praat (Boersma and Weenink 2017). The duration and intensity data were then submitted to linear mixed-effects regression (LMER) models in R (R Core Team 2017) in order to determine whether the duration and intensity values in different sandhi positions, citation positions, and sandhi-citation positions are significantly different from each other. Each of the subsections below addresses one of the hypotheses.

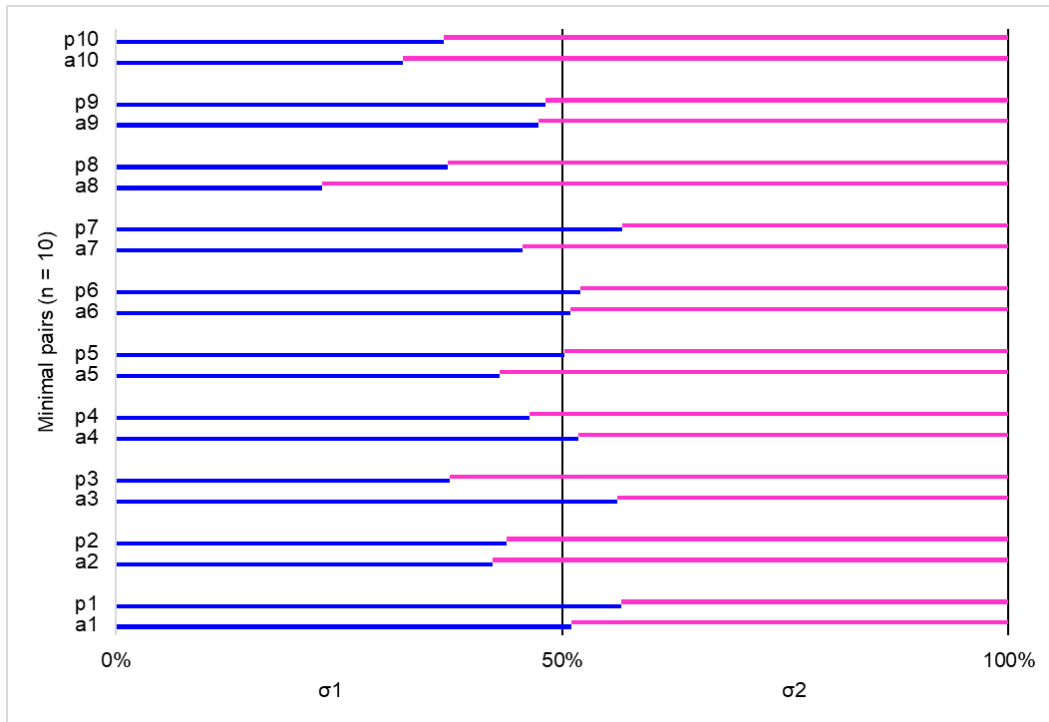
#### ***4.1 Within-unit duration contrast for anterior sandhi and intensity contrast for posterior sandhi***

The first comparison is the relative prominence in terms of duration within each unit for anterior type sandhi, and that of intensity within each unit for posterior type sandhi. The IT-Law predicts that the final syllable is longer than the first syllable in a trochee and the first syllable is louder than the final syllable in an iamb. If the predictions are correct, the second syllable [y] will be relatively longer than the first syllable [x] for anterior sandhi [TxTy], and the first syllable [x] will be relatively louder than the second syllable [y] for posterior sandhi [TxTy].

Figure 1 displays the duration and intensity measurements of the ten minimal pairs for both sandhi types. On the Minimal pairs axis, ‘a1-a10’ refer to the disyllabic units of anterior type sandhi which are numbered from 1 to 10 (see appendix for labelled tokens). The corresponding units of posterior type sandhi are ‘p1-p10’. For example, disyllabic unit 1 [sua tiŋ] refers to ‘the top of a hill’ when sandhi occurs on [sua] (anterior), but ‘(on) a hill’ when sandhi happens on [tiŋ] (posterior). ‘50%’ on the axis with ‘ $\sigma_1$ ’ and ‘ $\sigma_2$ ’ marks the midpoint of time (Figure 1a) and the midpoint of the total maximum intensity (Figure 1b) of the disyllabic units.

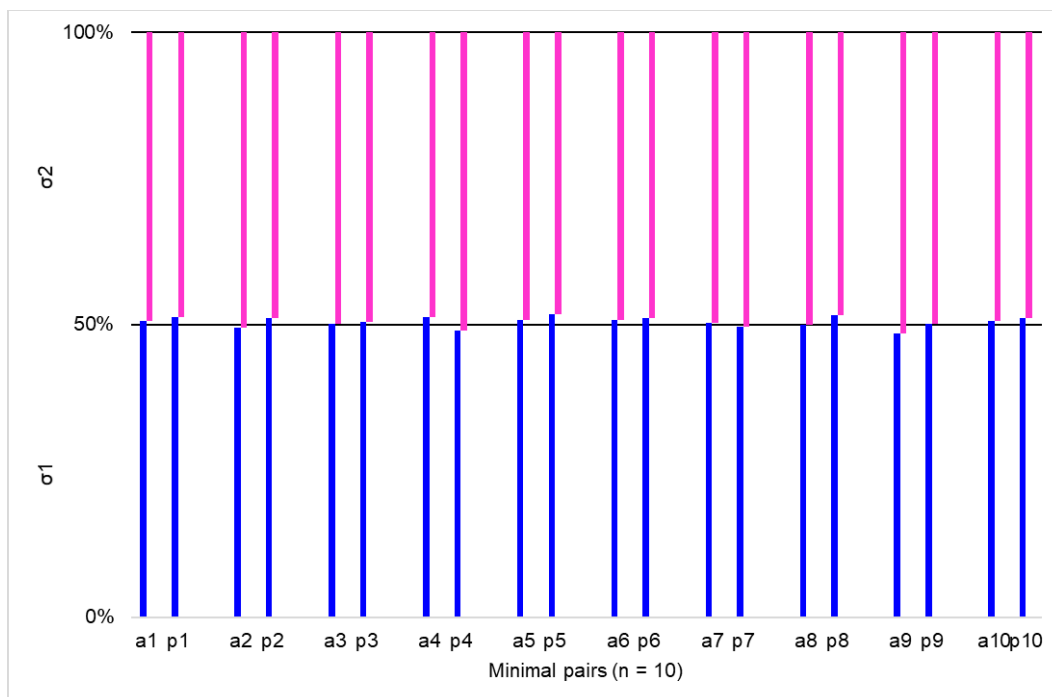
Figure 1: Duration and intensity of ten sandhi minimal pairs for speaker 2

a: Duration



As seen in Figure 1a, the second syllable overall tends to be proportionally longer than the first. Importantly, the second syllable tends to be proportionally longer in anterior type sandhi than posterior type sandhi (eight out of ten pairs). Also, when we take the data of both speakers into the LMER model which contained a fixed effect of syllable number (syllable 1, syllable 2) and random effects of word and speaker, results show that the second syllable is significantly longer than the first syllable for anterior types sandhi ( $p = 0.00123$ ).

b: Intensity



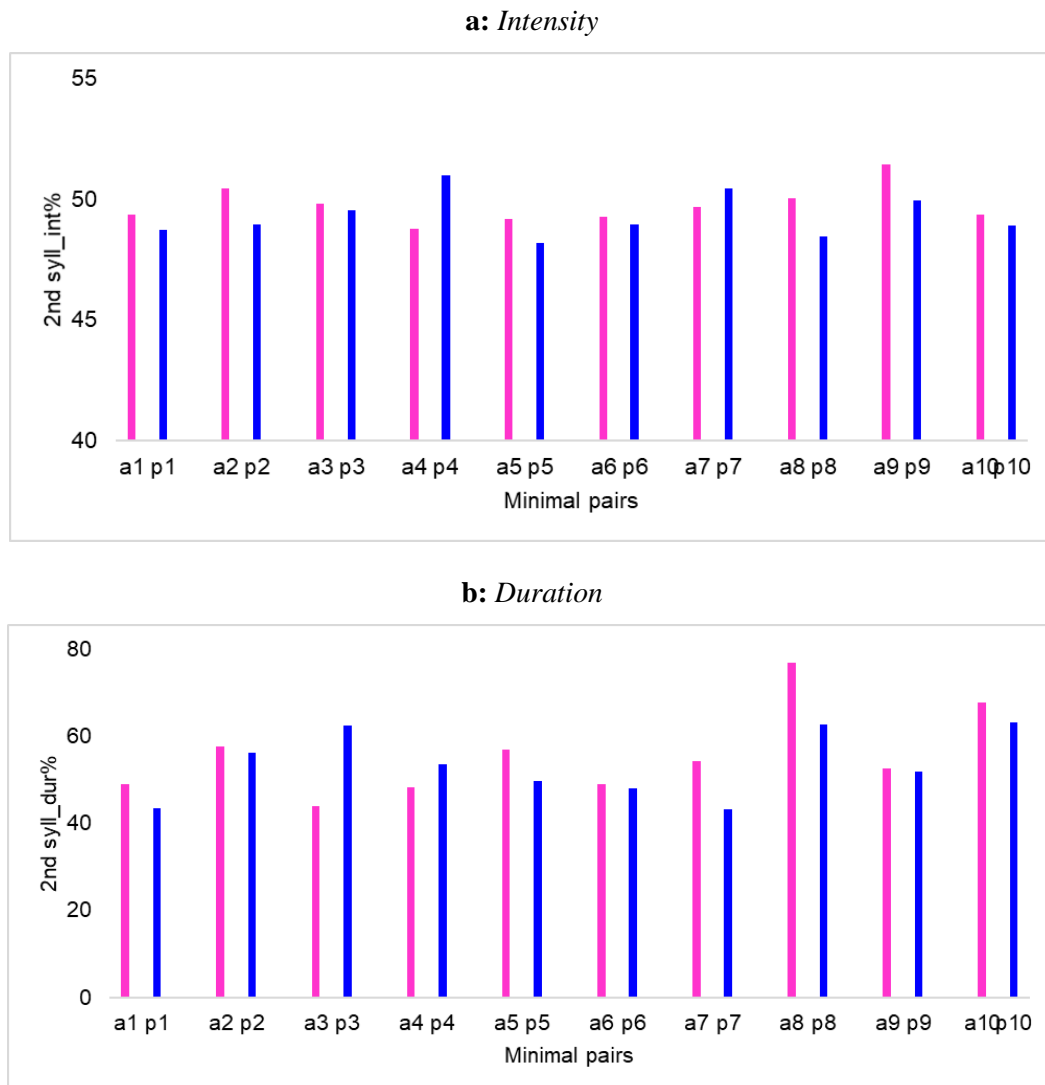
In Figure 1b, the first syllable tends to be proportionally louder in posterior type sandhi than in anterior type sandhi (eight out of ten pairs). And when we take the data of both speakers into the LMER model, results show that the first syllable is significantly louder than the second syllable for posterior sandhi ( $p = 2.11e-06$ ).

**4.2 Sandhi syllables are quieter and shorter than their citation counterparts**

With the minimal pairs, we can compare the sandhi syllables with their citation counterparts in the same position of a disyllabic unit. For instance, given a minimal pair [TxTy] and [TxTy], we can compare [Tx][Tx] and [Ty][Ty].

As sandhi syllables are considered to be less prominent than their citation counterparts based on lower tonal stability, they are expected to have shorter duration and lower intensity than the citation syllables. From the metrical perspective, it is also reasonable for syllables in the non-head position of a rhythmic unit to surface with shorter duration and lower intensity than those in the head position.

**Figure 2:** Intensity and duration of second syllables in posterior sandhi forms vs. their citation tone counterparts as found in anterior sandhi forms for speaker 2



For posterior type sandhi [TxTy], the sandhi syllable [y] is expected to be quieter and shorter than its citation counterpart in an anterior sandhi form [TxTy]. That is, the second syllable of the posterior type [y] is expected to be quieter and shorter than the second syllable of the anterior type [y].<sup>3</sup>

In Figure 2, the ten minimal pairs are shown from left to right. For each minimal pair, we compare the proportion of intensity and duration taken by the sandhi syllable in each posterior unit (p) against that taken by the citation syllable in each anterior unit (a).

The sandhi syllables are quieter (eight out of ten pairs in Figure 2a) and shorter (eight out of ten pairs in Figure 2b) than the citation syllables in general. In the LMER model which contained a fixed effect of sandhi type (anterior, posterior) and random effects of word and speaker, estimates of the posterior sandhi are lower than those of anterior sandhi for both intensity ( $52.751 < 54.762$ ,  $p = 0.225$ ) and duration ( $49.0347 < 49.5028$ ,  $p = 0.174$ ), even though the differences between the two sandhi types are not significant.

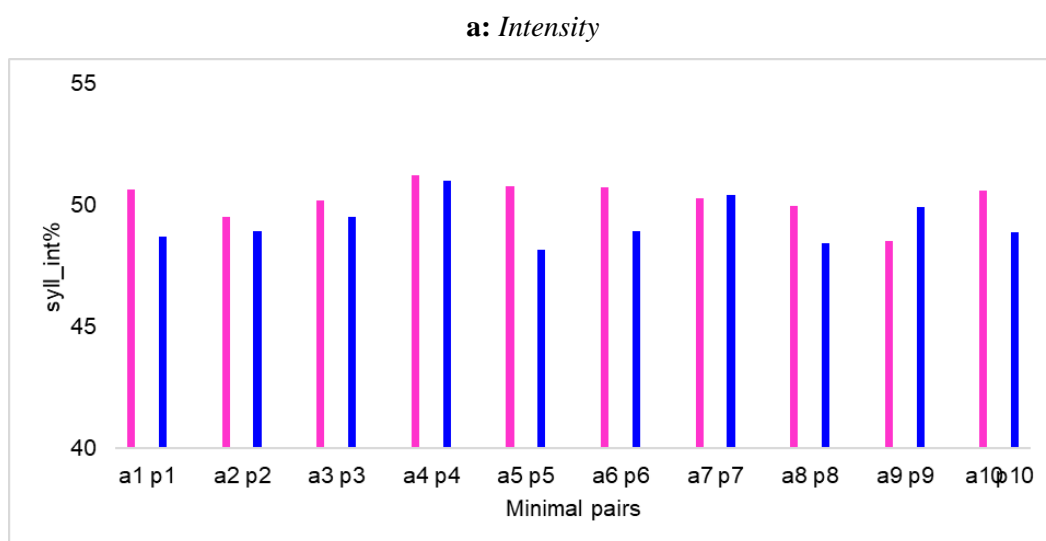
#### 4.3 Posterior sandhi syllables are quieter but longer than anterior sandhi syllables

Another cross-sandhi type comparison is comparing sandhi syllables in two sandhi types. When both syllables are prosodically weak in their respective sandhi type, we would like to know which position is weaker or stronger in terms of duration and intensity.

In terms of tonal stability, as the posterior type returns a narrower range of tones (two tones) than the anterior type (four tones) after sandhi, we expect the final position of the posterior type to be less prominent than the initial position of the anterior type, which will be reflected in the intensity and duration measurements, with no predictions on whether intensity and duration cues would be adopted differently for the two types of sandhi.

The metrical view makes hypotheses based on the IT-Law observed from how intensity contrast is associated with initial prominence (trochees) and duration contrast is associated with final prominence (iamb). Given that iambs prefer uneven feet, the non-head may suffer in terms of duration to meet the rhythmic ideal, but not in terms of intensity. Also, given trochees' tendency to adopt intensity to signal prominence, the non-head is weaker in terms of intensity, but not necessarily shorter. When all else is equal, this predicts that the final syllable of trochees is quieter but longer than the initial syllable of iambs. In other words, it predicts lower intensity for the second syllable of the posterior type [y] than for the first syllable of the anterior type [x] in [TxTy] (anterior) and [TxTy] (posterior). Duration is reversed, with greater duration for [x] than for [y].

**Figure 3:** Intensity and duration of the sandhi syllables in both sandhi types for speaker 2



<sup>3</sup> For the interest of space and presentation, we will see results of the comparison between syllables [x] and [x] in Section 5.

**b: Duration**

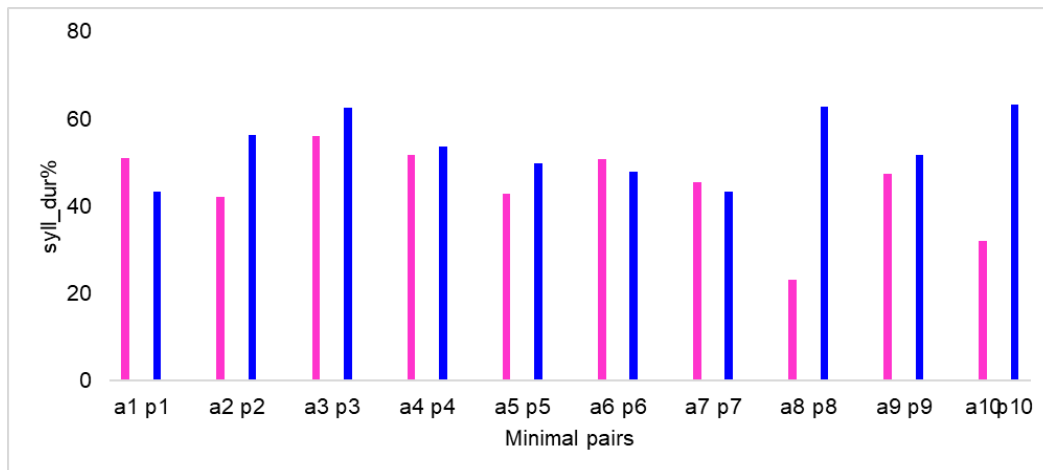


Figure 3a shows that sandhi syllables in posterior type tend to be quieter than those in anterior type by eight out of ten pairs. In spite of being quieter than the anterior sandhi syllables, Figure 3b shows that sandhi syllables in posterior type tend to be longer than those in anterior type by seven out of ten pairs, as predicted by the IT-Law. With the LMER model which contained a fixed effect of sandhi type (anterior, posterior) and random effects of word and speaker, results show that the differences in both intensity ( $p = 0.000117$ ) and duration ( $p = 0.00483$ ) are significant.

In sum, results presented in Section 4 largely agree with the durational asymmetry predicted by the metrical account, indicating that tonal prominence can be signalled by the typical phonetic correlates of stress along with pitch in terms of tone. This offers phonetic support for a metrical account of prominence in tone languages.

## 5 Discussion

In the previous section, I have presented results from comparing the relative prominence of syllables i) within sandhi types, ii) between sandhi syllables and their citation counterparts for the posterior type sandhi, and iii) between sandhi syllables of the two sandhi types. To complete a schematic representation of the relative prominence of different positions of a disyllabic unit, we still need comparisons between sandhi syllables and their citation counterparts for the anterior type sandhi, and between citation syllables in the two sandhi types.

Further measurements and LMER analyses show that results of the sandhi syllables and their citation counterparts for the anterior type sandhi are similar to those of the comparison for the posterior type presented in Section 4.2, in that the sandhi syllables tend to be quieter ( $50.4972 < 50.9653$ ,  $p = 0.174$ ) and shorter ( $45.238 < 47.249$ ,  $p = 0.225$ ) than the citation syllables, but the LMER results show the difference is not significant. Results of the comparison between citation syllables of the two sandhi types are similar to those for sandhi syllables of the two sandhi types in Section 4.3, where the citation syllable of the anterior type is quieter ( $49.5028 < 50.9654$ ,  $p = 0.000117$ ) but longer ( $54.762 > 47.248$ ,  $p = 0.00483$ ) than that of the posterior type. Incorporating these additional measurements and analyses, Table 2 displays a schematic representation of the relative prominence of different positions within a disyllabic unit in terms of duration (⌊) and intensity (x).

Table 2 shows that duration contrast holds for anterior type sandhi and intensity contrast for posterior type sandhi, despite a general tendency for higher intensity on the first syllable and longer duration on the second syllable across sandhi types. The phonetic results neatly reflect the durational asymmetry predicted by the IT-Law, where the anterior type sandhi and posterior type sandhi pattern with iambic rhythm and trochaic rhythm respectively. The duration and intensity measurements therefore support a metrical interpretation of prominence for complex tone languages like Chaoyang, in addition to the pitch-related cue in terms of tone sandhi.



**Table 2: Relative prominence within a disyllabic unit**

	$\sigma 1$	$\sigma 2$
Anterior type sandhi	_ xxx	_ _ _ _ xx
Posterior type sandhi	_ _ xxxx	_ _ _ x

For pitch, as shown in Table 1, the tone values of the posterior sandhi tones are much lower than their citation counterparts, indicating that the pitch on less prominent syllables is less marked. As for why the initial sandhi position in anterior type can accommodate more (and higher) tones than the final sandhi position in posterior type, positional faithfulness (Beckman 1998; Alderete 1999; Lombardi 1999) prefers the more prominent position of a prosodic unit to have more contrast. Even though both initial position of the anterior type and final position of the posterior type are the weaker position of their respective sandhi type in Chaoyang, the tone inventory supports the assumption that the initial position of the anterior type is more prominent between the two, and thus preserves tone contrast to a greater extent. Another possible explanation is foot-initial fortition, in that the initial position of a foot is strengthened, which agrees with the general cross-linguistic pattern of strengthening associated with the initial position of prosodic domains (Gordon 2011, citing Pierrehumbert and Talkin 1992; Byrd 1994; Keating et al. 2003, etc.)

## 6 Summary

This paper represents an exploration of a metrical approach to understanding how prominence is signalled in complex tone languages, motivated by the IT-Law via phonetic evidence from duration and intensity measurements. It has carefully gone through the two types of tone sandhi in the Chaoyang variety of Teochew, anterior type and posterior type, which correspond to iambic and trochaic rhythmic patterns respectively. A metrical interpretation of tone sandhi helps to understand the rhythm of complex tone languages, and to build a unified rhythmic account for languages as a whole.

Not only does Chaoyang represent a different case from the mainstream studies of linguistic rhythm on stress or simple tone languages with the use of metrical theory, its whole-tone replacement sandhi also differs from the ones previously reported for Chinese languages such as assimilation, dissimilation and spreading, enriching the pool of metrical phenomena.

With most descriptions of tone systems concentrating on the relatively well-recognised tone sandhi type in a language, the seemingly residual posterior sandhi type in Southern Min is often under-described and under-recognised for its importance in understanding the metrical structure of languages. Future studies may look into longer prosodic domains for more complex interactions between the two rhythmic types.

## References

- Alderete, John. 1999. Head dependence in stress-epenthesis interaction. In *The derivational residue in phonological Optimality Theory*, ed. by Ben Hermans and Marc van Oostendorp, 29-50. Amsterdam: John Benjamins.
- Beckman, Jill. 1998. Positional faithfulness. PhD dissertation, University of Massachusetts, Amherst.
- Boersma, Paul, and Weenink, David. 2017. Praat: doing phonetics by computer [computer program]. Version 6.0.19. <http://www.praat.org>. [accessed October 10, 2017].
- Byrd, Dani. 1994. Articulatory timing in English consonant sequences. Ph.D. dissertation, University of California, Los Angeles. [UCLA Working Papers in Phonetics 86].
- Chen, Matthew. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.
- Gordon, Matthew. 2011. Stress: phonotactic and phonetic evidence. In *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp et al., 924-948. Malden: Wiley-Blackwell.
- Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 11:429-446.

- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In *Phonetic Interpretation (Papers in Laboratory Phonology 6)*, ed. by John Local, Richard Ogden and Rosalind Temple, 143-161. New York: Cambridge University Press.
- Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17:267-302.
- Maddieson, Ian. 1978. Universals of tone. In *Universals of Human Language Vol. II*, ed. by Joseph Greenberg, 335-366. Stanford University Press.
- Pierrehumbert, Janet, and David Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, ed. by Gerard Docherty and Robert Ladd, 90-117. Cambridge: Cambridge University Press.
- R Core Team. 2017. R: A language and environment for statistical computing. Version 1.0.44. [accessed October 10, 2017].
- Yip, Moira. 2001. Tonal features, tonal inventories and phonetic targets. *UCL Working Papers in Linguistics*:161-188.
- Zhang, Sheng Yu. 1979. Chaoyang fangyan de chongdieshi. [Reduplication in the Chaoyang dialect]. *Zhongguo Yuwen* 2:106-114.
- Zhang, Sheng Yu. 1981. Chaoyang fangyan de yuyin xitong. [An outline of Chaoyang phonology]. *Fangyan* 1:27-39.
- Zhang, Sheng Yu. 1982. Chaoyang fangyan de xiangshengzi chongdie shi. [Reduplicated onomatopoeic particles in the Chaoyang dialect]. *Fangyan* 3:181-182.
- Zhang, Sheng Yu. 2016. *Chaoyang dialect study*. Social Sciences Academic Press (China).

## Appendix

Ten sandhi minimal pairs

		anterior		posterior	
1	[sua tɿŋ]	M.HM	‘hill top’	M.ML	‘on a hill’
2	[tsɛ̃ toi]	ML.HM	‘well bottom’	HM.ML	‘inside a well’
3	[lau hue]	M.HM	‘old partner’	LM.ML	‘old people’
4	[t <sup>h</sup> au toi]	L.HM	‘inside’	H.ML	‘inside the brain’
5	[ts <sup>h</sup> u tɿŋ]	H.HM	‘house top’	ML.ML	‘on a house’
6	[miŋ tɿŋ]	M.HM	‘surface of sth’	L.ML	‘on the face’
7	[tsi kai]	ML.H	‘the nearer one’	HM.L	‘the nearest one’
8	[zik t <sup>h</sup> au]	L.H	‘sun’	H.L	‘daytime’
9	[soi kai]	H.H	‘small thing(s)’	ML.L	‘small’
10	[tua t <sup>h</sup> au]	M.H	‘big head’	L.L	‘the bigger end’