

JSEALS Special Publication No. 11

**PAPERS FROM THE EIGHTH
INTERNATIONAL CONFERENCE ON
AUSTROASIATIC LINGUISTICS**



Edited by

Hiram Ring and Paul Sidwell



**UNIVERSITY of
HAWAII
PRESS**

© 2023 University of Hawai‘i Press
All rights reserved
OPEN ACCESS – Semiannual with periodic special publications
E-ISSN: 1836-6821
<http://hdl.handle.net/10524/52509>



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 are permitted, provided the author and the journal are properly credited.

Cover photo courtesy of Hiram Ring: Pnar speakers planting rice near Sohmynting, Meghalaya, North-East India.

JSEALS

Journal of the Southeast Asian Linguistics Society

Editor-in-Chief

Mark Alves (Montgomery College, USA)

Managing Editors

Nathan Hill (Trinity College Dublin, UK)

Sigrid Lew (Dallas International University, USA)

Paul Sidwell (University of Sydney, Australia)

Editorial Advisory Committee

Edith ALDRIDGE (Academia Sinica, Taiwan)

Nathan BADENOCH (Villanova University, USA)

Luke BRADLEY (University of Freiburg, Germany)

Marc BRUNELLE (University of Ottawa, Canada)

Christopher BUTTON (Independent researcher)

Kamil DEEN (University of Hawaii, USA)

Gerard DIFFLOTH (Cambodia)

Rikker DOCKUM (Swarthmore College, USA)

Ryan GEHRMANN (Payap University)

San San HNIN TUN (INALCO, 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)

Kenneth VAN BIK (California State University, Fullerton, CA)

Heather WINSKEL (Southern Cross University, Lismore, Australia)

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

Heather WINSKEL (Southern Cross University, Lismore, Australia)

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

<u>Introduction from the Volume Editors</u>	iv
<u>Statement from the JSEALS Editor-in-Chief</u>	v
Articles	
<u>A Typology of Grammatical, Local/Directional and Instrumental Markers in Kherwarian Languages</u>	1-14 Gregory D. S. ANDERSON and Bikram JORA
<u>Negation in Low Katu</u>	15-24 Thomas C. HUBER
<u>Subordination Strategies in Rucing Palaung</u>	25-52 LEE Wei-Wei
<u>Syllabification in Language Contact between French and Vietnamese</u>	53-71 Vera SCHOLVIN
<u>Proto-Pearic and the Role of Vowel Height in Creaky Register</u>	72-89 Paul SIDWELL
<u>Flagging and Indexing in Austroasiatic: A Typological Overview</u>	90-114 Tobias WEBER
<u>Secondary Verbs in Rumai, Palaung</u>	115-140 Rachel WEYMUTH
<u>Terminance in Bru</u>	141-154 Angus WHEELER

INTRODUCTION FROM THE VOLUME EDITORS

This special issue of the Journal of South-East Asian linguistics contains a selection of papers from the 8th International Conference on Austroasiatic Linguistics (ICAAL 8), held in Chiang Mai, Thailand on August 29-31, 2019 and hosted by The Myanmar Center at the Faculty of Humanities, Chiang Mai University. ICAAL conferences have been held since 1973, they provide an opportunity for scholars working on Austroasiatic languages to present and discuss their work. The production of this volume was delayed significantly by the Covid-19 pandemic, and so we are happy to finally make available these papers from that conference.

As noted in the previous proceedings volume (for ICAAL 7; JSEALS Special Publication No. 3), Austroasiatic languages are relatively diverse typologically and are located non-contiguously over a large geographical area stretching from eastern India to Vietnam. The papers in this volume reflect this diversity both in terms of geography (describing languages in India, South-East Asia, and in-between) as well as in terms of subject matter (phonological, morpho-syntactic, typological, and historical).

We would like to thank each of the authors and the anonymous reviewers who spent the time to assist with this publication. We would also like to sincerely thank Felix Rau for his work on this issue before he had to step down due to Covid-related complications. We also acknowledge the assistance and advice of Mark Alves, the JSEALS Editor-in-Chief, among others.

We hope that this publication will serve to showcase the diverse nature of Austroasiatic languages and highlight the interesting and important work of our colleagues. We also hope that it will spur fruitful discussion and encourage increased involvement in work on these understudied languages.

Hiram Ring
Singapore

Paul Sidwell
Nelligen, Australia

FROM THE JSEALS EDITOR-IN-CHIEF

This is the eleventh JSEALS Special Publication. The goal of JSEALS Special Publications is to share collections of linguistics articles, such as select papers from conferences or other special academic events, as well as to offer a way for linguistic researchers in the greater Southeast Asian region to publish monograph-length works.

This publication contains papers published for the International Conference of Austroasiatic Linguistics, its eighth meeting since its first in 1973 at the University of Hawai‘i. The volume contains eight papers in total: five on syntax, two on phonology, and one on historical linguistics. The languages covered in this volume are spoken in throughout the greater Southeast Asian region: Mainland Southeast Asia, Insular Southeast Asia, Southern China, and the Indian Subcontinent. Individual chapters cover languages in several branches of Austroasiatic: Mundic, Katuic, Bahnaric, Palaungic, Vietic, Pearic, and Palaungic. Moreover, one paper is a typological overview of Austroasiatic languages broadly. Thus, this volume represents a full geographic range of the Austroasiatic language family.

We are very pleased that JSEALS is able to support the half century of works coming from ICAAL and to contribute to the sharing of quality linguistic research in Greater Southeast Asia. We look forward to being able to produce such works for ICAAL in the future.

Mark J. Alves

January 15th, 2023

Montgomery College

Rockville, Maryland

A TYPOLOGY OF GRAMMATICAL, LOCAL/DIRECTIONAL AND INSTRUMENTAL MARKERS IN KHERWARIAN LANGUAGES*

Gregory D. S. ANDERSON and Bikram JORA

Living Tongues Institute for Endangered Languages

<livingtongues AT gmail DOT com>, <bikram DOT jora AT livingtongues DOT org>

Abstract

This paper describes the range of functional case-marking elements present in three Kherwarian languages (Santali, Mundari and Ho), one divergent dialect (Kera? Mundari) of the Mundari language and four endangered 'minor' Kherwarian languages: Asuri, Birhor, Bhumij and Korwa. The functions encompass grammatical case relations, local and directional concepts as well as notions of instrumentality and association. For some of these markers we also have clear indications of their historical source. Where possible, this description is further grounded in comparative studies with other Austroasiatic languages.

Keywords: case marking, Munda, Kherwarian, syntax, semantics

ISO 639-3 codes: sat, unr, hoc, asr, biy, kfp

1 Introduction

Kherwarian languages belong to the North Munda branch of the Munda language family. Kherwarian languages are mainly spoken in Jharkhand state and areas peripheral to this in border regions of Odisha, West Bengal and Chhattisgarh states and in diaspora in Tea Gardens of North-East India, e.g., in Assam. In this article, we give brief accounts of the case marking systems of three major Kherwarian languages (Santali, Mundari and Ho), one divergent dialect (Kera? Mundari) of the Mundari language and four endangered 'minor' Kherwarian languages: Asuri, Birhor, Bhumij and Korwa.

Kherwarian languages make use of a range of functional elements that encompass grammatical case relations, local and directional concepts as well as notions of instrumentality and association. Some of these grammatical elements are clearly secondary in origin, such as the obviously loaned objective case element *-ke* found in some dialects of Mundari and Ho and in Birhor, while others are clearly old, such as the genitive marker *-a(?)* with cognates in southern Munda languages as well. Santali (except varieties spoke in Assam) reflects the older system that lacks the *-ke* objective case, and rather has only a head-marking system of encoding subcategorized arguments, not a mixed head-/dependent-marking system seen in the other above-mentioned languages.

In Section 2 we discuss the different types of case systems and related phenomena that exist in Kherwarian languages. Section 3 overviews the multiple functions of *-te* and *-re* as elements which appears to function synchronically as case markers. Section 4 deals with an 'augment' morpheme which is found across varied branches of Austroasiatic family (Jenny et al. 2014) conventionally known as the linker element. Section 5 further elaborates the creation of a wide range of local and directional semantics where case elements may attach to relational nouns. Section 6 examines the adverbial use of words with case forms. Section 7 examines the functions of case markers and their uses as clausal

* The authors would like to thank the National Science Foundation Award #1844532 "Sora Typological Characteristics: Towards a Re-Evaluation of South Asian Human History" and a grant from the Zegar Family Foundation for work on Birhor that have made the research for this project possible. Data listed as field notes come from dozens of field trips over three decades across the Kherwarian-speaking area. NEH Grant PD-281083-21: Documentation and analysis of seven Munda languages and development of the Munda Virtual Archive and NSF grant 2041248: Words, phrases, and sentences at the interface of phonology and morphosyntax.

subordinators in complex sentences. Section 8 of the study presents a summary with a discussion of the findings and implications for further research.

2 Overview of the cases and related phenomenon in Kherwarian languages

Not all postnominal functional elements are to be considered case forms in Kherwarian. There are also true postpositions that encode various non-local/directional notions, such as the purposive postposition *nanġen* ‘for, because of’, e.g., Hasada? Mundari *kami nanġen* ‘for work (they went)’ (Osada 1992:49), which does not permit further case marking itself.¹

Mundari (Munda 1978:48)

- (1) *ale nanġen najal bai-ta-pe*
 1PL DAT/PURP/BEN plough share make-TAM-2PL
 ‘Make a plough share for us’.

2.1 Grammatical case marking of subcategorized arguments

In general, subject and object NPs of a sentence are unmarked but there is also a borrowed objective case marker (from Indo-Aryan Sadri/Sadani) *-ke* found in some dialects of Mundari and the rest of the Kherwarian languages except most Santali lects. In Santali, these two case roles are rather only marked in the verb in the form of agreement markers when the subject or object is animate.²

In sentences (2)-(3), *-ke* acts as the objective marker has Accusative function, marking the non-subject undergoer in a typical two argument transitive verb frame:

Objective = ‘Accusative’

Bhumij (Field notes)

- (2) *kula sukri=ke goi²=ke-d=i= a*
 tiger pig=OBJ kill=PFV.TR-TR/ACT=3OBJ=IND
 ‘the tiger killed the pig’

Bhumij (Field notes)

- (3) *iŋ hoŋo-ke goi²-ke-ji-a*
 I man-OBJ kill-PRF-3OBJ-IND
 ‘I killed the man’.

¹ We distinguish between three types of functional elements in Kherwarian languages that overlap to some degree in their semantic ranges but have formally distinct realizations. Terminology used in South Asian linguistic literature often fails to distinguish, in particular, between case marking and postpositions, a fact that is further complicated by the well-known grammaticalization-cum-prosodic integration cline whereby case forms often originate as postpositions and are drawn into the lexical head in a manner similar to the way auxiliary verbs often univerbate with lexical verbal heads as TAM suffixes. For the purposes of this study, we assume case markers to be phonologically bound functional elements, while postpositions at least in principle allow for prosodic independence even if their morphosyntactic locus is fixed and determined. A third functional element we call here relational nouns, these are typically spatial in sense, but may be further case marked by the bound case elements, while postpositions afford no such further inflection via case forms.

² Note that subject and object are typically encoded in a head-marking pattern on the verb across the Kherwarian languages. The dependent marking system encoded by case is typical in the higher contact varieties of Kherwarian. It is entirely lacking in Santali except in the speech of certain Assam Santali speakers who are heavily influenced by the regional Indo-Aryan tribal lingua franca Sadri/Sadani, where a mixed head/dependent encoding system is found.

However, as (4) shows, *-ke* also may mark the animate recipient in a ditransitive construction and thus its functional patterning is that of a canonical primary object marker in the Dryer (1986) sense, as in the following example from Ho. We therefore gloss this element as an objective case (OBJ).

Objective = 'Dative'

Ho (Pucilowski, 2013:147)

- (4) *dobro soba-ke fiʔi kul-e-i-ten-e*
 Dobro Soba-OBJ letter send-APPL-3OBJ-IPFV-IND
 'Dobro is sending a letter to Soba'.

2.2 Local/Directional and Instrumental case forms

Kherwarian languages make use of a number of case forms with local, directional and instrumental/comitative/sociative case-type functions. Several of these are common enough across the subgroup to infer that they were found in the proto-language system underlying the modern attested systems. These include an instrumental in *-te/-te*, a locative in *-re* and a comitative variably realized as *-loʔ/-loŋ/-la*; see Table 1.

Table 1: Common case forms found across Kherwarian languages

<u>Language</u>	<u>Instrumental</u>	<u>Locative</u>	<u>Comitative</u>
	<i>-te/-te</i>	<i>-re</i>	<i>-loʔ/-loŋ/-la</i>
Santali	+	+	-
Mundari	+	+	+
Keraʔ	(+) ³	+	+
Birhor	+	+	+
Bhumij	+	+	+
Ho	+	+	+

(**Key:** + marker present, (+) other uses; - marker absent)

Some examples demonstrating the functions of these case elements in various Kherwarian languages are offered in (5)-(7).

Instrumental

Ho (field notes)

- (5) *faku-te had-ke-d-a=e*
 knife-INS cut-PRF-TR/ACT-IND=3SG
 '(he) cut it with a knife'

Locative

Mundari (field notes)

- (6) *soma oʔaʔ-re menaʔ-i-a*
 Soma house-LOC COP-3OBJ-IND
 Soma is in the house'.

³ Note that Keraʔ Mundari has adverbial uses of *-te*, even with loans, so it appears to be quasi-productive.

Comitative

Birhor (field notes)

- (7) *ij era-ta?-la bazar sen=ka-n-a-ij*
 I woman-AUGM-COM market go=TAM-ITR/MDL-IND-1SG
 'I went to the market with the woman.'

Sticking for the time just to the functional categories that the cases instantiate and not the formal markers that realize them, we see a distribution of case elements across the Kherwarian languages as in Table 2.

Table 2: Functional case categories in Kherwarian

Language	dat	acc	obj	gen	ins	all	loc	abl	com
Santali	-	-	-/+ ⁴	+	+	+	+	+	+
Mundari	-	-	+	+	+	+	+	+	+
Kera? Mu.	+	+	-	+	?	?	+	+ ⁵	+
Birhor	-	-	+	+	+	?	+	+	+
Bhumij	-	-	+	+	+	+	+	+	+
Ho	-	-	(+)	+	+	+	+	+	+

(Key: + marker present; (+) other uses; - marker absent; -/+ marker variable; ? no evidence noted yet)

Turning to formal markers instantiating these case categories across the Kherwarian languages (and including Korku for North Munda comparisons too) we find the following distribution, as in Table-3:⁶

Table 3: North Munda Case markers

Language	PURP/DAT	ACC	OBJ	GEN	INS	ALL	LOC	ABL	COM
Santali	- ^h en			-ren -ak -reak	-te	-sen -sec	-te	-k ^h on	- ^h en - ^h ec
Mundari	(nan ^h gen)		-ke	-a?	-te		-re	-ate	
Kera? Mu.	-ta/-ke	-ke		-a? -ra?			-re	-se < IA	-lo ^h
Birhor			-ke	-a - ^h i(n) -ren	-te	-te	-re	-te	-lo
Bhumij				-ren	-te	-te	-re	-ate	-lo
Ho				-ren -a? -re-a?	-te	-te	-re	-(e)te	-lo?
Korku	-khe?	-khe?		-(a)?	-ten	- ^h ae	-en	-(a)ten -tan -te	-gon

⁴ Assam Santali uses this, Santali from other states does not.

⁵ In the form of *-se-* clearly a recent Indo-Aryan loan.

⁶ Given that we still lack clear data on some forms in some languages (Korwa, Asuri, Turi, Brijia, Karmali, Mahali), the reader should understand these observations to be preliminary. Moreover, given the lack of data in some areas, it is premature to offer too much in the way of speculation about the Proto-Kherwarian system yet, and we must be content with commenting simply on some of the more salient language contact effects such as borrowing of objective *-ke* from the tribal Indo-Aryan lingua franca Sadri/Sadani, or ablative *-se* from a Hindi-esque source (also likely Sadri/Sadani). The shift of *-ke* from an objective to a dative/accusative marker in Kera? Mundari might be explained as a substrate effect as the speakers originally were Kurukh speakers shifting to Mundari, since it is likely this is the reason that Kera? Mundari more consistently has subject markers at the end of verbs and not enclitic to the word immediately preceding the verb as in other Kherwarian languages. While plausible and likely, we cannot yet demonstrate a specific cause for this minor divergence in Kera? Mundari with respect to other Mundari varieties.

Thus, other directional cases may have a more restricted distribution but nevertheless occur in a range of Kherwarian languages such that one entertains the possibility that these may also be functional elements found in the ancestral proto-Kherwarian grammatical system. Such functional categories include the ablative (8) and the allative (9), both of which are of a sufficiently similar formal shape to the instrumental to ponder whether they may be historically connected to each other and to that older case element.⁷

Ablative

Korwa (Grierson LSI IV:158-159)

- (8) *buru-teŋ*
 mountain-ABL
 ‘from the mountain’.

Allative

Bhumij (Ramaswami 1986:105)

- (9) *Ramo kami-te sen-le-n-a*
 Ramo work-ALL go-ANT-ITR/MDL-IND
 ‘Ramo went to work’.

2.3 Genitive case forms in Kherwarian

Genitive case is complex in the Kherwarian languages. There are parallel sets of suffixes for genitive for animate and inanimate possessors and possessa. For example, the suffix *-aʔ* denotes alienable possession by an animate noun (10), with a likely cognate found in non-Kherwarian languages of Southern Odisha like Sora (11).⁸

Santali (field notes)

- (10) *iŋ ina[ʔ] dʒoɖʒom=ti ba=iŋ arub=aka-n=a*
 I I:GEN right =hand NEG=1 wash-PRF-ITR/MDL-IND
 ‘I didn’t wash my right hand’

Sora (Anderson and Harrison 2008:311)

- (11) *anin a-suʔuŋ-ban iar-re [...iar-le]*
 he GEN/POSS-house-ALL go-PST
 ‘He (elephant) went to his (big-frog’s) house.’

As for the other genitive suffixes in Kherwarian, both *-reaʔ/raʔ*, and *-ren* indicate alienable possession by an inanimate noun and likely historically incorporate the locative element *-re*. Furthermore, the distinction between *-reaʔ/-raʔ*, and *-ren* is made on the basis of the animacy of the possessum, the latter form *-ren* marking an animate possessum (12), the former *-reaʔ/-raʔ* an inanimate possessum (13)-(14) in such contexts.

⁷ Note also however that Mahali has *-kete* ABL.

⁸ Note that in Sora however, the genitive appears to formally be a prefix or proclitic on the possessed noun, not a suffix/enclitic on the possessor. This may reflect a secondary boundary reanalysis of a former clitic that appeared between the possessor and possessum: it was resolved to a suffix/enclitic on the possessor in Kherwarian but a prefix/proclitic on the possessum in Sora. Determining the morphotactics of this element, and thus resolving whether these should be considered affixes or clitics in the synchronically attested system is the subject of ongoing research.

Ho (Pucilowski, 2013:91)

- (12) *ente hatu-ren munda=ko keja-li?-je*
 then village-GEN chief=3PL call-ANT:TR:3SG-IND
 ‘Then they called the village chief’.

Ho (Pucilowski, 2013:47)

- (13) *ka:ni-reja? nutum=do...*
 story-GEN name=FOC
 ‘The story’s name...’

Kera? Mundari (Kobayashi & Murmu 2008:188)

- (14) *dura-ra? samne-re tinggu alo tinggu-n-am*
 door-GEN proximity-LOC stand PROH stand-ITR/MDL-2SG
 ‘don’t stand near the door (lit. in the door’s proximity) at all!’

Turning now to specific form-function pairings in the grammatical case domain, we find the following distribution of case elements (or their lack) across the Kherwarian languages in our data set summarized in Table-4.

Table-4: Form-function pairing distribution of grammatical case markers in Kherwarian languages

Language	-Ø.NOM	-Ø.ACC	-ke (OBJ)	-a? (GEN)	-re(C)a? (GEN)	-ren (GEN)	-ra? (GEN)
Santali	+	+	+/- ⁹	+	+	+	-
Mundari	+	+	+	+	+	+	(+)
Kera? Mu.	+	-	-	+	?	(+) ¹⁰	+
Birhor	+	+	+	+	?	+	?
Bhumij	+	-	+	+	+	+	-
Ho	+	+/-	+/-	+	+	+	-

(**Key:** + marker present, (+) other uses, - marker absent, +/- marker variable and ? no evidence noted yet)

3 Multiple functions of *-te* and *-re*

As noted above, in the directional/local and instrumental domains, we find at least two formal kinds of elements which appears to function synchronically as case markers, specifically the locative in *-re* and the multi-functional allative/instrumental in *-te*. This latter is sometimes attributed adverbial semantics in the analysis of specific forms in individual Kherwarian languages, and indeed expressions with both formal elements function as adverbial forms across the Kherwarian languages. Some examples demonstrating this range of functions are offered in (15)-(16).

Asuri (Grierson LSI IV:140-142)

- (15) (a) *kami-ku-te* (b) *maraj-re* (c) *tayom-re* (d) *kami-te* (e) *disum-te*
 worker-PL-ABL future/big-LOC after-LOC work-INS country-ALL
 ‘from the servants’ ‘in the future’ ‘after’ ‘with work’ ‘to the country’

⁹ As previously mentioned, *-ke* is found as an objective case marker in Santali varieties spoken in Assam where Sadri is strongly dominant, but not in other varieties of Santali.

¹⁰ Kobayashi and Murmu (2008: 176) consider *-ren* to be an adjectival possessive marker, not a case element per se.

Ho (Ramaswami 2007:81, 9-10, 109)

- (16) (a) *bugi-te* (b) *diri-te* (c) *gaḍa-te*
 good-ADV stone-INS river-ALL
 ‘well’ ‘with a stone’ ‘to the river’

In specific languages we find a widening of the functions of *-re* as well, not just as a locative or in some adverbial expressions. Thus, in the negative copula construction, the locative *-re* may function as a genitive (17) even when both possessor and possessum are animate (18).

Mayurbhanj Ho (Anderson et al. 2008: 206)

- (17) *ape hatu-re numu ʃie*
 2PL village-LOC/GEN name what
 ‘What is the name of your village?’

Mundari (Osada 1999:67)

- (18) *en gomke-taʔ-re hon-ko ban-ko-a*
 DEM gentleman-AUGM-LOC son-PL NEG.COP-3PL.OBJ-IND
 ‘That gentleman has no sons.’

The case element *-re* may also function as a comitative in Ernga Korwa in (19); in fact in the following example, we find both the genitive and the comitative functions of *-re*.

Ernga Korwa (Grierson LSI IV:164)

- (19) *tə bərah bahin-re mit'-hoṛ ədɔ-re dohon-a*
 then 12 sister-LOC one-person child-COM PST.COP-IND
 ‘only one of the twelve sisters was with a child.’

While canonically *-re* is locative and *-te* is allative (or ablative), the semantics of directional serial verb constructions in individual Kherwarian languages can find the use of *-re* replacing a form where functionally one would expect *-te* in an allative function, such as the following forms involving *nir* ‘run’ in Ho and Birhor (20)-(21).

Ho (Pucilowski 2013:98)

- (20) *daru-re=? nir-deʔ-eja-n-a*
 tree-LOC=3SG.SBJ run-climb-PRF.ITR-ITR/MDL-IND
 ‘He ran up into a tree’.

Birhor (field notes)

- (21) *sukri bir-re nir-bolo=a-n-a-e*
 pig forest-LOC run-run.into=TAM-ITR/MDL-IND-3SG.SBJ
 ‘The pig ran into the forest’.

In most Kherwarian languages (perhaps all?), there is separate morphosyntax for human and non-human possessa in copula constructions: Animate possessa in present tense forms are treated like objects whereas in negative past forms it is encoded as subject in the copular verb form. Compare (22) and (23). In the former, the referent is encoded as an object in the verb and in the latter as a subject and thus as is typical in Kherwarian languages, formally marked by an enclitic on the negative particle preceding the verb form.

Birhor (Field Notes)

- (22) *ij=ʈhi tin sukri menaʔ-ku-a*
 1SG=GEN.NHUM.ANIM three pig COP-3PL.OBJ-IND
 ‘I have three pigs’.

Birhor (Field Notes)

- (23) *ij-ren bəria majō kə=kin təhiken-a*
 1SG-GEN.HUM.PSM two daughter NEG=3DL.SBJ PST.COP-IND
 ‘I didn’t have two daughters’.

Korku shows similar patterns to Kherwarian languages, and thus present positive copular forms treat the animate possessa as objects, but negative copular forms remain unmarked, as subject marking is not found in Korku, only object encoding (24)-(25).

Korku (field notes)

- (24) *ij-en bari koŋje-kin ʈa-kin*
 1SG-GEN/DAT two daughter-DL COP-3.DL.OBJ
 ‘I have two daughters.’

Korku (field notes)

- (25) *ij-en bari koŋje-kin bən*
 1SG-GEN/DAT two daughter-DL NEG.COP
 ‘I do not have two daughters.’

4 Augment

The reader may have noted an element glossed as AUGM in some of the forms above. This augment takes the form *-ta(?)* and occurs between the stem and the case marker. It sometimes gets translated ‘side’ in context, but it is typically completely opaque or does not appear to convey any meaning at all, and thus we gloss it here as an augment [AUGM]. Case forms attach to this augmented stem rather than the bare stem as is otherwise found. This can occur with a range of case elements, such as the ablative (26), the objective (27) or the locative (28).

Ho (field notes)

- (26) *aiŋ era-taʔ-te ʈola=iŋ idi=ke-q-a*
 1SG woman-AUGM-ABL basket=1SG take=PRF.TR-TR/ACT-IND
 ‘I took the basket from the woman’.

Birhor (field notes)

- (27) *ij uni hoʈ-ʈaʔ-ke ʈoŋki ema-d-iŋ*
 1SG DEM man-AUGM-OBJ basket give:APPL-TR/ACT-1SG.SUBJ
 ‘I gave the basket to the man’.

Ho (field notes)

- (28) *raʈa-ta-re kula giti-tan-a=e*
 courtyard-AUGM-LOC tiger sleep-IPFV-IND=3SG.SUBJ
 ‘the tiger is sleeping in the courtyard’.

In Kera? Mundari this ‘augment’ or something meaning ‘-side, near’ can be used as a ‘dative’/‘allative’ or purposive element too. This does not require any further functional element (29).

Kera? Mundari (Kobayashi and Murmu 2008:171)

- (29) *nia kifri am-ta=ñ au-ta-r-a*
 this cloth you-AUGM/BEN=1SG.SUBJ bring-PRF-TR/ACT-IND
 ‘I brought this cloth for/to you’.

Compound elements like *-ta-re*, *-ta-te*, etc., may in part reflect a recently grammaticalized element in Kherwarian meaning ‘side, near’, but may also in part historically be conflated with the older ‘linker’ element found across varied branches of the Austroasiatic family (Jenny et al. 2014).¹¹

5. Relational noun plus case

Local/directional case elements in Kherwarian may attach to local/directional relational nouns to create a wide range of such local and directional semantics as is typically found in a range of OV languages of Eurasia, such as Turkic languages (32). Some of these forms marked in this manner function adverbially as well. Forms contrasting locational and directional meanings using the relational nouns in combination with the case markers may be found. The system found typically contrasts a stative locative *XX-re* with a directional allative in *XX-te*, where *XX* is the relational noun with more specified locational/directional semantics (30).

Santali (Ghosh 2008:49)

- (30) a. *latar-re* b. *latar-te* c. *ʃetan-re* d. *ʃetan-te*
 below-LOC below-ALL above/top-LOC above/top-ALL

Some examples of the use of relational nouns in case marked forms in Santali are offered in (31).

Santali (Ghosh 2008:49)

- (31) a. *oʃakʰ bʰitri-re* b. *oʃakʰ ʃetan-te*
 house inside-LOC house above/top-ALL
 ‘inside the house’ ‘to above the house/to the top of the house’
- c. *dʰiri latar-re* d. *ɖuŋri ʃetan-re*
 stone under-LOC hill above/top-LOC
 ‘under a stone’ ‘above/on top of the hill’

(32) *Xakas* (Turkic, Siberia; Field Notes, Anderson 1998:7)

ara ayas
 between tree
 ‘between’ ‘tree’

ayas ara-zun-da *ayas ara-zun-daŋ*
 tree between-3POSS-LOC tree between-3POSS-ABL
 ‘in between the trees’ ‘from between the trees’

¹¹ This issue is a complicated one and determining the developmental paths of functional elements inherited from proto-Austroasiatic into Munda and its various sub-branches requires better data in order to have a clearer picture of the range of functions it might have. Indeed, we could just be dealing with a chance correspondence given the commonness of the phonological segments involved (*ta*), or one that partly entails an innovation and partly a convergence with an older form. But given that neither the diachronic phonology nor morphology of Munda have been adequately resolved, whether or to what degree this augment in Kherwarian reflects the proto-AA ‘linker’ must be resolved in the future, so we stick to the neutral term ‘augment’ here, as the form is largely semantically opaque.

6 Adverbial uses

Adverbial uses of words with case forms involving *-te* (and *-re*) may also be found in various Kherwarian languages. This adverbial use is (or was) a quasi-productive formation, as in the following form from Kera? Mundari, which attaches to an obviously borrowed element.¹²

Kera? Mundari (Kobayashi and Murmu 2008:175)

- (33) *tʰi:k-te* *ajum-e-a-pe*
 good-ADV listen-3OBJ-IND-2PL.IMP
 ‘(you guys) listen [to him] carefully’.

In Santali, the sequence *-katε* can have an adverbial sense when used with elements not functioning as verbs. Historically this appears to be a TAM element *-ka* in combination with the case element or adverbializer *-tε*.

Santali (Ghosh 2008:50)

- (34) *phəsiara-katε* *hatao-ke-d-a-e*
 deceit(ful)-ADV get-PRF.TR-TR/ACT-IND-3SG
 ‘he got possession of it deceitfully’.

Note also in Santali some elements historically of the shape *XX-te/XX-te* function synchronically as postpositions, e.g., *iə-te* ‘owing to’; *-sāo ~ -sāo-te* ‘together, in association with’, this latter element formally consisting of a loan word plus *-te*.

7 Sentential uses of case in Kherwarian: Complex sentences, subordination, complementation

The last major set of functions of case markers in Kherwarian languages we examine here is their use in complex sentences, subordination and in some instances complements too. The most common such use of case elements in such complex sentence formations include the inclusion of *-te* (here glossed as an abstract adverbial case element as *-K*) with a TAM marker (e.g., *-ke-* or *-ka-*) and an optionally encoded voice/valence marker (*-n/-d-*) to form ‘conjunctive’ or perfect participles in an adverbial-type subordination system. In (35a) we find an intransitive/middle form (*-ka-n-te*) and in (35b) a transitive/active form (*-ke-d-te*) from Asuri:

Asuri (Grierson LSI IV:140-142)

- (35) a. *birid-ka-n-te* b. *džavar-ke-d-te*
 arise-TAM-ITR/MDL-K collect-TAM-TR/ACT-K
 ‘having arisen’ ‘having collected’

In the Korwa form in (36), we see an example of a form lacking the voice/valence marker. Note that several researchers previously have considered elements like *-ka-te* to be a single element marking conjunctive participles *-katε*, but forms like (35a-b) show that the internal structure as deriving from a case marked verbal form grammaticalized into this function also holds synchronically.

Korwa (Grierson LSI IV:152, 154, 158-159)

- (36) *džom-ñu-ka-te=bun* *susun-a*
 eat-drink-TAM-K=1PL ~dance-IND
 ‘having eaten and drunk we danced’.

¹² The word *tʰi:k* is from Sadani/Sadri *tʰikka* meaning ‘good, ok, nice, well done, all right’, etc.

languages like Ho in (41) are highly reminiscent of relative-type clauses with a null-head found in Sora (42).

Sora (Ramamurti 1931:50)

- (42) *anin-dʒi ə-ti'-l-am-ən-dʒi-ate*
 3PRON-PL DEP-give-PST-2SG.UND-N.SFX-PL-FOC
 'what they gave you'.

Finally, one last construction found in Kherwarian languages that seems to incorporate a subordinating function of case markers when used with an otherwise verbal element is the compound conditional element *-re-do* that appears to include the locative case element *-re* with the emphatic/focalizer *-do*, which attaches to a TAM- and voice/valence marked verb to encode conditional clauses (43).

Ho (Burrows 1915)

- (43) *agu-le-d-re-do-ij*
 bring-ANT-TR/ACT-COND[LOC-EMPH]-1SG.SBJ
 'if I will bring it'.

8 Summary

Kherwarian languages make fairly extensive use of grammatical, local/directional and instrumental type case forms in a wide range of functional domains. Grammatical cases that reflect subcategorized argument roles are quite restricted and of quite recent origin, and many varieties of Santali lack them altogether, retaining instead the original proto-Kherwarian head-marking preference. Typically showing a nominative-accusative alignment, the NP in the subject role typically remains unmarked in Kherwarian languages,¹³ while many have borrowed an objective case marker from Indo-Aryan Satri *-ke*, that encodes patients with two argument/transitive verbs and recipient with three argument/ditransitive ones, in a typical primary object pattern in the Dryer (1986) sense. In nominal dependency formations, Kherwarian languages show a typologically unusual system of genitive case marking that considers the animacy specifications of both possessor and possessum. Instrumental and comitative/sociative case forms are also found across Kherwarian languages. Two or three common local or directional case forms, a locative and an allative, and in many an ablative case form—the latter two often formally identical to the instrumental—may be found. These have a wide range of functions including adverbial or adjunct marking of local and directional noun phrases. Such local and directional case elements have also been incorporated into various relational nouns, some of these latter borrowed from Indo-Aryan, that further expand the range of semantic domains creating nuanced directional and local senses. The local and directional case elements *-re* and *-te* may also appear with a bridge or augment morpheme that itself may have deep historical roots in Austroasiatic. These same two case forms have further been grammaticalized for use in the verbal and sentential domains, encoding a range of complement, subordinate, relative and conditional clauses when used with TAM-marked or unmarked verb forms of various sorts. Some of these seem to be recent innovations likely based on areally common models for subordination strategies, but a few seem like they belong to an earlier historical layer of the languages, with possible parallels in non-Kherwarian languages. Determining the precise origin and paths of development of these remain an object for future research. The present study is simply a first step into expanding and refining our understanding of the nature of case marking in Kherwarian Munda languages.

¹³ Peterson (2017) has suggested some uses of the instrumental on agents in Santali may be an incipient development of ergative patterning based on local IA models; Hindi varieties, as well as many languages cross-linguistically have regrammaticalized former instrumental case markers as ergative case markers.

Abbreviations

ABL	ABLATIVE
ACC	ACCUSATIVE
ACT	ACTIVE
ADV	ADVERBIAL
ANIM	ANIMATE
ALL	ALLATIVE
ANT	ANTERIOR
APPL	APPLICATIVE
AUGM	AUGMENT
AUX	AUXILIARY
BEN	BENEFACTIVE
COM	COMITATIVE
COND	CONDITIONAL
COP	COPULA
DAT	DATIVE
DEM	DEMONSTRATIVE
DEP	DEPENDENT
DL	DUAL
EMPH	EMPHATIC
FOC	FOCUS
GEN	GENITIVE
HUM	HUMAN
.I	INCLUSIVE
IMP	IMPERATIVE
IND	INDICATIVE
INS	INSTRUMENTAL
IPFV	IMPERFECTIVE
ITR	INTRANSITIVE
K	CASE ELEMENT, FUNCTION VARIED/INDETERMINATE
LOC	LOCATIVE
MDL	MIDDLE
NEG	NEGATIVE
NHUM	NON-HUMAN
OBJ	OBJECT[IVE]
PFV	PERFECTIVE
PL	PLURAL
POSS	POSSESSIVE
PROH	PROHIBITIVE
PRON	PRONOMINAL
PRF	PERFECT
PSM	POSSESSUM
PST	PAST
PURP	PURPOSIVE
SBJ	SUBJECT
SFX	SUFFIX
SG	SINGULAR
TAM	TENSE/ASPECT/MOOD
TR	TRANSITIVE
=	CLITIC BOUNDARY
1	FIRST PERSON
2	SECOND PERSON
3	THIRD PERSON

References

- Anderson, Gregory D. S. 1998. *Xakas*. München: Lincom Europa.
- Anderson, Gregory D. S. 2002. Case Marked Clausal Subordination in Burushaski Complex Sentence Structure. In *Studies in Language* 26.3:547-71.
- Anderson, G. D. S. and K. D. Harrison 2008. Sora. In G. D. S. Anderson (ed.) *The Munda Languages*. Routledge Language Family Series. London: Routledge. 299-380.
- Anderson, G. D. S., T. Osada and K. D. Harrison. 2008. Ho and the other Kherwarian languages. In G. D. S. Anderson (ed.) *The Munda Languages*. Routledge Language Family Series. London: Routledge. 195-255.
- Burrows, Lionel B. 1915. *Ho grammar (with vocabulary)*. Calcutta: Catholic Orphan Press, 1915. Pp. vii, 194. [Reprinted as *The grammar of the Ho language: an Eastern Himalayan dialect*. New Delhi: Cosmo Publications 1980.]
- Dryer, Mathew 1986. Primary objects, secondary objects and antitativity. *Language*, 62:808-845.
- Ghosh, A. 2008. Santali. In G. D. S. Anderson (ed.) *The Munda Languages*. Routledge Language Family Series. London: Routledge. 11-98.
- Grierson G. 1898-1928. *Linguistics Survey of India Vol. IV*. Calcutta: Office of the Superintendent of Government Printing, India.
- Hoffmann, J. 1930-1978. *Encyclopedia Mundarica*. Patna: Government Press.
- Jenny, Mathias, Tobias Weber and Rachel Weymuth. 2015. The Austroasiatic Languages: A Typological Overview. In Jenny, Mathias and Paul Sidwell (eds.). 2014. *The Handbook of Austroasiatic Languages*. 2 volumes. Leiden: Brill. 13–143.
- Kobayashi, Masato & Ganesh Murmu 2008. *Kera Mundari*. In Gregory D. S. Anderson (ed.) *The Munda Languages*. London & New York: Routledge.
- Munda, Ram Dayal. 1978. *Mundari Grammar*. Ranchi: Mundari Sahitya
- Osada, Toshiki. 1992. *A Reference grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Osada, T. 1999. Experiential constructions in Mundari. *Gengo Kenkyu: Journal of Japanese Linguistic Society* 115:51-76.
- Peterson, John M. 2017. Jharkhand as a “linguistic area” - Language contact between Indo-Aryan and Munda in eastern-central South Asia.” Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press. 551-574.
- Pucilowski, A. 2013. *Topics in Ho Morphophonology and Morphosyntax*. Ph.D. Dissertation, University of Oregon.^[1]_[SEP]
- Ramamurti, G.V. 1931. *A Manual of the So:ra (Savara) Language*. Madras: Government Press
- Ramaswami, N. 1986. *Bhumij Grammar*. Mysore: CIIL.
- Ramaswami, N. 2007. *Ho Grammar*. Mysore: CIIL.

NEGATION IN LOW KATU

Thomas C. HUBER

University of Zurich

<thomas DOT huber3 AT uzh DOT ch>

Abstract

In Low Katu (or Western Katu; ISO 639-3: kuf) there are five common words used to mark negation: *kah*, *məʔ*, *jiəʔ*, *ʔeh* and *ʔɔ:ʔ*. This variety in negators hints at differential syntactic or semantic uses. In this paper I illustrate the syntactic properties of these negators and, where possible, describe what semantic or pragmatic backgrounds they might have. I do this by comparing negative sentences from Katu folk tales and stories and investigating how they behave with respect to the typology of negation. Understanding the negation of Low Katu can unveil aspects on the scarcely researched syntactic behavior of this language, for instance on the position of verbs. This paper is intended to be the groundwork for further, more corpus-based research on negation or other grammatical aspects of Low Katu.

Keywords: Katu, negation, typology, Austroasiatic, predicate, syntax

ISO 639-3 code: kuf

1 Introduction

1.1 Low Katu

Katu is a minority language in Vietnam and Laos belonging to the Katuic branch of the Austroasiatic language family. It is spoken by people who live in the mountainous jungle of Quang Nam province in Vietnam, and Xe Kong and Salavan provinces in the Lao People's Democratic Republic. The dialect of Low Katu, or Western Katu, has about 28,000 speakers (Lewis 2009)¹. There has not been much work on Katu syntax to date. Apart from phonological descriptions and dictionaries, there is only diachronic or classificatory literature on Katuic or Austroasiatic languages, where Katu is also briefly described (Sidwell 2005; Sidwell & Jenny 2014).

1.2 Standard negation and typological features of negation

This paper is a first attempt at investigating negation in Low Katu. When investigating negation in an unfamiliar language, it is necessary to know as much as possible of what one can expect in the area of negation. Since negation can range in function from simply negating a main lexical verb to various subtypes of negation (for a summary of different types of negation, see van der Auwera & Krasnoukhova 2020; 91), it is vital to know what exactly the nature of any negating feature is that one finds in a language.

Where does one then start to dig in such a broad topic? A good starting point is the negation of the simplest utterances, as Miestamo (2005) does in his work on standard negation. This is described as the type of negation that can apply to the most minimal and basic sentences, which are characteristically main clauses and consist of a single predicate with as few noun phrases and adverbial modifiers as possible. From this starting point one can demarcate further types of negation one encounters:

“A S[tandard]N[egation] construction is a construction whose function is to modify a verbal declarative main clause expressing a proposition *p* in such a way that the modified clause expresses the proposition with the opposite truth value to *p*, i.e., $\sim p$, or the proposition used as the closest equivalent to $\sim p$ in case

¹ <https://www.ethnologue.com/language/kuf>

the clause expressing $\sim p$ cannot be formed in the language, and that is (one of) the productive and general means the language has for performing this function.” (Miestamo 2005; 42)

Ideally, one would thus elicit very basic, short declarative positive propositions as p and elicit the negation $\sim p$ thereof. Since I couldn’t find any such direct negation of the most minimal and basic positive sentences in the corpus, I searched for sentences containing one of the five negators. Using the simpler utterances that came from that search, I investigated how the negative examples behave with respect to each other, and how they differ from otherwise syntactically comparable positive utterances.

A brief investigation of the typology of negation shows what one might expect in a language that is yet to be described in more detail. Negators that act as verbs show different types of verbal behavior, there are higher negative verbs and auxiliary negative verbs (Payne 1985; 208). Different types of non-verbal predication such as the expression of equation, existential or attribution, can require different constructions. Parallel to that, the negation of non-verbal predications may be constructed by different means (Payne 1997; 114 ff.). Besides that, non-verbal negators can have different positions in the sentence (Dahl 1979; 91 ff.), imperatives (Miestamo & van der Auwera 2007; 59, van der Auwera & Krasnoukhova 2020; 106) and different TAM categories (Miestamo 2019; 3) may have their own special constructions for negation, among other things (for an overall summary, see also van der Auwera & Krasnoukhova 2020). The negators used in the corpus for this analysis were investigated for these different types of variation and are described in detail in the next section.

The *Katu* Folktales in Costello & Sulavan (1993) are annotated word for word and have free translations but lack detailed glosses describing the syntax and morphology. As a result, it is often not straightforward whether a negator is for instance a verb, or whether it is only used in certain syntactic contexts (as opposed to negating basic declarative main clauses). Cross-linguistically, there is a considerable range of properties that change depending on negation. It can easily be the case that there is no literal translation for such words in English, or that the exact use isn’t clear even in a free translation.

1.3 Low *Katu* negators

This paper investigates the use of the negators *kah*, *məʔ*, *jiəʔ*, *ʔeh* and *ʔɔ:ʔ* based on the data in Costello & Sulavan (1993). Here are sentences showing each of the negators, and in the following it will be described how their usages vary. To see a summary which syntactic properties take which negator, or which negator is used for which context, Table 1 in Section 3 can be consulted.

- (1) a. *kah* *dɔŋ* *ku* *ʔom* *də:k*.
NEG give 1SG drink water
‘He did not give me water to drink.’ (Costello & Sulavan 1993; 48)
- b. *ʔama*, *məʔ* *ʔame:ʔ* *ʔfoʔ*
father NEG mother return
‘Father, why doesn’t mother return?’ (Costello & Sulavan 1993; 154)
- c. *la* *maniç* *ʔaŋa:j*, *jiəʔ*.
if person far NEG
‘If a person is far away, (he is) not (afraid).’ (Costello & Sulavan 1993; 88)
- d. *dɔ* *ʔeh* *mak* *tʰiʔ* *dom-baw*.
3SG NEG owe variety people
‘He does not owe a variety of things to people.’ (Costello & Sulavan 1993; 408)
- e. *ʔɔ:ʔ* *duoŋ* *leʔ* *pə:ʔ* *jiə*.
PROH go.down towards shelter 1DU.EXCL
‘Don’t go down to our shelter.’ (Costello & Sulavan 1993; 142)

2 Usage of negation in Low Katu

This section will show the different uses found for the negators in the corpus and propose their most likely syntactic properties by referring to known behavior found in the typology of negation. The data this research is based on consists of oral linguistic material containing false starts, repetitions, digressions and changes of audience. It has word-for-word translations and free translations, but there are inconsistencies, and the free translations can be loose. With the help of previous work on negation, this paper gives first insights on how negators are likely to behave in some contexts. Nevertheless, the precise usage of the negators needs closer identification through elicitation. The analysis presented here is intended to serve as a basis for future syntactic research on this largely undescribed language.

2.1 *Kah*

The negator *kah* is the most frequent negator in the corpus. It mostly appears immediately before the matrix verb of a sentence, which is a likely indicator that it may be an auxiliary verb (Dahl 1979; 93). Many auxiliaries in Low Katu immediately precede the Matrix verb. Furthermore, *kah*, is not separated by sentential boundaries from the matrix verb, an indicator that *kah* may be a negative auxiliary. The other type of negative verb is a higher negative verb, as argued below for the negator *moq* (Payne 1985; 208). Overall, *kah* appears to typically be used to negate simple main clauses (sentence 4a). I therefore argue that standard negation is expressed with *kah*.

There is a large class of stative verbs in Low Katu, which are often used to modify nouns. Usually this is done by placing stative verbs directly before the modified nouns, similarly to how adjectives function in many other languages. This means there is no copula used for attributive predication, as opposed to for example English *is* in *the girl is tall*. As a result, the typical predicate negator is used for stative verbs too, which here is the standard negator *kah* (sentence 4b).

It is possible to use *kah* as a copula, but only for negative existentials or possessives (sentences 4c-e), and presumably locatives as well, but it is not clear from the free translations if locative meanings are meant instead of existential meanings. In the positive, these typically are expressed with the verb *ve* (sentence 4c), and the similar behavior to the negative counterparts shows that they both could be analyzed as verbs. It is common for languages to express possession with predicate locative or existential clauses (Payne 1997; 125), and this seems to be the case in Low Katu (this also explains the ‘have’ gloss for positive *ve* in Costello & Sulavan 1993):

Low Katu

- (4) a. ***kah*** *dɔŋ* *ku* *ʔom* *də:k..*
 NEG give 1SG drink water
 ‘He did not give me water to drink.’ (Costello & Sulavan 1993; 48)
- b. ***kah*** *maʃʰiŋ.*
 NEG sure
 ‘We are not sure (why).’ (Costello & Sulavan 1993; 420)
- c. *la* ***ve*** *ʔakɔ:n* *taha* *ʔakɔ:n* *kum* *ʃʰɔ:t.*
 if have child grown child also harvest
 ‘If there were older children, the children also harvested.’ (Costello & Sulavan 1993; 488)
- d. ***kah*** *ʃana* *ʔadi, ...*
 NEG food here
 ‘There is no food here, ...’ (Costello & Sulavan 1993; 64)
- e. ...; *ʔakɔ:n* *darit* ***kah*** *diʔ* *dɔ* *ʃik* *tə:t*
 child half.orphan NEG place 3SG then stay
 ‘...; the half orphan child did not have anywhere to stay.’ (Costello & Sulavan 1993; 47)

2.2 *məʔ*

The negator *məʔ* fits into the typical structure of a higher negative verb. Higher negative verbs tend to appear in verb-initial and verb-final languages only, so they appear in the corresponding natural verb position at the beginning or the end of the sentence (Payne 1985; 208). In the corpus, there are often verbs in the initial position of a sentence, making the occurrence of higher negative verbs in the language plausible. However, more research is needed to investigate the exact constraints of word-order in Low Katu. Higher negative verbs appear in main clauses and introduce a relative clause, which is the negated content, hence the name *higher* negative verb. The negated clause depends on *məʔ*. The examples in 5 show that *məʔ* negates and introduces whole embedded clauses, which often are ‘why not’ questions:

- (5) a. *ʔama, məʔ [ʔame:ʔ ʃo]ʔ*
 father NEG mother return
 ‘Father, why doesn’t mother return?’ (Costello & Sulavan 1993; 154)
- b. *məʔ, [tamɔ:j ni mət no lɛʔ]ʔ*
 NEG visitor PLU enter this place
 ‘Visitors, why don’t you come into this place?’ (Costello & Sulavan 1993; 298)
- c. *məʔ [ʔave daj dɔŋ].*
 NEG have 1SG give
 ‘I did not allow them.’ (Costello & Sulavan 1993; 129)

2.3 *jiəʔ*

jiəʔ also can be said to be an auxiliary verb, but has a slightly different usage from *kah*. It appears to mostly introduce resultative ‘then not’ clauses or act as a negative complementizer ‘that not’, as the examples in 6 show. *jiəʔ* isn’t necessarily positioned immediately before the lexical verb, but there are some examples that hint to its use as an auxiliary verb. It is possible to replace a lexical verb that is known from the context with *jiəʔ*, similarly to English *isn’t* or *doesn’t*. In 6a, the previous sentence was talking about someone being afraid to stab people when close to them. *Yuroq* can be understood here as *Then he isn’t*.

It often happens that subordination requires a different negator in languages (Payne 1985: 240). Distinguishing main and subordinate clauses is not always so straightforward in Low Katu, as there are no morphological cues to changes in finiteness (Payne 1997; 306, Bisang 2007; 116) and there is very little information on other tendencies of subordination in Katu, like those cross-linguistic tendencies suggested by Cristofaro (2005; 24ff). Examples 6a and b show that *jiəʔ* appears both in complement clauses and in main clauses. Subordination does not seem to be the defining factor to use this negator, though it needs to be shown in future work what the exact nature of subordination is in Low Katu.

- (6) a. *la maniç ʃaŋa:j, jiəʔ.*
 if person far NEG
 ‘If a person is far away, (he is) not (afraid).’ (Costello & Sulavan 1993; 88)
- b. *nok lɔj lalua, jiəʔ ni bə:r ʃa*
 here see truly NEG PLU two eat
 ‘Here you see that truly we two have not eaten them.’ (Costello & Sulavan 1993; 110)
- c. *ʔəʔ hɛ jiəʔ raw kakʰin ʔəʔ*
 still 1PL.INCL NEG thing afraid still
 ‘Then we are not afraid anymore.’ (Costello & Sulavan 1993; 752)

2.4 *ʔeh*

ʔeh appears either before the lexical verb (sentence 7a) or forms negative equations, and therefore might also be an auxiliary verb (Payne 1985; 208). Positive equation is expressed with *NP NP* juxtaposition (sentence 7b). *ʔeh* is the only negator found to express negative equation (sentence 7c). *ʔeh* also appears both in complement and main clauses (sentences 7d and e). Both *ʔeh* and *jiəʔ* appear to be in an initial position of the subordinate clause, which is also the preferred position for verbs in subordinate clauses in Low Katu.

- (7) a. *dɔ ʔeh mak tʰiʔ dom-baw.*
3SG NEG owe variety people
‘He does not owe a variety of things to people.’ (Costello & Sulavan 1993; 408)
- b. *ŋa:j ʔama:ʔ ʔafɛkʔ*
Who grandmother that
‘Who is that grandmother?’ (Costello & Sulavan 1993:311)
- c. *ku ʔeh kaji:k diʔ maj.*
1SG NEG wife to 2SG
‘I am not your wife’ (Costello & Sulavan 1993; 396)
- d. *dala:ŋ dɔ nal ʔadik, ʔeh padɪ:l ʔadok.*
call 3SG know that NEG girl now
‘She called, she knew that they were not these girls (she was looking for).’ (Costello & Sulavan 1993; 96)
- e. *ŋə ʔeh ta:j dɔ, ...*
need NEG because 3SG
‘I remember it was not because of him, ...’ Costello & Sulavan (1993; 145)

2.5 *ʔɔ:ʔ*

ʔɔ:ʔ seems to consistently be used throughout the corpus for prohibitive imperatives (Aikhenvald 2010; 167). *ʔɔ:ʔ* always appears before lexical verbs (sentences 13b and c), like *kah*, as shown above. Since there is no trace of a sentential boundary between *ʔɔ:ʔ* and the lexical verb following it, it can also be assumed to be an auxiliary verb, rather than a higher negative verb (Payne 1985; 213). It is best translated with English *don’t*.

For positive imperatives, often the particle *ʔaʔ* is used (sentence 13a). However, *ʔaʔ* is only glossed as ‘emphasis’ or ‘yes’ and is used in situations that aren’t translated as imperatives as well, and the exact meaning is not clear from the free translations. Even if this is the typical construction for positive imperatives, it can well be that there are other constructions for positive imperatives that aren’t as easily identified in this corpus and can still be elicited.

- (8) a. *ŋi bə:r mɔ:n, ʔju:r ʔaʔ, ʔata:w; ...*
PLU two speak rise emphasis Ataa
‘The two (sisters) said, “Get up Ataa, ...” (Costello & Sulavan 1993; 146)
- b. *ʔɔ:ʔ kakʰin braw.*
PROH be.afraid killer
‘Do not be afraid of killers.’ (Costello & Sulavan 1993; 319)
- c. *ʔɔ:ʔ duon leʔ pə:ʔ jia.*
PROH go.down towards shelter 1DU.EXCL
‘Don’t go down to our shelter.’ (Costello & Sulavan 1993; 142)

2.6 Negating *ʔɔk*

Not a negator, but an important word to consider with negation, is *ʔɔk*, sometimes glossed by Costello & Sulavan (1993) with ‘yet’ and sometimes with ‘more’. In the free translations, the resulting meaning with negation is sometimes ‘not anymore’ and sometimes ‘not yet’ (see examples below). The purpose of example 9d is to show that the free translation shows a positive meaning of ‘still’.

The reason for these differences in translation might be due to differences in the scope of negation. Van der Auwera (2021) summarizes the behavior of phasal polarity, i.e., semantic domains served by the English adverbs *already* and *still* and the adverbial phrases *not yet* and *no longer*. Depending on whether *already* or *still* are negated internally or externally or both, the meanings change. In English we just happen to have other primitives for these expressions. Instead of *already not* or *not still*, we rather say *not anymore* etc. (van der Auwera 2021; 31), and this may have caused some of the confusing glosses in the Katu folk tales.

I thus argue that ‘still’ is the best gloss for *ʔɔk* and have accordingly changed the glosses for this paper, since Costello & Sulavan’s glosses are inconsistent. The translated meaning depends on the scope of negation: In example 9b., only *ɲa:j* seems to be negated, resulting in the meaning ‘no-one’, whereas in 9a. the scope of negation goes over the whole sentence, therefore including a negation of *ʔɔk*. Assuming the positive meaning of *ʔɔk* is ‘still’, the literal translation would be ‘not still’ as explained above, which is properly translated in English as ‘not anymore’. It appears that it is not the negator or the meaning of *ʔɔk*, but the scope of negation that determines the resulting change of tense and aspect. This behavior is very similar to that of the Lao postverbal aspectual-modal marker *ɲáŋ*, which in positive sentences usually means ‘still’, but when negated it means ‘not yet’ (Enfield 2007). It can well be that Low Katu *ʔɔk* is a calque of Lao *ɲáŋ*².

From these observations I decided a better fitting gloss for *ʔɔk* is ‘still’:

- (9) a. *kah* *fɔ:m* *jiəh* *ʔəʔ*.
NEG be.able cure still
‘They are not able to cure him anymore.’ (Costello & Sulavan 1993; 366)
- b. *kah* *ɲa:j* *fɔ* *ʔəʔ*.
NEG anyone return still
‘No-one returned yet.’ (Costello & Sulavan 1993; 309)
- c. *tʰoj* *ʔɔ:ʔ* *vo:c* *ʔəʔ*, *ni* *bə:r*.
Enough PROH go still PLU two
‘That’s enough, don’t go any more, you two.’ (Costello & Sulavan 1993; 170)
- d. ..., *maj* *kaŋi:ʔ* *ʔəʔ*.
2SG grunt still
‘..., you will still grunt’ (Costello & Sulavan 1993; 51)

2.7 Negative indefiniteness

Negating indefinites is another typologically interesting topic. Since the word-to-word translations aren’t detailed glosses in the Katu Folktales, it is hard to tell the part of speech and the exact use of words that might be indefinite pronouns. The only seemingly indefinite nominal that is directly negated multiple times in the corpus is *ɲa:j*, glossed as ‘anyone, who’ in Costello & Sulavan (1993). As can be seen in the examples in 10, using *ʔɔp* ‘every’ before *ɲa:j* results in the meaning ‘everyone’ (example 10a), but this meaning can also be conveyed if *ɲa:j* is followed by *du* ‘who, which’ (example 10b). When preceded by *kah*, presumably as a negative existential in cases such as here, the meaning ‘no-one’ is translated (example 10c). Interestingly, *ngaai* alone can also be translated as ‘no-one’ (example 10d). A better understanding of the different contexts *ngaai* can occur in is necessary to understand how

² I would like to thank Paul Sidwell for bringing this to my attention.

its meaning changes with negation, since there are many possibilities of how different combinations can change meaning (van der Auwera & Krasnoukhova 2020; 110 ff.).

- (10) a. *ʔəp ɲa:j mblɔ:j tɛ:ʔ kanɔ:n ɲan vo:c;*
 every anyone,who man.unmarried from village ask go
 ‘All the unmarried men from the village wanted to go to ask for them;’ (Costello & Sulavan 1993; 16)
- b. *haʔjəʔ dɔ lɔj padɪ:l du liem padɪ:l ɲa:j du kieŋ*
 then 3SG see girl which good girl anyone,who which want
 ‘then he saw a girl who was good looking, everyone wanted her.’ (Costello & Sulavan 1993; 236)
- c. *bət biəl kah ɲa:j tət diʔ dən.*
 all village NEG anyone,who stay,sit place,in,on house
 ‘The whole village, no-one should stay in the house.’ (Costello & Sulavan 1993; 182)
- d. *ɲa:j nal raw tuoj dɔ*
 anyone,who know thing age 3SG
 ‘No-one knows its age.’ (Costello & Sulavan 1993; 200)

3 Summary and Discussion

By understanding how negation works cross-linguistically and applying this knowledge to an unknown language like Low Katu, one can figure out how this language behaves in negation, and through this some other aspects of the language become clearer. For instance, no postverbal negators were found, and all negators have been observed to appear sentence or clause initially, except *ʔeh*. Thanks to these insights it seems probable that, despite Low Katu word order being usually verb medial or verb initial, an earlier stage have been more strictly verb initial, due to the tendency for negators to be sentence initial and for verb initial languages to have higher negative verbs and auxiliaries before the lexical verbs (Dryer 1988; 98). It should also be kept in mind though, that preverbal negators could also appear sentence initially because the subject, or agent, has been dropped.

In the following, a short summary and discussion of the negators is presented, along with some notes and ruminations about negators in related languages and potential historical links. Table 1 gives an overview of different types of clauses or predications and which of the Low Katu negators are used to negate them. The value ‘yes’ means this negator can be used to negate that type of clause or predication, and ‘-’ means no examples were found of this negator used in such utterances.

The most frequent negator in Low Katu is *kah*. It can be applied to declarative verbal main clauses, placed right before the matrix verb, to modify them and express the opposite truth value of the equivalent declarative. It thereby fulfills the definition Miestamo (2005) gives for standard negation. It can also be used for non-verbal predication, apparently for existentials only. These observations suggest that this negator is an auxiliary verb and is used for standard negation as Miestamo (2005) defines it. In Shorto (1972; 27, 54), a dictionary for Old Mon inscriptions from the sixth to the sixteenth centuries, the entries for *ka* or *kah* show the translations ‘not to be’ and ‘not to’. This means potential early cognates of this word behave as a verb or as a negative copula, and this fits very well with the result for Low Katu *kah*. Since borrowing of negators is widespread in the region, the possibility that *kah* is a Mon loan perhaps from the Dvaravati period should also be acknowledged³.

³ I would like to thank Paul Sidwell for pointing this out to me.

Table 1: Low Katu negators

property		kah	məʔ	jiəʔ	ʔɛh	ʔɔ:ʔ
Non-verbal predication	equation	-	-	-	yes	-
	existential	yes	-	-	-	-
	possessive	yes	-	-	-	-
Position	sentence initial	yes	yes	yes	-	yes
	pre verb	yes	yes	yes	yes	yes
	negative verb (higher)	-	yes	-	-	-
	negative verb (auxiliary)	yes	-	yes	yes	yes
Clause type	main clause	yes	yes	yes	yes	yes
	complement clauses	-	-	yes	yes	-
	questions	-	yes	-	-	-
	imperatives	-	-	-	-	yes

The word *məʔ* appears to be a higher negative verb in, which means it should always appear in main clauses. It is often used for negative ‘why not’ questions, but there are other uses as well. No non-verbal predication was found with *məʔ*. This negator may also have been borrowed between multiple languages. For instance, *maʔ* is a common negator in Bahnaric (also Austroasiatic) languages:

- (11) a. Alak *maʔ* ‘don’t’ (Huffman 1971)
 b. Halang *maʔ* ‘do not’ (Cooper & Cooper 1976)
 c. Bahnar *maʔ leej* ‘but’ (Banker, Banker & Mo’ 1979)
 d. Mnong Rölöm *məʔ* ‘not’ (Blood 1976)

Additionally, according to Shorto (1972; 296), there is an Old Mon counterpart of this negator that appears as *mu* or *mo*. It was used to form questions and is translated with ‘what’, or, when appearing together with *het* ‘reason’, it means ‘why, what cause’. The use of the negator in questions is very similar to the use of *məʔ* in Low Katu. As with *kah*, finding counterparts from several centuries ago can indicate cognacy from the common ancestor. However, especially if the negator is this similar in form and usage across geographically proximate languages such as the Bahnaric examples in 11, a more recent borrowing scenario should also be considered.

The word *jiəʔ* appears to be an auxiliary verb, but it does not necessarily precede a verb and it does not seem to be used for non-verbal predication. According to the free translations, *jiəʔ* appears both in main clauses and in subordinate clauses, but more work needs to be done to know the exact nature of subordination in Low Katu, as the word-to-word translations sometimes have very different word-orders from the free translations.

The word *ʔɛh* is the only negator that typically is not sentence initial. It can be used for negative equation, though further research might show it can be used for other non-verbal predications. Since it can appear before the lexical verb of a sentence, it was also marked as possibly an auxiliary verb in Table 1, but further research might be necessary to show it behaves more like a verb than a particle. It can appear in main clauses and in initial position in complement clauses.

As negator *ʔɔ:ʔ* only appears in negative imperatives and not in declaratives, it should be considered a prohibitive marker. It always precedes the lexical verb, so it is also likely an auxiliary verb. No examples for non-verbal predication were found.

Some points were not possible to be determined with certainty in this analysis. For one, no negative proper inclusions or locatives were found. I suspect that Katu does not distinguish between equation

and proper inclusion, and locatives probably negate like existentials, though this, and non-verbal predication in general, needs to be elicited from a native speaker for all details to be understood. This counts also for negative imperatives, which could be equations such as *don't be a fool*, but weren't found here.

The negator *məʔ* and *jiəʔ* have a slightly more complicated use in combination with subordination. A more detailed understanding of how subordination works in Low Katu is necessary, specifically of adverbial and relative clauses. This should already be possible with more exact word-to-word translations next to the free translations.

Regarding TAM categories, this paper investigates how *ʔɔk* 'still' behaves with negation. Beyond that, it is not uncommon that the verbal paradigms appear differently in the negative than in the positive due to paradigmatic asymmetry. Grammatical distinctions can change or get lost in the negative that are present in the affirmative, and this is an important point to keep in mind when talking about tense and aspect in negation (Miestamo & van der Auwera, 2011; 68). It can also be that certain forms only exist in the positive and need different constructions for the negative. It still needs to be investigated whether paradigmatic asymmetry is an issue, which isn't very easy due to the isolating morphology of Low Katu not making loss of distinctions as obvious as inflectional morphology does. Furthermore, identifying asymmetry through corpus analysis is greatly limited, as finding two sentences that differ only by polarity is unlikely without simply eliciting them from native speakers.

Despite the problems arising from the limitations of the corpus, the typological information gathered about negation as a phenomenon allowed some strong initial steps to be made towards understanding negation in Low Katu. Thanks to these insights, this paper can be used as groundwork for finding out more about Low Katu, also beyond its negation.

References

- Aikhenvald, Alexandra Y. 2010. *Imperatives and commands* Oxford studies in typology and linguistic theory. Oxford: Oxford University Press.
- Auwera, J. van der, O. Krasnoukhova, V. Déprez, and M. T. Espinal. 2020. *The Typology of Negation*. Oxford University Press. <https://hdl.handle.net/1887/3160737>
- Auwera, Johan van der. 2021. "Phasal Polarity – Warnings from Earlier Research." In Raija Kramer (ed.) *The Expression of Phasal Polarity in African Languages*, De Gruyter Mouton, 25–38. <https://www.degruyter.com/document/doi/10.1515/9783110646290-003/html>
- Banker, John, Elizabeth Banker & Siu Mo'. 1979. *Bahnar Dictionary, Plei Bong-Mang Yang Dialect*. Huntington Beach, Summer Institute of Linguistics.
- Bisang, Walter. 2007. Categories that make finiteness: discreteness from a functional perspective and some of its repercussions. In Irina Alekseevna Nikolaeva (ed.), *Finiteness: Theoretical and Empirical Foundations*, 114 – 137. Oxford: Oxford University Press.
- Blood, Evangeline. 1976. *Eastern Mnong Vocabulary: Eastern Mnong-Vietnamese-English Thesaurus (a computer print-out)*. Huntington Beach, California, Summer Institute of Linguistics.
- Blood, Evangeline. 2005. *Mnong-Vietnamese-English Dictionary*. [MS] <http://sealang.net/monkhmer/database/>
- Cooper, James & Nancy Cooper. 1976. *Halang Rhyming Dictionary: Halang-English*. Dallas, Summer Institute of Linguistics.
- Costello, Nancy A. & Khamluan Sulavan. 1993. *Katu folktales and society (Nôôq taruuih Katu)*. Vietnam: Institute of research on Lao culture and society.
- Cristofaro, Sonia. 2005. *Subordination* Oxford studies in typology and linguistic theory. Oxford: Oxford University Press reprinted with corrections edn.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17.1-2:79–106. <https://www.degruyter.com/view/j/ling.1979.17.issue-1-2/ling.1979.17.1-2.79/ling.1979.17.1-2.79.xml>

- Dryer, Matthew S. 1988. Universals of negative position. In Michael Hammond, Edith A. Moravcsik & Jessica R. Wirth (eds.), *Studies in syntactic typology*, vol. 17 Typological studies in language, 31. Amsterdam: Benjamins.
- Enfield, Nick J. 2007. 38 A Grammar of Lao. New York: Mouton de Gruyter.
<http://www.degruyter.com/view/books/9783110207538/9783110207538/9783110207538.xml>
- Huffman, Franklin E. 1971. *Vocabulary Lists (Mon-Khmer) 20 languages*. Archived online at:
<http://sealang.net/archives/huffman>
- Lewis, M. Paul. 2009. *Ethnologue: Languages of the World*. 16th ed. Dallas, Tex: SIL International.
<https://www.ethnologue.com/language/kuf>
- Miestamo, Matti. 2005. *Standard negation: the negation of declarative verbal main clauses in a typological perspective*, vol. 31 Empirical approaches to language typology. Berlin: De Gruyter.
- Miestamo, Matti & Johan van der Auwera. 2007. Negative declaratives and negative imperatives: Similarities and differences. In Andreas Ammann (ed.), *Linguistics Festival: May 2006*, Bremen, vol. 14 *Diversitas linguarum*, 59–77. Bochum: Universitätsverlag Brockmeyer.
- Miestamo, Matti. 2019. Questionnaire for describing the negation system of a language. [MS]
- Payne, John R. 1985. Negation. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. 1, Clause structure*, 197–242. Cambridge: Cambridge University Press 1st edn.
- Payne, Thomas E. 1997. *Describing morphosyntax: a guide for field linguists*. Cambridge: Cambridge University Press.
- Shorto, H. L. 1972. *A dictionary of the Mon inscriptions from the sixth to the sixteenth centuries, incorporating materials collected by the late C. O. Blagden*, vol. 24 London Oriental Series. London: Oxford University Press.
- Sidwell, Paul, and Mathias Jenny. 2014. *Handbook of Austroasiatic Languages (2 Vols)*. Brill.
- Sidwell, Paul. 2005. *The Katuic Languages: Classification, Reconstruction and Comparative Lexicon*. Vol. 58. München: Lincom Europa.

SUBORDINATION STRATEGIES IN RUCING PALAUNG

LEE Wei-Wei

Independent Researcher

<me DOT weiwei AT gmail DOT com>

Abstract

This paper presents an overview of morphosyntactic subordination strategies and their functions in Rucing Palaung. Rucing is a verb-medial language of the Palaungic branch of the Austroasiatic languages. Based on newly collected data, the study finds that subordinate clauses generally display verb-initial constituent order, including relative clauses and certain types of complement and adverbial clauses. These clauses may contain the same verbal features as independent clauses. In some cases, the differential constituent order is the only indicator of subordination. Other, verb-medial strategies all involve the verb *taih* ‘give’. Semantic extensions of *taih* are likely to have motivated the use of similar structures for a non-arbitrary range of functions. Further research is necessary to determine whether the subordinate verb-initial structures differ from independent clauses exclusively in the arrangement of verb and subject or whether additional formal processes are involved.

Keywords: subordination, constituent order, Palaungic, syntax, semantics

ISO 639-3 codes: pce, rbb, pll, shn

1 Introduction

Rucing¹ (*pce*) is a variety of Palaung, a language belonging to the Palaungic branch of Austroasiatic. Whereas Palaung is a name given by the Burmese, the ethnic Palaung people including the Rucing call themselves Ta’ang (/da.ʔa:ŋ/ or /da.ʔa:k/ in the variety described here). The Rucing variety investigated in this paper is spoken in Pein Ne Pin village, Kalaw township, in southern Shan State of Myanmar. Apart from the Kalaw area, Rucing is also spoken around Kengtung in southeastern Shan State, in Chiang Mai province of Thailand, and in western Yunnan province of China (see e.g., Deepadung 2011). According to Ethnologue (Eberhard et al. 2019), out of the total number of 273,000 Rucing speakers about 258,000, i.e., the large majority, were living in Myanmar in 2000. However, since no official census has been published in Myanmar since 1931 (Deepadung 2011) and it is unclear how Ethnologue’s data are obtained, these numbers are of unknown reliability. In Pein Ne Pin, the population mostly consists of native Rucing speakers, the younger generations of whom also speak Burmese. Although Shan (*shn*) is not spoken generally, the past Shan influence on the language is still visible, most clearly in the lexicon, in addition to the ongoing Burmese influence.

Among the 17 Palaung varieties recognized by native speakers themselves,² the three main varieties commonly mentioned in the linguistic literature are Samlong³ (*pll*), Rumai (*rbb*), and Rucing. Within these three varieties there is considerable dialectal variation in both grammar and phonology; indeed, they are mutually unintelligible. Only Rucing is spoken in southern Shan State; the Samlong and Rumai live in the northern parts. Phonological comparison suggests that the original Palaung homeland was in the border area of northern Shan State with Yunnan, west of the Salween river (Ostapirat 2009; Sidwell 2015:16–7, 21). The Rucing settlements are likely to be the result of later southward migrations.

¹ Also called Pale (e.g., Janzen & Janzen 1972), Silver or Ngwe (‘silver’ in Burmese) Palaung, or Da-ang (Ostapirat 2009).

² Ta’ang Official Language Commission, personal communication. They further divide these 17 groups into seven main groups, including Rucing, Rumai, and Samlong.

³ Also widely known as Golden or Shwe (‘gold’ in Burmese) Palaung or simply Ta-ang.

1.1 Data sources

Whereas Samlong Palaung is relatively well-described (Milne 1921; Mak 2012), there is no comprehensive grammatical description of Rumai thus far.⁴ As for Rucing, phonological and lexical studies likewise predominate (Janzen 1978; Shintani 2008; Buakaw 2009, 2012; Rattanapitak 2009). However, a grammar sketch does exist of a Rucing variety spoken in Chiang Mai province (Deepadung et al. 2015). As for the present Kalaw variety, several grammatical aspects have been described by Janzen & Janzen (1972) and Janzen (1976a; b, 1987). These include basic clause structures, aspectual markers, and topicalization in discourse. In addition, Janzen (1987) contains three glossed and translated stories.

Janzen and Janzen's work has been a valuable source for the present study by providing initial data of and insights into Rucing. Primarily, this study is based on data collected by the author in 2018 and 2019. These data were collected from various native speakers born in Pein Ne Pin village. The corpus consists of elicited sentences, spontaneous and visually guided narratives, and semi-guided conversation/interaction. Furthermore, there is a Rucing translation of the New Testament from 2013.⁵ The language variety used is both phonologically and grammatically very similar to the Pein Ne Pin variety. Unfortunately, little is known in terms of metadata. The main translator is presumably from Kengtung, Myanmar (Elizabeth Hall, p.c.). The NT translation provides valuable insight into a closely related variety as well as an additional genre and style that is absent from my own corpus.

1.2 Aim of the study

The aim of this paper is to describe subordination strategies in Rucing. As an Agent-Verb-Patient (AVP) language, Rucing displays AVP/SV constituent order throughout independent clauses. Many dependent clause types, on the other hand, are characterized by VAP/VS order, including relative clauses and some but not all complement and adverbial clauses. The remainder of this paper is organized as follows: section 2 describes basic clause structures in Rucing. Section 3 discusses the theoretical preliminaries; it provides the definition of subordination used in this study and takes up the impact that has been attributed to word order in signalling subordination cross-linguistically. Section 4 describes the form and functions of subordination strategies attested in Rucing, after which section 5 concludes the paper with a discussion of the findings and implications.

2 Simple sentences

2.1 Declarative sentences

Like most languages spoken in Mainland Southeast Asia, the basic constituent order in main declarative clauses in Rucing is AVP or SV, i.e., subject-verb order. The alignment is nominative-accusative.⁶ Core arguments (S, A, and P) and possessors are solely determined by their position in the clause; they are not otherwise overtly marked. Contextually understood arguments are very frequently omitted. There are no obligatory verbal categories; aspectual-modal and negation markers occur before the verb when present. Intransitive clauses are illustrated in (1-3); transitive clauses in (4-6).

⁴ Studies on specific aspects of Rumai grammar do exist, e.g., negation (Weymuth 2014) and verbal affixes (Weymuth 2018).

⁵ Also published in book form by Glendora, Christian Far East Ministry as *Bhp' jv' pheā 'ai kā māy'*. For the present study, the text was taken from <http://www.palaung-ruch.com/%E1%80%98%E1%80%95%E1%80%BA-%E1%80%87%E1%80%9D%E1%80%BA%E1%80%96%E1%80%BC>. The website also includes audio recordings. As a parallel English translation, the New International Version was used. In many cases, no corresponding or only partly overlapping content was available in the NIV. Therefore, for the purpose of clarity, and this purpose only, most English translations given in this paper are either strongly modified or newly added by the author.

⁶ For the sake of brevity, when indicating constituent order patterns in the remainder of this paper, A will be used for both A and S. AVP thus includes SV.

Intransitive

- (1) *madəbɔʊ? leh⁷*
 owl exit
 ‘An owl comes out.’ / ‘An owl came out.’
- (2) *ʔum ʔɛ leh*
 water NSIT exit
 ‘The water is coming out.’ / ‘The water has come out.’
- (3) *ka:k tu ta:k*
 house NEG be.big
 ‘The house is not big.’

Transitive

- (4) *maua ʔan cauh k^hruuk kamae*
 mother 3SG buy clothing be.new
 ‘His mother bought new clothes.’
- (5) *ʔan hoe? tih*
 3SG COMPL read
 ‘He has read [this book].’
- (6) *tu jau marɔʊ? gae tɔŋ*
 NEG see frog 3DU NVIS
 ‘[They] don't see their frog.’

Despite the strong pro-drop character of Rucing, the language also has an expletive/dummy subject clitic pronoun *ʔŋ=*.⁸ It is used in impersonal constructions expressing things like ‘one should’, ‘it is possible’, meteorological phenomena, existential constructions, among others. Examples are given in (7-8). In this paper, we will come across the function of *ʔŋ=* as a placeholder for subject complement clauses in section 4.2.

- (7) *ʔŋ=lɔu tauua kamai? daʔa:k*
 3SG.NHUM=need take,wear Palaung.scarf Ta'ang
 ‘One should wear Ta'ang scarves.’
- (8) *ʔŋ=mauh ge gun gɔŋ ʔm-p^huan*
 3SG.NHUM=EXIS 3PL father child NML-four
 ‘There was a family of four.’

⁷ The phonology of Rucing is not yet fully understood and is not part of this study's purpose. Therefore, the broad phonemic transcriptions provided here should be taken as tentative. Loan words are indicated by curly brackets. As for the New Testament translation, I have attempted to leave the phonology as intact as possible based on the Burmese orthography used in the original text. NT examples are marked as ‘NT’ plus book, chapter, and verse number.

⁸ The use of a dummy subject is cross-linguistically associated with the obligatoriness of the subject (Malchukov & Ogawa 2011:25; Schmidtke-Bode 2014:197). Interestingly, it also occurs in the related, likewise pro-drop, language Khmer (Haiman 2011:192ff., 209).

2.1. Pragmatic variation

In certain pragmatically marked contexts, the default AVP order may be replaced by an alternative order. In existential and presentational constructions, the subject follows the verb, as seen in (9-11). Presentative constructions also include those based on verbs like ‘remain’.

- (9) *mauh ʔum singuah*
 EXIS water wave
 ‘There are waves.’

- (10) *ʔaɔ mauh ʔɛ jam*
 NEG EXIS NML.SUBJ die
 ‘No one died.’ (Lit. ‘There is not someone who died.’)

- (11) *dɔh ʔɛ gi pɔ ca:ŋ doe dɔh*
 DIST.ADV NSIT have.left other bag DIST DIST.ADV
 ‘Look, there's another bag left there.’

Information structure affects clause structure in various other ways. Both topicalized and focused elements are generally fronted. If the topicalized element is the subject, an anaphoric, resumptive personal pronoun is used following the subject NP, as in (12). This is a common strategy in Southeast Asian languages (Jenny et al. 2015). As can be seen in (13), a topicalized object is simply fronted.

- (12) *nih, daua nei, ʔan buʔ ŋaɔ*
 PROX.ADV male PROX 3SG light fire
 ‘Here, this guy, he was making a fire.’

- (13) *ʔa-nei daɔ majanfaua nei je daʔa:ŋ mɔuk*
 3SG-PROX plant type.of.plant PROX 1PL.EXCL Ta'ang like
 ‘This, this flower here (*ma jan faua*), we Ta'ang like.’

Objects can also be fronted for focus. Examples (14-15) show instances of contrastive focus. The fronted element may or may not be followed by a topic-comment linker. In (14), the chicken’s fate and the speaker’s own fate are contrasted, whereas in (15), the preparation of wet tea leaves is compared with that of dry tea leaves.

- (14) *ʔi:ŋ muŋ ʔɔu ʔɛ bleh, ʔɔu muŋ ge di ŋaɔh*
 chicken TCL 1SG NSIT release 1SG TCL 3PL FUT kill
 ‘The chicken, I've released it. As for me, they're going to kill me.’

- (15) *mɛ:ŋ ʔaʔ mɛ naʔ pŋ bok ʔaŋau mɛ mɛ si-lɔu*
 tea.leaf be.soft 2SG still get do,make dried.tea.leaf 2SG 2SG want-need
rae di suda:ʔ
 dry OBL sunshine
 ‘You can still make wet tea leaves. Dry tea leaves still need to be dried in the sun.’

Another kind of topic appears as an afterthought or anti-topic (Chafe 1976; Lambrecht 1994:203), as an extra-clausal element at the end of the sentence. While the anti-topical element is typically a subject NP, the referent may have already been referred to inside the clause. The main pragmatic characteristic is that the referent is highly accessible to the addressee. This is another widespread phenomenon in the area (Jenny 2020). Accompanied by the pro-drop nature of these languages, the emerging construction is very often verb-initial, as the Rucing examples below show. A subject pronoun may, however, always

be added in front of the verb. A case where the subject pronoun has not been omitted is displayed in (16). In (17), *ka:k* ‘house’ is clearly an easily accessible referent, being mentioned in the directly preceding sentence. Similarly, example (18) is part of a conversation in which, directly preceding the given sentence, *kabɔu* ‘water pipe’ has already been mentioned by other speakers.

(16) *dɔ.tei ge vi tauua k^hruk mpra:ŋ, gɔŋ ge*
 now 3PL SEQ take,wear clothing Burmese.people child 3PL
 ‘Now they are wearing Burmese clothes, the children.’

(17) *ka:k tu gjaʔ pei dɔ.tei.nei ʔaiʔ nei. ʔece:k, ka:k.*
 house NEG be.good be.like now front PROX be.few house
 ‘The houses weren't as good as they are now, before. They were few, the houses.’

(18) *lup k^he:k groʔ, kabɔu*
 enter inside cut.bamboo water.pipe
 ‘It goes into the cut bamboo, the water pipe.’

Having seen how basic clause structure works in simple declarative sentences in Rucing, let us now turn to complex sentences involving subordinate constructions. Before proceeding with the findings, the next section first provides the definition of subordination used in the present study and discusses the relationship between subordination and clausal constituent order.

3 The notion of subordination

Subordination refers to a kind of clause-combining where one (subordinate) clause is in some way dependent on (a) the other (main) clause or (b) a constituent in the other clause. This dependency may be interpreted in terms of form (morphology, syntax, phonetics), function (semantics, pragmatics), or a combination of the two. Traditionally, a three-way classification of subordinate clauses is made, consisting of the relative, complement, and adverbial type. A grammatically dependent clause can be broadly defined as showing at least one of the following properties: (a) cannot stand by itself, (b) cannot be interpreted independently, and (c) does not form an intonation phrase of its own (Gast & Diessel 2012:4–5).

Functional approaches have arrived at the same three major subordinate types (of functions instead of clauses) based on semantic and pragmatic considerations (Langacker 1991, 2008; Croft 2001:ch. 9; Cristofaro 2003). In Cristofaro’s view, for instance, a ‘subordinate event’ is considered an event that is pragmatically non-asserted, with a backgrounded and dependent cognitive profile with respect to the ‘main event’.

While this three-way classification still forms a useful underpinning of much of contemporary functional-typological research (e.g., Gast & Diessel 2012; Herlin et al. 2014), the clause types are generally taken as prototypes or scalar concepts based on multiple parameters, rather than rigorous categories (Lehmann 1988; Van Valin & LaPolla 1997:ch. 8; Croft 2001:ch. 9; Givón 2001a:327–8; Bickel 2010). It has long been apparent that in many languages, the semantic functions associated with these clause types do not form a distinct or internally homogeneous grammatical class. In other words, a morphosyntactically dependent unit cannot necessarily be mapped onto a functionally dependent unit in a given language.

The present study adopts a data-driven, bottom-up approach and thus takes a formal view of subordination as the point of departure. For the purpose of this paper, a subordinate clause in Rucing is understood as a clause that (a) cannot form an independent utterance on its own and (b) structurally differs from an independent declarative clause. It will be shown that one formal feature, namely verb-initial constituent order, is shared by all three major functions traditionally identified: relative, complement, and adverbial relations. The next subsection reviews some literature on the role of constituent order in marking subordination.

3.1 The role of constituent order

Word order, here understood as the constituent order within a clause, is not widely discussed as a distinctive feature with respect to subordination. This does not mean that word order variation depending on the clause type is uncommon. In most cases of variation, however, either the order in subordinate clauses is restricted to one of the orders also found in main clauses or the special order is limited to a certain type of subordinate clause (Siewierska 1988:91–2; Hetterle 2015:127ff.; Salaberri Izko 2017:11–22). The former is often the result of word order change starting in the main clause, whereas the latter is usually derived from the particular construction or morpheme concerned. In all such cases, constituent order does not play a differentiating role between subordinate and main clauses. It is noteworthy that word order is nevertheless mentioned in the literature as such, e.g., as a “device [...] typically found among languages of the world for marking subordination” (Thompson et al. 2007:238) or a “widely acknowledged syntactic criterion for subordination” (Herlin et al. 2014:11). The examples provided along with such claims, however, are fragmentary and chiefly focus on Germanic languages, the most notorious case. What is widely believed to have happened in German, for instance, is that whereas the main clause changed its order from verb-final to V2, subordinate clauses remained verb-final (Vennemann 1975; Hock 1986:330–336). Indeed, in Cristofaro’s seminal typological study of subordination, she notes that “variation in word order [...] has been empirically proven to have little relevance for clause linkage strategies used within the domain of subordination” (2003:51). Along the same lines, in adverbial clauses Hetterle (2015:127–131) fails to find any cross-linguistic tendencies regarding differential word order and concludes that “word order permutations in adverbial clauses in general can be considered a “rarissimum”.”

The role of word order in signalling subordination in general thus seems to be contested among typologists. While the syntactic status of a clause is often relevant to word order variation within a language, there is little cross-linguistic support so far for claims that word order is a reliable indicator of subordination in general. What is more widely attested are situations in which (a) word order in main clauses is more flexible than in subordinate clauses, but the subordinate structure can also be found in main clauses, or (b) the deviant order is restricted to particular subordinate constructions only.

4 Subordination strategies

Rucing exhibits four major formal strategies to mark a subordinate clause,⁹ as listed and illustrated below. x^{SUB} stands for a subordinating morpheme.

x^{SUB} VAP (section 4.1)

e.g., object relative clause:

- (19) *jeʔ pi si-jeʔ be*
 ask NML.OBJ want-ask 2PL
 ‘Ask what you want to ask.’

VAP (section 4.2)

subject complement:

- (20) *bat ʔou mɛ:ŋ ʔe nɔuʔ ʔoi saŋe*
 pluck 1SG tea.leaf NSIT be.full three day
 ‘I have been picking tea leaves for three days.’
 (Lit. ‘My picking tea leaves has filled three days.’)

⁹ The present study limits its scope to clauses; monoclausal constructions, i.e., complex predicates where one component can be considered subordinate (Van Valin & LaPolla 1997:ch. 8), will therefore not be considered. In Rucing, these include i) complementation strategies used for modal and phasal verbs and ii) other types of serialization.

x^{SUB} AVP (section 4.3)

conditional clause:

- (21) *sai ʔəu ʔe bok ka:k ʔu saŋe, ʔəu di bok ka:k ta:k*
 if 1SG NSIT do,make house one day 1SG FUT do,make house be.big
 ‘If I build a house one day, I will build a big one.’

non-finite AVP (section 4.4)e.g., object complement of manipulation verb:¹⁰

- (22) *ʔiban ʔan tu taih ʔan diək {wain}*
 woman 3SG NEG give 3SG drink wine
 ‘His wife does not let him drink wine.’

The remainder of this section will discuss each of these structures and the function(s) it expresses. Finally, section 4.5 will briefly consider AVP structures functioning as object complements, which are shown to be main clauses.

4.1 x^{SUB} VAP

This is the most pervasive subordination strategy in Rucing, encompassing all three basic subordinate relations. All relative clauses, complement clauses of perception-cognition-utterance (PCU) verbs, and adverbial clauses of time, reason, and consequence manifest themselves in this form. The subordinating morpheme differs according to the function of the construction; they are displayed in Table 1.

Table 1: Functions of x^{SUB} VAP structures

<i>(ʔe)</i>	VØ(P)	subject relative
<i>(pi)</i>	VAØ	object relative
<i>(bɔ)</i>	VA(P)	complement of PCU verb
<i>dɔ (etc.)</i>	VA(P)	adverbial: temporal
<i>koʔ</i>	VA(P)	adverbial: reason
<i>kʰəu</i>	VA(P)	adverbial: consequence

All of these clause types can contain aspectual-modal and negation particles.

4.1.1 Relative

Relative clauses are introduced by a different relativizer depending on the role that is being relativized on. Rucing makes a binary distinction between subject relativization, using *ʔe/ʔi*, and object relativization, using *pi*. The object relativizer can be used for core and oblique objects. Since the relativizers also function as nominalizers within identical structures, as will be discussed below, they are glossed as nominalizers here.

Subject relative clauses are formed as [*ʔe/ʔi* V Ø (P)].¹¹ Examples are given in (23-25). The allomorphs of the relativizer show free variation. It is very likely that the morpheme can be traced back to proto-Palaungic **ʔi:ʔ*, a morpheme with generic human reference according to Sidwell (2015:105). Reflexes are visible in words like *ʔiban* ‘woman’, *ʔimae* ‘man’, *ʔikat* ‘older sibling’, *duʔi* ‘person’, and

¹⁰ In this construction, the *ʔan* ‘him’ following matrix verb *taih* has the double function of matrix-clause-P and subordinate-clause-A. There is, however, syntactic evidence that this argument forms a tighter unit with the subordinate VP: in reply to a polarity question, one could answer *heʔ taih*, literally ‘yes, [she] let/allowed’, eliding the entire complement clause *ʔan diək*. In contrast, it is ungrammatical to answer *heʔ taih ʔan* ‘yes, [she] let/allowed him’; the verb *diək* must then be added too. Moreover, no adverb can be inserted between *ʔan* and *diək*.

¹¹ I analyze the gapped subject as occupying a postverbal slot based on (a) the fact that relative clauses containing both subject and object (e.g., examples 29 and 30 with gapped indirect object) have VAP order, and (b) the assumption that all relative clauses are essentially instances of the same grammatical mechanism.

vi ‘CLF.HUM’. In all of these words, the realization of the high vowel in /ʔi/ likewise varies and may be lowered to /e/ or, when stressed, /ei/. If the subject relativizer is indeed related to this human referring morpheme, this semantic association has been lost due to generalization; it is used for both humans and non-humans, including inanimates (see e.g., (36) below).

- (23) *goeban ʔe leh haɔ tɔŋ*
 girl NML.SUBJ exit go NVIS
 ‘the girl who left’
- (24) *ʔaɔ mauh daua ʔe jɔuk magaua kʰe:k kalah ʔatau*
 NEG EXIS male NML.SUBJ sell fish inside market today
 ‘The man who sells fish isn’t there at the market today.’
- (25) *ʔɔu kudikʰam ʔe car ge ŋaoh silaʔ.rakʰidaua¹²*
 1SG king NML.SUBJ order 3PL kill PN
 ‘I am the king who ordered them to kill Sila Rakhidua.’

Object relative clauses take the form [*pi* V (A) Ø], with the subject after the verb. The non-subject relativizer is *pi*. While *pi* also happens to have a phonological variant *pɛ*, in this case the alternation is constrained. Due to vowel harmony, *pi* becomes *pɛ* before aspect markers *hɛʔ* and *hoeʔ* (Janzen 1987:199). By far the most frequently relativized role by *pi* is the core object, i.e., patient, as illustrated in (26-28). Examples of a relativized recipient are given in (29-30). Other oblique roles like instrument and location may also be relativized, as seen in (31-32).

Patient

- (26) *ʔabum pi di kʰɛ ʔɔu*
 story NML.OBJ FUT tell 1SG
 ‘the story I am going to tell’
- (27) *hmaɔ jak ʔi:ŋ pi duaʔ damusau di ʔɔu*
 INTER meat chicken NML.OBJ give hunter OBL 1SG
 ‘Where is the chicken that the hunter gave to me?’
- (28) *{ʔamouʔ} pi luiʔ mɛ muŋ ʔaɔ mauh ʔe nap*
 deed NML.OBJ do.wrong 2SG TCL NEG EXIS NML.SUBJ know
 ‘What crime you’ve committed, there’s no one who knows.’

Recipient

- (29) *ʔan mɔ di duaʔ di duʔi pi si-duaʔ ʔan*
 3SG TCL FUT give OBL people NML.OBJ want-give 3SG
 ‘He gives [life] to whom he is pleased to give it.’ [NT, John 5: 21]
- (30) *ʔŋ=mouh vi pi di duaʔ ʔɔu kʰamuj nei*
 3SG.NHUM=be.so person NML.OBJ FUT give 1SG snack PROX
 ‘It is the one to whom I will give this piece of bread.’ [NT, John 13: 26]

¹² In the nominal predicative construction, the copula *mouh* is generally omitted (as in this example), except when aspectual-modal or polarity features need to be expressed.

Other: instrument, location

- (31) *hmaɔ puəʔ pi haʔ mɛ makasoŋ*
 INTER knife NML.OBJ slice 2SG tomato
 ‘Where is the knife you sliced the tomatoes with?’

- (32) *{san} pouʔ.mak pi diək sɿ ʔe haɔ bjək pɔ plɔuʔ*
 shop milk NML.OBJ drink often 1PL.INCL go change other place
 ‘The tea shop we used to go to (*lit.* drink at) has moved to another place.’

Both relativizers may be omitted, as seen in (33-34). The agentive or patientive meaning can be inferred from the context.

- (33) *goemae hoŋ blɿ {tʰulɿn} tu groh, ʔan*
 boy eat fruit peanut NEG speak 3SG
 ‘The boy who ate the peanuts did not say anything.’

- (34) *hmaɔ ha dak mɛ*
 INTER curry cook 2SG
 ‘Where is the curry you cooked?’

The head noun that the relative clause is attached to may also be left out, resulting in a headless relative clause. Such clauses are very common in natural speech. Examples (35-36) display subject relativization, while example (37) illustrates object relativization.

- (35) *ʔɔu di kʰam di bɛ:ŋ ʔi kʰuʔ dɛ kʰuʔ ce:ŋ mɛ*
 1SG FUT bear LNK become NML.SUBJ wipe hand wipe foot 2SG
 ‘I can become the one who cleans your hands and feet, [...]’

- (36) *heʔ pɿŋ di vi sam ʔi gjaʔ*
 NSIT get LNK SEQ build,plant NML.SUBJ be.good
 ‘We can build good ones [i.e., houses] again.’

- (37) *mauh pi lɔu, pi si-kala:k*
 EXIS NML.OBJ need NML.OBJ want-speak
 ‘Is there something you need, something you'd like to talk about?’

These headless relative clauses are reminiscent of participant nominalizations, ubiquitous in Tibeto-Burman languages and languages spoken in the Americas, for instance (Genetti 2011; Zariquiey, Shibatani & Fleck 2020). Such nominalizations are referential expressions (e.g., ‘the one who Vs’), which look like relative clauses when modifying another noun. In Rucing, too, there are indications that the constructions [*ʔɛ/ʔi* V ∅ (P)] and [*pi* V (A) ∅] have functions extending further than a modifying one and can also be referential in themselves. They may refer to a non-specific, indefinite set of objects or people, as shown in (38-40). This is unlike what one would expect from a headless relative clause, in which case the elided head noun would supposedly refer to a known, specific entity that can be retrieved from the context.

- (38) *pi cauh muŋ ŋaoh*
 NML.OBJ buy TCL kill
 ‘Whatever you buy is expensive.’ (Lit. ‘What one buys kills.’)

- (39) *ʔan naʔ lɔu kʰam pi di kʰam ʔan*
 3SG still need bear NML.OBJ FUT bear 3SG
 ‘He still had to endure what he would endure.’

- (40) *dɔ.tei.nei ʔaɔ mauh ʔɛ he:k lɔt*
 now NEG EXIS NML.SUBJ raise pig
 ‘Now there’s no one who keeps pigs anymore.’
Not: ‘Now the one(s) who keep(s) pigs isn’t/aren’t there anymore.’

Moreover, the *pi* clause and *ʔɛ* clause as in these examples also constitute the most common, natural way to express ‘the one that is Ved’ or ‘the one that Vs’, i.e., with specific reference, in the corpus. Apart from contexts in which a lexically rich noun is required, they almost never occur with a (light) noun meaning ‘thing’ or ‘person’. If this were a headless relative clause, we would expect the head to be present at least sometimes and not omitted by default.

Closer investigation of the formal properties of [*ʔɛ/ʔi* V ∅ (P)] and [*pi* V (A) ∅] in referential versus modifying function will be necessary to make a conclusive statement about the nominal or clausal status of these constructions.

4.1.2 Complement: perception-cognition-utterance

Complements of verbs pertaining to perception, knowledge, propositional attitude, or utterance have the structure [*bɔ* VAP]. Such verbs in Rucing include *jau* ‘see, hear’, *nap* ‘know’, *kala:k* ‘speak’, *hɲim* ‘believe’, and *ʔɣhloh* ‘remember’ (41-45). As seen in (41), the complementizer *bɔ* is optional.

- (41) *tu dɔŋ jau bok mɛ hoŋ kʰahmau*
 NEG EXP see do,make 2SG eat evening
 ‘[I]’ve never seen you cook dinner.’

- (42) *kɔ heʔ nap bɔ jam ʔɔu*
 so,then NSIT know COMP die 1SG
 ‘I know for sure that I will die.’ [Janzen_RA_023 (Janzen 1987:490)]

- (43) *sɔi jau ʔan, ʔɔu di kala:k bɔ hrɔt mɛ di ʔan*
 if see 3SG 1SG FUT speak COMP search 2SG OBL 3SG
 ‘When I see him, I will tell him that you are looking for him.’

- (44) *ʔɔu tu hɲim bɔ mah ʔekat ʔɔu mɛ*
 1SG NEG believe COMP hit older.sibling 1SG 2SG
 ‘I don’t believe that my brother hit you.’

- (45) *sɪ.sɪ ʔɣhloh bɔ hap maua ʔan lasiɔk tɔŋ kʰe:k jok*
 PN remember COMP put.inside mother 3SG key NVIS inside basket
 ‘Zin Zin remembered that her mother put the key in the basket.’

These complement clauses freely take aspectual-modal and negation marking, as in independent clauses (46-47).

- (46) *ʔan heʔ nap bɔ di tah ʔekat ʔan*
 3SG NSIT know COMP FUT come older.sibling 3SG
 ‘He knows that his brother/sister will come.’

- (47) *ge kala:k di ʔəu ʔaɔ loʔ mɛ jangoʔ*
 3PL speak OBL 1SG NEG descend 2SG PN
 ‘They told me you’re not going to Yangon anymore.’

The verb-initial order is only allowed in a constellation where it can function as an embedded clause. That is, it needs to be preceded by the matrix verb in order to occupy the standard object position in an AVP clause (48). Preposing is not allowed.

- (48) a. *mɛ jəu pa-keh ɡəŋ ʔa-doe matuke*
 2SG see CAUS-fall child 3SG-DIST bicycle
 ‘Did you see that boy tip over the bicycle?’
- b.* *pa-keh ɡəŋ ʔa-doe matuke mɛ jəu*
 CAUS-fall child 3SG-DIST bicycle 2SG see
 Intended: ‘Did you see that boy tip over the bicycle?’

There are two PCU verbs that behave differently, *tʰa:k* ‘think’ and *taoh* ‘say’: these exclusively take verb-medial complements. In the case of *tʰa:k*, illustrated in (49-50), a plausible reason for this is that it is a Shan loanword, originally *tʰàŋ* ‘think’.¹³ Shan is a language with basic AVP/SV order. In a complement clause following *tʰàŋ*, it also displays SV order (51). Thus, Rucing probably borrowed the clause structure accompanying the verb along with the verb itself.

- (49) *naʔ tʰa:k marɔʔ ʔan koe kʰe:k saʔoʔ hɔi tɔŋ*
 still think frog 3SG stay inside hole tree NVIS
 ‘He again thinks that his frog is inside that tree hole.’
- (50) a. *ʔəu tʰa:k sɔ.sɔ mɔuk blai {la:k}*
 1SG think PN like fruit jackfruit
 ‘I think Zaw Zaw likes jackfruit.’
- b.* *ʔəu tʰa:k mɔuk sɔ.sɔ blai {la:k}*
 1SG think like PN fruit jackfruit
 Intended: ‘I think Zaw Zaw likes jackfruit.’

- (51) Shan (Tai-Kadai)
mén tʰàŋ wa tɔŋ.tɔŋ tɔŋ.mɔŋ kyá sě mén hây sě mén nɛy yù;
 3SG think say PN PN be.good EMPH 3SG be.bad EMPH 3SG ADV.PROX stay
 ‘He thought that Tawng Toy and Tawng Mawng were good and that he was bad.’

As for *taoh* ‘say’, it is not directly apparent why the complement can be verb-medial only. As shown by (52-53) versus (54), this is not affected by whether it is a direct or indirect speech complement.

¹³ The development from final /-ŋ/ to /-k/ is a known regular sound change in the present Rucing variety. Janzen (1978) described the language as having a tendency to preglottalize final nasals, especially /-ŋ/. Nowadays, phonetic realizations of final /-ŋ/ in Janzen’s wordlist vary on a continuum between /ʔŋ, ʔ, k/. The degree of denasalization depends on the speaker and the phonetic context in which it occurs, see for more details Buakaw’s (2012) phonological study. Since in my experience with my informants the nasal is usually barely audible, I have used /-k/ for all instances of former /-ŋ/ in the transcription.

- (52) *ʔan taoh ʔan tu mauh {pasẽ}*
 3SG say 3SG NEG EXIS money
 ‘He said he had no money.’
- (53) *sai mɛ ʔɛ taoh mɛ rɔuk ʔɔu, ʔɔu di tah kʰo mɛ*
 if 2SG NSIT say 2SG love 1SG 1SG FUT come by 2SG
 ‘If you tell me you love me, I will come to you.’
- (54) *hoeʔ tei kudikʰam taoh, ʔɔu muŋ di haɔ*
 COMPL MEDL king say 1SG TCL FUT go
 ‘After that, the king said: “I’m going.”’

Utterance or quotative verbs are special in terms of the possible functional complement types they can take. While there is cross-linguistic variation in the availability of these options, two types can be distinguished in general: indirect (reported) speech and direct speech. Most, but not all, languages make a morphosyntactic distinction between the two (Jäger 2007). Interestingly, in the case of *taoh*, it is used with both direct and indirect speech with no grammatical distinction (apart from prosodic features). This is in contrast to *kala:k* ‘speak, tell’ as shown before (see (43)), where an indirect speech complement can be marked clearly as subordinate to the matrix clause by the verb-initial order.

Looking at frequencies impressionistically, *taoh* is clearly the more general verbum dicendi among the two, being considerably more common in narratives and conversations. Furthermore, it occurs much more often with direct than indirect speech. This is often combined with a different, clearly unembedded clause position: the matrix X *taoh* frequently comes after the direct speech, as in (55). A speculative explanation for the verb-medial order could thus be that the direct speech construction, due to its extremely high frequency in use, was extended to the indirect speech environment, which may have been grammatically marked in the past.

- (55) *na hmaɔ marɔuʔ ʔae vaoʔ, gae taoh*
 side INTER frog 1DU.INCL go.to 3DU say
 ‘“Where did our frog go?” they say.’

4.1.3 Adverbial: temporal, reason, consequence

Temporal clause

Subordinate clauses with a time-setting function are generally introduced by a temporal morpheme, obligatorily followed by a VAP/VS construction. Examples are given below in (56-62) for clauses expressing ‘when’ (*dɔ*), ‘after’ (*vae* and *pan di*), ‘before’ (*ʔaiʔ*), and ‘since’ (*sin*) relations. These morphemes never occur with verb-medial structures, but do occur with (pro)nominals, e.g., *dɔ ʔan tei* [when-3SG-MEDL] ‘at that time’. As can be seen, the temporal clause may both precede and follow the matrix clause.

‘When’

- (56) *dɔ kʰɔu ʔɛ hlua, ʔjaiʔ kasaɔ hɿ*
 when fry 1PL.INCL cigar.leaf tear,break very.much tree
 ‘When we fried the cigar leaves we destroyed the trees.’
- (57) *dɔ ŋɿʔ ʔɔu masɔu, ʔan leh haɔ*
 when look 1SG dog 3SG exit go
 ‘When I looked at the dog, it walked away.’

‘After’ (past situation)

- (58) *vae hoeʔ ɲauk hle:ŋ ʔan, jauua ʔɔu cuh di kala:k ʔabum*
 after COMPL be.silent last.long 3SG female 1SG start LNK speak story
 ‘After being silent for a long time, my grandma started to tell a story.’

‘After’ (future situation)

- (59) *pan di leh ʔɔu kalah, ʔɔu di cauh tah {foun.kaʔ}*
 after FUT exit 1SG market 1SG FUT buy come SIM.card
 ‘After I leave the market, I will buy a SIM card.’

‘Before’

- (60) *ʔaiʔ ɲaoh ʔan malət, ʔan pak riʔ*
 front kill 3SG pig 3SG tie be.tight
 ‘Before he killed the pig, he tied it up tightly.’

- (61) *suʔ hra:k ʔaiʔ ʔiʔ mɛ*
 brush tooth front sleep 2SG
 ‘Brush your teeth before going to bed.’

‘Since/from’

- (62) *sin teʔ mɛ, mauh jok nei pae.nei*
 since be.small 2SG EXIS school PROX like.this
 ‘Has there been a monastery like this one since you were young?’

Occasionally the temporal subordinator is omitted, leaving the verb-initial order as the only marker of subordination, as in (63-64).

- (63) *ɲaʔ ʔɔu masɔu, ʔan leh haɔ*
 look 1SG dog 3SG exit go
 ‘When I looked at the dog, it walked away.’

- (64) *hoŋ ge boŋ tɔŋ, gun maua ge si-mouh ʔiʔ*
 eat 3PL rice NVIS father mother 3PL want-be.so sleep
 ‘When they were eating, their parents were still sleeping.’

Although the surface structure in all cases is [x^{SUB} VAP], these temporal subordinators do not necessarily form a syntactically homogeneous set. It is not clear for all of these what their diachronic origin is. For some words, a nominal source seems very likely based on their syntactic behavior in other contexts. For instance, when *ʔaiʔ* ‘front’ and *pan* ‘back’¹⁴ occur with a locative sense, they are preceded by oblique preposition *di* or light noun *na* ‘place’; see (65-67). The noun phrase following *ʔaiʔ* or *pan*, if present, is its possessor (‘in front/the back of X’), a role that always goes morphologically unmarked in Rucing. The history of the other morphemes is much less transparent.

- (65) *ge vi kʰrupkʰraɔ di ʔaiʔ ʔan*
 3PL SEQ kneel OBL front 3SG
 ‘Then they knelt in front of him.’ [NT, Matthew 27: 29]

¹⁴ Probably derived from one of its other meanings, ‘footstep’.

- (66) {si:k} maŋviʔ, ʔan koe na pan ʔa-doe
 machine rotate 3SG stay place back 3SG-DIST
 ‘The turning machine, is it behind him?’
- (67) jaɔ {jesu} vi jɔʔ di pan ʔan ʔu jeŋ
 holy PN SEQ look OBL back 3SG one time
 ‘Jesus turned around (*lit.* looked behind him) at once.’ [NT, Matthew 9: 22]

Reason clause

A reason or cause is expressed by *koʔ* ‘because, on account of’, a loanword from Shan *kɔp*.¹⁵ Similar to the temporal morphemes, *koʔ* introduces either a verb-initial construction (68-70) or a noun phrase (71). The prevalence of the verb-initial structure here is especially noteworthy considering its absence in the source language Shan. The reason clause may both precede and follow the matrix clause; see (70) for a postposed adverbial clause.

- (68) koʔ tu mɔuk ʔɔu jak, ʔɔu tu ha
 because NEG like 1SG meat 1SG NEG eat.curry
 ‘I’m not eating, because I don’t like meat.’
- (69) koʔ bɔ mouh mɛ guk dak ʔɔu, ʔɔu kɔ kala:k
 because TAM be.so 2SG friend bottom 1SG 1SG so,then speak
 pi k^hriək ʔɔu
 NML.OBJ hide 1SG
 ‘Because you are my best friend, I will tell you my secret.’
- (70) ʔɔu heʔ dɔŋ ja:ŋ koʔ hŋap ga:ŋ ʔɔu
 1SG NSIT EXP cry because be.difficult work 1SG
 ‘I have cried because my work has been hard.’ [NT, Acts 20:19]
- (71) kalaiʔ koʔ ʔa-tei
 break because 3SG-MEDL
 ‘They broke because of that.’

Consequence clause

The counterpart of the reason, which I will call the consequence, is introduced by *k^hɔu*. It appears in a range of syntactic constellations, but for the present discussion it suffices to limit ourselves to one of them: [*k^hɔu* VAP].¹⁶

Examples are provided below. Compare (72) with *koʔ* ‘because’ (68) above. These sentences express the same semantics but portray a different pragmatic perspective: here, the consequence is the presupposed part of the sentence, whereas the reason forms the asserted information. An imaginable context of utterance could be one where it is obvious that the speaker is refusing to eat, and the speaker then explains why that is the case.

¹⁵ The phonetic discrepancy is due to the glottal replacement of final stops in Rucing.

¹⁶ Apart from a verb-initial clause, *k^hɔu* may also be followed by a pronominal phrase, nothing, or a verb-medial clause. My provisional hypothesis for the latter two cases, however, is that a reanalysis of the clause boundary has taken place and *k^hɔu* has become a clause-final particle.

- (72) *ʔɔu tu mɔuk jak kʰɔu tu ha ʔɔu*
 1SG NEG like meat CONS NEG eat.curry 1SG
 ‘It is because I don’t like meat that I’m not eating.’

The presupposed nature of the *kʰɔu* clause is perhaps best illustrated by (73), which is the answer to the question ‘why (are you so sure)?’, after the speaker claims to know who the thief was. *hmaɔ ʔɨ pae kʰɔu nap ʔɔu* can be roughly translated as ‘why I know that is because...’, which is clearly repeating presupposed information that is already known to the addressee. The same holds for the question in (74), where the fact that the father asked the eldest brother to do something is already known to the speaker asking the question.

- (73) *hmaɔ ʔm=pae kʰɔu nap ʔɔu, ʔɔu hɛʔ jau dɔ dauh ʔan*
 INTER 3SG.NHUM=do,happen CONS know 1SG 1SG NSIT see when pick.up 3SG
 ‘(Why I know that is) because I saw him take it.’

- (74) *ʔekat ta:k mɛ tei, ʔan ploŋ hlɔŋ ge*
 older.sibling be.big 2SG MEDL 3SG be.strong be.more.than 3PL
kʰɔu {caŋ} gun mɛ
 CONS order father 2SG
 ‘Was it because your eldest brother is stronger than them that your father asked him?’

- (75) *kʰɔu saɔ katu bau ʔɔu ʔan li pah ca:ŋ can*
 CONS hurt back uncle 1SG 3SG like carry bag weigh,be.heavy
 ‘My uncle’s back is hurting because he (usually) carries heavy bags.’

As can be seen in examples (73-75), like temporal and reason clauses, the consequence clause also has a flexible position with respect to the matrix clause.

4.2 VAP

Verb-initial subordinate clauses that never contain a subordinating morpheme are subject complement clauses. Subject complementation constitutes a broad, heterogeneous category and has so far not been subject to the systematic treatment in typological research as object complementation has. This section looks at evaluative and descriptive clauses (section 4.2.1) and clauses in which the subject complement has an adverbial function (section 4.2.2).

4.2.1 Subject complement: descriptive and evaluative

When the subject complement consists of a verb or a verb and a patient only, it is zero-marked, as illustrated in (76-77). Recall that *ʔɨ=* is a dummy subject pronoun, which here serves as a placeholder until the actual subject appears.

- (76) *ʔɨ=hɨna kala:k groh daʔa:k*
 3SG.NHUM=be.difficult speak language Ta'ang
 ‘It is difficult to speak Palaung.’

- (77) *sai sɔ kasum ka:k nei, ʔɨ=hɨui vi gɔuk ɨaɔ, rɛ ɨaɔ*
 if MOD burn house PROX 3SG.NHUM=be.easy SEQ cover fire block fire
 ‘If this house would burn, it would be easy to put out the fire, to block the fire [...]’

If the complement-internal subject is present, however, it appears postverbally as in (78-81), creating the marked VAP structure. Example (80) suggests that the verb can take aspectual markers, but further

systematic study is needed to verify this. As seen in the examples here, the subject clause can both precede and follow the matrix clause. This is not unexpected typologically: AVP languages tend to have the option to express the subject clause in the object clause position, with variation in the ordering governed by information structure (Schmidtke-Bode 2014:170).

- (78) *haɔ ʔan hmʌŋ kaŋah*
 go 3SG be.good laugh
 ‘His walking is funny.’
- (79) *kalɔu sɔ.sɔ di guk ʔan tu ʔah*
 lie PN OBL friend 3SG NEG be.good
 ‘Zaw Zaw's lying to his friend was bad.’
- (80) *hɛʔ pae gɔŋ gua nei gae hɛʔ pae*
 NSIT do,happen child male.ADDR PROX and NSIT do,happen
gɔŋ gua tei ʔaɔ la-kʰahne:ŋ
 child male.ADDR MEDL NEG RECP-be.the.same
 ‘What this boy does and what that boy does are not the same.’
- (81) *huaʔ kalɔuʔ ɲarae, huaʔ ɲaɔ, hauh ɲaɔ si-pei ʔŋ=hauh*
 be.hot pot hell be.hot fire sear fire want-be.like 3SG.NHUM=sear
kʰe:k dɔu ʔɔu
 inside body 1SG
 ‘The heat of hell, the heat of the fire, the searing heat of the fire still seemed to be searing inside this body of mine.’

4.2.2 Subject complement as temporal adverbial

The main clause and the complement clause may also have a very different semantic relationship; the matrix verb sometimes has an adverbial function with respect to the subordinate event. Accordingly, this type has also been called adverbial complementation (Enfield 2007:463). The three verbs discussed here all describe a temporal duration: *hle:ŋ* ‘take, last (long); be a long time since’, *nɔuʔ* ‘be full’, and *paŋ* ‘get’.

In examples (82-83), *hle:ŋ* expresses the time that has elapsed since the subordinate event, while example (84) refers to the duration of the subordinate event itself. The two meanings are formally distinguished by the use of dummy pronoun *ʔŋ=* for the latter sense, as seen in (84, 85b). Again, the position of the subject complement is syntactically flexible and probably determined by information structural factors.

- (82) *vi gjaʔ ka:k nei ɲnam vi hle:ŋ kʰɔbaleh*
 SEQ be.good house PROX not.yet SEQ last very.much
 ‘It hasn't been very long yet since these houses became good.’
- (83) *ɲnam vi hle:ŋ ʔmprae ge ka:k nei*
 not.yet SEQ last repair 3PL house PROX
 ‘It hasn't been long yet since they repaired this house.’
- (84) *kʰa hmaɔ ʔŋ=di e:ŋ tu koe ʔekat mɛ di ka:k*
 amount INTER 3SG.NHUM=FUT last NEG stay older.sibling 2SG OBL house
 ‘How long will your brother be away?’

- (85) a. *ʔe hle:ŋ gaʔ so:so*
 NSIT last be.born PN
 ‘It’s been a long time since Zaw Zaw was born.’
- b. *ʔŋ=hle:ŋ gaʔ so:so*
 3SG.NHUM=last be.born PN
 ‘Zaw Zaw’s birth took a long time.’

The verbs *nouʔ* ‘be full’ and *paŋ* ‘get’ are used in the same way: both can take a lexical temporal duration expression as their object, with their grammatical subject expressing the event. Compare the literal use with the metaphorical, temporal use of *nouʔ* in (86) and (87-88), respectively. Note that the meaning of *nouʔ* is different from the English ‘be full’; it also means that the subject is filling a container (either expressed or unexpressed) to its full capacity.¹⁷ The verb-initial clause expressing the event thus takes the regular position of a noun.

- (86) a. *ʔum nouʔ ʔu kapaɔ*
 water be.full one bottle
 ‘The water fills one bottle.’
- b. *ʔum heʔ nouʔ*
 water NSIT be.full
 Approximately: ‘The water fills its container completely.’

- (87) *ʔou dihlaɔ kʰa ʔu ci:k, bat ʔou me:ŋ ʔe nouʔ ʔoi saŋe*
 1SG rest amount one a.little pluck 1SG tea.leaf NSIT be.full three day
 ‘I’m resting a bit, I have been picking tea leaves for three days.’ (Lit. ‘My picking tea leaves has filled three days.’)
- (88) *saɔ gan ʔou ʔe nouʔ ʔu sanam*
 hurt aunt 1SG NSIT be.full one year
 ‘My aunt has been sick for a year.’ (Lit. ‘My aunt’s being sick has filled a year.’)

The use of *paŋ* ‘get’ in the same contexts has an identical meaning, as illustrated by (89-90).¹⁸ In (90), the object ‘this book’ has been fronted for topicalization.

- (89) *do.tei.nei saɔpʰau gan ʔou heʔ paŋ ʔu sanam*
 now be.sick aunt 1SG NSIT get one year
 ‘My aunt has been sick for a year now.’
- (90) *bap bɔk nei tih ʔou ʔe paŋ {toun} {paʔ}*
 book CLF PROX read 1SG NSIT get three week
 ‘I’ve been reading this book for three weeks already.’

¹⁷ (Near-)equivalents in languages in the area are not uncommon; the same semantics are covered by e.g., *pié* ‘be full’ in Burmese and *təmɜ* ‘be full’ in Lao (Enfield 2008).

¹⁸ This function of a verb meaning ‘get, acquire’ has also been noted in Lao and Northern Zhuang (Enfield 2002:275, 2003:133–140), both Tai languages, and Khmer (Austroasiatic) (Haiman 2011:344).

The verb-initial order within the subject clauses is clearly triggered by the use of the verbs mentioned and the embedded status of the clause, since it is ungrammatical without the presence of the matrix clause (91).

- (91)* *saɔp^hauu gʌn ʔɔu*
 be.sick aunt 1SG
 Intended: ‘My aunt is sick.’

4.3 *x^{SUB} AVP*

This type of subordinate clause does not take on a different order, nor is it reduced in their finiteness, as is apparent from its ability to contain aspectual-modal and negation markers. The possibilities of aspect-mood marking are, however, restricted as opposed to in an independent clause. This strategy is used to express conditionals.

4.3.1 *Adverbial: conditional*

There are two conditional markers: the verb *taih* ‘give’¹⁹ and *sai*. The latter could be derived from the adverb *sai* ‘often, always’, which would be a plausible semantic change, or it could simply be a homonym. The subordinators *taih* and *sai* can be used interchangeably; there seems to be some dialectal variation.

A predictive conditional clause, expressing a potential future situation, is marked for the new situation aspect. Examples are given in (92-93). Irrealis mood marking with *di* is deemed ungrammatical here.

- (92) *taih ʔɛ paŋ, cɔuʔ bleh ʔu jɛŋ*
 give NSIT get raise release one time
 ‘If/when it's done, they take it out right away.’

- (93) *taih ʔɔu ʔɛ piu di daiʔ mɛ, mɛ vi daiʔ ʔɔu*
 give 1SG NSIT forget LNK call 2SG 2SG SEQ call 1SG
 ‘If I forget to call you, please call me.’

Conditions pertaining to general or present situations, in contrast, are not marked for any aspect or mood. This difference is illustrated by example (94a) versus (94b). In example (95), new situation aspect is unacceptable due to its incompatibility with the semantics of ‘like’, which is typically interpreted as a general rather than temporally bound property in this context.

- (94) a. *sai mɛ nap groh {ʔiŋgəliʔ}, manɔu taih bɔ di ʔɔu*
 if 2SG know speech English teach give also OBL 1SG
 ‘If you know English, please teach me (too).’
- b. *sai mɛ ʔɛ nap groh {ʔiŋgəliʔ}, manɔu taih bɔ di ʔɔu*
 if 2SG NSIT know speech English teach give also OBL 1SG
 ‘If/when you know English (in the future), please teach me (too).’

¹⁹ Although conditional marking is not a widespread grammaticalized function of a ‘give’ verb, it may be a tendency in the family or area: in Muak Sa-aak (Angkuic, Palaungic) the verb *wa:* ‘give’ likewise functions as a predictive conditional/temporal marker (Hall 2019).

- (95) *sai mɛ (*ʔɛ) mɔuk ha nei, ha ta:k*
 if 2SG NSIT like curry PROX eat.curry be.big
 ‘If you like this curry, eat a lot of it!’

Counterfactual conditionals are not formally distinct from factual generic/present conditionals; they likewise do not carry any aspectual-modal marking (as opposed to predictive conditionals), as shown in (96-97).

- (96) *sai mɛ tʰoŋ groh mauua mɛ, mɛ bɔ saɔ*
 if 2SG listen speech mother 2SG 2SG NEG be.sick
 ‘If you had listened to your mother, you wouldn't be sick.’

- (97) *sai mɛ nap di diək ʔan {bija}, mɛ di taih {pisa:ŋ} di ʔan*
 if 2SG know FUT drink 3SG beer 2SG FUT give money OBL 3SG
 ‘If you had known he would just drink beer, would you have given him the money?’

The fact that one of the conditional markers is *taih* ‘give’ may have had an impact on the structure of the clause. As will become clear in section 4.4, *taih* is associated with a following verb-medial structure in various functions. Among the subordinators discussed in this paper, *taih* is the only one that has an unambiguously verbal origin. This makes it plausible that the deviant verb-medial structure of the conditional clause has some connection to it.²⁰

Instead of a subordinate clause, a coordinated main clause is very often used, as in (98). The conditional semantics are then only implied.

- (98) *mɛ tu kala:k, ʔɔu tu nap*
 2SG NEG speak 1SG NEG know
 ‘If you hadn't told me, I would not have known.’

4.4 Non-finite AVP

Non-finite verb-medial clauses in Rucing are used as object complements of verbs of manipulation and as purpose clauses. These are schematically displayed in Table 2.

Table 2: Functions of non-finite AVP structures

(A <i>taih</i>)	P[/A V (P)] _{NFIN}	object complement of permissive-causative ‘let’
(A <i>si-taih</i>)	P[/A V (P)] _{NFIN}	object complement of desiderative different-subject ‘want’
(A <i>car</i>)	P[/A V (P)] _{NFIN}	object complement of manipulative ‘order, ask’
(clause) <i>taih/sai</i>	P[/A V (P)] _{NFIN}	adverbial: purposive ‘so that’

The verbs of manipulation, as termed by Givón (2001b:151–3), included here are *taih* ‘let (*lit.* give)’, *si-taih* ‘want (X to V) (*lit.* want-give)’, and *car* ‘order, command’. The purpose clause also makes use of the verb *taih* ‘give’, but it will be shown that its behavior differs from that of a finite declarative verb.

²⁰ A reviewer asked whether the conditional marker *taih* could be analyzed as a verb, which would make the conditional clause verb-initial. I would argue against such an analysis since conditional *taih* does not display any verbal properties: it cannot be negated or otherwise modified, and it can be followed by a preverbal aspect marker introducing a VP, as seen in example (92). While verbs can directly follow verbs in Rucing, a VP cannot.

The non-finiteness of these subordinate clauses is apparent from the fact that aspectual-modal markers never occur. The next subsections will shed more light on why these functions in particular share this clause structure.

4.4.1 Complement: manipulation

Conceptually, an enablement (e.g., ‘let’), a manipulative (e.g., ‘order’), and a different-subject desiderative (‘want X to V’) situation are much alike: in all cases the agent of the main event either enables, wants, or tries to cause someone else to do something. The affected entity of this main event (also ‘manipulee’) is thus simultaneously the agent of the enabled, desired, or caused subordinate event. Compare the schematic event structures of literal giving (A), an enablement event (B), a manipulative event (C), and a desiderative, different-subject event (D).

- | | | | | | |
|----|----------|----------------------------|------------|------------------|------------------|
| A. | GIVER | intentionally instigates | [RECIPIENT | gains control of | THING] |
| B. | ENTITY A | intentionally instigates | [ENTITY B | gains control of | EVENT/SITUATION] |
| C. | ENTITY A | (attempts to) instigate(s) | [ENTITY B | is involved in | EVENT/SITUATION] |
| D. | ENTITY A | wants (to instigate) | [ENTITY B | is involved in | EVENT/SITUATION] |

(cf. Newman 1996:179)

These relationships are also reflected in the syntax: the object of the main clause is coreferent with the subject of the subordinate clause and this argument is only expressed once. Another cross-linguistic characteristic of such complement clauses is a tendency to be non-finite (Givón 2001b:152). Based on their semantic and syntactic similarities, Givón (2001b:151–3) groups such verb types together as manipulation verbs. The Rucing manipulation verbs *taih* ‘let (*lit.* give)’, *si-taih* ‘want (X to V) (*lit.* want-give)’, and *car* ‘order, command’ all share the structure [A V MANIPULEE V].

As part of a widespread pattern in Southeast Asia and South China, Rucing has also developed a ‘give’ causative manifested as [A GIVE CAUSEE V] (Matisoff 1991:427–431; Jenny 2015). In the case of Rucing, the construction based on the verb *taih*²¹ has a permissive-causative meaning, as in (99–100).²²

- (99) *gun* *ʔan* *tu* *taih* *sĩ.sĩ* *leh* *dilaɔ* *na* *raɬ*
 father 3SG NEG give PN exit play side outside
 ‘Zin Zin’s father never lets her go out and play.’

- (100) *dəŋ* *ʔi* *lou* *taih* *gəŋ* *ʔan* *hoŋ* *pʰe* *di* *ʔaiʔ*
 every person need give child 3SG eat be.full OBL front
 ‘First everyone should let the children eat all they want [...]’ [NT, Mark 7: 27]

The question of how this [A GIVE CAUSEE V] construction arises in the first place is worth briefly considering here. Based on the conceptual structures of a ditransitive event (A) and an enablement event (B) above and data from several Southeast Asian languages, a plausible hypothesis centers on a direct structural mapping of CAUSEE to Recipient and EVENT to Theme (see Jenny 2015). Languages like Thai, Khmer, and as seen here, Rucing Palaung, display the same causative construction, despite having a seemingly unfitting default Theme-Recipient order in ditransitive clauses. Complex or heavy Themes are, however, generally placed after the Recipient in these languages, as the Rucing examples (101–

²¹ Two other ‘give’ verbs in Rucing, *hliu* and *duaʔ*, have undergone the same development. Whether there are any semantic or pragmatic differences is not yet clear. There does seem to be dialectal variation.

²² This has cross-linguistically been shown to be a more primitive semantic extension from ‘give’ than the causative; the presence of a ‘give’ causative in a language implies that of a ‘give’ permissive-causative (Heine & Kuteva 2002; Lord et al. 2002; Lee 2015).

102) show. Thus, the post-CAUSEE position of the EVENT, typically more complex than an NP, may be explained as the result of being treated as a complex Theme. However, in the case of Rucing, another issue in need of an explanation would be the role and fate of obligatory recipient marker *di* in the process. This will not be taken up in the present study and is left for future research (see also the brief diachronic discussion of the purposive construction in the next section).

(101) *sɔ.sɔ di dua? bap di mɛ*
 PN FUT give book OBL 2SG
 ‘Zaw Zaw will give you a book.’

(102) *sɔ.sɔ di dua? di mɛ bap pi taoh ʔɔu*
 PN FUT give OBL 2SG book NML.OBJ say 1SG
 ‘Zaw Zaw will give you the book I mentioned.’

Derived from the permissive-causative construction is the desiderative different-subject construction, expressing ‘want X to V’. This is illustrated in (103-104). The main verb *si-taih* is a combination of the desiderative prefix and the verb ‘give’. The path of change appears to go from ‘want to let X V’ > ‘want X to V’, involving the bleaching of the permissive meaning. The presence or absence of the permissive reading is context-dependent. When entirely bleached, *taih* takes on the grammatical function of marking the switch-reference instead (i.e., the agent of the subordinate event is not the agent of the ‘wanting’), as also observed in e.g., Mon (Jenny 2005:212–2), Lao (Enfield 2007:275, 447), and Khmer (Haiman 2011:309).²³

(103) *maua ʔan si-taih ʔan he.ŋ tʰam*
 mother 3SG want-give 3SG study text
 ‘His mother wants him to study.’ or ‘His mother wants to let him study.’

(104) *ʔɔu tu si-taih mɛ leh haɔ*
 1SG NEG want-give 2SG exit go
 ‘I don’t want you to leave.’

Manipulative events like ‘order’, ‘command’, ‘make’, and ‘ask’ are expressed by the verb *car*. Examples are shown in (105-107).

(105) *ʔɔu kudikʰam ʔɛ car ge ŋaoh silaʔrakʰidaua*
 1SG king NML.SUBJ order 3PL kill PN
 ‘I am the king who ordered them to kill Sila Rakhidua.’

(106) *ʔɔu car mɛ kʰriək*
 1SG order 2SG hide
 ‘I asked you to keep it a secret!’

(107) *gun ʔɔu car ʔekat ta:k ʔɔu gar hɔi tei*
 father 1SG order older.sibling be.big 1SG chop tree MEDL
 ‘My father asked my eldest brother to chop down the tree.’

²³ Bisang (1996:577) described this as the grammaticalization pathway from causative to ‘conjunctive’ verbs. Enfield (2007) called the function of ‘give’ here a ‘dummy causative’.

As mentioned above, considering the semantic similarity between verbs within the broad manipulation class, it is common for them to adopt the same syntax. The extension of one construction to a semantically close context is a common process in the development of complement constructions (Cristofaro 2014). The particular syntactic structure used by these three verbs is possibly motivated by the verb *taih* ‘give’, which gives rise to very similar structures in Southeast Asia. At the same time, in the case of *car*, the appearance of its complement may also have a different motivation, namely language contact. It is plausible that *car* ‘order’ is a Shan loanword. The Shan verb *c̣eu* means ‘employ, use, command, order, send’ and likewise takes a verb-medial complement clause, as seen in (108). Thus, Rucing may have copied the construction that comes with the verb.

(108) Shan (Tai-Kadai)

mén hěn náy.mǒ len má cú mén sě c̣eu
3SG see PN run come approach 3SG EMPH employ

mén k^hew kwà n̄ŋ néu hín k^ha ʔ;
3SG enter go sit in house POL ASRT.POL

‘... he saw Nang Mo come running toward him, and she told him to go and sit in the house.’

4.4.2 Adverbial: purpose

The polyfunctional *taih* ‘give’ is also used as a purpose clause marker in the construction [main clause] GIVE [subordinate clause]. The development of a ‘give’ verb into a marker of a purpose clause is typologically common, and even more in the areal context of Southeast/East Asia (Newman 1996; Heine & Kuteva 2002:149–52; Lord et al. 2002; Jenny 2015). The similarity in conceptual structure between the event of giving and a purposive construction is illustrated below in A and B, respectively.

- A. GIVER intentionally instigates [RECIPIENT gains control of THING]
B. ENTITY in EVENT intentionally instigates [EVENT B]

(cf. Newman 1996:179)

Examples of purpose clauses are given in (109–111). Interestingly, *taih* ‘give’ and the earlier mentioned *sai* (section 4.3 on the conditional) are interchangeable in the purposive context as well. It is unclear how this developed and what the origins of this *sai* are.²⁴

- (109) *ʔŋ=lɔu sam taih ʔŋ=k^hanaua taih ʔŋ=k^hahaʔ*
3SG.NHUM=need build give 3SG.NHUM=be.separate give 3SG.NHUM=be.separate
‘One needs to build [houses] so that they’re separate.’

- (110) *ʔɔu di {ga:ʔ} {sa:ʔsou} taih ʔm=bɛ:ŋ pi*
1SG FUT offer offering give 3SG.NHUM=become NML.OBJ
di ha saŋe me
FUT eat.curry day 2SG
‘I will offer them to you so that they will become your lunch.’

²⁴ While the possible scenario that both developed into conditional markers independently is not hard to imagine due to their semantics ‘give’ and ‘always, often’, it is much less obvious what the path of change may have looked like for these purposive markers. *sai* may have spread to this environment through analogy, being generalized to something identical to *taih*, or it may have other currently unknown origins.

- (111) *glaoh {kʰalok} ʔan di nei taih mabah tɿŋ bah haɔ dɔh,*
 loosen trigger 3SG FUT PROX give catapult NVIS shoot go DIST.ADV
taih ʔan coh kate ʔum, taih ʔan jam kate ʔum
 give 3SG fall middle water give 3SG die middle water
 ‘...then let go of the trigger here so that the catapult shoots him far away, he falls into the middle of the sea and dies in the middle of the sea.’

While apparently still verbal, *taih* in this structure does not behave like a declarative main verb, as opposed to when used as a permissive causative verb. This is evident from the negated counterpart of the purpose clause, expressing ‘lest’, as displayed in (112-113). Instead of the declarative negators *tu* or *ʔaɔ* (cf. (99)), *taih* takes the prohibitive particle used in imperative clauses, *kɔ*, together literally meaning ‘don’t give’.²⁵

- (112) *kukra:n jau pi tu lɔu ʔou tu cauʰ*
 thing kind NML.OBJ NEG need 1SG NEG buy
kɔ taih ʔŋ=toiʔ kuŋ rɿŋ
 PROH give 3SG.NHUM=be.used.up be.much money
 ‘I don’t buy unnecessary things not to waste money.’

- (113) *ʔau taua satɔu dui nei, kɔ taih giaʔ hmat me*
 INTJ take,wear sweater sheep.wool PROX PROH give cold catch 2SG
 ‘Wear this sweater so you don’t get sick.’

The polyfunctionality of *taih* gives rise to the question whether and how the purposive and the permissive are diachronically connected. As already mentioned, the phenomenon is shared by several other languages in the area (Lord et al. 2002; Jenny 2015). However, cross-linguistic data do not firmly point to one direction or the other, and the two functions may also have evolved independently.²⁶ From a language-internal point of view, too, none of these possibilities can be excluded. Until more in-depth investigation is conducted, the diachronic question shall thus remain unanswered.

4.5 Other AVP with PCU object semantics

The PCU verb complements, which have verb-initial order as previously shown (section 4.1.2), occasionally seem to occur in the main-clause-like verb-medial order, see for instance (114).

- (114) *ʔahmau ʔou jau ʔan bok ga.ŋ di {san}*
 yesterday 1SG see 3SG do,make work OBL shop
 ‘Yesterday I saw him working in the shop.’

Although the ‘complement clause’ express the semantics of a typically subordinate event, it does not show any signs of subordination. On the contrary, example (115a) shows that a temporal adverb like ‘yesterday’ can stand between the main event and the ‘subordinate’ event, providing evidence that the

²⁵ The exact pattern [PROH give] ‘lest’ is also found in Khmer (Haiman 2011:311).

²⁶ For both Thai and Khmer, a pathway GIVE > purposive complementizer > permissive-causative main verb has been posited based on primarily cognitive or pragmatic grounds (Iwasaki 2008:471ff.; Haiman 2011:325–7). On the other hand, see Jenny (2015:164, 166) for the view that the two functions developed independently in SEA languages. Likewise, a wider typological survey (Lee 2015) does not show any support for a cross-linguistic implicational hierarchy of the GIVE purposive and permissive.

perceived event is not embedded in the object position but is in fact an independent clause. In an embedded, verb-initial clause as in (115b), the same adverb cannot be inserted; it can only be in sentence-initial or -final position.

- (115) a. $\lambda\text{ɔu}$ *jau* λahmau , λan *bok* {*ga:ŋ*} *kʰe:k* {*san*}
 1SG perceive yesterday 3SG do work inside shop
 ‘I saw (it) yesterday. He was working in the shop.’
- b. $\lambda\text{ɔu}$ *jau* (* λahmau) [*bok* λan {*ga:ŋ*} *kʰe:k* {*san*}]
 1SG perceive yesterday do 3SG work inside shop
 ‘I saw him working in the shop (*yesterday).’

It is not surprising that perceived events like these are sometimes expressed by a main clause. Speakers of Rucing generally prefer juxtaposition of simple clauses over clause combining, especially in natural spontaneous speech. The fact that no orthography of the language existed until some 50 years ago (Kojima 2016) may be of relevance. Languages with no written form are often characterized by the predominance of paratactic structures over embedded structures, the latter of which arise later than the former (Karlsson 2009:195–6).

5 Discussion and conclusion

Table 3 summarizes the subordination strategies found in Rucing Palaung. They can be subdivided into four main patterns. The verb-initial structures clearly predominate in the subordinate domain, being used in all relative clauses and several types of complement and adverbial clauses. We have also seen that in these relative and complement clauses, the subordinator is optional, leaving the word order as the sole distinctive feature. Moreover, it is striking that nearly all of the verb-medial constructions occur together with *taih* ‘give’. This is unlikely to be a mere coincidence: the semantics of the original event structure of lexical ‘give’ has probably been a significant factor in driving these developments. The number of typological precedents provide sufficient support for this. The manipulative construction with verb *car* may have assumed the same form as the *taih* construction based on shared semantics, or it may have borrowed it from Shan along with the verb itself. As for the verb-medial clauses expressing the semantics of a PCU verb’s object (discussed in section 4.5), their syntactic behavior shows that they are main rather than subordinate clauses. With a plausible source-based motivation for the verb-medial subordinate structures at hand, it thus appears that in Rucing, verb-initial word order is strongly linked to the subordinate nature of the clause.

Viewed from a typological perspective, this finding poses more questions, since it appears to be the case that “variation in word order [...] ha[s] little relevance for clause linkage strategies used within the domain of subordination” (Cristofaro 2003:51). Apart from some Germanic languages, languages across the world do not tend to use the order of elements in a clause to signal backgrounded or non-asserted information. We might therefore wonder: (a) has Rucing perhaps undergone a similar process as the Germanic languages in question, namely word order change in main clauses only with subordinate clauses representing an older order? Or alternatively, (b) is the phenomenon that looks like word order variation in Rucing actually something else, in which case the emerging word order on the surface is merely epiphenomenal?

One potential candidate that would confirm the latter is nominalization. The linguistic environments in which verb-initial order turns up in Rucing are environments that are notoriously prone to nominalization in some language families and areas of the world (e.g., Tibeto-Burman, South American languages). These are syntactic slots that can but do not need to be filled by a nominal element. In addition to these slots, however, verb-initial structures have also been found in exclusively nominal positions, such as prepositional phrases. These occurrences require more investigation. There may be more indicators of nominality than the syntactic position: differential argument encoding, if present, may be well hidden in Rucing as possessors are structurally identical to postverbal subjects, being marked only by their position. Moreover, there are signs of non-main-clause-like marking of the

object in some verb-initial clauses, in Rucing as well as other Palaung varieties. Patients in verb-initial clauses are occasionally marked as an oblique object (see also Weber, this publication: §3.2). It is hoped that further research will reveal the extent of this special object-marking and whether the verb-initial structures are all instances of a single phenomenon or formally distinct.

Table 3: Subordinate structures and their functions in Rucing Palaung

Pattern	Construction	Specific function	Subordination type	
x^{SUB} VAP	(<i>?ε</i>)	VØ(P)	subject relative	relative
	(<i>pi</i>)	VAØ	object relative	
	(<i>bɔ</i>)	VA(P)	object of PCU verbs	complement
	<i>dɔ</i> (etc.)	VA(P)	temporal	adverbial
	<i>koʔ</i>	VA(P)	reason	
	<i>kʰɔu</i>	VA(P)	consequence	
VAP		VA(P)	subject	complement
x^{SUB} AVP	<i>taih/sai</i>	(A)V(P)	conditional	adverbial
non-finite AVP	(A <i>taih</i>)	P[/A V (P)] _{NFIN}	object of permissive-causative ‘let’	complement
	(A <i>si-taih</i>)	P[/A V (P)] _{NFIN}	object of desiderative different-subject ‘want’	
	(A <i>car</i>)	P[/A V (P)] _{NFIN}	object of manipulative ‘order, ask’	
	<i>taih/sai</i>	P[/A V (P)] _{NFIN}	purposive ‘so that’	adverbial

In sum, the present study has shown that in Rucing, verb-initial constituent order is a sufficient but not a necessary condition for a clause to be morphosyntactically subordinate. Where a subordinate clause is not verb-initial, considering the history of the construction yields a plausible explanation. The findings raise new synchronic and diachronic issues regarding clause structure and syntax in Rucing; a future comparison with geneologically and areally close languages would certainly be fruitful.

Glosses used in this paper

ADDR addressee, ADV adverb, CAUS causative, CLF classifier, COMP complementizer, COMPL completive, CONS consequence, DIST distal, DU dual, EMPH emphatic, EXCL exclusive, EXIS existential verb, EXP experiential, INCL inclusive, INTER interrogative, INTJ interjection, LNK linker, MEDL medial, MOD modal, NEG negation, NHUM non-human, NML nominalizer, NSIT new situation aspect, NVIS non-visible, OBJ object, OBL oblique, PL plural, PN proper name, PROH prohibitive, PROX proximal, RECP reciprocal, SG singular, SEQ sequential, SUBJ subject, TAM tense-aspect-modal, TCL topic-comment linker

References

- Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage. In Isabelle Bril (ed.), *Clause Linking and Clause Hierarchy*, 51–101. Amsterdam, Philadelphia: John Benjamins.
- Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and Mainland South East Asian languages. *Studies in Language* 20(3). 519–590.
- Buakaw, Supakit. 2009. The phonetic realization of denasalized final nasal consonants in No Lae Palaung. *Journal of Language and Culture* 28(1).

- Buakaw, Supakit. 2012. A phonological study of Palaung dialects spoken in Thailand and Myanmar, with focuses on vowels and final nasals. Salaya: Mahidol University PhD dissertation. <http://mulinet11.li.mahidol.ac.th/thesis/2555/cd470/4837496.pdf> (15 May, 2018).
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and Topic*, 22–25. New York: Academic Press.
- Comrie, Bernard & Sandra A. Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol. 3, 334–381. 2nd ed. Cambridge: Cambridge University Press.
- Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.
- Cristofaro, Sonia. 2012. Participial and infinitival complement sentences in Ancient Greek. In Volker Gast & Holger Diessel (eds.), *Clause Linkage in Cross-Linguistic Perspective: Data-Driven Approaches to Cross-Clausal Syntax*, 335–362. Berlin, Boston: Mouton de Gruyter.
- Cristofaro, Sonia. 2014. *Complementation hierarchies and the development of complement constructions*. <https://www.researchgate.net/publication/301690491> (29 November, 2018).
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford, New York: Oxford University Press.
- Deepadung, Sujaritlak. 2011. Names and sub-groups of the Palaung: A preliminary field survey in Yunnan and Kengtung. *MANUSYA: Journal of Humanities* 14(2). 88–104.
- Deepadung, Sujaritlak, Ampika Rattanapitak & Supakit Buakaw. 2015. Dara’ang Palaung. In Mathias Jenny & Paul Sidwell (eds.), *The Handbook of Austroasiatic Languages*, vol. 2, 1065–1103. Leiden, Boston: Brill.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the World*. 22nd ed. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Enfield, Nick J. 2002. On genetic and areal linguistics in Mainland South-East Asia: Parallel polyfunctionality of “acquire.” In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, 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, New York: RoutledgeCurzon.
- Enfield, Nick J. 2007. *A Grammar of Lao*. Berlin, New York: Mouton de Gruyter.
- Enfield, Nick J. 2008. “Case relations” in Lao, a radically isolating language. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 808–819. Oxford: Oxford University Press. www.oxfordhandbooks.com (9 December, 2018).
- Gast, Volker & Holger Diessel. 2012. The typology of clause linkage: status quo, challenges, prospects. In Volker Gast & Holger Diessel (eds.), *Clause Linkage in Cross-Linguistic Perspective: Data-Driven Approaches to Cross-Clausal Syntax*, 1–36. Berlin, Boston: Mouton de Gruyter.
- Genetti, Carol. 2011. Nominalization in Tibeto-Burman languages of the Himalayan area: A typological perspective. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian Languages: Diachronic and Typological Perspectives*, 163–193. Amsterdam, Philadelphia: John Benjamins.
- Givón, Talmy. 2001a. *Syntax: An Introduction*. Vol. 2. Amsterdam, Philadelphia: John Benjamins.
- Givón, Talmy. 2001b. *Syntax: An Introduction*. Vol. 1. Amsterdam, Philadelphia: John Benjamins.
- Haiman, John. 2011. *Cambodian: Khmer*. Amsterdam, Philadelphia: John Benjamins.
- Hall, Elizabeth. 2019. *Adverbial subordinate clauses in Muak Sa-aak*. Presentation at International Conference on Austroasiatic Linguistics, 29-31 August 2019, Chiang Mai, Thailand.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

- Herlin, Ilona, Jyrki Kalliokoski & Laura Visapää. 2014. Introduction. In Laura Visapää, Jyrki Kalliokoski & Helena Sorva (eds.), *Contexts of subordination: Cognitive, typological and discourse perspectives*, 1–16. Amsterdam, Philadelphia.
- Hetterle, Katja. 2015. The structure of adverbial clauses in cross-linguistic perspective. *Adverbial Clauses in Cross-Linguistic Perspective*, 67–146. Berlin, Boston: De Gruyter Mouton.
- Hock, Hans H. 1986. *Principles of Historical Linguistics*. Berlin, New York: Mouton de Gruyter.
- Iwasaki, Shoichi. 2008. Bipolar distribution of a word and grammaticalization in Thai: a discourse perspective. In Anthony V. N. Diller, Jerold A. Edmondson & Yongxian Luo (eds.), *The Tai-Kadai Languages*, 468–483. London, New York: Routledge.
- Jäger, Andreas. 2007. *Coding strategies of indirect reported speech in typological perspective*. University of Queensland Working Papers in Linguistics 1.
- Janzen, Hermann. 1976a. The system of verb-aspect words in Pale. *Oceanic Linguistics Special Publications*(13). 659–667.
- Janzen, Hermann. 1976b. Structure and function of clauses and phrases in Pale. (Ed.) Philip N. Jenner, Laurence Thompson & Stanley Starosta. *Oceanic Linguistics Special Publication 13*. Honolulu: University Press of Hawai‘i. 669–691.
- Janzen, Hermann. 1978. A phonological description of Pale in comparison with Gold Palaung and Rumai Palaung. *Paper presented at Second International Conference on Austroasiatic Linguistics, December 19-21, 1978*. Mysore, India.
- Janzen, Hermann. 1987. Form and function of topicalization in discourse in Thai and Pale. William Carey International University PhD dissertation.
- Janzen, Hermann & Magarete Janzen. 1972. Grammar analysis of Pale clauses and phrases. *Journal of the Burma Research Society* 55(1–2). 47–99.
- Jenny, Mathias. 2005. *The Verb System of Mon*. Zürich: ASAS.
- Jenny, Mathias. 2015. The far West of Southeast Asia: “Give” and “get” in the languages of Myanmar. In Nick J. Enfield & Bernard Comrie (eds.), *Mainland Southeast Asian Languages: The State of the Art*, 155–208. Berlin, New York: Mouton de Gruyter.
- Jenny, Mathias. 2020. Verb-initial structures in Austroasiatic languages. In Mathias Jenny, Paul Sidwell & Mark Alves (eds.), *Austroasiatic Syntax in Areal and Diachronic Perspective*, 21–45. Leiden, Boston: Brill.
- Jenny, Mathias, Tobias Weber & Rachel Weymuth. 2015. The Austroasiatic languages: A typological overview. In Mathias Jenny & Paul Sidwell (eds.), *The Handbook of Austroasiatic Languages*, vol. 1, 13–143. Leiden, Boston: Brill.
- Karlsson, Fred. 2009. Origin and maintenance of clausal embedding complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 192–202. Oxford, New York: Oxford University Press.
- Kojima, Takahiro. 2016. Palaung orthographies: writing and the politics of ethnicity in Shan State. *The Newsletter* 75. 38. https://ias.asia/sites/default/files/IIAS_NL75_38.pdf (28 November, 2018).
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, vol. 2, Descriptive Application*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2008. Subordination in cognitive grammar. In Barbara Lewandowska-Tomaszczyk (ed.), *Asymmetric Events*, 139–149. Amsterdam, Philadelphia: John Benjamins.
- Lee, Wei-Wei. 2015. Polygrammaticalization of GIVE in the domain of transfer: A typological survey. Unpublished manuscript.
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In John Haiman & Sandra A. Thompson (eds.), *Clause Combining in Grammar and Discourse*, 181–225. Amsterdam: John Benjamins.

- Lord, Carol, Foong Ha Yap & Shoichi Iwasaki. 2002. Grammaticalization of “give”: African and Asian perspectives. In Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 217–250. Amsterdam, Philadelphia: John Benjamins.
- Mak, Pandora. 2012. *Golden Palaung: A grammatical description*. Asia-Pacific Linguistics (SEAMLES).
- Malchukov, Andrej & Akio Ogawa. 2011. Towards a typology of impersonal constructions: A semantic map approach. In Andrej L. Malchukov & Anna Siewierska (eds.), *Impersonal Constructions: A cross-linguistic perspective*, 19–56. Amsterdam, Philadelphia: John Benjamins.
- Matisoff, James A. 1991. Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization*, vol. 2, 383–453. Amsterdam, Philadelphia: John Benjamins.
- Milne, Leslie. 1921. *An Elementary Palaung Grammar*. Oxford: Clarendon Press.
- Newman, John. 1996. *Give: A Cognitive Linguistic Study*. Berlin, New York: Mouton de Gruyter.
- Ostapirat, Weera. 2009. Some phonological criteria for Palaung subgrouping. *Journal of Language and Culture* 28(1). 63–76.
- Rattanapitak, Ampika. 2009. Palaung Wordlist. *Journal of Language and Culture* 28(1). 77–122.
- Salaberri Izko, Iker. 2017. Word order in subordinate clauses: innovative or conservative? A typology of word order change. Universidad del País Vasco BA thesis.
- Schmidtke-Bode, Karsten. 2014. Complement clauses and complementation systems: A cross-linguistic study of grammatical organization. Jena: Friedrich-Schiller-Universität Jena PhD dissertation.
- Shintani, Tadahiko L. A. 2008. *The Palaung Language: Comparative Lexicon of its Southern Dialects*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.
- Sidwell, Paul. 2015. *The Palaungic Languages: Classification, Reconstruction and Comparative Lexicon*. Munich: LINCOM Europa.
- Siewierska, Anna. 1988. *Word Order Rules. Language*. London, New York, Sydney: Croom Hel.
- Thompson, Sandra A., Robert E. Longacre & Shin Ja J. Hwang. 2007. Adverbial clauses. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol. 2, 237–300. 2nd ed. Cambridge: Cambridge University Press.
- Van der Auwera, Johan. 1985. The predicate relatives of French perception verbs. In A. Machtelt Bolkestein, Casper de Groot & J. Lachlan Mackenzie (eds.), *Predicates and Terms in Functional Grammar*, 219–234. Dordrecht: Foris.
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Vennemann, Theo. 1975. An explanation of drift. In Charles N. Li (ed.), *Word Order and Word Order Change*, 267–305. Austin: University of Texas Press.
- Weber, Tobias. 2023. In Hiram Ring & Paul Sidwell, *Papers from the Eighth International Conference on Austroasiatic Linguistics*, 90–114. JSEALS Special Publication No. 11. University of Hawai‘i Press.
- Weymuth, Rachel. 2014. Negation in Palaung. University of Zurich MA thesis.
- Weymuth, Rachel. 2018. Verbal affixes in Rumai, Palaung. In Hiram Ring & Felix Rau (eds.), *Papers from the Seventh International Conference on Austroasiatic Linguistics*, 87–100. University of Hawai‘i Press.
- Zariquiey, Roberto, Masayoshi Shibatani & David W. Fleck (eds.). 2019. *Nominalization in Languages of the Americas*. Amsterdam, Philadelphia: John Benjamins.

SYLLABIFICATION IN LANGUAGE CONTACT BETWEEN FRENCH AND VIETNAMESE

Vera Scholvin
Freie Universität Berlin
<vera DOT scholvin AT fu-berlin DOT de>

Abstract

This article is dedicated to the syllabification in two language contact settings between Vietnamese and French and discusses gemination and related phenomena. First, it contributes to the theoretical discussion on borrowing processes in loanwords and presents a model that includes all extant views and also accounts for other language contact situations. In our data analysis, we compare loan data to experimental data, and find similar processes in both scenarios. Syllable boundaries are shifted, consonants are geminated and consonant slots are doubled in certain syllabic environments. The emergence of gemination is especially noteworthy as it is only a marginal phenomenon in French and has not been found to exist in Vietnamese. Results provide evidence for our hypothesis that optimal Vietnamese syllables are closed, as these syllables have the highest frequency in the native lexicon. This goes beyond, but also corresponds to the idea that Vietnamese has a dimoraic minimum, and this paper provides more evidence for it.

Keywords: loanwords, prosody, syllabification, gemination, language contact

ISO 639-3 codes: vie, fra

1 Introduction

In this paper, we discuss one aspect of prosodic language contact between French and Vietnamese: syllabification. We examine and compare the results of two studies about two distinct language contact situations: borrowing processes in loanwords and experimental data in a second language acquisition context collected in Vietnam and France. Many studies have been undertaken about borrowing and loanword adaptation, especially for the source languages English and French and a high variety of target languages (Poplack, 2018; Kang, 2011 for an overview on loanword phonology; and Kang 2010 for an overview on suprasegmental adaptation). The language pair Vietnamese-French has also been studied previously (Scholvin & Meinschaefer 2018; Nguyễn & Dutta 2017; Kang, Pham & Storme 2016, Huynh 2010; Toàn 1992; Barker 1969), but to a much lower extent than many other language pairs. The same is true also for studies on second language acquisition for this language pair.

These two languages are especially interesting to consider because phonologically, Vietnamese is often typologically rather typical when French is not and the other way round: while Vietnamese syllables do not have clear consonant groups, French syllables do. And while most French syllables are open, most Vietnamese syllables are closed. Syllabification patterns have not been a clear focus of the just cited studies about this language pair, and the insights we gain with our studies deepen our understanding of syllabification processes in language contact in general, universal tendencies, and Vietnamese phonology.

Finally, it is promising to contrast two very multi-layered language contact situations, namely loanword borrowing and experimental settings, and also to look for parallels between the two. If certain structures emerge in both contact situations, this strengthens the evidence for phonetic and phonological processes which are activated in language contact between the two languages. The syllabification processes we focus on are *syllable boundary shift*, *gemination*, and a related phenomenon we call *featural gemination*. An example for syllable boundary shift from our loan data is Vietnamese *lắc-lê* [lǎk.le] ‘spanner’ from French *la clé* [la.kle] ‘theART key’. An example from the experimental data is the response *ích khǎng* [ɪk.xǎŋ] to the French stimulus *écran* [e.kʁɑ̃]. In both examples, the syllable boundary is different comparing the French source lexeme or stimulus to the loanword or the response given by a speaker of Vietnamese: in the latter, it is put one segment later. Two examples for gemination are the loan *ắc quy* [ǎk.kwi] from French *accu(mulateur)* [a.ky(my.la.tœʁ)] ‘battery’, and the response

ít tong [it. tɔŋ̃m] to the French stimulus *étang* [e.tã]. In both examples, one consonant is doubled. As to featural gemination, we will deliver examples as well when we dive into the topic in greater detail.

The paper is organized as follows: section 2 of this paper contains a theoretical discussion of two models for loanword adaptation and presents background information related to the context of French loanwords in Vietnamese as well as segmental and suprasegmental information about the two languages. In section 3, we present the methods of our two studies, and in 4.1 and 4.2 we present the results of the first study (French loanwords in Vietnamese), relating syllabification patterns to our model. We continue by presenting experimental data from our second study in 4.3 (816 occurrences in total) and comparing it to the loan data (54 occurrences in total). Finally, we summarize and discuss our results in sections 5 and 6, providing evidence for our assumption that Vietnamese syllables are preferentially closed. This has implications for processes of syllabification in language contact settings and explains cases of syllable boundary shift as well as gemination and related phenomena.

2 Background

In this section, we present in 2.1 models derived from theoretical discussions about borrowing processes and our own model which unifies ideas of other models and accounts for borrowing processes as well as second language acquisition contexts. We continue by providing background information on borrowing from French into Vietnamese in 2.2 and segmental and suprasegmental structure of these languages in 2.3.

2.1 Theoretical framework: Models of borrowing processes in loanword adaptation

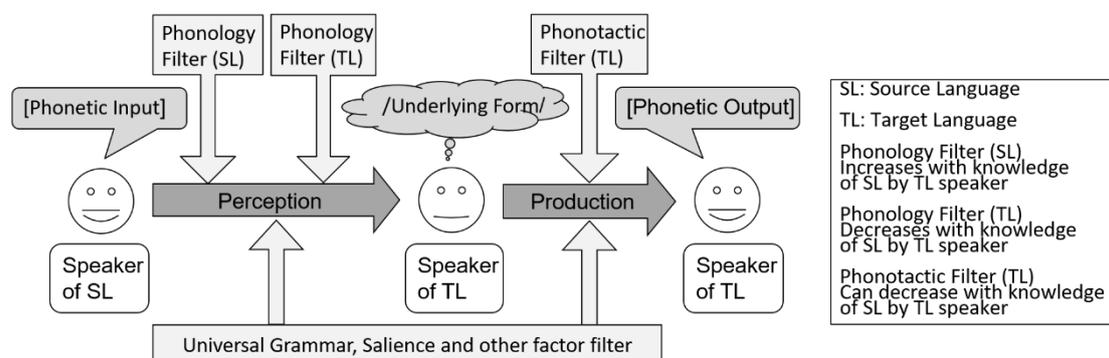
The point of departure for our research is the view on repair strategies as they occur in loanword adaptation, and the basic idea behind it is that speakers try on the one hand to be faithful to the source lexeme and on the other hand to the target language, which is, in the most cases, their native one (cf. Kenstowicz 2010). Basic principles behind this idea can also be applied to scenarios of second language acquisition (SLA), and we develop on this view and parallels between both language contact scenarios in the present article. Additionally, we give our view to the ongoing debate if phonetic or phonological processes are the driving force behind adaptation patterns found in loanword borrowing and SLA contexts and also consider the role of mono- and/or bilingualism linked to that. The idea that in borrowing, both perception and production has a role to play, date back to Silverman (1992), and the related question is: do specific processes take place in perception or production? The following questions are linked to the first question: are these processes of phonetic or of phonological nature? Are speakers who borrow speech material into their native language monolingual or do they have knowledge of the language from which they borrow?

A review of previous literature leads us to the observation that it is rather a question of degree than of binary distinctions. Consequently, it seems less important to clarify which of the two are true in general and in our specific case. We claim that phonetics *and* phonology is always involved, that it is always perception *and* production, and finally, that borrowers always have some knowledge of the language from which they borrow: When they borrow speech material, there must be some exposure, and this exposure is what we define as a very low degree of bilingual knowledge or the very first steps of second language acquisition, even if it stops in the moment of borrowing.

In contrast, many previous authors decide for one of the two possibilities, and two models have emerged. One of them is phonological, based on the idea that borrowers are bilingual, access phonological systems of both source and target language, and can abstract away from phonologically unnecessary phonetic details of the source language (cf. Paradis & LaCharité 1997 and 2005; Kang Pham, & Storme 2016). The alternate model is phonetic and different authors have elaborated on different versions of this model (cf. Peperkamp & Dupoux 2003; Broselow 2009; Yip 2006; Boersma & Hamman 2009 and Kenstowicz 2010). Still, some authors acknowledge that phonology also has a role to play, but only of the source language, as this model applies to contexts where speakers are monolingual. In line with our proposal, it has finally been claimed that there may not be a contradiction between the two models. The idea of degrees rather than the categorical view has therefore been proposed by various authors (cf. Kang, Pham, & Storme 2016, De Jong & Cho 2012, C. Ito 2014), and with our unified model applicable to all possible language contact settings, we continue our path in the

same direction. The scheme in Figure 1 is based on Kenstowicz (2003), and we now present our further development of his notions.

Figure 1: General scheme of borrowing processes in loanword integration



According to our model and in line with the accumulation of empirical research over more than thirty years (cf. Kang 2011) repair occurs either in perception, in production or in both. The new idea we bring in is that a distinction between a phonological vs. a phonetic input is not needed. According to our model, the input and its perception is always phonetic, but it is always filtered through phonology. The difference between different settings is just the degree to which it is filtered through which phonological system: The more a speaker has knowledge of the source language, the more the phonological filter of the source language increases. Consequently, the more the phonological filter of the target language decreases. Also, target language phonology intervenes for speakers with no perfect bilingualism, but our idea is that it can directly affect perception, even in the very first steps of second language acquisition. According to our view, the question should therefore not be if perception is phonetic or phonological, but instead which languages' phonological filter has a stronger influence on perception. Finally, even if the source languages' phonological filter is strong in perception, speakers may struggle with reproducing the speech material faithfully. Therefore, a third filter is active in production: the phonotactic filter of the target language. But phonotactic restrictions also apply when speakers are in principle able to reproduce the speech material: Speakers can adapt a loan for sociolinguistic or practical reasons to the target language's system. The fourth filter assembles other potential factors such as universal grammar and saliency (cf. Shinohara 2004, 2006; Davis & Cho 2006; Hsieh & Kenstowicz 2008, among others).

2.2 Specific situation: Borrowing processes in loanword adaptation from French to Vietnamese

In this section, we present and compare two specific case studies to each other and against the theoretical questions and our understanding of prosodical language contact presented so far. We will have a look at phenomena of syllabification which have been less in focus of research than other segmental and syllabic adaptation patterns and which can nourish the presented theoretical discussions with fruitful results. The starting point is the adaptation of French loanwords into Vietnamese, and we continue by comparing results of our study in the latter to an experimental study. Both studies share similarities and are synchronous in the sense that we look at both, loans and the experimental data, in a certain, present-day point in time. Still, we need to keep in mind that loans have entered into the Vietnamese language diachronically over a time span of about 100 years from the middle of the 19th to the middle of the 20th century, and that the experimental data has been collected in 2019 and early 2020. Therefore, the comparability is restricted, but the parallels which emerge suggest that there is the potential of similar processes in loanword adaptation and spontaneous language contact situations as created experimentally. For our second study, we have collected both oral and written data, but base this research only on the written data to have a better comparison to the loan data. A comprehensive presentation of the oral data is undertaken in our doctoral thesis (Scholvin, forthcoming).

2.3 Segmental and suprasegmental structure of French and Vietnamese

Before we dive into the studies themselves, we want to deliver the necessary background knowledge in order to understand the prosodic patterns in these language contact settings and describe segmental and suprasegmental properties of French and Vietnamese. The onset inventories based on the standard varieties of the respective languages are shown in Tables 1 and 2 (cf. Tranel 1987; and Walker 2001 for French; Kirby 2011; and Brunelle 2014 for Vietnamese). While Vietnamese and French have 14 correspondences for onset consonants, seven of the French onsets do not exist in Vietnamese, neither in onsets nor in codas. Vietnamese sounds we put in brackets, namely [p] and [r], are possible in loanwords (cf. Kirby 2011), but [p] is not part of the native inventory and [r] not part of this variety, but exists in other varieties of Vietnamese.

Table 1: French Onsets (Ile-de-France-Variety)

	Bilabial	Labio-dental	Dental	Post-alveolar	Palatal	Labio-palatal	Velar	Uvular	Glottal
Plosive	p b		t d				k g		
Nasal	m		n		ɲ				
Fricative		f v	s z	ʃ ʒ				ʁ	
Lateral			l						
Approximant	w				j	ɥ			

Table 2: Vietnamese Onsets (Hanoi-Variety)

	Bilabial	Labio-dental	Dental	Alveolar	Palatal	Velar	Uvular	Glottal
Plosive	(p) b		t tʰ	ɗ	tɕ	k		ʔ
Nasal	m		n		ɲ	ŋ		
Flap				(r)				
Frikative		f v		s z		x y		h
Lateral			l					
Approximant	w							

The strongest phonotactic differences between Vietnamese and French are in their coda consonant inventory. While the French coda consonant inventory is nearly identical to the inventory of onsets, only a very restricted number of eight Vietnamese onsets can also appear in coda position: [p t k m n ŋ j w]. The three plosives are unreleased in coda position. A restricted coda inventory is one of the parallels between Vietnamese and Cantonese (cf. Silverman 1992; Yip 1993), and we will come to more parallels later. After back rounded vowels [ɔ o u], the Vietnamese consonants /k ŋ/ are double articulated as [k̚ŋ̚] (cf. Kirby 2011). Interestingly, we find that in loans, this double articulation for /k ŋ/ following [u] does not have to take place which is why we transcribe the loan *xúc xúc* from French *saucisse* ‘sausage’ as [suk.sik], not as [suk̚ŋ̚.sik]. To the contrary, the double articulated [k̚ŋ̚] surface following the vowels [ɔ o] like in *hoóc-môn* [hɔk̚ŋ̚.mon] from French *hormone* ‘hormone’ or *xi phông* [si.fɔŋ̚m] from *siphon* ‘siphon’.

We continue with syllabic differences between Vietnamese and French. While in French, complex onsets and codas (CCV, CCVC, CVC, CVCC, etc.) exist, i.e., in *blé* [ble] ‘wheat’, the only possible cluster in Vietnamese is a consonant + the glide [w] in onset position, i.e., *hoa* [hwa] ‘flower’. Let us now come to gemination. It refers to the phonological distinction between phonetically long vs. short consonants. Gemination is a phenomenon in many languages such as Tashlhiyt Berber or Italian. A

minimal pair for Italian is *fato* [fato] ‘fate’ vs. *fatto* [fatto] ‘fact’. In early views such as presented in Chomsky and Halle (1968), geminates are represented by the feature [long], but later works such as Leben (1980) represent a syllable skeleton of geminated consonants with two consonant slots. Our paper is based on that view.

Bringing up the topic of gemination in a paper about prosodic language contact between French and Vietnamese is not obvious: Geminataion is only a very marginal phenomenon in French (cf. Delattre, 1971; Tranel, 1987) and has not been reported for Vietnamese in previous literature. Nevertheless, we can find minimal pairs in Vietnamese in word groups or compound words as in (1). In one of each pair, a syllable is closed by the same consonant with which the next syllable starts, and the two identical phonological consonants merge phonetically to a long consonant. The feature length is expressed phonetically in duration for sonorants and in closure for occlusives. In the second part of each pair, the first syllable is open and therefore there is only one onset consonant, which is realized phonetically shorter than the merge of coda and onset consonant in the other part of the minimal pair.

(1) Examples for word group or compound minimal pairs with differences between one vs. two consonants

[ka.kon]	vs.	[kək.kon]	or	[ka.məw]	vs.	[kam.məw]
cá con		các con		cá mèo		cám mèo
fish child		each child		fish cat		bran cat
‘little fish’		‘each child’		‘catfish’		‘cat food’ (special dry one)

A list of all French and Vietnamese vowels is not necessary for the understanding of this work. Still, it is important to keep in mind that French has no underlying distinction between long and short vowels. This is different from Vietnamese, where the two long vowels /a ɤ/ have their short counterpart. Other Vietnamese vowels can be phonetically long or short as well, and stand in complementary distribution depending on surrounding material (cf. Kirby 2011 and Hạo 2017). For instance, the vowel /i/ is long in *ly* [li] ‘cup’ and *lim* [lim] ‘ironwood’, but shorter and centralized in *linh* [liŋ] and *lich* [lik]. Finally, short vowels occur in Vietnamese only when the syllable is closed by a consonant, whereas long vowels occur in both open and closed syllables. And here is the second parallel to Cantonese, namely the dimoraic minimum that Vietnamese syllables have to fulfil, as there is no open syllable with a short vowel (cf. Brunelle 2017 for Vietnamese; Yip 2006, Yip 1993 and Cheung 1986 for Cantonese).

We now come back from the parallels between Vietnamese and Cantonese to the differences between Vietnamese and French: the reversed frequency of open vs. closed syllables. According to Thị Thúy Hiền Trần (2011), 74% of all Vietnamese syllables are closed, and the opposite ratio accounts for French (cf. Rousset 2004 and Vallée et al. 2001). CV structures are the most frequent syllable pattern universally (cf. Rousset 2004; Maddieson 1984; Jakobson & Halle 1956). While according to illicitness, Vietnamese goes more in line with universal tendencies than French not allowing for complex onsets and codas, we observe the contrary in French in terms of frequency for the preference of open syllables: With 73% CV syllables, French goes more in line with universal tendencies than Vietnamese.

Another difference between Vietnamese and French is that the number of Vietnamese syllables is quite restricted. “There are about 6500 distinct tonal syllables and 2400 distinct base syllables” (Le & Besacier 2009:1478). With the notion *base syllable*, the authors refer to segmental material of a syllable in Vietnamese, disregarding its tone assignment. In French, the number of syllables is larger for the reason alone that consonant clusters are possible, the vowel inventory is larger, and because all onset consonants are also possible in coda position. Finally, the unit of a syllable is very salient in Vietnamese. One of the reasons is that it is the unit which bears lexical tone. While lexical tone does not play a role in French, there is no toneless syllable in Vietnamese (cf. Michaud, Nguyen & Scholvin 2021 for a detailed literature review). In this paper, tones are not discussed further and therefore not marked in the IPA transcription, also because syllabification processes can be presented more clearly without distracting tone diacritics at the end of each syllable.

3 Methods

In this paper, we present small sub-studies from two larger-scale studies gathered for the author's PhD thesis (Scholvin, forthcoming), exploring them in more depth and relating them to each other. First, we introduce the methodology for study 1, a study on loanword adaptation, and then present the methodology for the experimental spontaneous imitation and adaptation task study 2, which in turn includes two sub-studies with different speaker groups.

3.1 Study on loanwords

Study 1 is about loanwords, and the corpus includes 532 nominal loanwords borrowed from French into Vietnamese during colonial times. We have selected these loans from a corpus of more than thousand loans, compiled of different corpora: Barker 1969; Huynh 2010; Nguyễn 2013. We selected only the number of loans which are still in use, even if not known and used by every speaker. We have checked current usage with the help of two native speakers, according to occurrences in the internet and Vietnamese lexicon entries (cf. Bui et al. 2003). Currently, we are extending the corpus by checking nominal nouns found in Nguyễn & Nguyễn (1992).

IPA transcriptions of the loans were generated automatically (Kirby 2008). This is possible because Vietnamese orthography reflects pronunciation in a very straightforward way. Also, syllable boundaries are marked by a space between two syllables, or, in loanwords, sometimes also by a hyphen. Finally, we checked automatic IPA transcriptions against two native speaker's pronunciation without any French knowledge. IPA transcriptions of the French source lexemes are based on a French dictionary (Rey-Debove and Rey 2013). For this paper, not all the 532 lexemes from our corpus are of relevance because the syllabification processes we focus on cannot be observed for all lexemes. For instance, if a lexeme is monosyllabic, it has no potential to enable the syllabification processes we focus on. In the first sheet of the excel file uploaded in Zenodo,¹ readers can find the list of lexemes to which syllabification processes have occurred. While we had no hypotheses about the syllabification patterns in focus before conducting our initial study and they were a surprising by-result, our initial hypothesis of our experimental study is that they reproduce herein. The motivation for the second study therefore is to experimentally enable the same processes in a different setting and with more data.

3.2 Study on spontaneous imitation and adaptation

Study 2 is a spontaneous imitation and adaptation task: Native speakers of Vietnamese listen to oral stimuli of French lexemes, repeat them in two conditions, and consequently write them down, adapting them to Vietnamese orthography. During an experimental run, each lexeme is repeated by each participant in two oral conditions, first in isolation and then embedded in a Vietnamese carrier sentence. What interests us in this paper is the third condition, the writing condition: After the two oral tasks, speakers are asked to write down the word as they have perceived and consequently produced it. Because the monolingual group does only access to one orthographic system, namely the Vietnamese one, we did not specify the instruction. Because the second group has access to both, the Vietnamese, and the French writing system, we specified the instruction and asked them to make use of Vietnamese orthography as if the word would have become a Vietnamese word. In this paper we added the IPA symbols corresponding to the written data, in parallel to the loan data. The writing of two of the monolingual participants is illegible and therefore, their written data is excluded from the analysis.

There are two reasons why we focus in this study only on the written data: First, a written loan corpus (study 1) can be compared best to a written experimental data corpus (study 2). Second, we can be 100% sure where participants put a syllable boundary only for the written data, because the space they put between two syllables is an unambiguous indicator of a syllable boundary, i.e., <cap bò lay du> [kap.bɔ.laj.zu] as a written response to the oral stimulus of the lexeme *cableuse* [ka.bløz] 'FEMwiring specialist'. Only when they forget the space, data remains unclear, i.e., in <chaplêro> [tɛaplɛɔ] as another response to the same stimulus lexeme.

¹ Accessible at: <https://zenodo.org/record/7038304>

The participants are two groups of native speakers of northern Vietnamese. The first group is 34 monolingual speakers of Vietnamese in Thái Nguyên, and the second group is made up of four native speakers of northern Vietnamese residing in Paris and speaking French as a second language. The second group is smaller because we stopped our research due to the outbreak of the COVID-19 pandemic. The stimuli are recordings of 49 French lexemes, but again only a subset in 20 of these stimuli were expected to enable the syllabification processes because again not all of the 49 stimuli have the potential to trigger them.

For the 20 lexemes we are interested in, all the responses are shown, no matter if syllabification patterns emerge or not, cf. column F in the second sheet of the excel file available on Zenodo. An example of its emergence is <am mặch> [am.mək] from French ‘hamac’ [a.mak] with gemination, and of its non-emergence another response <a mặch> [a.mək] to the same stimulus without gemination. Surprisingly, syllabification processes occur to more than the 20 stimuli lexemes and the additional, unexpected data can be found in column G, i.e., <bờ ghích cò> [bɔ̃.ɣik.kɔ̃] to the French stimulus *brique* [bʁik], where the [k] is geminated. Further evidence from data beyond syllabification patterns for our hypothesis to which we will come later is shown in column H, i.e., <pap> [pap] from French *pas* [pa] ‘step’, where a coda [p] is added.

4 Results

In this section we first consider the loan data (study 1) and in great detail the three syllabification types:

1. syllable boundary shift
2. complete gemination
3. featural gemination.

In the loan data, there is a structural system for when which type takes place. This is not the case for the experimental data (study 2): for the same stimulus lexeme, all three types of syllabification patterns are possible, and there is no system when which type occurs. While the loan data is quite restricted, the experimental data delivers more absolute and relative evidence for the occurrence of the three syllabification types. It additionally delivers more and unexpected evidence for our assumptions as we will see later in this section and in the discussion section.

Taking the loan data as our starting point, we can identify 93 instances where syllabification plays a role, in very different ways and for different motives, among them the insertion of epenthetic vowels to dissolve clusters (cf. Scholvin & Meinschaefer 2019 in a study of segmental and suprasegmental adaptation patterns). The complete number of attested syllabification types and items are treated in Scholvin (forthcoming), and in this paper we dive deeper only into the analysis of especially noteworthy patterns. The number of considered items is in total 54 and a much higher number will be given for the experimental data. These syllabification patterns have in common that they occur at the boundary between two syllables. Also, in all French source lexemes, the syllable before the boundary is open: CV.C(C)V(C), and the same syllable is closed in the Vietnamese loan: CVC.CV(C).

We can observe two different French source lexeme structures and two processes that result in this Vietnamese syllable structure: first, if the syllable behind the boundary has a complex onset in French as i.e., in *chou-fleur* [ʃu.flœʁ] ‘cauliflower’, the boundary is shifted between the two cluster consonants, as in *súp-lơ* [súp.lɔ̃]. This way, there is no intra syllabic consonant cluster in the loan. Second, if the onset of the syllable after the boundary is simple in French as in *potence* [po.tãs] ‘stem’ or *buffet* [so.sis] ‘buffet’, this consonant has the potential to either geminate as in *phót-tãng* [fot.tãŋ], or featurally geminate as in *búp phê* [búp.fe]. Consequently, the doubled consonant slot becomes the coda of the previous syllable. By introducing the term *featural gemination*² we refer to a subtype of gemination where features of the geminated consonant get lost either in the base or the double. In the just mentioned example, the double of [f] loses its manner of articulation.

² This term emerged during a conversation about this issue with Rachid Ridouane in early 2020.

4.1 Syllable boundary shift in loan data (study 1)

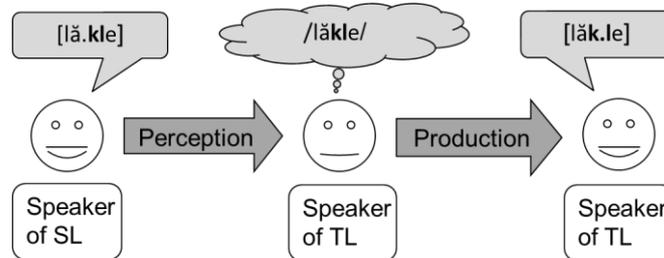
We start by presenting syllable boundary shift in loanwords: The first onset cluster consonant of a French source lexeme finds itself in the coda of the previous syllable in the Vietnamese target lexeme. Examples are shown in (2).

(2) Syllable boundary shift (25 items in total)

<i>French orthography</i>	<i>IPA</i>	<i>Vietnamese orth.</i>	<i>IPA</i>	<i>English translation</i>
<gicleur>	[ʒi.klœʁ]	<gíc-lơ>	[zik.lɤ]	‘jet’
<la clé>	[la.kle]	<lắc-lê>	[lăk.le]	‘spanner/the key’
<oeuf au plat>	[œ.fo.pla]	<ôp la>	[op.la]	‘fried egg’
<chou-fleur>	[ʃu.flœʁ]	<súp-lơ>	[sup.lɤ]	‘cauliflower’
<taxi>	[ta.ksi]	<tắc xi>	[tăk.si]	‘taxi’
<cyclopusse>	[si.klo.pus]	<xích lô>	[sik.lo]	‘pedicab’
<cable>	[ka.bl(ə)]	<cáp>	[kap]	‘cable’
<litre>	[li.tʁ(ə)]	<lít>	[lit]	‘liter’

While the segmental material of the French cluster consonants is also preserved after the syllable boundary shift in most of the examples, the first consonant of the cluster [fl] in *chou-fleur* and the [bl] in *cable* is replaced by [p]. This are typical consonant replacement patterns as summarized in Scholvin & Meinschaefer (2019). In our loan corpus, we find 25 items of the pattern described in (2) in total. From *cyclopusse*, the last syllable is truncated, additionally to syllable boundary shift between the first two syllables. The last two examples are somewhat special because there is a schwa after the consonant cluster in the French source lexeme which is truncated as well, but the first consonant of the truncated syllable is syllabified as the coda consonant of the previous syllable. After having illustrated the phenomenon in (2), Figure 2 offers an exemplary application to the model presented initially to the case of syllable boundary shift on the example of Vietnamese *lắc-lê* from French *la clé*. The idea is that sound material syllabified according to French syllabification rules pass via perception to the underlying representation in the borrower’s mind, and the string of sounds then gets syllabified according to Vietnamese syllabification rules in production.

Figure 2: Borrowing process for French *la clé* ‘the key’ to Vietnamese *lắc-lê* ‘spanner’



At first sight, one might consider syllable boundary shift as a newly discovered unique repair strategy to dissolve consonant clusters, but we continue by presenting first evidence for other reasons.

4.2 Complete and featural gemination in loan data (study 1)

In what follows, we describe the second syllabification pattern which delivers first evidence for assuming that syllable boundary shift is not a newly attested consonant cluster repair strategy but has other motivations. We deal with complete or featural gemination patterns. We start with the usual complete gemination process where a consonant is doubled. (cf. 3) and continue with featural gemination, where only some, but not all features of the base show up in the double (4).

The structure of the French source is similar to the structure which has the potential to enable syllable boundary shift. But there is one difference: There is no consonant cluster in the French source or stimulus lexeme structure CV.CV(C), and the second syllable starts with a simple onset. No cluster repair is therefore necessary, and still a process is enabled which results in the same syllable structure as for syllable boundary shift: the previous syllable gets closed. We arrive there by doubling the onset

consonant or at least its slot and some features of the second syllable, and while the base remains at its place, the double closes the first syllable.

The structure of the loan or of the response is CVC.CV(C), which is identical to the structure after syllable boundary shift previously discussed. As to the loan data, the process would principally be possible for in total 242 of the French source lexemes in our corpus. It can be attested for 29 items, which is at 12%. This shows that this adaptation process is not a necessary requirement of Vietnamese phonology, but an indication that there could be something in favour of it. The percentage is not extremely high, but the emergence of the pattern remains noteworthy. Among the 29 items in the loan data, there are 13 typical gemination cases (45%, cf. 3). The other 16 items are cases where not the consonant is perfectly doubled, but its slot and some of its features (55%, cf. 4). As said earlier, we call this featural gemination. Which features remain and which ones are changed follows the logic of Vietnamese phonotactics, in line with systematic consonant replacement patterns as described in Scholvin & Meinschaefer (2018).

(3) Complete gemination (13 items)

<accu(mulateur)>	[a.ky]	<ắc quy>	[ăk.kwi]	‘battery’
<automate>	[o.to.mat]	<át tô mát>	[at.to.mat]	‘circuit breaker’
<containeur>	[kõ.t(ə).nœʁ]	<công-ten-nơ>	[koŋm.ten.nɔ]	‘container’
<potence>	[po.tãs]	<phôt-tăng>	[fot.tãŋ]	‘stem’ (bicycle)
<pommade>	[põ.mad]	<pom-mát>	[põm.mat]	‘pommade’
<roquette>	[ʁõ.ket]	<rôc-két>	[zok.ket]	‘rocket’

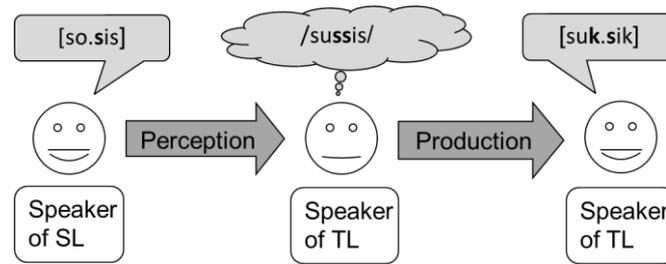
(4) Featural gemination (16 items)

<poupée>	[pu.pe]	<búp bê>	[bup.ʔe]	‘puppet’
<buffet>	[by.ʔe]	<búp phê>	[bup.ʔe]	‘buffet’
<essence>	[e.sãs]	<ét xăng>	[et.sãŋ]	‘gasoline’
<massage>	[ma.saʒ]	<mát xa>	[mat.sa]	‘massage’
<fascisme>	[fa.ʃi.sm(ə)]	<phát-xít>	[fat.sit]	‘fascism’
<saucisse>	[so.sis]	<xúc-xích>	[suk.sik]	‘sausage’

The reason for the emergence of featural gemination (4) instead of complete gemination (3) is that if the onset consonant in the second syllable remained as it was and became the coda consonant of the previous syllable, this would result in an illicit Vietnamese syllable because the consonant is not allowed in codas. Therefore, the double gets replaced systematically. For example, in Vietnamese *búp phê* from French *buffet*, the doubled labial fricative [f] is replaced by the labial plosive [p] in coda position, but the base remains [f]. This goes in parallel to patterns attested also in syllable boundary shift (cf. 2).

A special case is the loan *búp bê* from French *poupée*, where it is the base and not the double which changes features: the French onset consonant [p] gets voiced and becomes the Vietnamese onset implosive [b̥], while the double, which becomes the coda of the previous syllable, retains all features of [p]. This is because [p] is no native licit onset in Vietnamese, only became licit after a certain time of language contact with French, and still is not producible by all speakers of Vietnamese. Our idea of featural gemination, no matter if it is the base or the double which changes features, is that the consonant is underlyingly doubled and then passes the phonotactic filter in production and gets aligned to Vietnamese phonology. This is why we assume that it makes sense to classify this process as a special subtype of gemination, namely an incomplete one.

In Figure 3, we again present a schematic view of the application of our model to the subject of investigation, this time to featural gemination on the example of *xúc-xích* from French *saucisse*. Potentially, complete gemination takes place in perception, and the featural adjustment takes place consequently only in production passing via the phonotactic filter of Vietnamese.

Figure 3: Borrowing process for French *saucisse* to Vietnamese <xúc-xích> ‘sausage’

We observe that in the loan data, no phonological process might be going on but that simply orthography could be the factor of influence. This cannot completely be ruled out for instances where the double consonant appears twice in the orthography of the source lexeme as in *accu(mulateur)* <accu(mulateur)>. Still, in other source lexemes' orthography, there is no doubled consonant and this potential factor can be excluded (cf. 5). As in the experimental data (see the following section 4.3), there is no orthographic but only an oral stimulus, this factor can be completely ruled out for the whole dataset.

(5) Loans where orthography as a factor can be excluded

<automate>	[o.to.mat]	<áp-tô-mát, át tô mát>	[ap .to.mat], [at.to.mat]	‘circuit breaker’
<poupée>	[pu. pe]	<búp bê>	[bup. be]	‘puppet’
<cachet>	[ka.fɛ]	<cát sê>	[kat. se]	‘salary’
<containeur>	[kɔ̃.t(ə). nœʁ]	<công-ten-nơ>	[koŋm.tɛn. ny]	‘container’
<saucisse>	[so.sis]	<xúc-xích>	[suk .sik]	‘sausage’

The insights we gained considering the loanword data are noteworthy, but the total number of occurrences is not very high. With our experimental data, the lines of the loan results are reproduced with more data.

4.3 The three types of syllabification patterns in the experimental data (study 2)

In this section, we analyze the experimental data and compare it to the loan data presented previously. We start with a general difference, summarize the parallels, then move on with frequency distributions of the experimental data and finally provide additional evidence from related phenomena which only occur in the experimental settings. Generally, there is more variation in the experimental data than in the loan data, not only in terms of syllabification. Therefore, some patterns which are systematic in the loan data, are not in the experimental settings: while only certain syllable structures of the source lexeme enable specific syllabification process types, one and the same experimental stimulus has the potential to enable different types of syllabification processes. This provides more evidence for our assumption that the driving force behind all the three types, namely syllable boundary shift, featural and complete gemination, is the same. Five exemplary stimulus lexemes with several responses for each are provided in (6). On the right-hand side, we put the syllabification type, and the expected type for that stimulus lexeme is marked in bold.

(6) Examples for syllable boundary shift, complete and featural gemination

<i>écran</i>	[e.kʁã]	<ích khăng> <ít khom> <ích cợ>	[ik.xãŋ] [it.xəm] [ik.kɔ]	‘screen’	syllable boundary shift beyond syllable b. shift complete gemination
<i>étang</i>	[e.tã]	<ít tong> <ích tong>	[it.təŋm̃] [ik.təŋm̃]	‘pond’	complete gemination featural gemination
<i>chameau</i>	[ʃa.mo]	<săm mô> <săm ngu>	[săm.mo] <san mù> [săm.ŋu]	‘camel’	complete gemination featural gemination featural gemination
<i>épice</i>	[e.pis]	<íp pi> <íp bì> <íp phi> <ích bi>	[ip.pi] [ip.bi] [ip.fi] [ik.bi]	‘spice’	complete gemination featural gemination featural gemination beyond featural gem.
<i>coussin</i>	[ku.sɛ̃]	<cút xăng> <cúp sò>	[kut.sãŋ] [kup.sɜ]	‘cushion’	featural gemination featural gemination

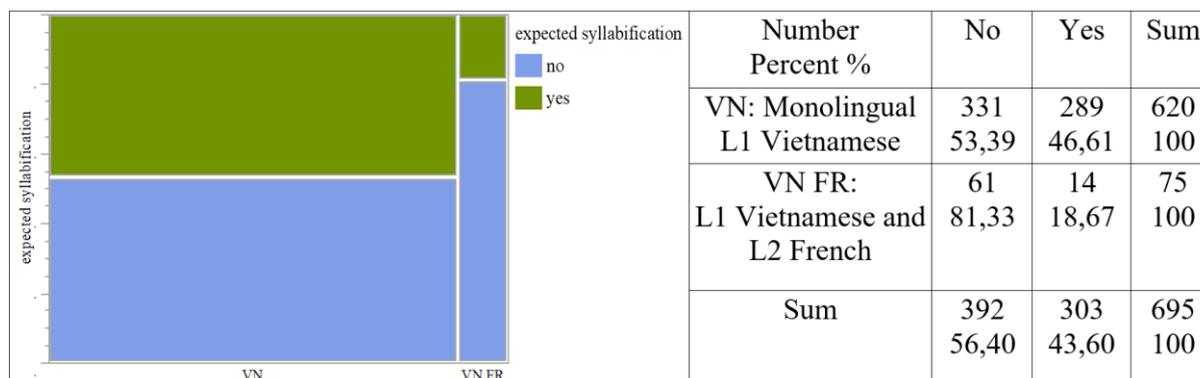
Starting with the stimulus lexeme *écran*, we exemplify several options with three responses. The expected option is syllable boundary shift as in the first response example. Additionally, the consonant [ʁ] is replaced by its voiceless counterpart [x], because [ʁ] is an illicit sound in Vietnamese. The second example goes beyond the expected pattern of syllable boundary shift, because both sounds are replaced, even though the first would not need to be replaced as [k] as it is a licit Vietnamese coda consonant. As to the third response, there is some room for interpretation, but the most convincing one is that the second consonant of the French stimulus onset cluster [kʁ] has not been perceived and therefore deleted. Additionally, the first consonant is geminated and the double becomes the coda of the previous syllable.

We have illustrated on the example of the French stimulus *écran* that not only the expected type syllable boundary shift, but also gemination can emerge in the responses. As to the rest of the exemplary stimuli, we either expect complete or featural gemination, but also the reverse can be found in our response data. For instance, we expect complete gemination for the stimulus *chameau* with the sound [m], because the double is a licit Vietnamese coda. But again, featural gemination surfaces in several cases, and either the double changes features as in [san.mu], or the base as in [săm.ŋu]. This could be interpreted as an enhancement of the process itself: Even though it would not be necessary to geminate in order to make the lexeme a licit Vietnamese one, it is not only geminated, but one of the sounds is additionally changed so that they are different from each other and so that it is unambiguously clear that we have two consonants instead of just one. The unexpected response [săm.ŋu] shows a parallel to the expected featural gemination pattern occurring to the French stimulus *épice*: the response is [ip.bi], and it is an expected Vietnamization pattern because the onset [p] is not a native Vietnamese onset. This response in turn replicates the pattern analysed in (4) for the loan *búp bê* from French *poupée*.

After having given examples, we now move on with frequency distributions for the experimental data. We excluded missing and unclear data (due to illegibility of the writing of participants) from our analysis: there are 24 occurrences of missing data (3,33%) and 2 occurrences of unclear data (0,28%). Table 3 shows results for the speaker group as an independent variable with the values monolingual Vietnamese (VN) and Vietnamese as first and French as second language (VN FR). The independent variable has the values *yes* and *no*, which means that the expected syllabification type can be attested or not. While the amount of data for the group of monolingual Vietnamese speakers is large enough to gather a statistical analysis, the amount of data for the second speaker group is too low, which is why running a chi² test for the two independent variable values is problematic in terms of statistical significance. Still, the Pearson chi² test gives out the number 25,081, and the probability is below 0,0001 which indicates that the distribution is not random. Looking at the numbers in Graph 1 and Table 3, we can see that the percentage of the attested syllabification patterns is quite high for the monolingual

speakers (46,61%) and lower for the speakers with French as a second language but still at 18,67%, which is higher than in the loan data (12%). In total, with a percentage of 43,60, the frequency is much higher than for loans, but again, the low amount of loan data would make the significance of a statistical analysis in a comparison between experimental and loan data problematic.

Graph 1 and Table 3: Mosaic chart and contingency table for distributions of expected syllabification patterns



In addition to the occurrences in expected environments, similar patterns also replicate in unexpected environments with other stimulus lexemes than the ones we considered as targets for studying syllabification, coming from the loan data. The unexpected responses fall into two categories: On the one hand, we are dealing with the same syllabification patterns as in gemination in the context of epenthesis, and on the other hand, we are dealing with monosyllabic stimuli with an open syllable which receive a coda consonant in the response.

Let us start with the first category: there are two processes. First, there is epenthesis, which is usually triggered by the presence of a consonant in the coda of a syllable in the stimulus lexeme which is illicit in Vietnamese. This illicit structure is broken up by the addition of an epenthetic vowel, so that the consonant now functions as the onset of the epenthetic syllable. This form of epenthesis is a common strategy in our experimental data while it is not attested in the loan data. Another possible explanation for the emergence of epenthesis could also be the perception of the released French coda consonants as a vowel, because Vietnamese occlusives in coda position are not released and the phonetic release could be interpreted as a vowel. But in addition to epenthesis, a syllabification process with complete or featural gemination can occur. In this case, the coda consonant in the stimulus syllable not only becomes the onset of the epenthetic syllable, but it is also doubled and replaced by a licit coda consonant in the coda, if illicit. Strikingly, we do not only find featural gemination but also complete gemination, and here we would not need neither epenthesis nor gemination to have the result of licit Vietnamese syllable structures. Noteworthy is also the fact that the number and percentage of these occurrences is higher in the speaker group with French knowledge (67 in total, which makes 55,37%) than for the monolinguals (54 in total and 44,63%). It is also interesting that in the corresponding oral data we do not concentrate on here, they often do not make use of epenthesis nor consequently of gemination, but do so when they are asked to write down the word according to Vietnamese orthography as if the word had become a Vietnamese word. This hints to the possibility that the emergence of syllabification patterns goes beyond perception or production difficulties. Examples are shown in (7), the last three examples occur within the data of native Vietnamese speakers with second language knowledge of French.

(7) Featural and complete gemination in epenthetic contexts (121 occurrences in total)

<i>crabe</i>	[kʁab]	<kháp bò>	[xăp. ɓ ɾ]	‘crab’	featural gemination
		<khóp pu>	[xɔp. p u]		complete gemination
<i>piaf</i>	[pjaf]	<pi áp phò>	[pi.ap. f ɾ]	‘sparrow’	featural gemination
<i>guignol</i>	[gijɔl]	<ghi nhôn lò>	[ɣi.jɔn. l ɾ]	‘joker’	featural gemination
<i>menotte</i>	[mənɔt]	<mờ nót tồ>	[mɾ.nɔt. t ɾ]	‘handcuffs’	complete gemination

We now move on to the other category of unexpected patterns. These patterns do not involve a syllabification process where material from one syllable switches to the other. The parallel to syllable boundary shift, complete and featural gemination is that an open syllable in the stimulus is closed in the response, even if the added coda consonant cannot be taken from other syllables in the surroundings. Examples can be found in (8).

(8) Open syllables get closed even without syllabification processes (148 occurrences in total)

					added consonant
<i>blé</i>	[ble]	<nét>	[net]	‘wheat’	[t]
		<bờ lít>	[bɔ.lit]		[t]
		<lim>	[lim]		[m]
<i>frai</i>	[fɾɛ]	<khách>	[xɛk]	‘frog egg’	[k]
		<khét>	[xɛt]		[t]
		<phép>	[fɛp]		[p]
		<phét>	[fɛt]		[t]
<i>pas</i>	[pa]	<páp>	[pap]	‘step’	[p]
		<báp>	[bap]		[p]
		<bợn>	[bɔn]		[n]
		<phặc>	[fak]		[k]
		<phạm>	[fam]		[m]
<i>bambou</i>	[bã.bu]	<bò bút>	[bɔ.but]	‘bamboo’	[t]
		<bòm bum>	[bɔm.bum]		[m]
		<con bún>	[kɔn.bun]		[n]
		<òm púp>	[ɔm.pup]		[p]
<i>boyau</i>	[bwa.jo]	<bò dúc>	[bɔ.zuk]	‘tubeless tire’	[k]
		<bờ giúp>	[bɾ.zup]		[p]
		<bờ rút>	[bɾ.zut]		[t]
		<buôi rúp>	[bwoj.zup]		[p]
<i>crapaud</i>	[kʁa.po]	<khấp phúp>	[xap.fup]	‘crab’	[p]
		<khọc phúp>	[xɔk.fup]		[p]
		<khập phút>	[xɿp.fut]		[t]

Among these examples, we find mono- and disyllabic stimulus lexemes and their responses. What unites them is that the syllable to which the process occurs is not followed by another syllable and the material of the added coda consonants cannot be taken from there. The last four responses including the last response to *boyau* and all responses to *crapaud* are special: we can attest featural gemination processes to close the first syllable and, in addition to that, the second syllable is also closed, but without being the subject of any syllabification process. The distribution between the two groups is different:

The pattern *hardy* occurs for the speakers with French knowledge (3 occurrences, 2.03%) and mostly for the monolinguals (145 occurrences in total, 97.97%).

5 Discussion

So far, we have presented three types of syllabification patterns occurring in two language contact settings between Vietnamese and French. As to syllable boundary shift, we had mentioned that it might look at first sight like a strategy to avoid consonant clusters which are illicit in Vietnamese. But having considered also complete and featural gemination, we gained initial evidence for our assumption that all the processes are similar and are enabled by the same mechanisms. Further evidence is delivered by the fact that in the experimental data, the three types of syllabification patterns can be found in different responses to one and the same stimulus lexeme: While the three types systematically occur in different environments in the loan data, they seem to merge into a single type with different manifestations in the experimental data, where generally more variability occurs.

We therefore suggest unified hypotheses which serve as explanations for syllable boundary shift, complete and featural gemination in both the loanword data and experimental settings. The most important parallel between the so far described processes is the pattern that we have an open first syllable in the disyllabic source or the stimulus lexeme and a closed one in the loan or the experimental response. The material of the added coda is taken from the onset of the second syllable. In this discussion, we deliver two explanations which are not mutually exclusive but go hand in hand and help us to get a better understanding of language contact mechanisms as well as native phonology.

The first explanation is phonetic perception. As to syllable boundary shift, speakers of Vietnamese might not even feel the urge to dissolve an illicit intrasyllabic consonant cluster because they do not perceive it as such, grouping the string of sounds according to the Vietnamese syllabification logic in perception and therefore differently from French syllabification rules. As to featural gemination, the same speakers might tend to perceive an illusory consonant due to phonetic cues. These might be the same for both syllable boundary shift and both types of gemination and have to do with our assumption that an average French vowel rather corresponds to a short Vietnamese vowel than to a long one, and therefore tends to be mapped as such in this language contact setting. This enables the so far discussed syllabification processes especially if otherwise, the syllable with only one mora would be incomplete according to the Vietnamese syllable structure requirements. Additionally, the perception of French syllable boundaries can be difficult for speakers of Vietnamese because French syllables might phonetically be less sharply separated from each other than Vietnamese ones. In line with this assumption goes the lower stability of the unit syllable in French compared to Vietnamese: A French syllable can be overridden by phonological processes while a Vietnamese syllable remains how it is. This may have to do with the fact that the Vietnamese syllable has the functionality of a tone bearing unit while the French syllable does not.

To get more evidence for our assumption on phonetic perception, phonetic measurements of average French vowel length compared to Vietnamese long and short vowels need to be done in the future. But phonetic perception must not be the cause or least not the only cause for enabling the analysed processes. A hint is that speakers with knowledge of French, who may have a good access to the French vowel system, reproduce the described patterns in the experimental setting as well. Therefore, we continue by presenting the *closed syllable template hypothesis*:

Speakers of Vietnamese have the tendency to adjust French material to their system such that French syllables which are open have the potential to be closed.

One piece of evidence for this is that the frequency of closed syllables is higher in Vietnamese than the frequency of open syllables in the native lexicon. This frequency in combination with the tendency of Vietnamese speakers to close open syllables in language contact leads us to the assumption that an optimal Vietnamese syllable is closed. In terms of distributions, the source or stimulus material is therefore vietnamized. More evidence for this hypothesis is delivered by the emergence of patterns in unexpected syllabic surroundings presented in (7) and (8), which partly go beyond syllabification, but also result in open stimulus syllables to be closed in responses.

These occurrences cannot be explained by linguistic economy and also call into question whether a reason for the occurrence of featural gemination can only be explained by phonetic perception. A parallel of both the phonetic perception explanation as well as the closed syllable template hypothesis is that they go in line with the fact that in native Vietnamese, well-formed syllables fulfil a dimoraic requirement: A syllable containing a short vowel must be closed, no matter if this syllable is prominent or not. Probably, this is the extreme of a universal tendency, which is not present for instance in many Slavic languages, but in Cantonese and Thai as well, and also has a role to play in German, several southern French varieties, and Limousin (cf. Meinschaefer, 2020). Still, French vowels can be shorter than Vietnamese vowels even when they stand in open syllables. This is so because in contrary to Vietnamese, there is no requirement of dimoraicity in French. Noteworthy is the fact that according to our data, not all open syllables which get closed have a short vowel as a nucleus in the loan or the experimental response: even when vowels are long, they can be closed, and the attested patterns do not only emerge when a dimoraic requirement needs to be fulfilled. Therefore, it cannot be the explanation for all the data. An example from the loan data is *cát sê* [kat.se] from French *cachet* [ka.ʃɛ] ‘salary, where the [a] is not short and still featural gemination takes place, and a similar example from the experimental data is the response *san mù* [san.mu] to the French stimulus *chameau* [ʃa.mo].

These examples are not an exception, and they provide stronger evidence for our closed syllable template hypothesis: it goes in line but beyond the dimoraic requirement, and while the latter is a hard constraint, the requirement to have closed syllables is a lower ranked and soft constraint. Postulating such a constraint goes against the universal tendency to prefer open syllables over closed ones and the oftentimes high-ranked constraint to have no coda in a syllable. Therefore, the possibility of such a constraint and its role and background should be explored further in native and language contact settings and for other languages and language pairs.

Because of the already mentioned parallels to Cantonese and Thai, we now deepen the debate having a look at similar language contact settings for these languages to see if similar patterns can be found here. Actually, processes of gemination are described for the adaptation of English loans into Cantonese by Silverman (1992). An example is [kəp.pi] from English *copy*, and by chance, we can directly compare it to the French loan in Vietnamese *cóp py* [kəp.pi]. As for Vietnamese, the process of gemination has not been described as a native process so far.

Silvermann’s analysis stops with complete gemination but he does not mention the existence of similar patterns as we describe under featural gemination. Instead, he points out that the process of gemination is blocked when the consonant which could be the subject of gemination is not possible in coda position. To the contrary, in Thai, both complete and featural gemination emerge, just that the patterns of featural gemination have not been labelled like that before but instead described as *homorganic patterns* (Rungruang, 2008).

It is quite striking that the patterns found in language contact between Thai and English are nearly identical to the patterns we find for contact situations between Vietnamese and French. Take the Thai loan [kʰlaat.sik] from English *classic* [klæ.sik], where the consonant slot of [s] doubles but without a conservation of all its features in the double. The fricative becomes a plosive and the coda consonant in the first syllable of the loan is [t]. A corresponding French loan with exactly the same pattern is *ét xăng* [ɛt.sãŋ] from French *essence* [e.sãs]. For the experimental setting, we can take the response [kut.sãŋ] which corresponds to the French stimulus *coussin* [ku.sɛ] ‘cushion’ (this and more examples for Thai with striking similarities to our data can be taken from Rungruang 2008:112).

But the parallels do not stop at complete and featural gemination. Syllable boundary shift emerges in the English-Thai borrowing setting as well, and it patterns identically as in the French-Vietnamese borrowing and experimental settings. Occasionally, we even have the chance to present these striking parallels for the example of English *taxi* [tæ.ksi], borrowed as [tʰɛk.sii] into Thai (example taken from Rungruang 2008:113) to the French lexeme *taxi* [ta.ksi], borrowed into Vietnamese as [tãk.si]. It would be interesting to investigate experimentally for these languages as well if we could find evidence for the closed syllable template also for them.

So far, we have studied language contact between Vietnamese and French and found parallels to other language contact settings between languages of European and East-Asian origin. It is striking that even though the three mentioned Asian languages Vietnamese, Cantonese and Thai do not belong to

the same language family, they pattern so similarly in prosodic language contact with French or English. This might be the case because language contact between these Asian languages has prosodically shaped them before entering into language contact with the mentioned European languages.

This opens a window into how multi-layered language contact scenarios are. The suggestion we take home from this comparison is that language contact between several very distinct East-Asian languages may have led to the fact that they pattern similarly in contact with European languages, which, in themselves, are also quite distinct, but also have a lot in common prosodically. It seems promising to go deeper into typological comparisons of different language pairs in contact to complete the picture of prosodical language contact and syllabification phenomenon such as complete and featural gemination from a typological stance.

We finally want to mention that gemination is a phenomenon which occurs not only in many languages as a native process, but has also been studied for loanword adaptation processes in the past. The studied languages are also quite diverse compared to each other. To mention just some, take into account studies about Japanese (cf. Kawahara 2015; Ito, Kubozono, & Mester 2017; and Tanaka 2017 and many more), Italian (Repetti 2009) and Hungarian (Kertész 2006). The difference between these languages as Vietnamese is that gemination is a common process in native phonology and has also been attested as such by many researchers. This does not seem to be the case for Vietnamese, nor did we find that it is so for Cantonese or Thai. And still, these patterns emerge, even if not systematically, then at least quite robustly in different language contact scenarios.

6 Conclusion

In this paper, we have focussed on two prosodic language contact settings, and considered the three syllabification processes: syllable boundary shift, complete and featural gemination. The results derived from the study of loanwords were replicated in experimental settings with monolingual speakers of Vietnamese and speakers who have acquired French as adults and live in France, which strengthens evidence the existence of these processes and patterns. Further evidence could be gained from the comparison to prosodic language contact scenarios with different languages involved.

While language contact such as borrowing processes in loanword adaptation as well as second language acquisition are a much-studied topic, the only syllabification process which has been paid much attention to is complete gemination. Inspired by a communication with Rachid Ridouane in early 2020, we present and shape in this paper the term *featural gemination*, describing another instance of incomplete gemination due to native phonotactic restrictions, and also find common sources for both types of gemination and syllable boundary shift. We came up with explanations as potential reasons for the emergence of such patterns. One of them is the phonetic perception of French vowels as short and linking them to short vowels according to the Vietnamese phonological system. The second explanation is what we call the *closed syllable template hypothesis*.

Both explanations are linked to the idea of a dimoraic minimum in Vietnamese and deliver more evidence for its existence: A coda consonant is added to a syllable which would otherwise be incomplete because a mora is lacking. While this is a hard constraint, we also consider the potential soft constraint that an optimal Vietnamese syllable is closed, and that while a monomoraic open syllable is not possible, a dimoraic closed syllable is better than a dimoraic open one. This is a quite unexpected constraint according to universal tendencies as it also outranks the commonly powerful and oftentimes quite high-ranked constraint that an optimal syllable is open. This paper shows again how fruitful the study of language contact can be in order to better understand native mechanisms. Finally, we suggest that it is time to definitively step back from the discussion if borrowing processes are phonetic or phonological and also to which extent. As phonetics and phonology is always linked in borrowing processes as well as in second language acquisition, it makes more sense to try to understand processes as complex as they are and give room for graduality instead of binary distinctions.

Acknowledgement

I would like to thank the audience at ICAAL 8 in Chiang Mai, two anonymous reviewers, Andrea Pešková, and Carolin Ulmer. I am grateful to Nguyễn Minh Châu, Nguyễn Thị Thu Hương, the

experiment participants, and everyone else who supported me with this research, including researchers at Trường Đại Học Khoa Học Thái Nguyên and LACITO Villejuif for their warm welcome during my field research. Remaining errors are my own.

7 Bibliography

- Barker, Milton E. 1969. The Phonological Adaptation of French Loanwords in Vietnamese. *Mon-Khmer Studies Journal* 3:138–47.
- Boersma, Paul and Silke Hamman. 2009. Loanword Adaptation as First-Language Phonological Perception. Pp. 11–58 in *Loan Phonology*, edited by A. Calabrese and L. Wetzels. Amsterdam: John Benjamins Publishing Company.
- Broselow, Ellen. 2009. Stress Adaptation in Loanword Phonology: Perception and Learnability. *Phonology in Perception* 191–234.
- Brunelle, Marc. 2017. Stress and Phrasal Prominence in Tone Languages: The Case of Southern Vietnamese. *Journal of the International Phonetic Association* 47.3:283–320.
- Brunelle, Marc. 2015. Vietnamese (Tiếng Việt). Pp. 909–54 in *The Handbook of Austroasiatic Languages. Volume 2*, edited by M. Jenny and P. Sidwell. Leiden, Boston: Brill.
- Bùi, Khắc Việt, Bích Thu Chu, Thân Đào, Hoàng Tuệ, Hoàng Văn Hành, Kim Chi Lê, Thanh Nga Nguyễn, Thuý Khanh Nguyễn, Văn Khang Nguyễn, Hùng Việt Phạm, Cẩm Vân Trần, Nghĩa Phương Trần, Vũ Ngọc Bảo, and Vương Lộc. 2003. *Từ Điển Tiếng Việt. Viện Ngôn Ngữ Học [Vietnamese Dictionary. Institute of Linguistics]*. Hà Nội - Đà Nẵng: Nhà Xuất Bản Đà Nẵng.
- Cheung, Kwan Hin. 1986. *The Phonology of Present-Day Cantonese*. University College London.
- Delattre, Pierre. 1971. Consonant Gemination in Four Languages: An Acoustic, Perceptual, and Radiographic Study (Part I). *IRAL: International Review of Applied Linguistics in Language Teaching* 9(1):31–52.
- Đoàn, Thiện Thuật, Khánh Hà Nguyễn, Như Quỳnh Phạm, Thiện Thuật Đoàn, Khánh Hà Nguyễn, and Như Quỳnh Phạm. 2009. *A Concise Vietnamese Grammar (for Non-Native Speakers)*. 3rd ed. Hà Nội: Thế Giới Publishers.
- Hạo, Cao Xuân. 2017. *Tiếng Việt. Mấy Vấn Đề Ngữ Âm, Ngữ Pháp, Ngữ Nghĩa [Vietnamese. Phonetic, Syntactic and Semantic Issues]*. Tái bản. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Hsieh, Feng Fan and Michael J. Kenstowicz. 2008. Phonetic Knowledge in Tonal Adaptation: Mandarin and English Loanwords in Lhasa Tibetan. *Journal of East Asian Linguistics* 17(4):279–97.
- Huynh, Sabine. 2010. *Les mécanismes d'intégration des mots d'emprunt Français en Vietnamien*. [Integration mechanisms for French loanwords into Vietnamese] Paris: Harmattan.
- Ito, Chiyuki. 2014. Loanword Accentuation in Yanbian Korean: A Weighted-Constraints Analysis. *Natural Language and Linguistic Theory* 32.2:537-592
- Ito, Junko, Haruo Kubozono, and Armin Mester. 2017. A Prosodic Account of Consonant Gemination in Japanese Loanwords. In *The Phonetics and Phonology of Geminate Consonants*, edited by H. Kubozono, 283-320. Oxford: Oxford University Press.
- De Jong, Kenneth and Mi-Hui Cho. 2012. Loanword Phonology and Perceptual Mapping: Comparing Two Corpora of Korean Contact with English. *Language* 88.2:341–68.
- Kang, Yoonjung. 2010. Tutorial Overview: Suprasegmental Adaptation in Loanwords. *Lingua* 120(9):2295–2310.
- Kang, Yoonjung. 2011. Loanword Phonology. in *The Blackwell companion to phonology*, edited by M. van Oostendorp, C. Ewen, E. Hume, and K. Rice, 2258–82. Malden MA & Oxford: Wiley-Blackwell.
- Kang, Yoonjung, Andrea Hoa Pham, and Benjamin Storme. 2016. French Loanwords in Vietnamese: The Role of Input Language Phonotactics and Contrast in Loanword Adaptation. in

- Proceedings of the 2014 Annual Meeting on Phonology*, 1–12. Edited by Adam Albright & Michelle A. Fullwood. Linguistic Society of America.
- Kawahara, Shigeto. 2015. Geminate Devoicing in Japanese Loanwords: Theoretical and Experimental Investigations. *Linguistics and Language Compass* 9.4:168–82.
- Kenstowicz, Michael. 2003. Review Article: The Role of Perception in Loanword Phonology. A Review of *Les Emprunts Linguistiques d'origine Européenne En Fon* by Flavien Gbéto, Köln: Rüdiger Köppe Verlag, 2000*. *Studies in African Linguistics* 32.1:96–112.
- Kenstowicz, Michael. 2005. The Phonetics and Phonology of Korean Loanword Adaptation. in *Proceedings of the First European Conference on Korean Linguistics*, 17–32.
- Kenstowicz, Michael. 2010. Loanword Phonology and Enhancement. in *Proceedings of the 2010 Seoul International Conference on Linguistics, Universal Grammar and Particular Languages*, 104–12. Seoul. Linguistic Society of Korea.
- Kenstowicz, Michael and Atiwong Suchato. 2006. Issues in Loanword Adaptation: A Case Study from Thai. *Lingua* 116.7:921–49.
- Kertész, Zsuzsa. (2006). Approaches to the phonological analysis of loanword adaptation. *The Even Yearbook* 7:1–15.
- Kim, Kyumin and Alexei Kochetov. 2011. Phonology and Phonetics of Epenthetic Vowels in Loanwords: Experimental Evidence from Korean. *Lingua* 121.3:511–32.
- Kirby, James. 2008. *VPhon: A Vietnamese Phonetizer* (version 0.2.6).
- Kirby, James. 2011. Illustration of the IPA. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41.3:381–92.
- LaCharité, Darlene and Carole Paradis. 2005. Category Preservation and Proximity versus Phonetic Approximation in Loanword Adaptation. *Linguistic Inquiry* 36.2:223–58.
- Le, Viet Bac and Laurent Besacier. 2009. Automatic Speech Recognition for Under-Resourced Languages: Application to Vietnamese Language. *IEEE Transactions on Audio, Speech and Language Processing* 17.8:1471–82.
- Meinschaefer, Judith. 2020. “Lange Vokale im Limousinischen.” 635–50 in *Contact, variation and change in Romance and beyond / Kontakt, Variation und Wandel in und jenseits der Romania. Studies in honor of Trudel Meisenburg / Festschrift für Trudel Meisenburg*, edited by C. Gabriel, A. Pešková, and M. Selig. Berlin: Erich Schmidt Verlag.
- Michaud, Alexis, Minh Châu Nguyễn, and Vera Scholvin. 2021. “L’intonation Dans Les Langues Tonales : Des Réflexions Générales et Deux Études de Cas.” *Études de Linguistique Appliquée (ELA)* 199.1:399-429.
- Nguyễn, Huỳnh Trang, & Dutta, Hermanga. (2017). The Adaptation of French Consonant Clusters in Vietnamese Phonology : *Journal of Universal Language* 18.1:69–103.
- Nguyễn, Quảng Tuân and Đức Dân Nguyễn. 1992. *Từ Điển Các Từ Tiếng Việt Gốc Pháp [Dictionary of Vietnamese Words of French Origin]*. Hồ Chí Minh City: Hội nghiên cứu và giảng dạy văn học Thành phố Hồ Chí Minh.
- Nguyễn, Văn Khang. 2013. *Từ Ngoại Lai Trong Tiếng Việt [Foreign Words in Vietnamese]*. Thành Phố Hồ Chí Minh: Nhà xuất bản Tổng hợp Thành phố Hồ Chí Minh.
- Paradis, Carole and Darlene Lacharité. 1997. Preservation and Minimality in Loanword Adaptation. *Journal of Linguistics* 33.2:379–430.
- Paradis, Carole and Jean-François Prunet. 2000. Nasal Vowels as Two Segments : Evidence from Borrowings. *Linguistic Society of America* 76/2:324–57.
- Peperkamp, Sharon and Emmanuel Dupoux. 2003. Reinterpreting Loanword Adaptations: The Role of Perception. in *Proceedings of the 15th International Congress of Phonetic Sciences*, 367–70.
- Pham, Andrea Hoa. 2009. The Identity of Non-Identified Sounds: Glottal Stop, Prevocalic /w/ and Triphthongs in Vietnamese. *Toronto Working Papers in Linguistics*. Vol. 34. Online: LGCU.

- Poplack, S. 2018. *Borrowing. Loanwords in the Speech Community and in the Grammar*. Oxford: Oxford University Press.
- Repetti, Lori. 2009. Gemination in English Loans in American Varieties of Italian. *Loan Phonology: Current Issues in Linguistic Theory* 307:225–40.
- Rey-Debove, J. and A. Rey, eds. 2013. *Le Nouveau Petit Robert. Dictionnaire Alphabétique et Analogique de La Langue Française. Version Électronique*.
- Rose, Yvan and Katherine Demuth. 2006. Vowel Epenthesis in Loanword Adaptation: Representational and Phonetic Considerations. *Lingua* 116.7:1112–39.
- Rungruang, Apichai. 2008. “English Loanwords in Thai and Optimality Theory.” ProQuestLLC, Ball State University, Muncie, Indiana.
- Scholvin, Vera and Judith Meinschaefer. 2018. The Integration of French Loanwords into Vietnamese: A Corpus-Based Analysis of Tonal, Syllabic and Segmental Aspects. in *Papers from the 7th International Conference on Austroasiatic Linguistics, JSEALS Special Publication No 3*, edited by H. Ring and F. Rau, 157–73. Honolulu: University of Hawai‘i Press.
- Scholvin, Vera. Forthcoming. Prosody in Language Contact: French and Vietnamese. PhD thesis submitted 2021 and defended 2022.
- Shinohara, Shigeko. 2004. Emergence of Universal Grammar in Foreign Word Adaptations. Pp. 292–320 in *Constraints in Phonological Acquisition*, edited by R. Kager, J. Pater, and W. Zonneveld. Cambridge: Cambridge University Press.
- Shinohara, Shigeko. 2006. Perceptual Effects in Final Cluster Reduction Patterns. *Lingua* 116.7:1046–78.
- Silverman, Daniel. 1992. Multiple Scansions in Loanword Phonology: Evidence from Cantonese. *Phonology* 9.2:289–328.
- Tanaka, Shin’ichi. 2017. The Relation between L2 Perception and L1 Phonology in Japanese Loanwords. An Analysis of Geminate in Loanwords from Italian. in *The Phonetics and Phonology of Geminate Consonants*, edited by H. Kubozono. Oxford: Oxford University Press.
- Trần, Thị Thúy Hiền. 2011. *Processus d’acquisition des clusters et autres séquences de consonnes en langue seconde : De l’analyse acoustico-perceptive des séquences consonantiques du vietnamien à l’analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE*. Université Stendhal - Grenoble III.
- Trần, Thi Thuy Hien, Nathalie Vallée, and Silvain Gerber. 2016. Syllabe CVC et Cycle Mandibulaire : Une étude articulatoire des ssymétries. Le cas du Vietnamien. in *Journées d’Etudes sur la Parole (JEP-TALN-RECITAL 2016)*, 705–13. Paris.
- Tranel, Bernard. 1987. *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press.
- Vương, Toàn. 1992. *Từ gốc Pháp trong tiếng Việt [Words of French origin in Vietnamese]*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Walker, Douglas C. 2001. *French Sound Structure*. Calgary: University of Calgary Press.
- Yip, Moira. 1993. Cantonese Loanword Phonology and Optimality Theory. *Journal of East Asian Linguistics* 2.3:261–91.
- Yip, Moira. 2006. The Symbiosis between Perception and Grammar in Loanword Phonology. *Lingua* 116.7:950–75.

PROTO-PEARIC AND THE ROLE OF VOWEL HEIGHT IN CREAKY REGISTER¹

Paul SIDWELL
University of Sydney
<paulsidwell AT gmail DOT com>

Abstract

The Pearic branch of the Austroasiatic phylum evidences 4 distinct voice quality settings functioning phonologically as four registers: clear, creaky, breathy, and breathy-creaky. While phonetic studies have established the nature of the registers, language shift by speakers is so extensive that it is difficult in practice to investigate the etymology of the phenomenon even with the best published sources. While comparative-historical reconstruction has already explained the origin of breathy voice in Pearic, the origin of creak is obscure, and it has been suggested as a primitive feature of Austroasiatic. Previous studies have noted some correlations may have a bearing on the problem: this includes non-occurrence of creak in syllables that have glottal onsets, and low frequency of creak in syllables with high vowels. It is proposed in this paper that in the history of Pearic, syllables with [-high] nuclei generally became creaky, subject to phonological restrictions. Additionally, some exceptions arose lexically due to mimicry of prestigious Khmer or other non-native forms. Additional proposals are made which refine the proto-Pearic reconstruction of Headley (1985), anticipating a much refined and expanded reconstruction of Proto-Pearic in due course.

Keywords: creaky phonation, comparative reconstruction, historical phonology
ISO 639-3 language codes: pcb, cpg, scq, sxm, smu, syo

1 Introduction

The Pearic branch of Austroasiatic (AA) is a small, highly endangered group of languages spoken mainly in western Cambodia, the Trat Province of Thailand, and pockets near Siem Reap and in Kompong Som and Kompong Thom. The group is known in the linguistic literature as *Pearic*, based on the Khmer forms *po:(r)* and *pəə(r)*, both from Sanskrit *varṇa* ‘colour, caste’, according to Ferlus (2011). This may be considered a derogatory and undesirable term to use to label the languages, but it has currency in the technical literature, so I use it here. The name Chong [tʰɔːŋ] and its variants (Song, Kasong etc.) appear to reflect the original autonym, reported as *Tchouang* in Tcheou Ta-Kouan’s 13th century description of Cambodia (Pelliot 1902:156 / 1951:70, cited by Ferlus 2011). The name also bears a suggestive resemblance that of the Zhuang (Tai) of Southern China and the Juang (Munda) of India, but that is not explored further here. Other lect names include Samre, Somrey being infixed forms of Khmer ស្រែ *srae* ‘wet field’, Suoy, and Sa’och/Sauch/Ut (see Ferlus 2011 for further discussion on Pearic lect names).

The population of speakers is perhaps only around 10,000, and there is little typological or lexical diversity within the branch. Both Baradat (1941) and Choosi (2007) report high intelligibility across all varieties except with Pear of Kompong Thom (KPT henceforth), and lexicostatistical studies (such as Martin 1974a) indicate that the lects other than KPT share a high percentage of cognate vocabulary consistent with being identified as lects of one language (between 77% and 98% of basic vocabulary). This suggests that from a comparative linguistic perspective the non-KPT lects are a single language in a mosaic of lects, and for the purposes of this study these will be referred to as the Chongic lects. Martin’s

¹ I would like to thank Ryan Gehrman for useful advice in preparation of this paper. Also I acknowledge with gratitude financial support I received from the Australian Research Council under Future Fellowship award FT120100241, and assistance from the Max Planck Institute for the Science of Human History (Jena), which supported work on this project, including conference travel.

study counted KPT sharing between 56% and 63% of basic vocabulary with the Chongic lects, indicating that the overall time-depth of Pearic is rather shallow, and we can suggest that historical unity of the group goes back to the Old Khmer period (1st millennium CE). It is striking that two other AA branches spoken in the immediate region, Khmer and Mon, are also not very internally diverse, and resolve to single languages in the first Millenium. Furthermore, Pearic, Mon, and Khmer, each shows no obvious close relations with each other or any another AA branches.² These characteristics probably correlate with the adoption of lowland rice farming and state formation, while upland swiddeners speaking other AA languages continued diversifying over time. This means that there were thousands of years of development between proto-AA and the reconstructable proto-Pearic and we must assume plenty of opportunity for linguistic changes to accumulate between these two stages of the language. The immediate practical effect of this is that proto-Pearic is expected to be structurally more similar to modern Pearic lects than to proto-AA or other AA branches.

Map 1: locations of Pearic languages (fragment of Mainland SE Asian language map compiled by David Bradley, in Wurm & Hattori's Language Atlas of the Pacific Area (1983).



² At the 2012 SEALS meeting (Argay, France) Gerard Diffloth proposed a handful of lexical isoglosses between Pearic and Khasi, and a shared sound change (loss of medial *h) with Khmu, suggesting a Northern-Austroasiatic affiliation, but this has not been elaborated in print nor been further developed by other scholars. Lexicostatistical analyses consistently favour grouping Pearic and Khmer (see Sidwell 2009 for a review) but those results arguably reflect effects of their close proximity.

Headley (1985) presented a preliminary reconstruction of proto-Pearic, relying on the manuscript vocabulary of Baradat, works by Martin (1974a,b), the Chong lexicon of Huffman, and various colonial era lexicons consolidated by Headley (1977, 1978). The 1985 study proposed vowel and consonant inventories, a model of syllable structure, and offered 148 proto-forms, including a set of distinctive Pearic lexical innovations replacing items otherwise well preserved in other AA languages. Most of the proto-lexemes proposed have good AA etymologies, and this fact provides confidence that many aspects of the proto-phonology were properly worked out, which is indeed borne out in more recent discussions such as Ostapirat (2009), Ferlus (2011), Sidwell & Rau (2014) which refine the reconstruction somewhat without extending the proto-lexicon. The reconstruction reveals that the Pearic branch has typical mainland Southeast Asian segments and phonological structures, and no essentially unique features in the AA context.

Notwithstanding the constructive contribution of Headley’s study, it had an essential shortcoming in that the question of breathy and creaky registers was not addressed, largely due to the lack of adequate phonetic detail in many of the sources and the modest number of etyma considered. This was an understandable oversight at the time: the origins of breathy phonation were already known and had been discussed at the 1973 ICAAL meeting (e.g., Huffman 1976) and consequently it was reasonably expected that breathiness would fall out of the proto-language with the reconstruction of a voiced/voiceless distinction among proto-Pearic onsets. At the same time, while creak was known as a phenomenon in AA languages, it was not well documented or understood, and had not yet attracted much attention.³

The Pearic registers are these days quite well documented. The most useful descriptions of the register phenomenon are based on Chongic lects spoken in Thailand and nearby areas of Cambodia, and these works include Suphanphaiboon (1982), Huffman (1985), Luang-Thongkham (1991), Ungsitipoonporn (2001), Ploykaew (2001), Premsrirat (2008) Premsrirat & Rojanakul (2014). It is important to understand the articulatory and acoustic aspects of these registers in order to ground the historical-phonological explanations. Given the clearly very close relations between the Pearic lects, and the consistencies between the various descriptions in respect of register in the works published since the 1980s, it is assumed for our purposes that the various close phonetic descriptions available to us are reasonably correct and representative of the branch.

The sources describe a maximal four-way register contrast in Chong, as diagrammed in Table 1. Note that the register numbering convention is as adopted in many works and has become standard.

Table 1: *Chong Registers*

high / tense series	R1 [CVC] clear	R2 [CV ^h C] clear-creaky
low / lax series	R3 [C ^h V] breathy	R4 [CV ^h C] breathy-creaky

Ferlus (2011) provides the following Chong register examples extracted from Ungsitipoonporn’s (2001) thesis:

Table 2: *Chong words illustrating register distinctions*

R1	ka:k ‘armpit’	t ^h a:k ‘span’	c ^h ɔ:k ‘pound’	kla:ŋ ‘branch’
R2	kəta:ʔk ‘tongue’	kəna:ʔk ‘crow’	kəsɔ:ʔŋ ‘ladder’	k ^h la:ʔŋ ‘owl’
R3	t ^h ə:k ‘water’	kəla:ŋ ‘ear’	c ^h ɔ:k ‘shine’	kəp ^h ɔ:k ‘hole’
R4	kəla:ʔk ‘swallow’	kəla:ʔŋ ‘sand’	c ^h ɔ:ʔŋ ‘Chong’	ŋa:ʔj ‘far’

³ A turning point was reached a little later, with Diffloth’s (1986) paper on AA creaky voice. That paper raised the issue of creak in Pearic, Katuic, and Vietic in comparative context but without solving its origin. The issue has remained a difficult one for AA comparative studies, for example it is striking that, while it is known that various Katuic language have creaky registers, none of the many Katuic reconstructions published to date pay regard to this aspect of the phonology.

Huffman (1985) discusses the phonetics of the registers based on his close aural-perceptual investigation. He notes that (p.361), “/p t c k/ are lax and slightly aspirated [...] before lax vowels.” However, this aspiration is different in character from the aspiration of the phonologically aspirated stops /p^h t^h c^h k^h/, which “... is much stronger than the breathiness which accompanies /p t c k/ before lax vowels.” Huffman’s analysis treats the creaky syllables as having final clusters of glottal stop+C, rather than being another dimension of register. This approach is particular to Huffman, other authors treat creak as a register component as it can be pronounced a little earlier in the syllable duration, creating something like the impression of splitting syllables in two. We see much of this in Baradat’s (1941) transcriptions (cf. Baradat’s KPT *sâaûch* ‘ant’ with Huffman’s *so:ʔc* ‘ant’). The variability in the timing of the glottal constriction, essentially anywhere in the second half of the syllable duration, clearly supports a supra-segmental analysis.

Huffman also noted that in the tense register following stops, there is (p.361), “...often conditioning a centered on-glide to a following vowel (especially before /i, ii, oo/)”. Luang-Thongkam (1991) investigated the register effects on vowel timbre and pitches using instrumental methods, being able to more precisely quantify the observations. She confirmed that the tense series are pronounced with higher pitch than the lax series, and graphs 1st and 2nd formants values for vowels in each register, finding timbre differences (p.147-7), “...clear voice and semi-clear voice vowels of the first register and second registers are more open than the breathy voice and semi-breathy voice vowels of the third and fourth registers.” This is consistent with Huffman’s observations such as /o:/ being pronounced [ao] in tense register.

An emphasis on pitch is found in Thongkam’s (2003) description of Kasong (a Chong lect spoken in Trat) as having contour tones combining with a clear/breathy phonation contrast. These map neatly to the Chong register scheme, although as Thongkam numbers the categories differently in the low register, her labels are given in brackets in the table 3:

Table 3: Kasong Registers / Tones

high / tense series	R1 (R1)	mid-level clear	R2 (R2)	high-rising-falling clear
low / lax series	R3 (R4)	low-falling breathy	R4 (R3)	high-rising-falling breathy

Thongkam does not report creak in Kasong, nor does she report glottal stop as a possible coda segment; this is understandable as creak/glottal stop condition a sharp drop and/or cessation of regular glottal fold movement heard as rapid pitch drop in terms of contour.

Priority is given to pitch differences in the Samre phonology of Ploykaew’s (2001). She characterises Samre in terms of three contour tones labelled A, B, C, and these map to the Chong scheme as shown in Table 4.

Table 4: Samre Registers / Tones

high / tense series	R1 / A	332 open syllables 334 long stopped syllables 334 short stopped syllables	R2 / C	451 open syllables
	R1 / C	451 open syllables		
low / lax series	R3 / B	22 stopped syllables 21 open syllables	R4 / C	451 open syllables
	R3 / C	342 long stopped syllables 451 open syllables		

The mapping between Samre pitch contours and Chong registers is broadly as expected although not entirely neat. Ploykaew does not report creak or glottal codas, yet like Thongkam’s Kasong all Samre etyma of the R2 and R4 categories are reported with high-falling tones (tone C). Some etyma expected to fall into R1 and R2 are also recorded by Ploykaew as having high-falling tone, and the explanation for this is not clear. Most of Ploykaew’s R1 and R2 etyma do fall out into her A and B tones with mid and low contours respectively, consistent with expected pitch values for historical voiceless and

voiced onsets respectively. She also reports breathy voice with all tone B words, and many tone C words, consistent with Chong values, but regards breathiness as non-contrastive and only gives the matter a brief mention.

What the above survey description achieves is to allow us to create a consolidated description of Pearic registers which is more richly detailed than the highly abstracted phonological representations that scholars seek to achieve in their grammars. This will be informative later when the origin of registers is discussed. The following consolidated table of phonetic correlates is proposed.

Table 5: Phonetic correlates of Pearic registers

high / tense series	R1	Phonation: clear Pitch: mid-level Timbre: lowered/centred	R2	Phonation: clear+creaky Pitch: falling Timbre: lowered/centred
low / lax series	R3	Phonation: breathy Pitch: low-level	R4	Phonation: breathy+creaky Pitch: falling

It is also significant to note that breathy phonation tends to be manifest on the left side of the syllable nucleus, and creaky phonation on the right. Hence in the 4th register, for example, the two phonations have distinct durations rather than overlap. This timing difference reflects the important insight that breathiness correlates historically with onset voicing, while creak—as a manifestation of tensed articulation—tends to associate with syllable codas, and can arise from tense coda segments. Consider the well-known fact that the creaky *hôi* and *ngã* tones in Vietnamese derive from historical *-h and *-s codas that rephonologised as increased glottal tension in vowel articulations (Ferlus 1998).

Diffloth (1989) offers several valuable observations on the distribution of creak in Pearic. They do not all bear directly on the question of origin, but they do provide insight into factors conditioning the distribution of creak which help to sort out historical and secondary factors.

Here also there are systematic gaps: eg. Chong words with final -h are never found with the glottalised registers R3 and R5; these two registers can never be preceded by Chong ʔ-, h-, b-, or d-; Chong long i: and u: are never found in the glottalised registers R3 and R5 with final velars, although they exist with other finals, or when short.

There are also a few cases of Chong register ablaut involving glottalisation due to derivational morphology: /khe:t/ (R1) “to comb” vs. /kanɛ:t/ (R3) “a comb”, /kɛ:p/ (R2) “to roast fish held with sticks” vs. /kanɛ:p/ (R4) “roasting sticks for fish”, /pɛ:k/ (R2) “to laugh” vs. /kanɛ:k/ (R4) “a joke”. (Diffloth 1989:145)

The observation regarding the contextual lack of long *i:* and *u:* is important and we will return to it below. The register ablaut is also important; Ferlus (2011) alluded to this suggesting that presyllables may have conditioned tenser articulation, but this is more likely to be a secondary factor that disrupted the distribution of creak in the lexicon than explaining its origin in Pearic.

2 Proto-Pearic Reconstruction of Headley (1985)

The Headley (1985) reconstruction (henceforth *Headley*) proposes that proto-Pearic had a sesquisyllabic syllable template that is essentially consistent with the conservative AA type. Main syllables are characterised as C₁(C₂)V₁(F₁).⁴ These could be preceded by presyllables which were assumed to have a similar maximal shape potentially, although in practice only a small set of pre-syllables are reconstructed (consisting mostly of a single consonant and non-contrastive vowel) and attention is directed to main-syllables as the most complex and important structure in the phonological word.

⁴ The same template is posited by Shorto (2006) for his proto-Mon-Khmer main-syllable.

Main-syllable C₁ is reconstructed by Headley as any of the following inventory:

*	p	t	c	k	ʔ
	P	T	C	K	
	b	d	ɟ	g	
	m	n	(ɲ)	ŋ	
	hm	hn		hŋ	
	v	l, r	j		
		hr,	hj		
		hl			
		s		h	

One can treat the *hC* items as devoiced sonorants rather than clusters, depending on the model of syllable structure one adopts. Note also that Headley was unable to find examples supporting onset *ɲ, but this segment can be confidently reconstructed on the basis of the database used for this study.

Proto-Pearic onset clusters formed with C₂ = *r, *l, *j, *v. Additionally, Headley posited complex word onsets (*kn, *rəg, *kd, *km, *gm, *gn, *ks, *pn, *bn, *pŋ, *ps, *sk, *sm, *sn, *sŋ, *tm, *tŋ, *tp) without making explicit whether the initial segments were syllabified, although this can be reasonably assumed (i.e., syllabification rules approximating those of Khmer).

Proto-Pearic mainsyllable codas (F₁) are reconstructed to follow the typical AA pattern of having no voicing distinction but matching the same places of articulation as the onsets, hence:

*	p	t	c	k	ʔ
	m	n	ɲ	ŋ	
	v	l, j			
		r			
		s		h	

The proto-vowel inventory was reconstructed as follows:

*	i	ɨ	u	
	e	(ə	o	
)		
	ɛ	a	ɔ	±
				long

The proto-schwa was only tentively reconstructed, Headley explains:

There is little good evidence for proto-Pearic /*ə/ and /*ɐ/, these two phonemes would be expected in view of the generally symmetrical nature of the proto-Pearic vowel system and the broad occurrence of /ə/ in other Mon-Khmer vowel systems. (Headley 1985:455).

This writer's Pearic lexical database extends Headley's etymologies to more than 400 comparisons, and the analysis finds no compelling case for his *ə or *ɐ. It appears that reflexes of the high-central and mid-central vowels are in complementary distribution, and we can posit a pair *ɿ, *ɿ: that broadly split *ɿ > ə, *ɿ: > ɨ: (with some variation), clarifying Headley's uncertainty regarding the central nuclei.

Headley's proto-vowel inventory is notable for the lack of diphthongs, and this is unlikely to be completely correct. Most Chongic lects are recorded without /iə/ or /uə/ diphthongs (or equivalents) while Samre (as in Ploykaew's 2001 lexicon) generally reflects /iə/ or /uə/ for *ɛ: and *ɔ:. This suggests that pAA *iə and *uə shifted and merged with *ɛ: and *ɔ: pre-proto-Pearic, and more recently *ɛ: and *ɔ: all diphthonsed in Samre. However, there is a small set of AA etyma which indicate proto-AA *iə, and traces of these are found in the Pearic sources, for example Baradat (1941) lists Western Chong *priəl* 'hail', and we can propose that—very marginally—proto-AA *iə was transmitted unchanged into

proto-Pearic. On the other hand, there is no support for ***uə**. Such an asymmetry may seem odd, yet the same pattern of only continuing pAA ***iə** while showing no indication of ***uə** is also reconstructed for proto-Palaungic (Sidwell 2015) and proto-Khasian (Sidwell 2018).

2.1 The Aspirated Stops and Breathy Registers

Headley's ***P *T *C *K** series is the most striking feature of the proto-phonology. The series is posited to account for the correspondence of plain stops in both KPT and Chongic, in apparent contrast to a correspondence of plain stops in KPT with aspirates in Chongic; consider the selected examples in Table 6:⁵

Table 6: Pearic *P *T *C *K** series**

proto-Pearic (1985)	Chong	Samre	Suoy	KPT	Other AA (indicative of pAA onsets)
* Kic 'small'	kic	kic ³⁴⁴	kich	kech	Laven <i>ke:c</i> 'small'
* Ce:v 'go/walk'	cɛ:w	ci:w ³³²	chév	--	Mon <i>cow</i> 'return'
* Tɔ:ŋ 'fear'	ta:ʔŋ	tuəŋ ⁴⁵¹	taâng	taâng	Katu <i>patuəŋ</i> 'fear'
* Po:t 'cut/hack'	paot	pu:t ²²	paût	--	Mon <i>pvt</i> 'break/part'
* koj 'chameleon'	k ^h o:ʔj	--	khoy	khnōy	Palaung <i>ʔakoj</i> 'large spiny lizard'
* ciʔ 'louse'	c ^h i:	c ^h i: ³³²	chhî	chei	Temiar <i>ceʔ</i> 'louse'
* ta:n 'to weave'	t ^h a:n	t ^h a:n ³³²	thanh	tanh	Katu <i>ta:n</i> 'weave'
* peʔ 'three'	p ^h e:ʔw	p ^h e: ⁴⁵¹	phé-i	pai	Laven <i>pe:</i> 'three'

This evidence of two contrasting correspondence series among the voiceless stops is challenging to interpret, and Headley tentatively suggested two distinct voiceless proto-series, wondering if ***P *T *C *K** might have been aspirates. However, as Ostapirat (2009) points out, this would suggest an odd combination and direction of changes, with one series losing aspiration while another acquired it. It is also significant that members of both series are reflected by plain voiced stops elsewhere in AA to the extent that cognates can be found, strongly suggesting that both ultimately derive from a single series.

Ferlus (2011) assumes the priority of the wider AA values, and does not recognise the need to distinguish ***P *T *C *K** and ***p *t *c *k** series in proto-Pearic. Instead, he regards the aspirated stops as the normal reflexes of ***p *t *c *k** in Chongic, having undergone a Germanic shift, such has also occurred independently in the Mal-Pray Khmuic languages, in Khasi, and in Angkuic lects of Palaungic (and identified as such by Haudricourt 1965). It is supposed that the Germanic shift failed to occur in KPT Pear altogether, while the presence of some words with plain stops in Chongic is explained as, "due to the Khmer language influence" (Ferlus 2011:42). Unconditioned Germanic shifts are known to occur in AA languages, and since we know that all Pearic lects have been under long-term Khmer pressure it is expected that Khmer words with plain stop onsets would be borrowed into Pearic languages, and that the pronunciation of some Pearic words would be influenced by the more prestigious phonology of Khmer cognates. Ferlus' hypothesis was augmented by Diffloth in 2012 at the Argay SEALS meeting,⁶ with further evidence that aspiration only developed in Chongic ('South-Pearic' in Diffloth's classification), and is only present in KPT words borrowed from Khmer.

Consequently, Ferlus' explanation for the ***P *T *C *K** is accepted here, with the additional suggestion that etyma reflecting this marked stop series may include borrowings from languages other than Khmer, such as Thai/Siamese, and other AA languages, and new indigenous words coined by Pearic speakers, as well as some native words that may have been expected to have aspirated onsets but do not due to analogical influence of cognate foreign words.

⁵ Note that Souy, Samray, and KPT data used are from Baradat (1941) and are transcribed in a non-IPA Romanisation, including *ch* for [c ~ tɕ]. Chong examples in tables are from Premssirat (2008) and Samre from Ploykaew (2001).

⁶ For the presentation handout see <http://jseals.org/seals22/diffloth2012fourh.pdf> (checked 15 Sept. 2022). Note that Diffloth refers to KPT Pear as *Chuang*.

Another important feature of Headley’s reconstruction is the voiced stop series **b, *d, *j, *g*; Pearic languages generally lack a full series of voiced stops, showing only /b, d/ which may be pronounced plain voiced or imploded. Headley (1985:432) speculated that the words with /b, d/ are borrowings from Khmer, and in some cases, this is clearly true; additionally, there are also /b, d/ words borrowed from Thai, and a small set of /b, d/ words in Chong lects which apparently are native forms reflecting an earlier preglottalised series **ʔm* and **ʔn* which we discuss further below. Otherwise, voiced stops are rare.

The evidence for a proto-Pearic voiced stop series is the existence of breathy phonation in the lax register. Breathily phonation occurs in a proportion of syllables with oral stop onsets, and in most syllables with sonorant onsets. The consensus explanation, consistent with generally understood principles of registrogenesis (at least since Huffman 1976)⁷, is that historically there was a full set of voiced oral stop and sonorant onsets, and these were pronounced with breathy or murmured phonation. At some point that marked phonation spread into the vowel, and later voicing in oral stops was lost leaving just the marked phonation in those syllables as the trace that there had once been a voicing contrast. This theory appears to be unassailable, although there are some considerations and distributional peculiarities that need to be mentioned.

Firstly, recall that Headley reconstructed a devoiced sonorant series. Pursuing the same logic that correlates breathiness with onset voicing, such devoiced sonorants were assumed to underlie various modern Pearic words with sonorant onsets but non-breathy phonation, such as in Table 7.

Table 7: Pearic tense(R1, R2) sonorant onsets

proto-Pearic (1985)	Chong	Samre	Suoy	KPT	Other AA (indicative of pAA onsets)
*hluk ‘salt’	luk	luk ³⁴⁴	luk	lok	Khasi <i>mlu?</i> ‘salt’
*hjok ‘breast’	--	jok ³⁴⁴	yǎk	--	--
*hle:k ‘chicken’	le:k	liək ³⁴⁴	lèk	lék	--
*hlɔ:ŋ ‘banana’	lɔ:ŋ	luəŋ ³³²	--	lāng	Palaung <i>rəhlɔ:ŋ</i> ‘hand of bananas’
*hliŋ ‘deep’	li’ŋ	luŋ ³³²	liŋg		Khasi <i>ɟliəŋ</i> ‘far, deep’
*hma:r ‘field’	ba:j	ma:ɽ ³³²	marr	--	Lawa <i>ʔma</i> ‘rice field’
*(m)o:t ‘younger sibling’	bo:t	mu:t ³³⁴	maût	mût	Mon <i>həmot</i> ‘young creature, child’
*lo:s ‘muntjak’	lo:t	lu:h ³³⁴	laûs		Khmer <i>clu:h</i> ‘muntjak’
--	dak	nak ²²	--	--	OldKhmer <i>’nak</i> ‘person’
--	wa:	wa: ³³²	ûva	va	Khmer <i>swa:</i> ‘monkey’

However, it is not obvious that simple devoicing is the best explanation for such words. The ‘field’ and ‘younger sibling’ etyma suggest that preglottalization, rather than devoicing, is a possible explanation for some or all of such cases. At the same time, external comparisons for ‘banana’, ‘deep’, ‘monkey’ and ‘muntjak’ variously suggest that a fricative or affricate preceded the sonorant historically, perhaps leniting in sequence **j > *c > *s > *h* yielding a devoiced sonorant series in proto-Pearic. At this stage it appears that we have to add at least **ʔm* and **ʔn* to our proto-onsets in addition to the hypothetical devoiced series which blocked breathy phonation.

In relation to the distribution of aspiration and breathy register, we would predict that these would not co-occur in Chongic, since the former originate with onset voicing and the latter from the Germanic shift among voiceless stop onsets. However, in Premsrirat’s (2008) Chong lexicon we note that all syllables with voiceless oral stop onsets are marked as aspirated, whether breathy syllables or clear. This appears to be a notational artifact of conflating aspiration and breathy phonation, if not an outright phonological merger. We noted above that in Huffman’s (1985) description of Chong oral stops he distinguished a slight aspiration in lax syllables in contrast to a strong aspiration in tense syllables with aspirated onsets. For our purposes we will treat this as a notational merger - it does not materially affect the reconstruction since this does not remove information necessary to distinguish historical values.

⁷ That paper based on a presentation and discussion at the 1973 ICAAL meeting, Hawaii.

However, the issue of the relative timing of the historical processes of devoicing and Germanic shift is problematic, as there is a potential clash in terms of feeding rules for the phonological changes.

Assuming that historical devoicing was a single change that occurred during or before proto-Pearic, and given that the Germanic shift did not occur in KPT, it is logical to assume that the Germanic shift occurred after the isolation of KPT from Chongic. However, that model implies that all oral stop onsets in early-Chongic were voiceless, and we would then expect all of them to undergo Germanic shift, which they did not. Without proposing independent devoicing in KPT and Chongic, we must assume that the devoiced stops in early Chongic remained distinct (inherently or contextually) from the other voiceless stops, in such a manner as to block the Germanic shift. Another way of looking at it is to say that the tense series stops had characteristics that favoured the Germanic shift, while the lax series stops did not. Additionally, depending on the timing of these changes, we must allow for the *P *T *C *K series items which are tense but show no aspiration. It seems that the ready explanation for the latter is that they were affected by Khmer (and other) influence at the time, or were acquired after the Germanic shift had occurred. The proposal offered here is that devoicing yielded a lenis series (*b̥, *d̥, *j̥, *g̥) with reduced duration of voicing and an association with breathy phonation, which later merged with plain voiceless stops in all lects.

It is noted that at a historically deeper level the proto-Pearic lenis stops derive from a merger of proto-AA implosive and plain voiced stops. According to the historical reconstruction (Shorto 2006, Sidwell & Rau 2014) proto-AA voiced stop onsets were *b, *d, *b, *d, *j, *g.⁸ The implosive labial and apical stops greatly outnumbered *b, *d in prevocalic position, essentially matching the pattern reconstructable for proto-Katuic (Sidwell 2005), proto-Vietic (Ferlus 2007), proto-Monic (Diffloth 1984). Consequently, the proto-Pearic lenis stops for which AA etymologies are evident are mostly traceable to original implosives.

3 Creak / Glottal Constriction in Pearic

It is notable that Headley reconstructed the glottal stop as a possible coda in proto-Pearic, contrasting with zero coda. Examples are given in Table 8:

Table 8: Phonetic correlates of Pearic registers

proto-Pearic (1985)	Chong	Samre	Suoy	KPT	Other AA
*ʒrəlaʔ ‘thorn’	kəlaʔ	sala: ⁴⁵¹	srala a	chrâléa	Khmu <i>crlaʔ</i> , Bahnar <i>jəla:</i>
*gmaʔ ‘rain’	kəmaʔ	kama: ⁴⁵¹	kmāk	khméa	Khmu <i>kmaʔ</i> , Bahnar <i>ʔmi:</i>
*pe(:)ʔ ‘three’	p ^h e:ʔw	p ^h e: ⁴⁵¹	phék	pai	Khmu <i>peʔ</i> , Brao <i>pe:</i>
*tʃiʔ ‘sun’	t ^h əŋiʔ	--	thnîk	thngei	Khmu <i>sŋiʔ</i> , Brao <i>taŋaj</i>
*ca: ‘to eat’	c ^h a:	c ^h a: ³³²	chha	cha	Khmu (<i>mah</i>), Bahnar <i>sa:</i>
*pli: ‘fruit’	p ^h li:	p ^h li: ³³²	phlî	plei	Khmu <i>pleʔ</i> , Bahnar <i>ple:j</i>
*blu: ‘thigh’	p ^h lu:	plu: ²¹	plû	plaûv	Khmu <i>bluʔ</i> , Bahnar <i>blu:</i>
*bri: ‘forest’	p ^h ri:	p ^h ri: ²¹	prî	prî	Khmu <i>briʔ</i> , Bahnar <i>bri:</i>

Table 8 includes comparisons with Khmu which always retain proto-AA glottal coda, and a Bahnaric language that always reflect the same etyma with open syllables. Although there are some minor discrepancies in the data, it is clear that the etyma for which Headley reconstructed final glottals belong to creaky registers R2, R4, while the open syllables belong to R1, R3. The best interpretation, consistent with Ferlus’ (2011) analysis, is that the glottal codas recorded for Chong (by Huffman 1985, Premsrirat 2008), in Suoy written as *k* (by Baradat 1941) reflect creaky phonation phonologised as glottal codas. The implication is that some time before proto-Pearic the etymological glottal codas were lost, a change

⁸ An implosive palatal *ʃ- onset is also tentatively posited by Sidwell & Rau (2014) on the basis of some Katuic forms.

also shared with Khmer, Katuic, Bahnaric⁹ (suggesting perhaps an areal feature of Indo-Chinese AA languages).

A further implication is that—contra the suggestion of Diffloth (1986)—it is unlikely that proto-AA had a creaky/non-creaky contrast which proto-Pearic inherited, and there are two strong pieces of evidence for this. Diffloth (1986) himself compared the lexical distribution of creak across Pearic, Katuic and Vietic,¹⁰ finding that the feature does not manifest in the same cognate lexicon across the three branches, such that regular correspondences cannot be established. his suggestion at the time was that somehow creak became redistributed differently in each group:

A long interim period evidently separates Proto-Austroasiatic from each of the proto-branch levels. I have mentioned a few systematic distributional differences between the creakiness of Proto-Katuic, Proto-Pearic and Proto-Vietic; these differences are probably the result of distinct innovations which occurred during intermediate periods. If we can discover more of these early innovations, and identify inter-branch borrowings, we may be able to explain the inconclusive etymological evidence shown above and reconstruct creakiness at the PAA level; for the moment, it is just a presumption. (Diffloth 1989:150)

Diffloth's hope has not been satisfied, and no evidence for linking Pearic, Katuic and Vietic creak etymologically has emerged over the past three decades, which brings us to our second consideration: all AA languages that use creak phonologically also contrast open and closed syllables.

There is a consensus among scholars that proto-AA word canon did not permit zero codas, a condition still reflected in some AA languages today (such as in various Khmuic and Aslian languages). Another way of characterising the issue is offered by Sidwell & Rau (2014:242-243) pointing out that across AA it is common for obstruent codas to be pronounced as unreleased and tense, with a coincident glottal constriction. In other words, all stop codas come with oral and glottal constriction specifications; a glottal stop coda lacks a specified oral place of constriction, and a zero coda (in languages that permit them) has neither oral place nor glottal constriction specified. AA languages with phonological creak do not contrast creaky and glottal final syllables, and they always have a robust contrast between closed and zero coda. The implication is that creak arises in languages that have already lost glottal stop codas, and/or glottal codas are lost as they merge with creaky phonation (and this may also see clear open syllable arise as creak is blocked or suppressed by other articulatory or perceptual factors). Thus, based on the synchronic typology and our understanding of creaky phonology, hypothetical proto-AA creak would not predict AA languages that prohibit zero codas.

Ferlus (2011) responds to Diffloth (1986) by pointing out that creak is used phonologically in Pearic, Vietic, but only a restricted sub-set of Katuic (lects of the Ta-oih sub-group such as Ong, Ir, Katang etc.). He relates this to the fact that these languages lay along a 1st millennium trade route that he calls the *Han Trail* (referred to in the 8th century *Great Treatise of Geography*, *shí dào zhì* 十道志). This route connected Chinese Nam Việt (Northern Vietnam) overland with the Gulf of Thailand and the way to India via the Isthmus of Kra, avoiding coastal piracy associated with navigating around Indo-China. This is proposed to have involved Chinese speakers along the route, at a time when the Chinese language was transitioning phonologically between Old to Middle Chinese, and had undergone a register split yielding tense versus lax syllables, on the path to the full development of tones in Middle Chinese. Ferlus supposes that AA speakers along the Han Trail acquired a tense/lax contrast in their speech by exposure to Chinese speaking with the same feature.

The Han Trail contact hypothesis is intriguing but difficult, being based on supposed geographical and temporal correlations without a body of compelling linguistic evidence. It is not obvious by what mechanism a phonological distinction (tense/lax) whose phonetic correlates in late Old Chinese are not clear, could have spread to quite different sets of vocabulary in three different AA branches; for example, we do not find a common body of Chinese lexical or other borrowings correlating with the

⁹ In Bahnaric the loss of -ʔ was complete in South Bahnaric, while in the rest of Bahnaric coda -ʔ is retained in some etyma, and the explanation for this is not clear and deserves further attention.

¹⁰ Evidence of creaky registers has since been found in some Palaungic languages, and there are suggestions of creak in some Munda lects, but any possible relation between these and the present problem are not established and in the view of the present writers considered unlikely.

trail geography. In effect, Ferlus has replaced Diffloth’s speculative genetic hypothesis with a speculative areality hypothesis, but the same fundamental problems remain. In short, neither approach has anchored their explanation in segmental or other structural correlates which might explain the distribution of creak in the languages.

3.1 Typological Parallels – Pacoh (Katuic), Sedang (North Bahnaric)

Assuming that Pearic creak is not inherited from proto-AA, but innovated in an early stage of Pearic, we return to the observation by Diffloth (1986:145) that, “the Chong long i: and u: are never found in the glottalised registers”. This fact can be related to other correlations between vowel height and register noted in regard to Pacoh (Katuic) and Sedang (North Bahnaric), and suggest an explanatory hypothesis that we will discuss in the next section.

Diffloth (1982) had proposed a historical explanation for the tense/lax registers in Pacoh, which appear to distribute independently of historical voice onset values (“registrogénèse hérétique”). The basic problem is that Pacoh high vowels are lax, low vowels are tense, and the contrast is functional for the mid vowels (see Alves 2006 for discussion of Pacoh registers). Diffloth suggested that sound changes had resulted in a crowding of the mid-vowel space and phonological mergers were avoided by speakers giving prominence to phonation differences among the old and new mid-vowels. The model does not explain the origin of the register differences, but does model the phonologization of the contrast. Since the 1980s our understanding of phonation types and their correlated has improved greatly, and instrumental studies (Kuang 2013 and others) confirm that tense vowels are inherently lower (more open) than lax vowels in register languages, with discussion since Gregerson (1976) linking this to effects of jaw and tongue position and perception of acoustic cues. Knowing this, we can be reasonably sure that the Pacoh system arose out of the vowel space splitting into tense and lax registers based on these inherent tendencies, with phonation becoming contrastive in the mid vowel space where opportunity or pressure for reanalysis would be greatest. Something like this also apparently happened in North Bahnaric.

The North Bahnaric languages are marked by a register system that in most cases contrasts modal and breathy voice, yet there was no devoicing of onsets and no correlation between onset voicing and register values (Smith 1972, Gregerson 1976). Smith (1972) remarked on the lack of low vowels in the North Bahnari lax register, and Sidwell (1998) proposed an historical phonology based on the correlations between vowel height and phonation, paralleling the explanation for the registrogenesis in Pacoh.

Within North Bahnaric, Sedang presents a special case; the register distinction is one of tense/creaky versus lax/modal, and it is clear that this arose from a general (unconditioned?) shift in register setting whereby the syllable generally became more tense; modal syllables became creaky and breathy syllables became modal. Logically, modal syllables became creaky first, which would avoid register mergers, and breathiness became less important as a marker of lax register. Another factor is that Sedang shows no glottal stop codas in any native syllables in either register (more recent loans can show exceptions). Additionally, Sedang devoiced all voiced onset obstruents (except contextually in *jr*, *dr* clusters).

This tensing of the phonology did not happen in isolation from, or without, other changes. Unlike its immediate neighbours, Sedang devoiced all its oral stop onsets, lost all oral stop codas in tense syllables, and lost glottal stop codas in all syllables. It is assumed that in lax rhymes the glottal constriction normally coincident with stop codas was relaxed, while in tense rhymes that glottal constriction merged with the strong creaky phonation characteristic of the tense phonation setting. Although it is not clear if the loss of voicing in obstruent onsets is linked to the Sedang creaky register since register values appear to be similarly phonologically blind to the voicing of onsets historically and synchronically (Smith 1972:76), the general loss of glottal stop codas is shared, and the distribution of creak does have specific segmental correlates.

It was pointed out by Smith (1972) that there is a clear correlation between register and vowel height in Sedang and North Bahnaric, although he did not investigate this further. Smith lays out 571

North Bahnaric etymologies in his reconstruction, and includes 473 Bahnar cognates.¹¹ The latter were originally included in the study as it was assumed at the time that Bahnar was a North Bahnaric language, but this idea was soon abandoned. Gregerson et al. (1976) demonstrated that Bahnar belongs within a separate Central Bahnaric sub-branch, which lacks register and instead shows an archaic unrestructured vocalism. Consequently, we can be confident that Bahnar vowels are reasonably indicative of proto-Bahnaric values, and we can treat them as a proxy for proto-Bahnaric for our purposes. On this basis, we align Smith's 473 vowel comparisons between Sedang and Bahnar, counting 314 tense and 159 lax tokens. Of the the tense tokens, 308 align with Bahnar low vowels ε , a , ɔ and diphthongs ia , ua , while only six align with higher vowels. Of the lax tokens, 153 align with Bahnar high and mid vowels i , i , u , o , ə , and only six align with low vowels. Half of the apparent exceptions appear to relate to nuclei lowered by dissimilation in the environment of palatal codas, so arguably this leaves only six tokens inconsistent with the hypothesis of a correlation between historical vowel height and phonation contrast in North Bahnaric, a 97.8% agreement, a good approximation of neogrammarian regularity.

The compelling interpretation of these facts is that North Bahnaric register reflects a phonologisation of the inherent phonation differences between higher and lower vowels, somewhat paralleling the explanation for Pacoh registers first tentatively suggested by Diffloth (1982). Additionally, while most North Bahnaric languages manifest the register contrast as one of modal-breathy, implying that the same is reconstructable to proto-North-Bahnaric, Sedang has tensed the overall system so that the register contrast is one of creaky-modal, apparently in a spontaneous/unconditioned shift.

Before proceeding to discuss the Pearic registers further, we take one more look at Katuic. Gehrman (2015) surveyed Kuy (Katuic) dialects and found a correlation between between vowel height and register, specifically among the high vowels. In the Kuy case we know that its breathy register initially arose in the conventional manner (Sidwell 2005 and others), and Gehrman has demonstrated that the tense/lax contrast in high vowels merged to lax in some lects, generalizing the inherent tendencies of high vowels to favour lax phonation. As such Gehrman has provided a type case of rephonologization of register to align inconsistently with historical onset voicing values. The possibility of rephonologization of registers was proposed by Ferlus (1974) to explain creak in the Ong dialect of Ta-oih, although the hypothesis was vaguely formed, and required multiple hypothetical onset devoicing phases. The ideas of both Ferlus and Gehrman are very interesting in terms of explaining how register values can become disconnected from onset voicing values, but they share the requirement that devoicing comes first. In both North Bahnaric and Pearic it seems clear that devoicing came after their initial register splits, so we have to look to other explanatory principles, in this case vowel height/openness.

3.2 Pearic Creak and Vowel Height

The above considerations were discussed by this writer with Robert Headley during a visit to Maryland (USA) in 2015 and out of that discussion the following proposals emerged:

- Pearic creaky register began at an early stage, either in conjunction with loss of glottal stop coda or at some stage after that, but before devoicing of obstruent onsets.
- Main-syllables with [-high] vowels mostly became tense > creaky; this will have been contextually blocked by phonological processes and borrowings.
- Pace Diffloth (1989) creak was blocked in syllables with a glottal stop or glottal fricative in the onset (simplex or clustered, including $*?m$, $*?n$ which are reflected as b , d in Chong).

To investigate the viability of these hypotheses, I compiled 73 etymologies from my Pearic etymological database which I am reasonably confident are direct reflexes of proto-AA forms, or at least reflect widely distributed AA roots that do not appear to have entered Pearic as borrowings (Tables 9-11). It would be possible to compile more, but the sheer volume of borrowings and lexical innovations

¹¹ This is my count after false cognates and loans are culled from the data.

renders it increasingly difficult to confidently stratify the lexicon and it is necessary to keep the project within a reasonably practical scale. The tabled examples are organized according to register classes and noted with a tick mark when they conform to the proposed rules. The discussion that follows separates the examples according to register and deals with the apparent exceptions (in terms of the rules listed above) and their potential implications.

Table 9: R1 – Clear Register Etyma

	Gloss	pAA	pPearic revised	pPearic Headley	Chong	Samre	Suoy	KPT	Notes
1	‘louse’	*ci:ʔ	*ci:	*ciʔ	c ^h i:	c ^h i: ³³²	chhî	chei	✓
2	‘wife/woman’	*kan	*kɤn	*kin	k ^h in	k ^h um ³³²	khün	--	✓
3	‘father’	*ku:ŋ	*ku:ŋ	--	--	k ^h u:ŋ ³³²	khûnh	--	✓
4	‘hair’	*suk	*suk	*suk	(kə)suk	suk	suk	sâsok	✓
5	‘hand’	*ti:ʔ	*ti:	--	ti:	ti: ³³²	tî	ktei	✓
6	‘fruit’	*pli:ʔ	*pli:	--	p ^h li: ³³²	p ^h li:	phlî	plei	✓
7	‘salt’	*-luk	*(ʔ/h)luk	*hluk	luk	luk ³⁴⁴	luk	lok	✓
8	‘breath’	*-hə:m	*-he:m	--	məhe:m	kahi:m ³³²	--	--	✓
9	‘blood’	*-ha:m	*-ha:m	*-ha:m	maha:m	ha:m ⁴⁵¹	hām	môham	✓
10	‘to fly’	*hər	*hɤr	*h(i:)r	hij	hu: ³³²	hürr	hoer	✓
11	‘die/dead’	*ho:c ‘decay’	*ho:c	*ho:c	ho:c	hu:c ³³⁴	haûch	hōch	✓
12	‘to bathe’	*hu:m	*hu:m	*hu:m	hu:m	hu:m ³³²	hûm	hôm	✓
13	‘field’	*ʔma:r	*ʔma:r	--	ba:j	ma:r ³³²	marr	--	✓
14	‘monkey’	*swa:ʔ	*hwa:	--	wa:	wa: ³³²	ûva	ûva	✓
15	‘wasp’	*ʔo:ŋ	*ʔo:ŋ	--	ʔo:ŋ	ʔuəŋ ³³²	--	--	✓
16	‘excrement’	*ʔic	*ʔic	*ʔic	ʔic	ʔic ³⁴⁴	ich	ech	✓
17	‘bran’	*ʔlɤk	*ʔlɤk	*lik	lik	luk ³⁴⁴	lük	lôlek	✓
18	‘firewood’	*ʔu:s	*ʔu:s	--	ʔu:t	ʔu:h ³³⁴	ûs	aûs	✓
19	‘to bark’	*ke:r	*ke:r	*Ke:r	ke:w	ki:r ²¹	kérr	kierr	?
20	‘dog’	*cə:ʔ	*cə:(?)	*c(ə)ʔ	c ^h ə:	c ^h ə: ⁴⁵¹	chhō	châ	?
21	‘eat’	*ca:ʔ	*ca:	*ca:	c ^h a:	c ^h a: ³³²	chha	cha	X
22	‘to spit’	*kcoh	*(k)coh	*c(u)h	c ^h uh	cuh ³⁴⁴	chhûs	kcós	X
23	‘neck’	*kə:k	*kə:k	--	kə:k	kuək ³⁴⁴	kâk	--	X
24	‘road/way’	*kraʔ	*kra:	--	k ^h ra:	k ^h ra: ³³²	kra	kra	X
25	‘pounded rice’	*ŋkoʔ	*rəko:	--	kək ^h o:	kak ^h o: ³³²	rôkhaû	rôkaû	X

It is apparent that in 16 out of 24 R1 examples, the condition of a high vowel and/or glottal stop/fricative onset is met, so these are regular. Note that for our purposes *ɣ is assumed to belong to the +high class. Also in terms of triggering breathy phonation, the presence of a voiceless segment in an onset cluster is sufficient to block this.

Table 10: R2 – Clear-Creaky Register Etyma

	Gloss	pAA	pPearic revised	pPearic Headley	Chong	Samre	Suoy	KPT	Notes
26	‘chin’	*ka:ŋ	*kma:ʔŋ	--	kəmə:ʔŋ	kama:ŋ ⁴⁵¹	<i>krâ mang</i>	--	✓
27	‘rat’	*kɔ:n ‘child’	*kɔ:ʔn	--	kʰɔ:ʔn	kʰuən ⁴⁵¹	<i>khôn</i>	--	✓
28	‘wind’	*kja:l	*kja:ʔl	--	kʰəja:ʔj	kʰaja:l ⁴⁵¹	<i>khyâl</i>	<i>khiel</i>	✓
29	‘hawk’	*kla:ŋ	*kla:ʔŋ	--	kʰla:ʔŋ	kʰla:ŋ ⁴⁵¹	<i>khan</i>	<i>klang</i>	✓
30	‘tongue’	*kltɑ:k	*kltɑ:ʔk	--	kəta:ʔk	kata:k ³⁴²	--	--	✓
31	‘ghost/corps e’	*kmo:c	*kmo:ʔc	--	kʰamu:ʔ c	kʰamu:c ³⁴²	<i>khmaûch</i>	<i>khmaûc h</i>	✓
32	‘crab’	*kta:m	*kta:ʔm	--	tʰa:ʔm	tʰa:m ⁴⁵¹	<i>tham</i>	<i>ktam</i>	✓
33	‘four’	*puən	*po:ʔn	--	pʰo:ʔn	pʰu:n ⁴⁵¹	<i>phoaûn</i>	<i>phôn</i>	✓
34	‘before’	*tu:j	*tɔ:ʔj	*Tɔ:j	tɔ:ʔj	tuəj ⁴⁵¹	<i>tây</i>	<i>tây</i>	✓
35	‘to fear’	*tu:ŋ	*tɔ:ʔŋ	*Tɔ:ŋ	tɔ:ʔŋ	tuəŋ ⁴⁵¹	<i>taâng</i>	<i>taâng</i>	✓
36	‘three’	*peʔ	*pe:ʔ	*pe(:)ʔ	pʰe:ʔw	pʰe: ⁴⁵¹	<i>phék</i>	<i>pai</i>	✓
37	‘earth’	*tiʔ	*teʔ	*teʔ	tʰeʔ	tʰe: ⁴⁵¹	<i>thek:</i>	<i>tai</i>	✓
38	‘leaf’	*slaʔ	*sla:ʔ	--	kʰalaʔ	kʰla: ⁴⁵¹	<i>slāk</i>	--	✓
39	‘stone’	*tmoʔ	*tmo:ʔ	*tmoʔ	kʰəmoʔ	tʰamo: ⁴⁵¹	<i>thmok</i>	<i>thmau</i>	✓
40	‘dream’	*ʔmpaw	*po:ʔ	*poʔ	pʰoʔ	pʰo: ³³²	<i>phok</i>	--	✓
41	‘skin’	*crloʔ ‘bark’	*trlo:ʔ	--	kʰəloʔ	talo: ⁴⁵¹	<i>trâlok</i>	--	✓
42	‘crow’	*kʔa:k	*ka:ʔk	--	kəna:ʔk	ka:k ³⁴²	<i>ka-ak</i>	<i>ka-ak</i>	✓
43	‘chameleon’	*ko:j	*ko:ʔj	*koj	kʰo:ʔj	--	<i>khøj</i>	<i>khñöy</i>	✓
44	‘ant’	*so:c ‘sting’	*su:ʔc	--	su:ʔc	(samu:c ³⁴²)	(<i>srâmaûch</i>)	<i>sâaûch</i>	?
45	‘bird’	*ci:m~ *ce:m	*ci:ʔm	--	cʰi:ʔm	cʰi:m ⁴⁵¹	<i>chhîm</i>	<i>châèm</i>	?
46	‘cooked’	*cin~ *ce:n	*ci:ʔn	--	ci:ʔn	cʰi:n ⁴⁵¹	<i>chhîn</i>	<i>chhâ-en (?<Kh.)</i>	< Khmer ?
47	‘sun/day’	*tɲi:ʔ	*tɲi:ʔ	--	tʰəŋiʔ	sani: ⁴⁵¹	<i>thnîk</i>	<i>thngei</i>	< Khmer ?

The onomatopoeic form #18 *ke:r ‘to bark’ belongs to the K correspondence and arguably does not belong here. Also, #20 ‘dog’ is recorded as clear voiced in Chong sources, but the tones noted for Kasong (453) and Samre (451) are consistent with creak, so this example is ambiguous.

The remaining exceptions have some commonalities: four have zero codas < proto-AA *ʔ; to this we can add #23 ‘neck’ assuming k < ʔ by analogy/assimilation (c.f. Mon kɔʔ ‘neck’) and # ‘spit’ with *-h coda since h is effectively a voiceless vowel. Nonetheless, this seems likely to be a factor that would have blocked creak, considering that multiple R4 items have coda from proto-AA *ʔ.

From these data it appears that the mid-vowels /o, e/ behave as [-high] in terms of creak. The R2 etyma show a smaller proportion of problematic forms (4/22) and it may be that these are not strictly irregular but have explanations:

Some of the specific etyma may have contact explanations:

- #46 ‘cooked’ may be influenced by Khmer ផ្លែ ឆ cʔən which has a media glottal stop (not found in reflexes in other AA branches, e.g., Mon *cin*, Sedang *cen*, etc.).
- #47 ‘day’ may be influenced by Khmer ថ្ងៃ tɲaj in which case R2 is potentially regular.

Items #44 ‘ant’ and #45 #’bird’ are common AA etyma but the vowel reconstruction is problematic. *i: and *u: are proposed for proto-Pearic based on the values common in the modern reflexes, however we note that the spellings of these items by Baradat (1941) for KPT suggest more open vowels (*sâaûch*, *châèm* respectively), as do comparisons with other branches (e.g., ‘bird’ Bahnar *se:m*, Mon *hacem*, etc., ‘sting’ Bahnar *so:c*, Katu *soc*, etc.) although some branches also reflect these etyma with high vowels.

Given these conflicting indications (internal and external), it could be that the proto-Pearic forms were **so:ʔc* and **ce:ʔm* respectively and thus derive from [-high] nuclei.¹²

Table 11: R3 – Breathy Register Etyma

	Gloss	pAA	pPearic revised	pPearic Headley	Chong	Samre	Suoy	KPT	Notes
48	‘thigh’	*blu:ʔ	*b̥lu:	*blu:	p ^h l̥u:	plu: ²¹	pl̥u	plaûv	✓
49	‘forest’	*bri:ʔ	*b̥ri:	*bri:	p ^h r̥i:	pi: ²¹	pr̥i	prî	✓
50	‘you’	*bu(:)ʔ ‘who’	b̥u:	*b(o:)	p ^h u:	po: ²¹	p̥u	poû	✓
51	‘drunk’	*bu:l	*b̥u:l	*bu:l	p ^h u:l	pu:l ²¹	p̥u:l	pûl	✓
52	‘winnow’	*gu:m	*g̥u:m	*gu:m	k ^h u:m	ku:m ³³²	k̥u:m	--	✓
53	‘foot’	*ɣy:ŋ	*ɣ̥y:ŋ	--	--	siŋ ²¹	siŋ ^h	--	✓
54	‘year’	*cnam	*n̥ɰm	*nim	n̥im	num ²¹	n̥im	nim	✓
55	‘root’	*riəs	*r̥y:s	*re:s	re:t	ri:h ³⁴²	r̥é:s	rüs	✓
56	‘to weep’	*ja:m	*j̥a:m	--	ja:m	ja:m ²¹	yam	yam	?
57	‘eye’	*mat	*m̥at	*m(a)t	m̥ət	mat ²²	m̥é:ət	mé:ət	X
58	‘red’	*ɲa:r ‘green’	*ɲ̥a:r	*ɲ(ə)r	ɲ̥əj	ɲa:r ²¹	ngoãrr	ngarr	X
59	‘sap’	*jar	*j̥a:r	*j̥o:r	--	sa:r ²¹	s̥arr	(chôrr < Kh.)	X
60	‘water’	*da:k	*d̥a:k	*da:k	t̥h̥a:k	ta:k ³⁴²	ta:k	téak	X
61	‘thread’	*bra:j	*b̥ra:j	*bra:j	--	pi:a:j ²¹	pr̥ay	pray	X
62	‘tiger’	*rəwa:j	*r̥əva:j	*rəva:j	kawa:j	lawa:j ²¹	r̥ôvay	rôvéay	X
63	‘liver’	*təlɔ:m	*t̥lɔ:m	--	lɔ:m	luəm ²¹	laûm	lôm	X
64	‘ashes’	*bəh	*b̥əh	--	p ^h əh	pəh ³⁴⁴	p̥ô-ûh	pâûs	?

The R3 items generally behave regularly in terms of breathy phonation and historical voicing values; however, nearly half the examples have [-high] vowels yet are not creaky and the reasons are not clear. Several items are ambiguous as to register:

- #56 ‘weep’ is recorded in Kasong as *ja:m*⁴⁵³ suggesting creak, but not so in Chong sources.
- #63 ‘ashes’ is recorded for Kasong as *paɽu:*³³ which would belong in R1. Suoy and KPT forms agree with Chong in the coda but their double vowel spellings suggest creak. My best suggestion is that there is a confusion of multiple roots.

This leaves us with seven items that are breathy but not creaky, despite having low vowels. There are no obvious phonological patterns of conditioning relating specifically to segmental values.

¹² An anonymous reviewer has suggested that in etyma such as ‘bird’ and ‘cooked’ some vowel raising may have been conditioned by their palatal onsets. This is worth further consideration, especially in terms of further refinement to the proto-Austroasiatic reconstruction, such that we have more confidence regarding the values of all segments in relevant etyma.

Table 11: R4 – Breathy-Creaky Register Etyma

	Gloss	pAA	pPearic revised	pPearic Headley	Chong	Samre	Suoy	KPT	Notes
65	‘two’	*ba:r	*ba:ʔr	--	pʰa:ʔj	pa:l ⁴⁵¹	paar	pierrl	✓
66	‘far’	*cɲa:j	*cɲa:ʔj	--	kʰəna:ʔj	sana:j ⁴⁵¹	--	chhngay	✓
67	‘thorn’	*crlaʔ	*jrla:ʔ	*jrəlaʔ	crli:ʔ	kaləʔ	salā:	chrâlêa	✓
68	‘rain’	*gmaʔ	*gma:ʔ	*gmaʔ	kəmaʔ	kama: ⁴⁵¹	kmāk	khméa	✓
69	‘sour’	*juʔ~ *jɔʔ	*jɔ:ʔ	--	cʰoʔ	cɔ: ⁴⁵¹	chok	--	✓
70	‘one’	*mo:j~ muəj	*mo:ʔj	--	mu:ʔj	mu:j ⁴⁵¹	maûy	muoy	✓
71	‘sweet’	*ɲa:m	*ɲa:ʔm	--	ɲa:ʔm	ɲa:m ⁴⁵¹	--	--	✓
72	‘a fly’	*rɔ:j	*rɔ:ʔj	*rɔ:j	rɔ:ʔj	ɭuəj ³³²	rôy	rôây	✓
73	‘thin/lean’	*rgəj	*rəgi:ʔ	*rəgiʔ	kiʔ	laki: ⁴⁵¹	rôkik	rôkî	?

The R4 items all appear to behave regularly in terms of creak and breathiness with the possible exception of #73 ‘thin/lean’. As with #43, #44 this is reconstructed with a high vowel based on the dominant reflexes, but this may not be appropriate. Shorto (2006 #1451) reconstructs a proto-rhyme *əj based on reflexes such as Sre, Chrau (Bahnaric) *rəgaj*, Vietnamese *gây*, Nancowry *ʔikēj*. Thus, the proto-Pearic rhyme may have been *əj or *ɛj* in which case the register fits our hypothesis. If this is correct, #73 passes our test.

4 Discussion of problematic register values

The overall impression is that R2 and R4 etyma have register values that are overwhelmingly consistent with our hypothesis. However, approximately 1/3 of [-high] vowel etyma do not show creak. In terms of our proposed rule the exceptions flow predominantly in one direction: failure to acquire/retain creak, the older of the two marked registers. No clear phonological principle is evident that may explain this situation as an outcome of regular sound change, and not wanting to posit segments or structures for which there is no direct evidence, a different explanation is required.

With this in mind we return to our earlier discussion of why among Chongic lects the Germanic shift is not found in all cases of voiceless stop onsets. Ferlus (2011) credits Khmer influences, not necessarily by direct borrowing, but that pronunciations may have been influenced by mimicking more prestigious Khmer forms (or other contact languages). A contemporary example in English would be mimicry of the American ‘Tuesday’ without the palatalised onset of Received Pronunciation, creating lexical violations of an otherwise regular rule. This is purely speculative, but it could be that a parallel social situation conditioned lexical exceptions to the otherwise general rule—proposed in this paper—that Pearic syllables with [-high] vowels became creaky (subject to the blocking rule regarding glottal segments in onsets).

This proposal is clearly difficult to test. Hypothetically, one could examine Pearic lexical innovations, assuming that lacking parallels in other languages of the area they would be less affected, but this runs into problems. Phonological rules apply at different times, and it is difficult or impossible to know if a specific lexeme was innovated before, during, or after specific rules applied. Another issue is that specific Pearic lexical innovations may be borrowings from languages that no longer exist, a circumstance that is quite likely given the leveling effect of the Khmer empire. A possible strategy is to compile apparently older borrowings (such as from Old Khmer) and examine their phonological patterns, but this is yet to be attempted.

In conclusion, while much potential work remains to be done, this paper has proposed a mechanism to explain the presence of creak in Pearic, and outlines revisions to Headley’s (1985) reconstruction

that should facilitate substantial qualitative and quantitative improvement to the proto-lexicon going forward. It is suggested that creaky registers in AA languages—alone or in combination with breathy registers—arise spontaneously from features inherent in the vocalism and are generally unlikely to be directly inherited from proto-Austroasiatic.

References

- Alves, Mark. 2006. *A Grammar of Pacoh: a Mon-Khmer language of the central highlands of Vietnam*. Canberra, Pacific Linguistics.
- Baradat, R. 1941. Les dialectes des tribus sâmrê. ms. l'Ecole Française d'Extrême-Orient, Paris.
- Choosi, Isara. 2007. Investigating Contact-Induced Language Change: cases of Chung (Saoch) in Thailand and Cambodia: PhD thesis Mahidol University (Thailand).
- Diffloth, Gérard. 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouïque. *Mon-Khmer Studies* 11:47-82.
- Diffloth, Gérard. 1984. *The Dvaravati-Old Mon Language and Nyah Kur* (Monic Language Studies 1). Bangkok, Chulalongkorn University Printing House.
- Diffloth, Gérard. 1986. Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies* 15:139-154.
- Diffloth, Gérard. 2012. The four registers of Pearic. Paper presented at the SEALS22 meeting, Argay, France. 12-page handout.
- Ferlus, Michel. 1974d. La langue Ong, mutations consonantiques et transphonologisations. *Asie du Sud-Est et Monde Insulindien* 5.1:113-121.
- Ferlus, Michel. 2007. *Lexique de racines Proto Viet-Muong (Proto Vietic Lexicon)*. ms. See <http://sealang.net/monkhmer/>
- Ferlus, Michel. 2011. 'Toward Proto-Pearic: Problems and Historical Implications.' In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3, part 1*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. 38-51.
- Gehrmann, Ryan. 2015. Vowel Height and Register Assignment in Katuic. *Journal of the Southeast Asian Linguistics Society* 8:56-70
- Gregerson, Kenneth J. 1976. "Tongue-Root and Register in Mon-Khmer." Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies, Two Volumes*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13), I:323-369
- Gregerson, Kenneth J., Kenneth D. Smith, and David D. Thomas. 1976. The place of Bahnar within Bahnaric. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta (eds.), *Austroasiatic studies, part I*. Oceanic Linguistics Special Publication, 13. Honolulu: University of Hawaii Press. 371-406.
- Haudricourt, André-Georges. 1965. Mutation consonantique en Mon-Khmer. *Bulletin de la Société Linguistique de Paris* 60:160-72.
- Headley, Robert K. Jr. 1977. A Pearic vocabulary. *Mon-Khmer Studies* 6:69-150.
- Headley, Robert K. Jr. 1978. An English-Pearic Vocabulary. *Mon-Khmer Studies* 7:61-94.
- Headley, Robert K. Jr. 1985. Proto-Pearic and the Classification of Pearic. In *Southeast Asian Linguistic Studies presented to André-G Haudricourt*. Ratanakul, Thomas and Premsirat eds., Bangkok, Mahidol University. 428-478.
- Huffman, Franklin E. 1976. The register problem in fifteen Mon-Khmer languages. Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13). Part I:575-589.
- Huffman, Franklin E. 1985. The phonology of Chong. *Southeast Asian Linguistic Studies presented to André-G Haudricourt*. Ratanakul, Thmas and Premsirat eds., Bangkok, Mahidol University. 355-88.

- Kuang, Jiangjing. 2013. Phonation in Tonal Contrasts. PhD thesis. University of California at Los Angeles.
- Luang-Thongkum, Theraphan. 1991. An Instrumental Study of Chong Register. In *Austroasiatic Languages. Essays in honour of H.L. Shorto*. Edited by J.H.C.S. Davidson. School of Oriental and African Studies, University of London:141-160.
- Martin, Marie A. 1974a. Remarques générales sur les dialectes pear. *Asie du Sud-Est et Monde Insulindien* 5.1: 25-37.
- Martin, Marie A. 1974b. Esquisse phonologique du sɔmree. *Asie du Sud-Est et Monde Insulindien* 5.1:97-106.
- Ostapirat, Weera. 2009. *Early and Modern Pearic Registers/Tones*. Paper presented at the 4th International Conference on Austroasiatic Linguistics, Mahidol University, Oct. 29-30 2009.
- Pelliot, Paul. 1902. Mémoires sur les coutumes du Cambodge. *Bulletin de l'École Française d'Étrême-Orient* 2.2:123-177. Traduction du texte de Tcheou Ta-Kouan.
- Ploykaew, Pornsawan. 2001. Samre grammar. PhD thesis. *Institute of Languages and Culture for Rural Development*, Mahidol University.
- Premsrirat, Suwilai and Nattamon Rojanakul. 2014. Chong. In Mathias Jenny and Paul Sidwell (eds.). 2014. *The handbook of Austroasiatic languages* (2 vols.). Leiden and Boston, Brill. Pp. 603-640.
- Premsrirat, Suwilai. 2008. Chong Dictionary. Center for Revitalization of Endangered languages, Institute of Language and Culture for Rural Development.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics.
- Sidwell, Paul and Felix Rau. 2014. *Austroasiatic Comparative-Historical Reconstruction: an overview*. In Mathias Jenny & Paul Sidwell (eds.) *The handbook of Austroasiatic languages*. Leiden, Boston: Brill. 221–362.
- Sidwell, Paul. 1998. A Reconstruction of Proto-Bahnaric. PhD thesis, University of Melbourne.
- Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich, Lincom Europa.
- Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas, Language Data Series, Summer Institute of Linguistics.
- Suphanphaiboon, Surekha. 1982. The phonological system of the Chong language, Muban Takhianthong, Tambon Thakhian-thong, Amphoe Makham, Changwat Chanuri. MA thesis, Srinakharinwirot University Prasarnmit Campus.
- Thyngkham, Noppawan. 2003. The phonology of Kasong at Khlong Saeng Village, Danchumphon sub-District, Bo Rai District, Trat Province, MA thesis, Mahidol University.
- Ungsitipoonporn, Siripen. 2001. *A phonological comparison between Khlong Phlu Chong and Wangkraphrae Chong*. MA thesis. Institute of Language and Culture for Rural Development, Mahidol University, Thailand.
- Wurm, S.A., and Shirô Hattori (eds.). 1983. Language atlas of the Pacific area, Part II: Japan area, Taiwan (Formosa), Philippines, Mainland and Insular South-East Asia. Canberra, Australian National University.

FLAGGING AND INDEXING IN AUSTROASIATIC: A TYPOLOGICAL OVERVIEW

Tobias WEBER

Christian-Albrechts-Universität zu Kiel
<tweber AT isfas DOT uni-kiel DOT de>

Abstract

Argument flagging and indexing are largely absent in the majority of Austroasiatic (AA) languages, especially in those spoken in (core) Mainland Southeast Asia. Still, flagging and indexing are found in subgroups outside that area, and the patterns are diverse. The present article provides a typological overview of these patterns and also shows that specific patterns are often restricted to some syntactic, semantic or pragmatic environment, i.e., many AA languages exhibit differential flagging or indexing. The article also seeks to establish how this diversity evolved in the different subgroups. It finds that many patterns are the result of language contact with non-AA languages. This includes flagging in Munda and Khasian and – albeit only partly – indexing in the subgroups where it is found. However, some patterns of different subgroups appear to be very old and should be reconstructed for Proto-AA, especially some flagging patterns found in Palaungic and Nicobarese languages. Furthermore, some patterns are neither very old nor can they be readily accounted for in terms of language contact. This suggests that these patterns are the result of language-internal developments (e.g., flagging patterns in Aslian). The data for this study are taken from grammars and other language descriptions, for a sample representing 32 AA languages of all subgroups that feature flagging or indexing and some non-AA languages that are in contact with AA languages. However, many AA languages of Mainland Southeast Asia are not included since they lack flagging or indexing and are thus not relevant for the present purpose.

Keywords: Austroasiatic, grammatical relations, case, flagging, indexing, typology, language contact, diachrony

ISO 639-3 codes: aml, bfw, caq, cdz, cwg, gaq, gbj, hoc, jhi, jun, kfq, kha, khm, khr, lwl, mnw, ncb, pac, pbv, pcj, pll, pce, prk, rbb, sat, scb, srb, sza, szc, tnz, tth, unr, zng

1 Introduction

The majority of Austroasiatic (henceforth AA) languages are spoken in (core) Mainland Southeast Asia, where grammatical relations are primarily established by word order (cf. Jenny et al. 2015), as the following example from Khmer illustrates:

- (1) Khmer (Jacob 1968:59)
mda:j thvɿ: mho:p.
mother make food
'Mother is making the meal.'

In this example, the argument preceding the verb ('mother') is the A argument (the more agent-like argument), whereas the argument following the verb ('food') is the P argument (the more patient-like argument). The arguments themselves exhibit no morphological marking indicating the grammatical relations, and the verb lacks any person marking. This is different in the following example from Semelai:

- (2) Semelai (Kruspe 2004:255)
ki=bukɔʔ la=kniek hn=pintuʔ.
 3A=open ERG=husband ACC=door¹
 ‘The husband opened the door.’

In this example, grammatical relations are coded in two ways. On the one hand, the arguments bear case clitics (or flags): The A argument (‘husband’) is marked by the ergative clitic *la=*, while the P argument (‘door’) is marked by the accusative clitic *hn=*. On the other hand, the A argument is cross-referenced by the bound clitic (or index) *ki=* on the verb.

The present article investigates how arguments are flagged and indexed in AA languages. And while most AA languages (such as Khmer) are not relevant for this purpose since neither argument flagging nor indexing is found, the paper shows that flagging and/or indexing is still found in a good number of AA languages. Moreover, the patterns that are found are very diverse, which begs the question as to how they evolved. Therefore, the paper also seeks to find explanations as to (i) whether the patterns were borrowed from languages which are or were in contact with AA languages, (ii) whether they were inherited from older stages of AA or even Proto-AA, or (iii) whether some other explanation is needed to account for them; for instance, the patterns can be the result of language-internal developments.

The structure of the paper is as follows: Section 2 provides some theoretical background concerning argument marking as it is defined for the present purpose, The following two sections investigate the different flagging (Section 3) and indexing (Section 4) patterns found in different languages and subgroups of AA. For more practical purposes, the investigations and diachronic discussions of these sections are subdivided into subsections concerning (i) Munda (spoken in central-eastern India), (ii) Khasian (spoken in Northeast India and Bangladesh) and Palaungic (spoken in Myanmar and adjacent areas of China, Laos and Thailand), (iii) Mainland Southeast Asian AA (without Palaungic and Aslian), and (iv) Aslian (spoken on the Malay Peninsula) and Nicobarese languages (spoken on the Nicobar Islands). Finally, Section 5 provides a summary of the findings and a conclusion.

The data for this study are taken from available grammars and other descriptions for a sample consisting of 32 AA languages (out of a total of 158 AA languages listed in the Glottolog, Hammarström et al. 2022) representing all subgroups that exhibit flagging and/or indexing (see Appendix). The sample was compiled with the aim of capturing as much of the variation of the flagging and indexing patterns in AA as possible. Therefore, Munda languages with their complex patterns are overrepresented, whereas languages belonging to the Bahnaric, Khmuic and Pearic subgroups, all of which are spoken in Mainland Southeast Asia, have not been included at all since they are expected to lack flagging (except possibly on some G arguments) and indexing and are therefore not expected to exhibit patterns

¹ The glosses follow the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>) and are sometimes adapted from the sources for consistency. In this particular example, I use ERG since it flags A (but not S or P) and ACC since it flags P (but not S or A). The full list of abbreviations used in this paper are: 1, 2, 3=first, second, third person; A=the more agent-like argument of bivalent predicates or the most agent-like argument of trivalent predicates; AA=Austroasiatic; ACC=accusative; ACT=active; ADD=additive (focus); ANA=anaphoric; ANIM=animate; APPL=applicative; BEN=benefactive; CAUS=causative; CLF=classifier; COP=copula; DAT=dative; DEF=definite; DU=dual; ERG=ergative; EXCL=exclusive; F=feminine; FIN=finite; FUT=future; G=the goal-/recipient-like argument of trivalent predicates; GEN=genitive; HAB=habitual; HUM=human; IMPERS=impersonal; INAN=inanimate; INCL=inclusive; IND=indicative; INGR=ingressive; INS=instrumental; IPFV=imperfective; ITR=intransitive; LOC=locative; M=male; MID=middle (voice); N=neuter; NEG=negation; NONVIS=nonvisual; NPST=nonpast; OBJ=object; OBL=oblique; OPT=optative; P=the more patient-like arguments of bivalent predicates; PASS=passive; POL=polite; PFV=perfective; PL=plural; PRF=perfect; PROX=proximate; PRS=present; PST=past; QUAL=qualitative (predication); SBJ=subject; SG=singular; T=the third argument of trivalent predicates (besides the A and the G argument); TAM=tense/aspect/mood; TOP=topic.

not already found in other AA languages.² Furthermore, descriptions of some non-AA languages which are in contact with AA languages were also considered, which was necessary in order to account for the patterns that evolved through language contact.

2 Definitions and some notes on the theoretical background

Arguments are defined in terms of semantic entailments (cf. e.g., Dowty 1991) rather than syntactic properties since these latter are often subject to language-specific properties (and hence criteria). Each argument has a generalized semantic role, which is dependent on the valence and the semantics of the predicate. The following generalized semantic roles are distinguished (cf. Bickel 2011, Witzlack-Makarevich 2011:57):

(3) Generalized semantic roles:

S: sole argument of monovalent predicates

A: the more agent-like argument of bivalent predicates or the most agent-like argument of trivalent predicates

P: the more patient-like arguments of bivalent predicates

G: the goal-/recipient-like argument of trivalent predicates

T: the third argument of trivalent predicates (besides the A and the G argument)

The present paper investigates argument marking in two different domains: flagging and indexing. Flagging comprises the coding of grammatical relations and semantic roles on noun phrases by case markers or adpositions. The morphemes in question are called “flags” and defined by Haspelmath (2019:96) as follows:

(4) Flag:

A flag is a bound form that occurs on a nominal and that indicates the semantic or syntactic role of the nominal with respect to a verb (in a clause) or with respect to a possessed noun (in a complex nominal). (Haspelmath 2019:96)

While case markers and adpositions can clearly be distinguished in many Indo-European languages, it is much less clear how to tell them apart from a cross-linguistic comparative perspective (cf. Haspelmath 2019), and this also holds for AA languages. In order to avoid an arbitrary decision on which markers should be included and which should not, the term flagging (and flags) is used in the present paper.

Indexing is the coding of arguments by bound person markers on the verb (or more generally on the predicate). The morphemes in questions are called “(person) indexes” and can be defined in the following way:

² Mainland Southeast Asian languages (especially east of Myanmar) are well known for their isolating and analytic morphology (cf. e.g., Enfield 2005: 187) and thus largely lack flagging and indexing. Four AA languages of this type, namely Eastern Lawa, Wa (both Palaungic), Khmer (Khmeric) and Mon (Monic), have still been included in the sample since absence of flagging and indexing also contributes to the variation; some other languages, namely Pacoh (Katuic), Ruc (Vietic) and Mang (Mangic) have also been included since they exhibit some morphological case marking (and Pacoh also exhibits some indexing), which is unusual for the area. Some more AA languages such as Ho (Munda) and Taoih (Katuic) are mentioned in the paper but have not been studied thoroughly enough to be included in the sample.

(5) Person index (=bound person marker):

A person index is a bound form denoting a speech role or a highly accessible third person referent that occurs on a verb (or in second position) to indicate a verb's argument, or on a noun to indicate its possessor. (Haspelmath 2019:96)

I largely follow this definition in this paper except that I also include bound person forms that are neither on the verb nor in second position but occur on another constituent.

The term “indexing” has the advantage over the related more traditional terms such as “agreement” and “pronoun” that it does not make problematic theoretical assumptions: “Agreement” suggests that a cooccurring noun phrase, a controller, is present, either overtly or at least virtually (which generative grammarians call “pro drop”). This view is Anglocentric and thus does not take into account the situation in many languages, where such controllers are usually not present (or cannot be present), and there is no evidence that they were “dropped”. The term “index” is open to the question as to whether or not a coreferential noun phrase is present together with the bound person marker. Indexes, especially the ones that do not occur together with a coreferential noun phrase, have also been called “bound pronouns”, but then the question arises how they are different from “agreement”. For a much more detailed discussion in favor of the term “index” see Haspelmath (2013; 2019).

Languages that have some sort of argument marking, either flagging or indexing (or both), usually have restrictions as to the environment in which particular markers occur. That is, they exhibit differential argument marking (differential argument flagging or differential argument indexing), which can broadly be defined as follows:

(6) Differential argument marking:

Any kind of situation where an argument of a predicate bearing the same generalized semantic argument role may be coded in different ways, depending on factors other than the argument role itself, and which is not licensed by diathesis alternations. (Witzlack-Makarevich & Seržant 2018:3)

The following two examples from Gutob illustrate differential flagging of the P argument:

(7) Gutob (Judith Voß, p.c.)

a. *lai maʔ dɔŋ-to=nen*
 rice curry cook-HAB=PL
 ‘They cook rice and curry.’

b. *bai=nen=lai laj akuɔ-tu*
 brother=PL=OBJ who hold.baby-ACT.NPST
 ‘Who will hold the brothers?’

In example (7a), the inanimate P argument (‘curry’) lacks flagging, whereas in (7b), the animate P argument is marked by the object flag =*lai*. Differential P flagging in Gutob is therefore conditioned by animacy.

Animacy is just one factor that can condition differential argument marking. Conditioning factors can broadly be grouped into three classes:

- a) Referential properties of the arguments, such as person, lexical class (e.g., pronouns vs. nouns), animacy, definiteness, information-structural properties (topicality, focality);
- b) Predicate classes, defined by shared argument structure
- c) Clause properties, such as TAM categories, clause types (e.g., main vs. different types of subordinate clauses).

In different AA languages, most of these factors are found, their distribution is outlined in Sections 3 (on flagging) and 4 (on indexing).

3 Flagging

3.1 Munda

Flagging is widespread in the predominantly verb-final Munda languages, and the patterns are very diverse (for a short overview see e.g., Anderson 2015:382–384). This section merely intends to provide a summary of the patterns, since a more comprehensive account would be beyond the scope of the paper.

S and A arguments usually lack flagging in Munda. However, it can still be found in some instances. First, like in many other languages of South Asia, experiencers (which are either S arguments of monovalent predicates or A arguments of bivalent predicates) sometimes receive flagging, which is illustrated by the following examples:

- (8) Korku (Zide 2008:289)
ij̄n-èn raram-bà
 1SG-LOC feel.cold-FIN.NPST
 ‘I will be/feel cold.’
- (9) Juang (Patnaik 2008:529)
aj̄n-te bəŋɔ-bə ɔn-gɔr-te betəŋ-ki-ŋ
 1SG-DAT forest-LOC go-PFV-PRF fear-PRS-1
 ‘I fear going alone to the forest.’

In (8), the experiencer argument (S) of the monovalent predicate ‘feel cold’ is marked by the locative, and in (9), the experiencer argument (A) of the bivalent predicate ‘fear’ is marked by the dative/oblique.³

The possessors of predicative possessive constructions, analyzed as A arguments, also often receive flagging in Munda languages. In Korku, the locative is employed:

- (10) Korku (Zide 2008:271)
dij̄-èn uphùn kon-ku ʈàkha-ku
 3-LOC four son-PL LOC.COP-3PL
 ‘He has four sons.’

In Kera’ Mundari, the possessor is marked by the genitive or the accusative:

- (11) Kera’ Mundari (Kobayashi & Murmu 2008:185)
- a. *aj̄n-a? bahut kami mena-?*
 1SG-GEN much work exist-IND.INAN
- b. *aj̄n-ke bahut kami hen-a?*
 1SG-ACC much work exist-IND.INAN
 both: ‘I have a lot of work.’

In Kharia, two case markers can cooccur in predicative alienable possession: the genitive and the oblique (otherwise marking Gs, definite Ps and Ts, and some obliques and adjuncts):

³ Note that in this example, the dative argument is indexed on the verb by an index also employed for P (rather than A) arguments, which usually does not occur in Juang (Patnaik 2008:529).

- (12) Kharia (Peterson 2011:148)
ɪn=aʔ=te saykal ayiʔj.
 1SG=GEN=OBL bicycle QUAL.PRS
 ‘I have a bicycle.’

Ergative flagging, which is common in western South Asia, is rare in eastern India (cf. Peterson 2017). However, it can still be found in single instances. Santali occasionally employs the instrumental to flag A (but not S or P) arguments:

- (13) Santali (Neukom 2001:27)
abo-te raj-do baŋ=bon bachao-ko-a.
 1PL.INCL-INS king-TOP NEG=1PL.INCL.SBJ choose-3PL.OBJ-IND
 ‘We shall not choose a king by ourselves.’

Neukom (2001:27) attributes the use of the instrumental in such instances to the “emphasis of subject”, which suggests that the use is conditioned by some pragmatic factors, which still need to be established.

Agents in passive constructions are also analyzed as A arguments here. Like in many other languages, they also receive flagging in Munda languages. In Santali, they are marked by the dative, and the verb exhibits middle voice marking:

- (14) Santali (Neukom 2001:25)
kaɖa-tʰɛn=e rɔk'-en-a.
 buffalo-DAT=3SG.SBJ gore-PST.MID-IND
 ‘He was gored by a buffalo.’

In Mundari, agents of passive constructions are marked by the instrumental, and the verb exhibits intransitive marking:

- (15) Mundari (Osada 2008:120)
maŋɖi-te hon-ko=ko bisi-ja-n-a.
 food-INS child-PL=3PL.SBJ poison-INGR-ITR-IND
 ‘The children have been poisoned by the food.’

But while S and A arguments usually lack flagging except in the environments just mentioned, P, T and G arguments receive flagging more often. A minority of Munda languages such as Ho generally lacks flagging for P and T arguments (Anderson et al. 2008:205). But most Munda languages have differential object (usually P and T, i.e., indirective alignment) flagging, i.e., a subset of P and T arguments receive flagging:

- (16) Remo (Anderson & Harrison 2008a:570–571)
- a. *nij nsuɾaʔ sum-t-iŋ*
 1SG banana eat-NPST-1
 ‘I eat bananas.’
- b. *remo a-gisiŋ ju-to*
 man OBJ-chicken see-NPST.1
 ‘The man sees the chicken.’

In (16a), the P argument receives no flagging, whereas the P argument of (16b) is marked by the prefix *a-*. In Remo, definiteness, animacy and predicate class are the relevant factors that condition the flagging of P and T (Anderson & Harrison 2008a:570). The exact factors vary from language to

language. In Juang, for instance, animacy is relevant but definiteness is not (Patnaik 2008:515), whereas in Kharia, definiteness is relevant while animacy is not: The oblique case enclitic *=te* marks definite P arguments, even inanimate ones (Peterson 2011:142–143), such as ‘water’ in the following example:

- (17) Kharia (Peterson 2011:144)
laʔ porha=jo col=ki ro yo=te, ho maʔa boʔ=te
 then Porha=ADD go=MID.PST and see=ACT.PRS that cave place=OBL
ɖam=ki ɖaʔ=te yo=yoʔ.
 arrive=MID.PST water=OBL see=ACT.PST
 ‘Then Porha also went and looks, he arrived at the cave and saw the water.’

However, the use of the oblique marker is also subject to speaker-specific variation, with some speakers omitting it in the case of mass nouns and often also on T arguments if an oblique-marked G argument is also present (Peterson 2011:144).

Munda languages also differ with respect to the position of the flag on P (and also T and G) arguments. Remo, for instance, just illustrated in example (16), employs a vocalic prefix. In some other South Munda languages such as Gutob, this vocalic prefix is restricted to pronouns, whereas nouns are marked by a suffix or enclitic, often borrowed from Indo-Aryan. Most North Munda languages only have a suffix or enclitic but no vocalic prefix (e.g., Kharia, just illustrated in example (17)). In Gutob (Judith Voß, p.c.) and Plains Gta’, pronominal P, T and G arguments can be marked by a combination of the vocalic prefix and the enclitic that is employed on nouns:

- (18) Plains Gta’ (Anderson & Gomango 2016:107)
a-nae=ke kmæ=hiŋ a-miaʔ
 OBJ-1PL=OBJ DEF.3=PL NEG-know
 ‘They don’t know us.’

Gorum employs the vocalic prefix and a suffix on both pronouns and nouns. The suffix, however, is optional on pronouns and inanimate nouns but obligatory on animate nouns (Anderson & Rau 2008:388).

Most Munda languages which flag a subset of P and T arguments also employ that same flagging on G, i.e., they exhibit neutral alignment. In the following example from Kharia, the T and the G arguments are marked in the same way, namely by the oblique enclitic *=te*:

- (19) Kharia (Peterson 2011:368)
ho=kaʔ am=te pothi=te ter=oʔ.
 that=SG.HUM 2SG=OBL book=OBL give=ACT.PST
 ‘S/he gave you the book.’

Example (17) already showed that *=te* also marks definite P arguments.

Some Munda languages also feature G arguments without flagging. Consider the following examples from Santali:

- (20) Santali (Neukom 2001:121)
 a. *maejiu-then gidraʔ=ɲ em-ked-e-a.*
 woman-DAT child=1SG.SBJ give-PST.ACT-3SG.OBJ-IND
 b. *gidraʔ-do maejiu=ijɲ em-ad-e-a.*
 child-TOP woman=1SG.SBJ give-APPL.PST.ACT-3SG.OBJ-IND
 both: ‘I gave the child to the woman.’

In example (20a), the G argument (‘woman’) carries the dative marker. In (20b), however, where the verb is in the applicative, the G argument lacks case marking. Note also that Santali is one of the few Munda languages that never employ flagging on P and T arguments (Neukom 2001:23–24).

Overall, the flagging patterns of the Munda languages fit in the geographical picture: The neighboring Indo-Aryan languages (and also the Dravidian languages) also have the following patterns (cf. also Peterson 2017):

- i. S and A arguments lack flagging, except some experiencers and possessors;
- ii. there is differential flagging of P and T arguments, conditioned mainly by referential properties of those arguments;
- iii. G arguments usually receive flagging, and this is identical to the flagging that is employed on the subset of P and T arguments that receive flagging.

In some of the Munda languages, the suffix or enclitic flagging of P, T and G arguments is borrowed from Indo-Aryan languages. Some Munda languages have a vocalic prefix as flag for P, T and G arguments. This prefix was presumably restricted to pronouns in Proto-Munda – or even in Proto-AA, since similar prefixes can be found on pronominal G arguments in some Katuic languages such as Taoih and Pacoh (see Solntseva 1996:34, Alves 2006:30; 2015:541–542, and Anderson & Rau 2008:388).

3.2 Khasian and Palaungic

Khasian and Palaungic languages have both verb-initial and verb-medial word orders. The distribution varies. Flagging can be found but is less common than in Munda.

Unlike other Khasian languages such as Pnar (Ring 2015:23), Standard Khasi is predominantly verb-medial (Nagaraja 2015:1168). Ps are optionally⁴ marked by the preposition *ja* (ACC):

(21) Standard Khasi (Nagaraja 2015:1168–1169, Rabel 1961:77)

- a. *ɲi je:jt ja ki=lum ki=wa?*
 1PL love ACC PL=mountain PL=river
 ‘We love the mountains, the rivers’
- b. *ɲa jo?i a:r ɲut khinna? fɪnraŋ*
 1SG see two CLF child male
 ‘I see two boys’

T arguments (like ‘book’ in the following example) are also marked by *ja*, while G arguments (like ‘1SG’ in the following example) are marked by *ha*:

(22) Standard Khasi (Nagaraja 2015:1170)

- u la a:j ja ka=kot ha ɲa*
 3SG.M PST give ACC F=book DAT 1SG
 ‘He gave the book to me.’

Differential P flagging is also found in Mnar, where it is conditioned by animacy:

(23) Mnar (Rymbai 2018:12)

- a. *u ʝɒn ieid=u ha ga meri*
 M John love=3SG.M DAT F Mary
 ‘John loves Mary.’

⁴ The factors that determine the differential use of the ACC flag still have to be established.

- b. *ga meri tieŋ haicbi:m=ka*
 F Mary cook food=3SG.F
 ‘Mary cooks food.’

Animate P arguments, as ‘Mary’ in (23a), are marked by *ha*, while inanimate P arguments lack flagging, as ‘food’ in (23b). *Ha* also marks G arguments, as ‘Ban’ in the following example:

- (24) Mnar (Rymbai 2018:13)
u ɟɔp leɔ ai ga bɔl ha u ban
 M Jop PST give F ball DAT M Ban
 ‘Jop gave the ball to Ban.’

Thus, in contrast to Standard Khasi, Mnar does not have an accusative flag that is different from the dative flag.

Pnar also has differential P (and T) flagging, where it is conditioned by lexical class and number: Singular pronouns (except third person singular neuter) have distinct accusative forms for P and T arguments, whereas plural pronouns and all nouns do not:

- (25) Personal pronouns of Pnar (Ring 2015:96)
- | | NOM.SG | ACC.SG | PL |
|---|---|---|-----------------------|
| 1 | <i>ɔ</i> | <i>ŋa</i> | <i>i</i> |
| 2 | <i>mi</i> (M), <i>p^hɔ</i> (F) | <i>me</i> (M), <i>p^ha</i> (F) | <i>p^hi</i> |
| 3 | <i>u</i> (M), <i>kɔ</i> (F), <i>i</i> (N) | <i>o</i> (M), <i>ka</i> (F), <i>i</i> (N) | <i>ki</i> |

G arguments are usually marked by the preposition *ja*, which governs the accusative (when available):

- (26) Pnar (Ring 2015:228)
e kɔ ka=hukum ja o.
 give 3SG.F.NOM F=command BEN 3SG.M.ACC
 ‘She gave the command to him.’

Palaungic languages also have both verb-medial and verb-initial structures (cf. Jenny 2015 and Lee 2023, in this issue). S and A arguments usually lack flagging. This often also holds for P and T arguments. However, P and T arguments can receive flagging under certain circumstances in West Palaungic languages. The following two examples from Shwe involving the verb ‘to keep’ show that the P argument lacks flagging in the verb-medial structure of (27a) but has flagging in verb-initial subordinate clause of (27b):

- (27) Shwe Palaung (Mak 2012:26, 132)
- a. *ɔ di ka ɔun ci kwən.hɔj din ta hɔ*
 1SG FUT NEG keep POL little.water.snail that OBL palace
 ‘I will not keep the little water snail in palace [sic].’
- b. *ɔ ka nəp ha ɔun ɔ ta ɔn*
 1SG NEG know where keep 1SG OBL 3SG
 ‘I can’t remember where I keep it.’

A similar situation is found in Rumai, where P arguments also receive flagging in verb-initial structures in subordinate contexts:

- (28) Rumai (Weymuth 2019)
di khə̀lɔh jê tɿʔ kâ dɿ
 place cross 1PL.EXCL OBL car ANA
 ‘place, where we cross the cars’

In Rucing, P arguments in verb-initial subordinate clauses occasionally also receive flagging, as shown in the following example:

- (29) Rucing (Wei-Wei Lee, p.c.)
sin dɔ vɛ:ŋ ʔəmae ʔan di ʔan, naoh ʔan ʔaɔ dak^hui
 since when throw.away husband 3SG OBL 3SG mind 3SG NEG be.well
 ‘Since her husband left her, she has been unhappy.’

However, the flagging of the P argument also appears to depend on animacy and other factors, which still need to be established. In verb-medial main clauses, the flagging of the P argument would be ungrammatical (Wei-Wei Lee, p.c.).

Waic languages also have verb-initial clauses but do not seem to employ flagging on P and T arguments (cf. e.g., Seng Mai 2012:21–27 for Wa or Blok 2013:24–30 for Eastern Lawa).

In West Palaungic languages, P arguments of certain verbs also receive flagging in main clauses with verb-medial word order. Examples for such verbs include ‘remember’ in Shwe (30) or ‘help’ in Rumai (31):⁵

- (30) Shwe (Mak 2012:45)
nwə:ɿuh ta ʔɔ ci
 remember OBL 1SG POL
 ‘Remember me (please).’
- (31) Rumai (Weymuth 2019)
ʔê gi:j-kəvê tɿʔ kê
 1PL.INCL IPFV-help OBL 3PL
 ‘We help them.’

G is marked by the same preposition as the P argument in the examples just mentioned, namely *ta* in Shwe, *tɿʔ* in Rumai and *di* in Rucing, while the T argument is unmarked:

- (32) Shwe (Mak 2012:30)
hɔj.blɔ dɛh k^hrɪr ta ʔan
 white.water.snail give gold OBL 3SG
 ‘White Water Snail gave gold to him.’
- (33) Rumai (Weymuth 2019)
ʔân dɛh kəbjok laʔ tɿʔ ʔâw
 3SG give shirt blue OBL 1SG
 ‘He gave me a blue shirt.’
- (34) Rucing (Lee 2023:45, this publication)
sɔ.sɔ di duaʔ bap di mɛ
 Zaw.Zaw FUT give book OBL 2SG
 ‘Zaw Zaw will give you a book.’

⁵ It is unclear whether these verbs share semantic features or whether they are merely ‘idiosyncrasies’.

In verb-initial subordinate clauses in Rumai, both the T and the G arguments are marked:

(35) Rumai (Weymuth 2019)

khaj dɛh ʔəw tɔʔ pôm tɔʔ bawpɦuŋ ʔəw ʔən pɔɔ khâj
 time give 1SG OBL rice OBL friend 1SG 3SG happy very
 ‘When I gave rice to my friend, he was very happy.’

Do the flagging patterns found in Khasian and Palaungic conform to the geographical picture? Khasian languages differ from their neighboring Sino-Tibetan languages in word order: The former have verb-initial and verb-medial structure, the latter are verb-final. However, the flagging patterns are fairly similar: For instance, in Bodo and Dimasa, S and A do not carry overt case marking, P flagging is restricted to animate noun phrases, which are marked by the accusative, T lacks case marking, and G is marked by the dative (Brahma 2013:182–187). In Atong, the situation is similar: A and S arguments lack overt case marking, the overt marking of P arguments depends on animacy and pragmatic factors, T lacks case marking, and G arguments are marked by the dative (van Breugel 2014:321–322, 337–340). Ergativity, which is absent in Khasian and the immediately neighboring Sino-Tibetan languages, is found in Sino-Tibetan languages further away, e.g., in Ao, where the marking of the A argument is conditioned by pragmatic factors (Coupe 2011).

Since differential P marking conditioned by referential properties such as animacy and pragmatic factors are found in both Khasian and their Sino-Tibetan neighbors but cannot readily be reconstructed for Proto-AA, it is likely that Khasian languages borrowed this pattern from Sino-Tibetan languages. Other shared patterns such as the lack of overt marking of S and A arguments and overt marking of G arguments were probably already present in Proto-AA, and therefore language contact is not needed to account for these patterns, although it might have played a role in maintaining these patterns in Khasian languages.

Palaungic languages are primarily surrounded by Tai and Sino-Tibetan languages, which have verb-medial word order and lack flagging of S, A and P arguments (G arguments can be marked). But while these patterns can also be found in Palaungic, at least West Palaungic languages additionally have verb-initial structures with overt flagging of P arguments. These additional structures cannot be explained by language contact. Furthermore, verb-initial word order is fairly uncommon, which makes a language-internal development of these structures unlikely. Therefore, the most parsimonious explanation for these structures is inheritance from (and therefore reconstruction for) Proto-AA, where P arguments in VAP structures presumably received flagging (cf. Jenny 2015, Sidwell 2019).

3.3 Mainland Southeast Asia

Flagging on S, A, P and T arguments is largely absent in AA languages of Mainland Southeast Asia (defined for the present purpose as excluding Palaungic and Aslian). This also holds for the G argument of the verb ‘give’ (and potentially a few other verbs) in Mon:

(36) Mon (Jenny 2015:585)

ʔəmè kɔ kon kwaj rəh ŋuə.
 mother give offspring sweets every day
 ‘The mother gives her child sweets every day.’

Otherwise, G arguments usually receive flagging, as in the following example, where the flag has grammaticalized from the verb ‘give’, which precedes the G argument (‘offspring’):

(37) Mon (Jenny 2015:586)

mìʔ rən kɔ kon kwaj.
 mother buy give offspring sweets
 ‘The mother bought her child sweets.’

Some AA languages employ morphological case marking on G arguments, e.g., in Ruc (Vietic; Solncev et al. 2001:73), Taoih (Solntseva 1996:34) and Pacoh (both Katuic; see Alves 2015:541–542):

- (38) Pacoh (Alves 2006:65, 67)
- a. *ʔiŋajkoh ki: tumɔ:ŋ maj*
 tomorrow 1SG meet 2SG
 ‘I will meet you tomorrow.’
- b. *ʔɔ:n ʔa-ki: fa:c ʔŋkoh*
 give DAT-1SG book that
 ‘Give me that book.’

In example (38a), the first person singular pronoun lacks flagging in A argument function. In (38b), on the other hand, it is marked by the dative prefix in G argument function. Besides pronominal A arguments, S, P and T arguments also lack flagging in Pacoh. Mang seems to be the only AA language of Mainland Southeast Asia where morphological case marking is not only attested on G arguments but also on P arguments. In example (39a), the P argument (‘cat’) is marked by the accusative case prefix *ta-*, while in example (39b), the G argument (‘chicken’) is marked by the dative prefix *ʔæŋ-*, while the T argument (‘rice’) lacks overt case marking:

- (39) Mang (Nguyễn et al. 2008:148, cited in Alves 2015:542)
- a. *əɔ^{42?} cǎp³⁵ ta-ŋo^{14?}*
 dog chase ACC-cat
 ‘The dog chases the cat.’
- b. *ʔu³²³ ʔɔ²² lam^{55?} ʔæŋ-ciɔj^{55?}*
 1SG give rice DAT-chicken
 ‘I give rice to the chicken.’

The fact that AA languages of Mainland Southeast Asia largely lack flagging on S, A, P and T arguments but can have flagging on G arguments fits in the geographical profile, as the Tai-Kadai, Sino-Tibetan (Sinitic, Karenic), Hmong-Mien and Austronesian (Chamic) languages of the area also exhibit these patterns.

3.4 Aslian and Nicobarese

Aslian languages show considerable variation with regard to flagging. Flagging is absent in some Aslian languages such as Ceq Wong or Maniq, regardless of word order (Kruspe et al. 2015:437, Wnuk 2016:99–100).

This holds for all arguments, including G arguments, as the following examples from Maniq shows, where ‘man’ lacks flagging:

- (40) Maniq (Kruspe et al. 2015:438, Wnuk 2016:99)
- ya? ʔe? ʔek kut ʔey*
 woman 3 give bottle man
 ‘The woman gave the man the bottle.’

On the other hand, flagging is found in languages belonging to different subgroups of Aslian.

Jahai shows differential S/A flagging conditioned by focality: postverbal (i.e., focal) S and A arguments (subjects) are marked by *ka=*, which coexpresses instrumental:

- (41) Jahai (Burenhult 2005:148)
ʔoʔ get ka=tmkal ka=tajiʔ
 3SG cut SBJ=man INS=knife
 ‘The man made a cut with the knife.’

Other S/A arguments do not receive flagging. This pattern is similar to the one found in Semaq Beri (Kruspe 2015:488).

Semelai has fairly different flagging patterns: Here, it is restricted to perfective clauses. Furthermore, different conditions apply for the flagging of S, A, P/T, and G: A arguments, which are restricted to animates, always receive flagging; S arguments have dedicated third person pronoun forms only, and on P/T arguments flagging is “optional”, i.e., conditioned by pragmatic factors (Kruspe 2004:258–266). G is flagged by the locative preposition *ʔen* or deictic locatives, independent of aspect (Kruspe 2004:263). Table 1 provides an overview of the distribution of the flags.

Table 1: Argument flagging in Semelai (Kruspe 2004:258, 263)

	Pronouns		Nouns	
	3 rd	1 st and 2 nd	animate	inanimate
A	<i>la=</i>		(n/a)	
S	<i>-n</i>	\emptyset		
P, T	<i>hn=</i> / \emptyset (conditioned by pragmatic factors)			
G	<i>ʔen</i> or deictic locatives			

Flagging is also found in Nicobarese languages, but the patterns still need more research. In mostly verb-initial Car, any immediately postverbal argument lacks flagging. In VPA order, the A argument also lacks flagging (as in (a)), unless it occurs in a passive sentence (as in (b)), where it is marked by the preposition *tə*:

- (42) Car (Braine 1970:145, 183)
- a. *ha-cát-ʔen=ʔək li:pəre cin*
 CAUS-be.lost-away=3SG book 1SG
 ‘I lost the book.’
- b. *ʔé:k-ə cin tə ʔané:k*
 bind-PASS 1SG OBL cord
 ‘I’m bound by the cord.’

In VAP word order, the P argument is marked by the preposition *tə*:

- (43) Car (Braine 1970:190)
ʔət kahúl-ʔ-uvə ʔəm tə ʔáʔã
 NEG cook-PFV-have 2SG OBL food
 ‘Didn’t you cook any food?’ (or ‘Don’t you have any cooked food.’)

The G argument lacks prepositional flagging, even when it is not in immediately postverbal position. Note, however, that pronominal G arguments are in the oblique case form:

- (44) Car (Braine 1970:239)
ké:ʔ-tə kanú:c cu
 take-toward pen 1SG.OBL
 ‘Give me a pen.’

The exact flagging patterns of Nancowry still have to be investigated. Based on Radhakrishnan's (1970) description and examples, the following patterns are found: S arguments either lack flagging or are marked by *ʔin* when they are postverbal and high on the animacy hierarchy⁶:

(45) Nancowry (Radhakrishnan 1970:75)

rián ʔin cã/cũã
run PROX 1SG
'I am running'

A arguments in preverbal position lack flagging. A and P arguments that are immediately postverbal often lack flagging, but when they occupy a later position, A is marked by *ʔin*, and P arguments either lack flagging, are marked by *ta* when they are not that high on the animacy hierarchy (as 'pig' in (46a)) or *tin* (etymologically *ta* + *ʔin*) when they are high on the animacy hierarchy (as '3SG.HUM' in (46b)):

(46) Nancowry (Radhakrishnan 1970:82, 83)

a. *yuána hew cã ta nət*
PST see 1SG OBL pig
'I saw the pig.'

b. *yuána hew cã tin ʔən*
PST see 1SG OBL.PROX 3SG.HUM
'I saw him.'

In passive constructions, the P argument lacks flagging or is marked by *ʔin* (or *-n*, phonologically bound to the verb stem), and the A argument is marked by instrumental *tay* (< 'hand').

(47) Nancowry (Radhakrishnan 1970:87)

ciawa=n me tay ʔən
be.called=PROX 2SG INS 3SG.HUM
'You are called by him/her.'

The G argument is marked by either *ta caʔ* (neutral with respect to the animacy hierarchy) or *tin* (high on the animacy hierarchy). The T argument is marked by *ta*, even if it is immediately postverbal as in (b) (Radhakrishnan 1970:85, 115):

(48) Nancowry (Radhakrishnan 1970:85)

a. *ʔumkuám ta nət ʔin ʔən tin cũã*
give OBL pig PROX 3SG OBL.PROX 1SG
'He gives me a pig.'

b. *ʔumkuám ta huŋʔðk ʔin ʔən ta caʔ nət*
give OBL food PROX 3SG OBL LOC pig
'He gives food to the pig.'

Do the flagging patterns found in Aslian and Nicobarese fit in the geographical picture? Verb-initial clauses and flagging are also found in some Austronesian languages of western Island Southeast Asia, especially in more remote areas such as the Barrier Islands off the coast of Sumatra (e.g., in Nias and

⁶ Following Rajasingh (2016) I gloss *ʔin* 'PROX' since it behaves more like a demonstrative (or an article) than a flag; it is not restricted to a particular argument role. However, the precise functional range still has to be investigated.

Enggano, cf. Brown 2001 and Edwards 2015 respectively), presumably retained from older stages of Austronesian, since flagging is a conservative Austronesian feature (cf. Ross 2002:35–37). Flagging can be expected to have been more widespread in the area before Malay and closely related languages that largely lack flagging came to dominate the area largely. Thus, the fact that Aslian and Nicobarese feature argument flagging suggests that they fit in the geographical picture. However, the flagging patterns and their developments differ. Nicobarese flagging – especially the flagged P argument in VAP clauses – is conservative in the AA context and bears some similarities to patterns found in West Palaungic (cf. Sidwell 2019). The conservation of these patterns might have been supported by Austronesian (or potentially also pre-Austronesian) languages with flagging. The patterns found in Aslian languages (postverbal differential S/A flagging as found in Jahai, separate flagging rules for S, A, P/T and G in perfective clauses as found in Semelai), on the other hand, are presumably an innovation: They are not very similar to the patterns found in Austronesian languages. Nias, for instance, exhibits a very rare marked-absolutive alignment for flagging (Brown 2001, Handschuh 2014), and Enggano has flagging on all arguments (Edwards 2015:75). The Aslian patterns might therefore constitute language-internal developments, but potentially supported by (Austronesian or pre-AA) contact languages which had flagging.

4 Indexing

4.1 Munda

Munda is the AA subgroup with the most complex verb morphology, and this includes indexing patterns (see e.g., Anderson & Zide 2001, Zide & Anderson 2001, Cysouw 2004, Anderson 2007: chapter 3, or Anderson 2015:390–392). Munda languages index either only the subject (S and A; e.g., Kharia, Gutob, Gta', Remo), only the object (P and T or G; Korcu; except for a few verbs such as 'eat' or 'be located', which have subject indexing, cf. Zide 2008:271), or both (e.g., Munda, Santali, Juang, Gorum, Sora).

The subject indexes are used for S and A arguments. Their position shows considerable variation across different Munda languages: Languages such as Remo exclusively employ suffixes:

(49) Subject indexes of Remo (Anderson & Harrison 2008a:589)

	SG	DU	PL
1	-niŋ	-naŋ	-naŋ
2	-no	-pa	-pe
3	-Ø, -ga	-Ø, -ga	-Ø, -ga

Languages such as Sora predominantly employ suffixes but also have circumfixes:

(50) Subject indexes of Sora (Anderson & Harrison 2008b:327)

	SG	PL
1	-ay	a-...-ay
2	-e(y)	a-...-e(y)
3	-e(y)	-ji

Languages such as Juang (Patnaik 2008:530), Gta' (Anderson 2008:719), Gorum (Anderson & Rau 2008:397) predominantly employ prefixes but also have suffixes, depending on person and number:

(51) Subject indexes of Gorum (Anderson & Rau 2008:397)

	SG	PL
1	ne-	le-
2	mɔ-	bɔ-
3	Ø	-ey, =gi

Yet other Munda languages employ enclitics, which can occur either on the verb or the word (of whatever lexical class) that immediately precedes the verb. This situation is found in Kharia (Peterson 2008:460–461), Gutob (Voß 2015), and Kherwarian languages. In the following two examples from Mundari, the subject index is encliticized to the noun phrase immediately preceding the predicate in (52a) and to the predicate in (52b):

- (52) Mundari (Osada 2008:121)
- a. *Soma hon-ko=e? lel-ko-a.*
 Soma child-PL=3SG.SBJ see-3PL.OBJ-IND
 ‘Soma will see (take care of) the children.’
- b. *Soma hon-ko lel-ko-ka=e?*
 Soma child-PL see-3PL.OBJ-OPT=3SG.SBJ
 ‘May Soma see the children.’

In expressive or contrastive discourse, Kera’ Mundari and Koḍa (cf. Anderson et al. 2008:219) can index the subject twice, i.e., both on the word immediately preceding the predicate and on the predicate:

- (53) Kera’ Mundari (Kobayashi & Murmu 2008:176)
- ini=ku lel-ki-a=ku*
 3SG=3PL.SBJ see-PST.3SG.ANIM.OBJ-IND=3PL.SBJ
 ‘They saw him/her.’

Object indexes are always suffixes in those Munda languages that have them. They usually index P and either T or G arguments.

Some Munda languages have differential object (P/T) indexing conditioned by either pragmatic factors or animacy. In Juang, for instance, object indexing is conditioned by pragmatic factors, whose nature still remains to be investigated (Patnaik 2008:529).

Santali has differential object indexing conditioned by animacy: Only animate objects are indexed (as in (54a)), whereas inanimate objects are not indexed (as in (54b)):

- (54) Santali (Neukom 2001:117)
- a. *hapen ceka-leka-te gidrə-dɔ=lan əsul-ko-a?*
 in.the.future what-like-INS child-TOP=1DU.INCL.SBJ support-3PL.OBJ-IND
 ‘How shall we be able to support our children in the future?’
- b. *dəlic’ daka=e dipil-ket’-a.*
 basket rice=3SG.SBJ put.on.head-PST.ACT-IND
 ‘He took the basket of rice on his head.’

Object indexes are not only used for P arguments, but also for either T or G arguments (depending on the language and construction), but never for both at the same time. In Santali, for instance, they usually index the T argument; when the applicative is employed, however, the G argument is indexed:

- (55) Santali (Neukom 2001:121)
- a. *maejiu-then gidrə=jn em-ked-e-a.*
 woman-DAT child=1SG.SBJ give-PST.ACT-3SG.OBJ-IND
- b. *gidrə-dɔ maejiu=jn em-ad-e-a.*
 child-TOP woman=1SG.SBJ give-APPL.PST.ACT-3SG.OBJ-IND
 both: ‘I gave the child to the woman.’

In (55a), the object index cross-references the T argument ('child'), whereas in (55b), it cross-references the G argument ('woman').

In Sora, object indexes are used for P and G arguments but not for T arguments:

- (56) Sora (Anderson & Harrison 2008b:329)
- a. *iando tid-t-ijn*
 why hit-NPST-1
 'Why are (you) hitting me?'
- b. *rban nen daʔa-n tiy-l-am*
 yesterday 1SG water-REF give-PST-[1SBJ-]2OBJ
 'I gave you water yesterday.'

In (56a), the P argument ('me') is indexed on the verb, and in (56b), the G argument ('you') is indexed on the verb.

Instances of object indexes cross-referencing dative/oblique-marked experiencers (analyzed as S and A arguments in the present approach) can occasionally be found, for instance in Juang:

- (57) Juang (Patnaik 2008:529)
- ajŋ-te bɔŋɔ-bɔ ɔn-gɔr-te betɔŋ-ki-ŋ.*
 1SG-OBL forest-LOC go-PFV-PRF fear-PRS-1
 'I fear going alone to the forest.'

Finally, instances where the object indexes are used for possessors ("possessor-raising") can also be found, e.g., in Gorum:

- (58) Gorum (Aze 1973:284, cited in Anderson & Rau 2008:399)
- mij pensil(-nɔm) ne-giʔ-t-ɔm*
 1SG pencil(-2POSS) 1-see-NPST-2OBJ
 'I will see your pencil.'

Now, how can this variation be accounted for? The diachronic scenario might have played out as follows: AA languages arriving in South Asia from Southeast Asia exhibited both AVP and VPA word orders. In South Asia, they developed verb-final word order under the influence of the autochthonous languages. However, the original word order patterns were not given up completely, but were morphologized by the development of the indexes (cf. Jenny 2015:319–320): AVP word order in the case of subject prefixes (and enclitics to preverbal words) and VPA word order in the case of subject suffixes (and enclitics to the verb).⁷ Note that the indexes in most cases still closely resemble free pronouns. Object indexes are more grammaticalized than subject indexes insofar as they are always suffixal and often more closely integrated within the verb morphology (e.g., closer to the stem than some TAM suffixes).⁸ This might be owed to the more fixed position of objects (P, T and G arguments) relative to the position of the verb: They basically always occurred after the verb, whereas subjects (S and A arguments) occurred either before or after the verb. The fact that indexing was developed is also owed to language contact: Indexing is found in all the other language families of South Asia: Most Indo-Aryan languages have subject indexing (but e.g., Maithili can index two arguments and in some

⁷ For some scenarios concerning the position of the subject indexes see Anderson (2007:79–82).

⁸ Note that this pattern does not go against Haig's (2018) observation that object person forms tend to become bound fairly rapidly, but remain at the stage of differential indexing, whereas subject indexing tends to become obligatory. The situation in Munda languages is largely in line with this observation: Object indexing is often differential (but obligatory e.g., in Korku), subject indexing is obligatory. The reverse pattern (differential subject indexing in combination with obligatory object indexing) could not be found.

instances adjuncts, cf. Bickel & Yādava 2000), and this also holds for Dravidian languages (some, e.g., Toda, lack indexing altogether, cf. Emeneau 1984). The Kiranti languages of the Sino-Tibetan family can index two arguments.

However, the patterns of these non-Munda languages of South Asia are fairly different from those found in Munda languages: They mostly exhibit templatic morphology (i.e., the position of the morphemes are often determined by factors other than syntax and semantics), while Munda languages have layered (or hierarchical) morphology (i.e., word-internal configurationality, cf. Bickel & Nichols 2007:219 and Peterson 2017:241). This difference suggests that while the Munda languages were subject to areal pressure to develop indexing, their indexing patterns are still much younger than those found in the other language families and therefore still reflect older syntax much more closely. The index morphology is also much more clearly concatenative than in the neighboring language families: Most index morphemes are clearly separable from adjacent morphemes rather than fused with them.⁹

4.2 Khasian and Palaungic

Indexing is marginal in Khasian languages at best. Pnar, for instance, lacks indexing (cf. Ring 2015). In Standard Khasi, however, indexing can be found: Third person forms preceding the verb occur together with coreferential lexical S/A nouns (but not pronouns):¹⁰

- (59) Khasi (Nagaraja 2015:1159)
- | | | | |
|---------------|----------|-------------|------------|
| <i>u=ksew</i> | <i>u</i> | <i>ba:m</i> | <i>do?</i> |
| M=dog | 3SG.M | eat | flesh |
- ‘The dog eats flesh.’

Mnar also has person marking cross-referencing subjects, which, however, does not always appear on the verb:

- (60) Mnar (Rymbai 2018:12)
- | | | | | | | |
|----|----------|------------|---------------|-----------|-----------|-------------|
| a. | <i>u</i> | <i>ɟɔn</i> | <i>ieid=u</i> | <i>ha</i> | <i>ga</i> | <i>meri</i> |
| | M | John | love=3SG.M | DAT | F | Mary |
- ‘John loves Mary.’
- | | | | | |
|----|-----------|-------------|-------------|--------------------|
| b. | <i>ga</i> | <i>meri</i> | <i>tieɲ</i> | <i>haicbi:m=ka</i> |
| | F | Mary | cook | food=3SG.F |
- ‘Mary cooks food.’

While in example (60a) the person marking cross-referencing the subject (‘3SG.M’) is an enclitic to the verb, in example (60b) it follows the P argument (‘food’). But some other clauses lack person markers altogether. More research is needed to establish their status and behavior.

These patterns are probably not the result of language contact: The Sino-Tibetan languages closest to Khasian such as Bodo (Brahma 2013:214) or Atong (van Breugel 2014:269) do not exhibit indexing, although it can be found in other Sino-Tibetan languages spoken further away in northeastern India such as some Kuki-Chin and Naga languages (see e.g., Peterson 2007:263–264 for Hakha Lai or Coupe 2007:150–169 for Mongsen Ao). Other contact languages such as Hindi or English are unlikely to have influenced the development of indexing in Khasi. Thus, the development of S/A indexing in Standard Khasian is probably a language-internal process, while the absence of indexing in other Khasian is shared with nearby Sino-Tibetan languages.

⁹ Some exceptions can be found in Sora, where the first-person subject suffix in the case of the singular and the suffix part of the circumfix in the case of the plural is dropped before the second person object suffix (Anderson & Harrison 2008b:328).

¹⁰ Although the person form is written separately, the fact that it occurs together with a coreferential noun phrase suggests that it is most probably an index.

Palaungic languages do not exhibit indexing, which is in line with the structures of the Sino-Tibetan and Tai languages which surround them.

4.3 Mainland Southeast Asia

The vast majority of the AA languages of Mainland Southeast Asia lack indexing. It is found, however, in some single instances: Pacoh has some verbs that take the prefixes λu - ‘3SG.HUM’ or λi - ‘IMPERS’, as the following examples illustrate:

- (61) Pacoh (Alves 2006:34–35)
- a. λu -to:ŋ ni:m məh naʔ λaca :j
 3SG.HUM-say only one UNIT brother
 ‘He said he has only one brother.’
- b. λi -taʔ pəllo: λal ɔ:ŋ
 IMPERS-make tube wood
 ‘One makes a wooden tube.’

The precise distribution of these prefixes and their productivity, however, have not been fully determined (Alves 2006:34). But it is clear that they cannot be combined with coreferential noun phrases, they are restricted to certain verbs, and there are no indexes for first and second person.

The absence of indexing in AA languages of Mainland Southeast Asia again matches the overall geographical pattern: Non-AA languages of the area (north of the Malay Peninsula) also lack indexing.

4.4 Aslian and Nicobarese

Indexing is found in most Aslian languages but is often restricted. Most Aslian languages differentiate between stative and dynamic clauses, and indexing only occurs in the latter, as the following two sentences from Ceq Wong illustrate:

- (62) Ceq Wong (Kruspe et al. 2015:437)
- a. tieʔ pasel
 ground be.dry
 ‘The ground is dry.’
- b. jɔ λu =ras kaʔ λ ɔs
 Jo 3SG=fall.over LOC fire
 ‘Jo fell into the fire.’

In (62a) ‘be dry’ is a stative predicate, which lacks indexing. In (62b), on the other hand, ‘fall over’ is a dynamic predicate, where subject indexing is found. Hence Ceq Wong exhibits differential S/A indexing conditioned by Aktionsart (or lexical aspect). Other Aslian languages have similar patterns. But Ceq Wong exhibits a further restriction: The indexes cannot occur with the irrealis prefix (Kruspe et al. 2015:437). Jahai has a very similar system, but it has two different irrealis forms: wa = for third person singular, and ya = for all other persons (Burenhult 2005:67). Maniq also has a similar system, where indexing is only found in dynamic clauses, but only with third person (Kruspe et al. 2015:436).

Semelai has a slightly different indexing pattern: Like flagging, indexing is restricted to perfective clauses, while imperfective/generic clauses lack indexing (Kruspe 2004:254). The following two examples from Semelai (Kruspe 2004:254–255) serve as an illustration:

- (63) Semelai (Kruspe 2004:254–255)
- a. *ki=bukɔʔ la=kniek hn=pintuʔ.*
 3A=open ERG=husband ACC=door
 ‘The husband opened the door.’
- b. *pɔdɔŋ ca smaʔ.*
 tiger eat person
 ‘Tigers eat people.’

In the perfective clause of (63a), the A argument is indexed, while the generic clause of (63b) lacks indexing.

A indexing shows further restrictions: It does not occur in imperative and irrealis clauses (Kruspe 2004:259). Third person Ps and Ts can be indexed as enclitics (formally identical with the P/T flag), but only in the absence of a coreferential noun phrase in the same clause:

- (64) Semelai (Kruspe 2004:263)
- ki=bantay=hn*
 3A=butcher=3P
 ‘He butchered her.’

S arguments are indexed in the same way as A arguments if the action or state is caused by some external source, which does not have argument status:

- (65) Semelai (Kruspe 2004:265)
- ki=ʔyam la=yɛ*
 3S=cry CAUSE=1
 ‘He cried because of me.’

G arguments are never indexed in Semelai (Kruspe 2004:263–264).

Indexing in Nicobarese languages appears to be marginal. For Nancowry, Radhakrishnan (1970) reports sentences such as the following:

- (66) Nancowry (Radhakrishnan 1970:97)
- ʔifě yuána ruk ʔin ʔifě*
 3PL PST come PROX 3PL
 ‘Did they come?’

Here, the third-person plural marker at the beginning of the sentence appears to be an index that cross-references the person marker at the end of the sentence. This pattern seems to be restricted to interrogative sentences but does not always occur there.

Car also has some sort of indexing. It has a complex system of person forms differentiating between core argument, oblique, interrogative, and subordinate forms; furthermore, for third person, visibility and (orthogonally) tense are additionally differentiated (cf. Sidwell 2015:1246; for a slightly different analysis see also Braine 1970:134–153). In some instances, person forms co-occur with coreferential full noun phrases:

- (67) Car (Braine 1970:242, Sidwell 2015:1239)
- lák ʔan ŋam máj*
 calm 3SG DEF sea
 ‘The sea is calm.’

While in this example, the person form (‘3SG’) cross-references the S argument, in the following example, the person form cross-references the P argument:

- (68) Car (Braine 1970:145, Sidwell 2015:1247)
Ha-cát-ηén ʔək lí pəre cin
 CAUS-be.lost-away 3SG.NONVIS.PST book 1SG
 ‘I lost the book.’

However, the glosses suggest that these pronouns are not bound, which would be necessary to be classified as indexing by definition. More research is needed to establish the properties of these person forms.

Thus, indexing can be found in Aslian and also (to some unclear degree) in Nicobarese languages. The indexing morphology was probably developed fairly recently, since the indexes resemble independent pronouns. But why was indexing developed here? Two scenarios are conceivable: Contact influence through pre-AA substrate languages or a language-internal development.

The fact that indexing is also present in some Austronesian languages of northern Sumatra and the Barrier islands (cf. Donohue & Denham 2010:233), i.e., Austronesian languages in the vicinity of Aslian and Nicobarese languages, points to the first scenario. These Austronesian languages must also have developed indexing, since it is not reconstructed for Proto-Austronesian (Ross 1995, 2002). Furthermore, indexing is also present in the Papuan languages further east, and subsequently the local Austronesian languages also developed it (cf. e.g., Schapper 2015). Thus, indexing is found in an area covering much of New Guinea and southern Island Southeast Asia.

But some facts also point to the second scenario: Indexing is especially developed in verb-initial constructions. This holds for both the Aslian languages and nearby Austronesian languages. From a global crosslinguistic perspective, verb-initial languages tend to have indexing (cf. Nichols 1992:Section 3.3, Hawkins 2004:Section 8.5 or Sinnemäki 2010). Thus, Proto-AA and Proto-Austronesian are unusual in lacking indexing despite being verb-initial, and the languages of western Island Southeast Asia (including Aslian) changed to being more in line with the worldwide tendency. But the two scenarios are not mutually exclusive: It is also conceivable that the development of indexing was pushed by both substrate influence and structural tendencies of verb-initial clauses.

5 Conclusion

The flagging patterns of Munda widely fit in the South Asian geographical profile. Presumably the Munda languages borrowed these patterns from other languages. In some instances, even the morphemes were borrowed from Indo-Aryan languages (cf. e.g., the Gutob object marker =*lai*). Similarly, the (rather limited) flagging present in Khasian languages was presumably borrowed from some neighboring Sino-Tibetan languages, which have similar patterns. On the other hand, flagging in West Palaungic is different from the patterns found in neighboring Sino-Tibetan and Tai-Kadai languages and bears some similarities to the patterns of Nicobarese. A likely explanation for this finding is that these patterns are conservative AA patterns which were (at least partly) inherited from Proto-AA. This suggests that the languages of (core) Mainland Southeast Asia largely lost argument flagging, which is also in line with the fact that languages of other families present in Mainland Southeast Asia also lack flagging. Finally, the presence of flagging in Aslian partly fits in the geographical picture; however, the language-specific patterns are unique in a genealogical and areal perspective and are therefore the result of language-internal (and subgroup-internal) processes.

The presence of indexing is more restricted than the presence of argument flagging in AA. The indexing morphology in all the AA languages where it is found is not very old and differs from subgroup to subgroup. This suggests that indexing cannot be reconstructed for Proto-AA. Indexing is especially prominent in Munda. But while in principle indexing is found in all language families of South Asia, the indexing patterns of Munda languages differ considerably from those found in languages of other families in the area. The reason for this discrepancy might lie in the fact that indexing is much younger and less grammaticalized in Munda than in non-Munda languages of the area. Indexing is also

(marginally) present in Khasian, where it might be a language-internal development, and Aslian and Nicobarese, where the development might have been fostered through contact influence with other languages of western Island Southeast Asia, where it is also present (especially in the more conservative Austronesian languages of northwest Sumatra and the Barrier Islands). The indexing patterns of Nicobarese languages are still under-researched and are much in need of more investigations.

Overall, more research is needed to establish argument marking patterns in older stages of AA and Proto-AA. Proto-AA presumably had some argument flagging, especially on the P argument in VAP clauses and also on G arguments, but it lacked indexing. This might have further implications for the investigation of the prehistory of greater Southeast Asia: Notably, Proto-Austronesian is also reconstructed as having flagging but lacking indexing (cf. Ross 2002). Thus, this might have been a pattern that was once much more widespread – not only in core Mainland Southeast Asia but also in parts of China, before the presently dominant analytical and isolating structures came into existence.

From a worldwide perspective, however, languages with dominant verb-initial structures most commonly have indexing and lack flagging (cf. Nichols 1992), which is thus a pattern not in line with reconstructed Proto-AA and Proto-Austronesian.

References

- Alves, Mark. 2006. *A Grammar of Pacoh: a Mon-Khmer language of the central highlands of Vietnam*. Canberra, Pacific Linguistics.
- Alves, Mark. 2015. Morphological functions among Mon-Khmer languages. Beyond the basics. In N. J. Enfield & Bernard Comrie (eds.), *Languages of Mainland Southeast Asia: The state of the art*, 531–557. Berlin: De Gruyter Mouton.
- Anderson, Gregory D. S. 2007. *The Munda verb*. Berlin: Mouton de Gruyter.
- Anderson, Gregory D. S. 2008. Gta?. In Anderson (ed.), 682–763.
- Anderson, Gregory D. S. 2015. Overview of the Munda languages. In Jenny & Sidwell (eds.), 364–414.
- Anderson, Gregory D. S. (ed.). 2008. *The Munda languages*. London: Routledge.
- Anderson, Gregory D. S. & K. David Harrison. 2008a. Remo (Bonda). In Anderson (ed.), 557–632.
- Anderson, Gregory D. S. & K. David Harrison. 2008b. Sora. In Anderson (ed.), 299–380.
- Anderson, Gregory D. S. & Opino Gomango 2016. On the current status and state of Juray in the Sora-Juray cluster. In Nicholas Ostler & Panchanan Mohanty (eds.), *Proceedings of the 20th FEL Conference*, 103–109. Hungerford: Foundation for Endangered Languages.
- Anderson, Gregory D. S., Toshiki Osada & K. David Harrison. 2008. Ho and the other Kherwarian languages. In Anderson (ed.), 195–255.
- Anderson, Gregory D. S. & Felix Rau. 2008. Gorum. In Anderson (ed.), 381–433.
- Anderson, Gregory D. S. & Norman H. Zide. 2001. Recent advances in the reconstruction of the Proto-Munda verb. In Laurel J. Brinton (ed.), *Historical Linguistics 1999*, 13–30. Amsterdam: Benjamins.
- Aze, Richard. 1973. Clause patterns in Parengi-Gorum. In R. L. Trail (ed.), *Patterns in clause, sentence and discourse in selected languages of India and Nepal, Part 1*, 235–312. Kathmandu: SIL and Tribhuvan University.
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar & Yogendra P. Yādava. 2000. A fresh look at grammatical relations in Indo-Aryan. *Lingua* 110.5:343–373.
- Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon*, 169–240. Cambridge: Cambridge University Press, 2nd edition.
- Blok, Greg. 2013. A descriptive grammar of Eastern Lawa. MA thesis, Payap University.

- Brahma, Partima. 2013. Phonology and morphology of Bodo and Dimasa: A comparative study. Doctoral dissertation, Assam University.
- Braine, Jean. 1970. Nicobarese grammar (Car dialect). Doctoral dissertation, University of California, Berkeley.
- Brown, Lea. 2001. A grammar of Nias Selatan. Doctoral Dissertation, University of Sydney.
- Burenhult, Niclas. 2005. *A grammar of Jahai*. Canberra: Pacific Linguistics.
- Coupe, Alexander Robertson. 2007. *A grammar of Mongsen Ao*. Berlin: Mouton de Gruyter.
- Coupe, Alexander Robertson. 2011. Pragmatic foundations of transitivity in Ao. *Studies in Language* 35.3:492–511.
- Cysouw, Michael. 2004. A history of Munda person marking. Manuscript, Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Donohue, Mark & Tim Denham. 2010. Farming and Language in Island Southeast Asia: Reframing Austronesian History. *Current Anthropology* 51.2:223–256.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67.3:547–619.
- Edwards, Owen. 2015. The position of Enggano within Austronesian. *Oceanic Linguistics* 54:54–109.
- Emeneau, Murray B. 1984. *Toda grammar and texts*. Philadelphia: American Philosophical Society.
- Enfield, N. J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34:181–206.
- Haig, Geoffrey. 2018. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56.4:781–818.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.7*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
<https://doi.org/10.5281/zenodo.7398962>
 (Available online at <http://glottolog.org>, accessed on 2022-12-09.)
- Handschuh, Corinna. 2014. *A typology of marked-S languages*. Berlin: Language Science Press.
- Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntax of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 197–226. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2019. Indexing and flagging, and head and dependent marking. *Te Reo* 62.1:93–115.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press
- Jacob, Judith M. 1968. *Introduction to Cambodian*. London: Oxford University Press.
- Jenny, Mathias. 2015. Modern Mon. In Jenny & Sidwell (eds.), 553–600.
- Jenny, Mathias. 2015. Syntactic diversity and change in Austroasiatic languages. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 317–340. Amsterdam: Benjamins.
- Jenny, Mathias, Tobias Weber & Rachel Weymuth. 2015. The Austroasiatic languages: A typological overview. In Jenny & Sidwell (eds.), 13–143.
- Jenny, Mathias & Paul Sidwell (eds.). 2015. *The handbook of Austroasiatic languages*. Leiden: Brill.
- Kobayashi, Masato & Ganesh Murmu. 2008. Kera? Mundari. In Anderson (ed.), 165–194.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Kruspe, Nicole. 2015. Semaq Beri. In Jenny & Sidwell (eds.), 475–516.
- Kruspe, Nicole, Niclas Burenhult & Ewelina Wnuk. 2015. Northern Aslian. In Jenny & Sidwell (eds.), 419–474.
- Lee, Wei-Wei. 2023 (this publication). Subordination strategies in Rucing Palaung. In Hiram Ring & Paul Sidwell (eds.), *Papers from the Eighth International Conference on Austroasiatic Linguistics*, 22–52. Journal of the Southeast Asian Linguistics Society Special Publication No. 11. Honolulu: University of Hawai'i Press.

- Mak, Pandora. 2012. *Golden Palaung: A grammatical description*. Canberra: Asia-Pacific Linguistics.
- Nagaraja, K. S. 2015. Standard Khasi. In Jenny & Sidwell (eds.), 1145–1185.
- Nguyễn, Văn Lợi, Nguyễn Hữu Hoàn & Tạ Văn Thông. 2008. *Tiếng Mảng* [The Mang language]. Hanoi: Nhà Xuất Bản Khoa Học Hà Nội.
- Neukom, Lukas. 2001. *Santali*. Munich: Lincom.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Osada, Toshiaki. 2008. Mundari. In Anderson (ed.), 99–164.
- Patnaik, Manideepa. 2008. Juang. In Anderson (ed.), 508–556.
- Peterson, David A. 2017. Hakha Lai. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 258–276. London: Routledge.
- Peterson, John. 2008. Kharia. In Anderson (ed.), *The Munda Languages*. 434–507. London: Routledge.
- Peterson, John. 2011. *A grammar of Kharia: A South Munda language*. Leiden: Brill.
- Peterson, John. 2017. Fitting the pieces together – Towards a linguistic prehistory of eastern-central South Asia (and beyond). *Journal of South Asian Languages and Linguistics* 4.2:211–257.
- Rabel, Lili. 1961. *Khasi, a language of Assam*. Baton Rouge: Louisiana State University Press.
- Radhakrishnan, R. 1970. *A preliminary descriptive analysis of Nancowry*. Ann Arbor: UMI. (Doctoral dissertation, University of Chicago.)
- Rajasingsh, V. R. 2016. Mūöt (Nicobarese). *Mon-Khmer Studies* 45:4–52.
- Ring, Hiram. 2015. A grammar of Pnar. Doctoral dissertation, Nanyang Technological University.
- Ross, Malcolm. 1994. Some current issues in Austronesian linguistics. In Darrell T. Tryon (ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies*, 45–120. Berlin: Mouton de Gruyter.
- Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice marking. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of Western Austronesian voice systems*, 17–62. Canberra: Pacific Linguistics.
- Rymbai, Ruth. 2018. Typological characteristics of Mnar. In Linda Konnerth, Stephen Morey & Amos Teo (eds.), *North East Indian Linguistics* 8, 1–16. Canberra: Asia-Pacific Linguistics.
- Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel* 90:99–151.
- Seng Mai, Ma. 2012. A descriptive grammar of Wa. MA thesis, Payap University.
- Sidwell, Paul. 2019. AA in Island SEA: Proto-Nicobarese syntax. Paper presented at the 24th International Conference on Historical Linguistics, Canberra, July 1–5.
- Sinnemäki, Kaius. Word order in zero-marking languages. *Studies in Language* 34.4:869–912.
- Solncev, V. M., N. V. Solnceva & I. V. Samarina. 2001. *Jazyk ruk* [The Ruc language]. Moscow: “Vostočnaja Literatura” RAN.
- Solntseva, Nina V. 1996. Case-marked pronouns in the Taoih language. *Mon-Khmer Studies* 26:33–36.
- van Breugel, Seino. 2014. *A grammar of Atong*. Leiden: Brill.
- Voß, Judith. 2015. Person markers in Gutob. *Journal of South Asian Languages and Linguistics* 2.2:215–240.
- Weymuth, Rachel. 2019. Grammatical relations in Rumai (Palaung). Paper presented at the 29th meeting of the Southeast Asian Linguistics Society, Tokyo, May 27–29.
- Witzlack-Makarevich, Alena. 2011. Typological variation in grammatical relations. Doctoral dissertation, University of Leipzig.
- Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking: Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *The diachronic typology of differential argument marking*, 1–40. Berlin: Language Science Press.

- Wnuk, Ewelina. 2016. Semantic specificity of perception verbs in Maniq. Doctoral dissertation, Radboud Universiteit Nijmegen.
- Zide, Norman H. 2008. Korku. In Anderson (ed.), 256–298.
- Zide, Norman H. & Gregory D. S. Anderson. 2001. The Proto-Munda verb system and some connections with Mon-Khmer. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Tokyo Symposium on South Asian Languages Contact, Convergence and Typology*, 517–540. New Delhi: Sage.

Appendix: Languages included in the sample

Note: The aim of this sample, consisting of 32 AA languages out of a total of 158 AA languages listed in the Glottolog (Hammarström et al. 2022), is to capture the variation of flagging and indexing patterns. For this reason, some subgroups, especially the Munda languages with their complex flagging and indexing patterns, have been overrepresented, whereas the Mainland Southeast Asian subgroups Khmuic, Bahnaric and Pearic have not been included at all since in the languages of these subgroups, flagging and indexing are (almost) completely absent (flagging is sometimes present on G arguments).

Language	Subgroup	Area	F/I? ¹¹	Source(s)
Gta', Plains	Munda	Eastern India	F/I	Anderson & Gomango (2016)
Gorum	Munda	Eastern India	F/I	Aze (1973), Anderson & Rau (2008)
Gutob	Munda	Eastern India	F/I	Voß (2015; p.c.)
Juang	Munda	Eastern India	F/I	Patnaik (2008)
Kharia	Munda	Eastern India	F/I	Peterson (2011)
Koḍa	Munda	Eastern India	F/I	Anderson et al. (2008)
Korku	Munda	Central India	F/I	Zide (2008)
Mundari	Munda	Eastern India	F/I	Osada (2008)
Mundari, Kera'	Munda	Eastern India	F/I	Kobayashi & Murmu (2008)
Remo	Munda	Eastern India	F/I	Anderson & Harrison (2008a)
Santali	Munda	Eastern India	F/I	Neukom (2001)
Sora	Munda	Eastern India	F/I	Anderson & Harrison (2008b)
Khasi, Standard	Khasian	Northeast India	F/I	Rabel (1961), Nagaraja (2015)
Mnar	Khasian	Northeast India	F/I	Rymbai (2018)
Pnar	Khasian	Northeast India	F	Ring (2015)
Lawa, Eastern	Palaungic	Thailand	(F)	Blok (2013)
Rucing	Palaungic	Myanmar	F	Lee (2023; p.c.)
Rumai	Palaungic	Myanmar	F	Weymuth (2019)
Shwe	Palaungic	Myanmar	F	Mak (2012)
Wa	Palaungic	Myanmar	(F)	Seng Mai (2012)
Mang	Mangic	Vietnam/Yunnan	F	Nguyễn et al. (2008)
Ruc	Vietic	Vietnam	F	Solncev et al. (2001)
Pacoh	Katuic	Vietnam	F/I	Alves (2006)
Khmer	Khmeric	Cambodia	(F)	Jacob (1968)
Mon	Monic	Myanmar/Thailand	(F)	Jenny (2015)
Ceq Wong	Aslian	Malay Peninsula	I	Kruspe et al. (2015)
Jahai	Aslian	Malay Peninsula	F/I	Burenhult (2005)
Maniq	Aslian	Malay Peninsula	I	Kruspe et al. (2015), Wnuk (2016)
Semaq Beri	Aslian	Malay Peninsula	F/I	Kruspe (2015)
Semelai	Aslian	Malay Peninsula	F/I	Kruspe (2004)
Car	Nicobarese	Nicobar Islands	F/I	Braine (1970), Sidwell (2015)
Nancowry	Nicobarese	Nicobar Islands	F/I	Radhakrishnan (1970)

¹¹ F=flagging is present; I=indexing is present; (F)=flagging is only present on (some) G arguments.

SECONDARY VERBS IN RUMAI, PALAUNG

Rachel WEYMUTH

Independent Researcher

<rachel DOT weymuth AT gmail DOT com>

Abstract

Rumai, a variety of Palaung (Palaungic, Austroasiatic), exhibits several secondary verbs (V2). Most of these verbs can occur as full verbs, but also have a grammaticalized meaning when used in combination with another verb in the same clause. Two groups of secondary verbs can be distinguished in Rumai, preverbal and postverbal ones. The preverbal V2 are the syntactic main verbs as they take the verbal affixes, while the postverbal V2 are syntactically dependent. There are also verbal prefixes, which have the same or a similar function as some postverbal V2. The aim of this paper is to give a descriptive account of the most common secondary verbs in Rumai including their functions and their connections to verbal affixes where appropriate.

Keywords: secondary verbs

ISO 639-3 codes: rbb

1 Introduction

Rumai is one of three varieties of Palaung so far mentioned in the linguistic literature, the other two being Shwe and Ruching. It is mainly spoken in Northern Shan State, but also in adjacent regions in Yunnan, China. There are some 140,000 speakers. The Rumai live mainly on the slopes and ridges of the Shan hills, but today many, especially young people, have moved to lower regions in Shan State and to the big cities in the Burmese plain for studies or for work. Their first language in the villages is Rumai, but in school, they are educated in Burmese and mainly at work and markets they have contact with Shan which also was the language of education until the introduction of the public school system in the 1960s. After this change, the influence of Burmese increased, and the language replaced Shan to a considerable extent as a donor language. This means that older loans in Rumai are mostly from Shan or Burmese via Shan while more recent ones are mostly directly from Burmese.

Rumai has an iambic intonation structure, that means, in a disyllabic word always the second syllable takes the stress, and pre-syllables in sesquisyllabic words, and prefixes are never stressed. The basic clause structure in Rumai is verb-medial, that is SV and AVP, but in subordinate clauses it is mainly VS and VAP. Grammatical relations are expressed by a) constituent order, b) the oblique marker $t\lambda^?$ (G arguments, P and T arguments in verb-initial dependent clauses, P arguments with certain verbs), c) the locative marker $p\lambda^?$ (adjuncts, S and P arguments ‘a place where one Vs’), and d) the complementizer $l\lambda^?$ (S arguments in existential clauses, P arguments with certain verbs). The verbal affixes play an important role in this paper, as a part of them together with a part of the secondary verbs form the TAM-system in Rumai, whereby this system only consists of aspectual and modality markers in Rumai. The main verbal affixes are listed in Table 1:

Table 1: Verbal affixes in Rumai

Affix(es)	Meaning	Source	Domain
<i>gi:j-</i>	imperfective	<i>gɔ̃:j</i> ‘stay’	aspectual
<i>ʔuN-</i>	durative	<i>ʔu:n</i> ‘keep’	aspectual
<i>ʔə-</i>	inceptive	-	aspectual
<i>hɔ̃j-</i>	new situation	<i>hɔ̃:j</i> ‘finish’	aspectual
<i>təm-</i>	experiential	-	aspectual
<i>nʌŋ-</i>	irrealis	-	modality
<i>siŋ-</i>	desiderative	-	modality
<i>bu:- / ʔa:w- / -maʔ</i>	negative	-	negation
<i>kʰu:-</i>	prohibitive	-	negation
<i>kə- / laj-</i>	reciprocal	- / <i>lɑ̃:j</i> ‘take’	reciprocal

Two of the aspectual prefixes, namely durative *ʔuN-* and new situation *hɔ̃j-* largely correspond in function to the two secondary verbs *ʔu:n* and *hɔ̃:j* which, like the prefixes, are derived from the lexical verbs *ʔu:n* ‘keep’ and *hɔ̃:j* ‘finish’ respectively. The new situation (newsit) markers denote a “new (but expected) situation after a limit has been transgressed” (Jenny 2001:125).¹

Based on their semantic priority in a clause, the Rumai lexical verbs can be divided into two functional categories: semantic primary and semantic secondary verbs. In this paper, the first category will be called “main verbs” and the second one “secondary verbs”. All the secondary verbs, except of preverbal *bɛ* ‘can’, a Shan loanword with only grammatical (modal) function, are also used as main verbs. As secondary verbs, they belong to the modality, valency, and aspectual domains. The use as main verbs and the grammatical functions are two of the twenty-two properties mentioned in Heine’s work about auxiliaries (Heine 1993:22-24). The preverbal secondary verbs show more of these properties than the postverbal ones and are more similar to auxiliaries in the traditional literature. As both share the clear basic property of secondary verbs, namely that they are grammaticalized verbs used in a clause to express grammatical functions which modify the main verb, they will be subsumed in this paper under the category “secondary verb” with two sub-categories instead of using the cross linguistically rather diverse and, where it is more consistent, Eurocentric category “auxiliary” (Steele 1978, Heine 1993, Aikhenvald 2006, Seiss 2009).

This study is based on a corpus collected by the author between 2013 and 2020. It includes about 2,450 sentences from questionnaires, picture stories, conversations, and written texts. Additional data was obtained from native speakers in Myanmar by elicitation in person and online.

Secondary verbs are a common feature in the world’s languages and several verbs with similar lexical meanings and used in similar grammatical functions like in Rumai are widespread in Southeast Asia. In this paper, the most common secondary verbs in Rumai will be described. This will also give more insight in the use and functions of secondary verbs generally, in Southeast Asia, and in contact scenarios.²

2 The secondary verbs

Secondary verbs are grammaticalized verbs which occur in addition to a main verb to express grammatical functions. The secondary verbs occur in the same form as they do as full verbs that is, they do not show phonetic reduction or coevolution between form and meaning, a feature that is

¹ For the verbal affixes see also Weymuth (2018).

² Abbreviations used are: 1 first person; 2 second person; 3 third person; A agent; ADD additive; ANA anaphoric demonstrative; ASRT assertive; CLF classifier; COMP complementizer; CONC concessive; COND conditional; CSP consent seeking particle; DEM demonstrative; DES desiderative; DISC discourse marker; DIST distal demonstrative; DU dual; DUR durative; EMPH emphasis; EUPH euphonic; EXP experiential; G goal; INCEP inceptive; INCL inclusive; IPFV imperfective; IRR irrealis; LOC locative; MEDL medial demonstrative; NEG negative; NFUT non-future; NML nominalizer; NSIT new situation; OBJ object; OBL oblique; P patient; PL plural; POL polite; PROH prohibitive; PROX proximal demonstrative; PTCL particle; PURP purposive; Q question marker; RESTR restrictive; S single argument; SG singular; T theme; TCL topic-comment linker; UP upper demonstrative; V verb.

characteristic for East and mainland Southeast Asian languages in comparison to many other (e.g., Indo-European) languages (Bisang 2004). While the preverbal secondary verbs are the syntactic primary verbs as they take the verbal affixes (Weymuth 2018), the postverbal ones are always syntactically dependent on the main or the preverbal secondary verb. An overview on the semantic and syntactic priority of the lexical verbs in Rumai is given in Table 2.

Table 2: *Semantic and syntactic priority of the Rumai lexical verbs*

	preverbal V2 + main verb		main verb (+ postverbal V2)	
semantic priority	–	+	+	–
syntactic priority	+	–	+	–

In a synchronic view, the secondary verbs form a closed class. However, the preverbal secondary verbs may develop into prefixes, like *səmâ* ‘know’, which is on the way from a preverbal secondary verb to a prefix (2.1.3). Table 3 shows the most common secondary verbs in Rumai including their basic functions and the grammatical domains to which they belong.

Table 3: *The most common secondary verbs in Rumai*

Verb	Lexical meaning	Position	Function	Domain
<i>lo</i>	‘need’	preverbal	obligation	modality
<i>bɛ</i>	‘can’	preverbal	possibility inherent ability	modality
<i>səmâ</i>	‘know’	preverbal	acquired ability	modality
<i>bɣn</i>	‘get’	preverbal	possibility permission	modality
<i>dɛh / tʰiw</i>	‘give’	preverbal	causative	valency
<i>dɛh / tʰiw</i>	‘give’	postverbal	benefactive	valency
<i>hò:j</i>	‘finish’	postverbal	new situation	aspectual
<i>ʔû:n</i>	‘keep’	postverbal	durative	aspectual
<i>hâ:w</i>	‘go’	postverbal	motion away from the center of interest	aspectual
<i>dɛəh</i>	‘come’	postverbal	motion towards the center of interest	aspectual

2.1 The preverbal secondary verbs

The verbs discussed in this subsection are *lo* ‘need’, *bɛ* ‘can’, *səmâ* ‘know’, *bɣn* ‘get’, and *dɛh* and *tʰiw*, both meaning ‘give’. The preverbal secondary verbs, like all syntactic primary verbs, have the following features:

- they take verbal affixes
- they are in the scope of negation
- the polar question marker *lɔh* follows them directly
- they occupy the first position, followed by the subject in verb-initial subordinate clauses
- they can be used as the only word in the positive, or negated in the negative answer to a polar question

2.1.1 *lo* ‘need’

A verb with the meaning ‘need’ is used in languages of different families to express the grammatical function of obligation, that means deontic modality (Heine and Kuteva 2002:215). Heine and Kuteva describe the process of grammaticalization as “an instance of a process whereby a verb, on account of some salient semantic property, gives rise to a grammatical marker highlighting that property” (Heine and Kuteva 2002:216). In Rumai as well, the verb *lo* ‘need’ is, besides its use as a main verb (1), a secondary verb which expresses an obligation. It can be translated with ‘have to’ (2) or ‘must’ (3).

- (1) *ʔáw lo ɲanwe təʔá:ŋ ká:j ɲap ʔə*
 1SG to.need bag Ta'ang ADD book POL
 'I need a Ta'ang bag and book.' (WPKT_14_L_016)
- (2) *ʔənoŋ lo kjwé ʔjɛ ká:j lɔ:j*
 first.son to.need to.feed chicken ADD pig
 'The son had to feed the chickens and the pigs.' (ECC_16_M_3_059)
- (3) *ʔáw lo kʰáj təgəən³ ʔá:j réən ʔáw tɔʔ ɲúŋ ʔu mjó*
 1SG to.need very to.consider before to.do 1SG OBL work one kind
 'Before I do a job, I carefully have to think about it.' (W4_15_M_3_027)

In Mak (2012:50) Shwe *lo* 'need' and in Deepadung et al. (2015:1082) Ruching *lo* 'must' are mentioned as secondary verbs with the same function as in Rumai. The verb *lo* is a loan either directly from Burmese *lo* 'want, need, lack' (4) or Shan *lò* 'need, be essential' (5), where Shan *lò* is a loan from Burmese.

- (4) *di ʔətʰe.ʔəne-go ɲan θə̀n.θaʔ-pʰó lo-de*
 this situation-OBJ return contemplate-for need-NFUT
 'We need to reanalyze the situation.' (Jenny and Hnin Tun 2016:187)
- (5) *lò pəu mój tʰuŋ kʰáw lǐ ʔú.tú lǐ kój*
 to.need to.wait to.look.forward to.arrive occasion be.good season be.good only
jəw nɔ
 finish CSP
 'It's necessary to wait and anticipate until it is a good time and good weather only, right?'
 (Glick and Tern Moeng 1991:200 (adapted))

Whether Palaung *lo* is a loan directly from Burmese or from Burmese via Shan cannot be detected by morphological or phonological features. However, *lo* as a secondary verb with the same function as today was already mentioned by Milne in her grammar of Palaung (Shwe) from 1921 (Milne 1921:75). And as Shan was the main contact language of Palaung until the 1960ies, the borrowing from this language is more probable. Another point, which supports this hypothesis, is the preverbal position in Shan and Rumai while in Burmese it is postverbal.

The secondary verb *lo* can take several affixes, among them are the negative ones (6).

- (6) *ma ʔáw bu-lo-maʔ réən ɲúŋ kjá:m*
 mother 1SG NEG-to.need-NEG to.do work be.rough
 'My mother doesn't have to do rough work anymore.' (WA_15_M_3_038)

There are several other negative affixes in Rumai (Weymuth 2018:94-96). All of them can be used in appropriate circumstances with *lo*, except of the prohibitive *kʰu-*.

The irrealis prefix *naŋ-* expresses, depending on the situation, future, possibility, intention, or supposition (Weymuth 2018:92). In example (7), *naŋ-* is used to express future.

- (7) *ʔuŋ.hjé ʔáw naŋ-lo há:w ɲaʔ ɲaŋuŋ*
 tomorrow 1SG IRR-to.need to.go LOC Yangon
 'I'll have to go to Yangon tomorrow.' (EGGD_16_M_3_146)

³ What is transcribed as *eə* is not a diphthong in Rumai, but a phonemic vowel between /e/ and /ɛ/.

- (8) *nay-lo lah kəlá:ŋ ʔəpom ní tɬ ma māj ʔəkʰáj*
 IRR-to.need Q to.tell story PROX OBL mother 2SG really
 ‘Should I really tell this story to your mother?’ (ECC_16_M_1_086)

Example (8) shows the question particle *lah* which – as usual – follows the syntactic primary verb. The irrealis *nay-* weakens the obligation. The same seems to be the case in (9). This example also contains the new situation (newsit) marker *hɔj-* which is used for a “new (but expected) situation after a limit has been transgressed” (Jenny 2001:125).⁴ This means here that the time limit has been passed until which it is appropriate to stay away from home.

- (9) *ʔé nay-hɔj-lo ví sam bá:j moh séəm*
 1PL.INCL IRR-NSIT-to.need to.return CONC to.happen be.so night
hɔj-luk
 NSIT-be.late.in.the.night
 ‘We should go back home because it is almost midnight.’ (WPKT_14_L_399)

Another prefix used with *lo* is the durative marker *ʔun-* (10). The prefix is derived from the verb *ʔú:n* ‘stay’ and its final nasal adapts to the point of articulation of the following consonant. It marks a situation of which neither the initial nor the final boundary is known or of importance (Weymouth 2018:89-90). With the prefix, the following question does not concern only the actual situation, but it is a general one.

- (10) *cəʔuʔ pɔŋ dih hɔ:j ʔɬw kiʔ dɛ pɔʔ di hméə*
 book NML to.read to.finish 1SG NML ANA LOC place which
ʔun-lo ʔú:n
 DUR-to.need to.keep
 ‘Where should I keep the books, I have read?’ (EMSR_19_M_3_023)

The answers in example (11) also show the independence of *lo* as a secondary verb: *lo* can be used as the only word in the positive answer and negated in the negative one.

- (11) a. *ʔán lo lah dih cəʔuʔ ní*
 3SG to.need Q to.read book PROX
 ‘Does he have to read this book?’
 b. *lo (ʔɛ)* c. *ʔa:w-lo-maʔ*
 to.need ASRT NEG-to.need-NEG
 ‘Yes.’ ‘No.’ (EGGD_16_M_3_149)

2.1.2 *bɛ* ‘can’

The secondary verb *bɛ*, ‘can’ expresses a possibility which is given by external circumstances or an inherent ability (e.g., physical power) of the subject. This verb is often used together with the complementizer *lɬ* which then precedes the lexical primary verb as it is shown in Table 4 which contains all relevant examples with *bɛ* in the sample. To detect the trigger for the use of *lɬ*, the most probable features, namely the main verbs and their transitivity, the functions of *bɛ*, the polarity of the clause, and the consultants are listed in the table.

⁴ For Rumai see also Weymouth (2018:91-92).

Table 4: The use of the complementizer *laʔ* together with *be*

Compl.	Verb	English	Transitivity	Function	Polarity	Consultant
+	<i>kʰl</i>	‘marry’	transitive	possibility	neg	3
+	<i>ʔəŋ</i>	‘pass’	transitive	possibility	neg	3
+	<i>ʔiʔ</i>	‘sleep’	intransitive	possibility	neg	3
+	<i>təbuʔ</i>	‘run’	intransitive	inherent ability	pos	3
+	<i>təbuʔ</i>	‘run’	intransitive	inherent ability	neg	3
–	<i>hôm</i>	‘eat’ (rice)	transitive	inherent ability	neg	3
–	<i>hôm</i>	‘eat’ (rice)	transitive	inherent ability	neg	1
+	<i>hâ</i>	‘eat’ (curry)	transitive	inherent ability	neg	6
–	<i>caʔ</i>	‘start’	transitive	possibility	pos	1
–	<i>samâ</i>	‘can’	intransitive	inherent ability	neg	3
+	<i>lɔt</i>	‘free’	intransitive	possibility	neg	3
+	<i>kəpak</i>	‘break’	transitive	inherent ability	pos	3
+	<i>kəpak</i>	‘break’	transitive	inherent ability	neg	3
+	<i>kâ:ŋ</i>	‘block’	transitive	inherent ability	pos	3
+	<i>rêən</i>	‘do’	transitive	inherent ability	pos	3
+	<i>cʰeəp</i>	‘wear’	transitive	possibility	neg	3
+	<i>deəh</i>	‘come’	intransitive	possibility	neg	3
+	<i>deəh</i>	‘come’	intransitive	possibility	neg	6
–	<i>nà</i>	‘know’	transitive	inherent ability	neg	unkn. 2014
–	<i>tʰîŋ</i>	‘control’	transitive	inherent ability	neg	unkn. 2014
+	<i>vɔh</i>	‘open’	transitive	possibility	neg	9
–	<i>dó</i>	‘carry’	transitive	inherent ability	neg	9
–	<i>deh</i>	‘give’	transitive	possibility	neg	9

In this sample, none of the last four criteria plays a role in triggering the complementizer. This leaves the semantic main verb as the trigger, what is supported by the verbs *təbuʔ* ‘run’ (+), *hôm* ‘eat’ (–), *kəpak* ‘break’ (+) and *deəh* ‘come’ (+), which take the complementizer or not, independent of other criteria. But see also the discussion in 2.1.4 for *bîn* ‘get’.

The secondary verb *be*, replicated from the Shan modal verb *pê* ‘be able’,⁵ is used in example (12).

- (12) *ʔáw be laʔ təbuʔ kʰjá:n*
 1SG be.able COMP to.run be.fast
 I can run fast. (EC_17_M_3_020)

Besides the negative affixes, three verbal prefixes occur in the corpus with *be*, namely irrealis *nəŋ-*, newsit *hɔj-* and durative *ʔuN-*.

⁵ In Shan, the verb *pê* occurs both as lexical verb ‘to win’ and as modal ‘be able’ replicating Burmese *nain* with the same range of meanings.

- (13) *ʔân ɣh deəh mōj.ja gò moh baj ɣeən kô:n deʔ*
 3SG to.buy to.come mattress because be.so how alike child be.small
ciʔ nɔŋ-be lɔʔ ʔiʔ
 EMPH IRR-be.able COMP to.sleep
 ‘She bought a mattress so that the child can sleep.’ (EMSR_19_M_3_017)

In example (13), a future meaning of *nɔŋ-* is most appropriate and in (14) it expresses a possibility.

- (14) *peŋ⁶-be ʔɣ lɔʔ kəpak hɔ:j ʔu təkaj*
 2PL.IRR-be.able ASRT COMP to.break wood one CLF
 ‘You could break a piece of wood.’ (W3_15_M_3_024)

The newsit marker in (15) is used to express the new situation, that the speaker can begin to learn now, because she got the book the day before.

- (15) *ʔɔw bɣn deəh cəʔuʔ ʔuŋ.hmêəw ʔɔw hɔj-be caʔ*
 1SG to.get to.come book yesterday 1SG NSIT-be.able to.start
pənɔw bəə ʔuŋ.bjâ:n jâ:m.ŋjêŋ
 to.learn text Burmese now
 ‘I received the textbook yesterday, so I can start learning Burmese now.’
 (EGGD_16_M_1_189)

In (16), the durative marker *ʔuN-* indicates a general property of the mentioned water.

- (16) *ʔéəm dī ʔuŋ-be bô lɔʔ kâ:ŋ jôka vat bu-hâ:w*
 water MEDL DUR-be.able EMPH COMP to.block disease stomach NEG-to.go
 ‘That water can protect from constipation.’ (WA_15_M_3_003)

All negative markers, except of the prohibitive *k^hu-*, can be used with *be*. Two of them occur in (17).

- (17) *kê ɲam be-maʔ lɔʔ vɣh dih*
 3PL not.yet be.able-NEG COMP to.open DISC
 ‘They cannot yet open (school).’ (C3_17_MS_9_022)

In Rumai, the main constituent order is SV/AVP, but the greater part of the dependent clauses has VS/VAP. As the preverbal secondary verbs are the syntactic primary verbs, they also precede the subject in appropriate dependent clauses as illustrated in (18).

- (18) *ma ɲɔ gò be kô:n ʔân lɔʔ p^hje bəə*
 mother be.happy because be.able child 3SG COMP to.take.an.exam text
 ‘The mother is happy that her son passed the exam well.’ (EGGD_16_M_1_174)

⁶ Some of the personal pronouns merge with the irrealis marker like in (14) where *peŋ-* is a merger of *pê* (2PL) and *nɔŋ-*. Other merged pronouns in the corpus are *ʔuŋ-* from *ʔɔw* (1SG), *ʔeŋ-* from *ʔé* (1PL.INCL), *mɔŋ-* from *mâj* (2SG) and *keŋ-* from *kê* (3PL). For the other personal pronouns, there are no examples of merged forms. This may be due of their rare use or, like in the case of the third person singular pronoun *ʔân*, due to its probable coincidence with another morpheme, in this case with the first singular irrealis *ʔuŋ-*. In the corpus, there are also two mergers with pronouns and the negative prefix *ʔa:w-*, namely *ʔéw* from *ʔé* (1PL.INCL) and *kêw* from *kê* (3PL). The reasons why there are not more examples for mergers with other pronouns may be the same as for those with the irrealis *nɔŋ-* or because the merging process is less advanced. In the mergers with the irrealis *nɔŋ-*, the falling tone of the pronouns is lost, but with the negative prefix *ʔa:w-* it is maintained. This may be caused by the different vowel lengths in the two affixes, namely short in *nɔŋ-* and long in *ʔa:w-*.

2.1.3 *səmə* ‘know’

The verb *səmə* ‘know’ expresses a skilled ability. It may be used as a main verb (19) as well as a secondary verb (20).

- (19) *ʔân səmə pəna kʰəlɔw pəŋ.ŋá:j*
 3SG to.know education to.lie eye
 ‘He has magic power.’ (WPKT_14_L_385)
- (20) *tá kûŋ háj ní səmə dɔh kjəh cʰeəm*
 master city other PROX to.know to.say language Shan
 ‘This foreigner can speak Shan well.’ (EGGD_16_M_3_167)

Except for the negative affixes, no verbal affixes occur with *səmə* in the corpus. Furthermore, *səmə* is not alone in the scope of negation like the other preverbal secondary verbs, but the main verb is also included (21).

- (21) *ʔɔw tʰaŋ ʔân ʔa:w-səmə-təŋ-maʔ tɛw*
 1SG to.think 3SG NEG-to.know-to.cook-NEG curry
 ‘I think, he cannot cook curry.’ (EMR_19_M_3_094)

This means, *səmə* does not show all the features of a syntactic primary verb (see 2.1) and therefore it seems to be on an advanced stage on the grammaticalization path towards a prefix. This is also shown in the following question (22) where the question marker follows the main verb which therefore is also the syntactic primary verb.

- (22) *máj səmə-dɔh lɔh tɔʔ ʔân ʔiʔ lój tɛəŋ kətá:j*
 2SG to.know-to.say Q OBL 3SG to.sleep only on ground
 ‘Can you make the dog lie down on the ground?’ (EGGD_16_L_6_091)

In the previous two examples, *səmə* is glossed as a prefix due to syntactic reasons. Intonation studies will give more insight about the status of *səmə* also in clauses like (20) and (24), as prefixes are never stressed. However, sometimes *səmə* is directly followed by the question marker *lɔh* and it may be used as the only word in answering a yes-no question (23).

- (23) a. *máj səmə lɔh ló:j ʔɛəm*
 2SG to.know Q to.swim water
 ‘Can you swim?’ (WPKT_14_L_234)
- b. *(ʔɔw) səmə (ʔɔ)* c. *(ʔɔw) səmə-maʔ*
 1SG to.know ASRT 1SG to.know-NEG
 ‘Yes.’ (pc_3) ‘No.’ (pc_3)

The two secondary verbs *səmə* ‘know’ and *bɛ* ‘can’ may also be used together. The meaning is here that someone is skilled and may have the possibility to steal. In (24), the complementizer *lɔʔ* is used due to *bɛ* as *lɔʔ* never occurs with *səmə* alone.

- (24) *ʔi səmə bɛ lɔʔ bjá bɔn nɔ*
 others to.know be.able COMP to.steal to.get POL
 ‘Someone can steal it.’ (W1_023)

2.1.4 *bân* ‘get’

Rumai verbs allow one (S), two (A, P), or three (A, T, G) core arguments, what results in intransitive, transitive, and ditransitive clauses. In ditransitive clauses, the order is usually agent-verb-theme-goal (AVTG), but a complex theme mostly follows the goal (AVGT). This is illustrated in (25) and (26) where *bân* ‘get’ is used as a main verb. The third argument (goal/source) is always preceded by the oblique marker *tɔʔ* which can be translated by ‘from’ with *bân* or ‘to’ with the verbs *dɛh* and *tʰiʔw* ‘give’. Clauses with the verb *bân* ‘get’ have the same structure as clauses with the verbs *dɛh* and *tʰiʔw* ‘give’ and are therefore also ditransitive (27) in Rumai.

(25) ʔɔw **bân** cəʔuʔ ʔu ʔuʔ tɔʔ ʔân
 1SG to.get book one CLF OBL 3SG
 ‘I got a book from her.’ (ECC_16_M_1_088)

(26) kɛ **bân** tɔʔ ʔɔw klô:j ʔɛ pôŋ kâ:j plâ:j məʔu:n ʔu pôŋ
 3PL to.get OBL 1SG banana two CLF ADD fruit coconut one CLF
 ‘They’ll get two bananas and a coconut from me.’ (ECC_16_M_1_090)

(27) ma ʔɔw **dɛh** rân tɔʔ ʔɔw
 mother 1SG to.give silver OBL 1SG
 ‘My mother gave me money.’ (WPKT_14_L_063)

The verb *bân* in its function as a main verb is also used as the host of the desiderative prefix *siŋ-* to express the wish to get or have something (28).

(28) ʔɔw **siŋ-bân** ple kʰji ʔu cuŋ ʔɔ
 1SG DES-to.get earring gold one CLF POL
 ‘I want a pair of earrings.’ (WPKT_14_L_014)

Heine and Kuteva (2002:143-149) mention for several languages different grammatical functions of verbs with the meaning ‘get’, ‘receive’ or ‘obtain’, among them possibility and permission, the two functions which also occur in Rumai. Enfield (2003) describes the function of preverbal ‘get’ (‘acquire’) for several Southeast Asian languages as indicating a “result of prior event” (Enfield 2003:292). This is also the case in Rumai where for most of the examples the prior event is obvious or can be guessed. The secondary verb *bân* ‘get’ expresses a possibility (29) or a (request for) permission (30a).

(29) vâ:j hɔj-ŋɔh ʔɔw tɔʔ soho ʔɔw **bân** ʔiʔ pɔɔ
 after NSIT-to.hit 1SG OBL mosquito 1SG to.get to.sleep be.happy
 ‘After I had killed the mosquitos, I could sleep well.’ (EGGD_16_M_1_144)

(30) a. ʔɔw nɔŋ-**bân** lah hâ:w jâ:m.njɛn
 1SG IRR-to.get Q to.go now
 ‘May I go now?’ (EGGD_16_M_1_132)

b. *bân* ʔɔ to.get ASRT ‘Yes.’
 c. ʔa:w-*bân*-maʔ NEG-to.get-NEG ‘No.’ (EGGD_16_M_3_131)

The question particle *lah* in (30a) follows, as usual, directly the syntactic primary verb and the answers (30b and c) show the independence of *bân* as a secondary verb. In (30a), the irrealis marker *nɔŋ-* precedes *bân*. Here the prefix may express future, or it may soften the syntactic main verb for courtesy.

Contrary to Burmese, Shan, and Milne’s (1921) Palaung (Shwe) that are mentioned by Jenny (2015), the Rumai secondary verb ‘get’ does not express obligation, but lo ‘need’ is used (2.1.1). Only in negative clauses it may express a negative obligation, a (weakened) prohibition (31).

- (31) *mâj ʔa:w-bʃn-maʔ deəh hla pʌʔ μúŋ*
 2SG NEG-to.get-NEG to.come be.late LOC work
 ‘You shouldn’t come late to work.’ (WPKT_14_L_097)

A negated *bʃn* can also express a general prohibition (32).

- (32) *ʔa:w-bʃn lɔ:j ʔéəm*
 NEG-to.get to.swim water
 ‘No swimming!’ (WPKT_14_L_092)

In Rumai, *bʃn* as a secondary verb can take the durative prefix *ʔuN-* as well as the postverbal secondary verb *ʔú:n* ‘keep’ (2.2.3). The prefix and the secondary verb may occur alone (33) and (34) as well as together (35). However, the function of their interplay is not yet clear (Weymouth 2018:90).

- (33) *kjɔh ma ʔum-bu-bʃn-maʔ lʌʔ kjɔh*
 language RESTR DUR-NEG-to.get-NEG COMP to.speak
 ‘He could not speak anymore (because his tongue had disappeared).’ (WA_15_M_3_046)

- (34) *tɔ:n cʰeh tɔ:n cʰeh bʃn ʔú:n deəh hle pʌʔ kʰəleap ní*
 every CLF every CLF to.get to.keep to.come to.visit LOC house PROX
 ‘Everyone can visit this house.’ (EMR_19_M_3_015)

- (35) *kʰʌ bâ:j jəən ʔum-bʃn ʔú:n ví hóm pʌʔ seŋ dí*
 COND happen alike DUR-to.get to.keep to.return to.eat LOC shop MEDL
 ‘In this case, we can go again to eat in that shop.’ (EQB_17_M_3_063)

As well as *bɛ* ‘can’, *bʃn* is sometimes used with the complementizer *lʌʔ*. However, a similar research like in Table 2 for *bʃn* shows no concrete feature as *lʌʔ* may be used (36) or not (37) with the same main verb.

- (36) *mâj ʔa:w-bʃn-maʔ lʌʔ hâ:w pʌʔ plɔ:j pʰəjâ ʔuŋ.dih*
 2SG NEG-to.get-NEG COMP to.go LOC festival pagoda today
 ‘You can’t go to the temple fair today.’ (ECC_16_M_3_051)

- (37) *mâj ʔa:w-bʃn-maʔ hâ:w pʌʔ plɔ:j pʰəjâ ʔuŋ.dih*
 2SG NEG-to.get-NEG to.go LOC festival pagoda today
 ‘You can’t go to the temple fair today.’ (ECC_16_M_1_050)

Therefore, *lʌʔ* apparently is obligatory neither with *bʃn*, and due to the discussion in 2.1.2, nor with *bɛ*. There may be tendencies for the use of *lʌʔ*, circumstances where it is more easily dropped than in other ones, but the rather small samples do not allow a reasonable inference.

When *bʃn* expresses a possibility (38), it is often interchangeable with *bɛ* ‘can’ (39).

- (38) *gò moh ʔân kɔ:t ʔân bʃn-maʔ lʌʔ lâ:j ʔi.pân pʌŋ kjʃk ʔân*
 because be.so 3SG be.poor 3SG to.get-NEG COMP to.take woman NML to.like 3SG
 ‘He can’t marry the girl he likes because of his poverty.’ (ECC_16_M_1_062)

- (39) *gə moh ʔân pʰlâ:n jâ:w ʔân ʔaw-be-maʔ lʔ kʰl*
 because be.so 3SG be.poor very 3SG NEG-be.able-NEG COMP to.marry
tʔ ʔi.lôj pʌŋ kjʃk ʔân
 OBL virgin NML to.like 3SG
 ‘He can’t marry the girl he likes because of his poverty.’ (ECC_16_M_3_063)

Enfield (2003) describes another function of ‘get’, namely “descriptive complementation” (Enfield 2003:250-289). Jenny (2015) summarizes this function as follows: “Between two predicative elements, the second of which can be either a verb or an adverbial expression, ‘get’ frequently conveys the meaning that the activity denoted by the first verb is carried out with the result or to the extent described by the second predicating element” (Jenny 2015:188). This function occurs also in Rumai as illustrated in (40).

- (40) *kêən ʔʌw sêw hɔj-bʃn tʂ pəkjʃ*
 head 1SG to.hurt NSIT-to.get six moon
 ‘I’ve got a headache for six months now.’ (WPKT_14_L_130)

However, unlike the description given by Jenny (2015), the second predicative element seems to be restricted to temporal adverbial expressions in Rumai.

2.1.5 *dêh* and *tʰiw* ‘give’

The two verbs *dêh* and *tʰiw*, both glossed as ‘give’, have only a slightly different meaning. While the semantic representation of *dêh* is plainly that “A passes control over T to G” (Jenny 2015:159) (41), *tʰiw* has, according to native speakers, an additional connotation of ‘donate’ (42) and it is also the more polite form. When they are used as secondary verbs, the only difference seems to be that *dêh* is more common than *tʰiw*. Sidwell (2015) reconstructs for ‘give’ a proto-Palaungic **dêh* and similar forms of the verb still occur in several Palaungic languages (Sidwell 2015:137). The origin of *tʰiw* is unknown.

- (41) *sra dêh cəʔuʔ tʔ kô:n cʰôŋ*
 teacher to.give book OBL child school
 ‘The teacher gave the book to the student.’ (EGGD_16_M_3_004)
- (42) *ʔuŋ.dih ʔʌŋ-tʰiw tʔ pê kəʔâ:ŋ plâ:j nî*
 today 1SG.IRR-to.give OBL 2PL bone fruit PROX
 ‘Today, I will give you this seed.’ (WA_15_M_3_011)

A secondary verb with the basic meaning ‘give’ is widespread in the languages of the world, expressing different functions. An overview of these functions is given by Newman (1976), who calls the constructions “constructions with figurative GIVE” (Newman 1976:134). Jenny (2015) describes the preverbal secondary verb as follows: “Preverbal (or more accurately pre-clausal) ‘give’ is widespread in Southeast Asia to encode permissive and in some languages also jussive causative expressions” (Jenny 2015:164). Both functions are also present in Rumai with permissive in (43) and (44) and jussive causative in (45) and (46).

- (43) *ʔʌw dêh ʔân hâ:w pʌʔ kəlah*
 1SG to.give 3SG to.go LOC market
 ‘I allow him to go to the market.’ (EGGD_16_M_1_062)
- (44) *ʔʌw tʰiw ʔân hâ:w pʌʔ kəlah*
 1SG to.give 3SG to.go LOC market
 ‘I allow him to go to the market.’ (EGGD_16_M_3_061)

(45) *ʔəw deh ʔən hɑ:w pʌʔ kəlah*
 1SG to.give 3SG to.go LOC market
 ‘I order him to go to the market.’ (EGGD_16_M_1_083)

(46) *ʔəw tʰiw ʔən hɑ:w pʌʔ kəlah*
 1SG to.give 3SG to.go LOC market
 ‘I order him to go to the market.’ (EGGD_16_L_6_084)

The previous four examples do not show any difference in the meaning of *deh* and *tʰiw* in their function as preverbal secondary verbs. The sentences are all from a questionnaire which was answered by three consultants. All of them were in their twenties. Two of them, one female and one male, are from different townships, but their use of *deh* and *tʰiw* is similar, while the third one, male and from the same township as the other male consultant, uses *tʰiw* significantly more than the other two. However, although the female and the first male consultants are from different townships and the distance between their villages is longer, they are roughly from the same valley, but the second male is from a different one what complicates the contact. Therefore, the different origin of the consultants may explain the different use of the two secondary verbs.

Newman (1996) distinguishes among others a construction with ‘give’ as “cause” which he describes as follows: “Where the construction involves a causee (not restricted to humans or animates) which is made to exist or act in a new way, then the sense becomes more like a true causative, as opposed to a “manipulative”” (Newman 1996:175). This construction also occurs in Rumai, and as illustrated in (47) and (48), without any difference between the two secondary verbs.

(47) *ʔəəm dī ʔun-deh hũ tɔ.pɔŋ ʔɛ ʔunɣɣ ʔunɣəm*
 water MEDL DUR-to.give skin body 1PL.INCL be.light be.soft
 ‘That water makes our skin light and soft.’ (WA_15_M_3_004)

(48) *ʔəəm dī ʔun-tʰiw hũ tɔ.pɔŋ ʔɛ ʔunɣɣ ʔunɣəm*
 water MEDL DUR-to.give skin body 1PL.INCL be.light be.soft
 ‘That water makes our skin light and soft.’ (pc_3)

Examples (47) and (48) show one of the rare uses of the durative marker *ʔun-* with the secondary verbs *deh* and *tʰiw*. Another aspectual prefix which is used, but also rarely, is the irrealis marker *nɔŋ-* as seen in (49) and (50) where *nɔŋ-* is *mâj nɔŋ-* (2SG IRR-).

(49) *mɔŋ-deh lɔh ʔəw tɣh kâ tɔj*
 2SG.IRR-to.give Q 1SG to.look fish DIST
 ‘Can you let me see that fish over there?’ (ECC_16_M_3_047)

(50) *mɔŋ-tʰiw lɔh ʔəw ŋɔh puŋ tɔʔ mâj*
 2SG.IRR-to.give Q 1SG to.hit picture OBL 2SG
 ‘Would you allow me to take a photo of you?’ (EGGD_16_L_6_057)

According to Jenny (2015), desiderative and prohibitive markers in Southeast Asian languages are, or diachronically were, verbal. Therefore, he calls ‘give’ in desiderative and prohibitive constructions “interverbal” (Jenny 2015:164). But as for the two markers in Rumai a verbal origin is not anymore traceable and they are affixes today, the two constructions are here described together with the preverbal secondary verbs.

- (51) *ʔəkʰaw di siŋ-dəh baw.pʰuŋ ʔân bjâ ʔjɛ̃*
 guy MEDL DES-to.give friend 3SG to.steal chicken
 ‘The boy wants his friend to steal a chicken.’ (EMSR_19_M_3_053)
- (52) *ʔaw siŋ-tʰiw ʔu:n mâj nà tʰa:n.gô:j sêw nɔh ʔaw*
 1SG DES-to.give to.keep 2SG to.know amount to.hurt mind 1SG
 ‘I want you to know how much it hurt me.’ (WPKT_14_L_290)

Enfield (2009) calls ‘give’ in examples like (51) and (52) “dummy causative” because in Lao ‘give’ is necessary to indicate switch-reference in such desiderative clauses (Enfield 2009:811). However, it is not clear whether this explanation also counts for Rumai. As Jenny (2015) states, dummy causatives and desiderative causatives do not differ in many languages (Jenny 2015:164). This is also the case in Rumai where the previous two examples also can be translated as ‘the boy wants to let his friend steal a chicken’ and ‘I want to let you know how much it hurt me’. Therefore, the dummy causative may be a secondary development from the desiderative causative (Jenny 2015:164). Then, the verb ‘give’ has been retained, what at least synchronically is necessary in Rumai as *siŋ-* is fully grammaticalized and not anymore a free form. Example (53) shows a negated desiderative causative.

- (53) *ma ʔaw siŋ-dəh-maʔ kô:n ʔân hâ:w pɔʔ ranɣuŋ*
 mother 1SG DES-to.give-NEG child 3SG to.go LOC Yangon
 ‘My mother does not want to let her children go to Yangon.’ (EMR_19_M_3_090)

In the prohibitive desiderative as well, there is apparently no difference between *dəh* (54) and *tʰiw* (55).

- (54) *kʰu-dəh ʔuŋ-kəməh*
 PROH-to.give DUR-to.disappear
 ‘Don’t let it disappear.’ (WA_15_M_3_079)
- (55) *kʰu-tʰiw ʔuŋ-kəməh*
 PROH-to.give DUR-to.disappear
 ‘Don’t let it disappear.’ (PC_3)

Different to the previous two examples, the direct object is not dropped in (56).

- (56) *kʰu-dəh kjaʔ mak leəp kʰêəŋ lôŋ njeəm*
 PROH-to.give buffalo cow to.enter in field tea
 ‘Don’t let the buffaloes and the cows come into the tea farm.’ (WPKT_14_L_112)

2.2 The postverbal secondary verbs

Unlike the preverbal secondary verbs, the postverbal ones are syntactically dependent on the main verb or the preverbal secondary verb. However, as they are phonologically independent, the free form is preferred in this paper. The secondary verbs *dəh* / *tʰiw* ‘give’, *hò:j* ‘finish’, *ʔu:n* ‘keep’, *hâ:w* ‘go’, and *deəh* ‘come’ are described and discussed in the following sections, including co-occurring verbal affixes.

2.2.1 *dəh* and *tʰiw* ‘give’

The use of *dəh* and *tʰiw* ‘give’ as full verbs is illustrated in examples (25) and (26) in section 2.1.5. Like in most Southeast Asian languages (Jenny 2015:161), the postverbal secondary verbs with the lexical meaning ‘give’, *dəh* (57) and *tʰiw* (58), have a benefactive function in Rumai.

(57) *kəlâ:ŋ dɛh hâ:ŋ ɹu pəca:ŋ*⁷
 to.tell to.give POL name monk
 ‘Please tell me your name.’ (I_15_M_1_001)

(58) *ʔân rêən tʰɪw ɹûŋ kô:n vâ*
 3SG to.do to.give work child younger.sibling
 ‘He worked for the children.’ (WA_15_M_3_015)

There is no obvious difference between postverbal *dɛh* and *tʰɪw*. Both can be used in similar environments and with nine (*dɛh*) respectively eleven (*tʰɪw*) examples in the corpus, none of them is more common there.

In presence of a preverbal secondary verb, postverbal ‘give’ keeps its position after the main verb (59).

(59) *ʔâw lo tâŋ tʰɪw ʔi.ŋê kê*⁸
 1SG to.need to.cook to.give guest 3PL
 ‘I need to cook for the guests.’ (EMR_19_M_3_084)

Another benefactive construction, about equally frequent and used in the same environments as the construction with ‘give’, contains the purposive particle *ca*. This particle always directly precedes the beneficiary (60).

(60) *ʔâw rêən ɹûŋ ca ʔuŋ.hôm.ʔuŋ.côŋ ʔâw*
 1SG to.do work PURP family 1SG
 ‘I work hard for my family.’ (EGGD_16_M_1_017)

A combination of postverbal ‘give’ and *ca* is also possible (61).

(61) *tâj ʔâw kəpa:k ʔân təkjeət dɛh ca ʔâw ɹaʔ səŋfâj*
 hand 1SG to.break 3SG to.tidy.up to.give PURP 1SG LOC kitchen
 ‘My hand was broken, so she cleaned the kitchen for me.’ (EGGD_16_M_1_020)

Example (62) illustrates *ca* in a purposive clause.

(62) *ʔâw ʔa:w-siŋ-sôm-maʔ rŋn kʰjî ca ɹh tɹʔ təsi*
 1SG NEG-DES-to.waste-NEG silver gold PURP to.buy OBL cloth
 ‘I don't want to waste my money on buying cloths.’ (WPKT_14_L_427)

As the purposive use of *ca* is more common than its benefactive use, the particle is constantly glossed as “PURP”.

2.2.2 *hə:j* ‘finish’

The verb *hə:j* ‘finish’ as a main verb, unlike in many languages in the region (e.g., Shan and Burmese), is a transitive verb in Rumai (63) and (64).

⁷ *pəca:ŋ* ‘monk’ is used as a second person singular possessive pronoun here. For more information about personal pronouns in Southeast Asia see Müller and Weymuth (2017). In Rumai, there is no morphological difference between personal and possessive pronouns.

⁸ The third person plural pronoun *kê* can be used as a plural marker for humans and sometimes also for animals.

⁹ The direct object in verb-initial dependent clauses is always preceded by the oblique marker *tɹʔ* (see also example (64))

- (63) *ʔân hɔj-hɔ:j c^hôŋ dih*
 3SG NSIT-to.finish school DISC
 ‘He has already finished school.’ (ENS_16_M_1_029)
- (64) *ʔà:j hɔ:j mâj tɔ? muŋ mâj ʔa:w-bʂn-ma? lɔ? hâ:w pɔ? plɔ:j*
 before to.finish 2SG OBL work 2SG NEG-to.get-NEG COMP to.go LOC festival
 ‘You cannot go to the festival before you have finished work.’ (EMR_19_M_3_105)

The secondary verb *hɔ:j* indicates a new situation (newsit), this means, a “new (but expected) situation after a limit has been transgressed” (Jenny 2001:125) or it emphasizes that an action has been completed. Which of the two functions is expressed in a clause seems only to be dependent on the context and sometimes both readings are possible. In example (65), *hɔ:j* is used as a newsit marker as the new situation, namely going to sleep, is still in progress.

- (65) *gò moh ʔân gi:j-ʔi? məmɔk ʔân ʔi? bî hâ:w hɔ:j*
 because be.so 3SG IPFV-to.sleep to.doze 3SG to.sleep to.forget to.go to.finish
 ‘As he is sleepy, he goes to sleep.’ (EMR_19_M_3_001)

The secondary verb *hɔ:j* as a newsit marker can also be used with stative verbs (66).

- (66) *ʔêəm dâ:ŋ hɔ:j t^hu.hlɛw vun do? pɔŋ tì sam kê*
 water be.big to.finish to.destroy to.throw be.exhausted NML to.plant to.plant 3PL
 ‘The flood damaged what they had planted.’ (WPKT_14_L_415)

A completive reading of *hɔ:j* occurs in (67) as, according to the consultant who translated the sentence, the prefix *hɔj-* which has exclusively a newsit function is not possible here. Moreover, the durative prefix *ʔun-* with the verb *lo* ‘need’ indicates a general question, not only concerning the books just read (see also (10)).

- (67) *cəʔu? pɔŋ dih hɔ:j ʔɔw ki dʂ pɔ? di hméə ʔun-lo ʔû:n*
 book NML to.read to.finish 1SG NML ANA LOC place which DUR-to.need to.keep
 ‘Where should I keep the books, I have read?’ (EMSR_19_M_3_023)

In (68), *hɔ:j* can have newsit or completive function, depending on the viewpoint:

- (68) *ʔân dih hɔ:j cəʔu? nî*
 3SG to.read to.finish book PROX
 ‘He has already read this book.’ (ETMA_17_M_3_017)

When a newsit reading is possible in a clause, the prefix *hɔj-* and the secondary verb *hɔ:j* are freely interchangeable. Examples (69) and (70) are elicited from the same clause but translated by different consultants.

- (69) *ʔân hóm hɔ:j do?*
 3SG to.eat to.finish be.exhausted
 ‘He has eaten it all.’ (EEN_16_M_3_022)
- (70) *ʔân hɔj-hóm do?*
 3SG NSIT-to.eat be.exhausted
 ‘He has eaten it all.’ (EEN_16_M_1_023)

For emphasis, sometimes the prefix *hɔj-* and the secondary verb *hɔ:j* are used together (71). Which of the two markers emphasizes the other one still must be clarified.

- (71) *ʔân hɔj-meeh hɔ:j dû:n*
 3SG NSIT-to.exist to.finish sweetheart
 ‘He already has a girlfriend.’ (WPKT_14_L_391)

Most of the postverbal secondary verbs are usually omitted in questions and in negative clauses. This is also the case for *hɔ:j*, but here, the particle *hɔʔ* can be used as a question particle to ask whether something has already been done (72a). According to native speakers, *hɔʔ* means ‘to stop’ or ‘be all’. This and the form of the particle are hints that *hɔʔ* may be derived from the verb *hɔ:j* ‘finish’. However, the postverbal position of *hɔʔ* in interrogative clauses implies a derivation from the postverbal secondary verb *hɔ:j*. This would be unusual as *hɔʔ* is an independent morpheme while *hɔ:j* in this function is not. Here also, more research is needed.

- (72) a. *mâj ɣh hɔʔ kâ*
 2SG to.buy PTCL car
 ‘Did you already buy the car?’
- b. *ɣh hɔ:j* c. *ɳjam ɣh-maʔ*
 to.buy to.finish not.yet to.buy-NEG
 ‘Yes.’ ‘No.’
- d. *hɔʔ* e. *ɳjam*
 PTCL not.yet
 ‘Yes.’ ‘No.’ (pc_3)

The positive answer in (72b) includes postverbal *hɔ:j* which shows that it is implicitly present in the question. Short answers to the question are *hɔʔ* ‘yes’ and *ɳjam* ‘no/not yet’ (72d and e). The negative answer shows that *ɳjam*, unlike it is stated in Weymouth (2018:96), is an independent morpheme and not a verbal affix.

hɔʔ is also used as a clause-final particle (73). For both functions, a derivation from the lexical verb *hɔ:j* ‘finish’ is reasonable. However, the position of *hɔʔ* in example (72) implies a derivation from the postverbal secondary verb, but this would be unusual as *hɔʔ* is an independent morpheme, while *hɔ:j* as a secondary verb is not. Another question is, whether postverbal and clause-final *hɔʔ* have the same source, but this needs more research.

- (73) *mâj moh baw.p^hûŋ kameh ʔlŋ ʔək^hâj hɔʔ*
 2SG be.so friend to.love 1SG really PTCL
 ‘You are a very good friend.’ (W1_010)

2.2.3 *ʔû:n* ‘keep’

Like most of the secondary verbs, *ʔû:n* can be used as a main verb (74).

- (74) *ʔân bu:-nɔ-maʔ ɳɔʔ di hméə ʔân ʔû:n cəʔuʔ ʔân*
 3SG NEG-to.know-NEG LOC place which 3SG to.keep book 3SG
 ‘He doesn’t know where he has put his book.’ (EMR_19_M_3_098)

The basic function of the secondary verb *ʔû:n* ‘keep’ is to express duration without any importance of the initial and the final boundaries. This may be an ongoing situation like the shining of the red light in (75) or an ongoing action (76).

(75) *nəŋ-meəh ɲáj rɬ:n pla:ŋ ʔú:n*
 IRR-to.exist fire be.red be.bright to.keep
 ‘There may be a red light shining.’ (W2_15_M_3_014)

(76) *ʔân sɬm cʰôŋ ʔú:n bô ʔərok*
 3SG CONC to.feed to.keep EMPH frog
 ‘However, he was also breeding a frog.’ (PFWAY_15_M_5_003)

For how long a situation or an action last does not matter, as a red traffic light (75) only shines for seconds or minutes. In the previous examples, although the boundaries are not important, there are logical ones. However, there are also situations marked by *ʔú:n* which have no obvious boundaries, and which therefore can count as a general truth (77).

(77) *rʃn ʔa:w-deh-maʔ tu.ʔi kê pɲɔ kʰɬ moh rêən ʔi pen*
 silver NEG-give-NEG person 3PL happy COND be.so do NML good
deh-ʔú:n tu.ʔi kê pɲɔ
 give-keep person 3PL happy
 ‘Money doesn’t make people happy, but good deeds do.’ (ECC_16_M_3_016)

In (78), *ʔú:n* converts the permission into a general one while in (79) the permission is only for a specific occasion.

(78) *ʔɬw tʰiw ʔú:n ʔân hâ:w pɬʔ kəlah*
 1SG to.give to.keep 3SG to.go LOC market
 ‘I allow him to go to the market.’ (EGGD_16_M_3_061)

(79) *ʔɬw deh ʔân hâ:w pɬʔ kəlah*
 1SG to.give 3SG to.go LOC market
 ‘I allow him to go to the market.’ (EGGD_16_M_1_062)

The secondary verb *ʔú:n* occurs with all the aspectual prefixes except the inceptive prefix *ʔə-* (Weymouth 2018:90-91) which, due to its property to focus on the beginning of an event, is not compatible with *ʔú:n*. Example (80) includes the newsit prefix *hɲj-* which refers to the new situation after a change of state, and *ʔú:n* indicates that this situation lasts for some time.

(80) *pəloh hɲj-vrɬ ʔú:n*
 door NSIT-to.open to.keep
 ‘The doors are open, at last.’ (ENS_16_M_1_102)

The experiential aspect is expressed by the prefix *tɬm-* (81).

(81) *ʔɬw tɬm-pənɬw ʔú:n kjɔh ɲpa:n*
 1SG EXP-to.learn to.keep language Japan
 ‘I have been learning Japanese yet.’ (EQB_17_M_3_067)

Both the imperfective prefix *gi:j-* (82) and the durative prefix *ʔuN-* (83) can be used with the secondary verb *ʔú:n*.

- (82) *mây gi:j-kəcu? ʔû:n lôj ʔuʔsa*
 2SG IPFV-to.gather to.keep only property
 ‘You only gather property.’ (W1_023)

- (83) *ʔum-bɛn ʔû:n pʰon pʰɑ:n héŋ*
 DUR-to.get to.keep four five thousand
 ‘We get four or five thousand.’ (C3_17_MS_9_150)

The two prefixes and the secondary verb are similar as they do not focus on the boundaries of a situation or an action, but there are also differences: While the prefix *gi:j-* and the secondary verb *ʔû:n* occur with activity as well as with stative verbs, the prefix *ʔuN-* is only used with stative verbs. But also, *gi:j-* and *ʔû:n* differ, in that *gi:j-*, in contrary to *ʔû:n* can be used as a continuous marker (84).

- (84) *caŋ.naj ʔɔw gi:j-jxh tɛw.hlâ*
 now 1SG IPFV-to.buy vegetable
 ‘Now I am buying vegetables.’ (ENEG_13_M_1/2_054)

The two prefixes are also used together (85).

- (85) *ʔuŋ-gi:j-moh lôj baj ɟəŋ*
 DUR-IPFV-be.so only how alike
 ‘It is always like this.’ (I_17_MS_7_029)

The previous examples show that any combination of the two prefixes and the secondary verb are possible, except of the combination of all three markers. The effect of the use of two of the markers together is not yet clear. Most probable is that one of them emphasizes the other one, but this needs more research.

The two modal prefixes, the irrealis *nɔŋ-* and the desiderative *siŋ-* also occur with *ʔû:n*. In (86), *nɔŋ-* expresses an intention and in (87), future.

- (86) *ʔuŋ.hmɛəw kʰɔ bu:-klâ:j ʔɔw nɔŋ-hâ:w ʔû:n pɔʔ kəlah*
 yesterday COND NEG-to.rain 1SG IRR-to.go to.keep LOC market
 ‘If it hadn’t rained yesterday, I would have gone to the market.’ (ECC_16_M_1_032)

- (87) *pəkjɛ bɔ:n ʔɔw nɔŋ-set ʔû:n pʰôŋ tɔʔ mâj*
 moon be.back 1SG IRR-to.ring.up to.keep phone OBL 2SG
 ‘I will call you up next month.’ (WPKT_14_L_295)

- (88) *ʔɔw siŋ-hâ:w ʔû:n pɔʔ kəlah ʔuŋ.hmɛəw*
 1SG DES-to.go to.keep LOC market yesterday
 ‘I wanted to go to the market yesterday.’ (ECC_16_M_3_081)

While in (88), the wish may have been ongoing during the day, an ongoing situation or action is not recognizable in (86) and (87). Duration, especially an unlimited one (general truth), can convey certainty. Therefore, one of the developments of *ʔû:n* may be an emphasis marker to convey more certainty about the situation or the action.

In clauses with a preverbal secondary verb, postverbal secondary verbs usually directly follow this one (89), but it is also possible to attach them to the main verb (90). According to native speakers, there is no change in meaning. However, *ʔû:n* may here emphasize either the secondary verb or the main verb.

- (89) *ʔáw siŋ-tʰíw ʔú:n máj nà tʰa:n.gô:j sêw nɔh ʔáw*
 1SG DES-to.give to.keep 2SG to.know amount to.hurt mind 1SG
 ‘I want you to know how much it hurt me.’ (WPKT_14_L_290)
- (90) *ʔáw siŋ-tʰíw máj nà ʔú:n tʰa:n.gô:j sêw nɔh ʔáw*
 1SG DES-to.give 2SG to.know to.keep amount to.hurt mind 1SG
 ‘I want you to know how much it hurt me.’ (PC_3)

Example (91) shows that a clear cut between the durative and the emphasizing function of *ʔú:n* is not always possible. *ʔú:n* implies here that the boy not only once pushed or pulled the door but made effort to open it for some time. This may be interpreted as a durative action as well as emphasis in the sense that the boy really tried to open the door.

- (91) *kô:n dɛʔ kjô.kjá ʔú:n lɔ vrh mòj ploh ʔa:w-bɿn-maʔ*
 child small to.make.effort to.keep COMP to.open mouth door NEG-to.get-NEG
lɔ vrh
 COMP to.open
 ‘The boy tried to open the door, but it wouldn’t open.’ (ECC_16_M_3_020)

Even if the secondary verb *ʔú:n* is present in a positive clause (92), it is usually dropped in the corresponding negative clause (93).

- (92) *ʔáw siŋ-há:w ʔú:n ranɡuŋ sənám bɛ:n*
 1SG DES-to.go to.keep Yangon year be.back
 ‘I want to go to Yangon next year.’ (ENEG_13_M_1/2_106)
- (93) *ʔáw siŋ-há:w-maʔ ranɡuŋ sənám bɛ:n*
 1SG DES-to.go-NEG Yangon year be.back
 ‘I don’t want to go to Yangon next year.’ (ENEG_13_M_1/2_107)

2.2.4 *há:w* ‘go’

The verb *há:w* ‘go’ generally denotes a movement away from the center of interest which usually is the speaker, but this center may be changed according to the context (Jenny 2005:167). In example (94), an inhabitant of a Rumai village talks about other villagers who went to China for work.

- (94) *ʔi há:w sənám ní*
 others go year PROX
 ‘Other people go this year.’ (C3_17_MS_9_093)

A motion away from the center of interest is also the basic meaning of the secondary verb *há:w* ‘go’ in Rumai which then, according to the context and the intent of the speaker, can take over several functions. These functions often cannot be assigned to a traditional aspectual category, but all of them are traceable back to the basic meaning. This is a widespread feature of secondary verbs with the meaning ‘go’ in Southeast Asia (Bisang 1992), occurring for example in Burmese (Jenny and Hnin Tun 2016:204-205), Shan,¹⁰ and Mon (Jenny 2005:167-170). When *há:w* is used with motion verbs, it denotes the spatial direction of the movement away from the center of interest and often towards an overt goal as seen in (95) and (96). In this function, *há:w* can be called a directional marker.

¹⁰ From data in a toolbox project for Shan drawn up by the author, and mainly based on written short stories.

(95) *kê ci? leəp hâ:w kətêw paŋla:j*
 3PL EMPH to.enter to.go in sea
 ‘They went into the sea.’ (PCHT_16_L_6_049)

(96) *ʔâw kətʰɔ:k hâ:w pa? kəjôŋ təmâ:j*
 1SG to.move to.go LOC village be.new
 ‘I moved to another village.’ (ENS_16_M_3_054)

With a temporal point of reference, *hâ:w* can be interpreted as a continuous marker denoting an ongoing action or situation which starts at the time of reference. In (97), this is the time when the Palaung had started to plant tea while in (98) it is the time of the spirit’s speech.

(97) *ʔê ti sɔm hâ:w kap pen*
 1PL.INCL to.plant to.plant to.go be.tight be.good
 ‘We plant well.’ (N2_17_MS_7_005)

(98) *kê ma gô:j hâ:w baj dɔh kənɔ:m rukkʰa? tɔ? kê*
 3PL RESTR to.stay to.go how to.say spirit arboreal OBL 3PL
 ‘They lived like the spirit of the tree had told them.’ (W3_15_M_3_027)

The center of interest of the first *hâ:w* in (99) is the place from where the person rides away (directional) while for the second one it is the time of the departure (continuous).

(99) *ʔân bak hâ:w pʰɔh ʔumbjɔŋ kʰôj-kʰêəm hâ:w*
 3SG to.ride to.go only horse be.glad-EUPH to.go
 ‘He rode away and was very happy.’ (WA_15_M_3_055)

With verbs describing an action other than a movement, sometimes two interpretations are possible: one taking V + *hâ:w* as a single event and the other describing two consecutive events in the sense of ‘to V and then go away’. In (100) this would mean something like “reaping away” or to reap and then go away with the harvest.

(100) *hɔŋɔw ki? si:n dʰ ma ʔi sok hâ:w*
 paddy NML be.ripe ANA RESTR others to.reap to.go
 ‘The paddy that was ripe, someone had reaped it.’ (W3_15_M_3_006)

In (101), the main verb *jɔ:m* ‘die’ itself implies a departure, and *hâ:w* may be used here to reinforce the impression of going away.

(101) *ʔân gô:j lôj pʰɔ.n pəkjɔ? ʔân jɔ:m hâ:w*
 3SG to.stay only five moon 3SG to.die to.go
 ‘When she was only five months old, she died.’ (C3_17_MS_9_046)

Several verbal affixes can be used together with *hâ:w*. In (102), the inceptive prefix *ʔə-* emphasizes the beginning of the action. The two interpretations mentioned for (100) are also possible in (102).

(102) *saw kətâ:j dī kê dɔh saw kətâ:j dī ʔə-jok hâ:w*
 dog ground MEDL 3PL to.say dog ground MEDL INCEP-to.lift to.go
kô.n ʔjɔ? dʰ
 child chicken ANA
 ‘“That fox”, they said, “that fox has taken away that chick.”’ (PCHT_16_L_6_014)

In (103), the irrealis marker *nəŋ-* denotes a possibility.

- (103) *kʰʌ kʰû deəh pjuh pʰʌh nəŋ-hlɛw hâ:w lɿj hʌ?*
 COND wind to.come to.blow only IRR-to.destroy to.go only PTCL
 ‘If the wind blows, it will destroy it.’ (W3_15_M_3_023)

The imperfective marker *gi:j-* can be used with the secondary verb *hâ:w*. This supports the hypothesis, that *hâ:w* has also a continuous function (104).

- (104) *gi:j-bâ:j hâ:w lôj ʔi hŋjêən*
 IPFV-to.act to.go only NML DEM
 ‘(So, we) continue like that (with this green tea). (N2_17_MS_7_017)

The postverbal secondary verbs *hâ:w* and *deəh* ‘come’ (2.2.5) occur, although rarely, in negated clauses as it is illustrated in (105) for *hâ:w*.

- (105) *bʌw.pʰûŋ ʔʌw siŋ-deh ʔʌw t̄ hâ:w k̄ ʔʌw ʔa:w-siŋ-t̄*
 friend 1SG DES-to.give 1SG to.follow to.go 3PL 1SG NEG-DES-to.follow
hâ:w-ma?
 to.go-NEG
 ‘My friends want me to go with them, but I don’t want to.’ (EMSR_19_M_3_036)

The secondary verb *hâ:w* always precedes the secondary verb *h̄:j* (106) which indicates a new situation or the completion of an action (see section 2.2.2).

- (106) *ʔʌn vī hâ:w h̄:j*
 3SG to.return to.go to.finish
 ‘He had already gone back.’ (ENS_16_M_3_058)

2.2.5 *deəh* ‘come’

The verb *deəh* ‘come’ is the conceptual opposite of the verb *hâ:w* ‘go’ (Jenny 2005:173) denoting a movement towards the center of interest. In (107), the center of interest (goal) is not explicitly mentioned, but it is the place where the speaker arrived at that day.¹¹

- (107) *ʔun.dih ʔʌw deəh pʌ? kûŋ mantələ*
 today 1SG to.come LOC city Mandalay
 ‘Today, I came from Mandalay.’ (WPKT_14_L_206)

The basic meaning, a motion towards the center of interest, is always present in sentences including the secondary verb *deəh* ‘come’. And similar to *hâ:w* ‘go’, dependent on the context and the intention of the speaker, the function of *deəh* can fulfill several functions which sometimes may be assigned to a traditional aspectual category, but often this is not the case. A secondary verb with the meaning ‘come’ is like the secondary verb ‘go’ widespread in Southeast Asia described for example for Burmese in Jenny and Hnin Tun (2016:387-388) and for Mon in Jenny (2005:173-177). With motion verbs, *deəh* denotes the spatial direction of the movement towards the center of interest (108) and it may therefore be called a directional marker.

¹¹ As *pʌ?*, dependent on the context, can mean ‘to’ or ‘from’, the sentence could also be translated as ‘Today I came to Mandalay’. In this case, the center of interest would be Mandalay.

- (108) *fâ ci vî deəh pʌʔ kʰəleəp ʔân*
 monkey TCL to.return to.come LOC house 3SG
 ‘The monkey came back to his house.’ (W1_018)

Like the construction V + *hâ:w* ‘go’ (see (100)), the construction V + *deəh* may sometimes be interpreted as a single event or as two consecutive actions. For (109) this would be a simple ‘take’ or to grip the mirror and then move it towards the face.

- (109) *ʔân teəh deəh cʌm ca hnâ:ŋ jêəw ʔân tʌʔ ʔân*
 3SG to.take to.come glass PURP to to.see 3SG OBL 3SG
 ‘He took a mirror to see himself.’ (EMSR_19_M_3_089)

The two different interpretations are also possible in the following imperative clause (110). Besides ‘give’ like in the translation, this would be ‘draw the water and then bring it to me’.

- (110) *dʌk deəh hʌ:ŋ ʔêəm tʌʔ ʔʌw ʔu bɔj*
 to.draw.water to.come POL water OBL 1SG one cup
 ‘Please, give me a cup of water.’ (WPKT_14_L_196)

In (111), an ongoing situation which holds at least until the temporal center of interest (the time of speaking) is described. This implies a continuous function for *deəh*.

- (111) *kê dʌh bɔ hla sʌw kətâ:j nî kâ:j ʔjɛ ci nî*
 3PL to.say when be.late dog ground PROX ADD chicken TCL PROX
kâ:j gô:j deəh ʔu di hɔj-hlʌ:n
 3DU to.stay to.come one place NSIT-be.long
 ‘They said later, this fox and the chicken, they both lived together for a long time.’
 (PCHT_16_L_6_072)

With quality verbs, *deəh* denotes a development towards the center of interest which is the quality itself (112).

- (112) *kê ʔjɔ deəh rêən kuso*
 3PL be.happy to.come to.do merit
 ‘They become happy because of good deeds.’ (ECC_16_M_1_064)

The same is illustrated in (113). Additionally, the newsit prefix *hɔj-* and the postverbal secondary verb *hɔ:j* ‘finish’ (see 2.2.2) denote the new situation that the quality is attained now.

- (113) *ʔəkʰʌw dî ʔân hɔj-dâ:ŋ deəh hɔ:j*
 guy MEDL 3SG NSIT-be.big to.come to.finish
 ‘That boy has grown up now (he is adult).’ (EMR_19_M_3_003)

Another extension of the basic meaning is illustrated in (114) where *deəh* denotes a situation which is still relevant at the time of utterance which means here that what the subject has seen in the forest is relevant for the speaker.

- (114) *si mâj jêəw deəh kətêw bjâj nɔj*
 what 2SG to.see to.come in forest UP
 ‘What did you see in that forest?’ (WPKT_14_L_021)

A similar function occurs in (115). The sentence is a reply to the utterance ‘I thought you would come over for lunch yesterday’. The inceptive prefix *ʔə-* denotes the upcoming of the situation which prevented the speaker from going to the lunch and *deəh* denotes that this reason is still relevant for the addressee.

- (115) *ʔə-meəh deəh μûŋ dâ:ŋ pʌʔ kʰêəŋ kûŋ*
 INCEP-to.exist to.come work be.big LOC in city
 ‘I had some urgent things to do in town.’ (EMSR_19_M_3_067)

As already mentioned, *deəh* can also be used in negative contexts. The main verb and the secondary verb are in the scope of negation (116).

- (116) *ʔʌw ʔa:w-ʃh deəh-maʔ pʌŋ dʌh mâj ŋəh dâ:ŋ jâ:w*
 1SG NEG-to.buy to.come-NEG NML to.say 2SG price be.big very
 ‘I did not buy what you asked me because it is very expensive.’ (WPKT_14_L_405)

In example (117), the durative prefix *ʔuN-* denotes a generic situation. The meaning may be here, that when someone is skilled in something, he knows everything about that topic. A further development and with this a motion towards the center of interest is not possible anymore.

- (117) *pʌŋ səmâ ʔân ʔum-bu-kətah deəh-maʔ*
 NML to.know3SG DUR-NEG-to.develop to.come-NEG
 ‘What he is skilled in cannot develop anymore.’ (WA_15_M_3_010)

3 Summary

Rumai secondary verbs can be divided into two categories, preverbal and postverbal. The preverbal secondary verbs mainly belong to the modality domain while the postverbal ones mainly belong to the aspectual domain. Two of the verbs, *dəh* and *tʰiŋw*, both meaning ‘give’, belong to both categories. Their domain is valency with causative function in preverbal and benefactive function in postverbal position. All the secondary verbs, except *bɛ* ‘can’ which is a loan from Shan, are also used as main verbs and, although they are grammaticalized, there is no coevolution between form and meaning (Bisang 2004), as the forms of main and secondary verbs don’t differ.

The preverbal secondary verbs are the syntactic primary verbs in a clause, and they therefore behave like main verbs in absence of a preverbal secondary verb. This mainly means that they take the verbal affixes, are directly followed by the question marker *lʌh*, and can be used as the only word in the positive answer to a yes-no question. One of the preverbal secondary verbs, *səmâ* ‘know’, is sometimes included together with the main verb in the scopes of negation or interrogation, and therefore, it seems to be on an advanced stage towards a prefix.

Unlike the preverbal secondary verbs, the postverbal ones are syntactically dependent either on a main or a preverbal secondary verb. Furthermore, the verbs of the aspectual domain can hardly be assigned to a traditional aspectual category, but there are basic functions from which, dependent on the context, and the intent of the speaker, quite different interpretations are derived.

As this paper is work in progress, many questions are still open to discussion. Among them are the different interpretations of the aspectual postverbal secondary verbs and these verbs’ interplay with the aspectual prefixes.

Sources of the examples

Short name	Description
C3_17_MS	Conversation about the life in a village
EC_17_M	Questionnaire comparison; English
ECC_16_M	Questionnaire causative constructions; Burmese/English (Checklist causative constructions 2012)
EEN_16_M/L	Questionnaire about emphasis and nominalization, Burmese/English, (Checklist EMPH/NMLZ 2016)
EGGD_16_M/L	Questionnaire about the use of “give” and “get” and about ditransitive constructions, English
ENEG_13_M	Questionnaire about negation, Burmese/English
ENS_16_M	Questionnaire “new situation”, Burmese/English (Checklist: NEWSIT 2012)
EQB_17_M	Questionnaire about Burmese constructions, Burmese/English, (Questionnaire Burmese 2017)
ETMA_17_M	Questionnaire tense, mood, aspect; English Östen Dahl, 1985, <i>Tense and Aspect Systems</i> , Oxford, New York: Basil Blackwell, pp. 199-206
EMR_19_M	Questionnaire resumptive pronoun
EMSR_19_M	Questionnaire subordinate clauses
I_15_M	Interview of a young woman with a monk, Rumai
I_17_MS	Interview of a monk and two villagers about their life in the village
N2_17_MS	About green tea cultivation in a village
PCHT_16_L	Picture story: <i>The chicken thief</i> Béatrice Rodriguez, 2008, <i>Der Hühnerdieb</i> , Wuppertal: Peter Hammer Verlag
PFWAY_15_M	Picture story: <i>Frog, where are you?</i> Mercer Mayer, 1969 New York: Dial Books for Young Readers
WA_15_M	Sentences containing aspectual markers from: <i>Being clever with texts</i> Sa Pe 2015, Mandalay: self-publishing
WPKT_14_L	Phrase book Rumai – Shwe – Burmese – English, 2014, Lashio: Ta’ang Students and Youth Union
W1	Story in <i>Our Ta’ang magazine</i>
W2_15_M	Short story in: <i>Being clever with texts</i> Sa Pe 2015, Mandalay: self-publishing, pp. 78-79
W3_15_M	Short story in: <i>Being clever with texts</i> Sa Pe 2015, Mandalay: self-publishing, pp. 48-49
W4_15_M	Short story in: <i>Being clever with texts</i> Sa Pe 2015, Mandalay: self-publishing, pp. 38-39

Structure of the labels indicated by the examples:

1. Short name: shortcut_year_place of recording/edition
2. _consultant(s)/author_example in toolbox
3. pc_X: personal conversation with consultant X

Key to Example Label abbreviations

E	elicitation
P	picture story
I	interview
C	conversation
N	narrative
W	written text

Places

M	Mandalay
MS	Man Sat village (Namkham Township)
L	Lashio

Consultants

Nr.	Gender	Y.o.b	Origin	Education	Occupation	Languages
1	f	1996	Na Aw Gyi village, Man Ton township		bachelor student	Rumai, Burmese, English
2	m	1994			bachelor student, novice	Rumai, Burmese, English
3	m	1988	Man Sat village, Nam Hkam township	Bachelor diploma 2017	monk	Rumai, Burmese, English, Pali
5	f	1997	Nam Hkam township		student	Rumai, Shan, Burmese (English)
6	m	1993	Sar Lu village, Nam Hkam township		student	Rumai, Burmese
7	f	1959	Man Sat village, Nam Hkam township	public school in Shan, grade?	housewife, farmer	Rumai, Shan, (Burmese)
9	f	1998	Pha Daen village, Nam Hkam township	public school in Burmese, grade 10	housewife, farmer	Rumai, Burmese, (English)

References

- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In Alexandra Y. Aikhenvald and R. M. W. Dixon (eds.) *Serial Verb Constructions*:1-68.
- Aikhenvald, Alexandra Y. and R. M. W. Dixon (eds.). 2006. *Serial Verb Constructions*. Oxford: Oxford 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: Narr.
- Bisang, Walter. 2004. Grammaticalization without coevolution of form and meaning: The case of tense-aspect-modality in East and mainland Southeast Asia. In Walter Bisang, Nikolaus P. Himmelmann and Björn Wiemer (eds.) *What makes Grammaticalization? A Look from its Fringes and its Components*:109-138.
- Bisang, Walter, Nikolaus P. Himmelmann and Björn Wiemer (eds.). 2004. *What makes Grammaticalization? A Look from its Fringes and its Components*. Berlin: Walter de Gruyter GmbH & Co.
- Butt, Miriam and Tracy Holloway King (eds.). 2009. *Proceedings of the LFG09 Conference*. Stanford: CSLI Publications.
- Deepadung, Sujaritlak, Ampika Rattanapitak and Supakit Buakaw. 2015. Dara'ang Palaung. In Mathias Jenny and Paul Sidwell (eds.). *The Handbook of Austroasiatic Languages*:1065-1103.
- Ebert, Karen H. and Fernando Zúñiga (eds.). 2001. *Aktionsart and Aspectotemporality in Non-European Languages*. Zurich: University of Zurich, ASAS 16.

- Enfield, N. J. 2003. Linguistic epidemiology. Semantics and grammar of language contact in Mainland Southeast Asia. London and New York: RoutledgeCurzon.
- Enfield, N. J. 2009. 'Case relations' in Lao, A Radically Isolating Language. In Andrej Malchukov and Andrew Spencer (eds.) *The Oxford Handbook of Case*:808-819.
- Enfield, N. J. and Bernard Comrie (eds.). 2015. *Languages of Mainland Southeast Asia. The State of the Art*. Berlin, Boston: Walter de Gruyter Inc.
- Glick, Irving I. and Sao Tern Moeng. 1991. *Shan for English Speakers*. Wheaton: Dunwoody Press.
- Greenberg, Joseph H., Charles A. Ferguson and Edith A. Moravcsik (eds.) *Universals of human languages*. Stanford, CA: Stanford University Press.
- Heine, Bernd. 1993. *Auxiliaries: cognitive forces and grammaticalization*. Oxford: Oxford University Press.
- Heine, Bernd and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Jenny, Mathias. 2001. The aspect system of Thai. In Karen H. Ebert and Fernando Zúñiga (eds.) *Aktionsart and Aspectotemporality in Non-European Languages*:97-140.
- Jenny, Mathias. 2005. *The verb system of Mon*. Zurich: University of Zurich, ASAS 19.
- Jenny, Mathias. 2015. The far West of Southeast Asia – 'give' and 'get' in the languages of Myanmar. In N. J. Enfield and Bernard Comrie (eds.). *Languages of Mainland Southeast Asia. The State of the Art*: 155-208.
- Jenny, Mathias and Paul Sidwell (eds.). 2015. *The Handbook of Austroasiatic Languages*. Leiden: Brill.
- Jenny, Mathias and San San Hnin Tun. 2016. *Burmese. A Comprehensive Grammar*. Abingdon and New York: Routledge.
- Mak, Pandora. 2012. *Golden Palaung. A grammatical description*. Canberra: Asia-Pacific Linguistics. Open Access Monographs: <http://pacling.anu.edu.au/materials/SEAMLES/Mak2012GoldenPalaung.pdf>
- Malchukov, Andrej and Andrew Spencer (eds.). 2009. *The Oxford Handbook of Case*. Oxford: Oxford University Press.
- Milne, Mary Lewis Harper. 1921. *An Elementary Palaung Grammar*. Oxford: Clarendon Press.
- Müller, André and Rachel Weymuth. 2017. How society shapes language: Personal pronouns in the greater Burma zone. *Asiatische Studien* 71.1:409-432.
- Newman, John. 1996. Give. A cognitive linguistic study. Berlin and New York: Mouton de Gruyter.
- Seiss, Melanie. 2009. On the difference between auxiliaries, serial verbs and light verbs. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG09 Conference*:501-519.
- Sidwell, Paul. 2015. The Palaungic Languages. Classification, Reconstruction and Comparative Lexicon. München: LINCOM GmbH.
- Steele, Susan. 1978. The category AUX as a language universal. In Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.) *Universals of human languages*, vol 3:7-45.
- Weymuth, Rachel. 2018. Verbal affixes in Rumai, Palaung. In Hiram Ring and Felix Rau (eds.) *Papers from the 7th International Conference on Austroasiatic Linguistics*, JSEALS Special Publication No. 3, pp. 87-100. Honolulu: University of Hawai'i Press.

TERMINANCE IN BRU

Angus Wheeler
University of Sydney
<angus DOT wheeler AT sydney DOT edu DOT au>

Abstract

This paper discusses the manifestations of ‘terminance’ (following Enfield and Diffloth 2009) in a variety of Bru, an Austroasiatic (AA) language spoken in Nakai District, Khammouane province, Laos. It explores the idea that Bru is at an intermediate stage between coda phonation contrasts and contrasts of tone or register. Phonological data on terminance in word reduction and contrastive length is presented to support this idea.

Keywords: terminance, Bru, phonation, phonology

ISO 639-3 codes: bru

1 Overview

This paper presents new data from a variety of Bru, an Austroasiatic (AA) language spoken in Nakai District, Khammouane province, Laos. I discuss the hypothesis that laryngeal articulations involving voiceless and checked glottis in syllable codas should, in this language, be recognised as manifestations of a distinct phonological category which I call ‘terminance’ (following Enfield and Diffloth 2009). Viewed in the context of recurrent historical transfer of phonemic complexity from coda segments to suprasegments in Mainland Southeast Asian (MSEA) languages, the purpose here is to explore the idea that Bru is at an intermediate stage between segmental coda phonation contrasts and suprasegmental contrasts of tone or register. This idea is supported by the fact that terminance is persistent when a word is reduced and procliticised, and that contrasts in length are conspicuously absent where there are contrasts in terminance.

The speakers of this variety of Bru are in regular contact with speakers of Lao (Southwestern Tai, Enfield 2007), Saek (Northern Tai, Meng, in preparation), and Kri (Vietic, Enfield and Diffloth 2009), and Bru people also tend to be multilingual. In all of these languages, laryngeal articulations at the right edge of major syllables show special properties that interact with features such as vowel length and/or register, and which arguably favour the recognition of a distinct phonological category. In Lao, a phonetic glottal stop is used to mark an open syllable as being phonologically short. This means that the vowels in short open syllables can vary in their actual duration, sometimes to the extent that they are as long or even longer than a phonologically long vowel. In this variety of Saek, [ʔ] in word-final position does the same thing with length as in Lao, and is also one of the acoustic signals of tones 3 and 6. In Kri, open syllables and most coda consonants have more than one contrastive laryngeal setting. Enfield and Diffloth (2009) analyse laryngeal settings in Kri as being the phonetic realisations of a rime-level distinction they call ‘terminance’, which requires all major syllable rimes to be specified as having one of three possible values. Those values are ‘checked’ (full obstruction of airflow), ‘voiced’ (a normal laryngeal state), and ‘voiceless’ (spread glottis). Their main argument for this analysis is that terminance better preserves the symmetry of the system, and eliminates the need for final consonant clusters in the phonotactic template.

In languages spoken further afield throughout MSEA, word-final laryngeal articulations often show special properties. In Bru Wœn Beuk (Thongkhum 1979) and Sô (Migliazza 1998), /h/ and /ʔ/ are the only phonemes that can occur in C2 position of coda clusters. In Ta’oiq (Katuic, Gehrman 2019), glottalised codas of any kind (eg. /ʔ/, /mʔ/, or /k̚/) do not co-occur with creaky vowel phonation. Gehrman (2019) concludes that terminance types show particular patterns of distribution, even though terminance contrasts themselves are subsumed under phonemic coda consonant contrasts (due to the fact that syllables with glottalised or voiceless terminance show length distinctions). Voiceless terminance is a grouping of final /h/ and /s/; glottalised (=checked in Enfield and Diffloth’s 2009 terminology) terminance includes final /ʔ/, glottalised final sonorants, and unreleased final stops; and

voiced terminance includes open syllables, and those ending in voiced consonants. Similarly, Diffloth (1989) describes glottalised sonorant coda consonants in two East-Katuic languages Talan and Ong, and notes that they do not co-occur with creaky-voiced vowels in the same syllable.

It is no coincidence that laryngeal articulations are often singled out as the exceptions to the rule, given that they are strongly associated with suprasegmental tone and register systems of this region (cf. Brunelle and Kirby 2016). Enfield and Diffloth (2009) argue that in the case of Kri, combinations of terminance and register form a six-way distinction in syllable types, which mirrors other historical developments in AA languages, eg. development of tone/register in Vietnamese and Khmer, (Haudricourt 1954; Thurgood 2002; Huffman 1967). This paper explores the possibility that the Bru system is developing towards a similar model as Kri, where phonation-type contrasts are expressed in two discrete phases of the rime, corresponding to register and terminance, yielding six unique combinations.

This paper is structured as follows. Section 2 provides some background about the language, as well as information about the data used in this paper. Section 3 defines terminance as it occurs in this language. Section 4 explores the distributional problems regarding terminance types, and the interface those have with length contrasts. Section 5 is a summary and discussion of the implications of these data.

2 Language Background, Data, and Phonology Overview

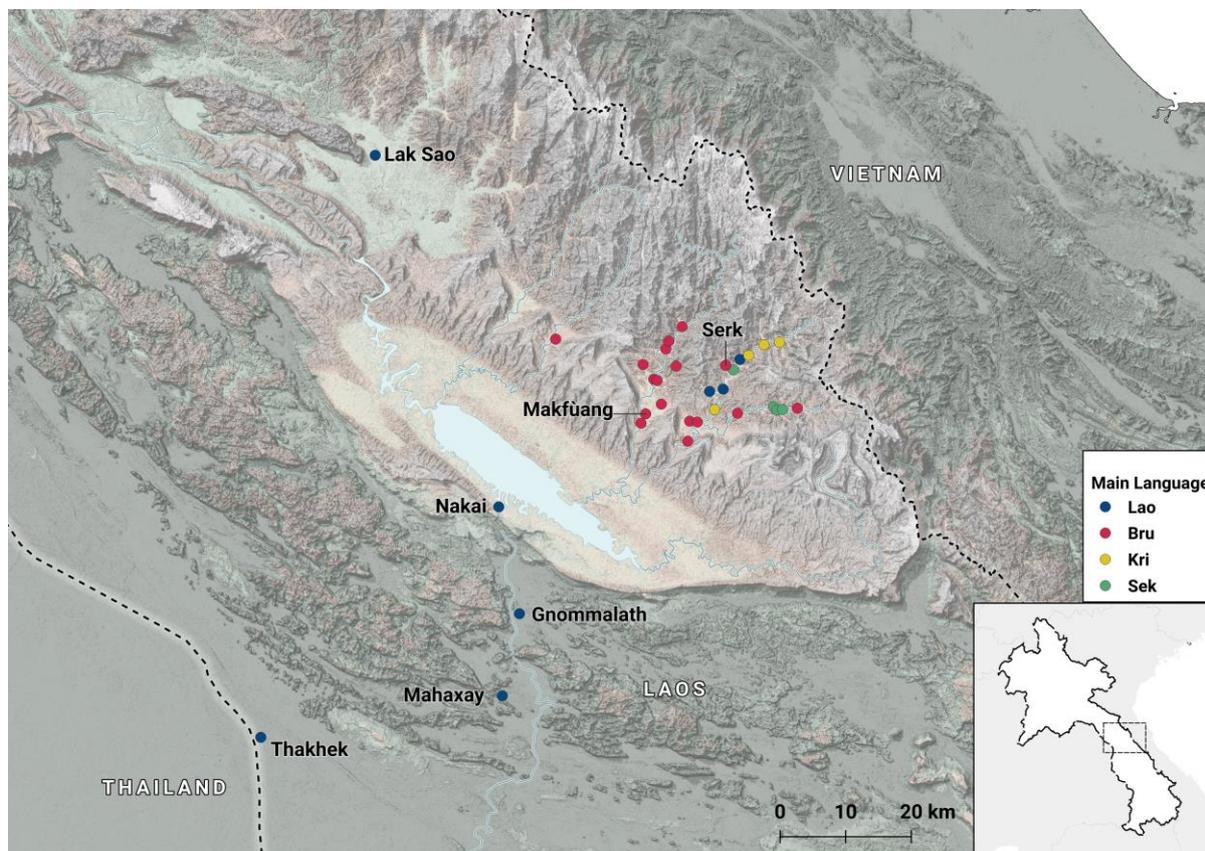
The Bru language that appears in this paper is spoken in Nakai district, the mountainous eastern part of Khammouane province of Laos. It is part of a larger West Katuic dialect chain that extends from this northern region, west into North East Thailand, South into Savannakhet and Salavan provinces, and East into Vietnam. Previous descriptions of West Katuic languages spoken in Thailand include¹ Green (1996); Jimreivat (1981); Migliazza (1998); Tebow and Lew (2013); and Thongkhum (1979). Descriptions of Vietnamese West Katuic languages include Hoàng and Thông (1998); Miller and Miller (1996); and Miller (2017b, 2017a); and Phillips, Miller, and Miller (1976). However, this is not a study of Bru languages in general, and the data used for this paper come entirely from my own fieldwork in Nakai District. Throughout this paper, I use the word ‘Bru’ in the narrow sense, referring to the communities (shown using red dots in Figure 1) located upstream of Nakai district centre.²

The data for this study comes from an ongoing project to produce a descriptive grammar of Bru, which was begun in 2017. It consists of 5+ hours of transcribed audio-visual recordings of spontaneous and elicited speech, including both monologue and dialogue, a lexicon with 2000+ entries, and documentation of 9 months of fieldwork. Two fieldwork trips have been undertaken so far, the first in 2017-18, the second in 2019.

¹ This list is not exhaustive, although West Katuic languages are relatively under-documented in print.

² This map was created using data from the Shuttle Radar Topography Mission 1 Arc-Second Global. DOI: /10.5066/F7PR7TFT. Accessed online, July 17, 2021.

Figure 1: Relief Map of Central Laos, showing the NT2 reservoir and significant rivers (light blue), villages in the NT2 watershed, and the nearby large lowland-Lao towns. Dot-colour illustrates the approximate ethno-linguistic makeup of watershed, based on provisional surveys conducted during fieldwork.



The first field site was the village of Serk on the Nam Noi River. This village is undergoing ethnolinguistic shift, and Bru is no longer the main means of talk, although many inhabitants trace lineage to Bru people. The second was the village of Makfùang along the Nam Theun. Along the upper extent of this river, all villages are predominantly L1 Bru speaking. Most people also speak Lao, and some know Kri and/or Saek.

Here is a brief overview of the sound system. Bru distinguishes minor and major syllables, where minor syllables are unstressed, have fewer vowel quality contrasts, no independent register, no coda consonants, and a reduced set of onset contrasts. However, minor syllables are phonologically independent of the major syllable (cf. voiced excrement transitions between consonants in onset clusters, Thomas 1992; Butler 2015). A phonetic [ə] might surface between /p/ and /l/ in both words of the following example, yet Bru speakers report that the words' forms are not the same.

- (1) /plɤ:/ /palɤ:/
 'head' 'play'

Bru also has three non-predictable lexically contrastive vowel qualities in minor syllables:³

- (2) /sinəh/ /sula:/ /sali:/
 'hive' 'leaf' 'corn'

³ The number of minor syllable vowel contrasts is constrained by particular onset consonants which do not allow all three, eg. /ʔ/ contrasts /a/ and /i/, but not /u/.

Major syllable vowels are shown in **Table 1**.

Table 1: *The major syllable vowel inventory*

	<i>front</i>	<i>central</i>	<i>back</i>
<i>high</i>	<i>i</i>	<i>u</i>	<i>u</i>
<i>close mid</i>	<i>e</i>	<i>ɤ</i>	<i>o</i>
<i>open mid</i>	<i>ɛ</i>	<i>ʌ</i>	<i>ɔ</i>
<i>low</i>	<i>a</i>	<i>a</i>	
<i>diphthongs</i>	<i>iə</i>	<i>uə</i>	<i>uə</i>

As will be discussed in Section 4, the type of coda affects whether vowel length is contrastive. The following example illustrates minimal or near-minimal lexical contrasts made by the duration of the vowel in the major syllable, for syllables with stop, liquid, nasal, and voiced-glide codas.

(3)	Short	Long
	[qet̚]	[qe:t̚]
	‘blue’	‘take’
	[pal]	[ʔapa:l]
	‘fly’	‘shoulder’
	[tam]	[ta:m]
	‘strike forwards’	‘go along’
	[tamaj]	[tama:j]
	‘new’	‘rattan’

The final pair [tamaj] vs [tama:j] illustrate that coda-glides are consonants, and not vowels. If they were vowels, then three vowel slots in the phonotactic template would be required for words like [tama:j] (ie. /tamaai/). Note also that length is non-contrastive for diphthongs, and length contrasts are neutralised to long in open syllables. The duration of a monophthong in an open syllable tends to be roughly equivalent to the duration of a long vowel in a closed syllable. Diphthongs pattern like long vowels in this respect, as they fulfill the length requirements of open syllables:

(4)	[ntu:]	[ntu̠a]
	‘season’	‘resemble’

All vowels have two contrastive register variants, which I refer to as ‘heavy’ and ‘light.’⁴ Heavy vowels are realised with breathy phonation; light vowels with modal phonation, and slight pre-centralisation for the vowels /i/ and /u/. Speakers have a metalinguistic trope which they use to talk about what are essentially minimal pairs, which translates to English as ‘heavy sound light sound’ (see Zuckerman and Enfield 2020), and they report that the following examples are two such pairs. This implies that they recognise that the two items in each pair have the same vowel nuclei, despite minor differences in surface quality:

⁴ I use the terms ‘heavy’ and ‘light’ as opposed to ‘lax’ and ‘tense’ or ‘low’ and ‘high’ in keeping with how Bru people talk about the phenomena.

- (5) /mi:^{HEAVY}/ /mi:^{LIGHT}/
 [mi:] [mɔi]
 ‘have’ ‘jackfruit’
- (6) /ʔacu:^{HEAVY}/ /ʔacu:^{LIGHT}/
 [ʔacu:] [ʔacəu]
 ‘bring back’ ‘knife’

Bru morphology is relatively productive, and is exclusively prefixing (aside from some non-affixing processes, eg. reduplication), in contrast to its infixing ancestor languages (cf. Gehrman 2018). The relative productivity and general significance of derivational morphology in Bru (compared to other described Western Katuic languages) should be stressed, because it means that minor syllables are often responsible for making key lexical and/or grammatical contrasts, and so presumably are unlikely to elide. A simple example is below. The prefix *ʔa-* (in bold) marks the verb as an accomplishment (an aspectual class, see Dowty 1979), significantly affecting, for example, how it serialises with other verbs.⁵

- (7) paŋ sɛ:ŋ
 dive descend
 ‘(He) dived down’
- (8) kwɛ:l **ʔa**-sɛ:ŋ
 throw ACCMP-descend
 ‘(He) threw (it) down’

3 Defining terminance

Terminance refers to the laryngeal setting of the final part of the syllable. For some syllable types, there are three phonetically distinct, lexically contrastive settings: ‘voiceless’, ‘voiced’, and ‘checked’. The first question is: how should these contrasts be represented phonologically? Consider for example, a coda that ends in a palatal glide with glottal occlusion [jʔ]. Below are the three alternative analyses that I have come across in descriptions of related languages, which explain what that sound is in phonological terms:

1. Glottalised glide phoneme /-jʔ/ eg. Bru Sakon Nakhorn (Tebow and Lew 2013), Ta’oiq (Gehrman 2019), Talan and Ong (Difflloth 1989)
2. Consonant cluster /-jʔ/ eg. Bru Wœn Beuk (Thongkhum 1979), Sô (Migliazza 1998)
3. Regular glide phoneme with suprasegmental terminance /-j^{CHECKED}/, eg. Kri (Enfield and Difflloth 2009)

Throughout the rest of this paper, I explore the possibility that phonemic coda contrasts and terminance contrasts are partially independent of one another, in this particular language variety (ie. somewhere between the second and third analyses). Evidence for this can be seen when words reduce and become proclitics,⁶ forming a phonological word-like unit with their host stems (although not entirely integrating as a single phonological and grammatical word). In Bru, cliticisation is an iambic process (cf. Keller, Jordi, and Gregerson 2011), meaning that clitics are unstressed, and their rightmost edges are butted against their host stems. The process of reduction causes deletion of coda consonants, neutralisation of length and register contrasts, and simplification of the onset. However, the laryngeal

⁵ Full description of Bru morphology will be provided in the forthcoming grammatical description.

⁶ For the forms presented here, it is unclear whether reduction takes place over the course of years, or if it is spontaneously produced as a signal of something like head-dependent relations. The term ‘proclitic’ as it is used here refers to a ‘phonologically reduced grammatical word’.

settings during the articulations of the non-reduced forms' final segments [j̥], [j], and [k̥] are retained when those forms procliticise (as per the current hypothesis, I represent laryngeal settings in superscript):

- (9) /mpj̥aj̥^{VOICELESS}/
 [mpj̥aj̥] [məh=m̩i:]
 not not=have
 'not/no' 'not have'
- (10) /mu:j̥^{VOICED}/x
 [mu:j̥] [mu=naʔ]
 one one=clf.person
 'one' 'alone'
- (11) /ma:k̥^{CHECKED}/
 [ma:k̥] [məʔ=m̩i]
 fruit CT.fruit=jackfruit
 'fruit' 'jackfruit'

The distinction between coda consonants and terminance is useful for concisely stating the rule for what is retained and what is dropped when a morpheme is procliticised. Namely, the laryngeal setting of the second phase of the rime is unaltered by reduction. An alternative account is that voiceless and checked terminance in the non-reduced forms are functions of coda segments /h/ and /ʔ/, which are excluded from reduction rules. Combined with debuccalisation (which is reported to be common in Bru languages, Gehrman 2016, page 93) of /k/ → /ʔ/ when Example 11 was borrowed from Lao, this alternative adequately predicts the forms of the proclitics.

Therefore, the persistence of terminance in proclitics does not prove that it should be abstracted away from segmental structure, but it does indicate that in this language, laryngeal contrasts can occur in contexts that other consonants can't. Given this, I would argue against the treatment of voiceless or checked glides as representing single phonemes in Bru, /j̥ʔ/, /j̥/, /wʔ/, and /w̥/ (cf. the first analysis on the previous page). Although they are co-articulated when they occur in the same coda, proclitics show that the laryngeal components are separable from the glides.

In regards to non-reduced major syllables, terminance and register are realised in separate phases throughout the articulation of the rime, and combine freely within a single syllable. Table 2 gives examples for each type, using syllables without oral coda consonants for illustration.⁷

Table 2: Six combinations of register and terminance

	VOICED	CHECKED	VOICELESS
LIGHT	[m̩i]	[t̩ʔ]	[cih]
	jackfruit	kick	be born
HEAVY	[m̩i:]	[t̩ʔ]	[cih]
	exit	arrive	ride

Ignoring differences in vowel quality and onset phonemes,⁸ it is possible to describe the contrasts in the state of the larynx as being distributed across two separate phases.⁹ Like Henderson's (1985) concept of 'feature shuffling', the representation below considers only the maximum complexity and relative

⁷ The fact that the light-checked vowel has creaky voice is not accidental, see Section 5.

⁸ A perfect minimal set showing all six types without varying onset or vowel segments was not found, but the point can still be illustrated if we set these things aside.

⁹ The idea of 'phases' is borrowed from Gehrman (2019), a description of Ta'Oiq rimes.

4 Distribution of terminance types and interface with length

This section discusses the distributions of terminance types, and what they mean for the rime’s possible contrasts in length (and therefore, the syllable’s).¹⁰ First, not all syllable types have contrasts in terminance. The simplest way to discuss syllable types is in terms of their final segment, which can be a coda consonant (for closed syllables) or a vowel (for open syllables), and for these purposes does not include [ʔ] or [h]. In Bru, syllables whose final segment is a vowel or glide make a three-way contrast in terminance. All other types of syllable have just one way that they can terminate, which is therefore not lexically contrastive. **Table 3** shows the distribution of terminance contrasts per final segment type:

Table 3: *Terminance types in Bru across syllables with different final segments.*

	Stops	Nasals + Liquids	Glides + Vowels
CHECKED	✓	✗	✓
VOICED	✗	✓	✓
VOICELESS	✗	✗	✓

The Bru system is demonstrably less elaborate than the Kri system, which has a three-way terminance contrast for all oral sonorants (including vowels), and a two way contrast in nasals, shown in **Table 4** below:

Table 4: *Terminance types in Kri across syllables with different final segments. Adapted from Enfield & Diffloth (2009).*

	Stops	Nasals	Oral sonorants
CHECKED	✓	✓	✓
VOICED	✗	✓	✓
VOICELESS	✗	✗	✓

Although less elaborate, the Bru terminance system has an additional property, which is that the subset of syllables that have a contrast in terminance do not also contrast length for their checked and voiceless variants. This issue has been noted for other Bru varieties, eg. Sakon Nakhorn (Tebow and Lew 2013), and for AA languages elsewhere, eg. Eastern Kmhmu’ (Osborne 2018). In regard to Bru, a key point is that contrasts in checked or voiced terminance do not just neutralise length contrasts to a particular value (long or short); they eliminate contrasts based on vowel duration entirely. Therefore, while open syllables with modal terminance must realise their vowel with a longer duration (given that open syllables neutralise to long); any syllable with lexically contrastive checked or voiceless terminance can actually vary the duration of its vowel.

One of the clearest illustrations is /mpjaj^{VOICELESS}/ ‘not’, which as previously mentioned can be reduced to something close to an unstressed minor syllable, with a phonetically short vowel. However, it may also be used as a standalone utterance (expressing disagreement with what was just said), and speakers sometimes extend the duration of the vowel beyond that of even a phonologically long vowel (I represent this below as a doubled ‘:.’).

¹⁰ I use the word ‘syllable’ here as it feels more natural to talk about their being open and closed compared to ‘rimes’. Because syllables have only one rime, there is no ambiguity. However, note that the idea of ‘phases’ applies strictly to rimes, as there is no evidence of syllable-onset complexity affecting length or terminance contrasts.

- (15) [məh=mi:] [mpja:ːj]
 not=have not
 ‘(I) don’t have (it)’ ‘Naaaaaaah’

11

I see no reason to dismiss this fact as a result of ‘performance’, given that the option of extending a word like this is simply not available when the major syllable’s vowel duration is contrastive (and especially when it is short). For example, if a speaker wanted to place emphasis on the word /tamaj/ ‘new’, which has a short vowel and voiced terminance, they could not extend the duration for that purpose, and would have to rely on other cues eg. pitch and amplitude.

One way to explain why length contrasts are conflated in syllables with no oral coda consonants is to say that laryngeal settings are completely transparent in terms of syllable-closure; eg. [cɔʔ] ‘put’ is phonologically-speaking an open syllable with a long vowel, /cɔ:^{checked}/. Because open syllables do not contrast in length, the vowel’s duration is irrelevant for lexical distinctions, making the following forms a minimal set.¹²

- (16) /pɛ:^{checked}/ /pɛ:^{voiced}/ /pɛ:^{voiceless}/
 [pɛʔ] [pɛ:] [pɛh]
 ‘slaughter’ ‘rafters’ ‘stop’

However, this does not account for syllables with glides in their codas. With voiced terminance, rimes with glide-codas contrast long and short monophthongs, and diphthongs:

- (17) /paj^{voiced}/ /pa:j^{voiced}/ /t^hruaj^{voiced}/
 ‘three’ ‘say’ ‘fall’

With voiceless or checked terminance, glide-final syllables do not show duration-based contrasts, although they can take both monophthongs and diphthongs as their nuclei (which indicates that two vowel slots are available). The previous explanation does not work here, because the major syllables in the examples below would be closed even with the caveat of laryngeal transparency.

- (18) /quj^{voiceless}/ /qajuaɹ^{voiceless}/
 ‘fire’ ‘husband’

- (19) /ntruɹ^{checked}/ /cuaj^{checked}/
 ‘chicken’ ‘search’

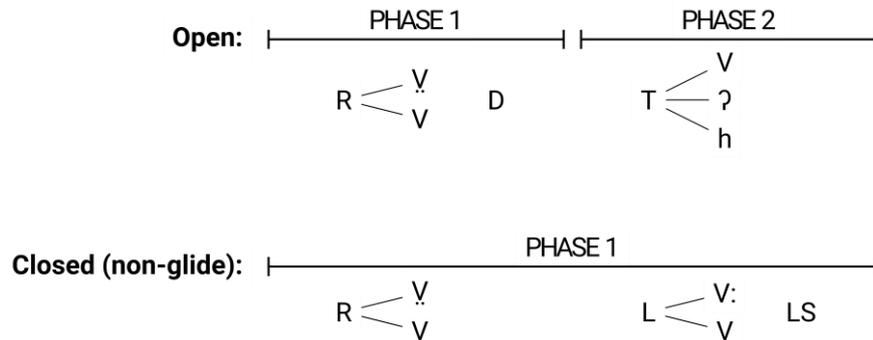
Let me address another explanation that plainly does not work. If we treat /ʔ/ and /h/ as if they were vowels in this language, then words like /pɛʔ/ are equivalent to CVV, and therefore satisfy the minimum weight requirements of a major syllable. Aside from the fact that it would mean words like /siaʔ/ require CVVV (which is not otherwise attested), this analysis also implies that words like /pɔjh/ are actually CVCV (with stress on the first syllable), which is a violation of the most basic principles of the Bru phonological word.

Instead, I suggest that the system groups terminance and length as manifestations of the same phonological category. This idea requires us to strain the definition of a phase somewhat, but I think doing so is motivated by the data. First, Table 3 compares the rime-internal contrasts that originate in the larynx for open and closed syllables, excluding glide-final syllables. Each representation includes one piece of redundant phonetic information: open syllables have some duration (‘D’), which is not lexically contrastive; the terminus of closed syllables has some obligatory laryngeal setting (‘LS’), checked for stops and voiced for nasals and liquids.

¹¹ This translation is intended to capture a piece of casual conversation, and would sound quite normal in my dialect of Australian English.

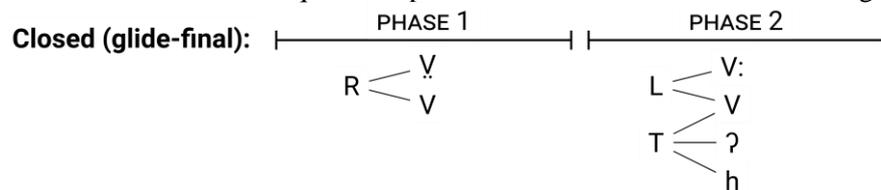
¹² Vowels in checked or voiceless rimes tend to be somewhat shorter (cf. Osborne, 2018), but again, their duration is not lexically contrastive.

Figure 3: Comparison of rime phases in open and closed (stop, nasal, or liquid-final) syllables. 'R' = register, 'T' = terminance, 'LS' = laryngeal setting, 'L' = length, 'D' = duration.



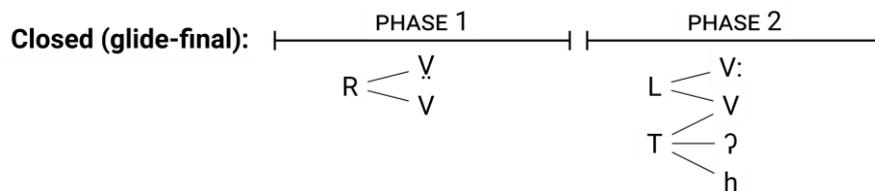
What we are left with is two lexically contrastive features for each syllable type. Both types have register, as well as a secondary contrast: one is expressed via laryngeal settings; the other by the duration of voicing, irrespective of setting. Because duration measures the entire voiced component of the rime, it is not a truly separate phase. However, if we allow the idea of the second phase to represent any secondary contrast regardless of temporal distribution, terminance and length are near-equivalents of one another (this equivalence is not perfect, eg. there are three options in open syllables vs two in closed ones, but we do at least see that both kinds of syllable have more than one contrastive type).

I do not mean to imply that terminance has taken up the functional load of an erstwhile length contrast in open syllables. If anything, the story may be the other way around: it could be that the pre-existence of terminance contrasts inhibited the development of new length contrasts in open syllables, owing to the fact that the number of laryngeal contrasts in the rime was already at maximum. Another way to say this is that grouping length and terminance has the consequence of making them compete for the same structural valence. This becomes quite complicated when we consider rimes with glides in



their codas. Table 4 shows the rimes of glide-final syllables (using the flexible definition of a phase mentioned above):

Figure 4: Rime phases in glide-final syllables.



Although both length and terminance are lexically contrastive in the same syllable type, there are no two rimes that contrast with one another in both length and terminance. For example, the rimes of the following pair contrast in terminance alone (their duration could vary without affecting the comparison):

- (20) [nʌj] [kʌj?]
 'tall' 'small'

And the following rimes contrast in length alone. If one had voiceless or checked terminance, duration would no longer matter:

be realised with, and which is essentially predictable from iambic stress patterns.¹⁴ Added to this equation is the fact that minor syllables are not grammatical words (cf. Dixon, 2010), while proclitics and open syllables are, and so are required to fulfill a quantity constraint (ie. they must have at least two mora). It would seem that having a three-way contrast in terminance is a reasonable approximation of a quantity constraint. It offers a formal contrast between minor syllables and proclitics which allows proclitics to be interpreted as grammatical words, even though they are unstressed, short, and are not allowed to have coda consonants.

Table 5: Comparison of redundant features in the rimes of open syllable types.

<i>Minor Syllable</i>	<i>Proclitic</i>	<i>Open Syllable</i>
<i>voiced-short</i>	<i>voiced-short</i>	<i>voiced-long</i>
	<i>checked-short</i>	<i>checked-short/long</i>
	<i>voiceless-short</i>	<i>voiceless-short/long</i>

2. *The interface between register and terminance*

The hypothesis presented here is obviously based on conventional phonological analysis; the forms that speakers report to be lexically contrastive; and also the forms that they produce and judge to be acceptable in natural conversation as representations of a lexical category. However, in a previous footnote I alluded to an observed tendency for the vowels of words containing some like-with-like combinations of register and terminance to be realised further towards the edge of the phonation type continuum (cf. Gordon and Ladefoged 2001), especially those with light register with checked terminance. Also, speakers would sometimes elaborate on the sounds during consultation, describing them as eg. ‘very light’, in reference to their ‘heavy-light’ cultural trope. In other words, light register vowels often seem creaky and higher pitched when they occur before checked terminance, in anticipation of the full glottal closure that is to come. Their heavy register counterparts may have shown a similar pattern when combined with voiceless terminance, but it was less obvious to me. For example:

- (22) /ci:^{light,voiceless}/ /ci:^{light,checked}/
 [cih] [ciʔ]
 ‘birth’ ‘throw away’
- (23) /mpre^{heavy,voiceless}/ /mpe^{heavy,checked}/
 [mpreḥ] [mpeʔ]
 ‘fight’ ‘mother’

To be clear, the above are a sample of observed cases, which form the basis of a hypothesis that is yet to be tested by way of acoustic analysis. The hypothesis predicts that certain two-phase patterns may realise new phonation types in the first phase. However, even if the above observations were upheld by a detailed acoustic study, I would not claim that the two-phase system has or will combine into a single suprasegmental feature, with a larger number of contrastive phonation types. That may be one possible direction for change, but in its current state, there are two clearly distinct phases of laryngeal setting in the rime, meaning that register and terminance themselves are distinct.

Gehrmann (2022) argues that in many AA languages, the development of suprasegmental contrasts from erstwhile segmental contrasts in coda phonation may have occurred after those that came from contrasts in onset phonation. This case would appear to support his claim, and offer an example of what happens a short way in to the process of reanalysis of coda phonation. In the social context where Bru is spoken, four languages use their own calques of the phrase ‘heavy sound light sound’ to talk about lexical contrasts, and speakers primarily anchor their judgments to whichever tone and register systems

¹⁴ There is a grey area in voiceless and checked open syllables, but otherwise unstressed syllables are short, and stressed syllables are long.

their languages have. However, in my experience the contrasts do not always fall neatly into two categories: speakers can disagree on categorisation; or use non-canonical terminology, eg. ‘very light’ (as mentioned above for light-checked words), or ‘normal sounding’ (it is unclear what prompted this description, on the occasions that I heard it used). I think this illustrates one reason why there may not be a clear distinction between segmental and suprasegmental features; that is not always the reality that speakers of a language experience.

References

- Brunelle, Marc, and James Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistic Compass* 10.4:191–207.
- Butler, Becky. 2015. Approaching a Phonological Understanding of the Sesquisyllable with Phonetic Evidence from Khmer and Bunong. In *Mainland Southeast Asian Languages: State of the Art and New Directions*, edited by N.J. Enfield and Bernard Comrie, 437–93. Berlin: De Gruyter.
- Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies* 15:139–54.
- Dixon, R. M. 2010. Basic linguistic theory volume 2. Oxford University Press.
- Dowty, David. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Enfield, N. J. 2007. *A Grammar of Lao*. Berlin: Mouton de Gruyter.
- Enfield, N. J., and G. Diffloth. 2009. Phonology and Sketch Grammar of Kri, a Vietic Language of Laos. *Cahiers de Linguistique - Asie Orientale* 38.1:3–69.
- Gehrmann, Ryan. 2016. The West Katuic Languages: Comparative Phonology and Diagnostic Tools. Master’s thesis, Chiang Mai: Payap University.
- . 2018. Katuic Presyllables and Derivational Morphology in Diachronic Perspective. In *JSEALS Special Publication No. 3, Papers from the Seventh International Conference on Austroasiatic Linguistics*, edited by Hiram Ring and Felix Rau, 132–56. University of Hawai’i Press.
- . 2019. On the Origin of Rime Laryngealization in Ta’oiq: A Case Study in the Phonatory Restructuring of Vowel Quality Contrasts. Paper presented at the International Conference on Austroasiatic Linguistics 8, Chiang Mai University, Chiang Mai, Thailand, 2019.
- . 2022. Desegmentalization: Towards a Common Framework for the Modeling of Tonogenesis and Registrogenesis in Mainland Southeast Asia with Case Studies from Austroasiatic. PhD thesis, University of Edinburgh.
- Gordon, Matthew, and Peter Ladefoged. 2001. Phonation Types: A Cross-Linguistic Overview. *Journal of Phonetics* 29:383–406.
- Green, Julie. 1996. A Preliminary Description of Bru (Khong Chiam) Phonology. PhD thesis, Thammasat University; the Summer Institute of Linguistics.
- Haudricourt, A. G. 1954. De ;’origine des tons en Vietnamien. *Journal Asiatique* 242:69–82.
- Henderson, E. J. A. 1985. Feature Shuffling in Southeast Asian Languages. In *Southeast Asian Linguistic Studies Presented to André G. Haudricourt*, edited by S. Ratanakul, 1–22. Institute of Language; Culture for Rural Development, Mahidol University.
- Hoàng, Văn Ma, and Tạ Văn Thông. 1998. *Tiếng Bru-Vân Kiều (The Bru-Van Kieu language)*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Huffman, Franklin E. 1967. An Outline of Cambodian Grammar. *ProQuest Dissertations and Theses*. PhD thesis, Cornell University.
- Jimreivat, Pattiya. 1981. Clauses and Phrases in Bruu. Master’s thesis, Mahidol University.
- Keller, Charles, Jacqueline Jordi, and Kenneth Gregerson. 2011. Merge Right: Iambic Effects, Grammaticalization and Cliticization in Brao Grammar. *Mon-Khmer Studies*, Mon-khmer studies journal special issue no. 3: Austroasiatic studies: Papers from icaal 4, 2:162–77.
- Meng, Weijian. In Preparation. A Grammar of Saek. PhD thesis, University of Sydney.

- Migliazza, Brian. 1998. A Grammar of so: A Mon-Khmer Language of Northeast Thailand. Master's thesis, Mahidol University.
- Miller, Carolyn. 2017a. Eastern Bru Grammar Sketch. <https://www.sil.org/language/bru>.
- . 2017b. Eastern Bru Orthography Statement. <https://www.sil.org/language/bru>.
- Miller, John, and Carolyn Miller. 1996. Lexical Comparison of Katuic Mon-Khmer Languages with Special Focus on so-Bru Groups in Northeast Thailand. *Mon-Khmer Studies* 26:255–90.
- Osborne, R. Anne. 2018. A Phonology of Eastern Kmhmu' with Special Reference to Palatal Continuant Codas and Neutralisation of Vowel Length Contrast. *JSEALS* 11.2:67–86.
- Phillips, Richard L., John Miller, and Carolyn Miller. 1976. The Brū Vowel System: Alternative Analyses. *Mon-Khmer Studies* 5:203–17.
- Tebow II, Charles, and Sigrid Lew. 2013. A Phonological Description of Western Bru, Sakon Nakhorn Variety, Thailand. *Mon-Khmer Studies* 42:127–39.
- Thomas, David. 1992. On Sesquisyllabic Structure. *Mon-Khmer Studies* 21:206–10.
- Thongkhum, Threaphan L. 1979. The Distribution of the Sounds of Bruu. *Mon-Khmer Studies* 8:221–93.
- Thurgood, G. 2002. Vietnamese and Tonogenesis: Revisiting the Model and the Analysis. *Diachronica* 19:333–63.
- Zuckerman, Charles H. P., and N. J. Enfield. 2020. Heavy Sound Light Sound: A Nam Noi Metalinguistic Trope. In *JSEALS Special Publication No. 6: Studies in the Anthropology of Language in Mainland Southeast Asia*, edited by N. J. Enfield, Jack Sidnell, and Charles H. P. Zuckerman, 85–92. University of Hawai'i Press.

