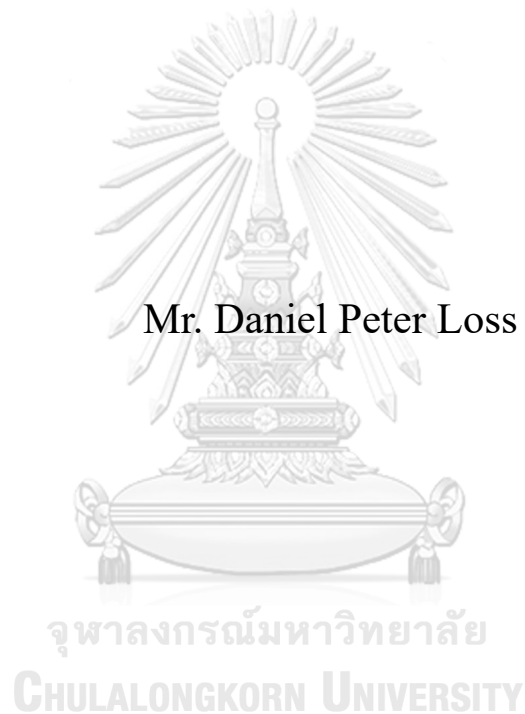


INFORMATION STRUCTURE IN MOKLEN



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Linguistics
Department of Linguistics
Faculty Of Arts
Chulalongkorn University
Academic Year 2023

โครงสร้างสาระในภาษามอแกน



นายแคเนียล ปีเตอร์ ลอส

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2566

Thesis Title	INFORMATION STRUCTURE IN MOKLEN
By	Mr. Daniel Peter Loss
Field of Study	Linguistics
Thesis Advisor	Associate Professor Dr. PITTAYAWAT PITTAYAPORN
Thesis Co Advisor	Assistant Professor Dr. NATTANUN CHANCHAOCHAI

Accepted by the FACULTY OF ARTS, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

..... Dean of the FACULTY OF ARTS
(Associate Professor Dr. SURADECH
CHOTIUDOMPANT)

DISSERTATION COMMITTEE

..... Chairman
(Assistant Professor Dr. Pattama Patpong)

..... Thesis Advisor
(Associate Professor Dr. PITTAYAWAT
PITTAYAPORN)

..... Thesis Co-Advisor
(Assistant Professor Dr. NATTANUN
CHANCHAOCHAI)

..... Examiner
(Assistant Professor Dr. SUJINAT JITWIRIYANONT)

..... Examiner
(Assistant Professor Dr. THEERAPORN
RATITAMKUL)

..... Examiner
(Assistant Professor Dr. PAVADEE SAISUWAN)

CHULALONGKORN UNIVERSITY

แดเนียล ปีเตอร์ ลอส : โครงสร้างสาระในภาษามอแกน. (INFORMATION STRUCTURE IN MOKLEN) อ.ที่ปรึกษาหลัก : พิทยาวัฒน์ พิทยากรณ์, อ.ที่ปรึกษาร่วม : ญัฐนันท์ จันทร์เจ้าฉาย

งานวิจัยนี้ศึกษาลักษณะของโครงสร้างสาระ (information structure) ในภาษามอแกน ซึ่งเป็นภาษาตระกูลออสโตรนีเซียนที่ใช้อยู่ในบริเวณภาคใต้ของประเทศไทย วัตถุประสงค์คือเพื่อศึกษาคุณสมบัติด้านสาระและวากยสัมพันธ์ของหน่วยทำนองเสียง (intonation unit; IU) และเพื่อศึกษาความสัมพันธ์ระหว่างสถานะของสาระและการเปลี่ยนแปลงรูปคำของหน่วยศัพท์ ในส่วนของวัตถุประสงค์ข้อแรก ประกอบด้วยสมมติฐาน 2 ข้อ คือ สมมติฐานข้อที่ 1 คือ หน่วยทำนองเสียงระดับอนุพากย์ (clausal information units) จะสอดคล้องกับข้อกำหนดหนึ่งข้อกำหนดต่อความคิดใหม่หนึ่งความคิด (one-new-idea constraint) สมมติฐานข้อที่ 2 คือการแปรของโครงสร้างอาร์กิวเมนต์จะสามารถอธิบายได้ด้วยหลักการสาระเดิมปรากฏก่อนสาระใหม่ (given-before-new principle) สำหรับวัตถุประสงค์ข้อสอง ด้านการเปลี่ยนแปลงรูปคำของหน่วยศัพท์ เกี่ยวข้องกับสมมติฐานข้อที่ 3 คือการใช้รูปแปรพยางค์เดียวจะสอดคล้องกับสถานะสาระเดิม ผู้วิจัยดำเนินการศึกษาเพื่อพิสูจน์สมมติฐานทั้งสามข้อ ด้วยการออกแบบภาพสิ่งเร้าและใช้เก็บข้อมูลภาษาจากผู้ออกภาษามอแกนจำนวนรวม 24 คนจากชุมชนชาวมอแกน 13 ชุมชน ครอบคลุมบริเวณจังหวัดพังงาและภูเก็ต ผลการศึกษาส่วนแรกมาจากภารกิจที่ให้ผู้บอกภาษาเล่าเรื่องจากหนังสือภาพปลาที่ถูกขโมย (Stolen Fish picture book) พบว่าหน่วยทำนองเสียงระดับอนุพากย์ของผู้บอกภาษาสอดคล้อง ข้อกำหนดหนึ่งข้อกำหนดต่อความคิดใหม่หนึ่งความคิด ผลการศึกษาส่วนที่สองซึ่งเป็นการกิจให้ผู้บอกภาษาเล่าเรื่องจากภาพลำดับเหตุการณ์กระทำก่อผล (Transitive Event Picture Sequences) พบว่าการเรียงลำดับคำแบบ AVO (ผู้กระทำ-กริยา-กรรม) เป็นโครงสร้างอาร์กิวเมนต์ที่ใช้มากที่สุดใอนุพากย์ที่มีกริยากรรมเป็นแก่น กล่าวคือการแปรของโครงสร้างอาร์กิวเมนต์พบได้น้อย ไม่สอดคล้องกับหลักการสาระเดิมปรากฏก่อนสาระใหม่ นอกจากนี้ เมื่อพิจารณาการเปลี่ยนแปลงรูปคำของหน่วยศัพท์ ผู้วิจัยพบว่าการเปลี่ยนแปลงรูปคำดังกล่าวไม่ได้มีเหตุปัจจัยมาจากเรื่องความเป็นสาระเดิมเท่านั้น หากแต่ยังมีความสัมพันธ์สอดคล้องไปกับการสลับสาระไปเป็นหัวเรื่อง (informational shift towards topics) จากข้อค้นพบนี้ ผู้วิจัยเสนอว่าภาษามอแกนมีแนวโน้มที่จะละหรือตัดพยางค์รอง (minor-syllable elision) ส่วนหนึ่งเนื่องมาจากการลดความเด่นของสาระ ซึ่งถูกกำหนดโดยตัวปริงเอง โดยภาพรวม ผลการวิจัยนี้แสดงให้เห็นลักษณะด้านหน่วยคำ-วากยสัมพันธ์ของภาษามอแกน และแสดงให้เห็นภาพปรากฏการณ์ของโครงสร้างสาระจากภาษามอแกนซึ่งเป็นภาษาที่มีผู้ศึกษาวิจัยยังไม่มากนัก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา ภาษาศาสตร์
ปีการศึกษา 2566

ลายมือชื่อนิติศ
ลายมือชื่อ อ.ที่ปรึกษาหลัก
ลายมือชื่อ อ.ที่ปรึกษาร่วม

6281006822 : MAJOR LINGUISTICS

KEYWORD Moklen, Moklenic, information structure, intonation units, argument structure, information status, word-form, elision

D: Daniel Peter Loss : INFORMATION STRUCTURE IN MOKLEN.
 Advisor: Assoc. Prof. Dr. PITTAYAWAT PITTAYAPORN Co-advisor:
 Asst. Prof. Dr. NATTANUN CHANCHAICHAI

This study investigates aspects of information structure in Moklen, an endangered Austronesian language of Southern Thailand. The two objectives were to study the syntactic and informational properties of intonation units and to study the relationship between information status and changes in word-form. With respect to the first objective, Hypothesis 1 held that clausal intonation units would conform to the one-new-idea constraint, while Hypothesis 2 held that variations in argument structure could be accounted for by the given-before-new principle. Regarding changes in word-form, Hypothesis 3 held that use of monosyllabic alternants would correspond to “given” information statuses. To assess these hypotheses, custom-designed picture-based stimuli were used in language elicitation tasks with 24 different Moklen speakers from 13 different Moklen communities across Phang Nga and Phuket provinces. Findings from the Stolen Fish picture book narration task showed that speakers’ intonation units conformed to the one-new-idea constraint. Results from the Transitive Event Picture Sequences task demonstrated that AVO (Agent - Verb - Object) was the preferred argument structure of transitive clauses, meaning that despite givenness conditions variation in argument structure is rare. Findings on changes in word-form showed that word-form alternations were not motivated merely by factors of givenness but instead corresponded to a broader informational shift towards topics. This suggests that minor-syllable elision is motivated at least in part by discourse-conditioned reductions in prominence. Overall, findings from the study reveal features of Moklen morpho-syntax and provide a picture of information structure phenomena in a lesser-described language.

จุฬาลงกรณ์มหาวิทยาลัย
 CHULALONGKORN UNIVERSITY

Field of Study: Linguistics

Student's Signature

Academic Year: 2023

.....
 Advisor's Signature

Year:

.....
 Co-advisor's Signature

.....

ACKNOWLEDGEMENTS

This research project was supported by the Second Century Fund (C2F), Chulalongkorn University. I am grateful to have had the luxury and privilege of its support and to be part of Center of Excellence in Southeast Asian Linguistics, Faculty of Arts, Chulalongkorn University. Throughout the duration of the project, I was fortunate enough to have two advisors: Pittayawat Pittayaporn and Nattanun Chanchaochai. From inception to completion, their input and guidance greatly impacted this thesis. This research is part of the project “Research and Documentation of the Moklen Language and Culture in the Southeast Asian Context” supported by the Institute of Suvarnabhumi Studies, Thailand Academy of Sciences, Humanities, and Arts, and the Ministry of Higher Education, Science, Research, and Innovation, headed by Associate Professor Dr. Pittayawat Pittayaporn. Many researchers are connected to this effort; however, I would especially like to thank Phongbhorn Prayongdravya for accompanying me on my first trip to Phang Nga and Athikhom Saengchai for photographing a collaborative session of the study.

Moklenic scholarship from Michael Larish was an invaluable resource. Furthermore, upon learning of my project, “Osh” shared field recordings and continued to check in on my progress all along the way. In surveying Moklenic linguistic research, I was led to Peter Jenks, Naw Say Bay, and Lyn Jaw Zaw, each of whom provided Moken data. Their contribution was formative in situating my ideas within a broader Moklenic context. Many of my fellow students at Chulalongkorn University deserve special recognition. For all their help, I would like to thank Kongweha Intaranuch, Naruemol Jirapanakorn, Ashley Laughlin, Lena Maluleem, Apinya Singsopa, and Santhawat Thanyawong. Also, Tyler Davis and Jacob Watson, friends and fellow Payap University Linguistics alumni, contributed feedback to individual chapters. Deserving a very special acknowledgement is Paul Hoch Myers. Paul generously donated his artistic talents towards the creation of the Stolen Fish picture book, proving, yet again, his faithful friendship and status as creative soulmate. Similarly, Nittaya Intapong must also be thanked for helping to compose the ever-evolving Transitive Event Picture Sequences. These artistic contributions served as cornerstones for conversations with Moklen speakers and, therefore, were a real foundation for the whole project. Through

ups and downs, I've always had support and encouragement from families in Pasang and Pensacola. Rita, เป็ล, all my love. Finally, I must express my sincere gratitude and appreciation to all the Moklen communities, individuals, and contributors for sharing their time, care, and interest—gifts for which I'm truly grateful.

Daniel Peter Loss



TABLE OF CONTENTS

	Page
.....	iii
ABSTRACT (THAI)	iii
.....	iv
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
1 Introduction.....	1
1.1. Background.....	2
1.1.1 Moklen.....	2
1.1.2 Moklen’s variant linguistic forms.....	3
1.1.3 Information structure.....	5
1.2 Research questions, objectives, hypotheses.....	9
1.2.1 Research questions	9
1.2.2 Objectives.....	9
1.2.3 Hypotheses	9
1.3 Significance	9
1.4 Definitions	9
1.5 Abbreviations.....	11
1.6 Thesis overview	12
2 The Moklen people and language	14
2.1 The Moklen people	14
2.2 Previous scholarship and path to recognition	18
2.3 The place of the Moklenic languages	24

2.4 Moklen language	27
2.4.1 Phonology.....	28
2.4.2 Basic syntax.....	31
2.5 Summary.....	34
3 Information structure	35
3.1 Introduction.....	35
3.2 Background and interfaces of information structure	36
3.3 Information status	38
3.3.1 Models of givenness.....	38
3.3.2 RefLex Scheme	42
3.3.3 Prominence.....	44
3.4 Argument structure	47
3.4.1 Word order vs. argument structure	47
3.4.2 Preferred argument structure	48
3.4.3 Given-before-new principle	50
3.5 Intonation units	53
3.5.1 Basic units of speech	54
3.5.2 Syntactic properties	60
3.5.3 One-new-idea constraint.....	62
3.6 Topic and focus.....	65
3.7 Summary.....	67
4 Methodology	69
4.1 Introduction.....	69
4.2 Data collection	70
4.2.1 Participants	70
4.2.2 Session recording	71
4.3 Stimulus 1: Stolen Fish Picture Book	72
4.3.1 Design.....	72
4.3.2 Procedure.....	76

4.3.3 Data	77
4.3.4 Annotation	78
4.3.5 Analysis	79
4.4 Stimulus 2: Transitive Event Picture Sequences	79
4.4.1 Design.....	79
4.4.2 Procedure.....	90
4.4.3 Data	92
4.4.4 Annotation	93
4.4.5 Analysis	96
4.5 Data presentation	97
4.6 Summary.....	98
5 Informational properties of intonation units.	99
5.1 Introduction.....	99
5.2. The one-new-idea constraint.....	99
5.2.1 New referring expressions in clausal intonation units.....	102
5.2.2 New referring expressions in phrasal intonation units	106
5.2.3 Violating the one-new-idea constraint.....	108
5.3 Discussion.....	112
5.4 Summary.....	117
6 Syntactic properties of intonation units	118
6.1 Introduction.....	118
6.2 Argument structure	118
6.2.1.Variation in argument structure	127
6.2.2 Co-referring expressions	132
6.2.3 The given-before-new principle	139
6.3 Discussion.....	144
6.4 Summary.....	147
7 Monosyllabic alternants	148
7.1 Introduction.....	148

7.2 Word-form shifts	149
7.2.1 Nouns.....	151
7.2.2 Verbs	160
7.2.3 Closed-class items	166
7.3 Topical shift	169
7.4 Discussion.....	174
7.5 Summary	181
8 Discussion & Conclusion.....	182
8.1 Introduction.....	182
8.2 Informational properties	182
8.3 Syntactic properties	183
8.4 Moklen word-form.....	185
8.4.1 Minor-syllable elision and information structure.....	186
8.4.2 Diachronic implications	187
8.5 Limitations.....	189
8.6 Conclusion	191
Appendix A Stolen Fish Text	192
Appendix B TEPS Monosyllabic alternants and word-form shifts	210
Appendix C Stolen Fish Illustrations	226
Appendix D Transitive Event Picture Sequences	244
REFERENCES	269
VITA	297

LIST OF TABLES

	Page
Table 1 Initial Consonants in Moklen.....	28
Table 2 Final consonants in Moklen	29
Table 3 Variation of initial consonants.....	29
Table 4 Moklen vowels.....	30
Table 5 Annotation tags of the Reflex Scheme (Riester & Baumann 2017).....	44
Table 6 Preferred Argument Structure constraints (Du Bois 2003).....	49
Table 7 Mithun’s intonation unit cues (based on Chafe 1994)	55
Table 8 Principal intonation unit cues (Du Bois, 2008).....	57
Table 9 The syntax-lexicon continuum (Croft, 2005).....	60
Table 10 Titles and descriptions of Stolen Fish illustrations	75
Table 11 Stolen Fish sessions selected for analysis	77
Table 12 Stolen Fish corpus intonation unit types	79
Table 13 Breakdown of referent types in the 24 TEP sequences.....	83
Table 14 Referents and events of Transitive Event Picture Sequences	85
Table 15 TEPS Target Scenes Versions A1 and B2	87
Table 16 TEPS Target Scenes Versions A2 and B2	88
Table 17 TEPS Sessions selected for analysis	92
Table 18 Word Classes and levels of information status analysis.....	95
Table 19 Given and new in the RefLex Scheme.....	96
Table 20 Clausal IUs’ given and new R-EXPs	102
Table 21 Clausal IUs with new R-EXPs.....	102
Table 22 Argument type: Clausal IUs with only a single new R-EXP	106
Table 23 Valid responses to TEPS target scenes	119
Table 24 Clausal IUs with new R-EXPs	140
Table 25 Argument type: Clausal IUs with only a single new R-EXP.....	142
Table 26 TEPS Monosyllabic alternants by lexical class.	149

Table 27 Information Status of monosyllabic alternants in TEPS	150
Table 28 Information status of nominal monosyllabic alternants	152
Table 29 Verbal monosyllabic alternants	160
Table 30 Closed class monosyllabic alternants.....	166



LIST OF FIGURES

	Page
Figure 1 Contrasts between disyllabic and monosyllabic form of /ma4nut4/ in (1.1)..4	
Figure 2 Moklen Communities (Adapted from CUSRI, 2016)	14
Figure 3 Spirit poles at Ban Khanim, Lampi, and Bangsak.....	16
Figure 4 Houses: traditional style (left), post-tsunami reconstruction style (right)....	17
Figure 5 Moklen Cultural Centers at Tap Tawan, Tung Waa, and Thap Pla	18
Figure 6 Larish's (1999, p.364) PMACM Subgroup Tree Diagram.....	25
Figure 7 Smith's (2017) Malayo-Polynesian higher-order subgroupings	26
Figure 8 Layers of analysis.....	37
Figure 9 Alignment of activation, information status, and linguistic form.....	39
Figure 10 Prince's (1981) Taxonomy of Assumed Familiarity.....	40
Figure 11 Lambrecht's (1994) information status categories	40
Figure 12 Ariel's (1990) Accessibility Marking Scale.....	41
Figure 13 Gundel et al.'s (1993) Givenness Hierarchy.....	42
Figure 14 Conceptualization of activation and prominence alignment	47
Figure 15 Intonation unit categories and constraints.	53
Figure 16 Mithun's (2021) Mohawk IU segmentation.	56
Figure 17 Stolen Fish and TEPS sessions (Photographs by Athikhom Saengchai)....	71
Figure 18 First six illustrations of Stolen Fish.....	74
Figure 19 Photograph of the Stolen Fish picture book	76
Figure 20 Screenshot of Stolen Fish annotation in ELAN	78
Figure 21 Illustrations for three Transitive Event Picture Sequences.....	80
Figure 22 TEP Sequence experimental design.....	81
Figure 23 Comparison of TEPS sets: A1's and B1's first three sequences.....	87
Figure 24 Photograph of the 16th sequence of TEPS sets A1 and B1	90
Figure 25 Screenshot of TEPS annotation in ELAN	93
Figure 26 Percentage of clausal IUs by # of R-EXPs.....	101

Figure 27 Deictic gesture accompanying verbless copular construction.....	104
Figure 28 Secondary characters in Scene #27 ‘Eating together.’	107
Figure 29 Stolen Fish scene #1 ‘Line in water’	110
Figure 30 Percentage of response types per target scene condition.....	121
Figure 31 Initially-detached patient in preposed phrasal IU.....	122
Figure 32 Linear order of arguments in clausal IUs with two R-EXPS	128
Figure 33 Stolen Fish scene #24 ‘Nothing to see here.’	136
Figure 34 Waveform attested OVA order.	137
Figure 35 Patterns of word-form shifts.....	151
Figure 36 Patient-initial sequence #19.....	154
Figure 37 Information status of nominal alternants and co-occurring disyllables....	155
Figure 38 Patient-initial sequence #14.....	156
Figure 39 Full sequence #19 illustrations	161
Figure 40 Information status of verbal alternants and co-occurring disyllables.....	162
Figure 41 Full sequence #7 illustrations	163
Figure 42 Resolution Scene sequence #12.....	167
Figure 43 IU-initial low prominence with minor-syllable elision.	176

1 Introduction

Studying lesser-known spoken languages can be a challenge. Unlike large major languages, small minority languages may not only lack linguistic resources, but at the same time they are likely subject to an array of pressures leading to language loss. Moklen, a language of Thailand, is one such language. The population of the Moklen people is estimated to be 4,000 (Arunotai, 2017), but this does not reflect the number of active speakers, as most have shifted towards Thai (Larish, 1999; Pittayaporn & Choemprayong, to appear). Moklen is an endangered language; in all likelihood the last generation of Moklen speakers has already been born. However, active Moklen speakers from older generations still maintain a Moklen cultural identity and use the language amongst each other. This means that the window for linguistic research on this interesting Mainland Southeast Asian language of Austronesian origin is not completely closed.

Previous research on Moklen provides a solid foundation, with a description of basic grammatical features and an extensive inventory of the lexicon (Larish, 1999; Pittayaporn, Pornpottanamas, & Loss, 2022; Swastham, 1982). However, previous grammars are based mostly on a methodology of translation-based direct elicitation, leaving Moklen speech without any systematic framing within communicative contexts. Furthermore, study of discourse was a gap within previous Moklen research, with only a solitary Moklen text available in Larish (2005). Additionally, there were questions remaining around the use of some previously reported variant linguistic forms. One question concerned claims of alternant “word orders” in addition to Moklen’s canonical SVO order. Another area of interest centered around reports of Moklen’s typically disyllabic lexemes frequently displaying reduced monosyllabic forms—for example *ʔeká:n* ‘fish’ being realized as *ká:n*. The Moklen language was already of interest in-itself, but because these areas of variation had never been the direct target of study, morphosyntactic variations in Moklen and moreover discovering their cognitive bases presented an interesting puzzle.

To begin studying Moklen’s alternant word orders and changes in word-form there were two problems. The first one was deciding how to capture the relevant spoken variants while avoiding “unnatural” data. The second problem was whether or not linguistic theory could offer a principled explanation for use of the variant forms

in question. A solution to the first problem was to use picture-based stimuli in staged communicative events, a compromise offering relatively spontaneous speech in a semi-controlled context (Himmelmann, 1998). As for functional explanations, the field of *information structure* offered ideas and methods that promised to, if not explain, at least contextualize linguistic forms within their discourse context. More generally, with this approach, the aim was to provide a *descriptive* account of Moklen speech and linguistic variants and thereby contribute one picture to a documentary record.

In sum, the objective of this study of information structure in Moklen was to describe variations in Moklen speech and investigate potential discourse factors leading to variant linguistic forms. Ideas for the study were spurred by periods of preliminary fieldwork by the researcher during which both the approach and research questions were formulated. Primary data for this study comes from collaborative language elicitation sessions with Moklen speakers, which was then analyzed with reference to hypotheses concerning *intonation units*, *argument structure* and *changes in word-form*.

1.1. Background

1.1.1 Moklen

The Moklen are a group of people living in southern Thailand. Modern-day Moklen are the descendants of Austronesians who some time ago settled in small communities along a 250km strip of Thai coast facing the Andaman sea. Along with the *Moken*, a nearby and closely related group, the Moklen are part of the *Moklenic* group of people and languages (Larish, 1999). Previous linguistic research on the Moklen language includes a basic descriptive grammar (Larish, 2005; Swastham, 1982), a historical reconstruction (Larish, 1999), and most recently a dictionary (Pittayaporn et al., 2022). Until now, there has never been a discourse-based analysis of Moklen speech. Interest in researching Moklen is driven in part by its unique history and place among other Mainland Southeast Asian languages, but it is also motivated by an awareness of its endangered status. Research prioritizing discourse phenomena, therefore, is an opportunity to document the language.

1.1.2 Moklen's variant linguistic forms

Moklen canonical word order is acknowledged to be SVO, but variant orders have also been put forth. Swastham (1982) provided SVO as the normal order for transitive clauses but also listed VSO, OSV, OVS as other possible orders—saying that word order can vary according to the speaker's "emphasis". Larish (1999) also said that Moklen has a fairly rigid SVO order but added that alternate sequences such as VSO are not uncommon, maintaining that these constructions are relics from a Proto-Austronesian verb-initial order. Despite a consensus on canonical order, claims for and examples of alternate orders were not derived from a corpus of naturally occurring speech. Instead these claims were founded upon a methodology of direct elicitation. Therefore, to the extent that alternate word orders were indeed used by speakers, we were without any discourse framing for understanding their use.

Most Moklen lexemes take the form of an iambic disyllable made up of an initial minor syllable and a stressed major syllable (e.g., *tə'ʔáw* 'sea', *ka'ba:ŋ* 'boat'). While this is the most prevalent word-form and the one produced in citation, Moklenic disyllables had also long been noted to exhibit variant monosyllabic "colloquial forms" consisting of just the major syllable (Court, 1971; Lewis, 1960). Larish (1999) pointed out that a common context for these reduced word-forms was compounds, wherein disyllables like *ʔəlá:ŋ* 'people' and *ʔəká:n* 'fish' could be realized as monosyllables, such as *lá:ŋ pələ:w* 'island people' and *ká:n melá:k* 'red-bellied fusilier'. Noting additional reductions of disyllables within connected speech, Larish coined the term "non-ultimate syllabic aphaeresis" to describe a synchronic phenomenon of there being "optional" deletion of minor syllables. With this term, Larish's aim was to distinguish between the reduced monosyllabic forms occurring in speech from an additional diachronic shift of disyllables into monosyllables. For example, monosyllabic verbs such as *dín* 'to come', *káw* 'to go', *dán* 'to know' had already exhibited a permanent loss of the minor syllable, while Moklen's sister language Moken still maintained the corresponding disyllabic forms (*ŋadin*, *lakaw*, and *mədan*, respectively). Obviously, the variable omission of minor syllables synchronically and their loss diachronically suggest a link. Wolff (2010), speaking of Moken, made such a connection explicit when he placed the "weakening of the penult" as part of a "movement towards monosyllabization [that] is most prominently

manifested in the tendency in colloquial styles to pronounce only the major syllable of the word” (p.526). Although the diachronic implications of minor-syllable loss are interesting—especially for a discussion about monosyllabization within the Mainland Southeast Asian linguistic area (Matisoff, 1990; Michaud, 2012)—the appearance of Moklen’s reduced monosyllabic forms, hereafter *monosyllabic alternants*, within actual speech had never been studied.

During preliminary fieldwork by the author, monosyllabic alternants were observed to be a common occurrence. What seemed especially interesting, however, was that word-form alternations often appeared during connected speech in a pattern of the disyllabic *before* monosyllabic. Furthermore, the alternation happened across word classes, such as nouns and verbs like in (1.1) and (1.2).

(1.1) *ticúm nəŋé:n ma4nut4 ... nut4 ləbút*
 bird chase person person run

‘The bird chases the person. The person runs.’

(1.2) *lɛ:w4kɔ:3 didú:n ... dú:n lɛʔ4 ʔɨj wà: luj*
 CONN sleep sleep with dog two CLF

‘so (he) sleeps – (he) sleeps with the dog, the two of them.’

Also notable was that the change in word-form also seemed to correspond with a noticeable contrast in prosodic prominence of each word-forms’ major-syllable (see Figure 1). Crucially, this was not something that could be discovered through direct elicitation but rather was a common feature in free-flowing speech.

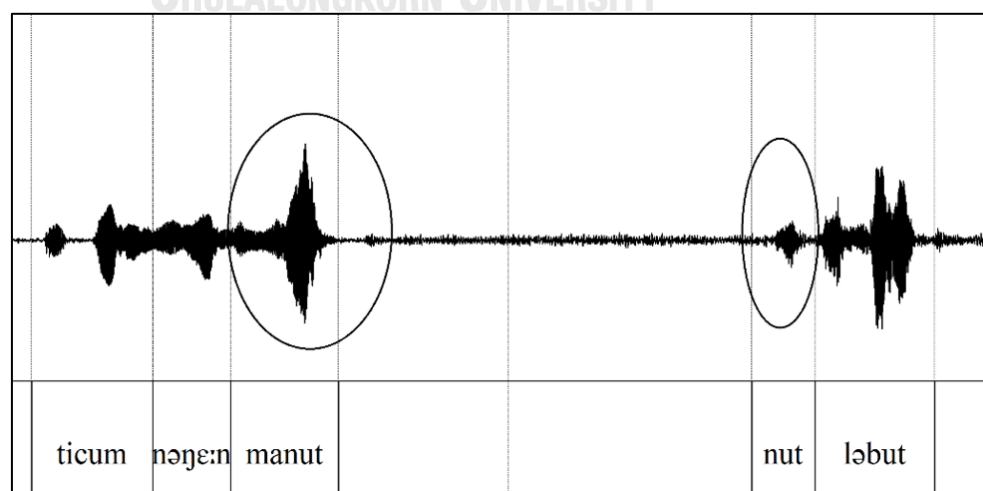


Figure 1 Contrasts between disyllabic and monosyllabic form of /ma4nut4/ in (1.1)

Discovery of this pattern and the emergence of the monosyllabic alternants when they were “old” information raised some unique and interesting questions. After surveying the literature the broadest explanation seemed to be a purported iconic relationship between prominence and communicative importance during discourse (Chafe, 1994; Lambrecht, 1994). Ultimately, this led to the hypotheses and main idea of the study: variant linguistic forms and changes in word-form could be accounted for in terms of *information structure*.

1.1.3 Information structure

“Linguists like to begin with a form, and ask in what ways it can be used... Speakers, on the other hand, start with something to say, and ask what forms will help them say it. And in the actual act of speaking, there is not only an event to be verbalized, but also information to be managed.” (Du Bois, 2003, p.52)

Information structure can be defined as “that component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts” (Lambrecht, 1994, p.5). For information structure studies, the main target of interest is not the propositional or truth-value content of individual messages but instead the manner in which they relate to the situation as a whole; in other words, not so much *what is said* but rather *how it is said*. For example, consider the pairs of sentences in (1.3).

- (1.3) i. a. *John fixed the house.* b. *The house was fixed by John.*
 ii. a. *Two dogs were on the shore.* b. *There were two dogs on the shore.*
 iii. a. *We cooked six of the chickens.* b. *Six of the chickens we cooked.*

While the allosentences of each pair essentially convey the same propositions, within an information structure approach, use of either syntactic schema (a) or (b) can be understood as arising out of contrastive discourse contexts. Differences, however, are not limited solely to sentence grammar, for another layer of complexity is added once we consider variable locations of prosodic accentuation. For example, accentuation of any one of the words in (1.3) could signal informationally salient distinctions e.g., *JOHN fixed the house* vs. *John FIXED the house*. Furthermore, a formulation like *He*

fixed it more directly showcases an informational component as pronominals often replace lexical arguments when discourse referents are “given” information.

Fundamentally, there is no message without information structure. Although any two allosentences can have equivalent propositional content, contrasts in form (e.g. schematic grammatical frames, prosodic prominence, or word choice) ultimately stem from differences in which presuppositions were present at the moment of speech. Using our imaginations we can see how all sorts of preceding discourse contexts could produce any one sentence in (1.3) and correspondingly how it would fit within the informational nexus of the situation. The study of information structure, therefore, represents an attempt to connect how particular linguistic forms structurally encode information to a shared discourse context.

More generally, as a field of linguistics, information structure can be seen as a form of discourse pragmatics (Lambrecht, 1994). Although the field is sometimes characterized as relatively new, it can also be a bit of what Vallduví and Engdahl (2013) call a “terminological quagmire”. Overviews of the field, therefore, commonly acknowledge a lack of consensus around information structure notions and need to devote space to untangling previous theory (Krifka & Musan, 2012; Maticć, 2015). Recent critiques by Maticć (2022) and Ozerov (2018) are clear in calling for more bottom-up research in the field from a more diverse range of languages. In a bottom-up approach, researchers are encouraged to start with observable linguistic phenomena (e.g., alternant word orders or word-forms), see how they might contribute to information management, and then generalize across the data in a dialectical process, all while continually acknowledging their methodological commitments.

Among methodological commitments that can provide a “scaffold” (Haspelmath, (2020) for information structure research, three are central to this study. *Intonation units* are taken as the basic units of speech, serving as a working unit for segmentation of spoken discourse data. Use of intonation units is common within discourse-functional traditions (Barth-Weingarten, 2016), where they are also thought to outline a range of syntactic and informational properties. *Argument structure* concerns the order of clausal constituents (Goldberg, 2006b), a topic commonly

covered by “word order”. Use of argument structure differs, however, in its labeling of clausal constituents in terms of semantic core arguments (e.g. A=AGENT, O=PATIENT). Additional contrast is also evident in the association of argument structure, especially “preferred argument structures” (Du Bois, 2003), for framing syntactic preferences within discourse. *Information status* represents the idea that linguistic expressions (e.g., words or noun phrases) can be classified in terms of their “givenness”—the degree to which they are “old” or “new” information. Several models of information status exist (e.g., Riester & Baumann, 2017), but common to them all is the basic distinction between “given” and “new” information. Finally, one additional methodological concept is the idea that utterances are composed of two informational units. For example, the utterance-level units of *topic* and *focus*, are said to work together in an additive process of communication, whereby common ground content is signified and updated (Matić, 2015). In this study of information structure, intonation units, argument structure, information status were thought of as working concepts for building a *language-particular description*—and not a direct endorsement of some broader linguistic framework, nor a presumption of any notion’s cross-linguistic universality.

As for theoretical ideas, this study looks to three proposals for the role of information management in shaping linguistic forms. The *one-new-idea constraint* proposes that there are limits to the amount of overt information signaled within intonation units (Chafe, 1994). In its broadest conception the one-new-idea constraint is a claim about a natural alignment between consciousness, syntax, and prosody, but in a more practical application the constraint predicts that “new” referring expressions within an intonation unit will be limited to one (Matsumoto, 2003). The *given-before-new principle* proposes that there is a psychological preference for the relative ordering of linguistic elements with regards to their informational properties (Clark & Haviland, 1977). This principle is often connected to research on discourse factors impacting changes in word order (Skopeteas & Fanselow, 2009). The main generalization from this research is that if a language marks givenness using word order alternations, then the marked variant order will be in a given before new pattern (Neeleman & van de Koot, 2015). Lastly, there is the idea that there is a correspondence between the givenness of a linguistic element and its prominence,

whereby things that are more salient or firmly established within a common ground context require less prominence (Lambrecht, 1994). As mentioned earlier for information status, there exist various models for classifying givenness. “Prominence”, however, can be taken as a general concept encompassing any relative differences in amounts of phonetic material by which one element stands out more than another (Baumann & Cangemi 2020). The take-away from this supposed *givenness-prominence relationship* is that predictable or “given” entities are aligned with reduced or comparatively shortened forms, as there is a lower mental cost for “activating” the concept in the mind of a hearer.

Even in this brief introduction, the field of information structure can be seen to encompass several ideas and domains of linguistics. Essential for further development of the field, besides rigorous empirical testing of information structure notions, is the need for more research on lesser-described languages (Adamou et al., 2018; Güldemann et al., 2015). Research on a more diverse range of languages has the potential to broaden perspectives of information structure phenomena as well as substantiate any commonalities. Although the value of such studies is unquestioned, often studies of lesser-described languages face challenges, such as low language vitality, difficulties in access, and a lack of linguistic resources. Information structure researchers especially, may need to develop unique methods for uncovering information structure processes within their specific research context.

Overall, the rationale for the study of Moklen was that it could simultaneously fill a research gap, investigate variant forms, and contribute to some modest documentary aims. Preliminary research by the author led to the idea that an information structure approach would help achieve these aims. First, an intonation-unit based approach would establish a baseline of typical Moklen utterances, beyond what was previously available. And, adopting a methodology with picture stimuli would allow for the targeting of certain Moklen forms, while also aid in framing and understanding speakers’ communicative intentions.

1.2 Research questions, objectives, hypotheses

1.2.1 Research questions

What information structure factors influence the realization and use of variant Moklen linguistic structures? Do Moklen intonation units conform to the one-new-idea constraint? Do Moklen's alternant word orders adhere to the given-before-new principle? Are Moklen's monosyllabic alternants a result of being "given"?

1.2.2 Objectives

Objective 1: To study the syntactic and informational properties of Moklen intonation units.

Objective 2: To study the relationship between changes in information status and Moklen word-form.

1.2.3 Hypotheses

Hypothesis 1: Clausal intonation units will conform to the one-new-idea constraint.

Hypothesis 2: Variations in argument structure and associated grammatical constructions can be accounted for by the given-before-new principle.

Hypothesis 3: Use of monosyllabic alternants will correspond to "given" information statuses.

1.3 Significance

This study attempts to account for Moklen's variant linguistic forms in terms of their relation to information structure. Findings from the study can contribute to broadening the scope of languages within the purview of information structure research, by reporting information structure phenomena in a lesser-known language.

1.4 Definitions

argument structure – (also known as word order or constituent order) the schematic syntactic order of elements of a clause i.e., the linear order of a verb and its core arguments in terms of their semantic roles.

focus – a discourse strategy of broad evidentiality encoding information conveyed by the speaker as her communicative intention and as individual knowledge of which she is the only epistemic source (Masia 2022).

givenness – a key information structure notion representing the degree to which something is either "activated" or known within a common ground context.

Commonly, a distinction between “given” and “new” represents the two ends of the spectrum, wherein “given” is either old or activated information and “new” represents updates to common ground knowledge with previously “inactive” information (Baumann & Riester, 2012). Several classification systems of “givenness” have been put forth in the literature (see information status).

information status – any one givenness classification for a linguistic expression (Riester & Baumann, 2017). Several models of information status categories have been put forth (see §3.3), but at their core is a distinction between “given” and “new” information.

intonation unit (IU) – a prosodic unit of natural speech characterized by a single coherent intonation contour, often marked by cues such as a pause. While comparable to some other corresponding prosodic units discussed in the literature, intonation units and their interfaces are a significant component within discourse-functional linguistic traditions (Barth-Weingarten, 2016). In practice IUs are segmented and delineated by a bundle of perceptual and acoustic cues, such as pausing, pitch reset, and coherent intonational contour (Du Bois, 2008).

minor-syllable elision – a form of clipping in discourse whereby Moklen disyllabic words become abbreviated through omission of the initial minor-syllable (see monosyllabic alternant).

monosyllabic alternant – a reduced monosyllabic form of a Moklen lexeme for which there is still an attested corresponding disyllabic form. In this form the initial minor-syllable is omitted, and the major syllable remains.

prominence – a general concept encompassing any relative differences in amounts of phonetic material by which one element stands out more than another (Baumann & Cangemi 2020).

referring expression (R-EXP) – a linguistic expression that corresponds to a discourse referent. In this study referring expressions are used to encompass what are the core and non-arguments of clauses and what are grammatically outlined as noun-phrases (Baumann & Riester, 2012; Gundel et al., 1993).

topic – a discourse strategy of broad evidentiality encoding information not conveyed as the speaker’s communicative intention and which represents mutual knowledge

established as shared conceptual grounding with both speaker and hearer as committed source (Masia 2022).

word-form shift – an alternation between the disyllabic and monosyllabic word-form of a Moklen lexeme within a singular discourse context.

1.5 Abbreviations

...	=	intonation unit break
1	=	first-person pronoun
2	=	second-person pronoun
3	=	third-person pronoun
A	=	agent-like argument of a transitive clause
AFFR	=	affirmative
C	=	consonant
CLF	=	classifier
CoCl	=	complement clause
COM	=	complementizer
COP	=	copular
CONN	=	connective
DEM	=	demonstrative
DIS	=	distal
DO	=	direct-object
E	=	extension to core
G	=	given
G>N	=	given-before-new
GU	=	grammatical unit
INTS	=	intensifier
IO	=	indirect-object
IU	=	intonation unit
IRR	=	irrealis
M	=	male
MED	=	medial
MSEA	=	Mainland Southeast Asia
O	=	patient-like argument of a transitive clause

N	=	new
NEG	=	negator
NP	=	noun phrase
P	=	plural
PROG	=	progressive
PRF	=	perfect
PRT	=	particle
PROX	=	proximate
QPLR	=	polar question
R-EXP	=	referring expression
REF	=	relativizer
S	=	subject-like argument of intransitive clause
TOP	=	topic
SG	=	singular
V	=	verb
Vi	=	intransitive verb
V	=	vowel (in context of syllable structure)
VOC	=	vocative

1.6 Thesis overview

Chapter 2 provides essential background and history of the Moklen people and language. The aim was to synthesize previous scholarly literature on the topic alongside a view from first-hand experiences within Moklen communities. Chapter 3 is a literature review on the topic of Information Structure. The main focus concerns the methodological concepts and explanatory theories used in constructing this study's description of Moklen. Chapter 4 describes in detail the two custom-picture stimuli used in the study, the *Stolen Fish* picture book and the *Transitive Event Picture Sequences*. Both were designed and implemented by the researcher and served as the discourse contexts for all primary data of the study. All illustrations from these stimuli are viewable in the appendices. Chapter 5 provides findings on informational properties of Moklen intonations units, wherein the role of the one-new-idea constraint is assessed. Chapter 6 reviews findings on the syntactic properties of Moklen and their relation to the given-before-new principle. Chapter 7 shares findings

on the relationship between changes in information status and Moklen word-form. The main idea investigated here is the extent to which “given” information statuses can account for the use of Moklen’s monosyllabic alternant word-forms. At the end of Chapters 5-7 an intermediate discussion of each chapter’s respective findings is provided. Chapter 8 offers a final summary and discussion of the study findings before concluding with implications and suggestions for future research.



2 The Moklen people and language

2.1 The Moklen people

The Moklen are a minority group of Southern Thailand, whose language and ancestry are of Austronesian origin (Larish, 2005). In total, there are around two dozen Moklen communities, most of which are located along the coast by the Andaman Sea in Phang Nga province, with a few on nearby islands Koh Phra Thong and Phuket (CUSRI, 2016), see Figure 2.



Figure 2 Moklen Communities (Adapted from CUSRI, 2016)

The most recent population figure for the Moklen puts their numbers at around 4,000 (Arunotai, 2017), but the number of active speakers of the Moklen language is much fewer. All Moklen speak Thai, and language shift continually heads in towards that direction (Larish, 1999; Pittayaporn & Choemprayong, to appear). Additionally, what

was already a precarious language situation for Moklen was further impacted by devastation from a tsunami in 2004, directly affecting the Moklen people through the loss of life, property, and the displacement of communities. Current use of the Moklen language is limited to older adults, with extremely low transmission to younger generations (Pittayaporn et al., 2022). As an ethnic minority, the Moklen have historically been outsiders, but prolonged interaction with mainland Thai society has resulted in a greater level of integration in all facets of life. Older adults attended local government schools up to grade 4, typical of most rural Thais of this generation. Younger generations have had access to more formal education, and there are seemingly more and more examples of people finishing secondary education and attending university. Not long ago many Moklen may have worked as laborers in the tin mining industry, but now many work as laborers on rubber plantations near the lands on which they live (Arunothai, 2017; Larish, 1999).

All members of Moklen communities were born in Thailand and have Thai citizenship. However, for the oldest Moklen, some still report of not having had it at birth and only obtaining it and a legal surname later. Modern Moklen live and work alongside their Thai neighbors, and commonly have a Thai parent or spouse, so in large part it is natural for them to view themselves in a large part as also being “Thai”. This sentiment is reflected in the use of *Thai Mai* ‘new Thais’, a term used for all of Thailand’s “Sea People”, signifying what Arunothai (2017) calls a process of “Thaisation”. The origin of the term is unclear, but it may have originated with King Rama IX, whose famed visit to the Sea-Peoples in Phuket in 1959 still holds modern relevance (Mueanhawong, 2017). For the Moklen, especially given the multi-generational integration into Thai society, use of the term “Thai Mai” already seems be a misnomer, especially for younger generations whose Moklen ancestry would not exclude them any more than other of members of diverse groups who live throughout the nation (Smalley, 1994).

With Thai citizenship comes identification as a Buddhist, which in some way serves as a nominal contrast with Islam, a major religion of the southern region. Many Moklen do occasionally attend local Buddhist temples, and keep Buddhist iconography, but they also maintain several of their own unique religious practices in the form of ancestor worship, annual ceremonies, and the honoring of spirits (Ferrari

et al., 2006) (see Figure 3). One major difference is the Moklen practice of burying their dead in sacred funeral plots, as opposed to the Buddhist practice of cremation (Larish, 1999). Additionally, some Moklen also observe Chinese traditions and rituals, such as the celebration of the Chinese New Year. Although this is common throughout the broader region more generally, some Moklen still acknowledge specific Chinese ancestors. Adding further context here, historically Moklenic people have been reported to have often relied on close associations to a *taukay*—Chinese mercantile middleman—for assistance in bartering in Thai and Burmese markets (Hogan, 1972; Ivanoff, 1997).



Figure 3 Spirit poles at Ban Khanim, Lampi, and Bangsak

Most Moklen communities still have access to the sea, from which they usually harvest for their own consumption and also for sale in small quantities to local markets. Legal issues surrounding land rights were brought into view particularly during reconstruction and resettlement after the 2004 tsunami (Attavanich et al., 2015). Post-tsunami relief efforts brought in aid, but it also transformed the shape of Moklen homes and communities. Whereas previously the Moklen would have used traditional methods of construction (Attavanich & Kobayashi, 2014), in the aftermath of the tsunami many Moklen communities received homes which utilized concrete and uniform designs in grid-like layouts (see Figure 4).



Figure 4 Houses: traditional style (left), post-tsunami reconstruction style (right)

Modern Moklen homes and communities now have basic utilities, are connected to major roads, and have access to major media outlets both through television and mobile devices. Additionally, during the post-tsunami reconstruction period several communities received community cultural centers aimed at helping to highlight and preserve their unique cultural identity (see Figure 5). These centers keep small collections of traditional handiwork, pamphlets, murals and other displays aimed at encapsulating and communicating both a Moklen identity and broader affiliation with the “Sea Peoples” of Thailand (see §2.2). These sites serve as the location for community meetings, hosting of tours groups, and maybe even a rare language lesson for community children. The Moklen language, however, has no indigenous script, and so when put in writing it is usually done impromptu using the Thai script, an abugida of Old Khmer origin. There has never been an active Moklen literary tradition, and most Moklen speakers have low levels of literacy. Recently, there has been a proposal for a formalized use of the Thai orthography for Moklen (Pittayaporn & Choemprayong, to appear) and the compilation of a Moklen dictionary (Pittayaporn et al., 2022).¹

¹ Historically, there have been three other forays into Moklenic writing. The oldest of these is the Burmese-Karen-missionary line starting with Brayton (1846), which includes literacy primers and a complete translation of the Bible by Naw Say Bay (p.c.) for Burmese Moken, see also Koh (2016). Hogan (1983) is an unpublished manuscript for using Thai for Moken. Premsirat et al. (2013) proposed a Thai-based script for the Moken of Koh Surin.



Figure 5 Moklen Cultural Centers at Tap Tawan, Tung Waa, and Thap Pla

The post-tsunami period marks a revitalization of interest in all of Thailand's Sea Peoples, which besides Moklen includes the Moken and Urak Lawoi. Continued advocacy domestically and internationally, as well as Moklen's own community representatives have continued to spotlight issues these communities face (Arunotai, 2017; IRASEC, 2009; Robinson et al., 2021). Recent interest in understanding the Moklen's traditional way of life and language is due in part to the recognition that the Moklen have likely passed the precipice of a cultural and linguistic decline. In all likelihood, the last generation of active speakers has already been born, and there are no formal language revitalization efforts in place. Many are now aware that given this current state of affairs pursuing any interest in documenting and describing the Moklen language is a linguistic emergency.

2.2 Previous scholarship and path to recognition

To understand the Moklen, one must also know the *Moken*, a closely related group. Together, they form a larger *Moklenic* group (Larish, 1999). The Moken mostly live north of the Moklen, on the islands of Burma's Mergui Archipelago, but there do exist Moken communities on the Thai islands of Chang, Lao, Phayam, Phuket, Sin Hai, and Surin (Arunotai, 2008). Presently, the island dwelling Moken's way of life is largely sedentary, however, it was still within the 20th century that they lived the life of "gypsies" or "nomads", who sailing their *kabangs* traversed a maze of waterways and coastal islands (Ivanoff, 1997). The terms "Sea Gypsies" or "Sea Nomads" have often been used when talking of Thailand's "Sea Peoples" or for other Austronesian groups. In the anthropological literature, "sea gypsies/nomads" refers to

any of the small groups of nomadic boat people found scattered throughout what Sopher (1965) called “the Malaysian littoral”. Besides all these groups being Austronesian in origin and having some similar cultural traits, they are functionally unrelated groups. Chou (2020) says that the term Sea Nomads can refer to three groups. The largest, the *Bajau Laut*, are located in the areas east of Borneo, the *Orang Laut* are in the Riau-Lingga archipelagoes and east coast of Sumatra, and then there are the *Moken* who are the furthest northern reaching, residing on islands of Burmese waters down to the southwestern coast of Thailand. And although within living memory the lifestyle of the Moken could be described as nomadic, that they and their sister group, the *Moklen*, have had permanent settlements for some time now makes contemporary use of exonyms like gypsies or nomads inaccurate. Therefore, use of these terms, at most, should be left for describing the lifestyle of the Moken’s recent ancestors, and the Moklen’s more distant ones. Moreover, use of each group’s respective endonym—*Moken* or *Moklen*—not only acknowledges a divergence in their histories, but it emphasizes the recognition that they are at the same time related, yet distinct, groups.

Historically, it took some time before the geographic, cultural, and linguistic boundaries distinguishing the Moklen from the Moken were ever noticed. But, tracing the path leading up to their recognition offers insight into many issues and themes still relevant for understanding the Moklen. The first recorded report of any Moklenic people in the Western scholarly tradition² is in an 1827 report from colonial Burma acknowledging the existence of the “Chalome”, an Anglicized Burmese term for the *Moken* (elsewhere rendered: Chillones, Selong, Selung, Salong, Salones or Salons). Revealed here is not only a depiction of the Moken’s nomadic way of life, but reference to their fears of the nearby majority cultures, a consistent theme in later ethnographic reports as well. The report states:

"A race of people termed, by the Burmans, Chalome and Pase, are to be found scattered throughout the Mergui Archipelago. But their dread of Malayan and other pirates has compelled these poor creatures to adopt an unsettled mode of life. During the north-east monsoon, they are obliged to remove from the vicinity of those islands which are most

² Hinshirana (1996) reports Thai documentation for *chaaw nam* ‘water people’ and *chaaw thalee* ‘sea people’ in the late 19th and early 20th centuries. Both are general Thai terms used for any of Thailand’s ethnic sea people and therefore the precise group is unknown.

frequented, to escape being carried off as slaves by Siamese, Burmans and Malays, who visit them in quest of the valuable commodities they afford. They appear to be a harmless, and, from necessity, an industrious race. The whole tribe consists of no more than four hundred souls. They exchange mats and produce of the islands, for clothes and other articles, conveyed to them from Mergui. Another tribe of this race is thinly spread over the islands lying close in front of Mergui. They all seem to have adopted the religion of Buddha, and to have conformed, in a great degree, to the Burman mode of dress. They scarcely know the value of money, and are, therefore, losers in the bartering trade with the Chinese and others who visit them. Perhaps they think themselves the greater gainers, since they give away products of no use to them for others of vital importance, and are, thereby, enabled to maintain a degree of wild independence" (Wilson, 1827, p.434).

The first recorded instance of the word "Moken" is found briefly in a Pwo Karen script' as part of a subtitle to the 1846 *A Primer of the Selong Language*, by American Baptist Mission Press, but use of the Burmese exonym would persist for the remainder of the 19th century. For example, in a word list of approximately 100 "Selung" words, Spearman (1880) lists the item *maw-keng* only as a word meaning 'man'. Anderson's (1890) full treatise on the Selungs summarized the day's current knowledge about the Moken in Burma, including a compilation of linguistic data and even speculation about the language's Malay-like affinities. Still, despite also contributing original fieldwork the name "Moken" was overlooked. The first real explicit acknowledgment of the group's endonym "Maw-khen" appears in Carrapiett, (1909), who upon discovering it, also provides a purported etymological meaning of 'drowned people'. The first prolific and insistent use of "Mawken" appears in White's (1922) *The Sea-Gypsies of Malaya*. Here, White also expanded upon the apparent etymological root, sharing that the disyllables *l'maw* means 'to drown' and *o'ken* 'salt water', and moreover he hypothesizes that these two words coalesced into 'the sea-drowned'. Ainsworth (1930), an intimate look at the Moken of Mergui during this period, further clarified the purported etymological origin of 'the sea-drowned' by saying that *maw* also refers to the *kabang*'s strake (a long piece of planking which runs the entire length of a boat). Bernatizik (1938) is a notable work of this era, wherein he documented several aspects of the Moken of Burmese waters (republished as Bernatizik, 2005). Bernatizik offers detailed ethnographic reports, stunning

photographs, and even several Moken folktales. And it is here where we see a shift to the “Moken” spelling. By the end of this pre-war era, the name Moken had been established, but these reports were all seaward facing, and it would still be several decades before any recognition of the Moklen of the Thai mainland.

The first explicit mention of the name *Moklen* appears in Hogan (1972). Here, Hogan was in part responding to Court’s (1971) “fleeting encounter” with the *Moken* of Thailand, in which he pointed out that up to that point no one had seemed to have studied the Moken of Thai territorial waters. Relaying his experience, Court reported that while on a trip to the island Koh Phra Thong and back to the mainland, he became aware of the distinction made by Thais of there being *chā:w kòʔ thé:* ‘real islanders’ and the *Thai Maj /tʰāj māj/* ‘new Thais’. Court explains that ‘real islanders’ referred to those Moken still leading a nomadic existence between islands, while the Thai Mai had settled on coastal villages and adopted Thai surnames. Also in his brief encounter, Court, a linguist, was able to notice differences between the two groups from wordlists he collected. Hogan who by that time had spent over a decade working with Thailand’s Sea Peoples, especially the Urak Lawoi, was able to elucidate upon Court’s encounter with the introduction of the slightly different yet important endonym *Moklen*. Hogan explains,

“In their own language the Sea Gypsies call themselves Moken or Moklen (*məken, məklen*) according to their dialect. The word Moken has sometimes been written Mawken. So far as I know the name Moklen has not been mentioned in the literature before” (p.206).

Hogan also added that the Moklen refer to themselves in Thai as *chā:w bòk* ‘land people’, as opposed to *chā:w kòʔ* ‘island people’. Additionally, Hogan indicated the specific locations of several Moklen communities along the coast of Phang Nga and in the northern part of Phuket, offering the first map with precise locations of Thailand’s Sea People. As the first explicit recognition of the Moklen, Hogan (1972) is a key piece in the historical recognition of the Moklen, especially considering that up to this time, the Moken had already been known for over a century. In addition to providing a rough population estimate of 1,000 people, Hogan noted that the Moklen in this area had taken to the cultivation of agricultural produce, and at this point a school for only Moklen children had already begun in Thai Muang district of Phang

Nga—showing that many communities of Moklen had begun the acculturation process for some time. Hogan (1972), therefore, not only marks the first explicit recognition of the Moklen, as distinct from the Moken, but is the first published account showing that the Moklen had been settled on the Thai mainland for quite some time.³

Post-war research on Moklen's sister group, the *Moken*, had actually resumed briefly in 1957 with work of Pierre Ivanoff (See Ivanoff, 1997). Later his son, Jacques Ivanoff, began his own ethnographic work on the Moken at the Surin Islands of Thailand in 1982. In the years since, Ivanoff and colleagues have continued to provide ethnographic reports and some of the most intimate looks at Moklenic life and culture, within both Burmese and Thai territories (e.g., Ferrari et al., 2006). Although these works are mostly anthropological in nature, throughout their work there are several insights into the language, with several works including glossaries and discussion of both Moken and Moklen ideas and practices. Ivanoff (2001) is a large volume, offering translations of Moklenic folktales. Interestingly, with regards to the etymology of the name Moken, Ivanoff argues that the true source of its meaning is found in a Moken origin myth, which he also argues encapsulates the “mytho-historical” motives for their traditional nomadism. Briefly, the tale goes: there was once a powerful queen who condemned *Kèn*, a Moken woman, to be ‘immersed in the sea’ with the line *lemò kèn* ‘drown *Kèn*’. The condemnation caused *Kèn* and her people to build and flee on *kabangs*—boats whose design is also full of symbolism (Ivanoff, 1999). On Ivanoff's account, the name *Moken* then is a colloquial rendering of *lemo ken* ‘immerse’ or ‘drown’ *Kèn*, a construction which Ivanoff points out features loss of the verbal prefix *lè*, old Austronesian morphology which is often elided in normal speech. Ivanoff, in interpreting the endonym's significance, believes the name *Moken* is a source of cultural identity as it reveals the mythological roots of their nomadic existence, something he says the land-dwelling *Moklen* have lost the connection to.

³ Mason (1860), writing from colonial Burma, relays the following report from a Moken family of Burma “The family state that the islands to the southward of the British territories are frequented by Salones in greater numbers than those in the Mergui Archipelago, and that some of the southern Salones have taken to cultivation, and form permanent villages. The language is the same with that of the Salones of the Mergui Archipelago” (pp. 100-101).

Considering their current location, one scenario for the origin of the Moklen is that they are descendants of a Moklenic group who settled on the coast during a migratory path towards the northern waters of the Mergui Archipelago. For example, Sopher (1965, p.198) cites reports from Portuguese explorers, arguing that the Moken could have been in the Mergui Archipelago as early as the 16th century. If Moken groups sailing along a south to north route had reached Mergui by this time (see Figure 2), this means they would have already traveled past the locations of modern Moklen communities in Thailand. However, the history of Austronesians in maritime Southeast Asia stretches far beyond any available report about Moklenic people.

One intriguing piece to the puzzle of the Moklen's origin comes in the form of a traditional folktale about their ancestor Sampan. The basic story goes: once a Moklen man named Sampan and his kin were forced to migrate from Nakhon Si Thammarat, a city on the eastern coast of the Thai peninsula that faces the Gulf of Thailand (see Figure 2). After a conflict with a local leader, perhaps related somehow to being subjected to forced labor for construction at the site of the Phra Mahathat Temple (Ivanoff, 2005), the Moklen people escaped and eventually ended up on the opposite western coast—a trip which is either a 200 km trek over land or a 2000 km *kabang*-trip around the Malay peninsula by sea.⁴ Deciding to not go any further, these Moklen people settled and dispersed into different seaside communities throughout Phang Nga and Phuket provinces. Today at the beach of Bangsak, Phang Nga, the Moklen still maintain an altar and spirit poles in honor of their ancestor Sampan, and this location still serves as the site of an annual ceremony and rituals. Other Moklen communities, who all acknowledge their relation to Sampan, also construct similar spirit poles and annually honor a unique, yet related, Moklen progenitor of their own community (see Figure 3).

The Moklen and the Moken's histories are obviously intertwined. Mutual intelligibility testing of the two languages by Bishop and Peterson (1987) and further research by Larish (1999) established that they were in fact two distinct languages.⁵

⁴ A potential land crossing is explored in Larish (1999) and Benjamin (forthcoming).

⁵ In the interest of informally checking mutual intelligibility, I played recordings of Moken missionary materials created for Burmese Moken to some Thai Moklen. These are materials from projects led by Naw Say Bay (See Koh 2007). While some Moklen with closer relations to Moken could recognize some similarities, most perceive it immediately as markedly different and unintelligible.

Additionally, both within Moken and Moklen linguistic variations are found across communities (Larish, 1999; Naw Say Bay, 1995). Describing the situation, Larish (1999) uses the dialect chain model, saying that Moken and Moklen have become two separate chains, with the break happening at the Thai-Burmese border. However, despite linguistic differences, which in part explain the variation in pronunciation of the two endonyms: *Moken* and *Moklen*, these names are *not* the most salient feature by which either population contrasts themselves with the other. Hence, it is possible in some places to see the word “Moken” (or some variant spelling) in English signage and community materials (see Figure 5). But, during the author’s fieldwork, it was found that some literate Moklen speakers have a metalinguistic awareness of the different Thai spellings (มอแกน and มอเกลน). Historically, the Moklen in their own language have framed the distinction in terms of location or geography, with there being *ʔólá:ŋ pólà:w* ‘islanders’ vs. *ʔólá:ŋ datá:* ‘mainlanders’ (lit. ‘people island’ and ‘people on [the mainland]’) (Larish, 1999).

Presently, there is still interaction among the Moklen and Moken of Thailand. For example, cases of intermarriage between Moklen and the Moken of Thailand’s Surin Islands were noted during the author’s fieldwork. However, based on all the geographic, cultural, and linguistic factors the Moken and Moklen are two distinct groups. Unfortunately however, because the Moken were recognized first, and perhaps because of popular romanticization of their once traditional nomadic way of life, the Moklen of the Thai mainland have seemingly been overshadowed. One result of this is that even in contemporary academic literature the term Moken can be found as a heading for the broader group as a whole with no acknowledgment of the Moklen (e.g., Smith, 2017). Larish (1999), however, made the important point of referring to both Moken and Moklen as forming the *Moklenic* group. Using the term *Moklenic* is significant going forward for further discussion of these people, not only because it markedly recognizes the Moklen, but it also frames the broader context with which we can more accurately capture all the factors which have, and continue, to impact all Moklenic peoples.

2.3 The place of the Moklenic languages

Besides the precise details of a Moken-Moklen split, a bigger question is where exactly does Moklenic fit within the Austronesian family, a topic that has been

broached from the fields of historical linguistics (Larish, 1999; Smith, 2017), ethnography (Ivanoff, 1999), and genetic phylogeography (Dancause et al., 2009; Seetaraso et al., 2020). Modern Austronesian languages resulting from the migratory expansion out of Taiwan are grouped under the heading Malayo-Polynesian, which includes more than 1,000 languages (Adelaar, 2005). Malayo-Polynesian languages—and most Austronesian languages for that matter—are found throughout the Oceanic world and maritime Southeast Asia. Moklenic languages, however, are of the very few that are located in *Mainland* Southeast Asia. Chamic languages, the earliest attested Austronesian languages with 4th century inscriptions in what is modern day Vietnam, is another notable case (Thurgood, 1999). Other nearby Austronesian languages include Malay, south of current Moklenic populations, and Aceh on the northern part of the island of Sumatra. In his reconstruction, Larish (1999) concluded that Proto-Moklenic and Proto-Malayo-Aceh-Chamic could possibly be grouped at a level called *Proto-Malayo-Aceh-Chamic-Moklenic* (see Figure 6).

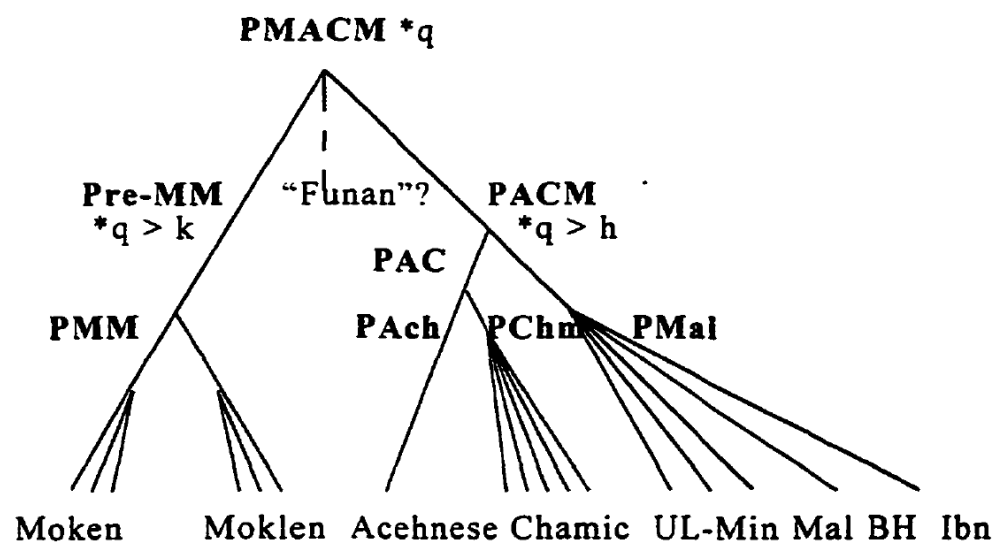


Figure 6 Larish's (1999, p.364) PMACM Subgroup Tree Diagram

(PMACM=Proto-Malayo-Aceh-Chamic-Moklenic; PACM=Proto-Malayo-Aceh-Chamic; PMM=Proto-Moken-Moklen)

Other authors, however, have tended to distance Moklenic from Chamic (Blust, 2010; Sidwell, 2005; Thurgood, 1999). Most recently Smith (2017) challenged many previous higher-order subgroupings of Malayo-Polynesian, and proposed several primary branches, with Malay, Aceh, and Chamic falling within a Western

Indonesian branch, while Moklenic represents a group of “unknown origin” and an outsider to any known group (see Figure 7).

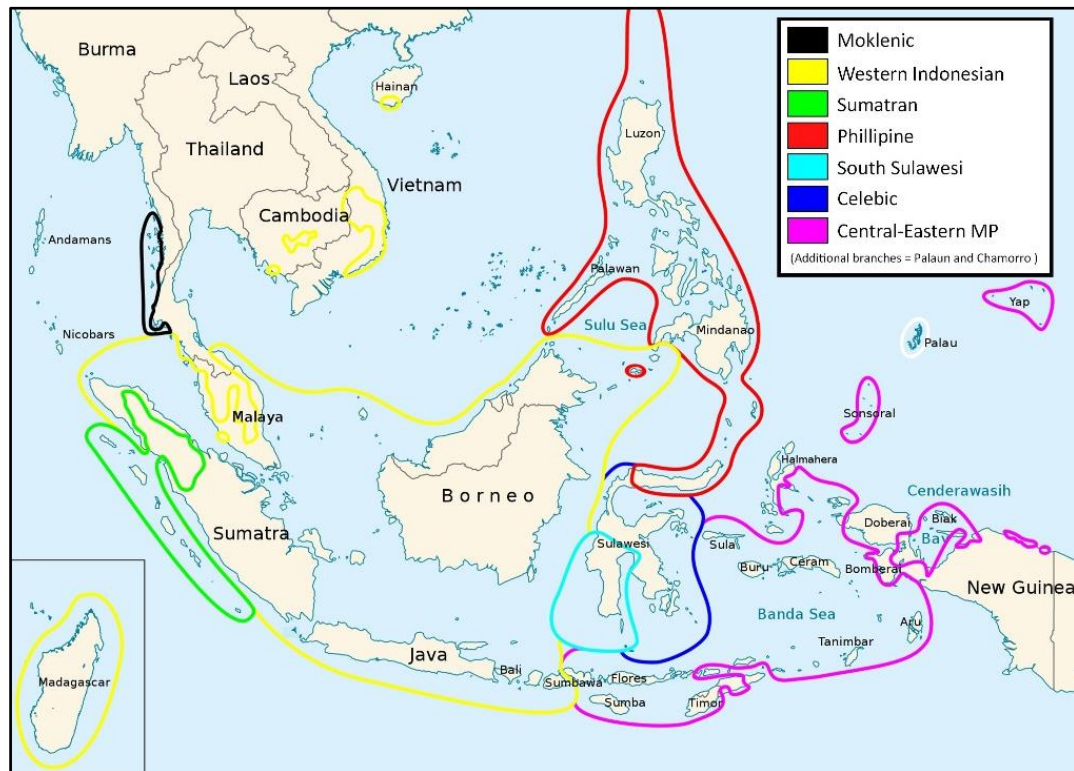


Figure 7 Smith's (2017) Malayo-Polynesian higher-order subgroupings
(Adapted from "[Malayo-Polynesian-en.svg](#)" by Kwamikagami licensed under CC BY-SA 3.0)

This current view means there is no meaningful genetic affiliation between Chamic and Moklen languages. However, both languages' location on Mainland Southeast Asia (MSEA) still holds some significance. One reason for Larish's (1999) linking of Moklenic to Chamic were the signs of Austroasiatic (a major language family of the region) influence on both of these languages, a hypothesis originally put forth by Cowan, (1948). Beyond the borrowing of lexical material, other changes attributed to Austroasiatic contact included a more isolating syntax, reduction of the phonological word (e.g., monosyllabization), increasing complexity of phonemic distinctions, even the development of register systems and sets of tonal contrasts, all features characteristic of languages of the region (Grant & Sidwell, 2005). Given current Moklenic speaking communities' locations, Larish pointed to Aslian languages, Austroasiatic languages of the Malay peninsula, as a potential historical

influence (see also Benjamin, forthcoming), all the while seeing Moklenic convergence to MSEA linguistic norms as analogous to the hypothesized contact-induced changes in Chamic. However, continued discussion on the precise role of language contact has painted a more complex picture on the potential influence of Austroasiatic languages for these Austronesian languages of MSEA (Brunelle, 2020; Brunelle & Pittayaporn, 2012; Pittayaporn, 2005). In this more nuanced view, mere language contact is insufficient for explaining a convergence to MSEA prosodic norms. Instead, structural constraints and internal phonetic/phonological pressures can be placed at the forefront of what are largely language-internal developments.

2.4 Moklen language

Despite Moken and Moklen being two separate languages, broad structural similarities mean they have often been looked at together. Purely linguistic work on Moklenic languages has focused almost exclusively on the Moken of Thailand (Baclawski Jr. & Jenks, 2016; Jenks, 2010; Kraissame, 2016; Pittayaporn, 2005, 2006; Premsirat et al., 2013; Thavisak, 2004). Linguistic research on Moklen is mainly limited to Swastham (1982) and Larish (1999). Swastham (1982) is a short grammatical description. Larish (1999) is a historical reconstruction of Proto-Moklenic which includes an extensive compilation of Moklenic word lists, along with original fieldwork with Moklen speakers. Most recently is the release of a dictionary based on the Moklen dialect of Bangsak (Pittayaporn et al., 2022). The language of most contemporary influence on Moklen is Thai, particularly its Southern Thai varieties. Larish (1999, 2005) emphasized the role of Moklen-Thai bilingualism in shaping Moklen syntax, and the traces of Thai influence are apparent in Swastham (1982) as well. Present-day influence is immediately apparent in the large influx of Thai lexical material as well as the extensive use of Thai function words and grammatical constructions. Fieldwork for Larish (1999) took place from 1991-1992, during which Larish could already see language shift towards Thai. The current situation, a generation later, proves this to be a continued trend.

2.4.1 Phonology

This section draws from the phonological analysis in Pittayaporn et al. (2022), based on the Moklen dialect around Bangsak. Moklen phonological words—of Austronesian origin—can either be a monosyllable, like *ʔɣj* ‘dog’, or an iambic disyllable such as *ʔeká:n* ‘fish’. For disyllabic words, the second ‘major’ syllable receives primary stress while the initial ‘minor’ syllable is unstressed. The major syllable can be an open syllable with a long vowel (e.g., *dagà:* ‘basket’) or it can be a closed syllable with either a short or long vowel (e.g., *niɣít* ‘to cut’, *buwá:k* ‘fruit’). The minor syllable is always open and unstressed. The disyllabic structure of Moklen words is made up of consonants and vowels in the form of (CV).(C)CV(C), while monosyllables have the form of (C)CV(C). In total, including some dialectal variation, there are twenty consonants which are possible in syllable initial position (see Table 1).

Table 1 Initial Consonants in Moklen

	bilabial	alveolar	palatal	velar	glottal
	p-	t-	c-	k-	ʔ-
plosive	p ^h -	t ^h -	c ^h -	k ^h -	
	b-	d-		g-	
fricative					h-
nasal	m-	n-	ɲ-	ŋ-	
approximant	w-	l-	j-		

Moklen consonant clusters /pl/ and /kl/ are possible in syllable initial position such as in *plá:ŋ* ‘sunlight’ and *klá:n* ‘bone’. There are ten possible consonants found in the coda of stressed syllables (see Table 2).

Table 2 Final consonants in Moklen

	bilabial	alveolar	palatal	velar	glottal
plosive	-p	-t		-k	-ʔ
fricative					-h
nasal	-m	-n		-ŋ	
approximant	-w		-j		

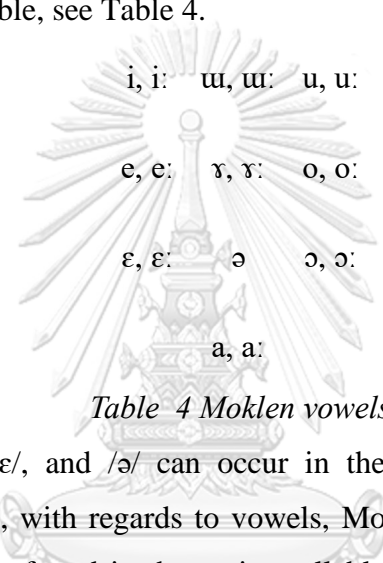
The left-edge of Moklen disyllables is a place of considerable variation, both in terms of dialectal and allomorphic variation. For example, for alveolar phonemes the manner of articulation may vary so that [d], [l], and [n] are variably realized in some words. Similarly, bilabial phonemes /b/ and /m/ are found to create allomorphic forms. Variation of initial consonants can also include changes in place of articulation so that allomorphs of disyllables beginning with [t] and [k], and [c] and [k] are also found. Another, slight variation is with /c^h/ where varying levels of frication create the potential of it being realized as [s], an allophone which can also appear in the initial consonant of major syllables as well. Examples of these variations are in Table 3.

Table 3 Variation of initial consonants

[d]			
[l]	<i>daʔón, laʔón</i> ‘leaf’	<i>lɔc^há:ʔ, nɔc^há:ʔ</i> ‘deer’	<i>didú:n, nidú:n</i> ‘to sleep’
[n]			
[b]	<i>bináj, mináj</i> ‘woman’		<i>buná:ʔ, muná:ʔ</i> ‘flower’
[m]			
[t]	<i>ticúm, kicúm</i> ‘bird’		<i>təmán, kəmán</i> ‘to fish.with.rod’
[k]			
[c]	<i>caná:t, kaná:t</i> ‘child’		<i>cəwá:t, kəwá:t</i> ‘clothes’
[k]			
[c ^h]	<i>c^hé:m, sé:m</i> ‘Thai’	<i>c^həbá:j, səbá:j</i> ‘meal’	<i>peç^háŋ, pesáŋ</i> ‘banana’
[s]			

One additional variation to note is a conservative pronunciation of an initial /g/ in Moklen communities in Phuket, such as /gajáh/ vs. /dajáh/ ‘elephant’. The appearance of /g/ can be seen as a retention of a form still found in Moken, such as in Moken /gací:/ ‘larva’ versus Moklen [dací:] or [lací:] ‘larva’. Larish (2005) points out that, the loss of the voiced velar plosive /g/, as well as the appearance of [s], can be accounted for by a shift towards the Thai phonological inventory.

Altogether, Moklen has nineteen vowels, nine of which can also contrast for length in the major syllable, see Table 4.



i, i:	ɯ, ɯ:	u, u:
e, e:	ɤ, ɤ:	o, o:
ɛ, ɛ:	ə	ɔ, ɔ:
	a, a:	

Table 4 Moklen vowels

Only /a/, /i/, /u/, /ɔ/, /ɛ/, and /ə/ can occur in the unstressed minor-syllables of disyllabic words. Lastly, with regards to vowels, Moklen has three diphthongs. /iə/, /uə/, and /uə/ which are found in the major syllable like in *ʔiák* ‘shellfish’, *bəluəj* ‘sore’, and *bitiək* ‘star’.

One interesting feature of Moklen is its lexical tone. Court (1971) had noticed that a few Moklen words seemed to have fixed pitch patterns, but the first explicit mention of Moklenic lexical tone is in Chantanakomes (1980), who noted a few tonal minimal pairs in Moken of Rawai, Phuket. Larish (1997) added some additional potential tonal minimal pairs for Moklen along with a proposed analysis of possible pitch contours. Recent analysis in Pittayaporn et al. (2022) holds that Moklen has two tones. Tone 1 is usually realized first with high pitch in and then a falling contour [V] (e.g., *kəlát* ‘hot’ and *ʔá:k* ‘to place’), while Tone 2 often has a mid-level pitch followed by a rise [ɿ] (e.g., *kələt* ‘mushroom’ and *ʔà:k* ‘crow’). Tone 1 appears to be much more common, but overall lexical tone has a low functional load.

2.4.2 Basic syntax

Moklen shares many of the general syntactic properties of other Mainland Southeast Asian languages. It has serial verb constructions, a range of utterance final particles, zero-anaphora is widely used, and there is no productive inflectional morphology (Enfield, 2021). However, historical remnants of productive prefixes are evident in some words (Larish, 2005). For a more general overview of Moklen syntax see Swastham (1982) and Larish (1999, 2005). Both Swastham (1982) and Larish's (1999) grammatical descriptions were based on a methodology of translation-based direct elicitation.⁶ Swastham (1982) was a short Tagmemic grammatical description based mostly on data elicitation from a single informant. Larish's (1999) grammatical analysis was based mostly on Moklen and using Lexicase Theory. The only previously published Moklen text is in Larish (2005). As a large portion of this current study is focused on aspects of clausal syntax, discussion in this section is limited to basic features of noun phrases based on the current research and a brief review of previous accounts of clausal syntax.

Grammatically, referring expressions take the form of head-initial noun phrases. For example, nominal heads may take a verbal adjective:

(2.1) *pəkón dajó:ŋ*

tree tall

'tall tree'

(2.2) *ʔé:n cəŋá:m*

water cold

'cold water'

Or it may be followed by another nominal to form a compound or noun phrase:

(2.3) *bul̀j matá:ʔ*

hair eye

'eyelashes'

⁶ Larish (1999) emphasizing the role of Moklen-Thai bilingualism in shaping Moklen syntax says, "When Moklen consultants translate their Moklen sentences into Thai, the Moklen and Thai constructions are normally congruent...the majority of Moklen sentences analyzed above have word-for-word counterparts in Thai" then adding in a footnote "To a certain extent, this may be partly an artifact of the author's fieldwork technique of eliciting Moklen sentences through use of Thai prompts and requesting Thai translations. Further research will probably reveal additional instances where the two languages do not have isomorphic constructions" (p. 271).

- (2.4) *matá:ʔ təmán*
 eye fishing.rod
 ‘fishhook’

Possession works similarly as an initial possessed noun is modified by the possessor. In (2.5) the possessor is a pronoun (see Larish (2005) for Moklen pronominals). In (2.6) the possessor is indicated with a proper name fronted with the kinship term *ʔibú:m* ‘elder female’.

- (2.5) *ʔəmá:k bɿj*
 home 2SG
 ‘your house’
- (2.6) *má:k bú:m nɿj*
 home elder.female Noi
 ‘Elder Noi’s house’

Kinship terms like *ʔibú:m* ‘elder female’ or *ʔebá:p* ‘elder male’ can also serve as independent nouns, and take other modifiers such as demonstratives, which also follow a nominal head.

- (2.7) *bá:p nú:*
 elder.male DEM.MED
 ‘that male elder’
- (2.8) *kanáj nɿj*
 man DEM.PROX
 ‘this man’

Numerical quantification of nominal expressions is accomplished with a closed-class set of classifiers. Interestingly, Moklen general classifier *bulàt* and is used for numerical values with 0, 1, or 2 in the ones place, but *p^hóh* is used for values with 3-9 in the ones place, as shown in (2.9). Additionally, a proclitic *ʔa-* can be used along with the classifier for quantities of one.

- (2.9) a: *manók ʔa=bulàt*
 chicken one=CLF
 ‘one chicken’
- b: *manók wà:ʔ làt*
 chicken two CLF

'two chickens'

c: *manók talíj p^hóh*

chicken three CLF

'three chickens'

Classifiers may also appear without an overt nominal head; however, given some classifiers specific relation to semantic classes, the referent type of which they enumerate is still retrievable. For example, the classifier *lùj* is used only for people:

(2.10) Ø *pá:t lùj*

four CLF

'four people'

Both Swastham (1982) and Larish (1999) put forth SVO as Moklen canonical word order; however, they also report that variant orders are possible. For transitive clauses Swastham (1982) provides SVO as the normal order but adds that word order can vary according to the speaker's "emphasis" before listing VSO, OSV, OVS as other possible orders. Swastham's examples of Moklen variation are presented in a "transformational battery", wherein identical propositions are just reordered to illustrate a list of argument structure possibilities, as in (2.11).

(2.11) Swastham's (1982:67-68) "transformational battery"

(a) *wɔ:ŋ mətət kaʔɛ:w*

Wong to.cut wood

'Wong cuts wood.'

(b) *kaʔɛ:w wɔ:ŋ mətət*

wood Wong to.cut

'Wong cuts wood.'

(c) *mətət wɔ:ŋ kaʔɛ:w*

to.cut Wong wood

'Wong cuts wood.'

Larish (1999) also upholds SVO as the main word order for Moklen. However, he says that alternate sequences such as VSO are not uncommon, especially with clitic

pronouns, maintaining that these constructions are relics from a Proto-Austronesian verb-initial order, an example of which he offers in (2.12)

(2.12) [Larish, 1999:212]

kaw laŋ met ka?
 go they all already
 ‘They all went already.’

Larish (1999, 2005) frames Moklen syntax as conforming to many of the morphosyntactic norms of Mainland Southeast Asian languages, such as serial verbs and zero-anaphora. One interesting feature which distinguishes Moklen from neighboring languages is its postverbal negation, with the negator *hah*, such as (2.13) which Larish aligns with a Thai example.

(2.13) [Larish 2005:253]

(Moklen)	<i>bəh hãh ka:n ləy nəy cʰuba:y hãh</i>
	do NEG work day PROX well NEG
	‘I am not working today, (for) I do not feel well.’
(Thai)	<i>mây tʰam ɲa:n wan nî: mây saba:y</i>
	NEG do work day PROX NEG well
	‘I am not working today, (for) I do not feel well.’

Further examples of syntax from Swastham (1982) and Larish (1999; 2005) mainly serve as a cursory outline of the placement of clausal elements.

2.5 Summary

The Moklen are a people of Southern Thailand whose language is of Austronesian origin. One of two Moklenic languages, Moklen is a lesser-described language and in need of description and documentation, particularly in matters of discourse. Previous research has laid a solid foundation in terms of an inventory of lexical items (Larish, 1999, Pittyaporn et al., 2022) and major syntactic functions (Larish, 2005; Swastham, 1982), but left a gap for research on Moklen discourse.

3 Information structure

3.1 Introduction

Information structure is a field of linguistics that looks at how the packaging or management of linguistic forms encodes the exchange of information between interlocutors. Essential to the view of this current study is the primacy of spontaneous spoken data and an interest in a relationship between language and the mind. Other sources of data such as writing, direct elicitation, or grammaticality judgments are seen as removed from the types of natural communicative contexts and cognitive factors that impact speech. To develop an analysis of information structure it is essential to frame both discourse context and speakers' communicative intentions. Because this study does not adopt any one particular framework, I would like to distinguish between *methodological commitments* and *theoretical claims*. Three notions represent major methodological commitments: intonation units, argument structure, and information status. These are the study's assumed working concepts used to classify types of linguistic forms.

Analysis of speech begins with *intonation units*, the naturally occurring chunks and spurts that characterize the flow of speech (Chafe 1994). Intonation units represent a basic reference unit from which further analysis of speech can be done. *Argument structure*, the schematic order of clausal elements, offers a semantic approach to analyzing clausal relations in terms of their core arguments (Goldberg 2006). Next are classifications of a linguistic expression's givenness through *information status* categories. Many information structure notions are rooted in an idea of "givenness"—the extent to which an item is "present" or "salient" in the common ground (Krifka and Musan 2012). Degrees and types of givenness are captured through information status categories wherein, typically, *given* versus *new* form the main polar distinction. Finally, one additional idea appealed to within the study is an informational bifurcation of the utterance. Specifically, I adopt Masia's (2022) framing of *topic* and *focus* as discourse strategies of broad evidentiality. In Masia's model, topics encode information presented as interlocutors' mutual knowledge, while focus encodes information presented as a speaker's individual knowledge.

With regards to theoretical claims, the study considers three; each of which is purported to affect the shape of linguistic forms. The first one, the *one-new-idea constraint*, proposes a limit on the amount of *new* information within an intonation unit. The second one, the *given-before-new principle*, holds that, in regards to the relative ordering of linguistic elements (e.g., clausal arguments) “given” items will precede “new” ones. The third claim is a general idea that there is correspondence between prosodic prominence and information status, wherein “given” items require less phonetic material than “new” ones. Correspondingly, you will find each of these claims figures into one of the study’s three hypotheses (§1.2.3).

Together, these ideas and methods configure the larger theoretical background upon which this investigation of Moklen was done. For this study, it is important to take methodological commitments as constituting a *scaffold*, rather than a subscription to a *framework* (Haspelmath, 2020). My emphasis here is to say that the concepts of intonation units, argument structure, and information status should be seen mainly as tools for building a *language-particular description*. In other words, these notions represent “descriptive theories” rather than “explanatory theories” (Dryer 2006). Explanatory concepts would be those of the *one-new-idea constraint*; the *given-before-new principle*; and a purported *givenness-prominence* relationship—each of which is a component of a hypothesis. In walking this line, I have tried to heed warnings from critiques of information structure research (Matić, 2022; Ozerov, 2018, 2021), and at the same time provide a useful description in terms of basic linguistic theory (Dixon 2009).

3.2 Background and interfaces of information structure

Historically, the field of information structure has roots in the *communicative dynamism* of the Prague School (McElvenny, 2022-present) as well as *information theory* (Shannon, 1948), but coinage of the term and its modern application to linguistics originates with Halliday (1967). A major figure in information structure studies is the linguist Wallace Chafe, who emphasized a connection between the mind and language through the analysis of spontaneous narrative speech (Chafe 1994, 2018). As for significant singular publications, Lambrecht (1994) is a seminal work bringing together Chafe’s and others’ ideas to construct a unification of several information structure concepts, creating a work that is still highly cited.

Information structure theory offers several heuristics for recognizing patterns within language. Many, however, provide overlapping coverage of analysis. Therefore, before embarking on a literature review, it would be useful to distinguish some layers of analysis. To do this let us consider a hypothetical response to the question: *What did Mary eat?* shown in Figure 8.

intonation unit	<i>She ate a can of beans</i>
information units	<i>She ate</i> ^{TOPIC} <i>a can of beans</i> ^{FOCUS}
argument structure	[<i>She</i>] ^A <i>ate</i> ^V [<i>a can of beans</i>] ^O
information statuses	[<i>She</i>] ^{GIVEN} <i>ate</i> [<i>a can of beans</i>] ^{NEW}

Figure 8 Layers of analysis

Each layer represents a distinct analysis. Starting from the top we can note that the intonation unit (IU) took the form of a grammatical clause—a clausal IU—and not a mere phrasal IU, such as an answer: *a can of beans*. Uttered as a clausal IU, we would expect it to be one perceptual prosodic chunk and feature some typical acoustic cues. Next, at the next level of information units (or “informatics”, see Vallduví, 1993), we can see that *She ate* is marked as the *topic* as this is what the utterance is about and represents a shared conceptual grounding of both speaker and hearer. At the same time *a can of beans* is designated as the *focus* as this being added to the common ground, and is information of which the speaker is the epistemic source. Consider an alternative question: *What did Mary do?*; in this case, the topic would just be *She*, as the remaining focus of the utterance (now including the verb *ate*) would constitute an informational update to the common ground and achievement of the speech act. The contrast highlights the usefulness of question-and-answer sequences, because they explicitly outline the way in which common ground content develops (Aissen, 2023)

Third down in Figure 8 is argument structure, which is used to mark the syntactic elements of a clause, namely a verb and its core arguments. You can see that linear argument structure is AVO, where A=AGENT, V=verb, and O=PATIENT; and furthermore, that argument structure is more of a semantic-syntactic concept than a purely information structure one. Finally, in the bottommost layer we see that two referring expressions *she* and *a can of beans* are marked in terms of information status: *given* and *new*. As discussed §3.3, there are several proposals for information status taxonomies, which differ in their classifications of degrees of givenness. For

our purposes here it is enough to note that terms like “given” and “new” are built on a similar intuition about common ground context as topic and focus, so that “given” represents something already in the common ground while “new” is an addition to the common ground knowledge, however they are distinct levels of analysis.

3.3 Information status

3.3.1 Models of givenness

Throughout this study, terms “given” and “new” are repeatedly referred to. “Givenness” is generally taken to represent the degree to which something is present within the *common ground* space between interlocutors (Krifka & Musan, 2012).⁷ Something “given” is already in the common ground; “new” is yet to be established. For example in (3.1), items in small caps could be aligned to “new”, while other items could correspond to “given” (e.g. *there* = ‘PARTY’ and *he* = ‘PAUL’).

(3.1) *I went to a PARTY. PAUL was there. He says HELLO.*

Baumann and Riester (2012) characterize the central aspect of givenness within dynamic discourse as a “dimension of consciousness” or “cognitive activation” of discourse referents and information about them. Chafe (1994) uses the term “activation” to represent the level to which speakers and listeners are actively thinking of a concept, such as a discourse referent. Put another way, Chafe says activation is the extent to which something is “currently lit up” in consciousness. For Chafe, the degree to which a concept is activated—existing in either the focus or periphery of consciousness—represents its level of accessibility for both speakers and hearers. While clearly a psychological concept, significant for linguists is the idea that different levels of marked linguistic “prominence” (Baumann & Cangemi, 2020; von Heusinger & Schumacher, 2019) can be correlated to a spectrum of mental states of activation. The logic here is that the physical effort needed to signal a concept has an *iconic relationship* with an underlying mental state (Lambrecht, 1994), (see also “The Effort Code” (Gussenhoven 2004). In other words, a speaker’s assessment of a

⁷ Stalnaker (2002) frames common ground in terms of presuppositions, “To presuppose something is to take it for granted, or at least to act as if one takes it for granted, as background information – as *common ground* among the participants in the conversation. What is most distinctive about this propositional attitude is that it is a social or public attitude: one presupposes that ϕ only if one presupposes that others presuppose it as well.” (p. 701).

listener's mental state determines signal strength requirements and therefore the use of particular linguistic forms. For example, in (3.1) accentuated lexical forms (*PARTY*, *PAUL*, *HELLO*) would correspond to new "inactive" referents, while deaccented pronominal forms (*there*, *he*) mark "active" given ones.

The term *information status* refers to the data-oriented classification of linguistic expressions in relation to their degree of givenness (Riester and Baumann, 2017). Categories of information status serve as labels by which we can map the relationship between linguistic forms and any supposed *mental states*. Information statuses in Chafe's model consist of the three categories as shown in Figure 9.

ACTIVATION	activated	semi-activated	inactive	
INFORMATION STATUS	<i>given</i>	<i>accessible</i>	<i>new</i>	
PROMINENCE (linguistic form)	less	←—————→		more

Figure 9 Alignment of activation, information status, and linguistic form

With any conception of activation states it is best to think of the differences as being gradient in nature. As Lambrecht (1994, p.100) points out, "From the psychological point of view, there is no theoretical upper limit to the number and kinds of cognitive states which mental representations may have in the course of a conversation." Therefore, for linguists what is ultimately of interest is the extent to which any purported cognitive states (as coded for by information status tags) do indeed correlate with some linguistic form. Information status categories thereby serve as models for capturing a relationship between mental states and linguistic expressions.

Several frameworks of information statuses have been put forth (Röhr, 2016). Development of information status modeling outside of Chafe's ternary model not only reframes the concept of givenness, but also adds multiple intermediate categories between two extremes. Prince's (1981) taxonomy of Assumed Familiarity (see Figure 10) challenged some previous definitions of givenness, but it is still influenced by the Chafean approach, with a core three-way distinction between "evoked", "inferable", and "new" information status categories.

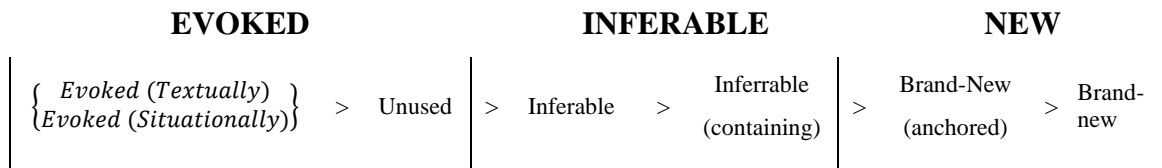


Figure 10 Prince's (1981) Taxonomy of Assumed Familiarity.

One feature of Prince's approach is the inferrables category. This covers discourse entities which hearers could infer via other evoked discourse entities. For example, Prince would say that a referring expression like *the driver* is inferable from *a bus*, because speakers can assume that a hearer's knowledge of 'buses' includes the knowledge that 'buses have drivers'.

In a model put forth by Lambrecht (1994), a distinction between knowledge of referents ("identifiability") and actively thinking about referents in a discourse ("activation") is put forth. Drawing many nuanced distinctions of how speakers potentially conceive of discourse referents, Lambrecht (1994) offers the model of mental representations of discourse referents in Figure 11.

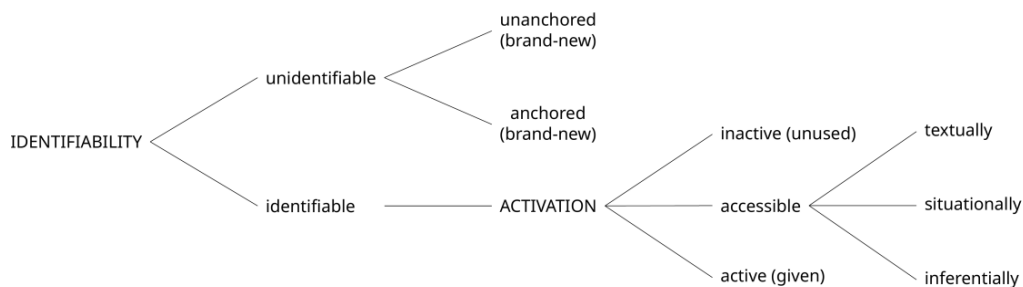


Figure 11 Lambrecht's (1994) information status categories

Models from Chafe, Lambrecht, and Prince all emphasize a conceptual inventory which carves up levels of "givenness", "activation", or "familiarity" at the level of cognition. A different approach to information status reframes the discussion by beginning analysis with the various types of *linguistic forms* used as referring expressions. Two models which take this approach are Ariel's (1988) Accessibility Hierarchy and Gundel et al.'s (1993) Givenness Hierarchy. The motivation for an approach based on linguistic forms can be seen in Ariel's (1988) criticism of Prince:

"The categories of [Prince's] scale... are not linguistic categories. Rather, they are properties of referents which are potential discourse entities. However, for givenness to be a proper linguistic term, accounting for the

distribution of referring expressions, it should be defined according to the way language CODES this scale... I suggest that natural languages primarily provide speakers with means to code accessibility.” (pp.67-68).

While Ariel’s Accessibility Hierarchy is founded on similar intuitions about givenness, the proposed model is populated with specific types of linguistic expressions across a continuous scale, as seen in Figure 12.

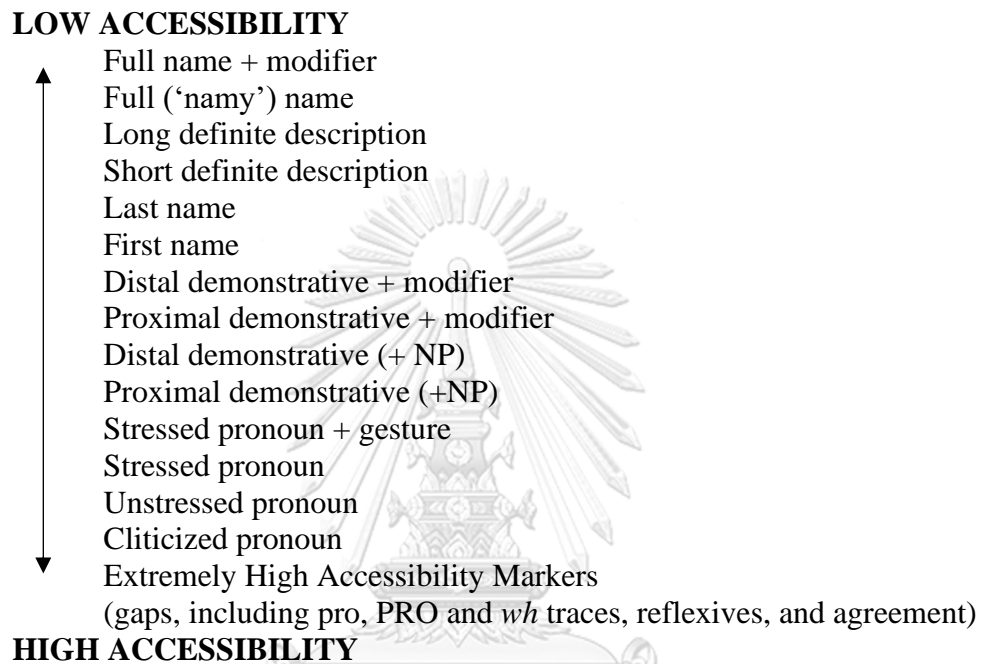


Figure 12 Ariel's (1990) Accessibility Marking Scale.

Types within this scale are based in part on Givón's (1983) cross-linguistic study of topic continuity and gradation of referring expressions. Ariel (1990) proposes that “All languages define the appropriate distribution of their referring expressions based on the cognitive concept of Accessibility” (p.92). Illustrating the full range of “accessibility” in English referring expressions, Ariel provides the following constructed data showing movement from low to high accessibility:

Joan Smith, the president > *Joan Smith* > *The president* > *Smith* > *Joan* >
That/this hat we bought last year > *That hat* > *This hat* > *That* > *This* > **SHE** >
she > *herself* > \emptyset

Another model which targets information status from the perspective of linguistic form is Gundel et al.'s (1993) Givenness Hierarchy. With this model it is also assumed that all natural languages have determiners and pronouns which encode information about the cognitive status of discourse referents. Based in part on a study

of spoken and written data from five languages, Gundel et al. proposes the following hierarchy, which is typically presented using English forms (see Figure 13).

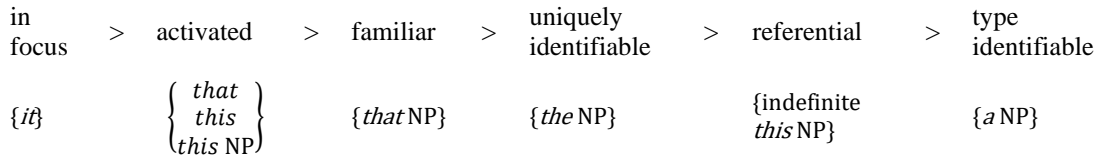


Figure 13 Gundel et al.'s (1993) Givenness Hierarchy

Further applications of the Givenness Hierarchy to more languages led Gundel et al. (2010) to hypothesize that patterns of entailments within the Givenness Hierarchy allow for predictions about the marking of information status in the world's languages. For example, it is hypothesized that all languages explicitly encode "in focus" and "activated", and that "if a language explicitly encodes a distinction between two adjacent statuses on the [Givenness Hierarchy], it will also encode distinctions between higher statuses" (p.1773). And though it is not expected that all languages will conform exactly to this model, the authors believe it outlines strong tendencies.

3.3.2 RefLex Scheme

As shown in the previous section, there are several models and taxonomies of information status classifications within the literature. Despite their differences, at the core of most proposals is the dichotomy between given and new. However, Baumann and Riester (2012) argue that previously notions of givenness have been used inconsistently, and moreover, they contend that previous models are inadequate at capturing various informational distinctions. For example, they present (3.2) and (3.3) as illustrative of the types of patterns annotators have previously struggled with.

(3.2) A: *Did you see Dr. Cremer to get your root canal?*

B: *Don't remind me. I'd like to STRANgle the butcher.*

(3.3) A: *Why do you study Italian?*

B: *I'm MArried to an Italian.*

Baumann and Riester hold that for the underlined elements in these examples two different types of givenness are involved. They in turn propose the RefLex Scheme (Riester and Baumann 2017), a two-dimensional approach to annotating information status. The key feature of this proposal is the classification of information

status at both a *referential* and *lexical* level (*r-level* and *l-level*, respectively). Interpreting the previous examples, they argue that in (3.2) *the butcher* is deaccented as it is referentially given, meaning it is coreferential with the previously mentioned discourse referent *Dr. Cremer*, while for (3.3) *an Italian* is deaccented because it is lexically given. Therefore, within the RefLex Scheme, a “given” element is either a referring expression with a coreferential antecedent, like (3.4), and/or a word for which the same (or similar) lexical expression was previously used as in (3.5).

(3.4) [Riester & Baumann, 2017:6]

I met a man yesterday. [The man]^{r-given} told me a story.

(3.5) [Riester & Baumann, 2017:23]

Look at the funny dog over there! It makes me think of Anna's [dog.]^{l-given}

Conversely, a “new” element is either a discourse referent without a coreferential antecedent, like (3.6) or an unused or unrelated lexical concept, as in (3.7).

(3.6) [Riester & Baumann, 2017:11]

I'm looking for [a friend.]^{r-new} He owes me [money.]^{r-new}

(3.7) [Riester & Baumann, 2017:26]

[Pakistan's]^{l-new} [highest]^{l-new} [court]^{l-new} has [declared]^{l-new} that the country's [prime minister]^{l-new} is [disqualified]^{l-new} from [office.]^{l-new}

Because the scheme looks at both referring and non-referring expressions, lexical classes are covered differently. Nouns, words that can act as referring expressions, can be classified at the referential level and the lexical level. Non-referring expressions, such as verbs, do not pick out discrete discourse referents, are therefore treated only as lexical concepts classified at the lexical level. That verbs and other lexical classes can be included is a crucial and unique aspect of the RefLex Scheme, as most information status annotation systems focus simply on discourse referents via referring expressions. The full breadth of the RefLex Scheme and its tags are shown in Table 5.

Table 5 Annotation tags of the Reflex Scheme (Riester & Baumann 2017).

Referential level	
Tag	Contextual class
<i>r-given-sit</i>	Referents immediately present in the text-external context (no gesture).
<i>r-environment</i>	Referents immediately present in the text-external context (with gesture).
<i>r-given</i>	Coreferential with an antecedent in the previous discourse.
<i>r-given-displaced</i>	Coreferential antecedent of expression occurs earlier than the previous five intonation units.
<i>r-cataphor</i>	Referent is established later on in the text.
<i>r-bridging</i>	Non-coreferential anaphoric expression dependent on previously introduced scenario.
<i>r-bridging contained</i>	Non-coreferential anaphoric expression anchored to an embedded phrase.
<i>r-unused-unknown</i>	Discourse-new identifiable but not generally known.
<i>r-unused-known</i>	Discourse-new but generally known.
<i>r-new</i>	Discourse-new, non-unique referent.
<i>r-expletive</i>	Non-referring expression
<i>r-idiom</i>	
+ <i>genetic</i>	Optional features
+ <i>predicative</i>	
Lexical level	
<i>l-given-same</i>	active, i.e., salient concepts (includes synonyms, hypernyms)
<i>l-given-syn</i>	
<i>l-given-super</i>	
<i>l-given-whole</i>	
<i>l-accessible-sub</i>	semi-active, i.e., derivable concepts (includes hyponyms, meronyms, or recurring stems)
<i>l-accessible-part</i>	
<i>l-accessible-stem</i>	
<i>l-new</i>	inactive concepts, i.e., word is not related to another within the last five intonation units.

3.3.3 Prominence

Deaccentuation has long been put forward as a linguistic correlate of givenness, especially for Germanic languages (Cruttenden, 1993, 2006). Studies featuring the RefLex Scheme support this view by showing how various information status distinctions correlate with differences in the prosodic realization of

lexical/referring expressions in speech.⁸ For example, using a semi-spontaneous dialogue task Röhr et al. (2016) found a stepwise increase in prosodic prominence from *given* to *new* at both the referential and lexical levels, as well as some combined effects. Thies et al. (2018) analyzed data from a picture story description task and found that German accent placement was a decisive prosodic marker of information status, with the distribution of accents showing an increase in prosodic prominence from *given* and *accessible* to *new* referents. Baumann and Schumacher (2020) present neurophysiological results linking perception of German prosodic prominence to information status, showing increased processing power in posterior brain regions was attributable to *new* rather than focused information. Notable in this study, with regards to the RefLex Scheme's two dimensions of givenness was that it was *lexical newness* as opposed to *referential newness* which appeared to be the trigger.

Another notable aspect of givenness is its specific effects on word duration. Basically, the insight from this line of research is that there is a correlation between repetition of a word and a reduction in its duration (Aylett & Turk, 2004; Bard et al., 2000; Bell et al., 2009; Fowler, 1988; Fowler & Housum, 1987; Shields & Balota, 1991; Trón, 2008). And although the phenomenon is often phrased merely in terms of “repetition” or “intelligibility”, in accounting for the reduction researchers usually appeal to the same cognitive mechanisms (e.g., activation) underlying the information status classifications. One example from Shield and Balota (1991) in (3.8) can convey the basic idea. In this study participants produced sentences that included either repetition of the same word (a), a related word (b), or an unrelated word (c). For the repetition condition (a) they found significant effects on duration and amplitude measures, while for relatedness (b) there were effects on duration only. But most interesting is that second mentions of *cat* in (a) and (b) were shorter than *cat* in the unrelated condition (c).

(3.8) [Shield & Balboa 1991:49]

- a. Her cat chases our *cat* under the table.
- b. Her dog chases our *cat* under the table.

⁸ Several similar findings for non-Germanic languages are also available, for example studies on Mandarin (Bi et al., 2016; Ouyang & Kaiser, 2015; Pan et al., 2005) however, these studies use different models of information status and therefore make direct comparison cumbersome.

c. Her son chases our *cat* under the table.

Review of the literature clearly establishes the general effect of repetition leading to a phonetic reduction (Kaland & Himmelmann, 2020). The overall picture, however, is complex, as factors such as overall frequency, informativity, or predictability are also thought to contribute to the effect (Kanwal et al., 2017; Lam & Watson, 2010; Seyfarth, 2014). Overall, the core intuition surrounding reduction of linguistic expressions is that it stems from drivers for efficient communication. Encapsulating this idea, Aylett and Turk (2004) put forth the “Smooth Signal Redundancy Hypothesis”, arguing that “the constraint of producing robust communication while efficiently expending articulatory effort leads to an inverse relationship between language redundancy and duration” (p.31). And with respect to information structure in particular, Aylett and Turk (2004) place “givenness” as a high-level factor in the reduction of prosodic prominence and the shortening of word-form.

Whether it be accent placement, amplitude, or duration, the broadest way of encapsulating the differential linguistic marking of information status is with reference to *prominence*. According to Baumann and Cangemi (2020), “prominence” is one of the most vaguely defined notions in phonetics and phonology research. They, therefore, offer the following minimalist definition: prominence “a unit is prominent when it stands out from its environment” (p.1). Working through its implications, Baumann and Cangemi say this definition does at least two things. First, it offers insight into the relational properties of prominence, highlighting the possibility of prominence to be understood in relative terms as opposed to absolute phonetic terms. Second, it does not specify any particular mechanism by which a unit is made to stand out, thereby opening changes in prominence to a larger number of linguistic phenomena. Typically, investigation into prominence looks for fine-grained phonetic contrasts, i.e., what the salient acoustic cues are. But, with the *minimalist* definition something like the difference between pronouns—which are usually shorter and indicate given information—and full lexical noun phrases could also be framed in terms of relative differences of prominence as they generally contrast in amounts of phonetic material.

The resulting picture, at the nexus of information status and linguistic form, (see Figure 14) is on one axis there is a continuum of cognitive states, with “given” and “new” representing diametric ends of the spectrum. On the other axis, there is the potential for variant linguistic forms to reflect a dimension of givenness through relative differences of prominence, that is more or less phonetic material. For interested linguists, the goal then is to discover any marked alignment of linguistic forms with any supposed underlying cognitive state, of which information statuses are proxies for.

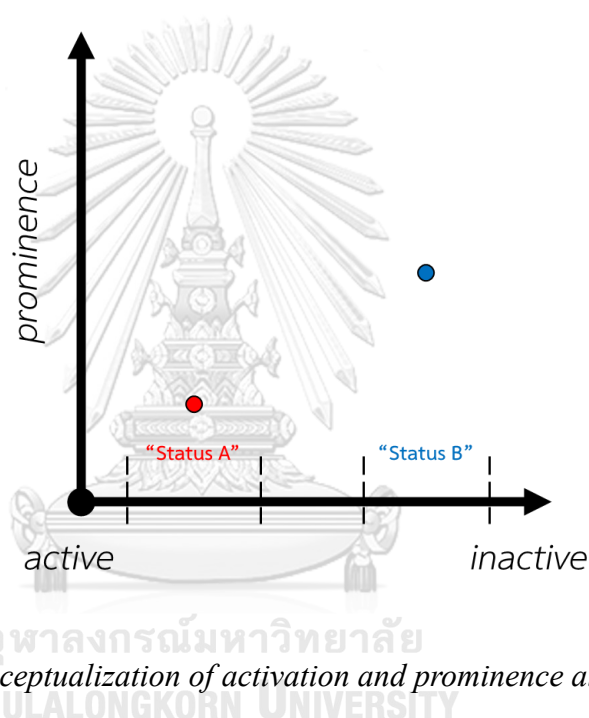


Figure 14 Conceptualization of activation and prominence alignment

3.4 Argument structure

3.4.1 Word order vs. argument structure

Word order (also known as “basic word order” or “basic constituent order”) describes the sequential patterns of major clausal constituents and is a common feature of linguistic description and typology (Tomlin, 1986). Word orders are typically provided as a three-letter acronym, such as SVO, which stands for grammatical terms: subject, verb, and object. Generally, a language’s “word order” is meant as a description of its prototypical declarative transitive clause with two overt arguments. Use of word order, however, is considered by some as a somewhat

informal representation of a clause type, moreover one that is actually infrequent in spoken language (Dryer, 2005). Dixon (2009) proposes the use of *argument structure* for describing clausal syntax. Use of “argument structure” is found in several theoretical frameworks, but it can also be seen as a “pretheoretical cover term” for linguistic phenomena involving the realization of clausal arguments as determined by verbal predicates (Levin, 2018). Use of argument structure as a syntactic frame is also associated with constructionist approaches to syntax (see §3.5.2).

In Dixon’s framing, there are two basic varieties of clauses: transitive and intransitive. Clauses can be described in terms of the order of their core arguments: A, S, O, E (AGENT, SUBJECT, OBJECT, EXTENSION to the core). For example, in (3.9), “extended transitive” clauses (a) and (b) display two different argument structures: AVEO and AVOE, despite having the same general semantics.

- (3.9) a. He^A gave^V them^E a fish^O.
 b. He^A gave^V a fish^O to them^E.

The notion of argument structure relies on the idea that semantic arguments are part of a predicate calculus that all together forms a proposition (Crystal, 2008). For example, in (3.9) the verb ‘to give’ constitutes a predicate, which semantically entails three semantic roles or “arguments”: AGENT, THEME, and RECIPIENT. And as shown with variants (a) and (b), there are two sequential orders in English that make clear who gave what to whom. In other words, these orders reflect grammatically appropriate patterns for a correct semantic interpretation. Argument structure, therefore, is a *semantic-syntactic* notion encapsulating the sequential order of elements of a clause—a predicate and its core arguments.⁹

3.4.2 Preferred argument structure

A discourse preference for an argument structure is an idea at the crossroads of syntax and information structure. And, variation in argument structure (or “word order”) has long been acknowledged to be susceptible to a range of discourse factors (Downing & Noonan, 1995). For example, Bresnan et al. (2007) showed that a hierarchy of information structure factors, such as the givenness of the recipient, influences the choice of dative alternations in English, see (3.9) above. To the extent

⁹ See Mithun & Chafe, 1999 for a critique of this approach.

any argument structure can be connected to a particular discourse context, we can say it is a *preferred argument structure* (Crystal 2008).

With the term “preferred argument structure” we must also acknowledge the research program of the same name *Preferred Argument Structure* (PAS) (Du Bois, 1987, 2003). Of PAS, Du Bois (2003) says it “represents a hypothesis that in spontaneous discourse, certain configurations of arguments are systematically preferred over other grammatically possible alternatives” (p.33). The core component of PAS studies are the four soft constraints shown in Table 6, which are put forth as general tendencies of grammatical clauses when spontaneous language use is investigated systematically.

Table 6 Preferred Argument Structure constraints (Du Bois 2003)

	Grammar	Pragmatics
Quantity	Avoid more than one lexical core argument	Avoid more than one new core argument
Role	Avoid lexical A	Avoid new A

Notable here is the inclusion of *argument realization*, with specifications on the number of lexical arguments. This brings aspects such as pronominalization and ellipsis into the matrix of argument structure considerations. For example, Ratitamkul (2007) reports that for Thai sentences, only a small minority of sentences violated the constraint *avoid more than one lexical argument*. Ratitamkul provides (3.10) as an example of this. But also note that a third discourse referent corresponding to the A argument was ellipsed.¹⁰

(3.10) [Ratitamkul 2007:43]

láew thii níi Ø kᵒᵛ-ləəy pàn càk-ka-yaan pay sa-dùt kᵒᵛn-hĩn

Conn. time De. Ø Conn. peddle bicycle go trip rock

‘Then, this time, (the boy) peddled his bicycle and tripped over a rock.’

¹⁰ Also notable with this example, is its adherence to the *avoid more than one new core argument constraint* (See one-new-idea constraint in §3.5.3). Ostensibly, the *new* argument is the ‘rock’, given the indefinite article in the translation line and that Ratitamkul provides an example of the only one violator of this constraint elsewhere (p. 43). Note also then that the order of overt referring expressions would therefore be in accordance with a given-before-new pattern (See §3.4.3).

The underlying motivation for the tendencies outlined in the four constraints of PAS are explained as arising from cognitive-pragmatics discourse factors, namely the manifestation of speakers' and languages' resources for information management.¹¹ One important point about PAS studies is that the starting point of analysis is the clause. Although not directly opposed to an IU-based approach (as IUs are often clausal), in some PAS studies the lines between syntactic and prosodic units are blurred as researchers begin from grammatical "sentences" or even written data. Moreover, while "new" appears in two of the four constraints, use of different information status taxonomies and models are found across PAS studies, effectively meaning researchers in this paradigm are working with different models of givenness.

3.4.3 Given-before-new principle

Of the possible information structure factors which influence variation in argument structure a major one is the *given-before-new principle* (Neeleman and van de Koot 2015). Reference to a general given-before-new (G>N) pattern is present with early information structure progenitors (e.g., theme and rheme in the Prague school), but experimental testing by Clark and Haviland (1974, 1977) substantiated the hypothesis by showing a psychological preference in discourse for given information to precede new information. In one iteration of their experimental investigation, Clark and Haviland, (1977) presented participants with pairs of sentences like in (3.11), finding that in cases where there was a direct antecedent, like (3.11a), participants displayed shorter comprehension times than they did for (3.11b). Based on findings like these, Clark and Haviland argue that a G>N order manifests itself as a discourse strategy for ordering linguistic elements.

(3.11) [Clark & Haviland 1977:21]

- a. Horace got some beer out of the car. The beer was warm.
- b. Horace got some picnic supplies out of the car. The beer was warm.

Experimental approaches to givenness effects on word order offer some interesting methodological tools. For example, early empirical investigation into

¹¹ Despite general confirmation of PAS tendencies, there is ongoing discussion on the interpretation of these findings, one critique being that the tendencies displayed are a sort of epiphenomenon of broader factors of overrepresentation of human/animate topics rather than pure information packaging pressures (See Everett, 2009; Matter, 2020).

givenness effects on word order in English by Prentice (1967) showed that word order varied as a function of a cue, wherein the cued element or given element was said earlier in the description of an event. In this study, Prentice (1967) used cartoon illustrations of transitive events in an experimental setting. Here, participants were cued beforehand with pictures of either the agent or patient of the transitive event. Results of the study showed that active and passive sentences were more likely elicited in a pattern consistent with the G>N principle.

Similar investigation by Skopeteas and Fanselow (2009) also showed the G>N principle was related to changes in word order across different languages. In Skopeteas and Fanselow's (2009) study, before describing a static picture of a transitive scene, participants also first described a picture of either the agent or patient of the following transitive event. The experiment was conducted with participants from 12 languages to test whether there were givenness effects on word order. While they were able to report on the language-specific strategies for managing the target conditions, overall, it was found that *given* agents induced canonical word order across languages, while *given* patients licensed deviations (e.g., passivization, non-canonical orders). One example from German in (3.12a) shows the agent-given condition elicited canonical word order, but the patient-given condition in (3.12b) elicited a passive sentence.

(3.12) [Skopetas and Fanselow 2009:321]

- (a) [Scene 1] {A boy stands on a carpet...}
 [Scene 2] ...dieser Junge schubst eine grüne Sektflasche um...
 '...this boy pushes a green champagne-bottle.'
 (decoded as AG=SBJ/first; condition AG/GIV)
- (b) [Scene 1] {A girl is running...}
 [Scene 2] ... das Mädchen wird von einem Mann gegriffen und umgeschmissen...
 '...the girl is grasped and knocked down by a man.'
 (decoded as AG=NON-SBJ/non-first; condition PAT/GIV)

Across Skopetas and Fanselow's (2009) cross-linguistic sample, different languages presented various response types and strategies. Generally, the G>N principle can be held as a general strategy within discourse. But, in connecting it to argument structure variation, Skopetas and Fanselow say that givenness-induced

variation depends on the availability of argument positions in syntax. In other words, to whatever extent givenness factors can elicit variant argument structures, these must first be *viable syntactic options* for an expression of information structure within the language. Or as put in Neeleman and van de Koot's (2015) generalization: *If a language uses word order alternation to mark givenness, then in the marked order the given material precedes the new material.*

Other studies into English word order have provided general support for the G>N principle while adding some further nuance. For example, Arnold et al. (2000) showed that both *newness* and *heaviness* (i.e., word length of a constituent) were significantly correlated to later constituents in a sentence, but where *givenness* is equal between constituents, *heaviness* had more of an effect. Clifton and Frazier (2004) looked at constituent ordering in English using grammaticality judgements on different constructions. Here, they found that although there was a preference for definite objects (i.e., *given*) before indefinite objects (i.e., *new*) in double-object sentences, there was a preference for indefinite before definite in NP-PP sentences, thereby showing that certain constructions flout the G > N principle for informational purposes.

In sum, the G>N principle represents one informational aspect often connected to argument structure variation. Some languages may exhibit alternant word orders to mark givenness to differing degrees, while others may not (Skopeteas & Fanselow, 2009). The basic take-away for transitive clauses is that if variations in the order of clausal constituents is used to mark givenness, we can expect the following: A precedes O if A="given" and O="new" and, O precedes A if O="given" and A="new".

3.5 Intonation units

Intonation units are a prosodic reference unit used for segmenting naturally spoken discourse data. Intonation units (IU) connect to an array of functional, grammatical, and informational features. The principles by which individual IUs are delimited within speech are covered more in §3.5.1 But first, it would be useful to sketch out where IUs relate to matters of information structure.

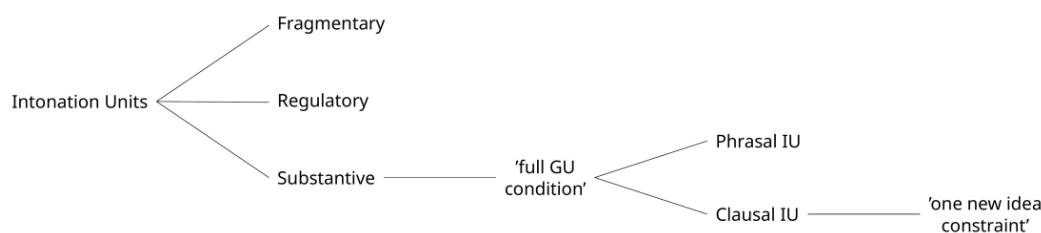


Figure 15 Intonation unit categories and constraints.

Going through Figure 15, we see Chafe's (1994) functional classification of IUs into the three types: *fragmentary* (truncated or unsuccessful speech), *regulatory* (regulating interaction or information flow and discourse markers), and *substantive* (conveying ideas of events, states, or referents). As substantive IUs convey propositional content, they encapsulate the nucleus of information structure research. Further examination of substantive IUs has established there is a general alignment between prosodic and grammatical constituents, such clauses and phrases (Croft, 1995, 2007). A *clausal IU* contains a verbal predicate and its associated core arguments, while *phrasal IUs* contain disjointed syntactic dependencies, such as an isolated noun phrase conveying only a fragment of a proposition. Finally, many researchers have also explored the idea that spoken linguistic units like utterances, sentences, or intonation units seem to be constrained in their amounts of overt information (Chafe, 1994; Du Bois, 2003; Givon, 1975; Halliday, 1967; Matsumoto, 2003). Chafe, specifically for intonation units, posits such an idea with the *one-new-idea constraint* (see §3.5.3). Intonation units, therefore, represent not only a key reference unit in the flow of speech but an interesting means with which to frame information structure.

3.5.1 Basic units of speech

Chafe (1994) holds that if we look at long stretches of speech we will see that they are not characterized by one continuous flow but instead occur in “chunks” and “spurts.” For example, consider the transcription of a sequence from a narrative text in (3.13).

(3.13) [Chafe, 1980:13]

- (a) ..So he takes the whole basket,
- (b) ..and puts it near his bike,
- (c) ..lifts up the bike,
- (d) ..puts the basket on..the front part of his bicycle
- (e) [.5] and rides off.

Each item (a)-(e) corresponds to what Chafe calls an *intonation unit*: a prosodic reference unit of speech. At a methodological level, intonation units are a means by which utterances are segmented into smaller units—the aim of which is to create “a written representation of a speech event so as to make it accessible to discourse research” (Du Bois et al. 1993, p.45).

A first assumption behind approaches and frameworks which heed a call for *basic units of speech* is to first view a linguistic sequence—like a string of words—as stripped of its syntactic and semantic structure (Izre’el, 2020). Something like the IU can then act as a reference unit that forms the foundation of later analyses. For example, Izre’el, (2020) has proposed use of “prosodic module” as an alternative term to the Chafean IU, as intonation is more restricted in scope and module suggests the capacity for these units to work both independently or in combinations. Regardless of whatever label is used, researchers who are interested in tackling this problem are very clear in contrasting their approach to another well-known approach: the Intonational Phonology approach in the Generative tradition, which posits an abstract phonological hierarchy (see Barth-Weingarten, 2016). The foundation for the differences is clear; the Intonational Phonology approach starts off from a theoretical unit rather than a phonetic observation. Furthermore, many of the findings in the generative tradition are based on constructed sample sentences or on the basis of read-aloud made-up sentences. Similarly, Izre’el et al. (2020) says that in the Generative

tradition the departure point of analysis for researchers is small structural domains (e.g., mora, syllable, foot etc.). From this there then is a leap from structure to function. In contrast, functional approaches begin from spontaneous discourse data and identify cohesive communicative units *before* looking to describe their structure.

Identification of intonation units is essentially a task of speech segmentation, but it is at first a perceptual process. However, those tasked with segmenting speech often cite the same set of acoustic cues for delineating intonation units. Mithun (2021), a practitioner of the IU-based approach to speech segmentation provides the list of cues in Table 7.

Table 7 Mithun's intonation unit cues (based on Chafe 1994)

	Coherent intonation contour
Pitch	Initial pitch reset
	Final boundary intonation
Timing	Potential pauses at boundaries
	Possible initial rush
	Possible final lag
Phonation	Possible non-modal phonation
	Possible final creaky voice

Cues within Table 7 are the core prosodic features of intonation units cited throughout the literature for typologically different languages. For example, Mithun reports the same bundle of cues for English (Mithun 2020) and Mohawk (Mithun 2021). In (3.14) Mithun (2020) highlights the variability of pausing between IUs, pointing to both the length of pauses between IUs but also that IUs need not be preceded by pauses (see IU timing).

(3.14) [Mithun 2020:350]

IU duration	Transcription	IU timing
1.38ms	<i>I said this is terribly awkward,</i>	00:15.205 – 00:16.434
	<i>I've just been promoted from,</i>	00:16.434 – 00:17.633
0.05ms	<i>third mate,</i>	00:17.683 – 00:18.204
0.15ms	<i>to sécond mate,</i>	00:18.355 – 00:18.990
0.62ms	<i>and --</i>	00:19.609 – 00:20.185
0.04ms	<i>could we,</i>	00:20.223 – 00:20.526
0.24ms	<i>possibly postpone these orders,</i>	00:20.769 – 00:22.280
	<i>for a little bit.</i>	00:22.280 – 00:22.721

Use of orthographic transcriptions and discourse transcription more broadly would be very familiar in a conversational analysis setting, but speech segmentation and IUs also serve as an important unit of analysis for researchers focused on languages without literary traditions or even a writing system (Adamou et al., 2018). For example, in Figure 16 we see how Mithun (2021) segmented a stretch of Mohawk speech into several IUs (the pitch contours in between dotted lines) many of which display coherent contours as well as variable pausing.

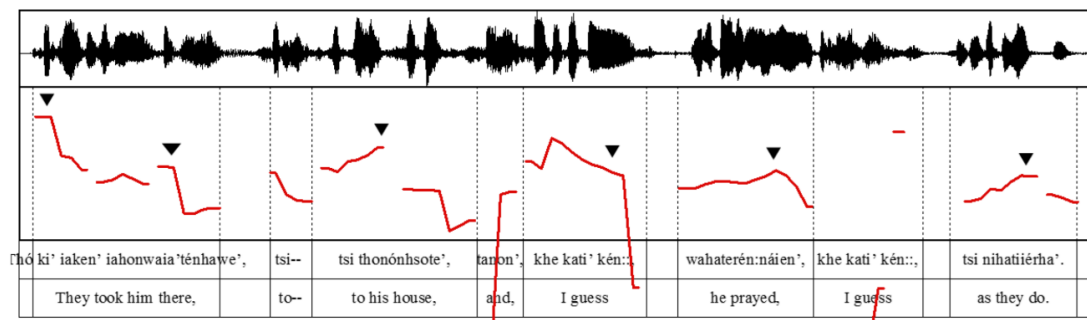


Figure 16 Mithun's (2021) Mohawk IU segmentation.

Researchers from a discourse-functional tradition (Barth-Weingarten, 2016) all commit to the same basic acoustic cues (Table 7), while also admitting of the potential for IUs' acoustic characteristics to vary across language, genre, and even speaker. This variability therefore requires the often repeated acknowledgement that no one cue is “necessary nor sufficient” (Du Bois et al., 1992). Characterizing the general approach to IU-based segmentation Wahl (2015) cites the following recommendations from Du Bois (2008):

[Du Bois] has proposed that a fruitful approach to IU parsability is to think about cues as exhibiting different hypothetical degrees of sufficiency, necessity, and reliability with regard to their capacity to signal the presence of an IU and its boundaries. Certain cues would thereby land higher on one or more of these spectra while other cues would thereby land lower. ...a cue such as lag is theoretically highly sufficient, necessary, and reliable, while a cue such as a turn boundary is sufficient and reliable, though not highly necessary. Moreover, under this view, only a varying subset of possible cues will be present for any given IU. During the task of transcription, then, it is the job of the analyst to identify clusters of cues, and the adequacy of these clusters for the identification of IUs is based on the sufficiency, necessity, and reliability of the individual cues that constitute them. (pp.194-195)

Wahl (2015) then provides the following expanded list of intonation unit cues shown in Table 8.

Table 8 Principal intonation unit cues (Du Bois, 2008)

Cue	Definition
1. lag	tempo lag or prosodic (non-lexical) lengthening
2. rush	rapid tempo unstressed syllables (anacrusis)
3. closure tone	IU-final boundary tone distinguishing intonational finality vs. continuity
4. pitch reset	rise/drop in overall baseline pitch level for IU (esp. on unstressed syllables)
5. pause	noticeable absence of speech by discourse participants
6. creak	creaky voice on final portion of Intonation Unit (not consistent)
7. breath	breathing in (and other vocalisms: exhale, throat-clear, sniff, click, etc.)
8. tune gestalt	coherent intonation contour perceived as unified (holistic) gestalt for the unit
9. isotony	repeated tunes across sequence of Intonation Units (intonational parallelism)
10. turn start	next speaker (new voice) begins
11. turn end	current speaker (current voice) ends
12. accent	IU 'size' in primary accents per IU (tends to be 1, 2, or 0 – in that order)
13. register	overall register shift (of pitch and/or amplitude) for whole Intonation Unit
14. truncation	truncation masks normal end cues, but sometimes is signaled by glottal stop

Succinctly and precisely describing what makes one language sound prosodically different from another is still difficult (Hirst & Di Cristo, 1998). Therefore, searching for actual differences in terms of a hierarchy of cues between different languages reveals only anecdotal reports from various speech segmenters. For example, for Japanese, Matsumoto (2003, p.46) says that the presence of a unified intonation contour was the “single most reliable indicator of an IU boundary”. Or in a small study on 54 Hebrew IUs in Amir et. al. (2004) reports that final lengthening was higher in the hierarchy of cues than anacrusis, pitch reset, and pause. Additionally, Izre’el and Mettouchi (2015) say that many contributors to the *CorpAfroAs Corpus* reported some different hierarchies among acoustic features, such as pitch reset being the most frequent cue in the Tx’amakko and Juba Arabic subcorpora, while in the Moroccan Arabic subcorpus pause was the most frequent cue. Pauses have also been upheld as prominent cues in perception of IU boundaries in Hebrew and Kabyle (Mettouchi et al., 2007), French (Simon & Christodoulides, 2016), and German and Papuan Malay (Riesberg et al., 2018). However, despite these reports of potential cross-linguistic differences in the hierarchy of cues, all researchers, more or less, still acknowledge *the same basic bundle* (Table 7) in segmenting speech.

Despite the methodological uniformity by which researchers use cues for segmentation, the question as to whether or not there are significant cross-linguistic differences for the delimitation of IUs still lingers. However, Himmelmann et al.'s, (2018) Universal Phonetic Intonational Phrase Hypothesis makes a strong claim that tackles the question head on.¹² The Universal Phonetic Intonational Phrase Hypothesis (UPIPH) holds that, “all natural languages make use of the same kinds of phonetic cues for IPs, and that these cues can be perceived by speaker-hearers even in unfamiliar languages” (p.239). This hypothesis is based on significant levels of inter-rater agreement on the segmentation of spontaneous speech of sixty retellings of the Pear Film (Chafe, 1980) in German and three languages from Eastern Indonesia, including languages which were unknown to the German annotators. Adding further

¹² The use of “intonational phrase” here should not be taken as support for the range of phenomena investigated under the rubric of prosodic phonology which proposes an abstract phonological hierarchy. (See Himmelmann & Ladd, 2008, p. 244).

support of the UPIPH, Riesberg et al. (2020) repeated the unfamiliar-language condition of this method to find that Papuan Malay speakers were able to agree on the prosodic boundaries of German tellings of the Pear Film. Speaking directly of the cues which support UPIPH and the implications of their findings, Himmelmann et al. say that “Melodic coherence, pauses, unit-final lengthening and increased unit-initial speaking rate are *universal cues* for intonational phrase boundaries. On the basis of these cues, it is possible to segment narratives in unknown languages with roughly the same reliability as in one’s native language. [emphasis mine]” (p.242).

Not only does the UPIPH refer to the same bundle of cues which are found throughout the discourse-functional tradition (Table 7), but it offers a bold argument against any claim that there are major differences between the acoustic and perceptual cues for the delimitation of IUs in disparate languages. Recent, peripheral support for the UPIPH also comes from interesting neuro-linguistic findings in Inbar et al. (2020). Here it was discovered that there was a uniform rhythm in the temporal structure of IUs in the languages of dramatically different prosodic systems. Inbar et al. (2020) interpret this finding as suggesting that there is a universal structural cue for the cognitive dynamics of speech production and comprehension.

An argument that IUs are universal and can reliably be identified, despite your linguistic background, would offer comfort to those interested in the segmentation of speech into natural chunks. In practice though, IU-based studies rarely, if ever, even seek to strictly define the acoustic properties of the IU itself; most often indicators like pauses or turn-taking are just cited at the top of the hierarchy (Izre’el & Mettouchi, 2015; Mettouchi et al., 2007; Simon & Christodoulides, 2016). Furthermore, nearly all IU-based studies are focused on other objectives, such as language grammars (Kiser, 2014; Yoder, 2020), wherein the main interest is just in accurately presenting the spontaneous and temporal nature of actual spoken texts. In the end, despite some potential variation in their acoustic properties, IUs are mainly just a *practical* means for segmentation of speech into reference units—which later serve as the foundation of additional analysis.

3.5.2 Syntactic properties

With regards to syntax, one idea regarding intonation units, is that for something to be counted as an attested grammatical construction it must be demonstrably evident within IUs in spontaneous speech, and furthermore to some reliable extent in spoken corpora. Essentially, this view holds that beyond sheer physical contiguity of grammatical elements, there are several reasons for basing a language's grammatical units on their relation to prosodic contiguity (Croft, 2005).

For those intonation units that contain propositional content, often they have been noted to take the form of complete grammatical units (GUs). Accounting for the alignment of syntax and prosody, Croft (1995) posits “the full GU condition”, which says that all other things being equal, IUs prefer to be in full grammatical constructions. In Croft's framing, IU-GU mappings represent “stored/precompiled syntactic structures [i.e., constructions] from which more complex structures, usually broken across IUs, are computed in language processing” (p.875). Ultimately, Croft holds that IUs set the prosodic boundaries for grammaticalization, an idea which forms a part of his “Radical Construction Grammar” wherein a language—and therefore its grammar—is composed of entrenched routines within a population of actual utterances (Croft, 2001). In this constructionist framework, syntactic structure is both construction-specific and language-specific. From words to clauses no other representation is needed except for indications of correspondence between syntactic elements and semantic components. Or as put by Goldberg (2006a, p.18) from clausal structure to individual words, “it's constructions all the way down” (see Table 9).

Table 9 The syntax-lexicon continuum (Croft, 2005)

Construction type	Traditional name	Examples
Complex and (mostly) schematic	syntax	[SBJ <i>be</i> -TNS VERB- <i>en</i> by OBL]
Complex and (mostly) substantive	idiom	[<i>kick</i> -TNS <i>the bucket</i>]
Complex but bound	morphology	[NOUN- <i>s</i>], [VERB-TNS]
Atomic and schematic	syntactic category	[DEM], [ADJ]
Atomic and substantive	word/lexicon	[<i>this</i>], [<i>green</i>]

One line of analysis concerning IU-GU mappings looks at the function of clausal and phrasal IUs (Iwasaki, 1993, 1996; Tao, 1992). The basic distinction here is as follows: a *clausal IU* is an intonation unit that contains a verbal predicate and its associated core arguments (either explicitly or implicitly); a *phrasal IU* is an intonation unit that contains a disjointed syntactic dependency, such as an isolated noun phrase which conveys only a fragment of a proposition. For example, Croft (2007) shows how after the clausal IU in (3.15a), a series of phrasal IUs (3.15b)-(3.15d) in the form of NPs serve to describe an already introduced referent.

- (3.15) [Croft 2007:32]
- (a) [1.25] And then three boys happen by.
 - (b) [.65] Three boys,
 - (c) three different sizes,
 - (d) three different coloured shirts,

Iwasaki and Tao (1993) examined clausal and phrasal IUs in small corpora of English, Japanese, and Mandarin, finding that a high proportion of all IUs were clausal, but also that Japanese and Mandarin exhibited more phrasal IUs. Additionally, they identified one salient role of phrasal IUs was *referent establishment*, such as in the first IU of the Mandarin example in (3.16).

- (3.16) [Iwasaki and Tao 1993:8]
- ‘**Bao zhiye** **gaozhong de** ,
 apply vocation school NOM
 ‘those who have applied for vocational schools.’
 hai you hao ^duo
 still have very many
 ‘there are still a lot of them.’

Interpreting such types of syntactic detachment, Tao (1996) says, “there is no syntactic rule that specifies the [prosodic] separation of the argument NP from the predicate, yet this is a common phenomenon in Mandarin discourse. To capture this fact in Mandarin discourse, only the proposal of a dynamic process makes it possible.” (p.182). The distinction between phrasal IUs and clausal IUs, therefore, brings out an aspect overlooked by mere syntactic entailment. Put another way, the

IU-based approach provides a means of highlighting types of prosodic disjuncture across syntactically connected items within speech. Furthermore, there is the potential for specific grammatical markers to also be identified with a particular IU type. For example in Nakagawa's (2020) investigation into spoken Japanese, it was found that grammatically-marked topics also tend to occur in phrasal IUs. Such as in (3.17) where the bolded item is a phrasal IU featuring the topic marker *wa*.

(3.17) [Nakagawa 2020:219]

koo it-ta ... **kaisyuu hoo-hoo-wa** ... mazui-to
 this.way say-PAST collecting method-*wa* wrong-QUOT
 ‘This way of collecting (debt) is wrong...’

An IU-based approach to syntax sheds light on discourse by more clearly outlining the timing and constructions with which speakers organize reference points in discourse. Moreover, by conducting syntactic analysis through IU-GU mappings grammatical description is founded upon spoken utterances within communicative contexts and are not idealized sentences.¹³ The IU-as-reference-unit, therefore, becomes a crucial principle for documentation and description of spoken languages’ syntactic properties, as it lays out a criterion for what qualifies as a syntactically and prosodically integrated grammatical construction (Simard & Schultze-Berndt, 2011).

3.5.3 One-new-idea constraint

While the prosodic aspect of “intonation” units gives them their name, that the terminological predecessor was Halliday's (1967) “information unit” is not insignificant. Chafe (1980) in fact, first used the term “idea unit” before later using the single intonation contour as the main descriptive feature (Chafe, 1984). And while IUs’ prosodic and syntactic features are of interest, key for Chafe (1994) was how IUs outlined the flow of information within discourse and therefore offered potential insights into the relationship between the mind and language.

Chafe (1994) believes each IU can be taken as a verbalization of a speaker’s singular focusing of consciousness—that is each IU seems to have *one* aspect directly

¹³ Linell (1982), with regards to the *written language bias in linguistics* says, that “Historically, there are two categories of linguistic units that have been considered much more important than others, i.e., words and sentences” (p. 63).

in focus, while others are in the periphery. Surrounding Chafe's understanding here is a range of psychological speculation, analogies to the visual system, and an earnest appeal to introspection of subjective conscious experience. Chafe, however, was not the first to put forward such ideas. Prior to Chafe, Halliday (1967) sketched out a picture of a "one information unit to one tone group" relationship, while Givon (1975) characterized a "one bit of new information per proposition" strategy. In a similar vein, Chafe formulated the *one-new-idea constraint* to characterize the informational brevity or low-information load of most IUs. The broader picture for discourse is that it consists largely of brief incremental communicative moves in small structures lasting 1-2 seconds (Chafe, 2018).

The one-new-idea constraint, basically, is a general observation that within the flow of naturally occurring spoken discourse, IUs proceed in a manner whereby there are few cases in which an IU has two or more "separately activated *new* ideas". Chafe goes on to say that since grammatical subjects are normally "given", the predicate is usually the locus of new information. In the case where grammatical objects are "given" to some extent, well then it is the relation between all clausal constituents which serves as the "one new idea". However, as there can often be "new" grammatical objects, Chafe says these will likely occur with either predicative low-content verbs (e.g., *have, get, give, do, take, use, say* etc.) or some lexicalized phrase, arguing that many verb-object pairs should be viewed as functionally singular constructions which do not exact separate activation costs.

Chafe (2007) admits that "the validity of the hypothesis requires careful specification of the terms *one, new, and idea*"(p.139), however several researchers have placed the one-new-idea constraint as having explanatory power for utterances within discourse. For example, Degand and Simon (2008) cite the one-new-idea constraint, as one reason for explaining a dislocation strategy in spoken French, like in (3.18) where they say two IUs were separated as they each contained a unit of new information.

(3.18) [Degand and Simon 2008:38]

et il a conservé ce billet/ dans la doublure de son vêtement/

'and he has kept this note/ in the lining of his cloth/'

Here, though, and in other studies (Kiser, 2014; Tao, 1996) appeals to the one-new-idea constraint (or critiques of, see Ross et al., 2016) often seem to merely rely on an *impressionistic* interpretation of the informational contribution each IU makes. Essentially, this means that the one-new-idea constraint is never usually examined rigorously. One exception is Matsumoto's (2003) study of Japanese IUs. Crucially, Matsumoto's methodology included information status tagging for all noun phrases within a corpus of speech. From this, Matsumoto operationalized the one-new-idea constraint as a measurement of the number of new NPs within an IU. Ultimately, it was found that speakers avoided introducing more than one new NP per IU, like in (3.19), with *new* NPs bolded. Findings from this study led Matsumoto to put forward a "one new NP per IU constraint" as a basically equivalent operationalization of the one-new-idea constraint.

(3.19) [Matsumoto 2003:122]

(3.19.1) **shusseki** **ritsu** takai mono

attendance percentage high because

'Cause the percentage of (my class) attendance (is) high.'

(3.19.2) **samaa** **sesshon** totta no yo

summer session take-PAST FP FP

'(I) took summer session (courses).'

In most research, where the one-new-idea constraint is cited it was not a main emphasis of study (Degand & Simon, 2008; Kiser, 2014; Tao, 1996). However, the flow of information across discourse is at the heart of information structure research. And, notable for intonation units (or other strands of linguistic elements) is that they have been seen as aligning with a specific quantity of information, namely "one". There have been some forays into the underlying cognitive basis for the one-new-idea constraint looking at levels of the brain activity (Vallauri & Masia, 2018) and psycholinguistic testing of short-term memory (Simpson 2016), but this type of research goes beyond mere linguistic analysis. For further research with regards to linguistic description, Matsumoto's (2003) operationalization provides one means by which to assess the one-new-idea constraint. Basically, in this framing, the one-new-idea constraint predicts clausal IUs will contain no more than one new *referring*

expression—noun phrases which pick up an identifiable entity as a discourse referent (Baumann & Riester, 2012; Gundel et al., 1993).

3.6 Topic and focus

A major notion within the field of information structure is the idea that utterances typically consist of two informational units: one based upon common ground and one that updates it (Vallduví & Engdahl, 2013). A whole host of terminology has been put forth regarding this dichotomy, e.g., presupposition-assertion, background-foreground, theme-rheme, topic-comment, but now these units are being grounded in cross-linguistic analysis of corpora of spontaneous speech. Working with such data, Cresti's (2018) illocutionary model of information structure not only puts forth prosodic criteria for speech segmentation, but it also aligns the classic informational bifurcation of utterances within the pragmatic framework of speech act theory. Briefly, the view starts with the idea that a core part of utterances is their *illocutionary force*—the effect of the speech act intended by the speaker. Utterances may also have a *topic* that serves as the conceptual domain or field of application for the illocutionary force. Based on Cresti's model, Masia (2022) has argued for a recharacterization of the informational units of topic and focus as discourse strategies of broad evidentiality, with the following definitions:

Focus encodes information conveyed by the speaker as her communicative intention and as individual knowledge of which she is the only epistemic source.

Topic encodes information not conveyed as the speaker's communicative intention and which represents mutual knowledge established as shared conceptual grounding with both speaker and hearer as committed source.

Topic and focus then are not just holders of contents but a reflection of how speakers present these contents. And so, while the thrust of information status marking is to reveal a speaker's cognitive commitment to a discourse referent, the topic-focus distinction targets the degree to which a speaker is manifesting this within discourse.

Generally, topics are seen as correlating with common ground content, but they are also associated with what an utterance is “about” and even the grammatical

category of “subjects”. However, for Cresti and Moneglia (2018) semantic and grammatical correlates of topics are a sort of epiphenomena, and instead they emphasize the pragmatic nature of topics as providing an addressee adequate reference for the speech act the speaker is about to accomplish, i.e., the illocutionary force of the utterance which is accomplished through the focus, like in (3.20).

(3.20) [Masia 2022:74]

- a. [*These shirts*]_{TOPIC}, [*don't take them*]_{FOCUS!} ORDER
- b. [*These shirts*]_{TOPIC}, [*can you take them*]_{FOCUS?} QUESTION
- c. [*These shirts*]_{TOPIC}, [*I'll take them*]_{FOCUS} ASSERTION
- d. [*These shirts*]_{TOPIC}, [*feel free to take them*]_{FOCUS!} EXCLAMATION

Previously, informational units of topic and focus¹⁴ (or any other of the dichotomies) would be equated to language specific features at phonological or morpho-syntactic levels (e.g., grammatical position, topic markers, or prosodic characteristics). While such language-specific features are expected and can still be acknowledged, in this view topic and focus are ultimately determined with regards to a speaker's *communicative aim*.

In Masia's view, the point is that information contained within topics may in fact be “given” or “new” in strict *information status* terms, however, with information presented as a topic, speakers will be providing “linguistic clothing” which discursively commits the listener to its truth—that is its place as part of common ground knowledge. Conversely, information in focus, whether “new” or “given”, will be conveyed as deriving solely from the knowledge base of the speaker. With this framing, the topic-focus distinction, in addition to information status, further underscores how information packaging processes are a complicated multifactorial process of categories and strategies (Ozerov, 2021). The point is that despite sometimes being conflated with information status, topic/focus represent a separate level of analysis, as they rely on interpreting a speaker's communicative intention towards an addressee, an aspect which is not revealed simply by connecting anaphoric links and counting uses of a lexeme.

¹⁴ Like other information structure terms, “focus” is also terminologically troublesome. According to Vallduví (2016) it is “one of the most (*ab*)used labels in information structure research” [parenthetical indictment in the original, emphasis mine].

3.7 Summary

The field of information structure looks at how the packaging or management of linguistic forms encodes the exchange of information between interlocutors. The study of information structure covers several overlapping domains and notions. This study adopts intonation units, argument structure, and information status as methodological commitments.

Intonation units are a reference unit of analysis when segmenting stretches of discourse data. The one-new-idea constraint holds that within intonation units there will be few cases in which an intonation unit has two or more separately activated new ideas. Based on Matusomoto (2003), the one-new-idea constraint can be operationalized as a constraint on “new” referring expressions.

Argument structure is a syntactic schema for elements of a clause: verbs and core arguments. One factor thought to impact variation in argument structure is the given-before-new principle, which holds that there are psychological processes which prefer given information to precede new information. Previous research has shown that if argument structure variants are used to mark givenness, then a given-before-new order will be the preferred structure of clausal arguments.

Givenness represents the degree to which something is present within the common ground space between interlocutors. The term information status refers to the data-oriented classification of referring and non-referring expressions in relation to their degree of givenness. The *RefLex Scheme* (Riester & Baumann, 2017) is a recent proposal of an information status taxonomy that looks simultaneously at both *referential* and *lexical* givenness. The broadest way of encapsulating the differential linguistic marking of information status is with reference to linguistic forms’ prominence, generally meaning relative differences in phonetic material. In the information structure literature, there is a general acknowledgment of an alignment of given elements to require less prominence and therefore result in forms with less phonetic material.

Lastly, there is also a level of analysis of at an utterance’s information units, which is based on inference about the management of common ground content. Of this distinction, Masia’s (2022) puts forth topic and focus as broad strategies of

evidentiality within discourse. Despite covering, a similar conceptual space as information status, positing topic and focus represents a distinct level of analysis as the informational bifurcation of utterances relies on interpreting a speaker's communicative intention, while information status is based on a narrower annotation of individual linguistic expressions.



4 Methodology

4.1 Introduction.

Studying a language in the field presents conditions and issues different from traditional experimental settings. First, it not only requires the researcher to manage their own time and resources, but it also requires the researcher to consider the needs of the host community. A study of information structure, moreover, is especially challenging, as it requires an intimate understanding of discourse contexts and speakers' communicative intentions. The foundation for this study are the collaborative sessions of language elicitation between the researcher and Moklen speakers using custom stimuli designed for investigation of information structure.

Previous research on Moklen (Larish, 1999; Swastham, 1982) was based largely on translation-based direct elicitation. Therefore, there were remaining questions about the discourse contexts in which variations of argument structure and word-form occurred. As the general hypothesis was that information structure factors might account for some of the reported variation, I decided to collect data through *staged communicative events*—communicative events enacted for the purpose of recording (Himmelman, 1998). The catalyst for these staged communicative events would be custom-designed picture-based stimuli, which would offer a semi-controlled discourse context. The core data for this study therefore consist of one text type: *picture descriptions*. While this is a limitation, it is the framing of Moklen speech within these contexts that facilitated inferences about speakers' communicative intentions and in turn offered insight into the use of particular linguistic structures.

Two custom-designed stimuli were created for pursuit of the research objectives. However, design of each stimulus also had to balance the following needs: (1) portrayal of culturally salient content¹⁵ (2) feasible implementation in the field, and (3) capable of generating the relevant data. The first stimulus is the *Stolen Fish Picture Book* (§4.3). The second stimulus is the *Transitive Event Picture Sequences*

¹⁵ My use of 'culturally salient' is not intended to suggest materials here will exemplify Moklen culture or even transmit any particular Moklen themes or notions. The main sense here is to signify that these materials will aim to avoid domains which might induce lexical material from the language of the dominant culture (i.e., Thai).

(§4.4). Data elicited with the Stolen Fish Picture Book went into a corpus from which clausal intonation units' adherence to the one-new-idea constraint was assessed (Hypothesis 1). Data from the Transitive Event Picture Sequences was used to investigate givenness effects on argument structure (Hypothesis 2). Additionally, the corpus of Stolen Fish texts was also consulted to further explore argument structure variations and adherence to the given-before-new principle. Data from the Transitive Event Picture Sequences was used to assess the extent to which monosyllabic alternants corresponded to "given" information statuses (Hypothesis 3). Further contextualization of word-form shifts was also provided from data from the Stolen Fish corpus and other sources.

4.2 Data collection

4.2.1 Participants

Working with Moklen language consultants, primary data was collected by the researcher from January to April 2022. This was preceded by periods of preliminary collaborative language work with Moklen speakers from 2019-2021. Having established working relationships with Moklen speakers and community leaders, many collaborators were already familiar with the author's interest in the Moklen language. All participants were Moklen speakers in communities in Phang Nga and Phuket provinces. In total 24 Moklen speakers (11 female 13 male) from 13 different communities participated in sessions for one or both tasks that made it into the final analysis. The ages of speakers ranged from 43-77, with an approximate average of 60. I would also like to acknowledge that many other Moklen speakers also took time to participate in these tasks and to discuss and review matters of language concerning them.

The tasks of the study were conducted as sessions of language elicitation between the researcher and a Moklen language consultant. Names or personal information of participants are not included in any published materials. Where a participant's information (e.g., age, gender, location) is presented, two-letter coded tags are used. Research approval for this study was obtained through the project IRB (COA no. 220/2564). Documentary language data from this project will also be

deposited in a Moklen Language Documentary Corpus¹⁶ under the auspices of the research project “Research and Documentation of the Moklen Language and Culture in the Southeast Asian Context”, led by principal investigator Associate Professor Pittayawat Pittyaporn, Department of Linguistics, Chulalongkorn University, Bangkok, Thailand.

4.2.2 Session recording

Implementation of the stimuli and session recording was done by the researcher. Every session occurred in the participants’ respective community usually at or nearby the participant’s residence. Participants provided verbal consent to participating in the task and for the sessions to be recorded with both audio and video devices. Audio was recorded with a ZOOM H1 Handy Recorder outfitted with a foam mic wind cover at sampling rate of 44.1 kHz/s (16-bit wav. file format). Video was recorded with a GoPro Hero 7 in an mp4 file format. The typical set up and orientation of the participant, the stimuli, the recording devices, and the researcher during language consultation sessions is shown in Figure 17.



Figure 17 Stolen Fish and TEPS sessions (Photographs by Athikhom Saengchai).

¹⁶ <https://www.arts.chula.ac.th/moklen/>

4.3 Stimulus 1: Stolen Fish Picture Book

4.3.1 Design

The *Stolen Fish* picture book was designed for the elicitation of Moklen narrative discourse data. The main purpose was to investigate the syntactic and informational properties of Moklen intonation units (Objective 1), but it was also conceived as a means to also accomplish some documentary aims. *Stolen Fish* was written and designed by the researcher. Illustrations were done by Paul Hoch Myers. Inspiration for the story and the stimulus comes directly from use of “Frog Stories” (Berman et al., 1994; Mayer, 1969; Strömquist & Verhoeven, 2004) and the *Pearl Film* (Chafe, 1980) for language research (see also San Roque et al., 2012).

The *Stolen Fish* picture book presents participants with a single story containing multiple interactions of various characters. This story is presented through 33 illustrations depicting a relatively simple yet engaging plotline. Taking place in a culturally salient setting, relatable characters enter and exit the storyline in a manner calculated to contrast their information status. As participants narrate the story, they are presented with *new* and *given* referents, along with the reappearance of characters after a period of “non-activation”. The basic outline of the story as intended by the author is as follows:

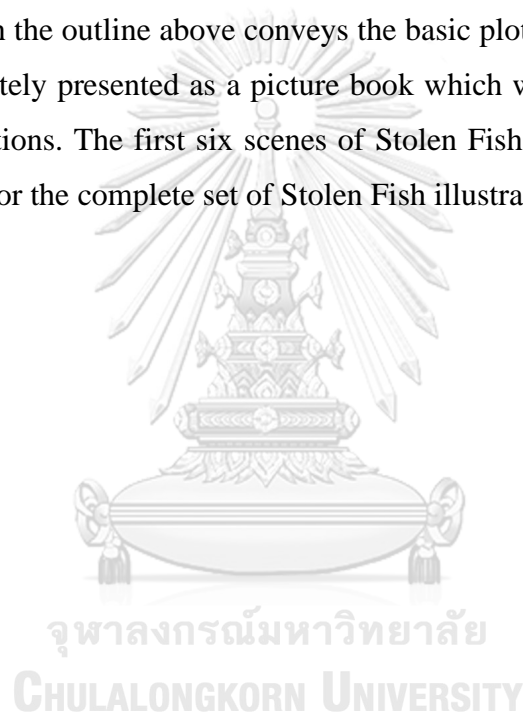
The setting is the seaside. After failing to catch any fish, a young boy (the thief) decides to steal fish from a nearby fisherman. After taking the fish he encounters two girls picnicking who tell him that he has dropped a fish. Then, we see three boys on a path who encounter the thief as he trips and drops all his fish. The boys help him up, and the thief gives them two fish. As the thief and the boys begin to go their separate ways, one of the boys gives a fish to a nearby bird. The three boys then go to the site of the two girls, with one of the stolen fish in hand. The entire group then proceeds to cook the fish. The thief is then depicted returning to his original fishing spot with one fish in hand.

The story then returns to the fisherman who now realizes that all his fish are gone. Coincidentally, he encounters the bird who received the one fish from the boys. The fisherman assumes the bird has stolen the fish and chases it. Chasing the fleeing bird, the fisherman is led to the location of the thief. The thief, seeing the fisherman approaching, attempts to avoid suspicion by hooking the stolen fish onto his fishing line. When the fisherman comes near, the thief presents the fish as if he had caught it on his own. The fisherman perhaps suspects something but leaves empty handed. As the fisherman is walking, he encounters

the picnicking girls and boys. They invite him to sit and join them as they begin their feast. The fisherman and the group sit down together happily, and the fisherman is depicted gesturing, apparently telling the tale of having his fish stolen.

The final sequence of the story shows the thief who has now returned to his fishing spot, and gotten away with theft, placing the last of the stolen fish in a basket. However, while he is not looking a bird comes from behind and steals the fish. In the final scene the boy is fishing, line in the water, unaware that his stolen fish is now gone.

From this overview you can see that the Stolen Fish picture book presents many events and interactions to narrate and that all characters appear and reappear at least once. Though the outline above conveys the basic plot as intended by the author, the story is ultimately presented as a picture book which was expected to be open to varying interpretations. The first six scenes of Stolen Fish are provided in Figure 18 (see Appendix C for the complete set of Stolen Fish illustrations).



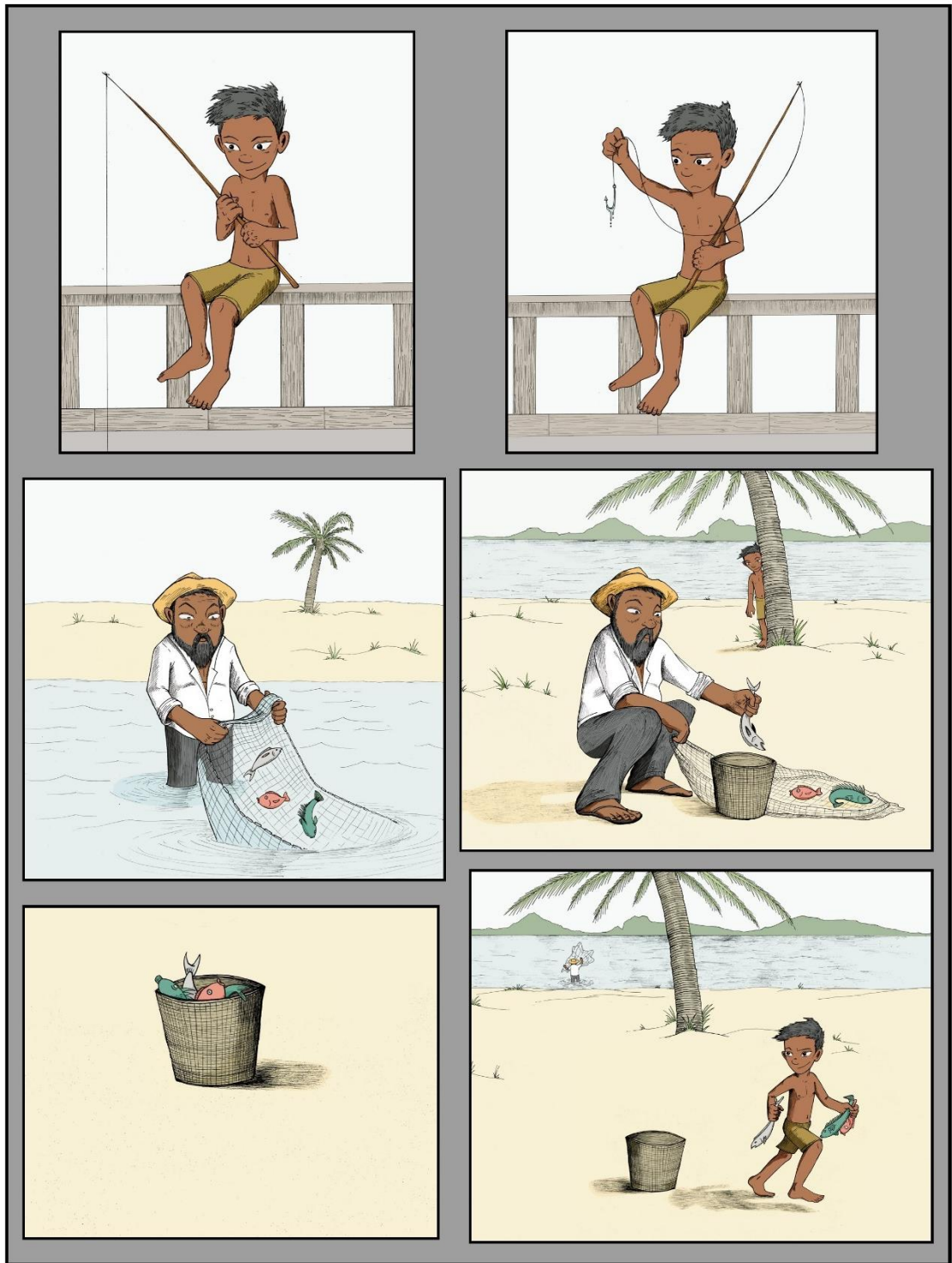


Figure 18 First six illustrations of Stolen Fish

The Stolen Fish picture book has a total of 33 illustrations. A title and basic description for each of the 33 illustrations is provided in Table 10.

Table 10 Titles and descriptions of Stolen Fish illustrations

#	Title	Description
1	Line in water	A boy (the thief) sitting on a bridge fishing.
2	No luck	The thief pulls up fishing line. He has not caught any fish.
3	Today's catch	A fisherman is at the seashore. He has caught three fish in a net.
4	I spy	The fisherman is putting his fish into the basket. The thief is secretly watching the fisherman.
5	Full basket	A close-up of all 3 fish in the basket.
6	The theft	The fisherman goes back out to sea with the net. The boy has stolen the fish.
7	Picnic spot	Two girls are sitting on a mat by a tree.
8	Hey you!	The girls are sitting and looking as the thief runs by and drops a fish.
9	Returning the fish	One girl has picked up the fish and is returning it to the thief.
10	Three boys	Three boys are walking down a path.
11	Down the path	The thief is running down the path.
12	Have a nice trip	The thief trips over a rock, falling and dropping the fish in front of the three boys.
13	Let's help	The boys help the thief pick his fish up.
14	Sharing is caring	The thief gives two of the fish to the boys.
15	Feed the birds	The boys have two fish, but one is given to a nearby bird.
16	Look what we got	The boys go to the site of the two girls and show their one fish to them.
17	Cookout	Everyone sits down while the fish is being cooked over fire.
18	Back at the bridge	The thief has returned to the bridge with one fish.
19	Where's my fish?	The fisherman returns to his basket and notices his fish are gone.
20	Pesky birds	The fisherman suspects a nearby bird has eaten his fish.
21	Bird flees	The bird flies away, fish in mouth, while the fisherman chases it along the path.
22	Look who's coming	The fisherman has been led to the bridge where the thief is located.
23	Trickery	The thief takes the stolen fish and hooks it onto his fishing line.
24	Nothing to see here	The fisherman is next to the boy looking at him and "his fish" suspiciously.
25	Poor fisherman	The fisherman walks back the way he came, appearing unhappy.
26	Join us	The fisherman encounters a group picnicking. They wave and invite him over.

27	Eating together	The boys, the girls, and the fisherman are sitting and talking together after their meal.
28	Off the hook	The thief is removing the stolen fish from his fishing line.
29	Stolen goods	A close-up of the stolen fish in a basket.
30	Return of the bird	The thief is looking away, fishing, while a bird is landing nearby.
31	Bird creeping	The bird comes near the basket with the fish.
32	Bye, bye bird	The bird flies away with the stolen fish in his mouth.
33	Just deserts	The thief is standing, line in the water, unaware that the last fish has been stolen.

The *Stolen Fish* picture book was printed as a single book containing 33 pages. The pictures were printed in color in a landscape orientation onto a single side of A3 (297 x 420 mm) 260gsm paper. Figure 19 is a photograph of a *Stolen Fish* illustration as presented within the bound picture book.

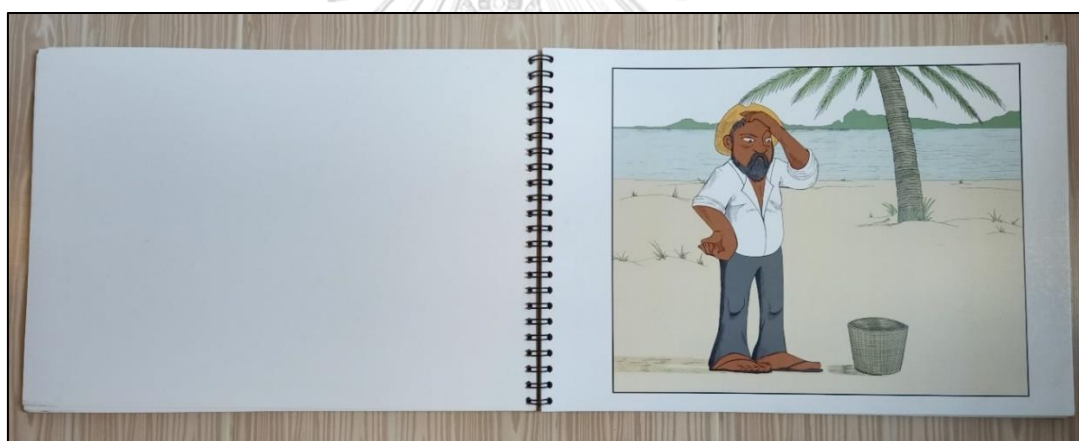


Figure 19 Photograph of the *Stolen Fish* picture book

4.3.2 Procedure

The *Stolen Fish* picture book was used to elicit a brief narrative text. It was presented by the researcher/interviewer as a picture book that was to be freely narrated and described by participants. Before providing a *Stolen Fish* narration, participants could become acquainted with picture-book narrations through sessions with other similar materials e.g., *Frog, where are you?* (Mayer, 1969) or *The jackal and crow picture task* (Carroll et al., 2011). For this study before providing their narration, participants were given the opportunity to preview the entire book alongside the interviewer (see Figure 17). This initial inspection was led by the

participant, but the interviewer was free to draw attention to certain parts, for example asking “*What is that called?*” while pointing to the rock in scene #12. After the participant viewed all the illustrations, they were then asked to narrate the story from beginning to end. Participants were allowed to handle the book, but the interviewer could also act as a page turner. The aim of the sessions were for the elicitation of monologic texts from single participants. Some sessions, however, occurred in the company of an active Moklen audience, who at times responded to the narration and perhaps even interjected from time to time. Once the participant completed the narration the task was completed.

4.3.3 Data

A total of 16 Stolen Fish texts were included in the Moklen *Stolen Fish* corpus, as shown in Table 11. Texts were collected from 8 females and 8 males from 9 different communities (age range=43-77 approximate average=59). The average Stolen Fish text runs about 5 minutes and has an average of 127 intonation units. The total running time of the Stolen Fish corpus is 1 hour 26 minutes and is composed of 2,033 intonation units.

Table 11 Stolen Fish sessions selected for analysis

Contributor	Gender	Community	Length	# of IUs
TG	F	Tap Tawan	0:06:33	169
NK	M	Tha Yai	0:05:03	121
UN	F	Thap Plaa	0:05:15	140
YN	F	Thap Tawan	0:06:18	112
NA	F	Bon Rai	0:04:18	107
YG	M	Thap Plaa	0:05:03	101
CN	F	Laem La	0:04:00	120
NG	M	Laem La	0:08:29	234
NJ	M	Bon Rai	0:06:03	118
DW	F	Bon Rai	0:04:48	93
GP	F	Baan Thung Wa	0:04:58	84
WN	F	Baan Thung Wa	0:05:35	136
KG	M	Bang Khaya	0:05:22	117

LN	M	Thap Tawan	0:05:50	131
YT	M	Lam Pi	0:04:28	97
HJ	M	Tha Pae Yoj	0:04:42	153
Total			1:26:45	2,033

4.3.4 Annotation

Stolen Fish texts were used for study of the syntactic and informational properties of Moklen intonation units (Objective 1). Intonation units were annotated for their communicative type (substantive, regulatory, fragmentary), grammatical class (phrasal or clausal), transitivity of main predicate, along with additional coding for construction type and the number of overt referring expressions. Referring expressions also received annotations for their information status, with tags based on the *RefLex Scheme* (Riester & Baumann, 2017). The referential distance of referential/lexical antecedents was also noted (see §4.4.4 for a more detailed account of the annotation of information status within this study). One factor was the appearance of co-referring expressions (e.g. pronominal clitics and classifier phrases). For annotation of core arguments, lexical arguments were prioritized. Co-referring expressions were only counted as an additional independent referring expression if they were the only overt information for an argument within a clausal IU. Figure 20 is a screenshot of tiers and annotations of a Stolen Fish text within ELAN (*ELAN (Version 6.4) [Computer Software], 2022*).

The screenshot displays the ELAN software interface with the following components:

- Video Window:** Shows a scene with two people in a boat.
- Timeline:** Shows a selection from 00:05:15.836 to 00:05:19.678 (1434 ms).
- Annotation Grid:**

Tier	Annotation 1 (148)	Annotation 2 (149)	Annotation 3 (150)	Annotation 4 (151)
PIC# (p1)	#22 5:16			
NG (p24)	lek: nanj: n: din	ticum nrj: din	pat'a? sa?p'a: n t'i: na ? 7ana: t	do: k neman
IU (p24)	S	S	S	S
Phrasal				
Clausal	Vt-Vi	S-Vi	Vt-O	RC
TRSV	Vt	Vi	Vt	Vt
CL type	B-Vt	1 S-Vi	1 Vt-O	B-Vt
R-EXP		ticum nrj:	sa?p'a: n t'i: na ? 7ana: t	
ARG (p24)		S	O	
REF (p27)		[r-given	[r-given-displaced	
LEX (p29)		[r-given	[r-new	
RD (p34)		3/3	149/148	
MSE (p1)				
# (p24)	148	149	150	151

Figure 20 Screenshot of Stolen Fish annotation in ELAN

4.3.5 Analysis

Annotations from ELAN were tabulated in EXCEL. There, intonation units (IUs) could be organized by type, grammatical class, and their number of referring expressions. After functional classification, substantive IUs were then classified by grammatical type: phrasal or clausal. Clausal IUs are intonation units containing a verbal predicate and its associated arguments (Croft, 1995). In total there were 1,517 clausal IUs (see Table 12).

Table 12 Stolen Fish corpus intonation unit types

Functional Type	<i>n</i>	Grammatical Type
Substantive	1,517	Clausal
	348	Phrasal
Regulatory	134	
Fragmentary	34	
Total	2,033	

To assess whether clausal intonation units conformed to the one-new-idea constraint (Hypothesis 1), the extent to which all 1,517 clausal IUs had no more than one “new” referring expression was examined, essentially the same operationalization as Matsumoto (2003) (see §3.5.3). In this analysis “new” corresponds to the *r-new* tag of the RefLex Scheme i.e. a discourse-new referent. Findings of this analysis are presented as descriptive statistics, along with a description of syntactic and informational properties of Moklen intonation units (Objective 1).

4.4 Stimulus 2: Transitive Event Picture Sequences

4.4.1 Design

The Transitive Event Picture Sequences stimulus was designed for the elicitation of more narrowly framed speech data. One purpose was to test givenness effects on argument structure (Objective 1), but it was mainly conceived as a novel means to investigate the relationship between changes in information status and Moklen word-form (Objective 2). Initial inspiration came from a task in Skopeteas et al. (2006) and implementation in Skopeteas and Fanselow (2009) for investigation

into givenness effects on word order (see §3.4.3). Design of the stimulus was done by the researcher. Illustrations were done by Nittaya Intapong.

The Transitive Event Picture Sequences (TEPS) stimulus presents participants with the task of describing a series of 24 transitive events depicted through three-picture sequences. Participants describe each picture individually, as the sequence unfolds through three scenes: a *context scene*, a *target scene*, and a *resolution scene*. The first scene, the context scene, serves as a cue by presenting one of two possibilities, either the *agent* or the *patient* of a following transitive event. Next, the target scene presents a depiction of both referents engaged in the transitive event. Finally, the resolution scene depicts both the designated agent and patient in a context that implies the completion of the transitive event. Each of 24 sequences has two versions, an *agent-initial* one and a *patient-initial* one. The full set of illustrations for three sequences of TEPS are shown in Figure 21 (see Appendix D for the full TEPS instrument).

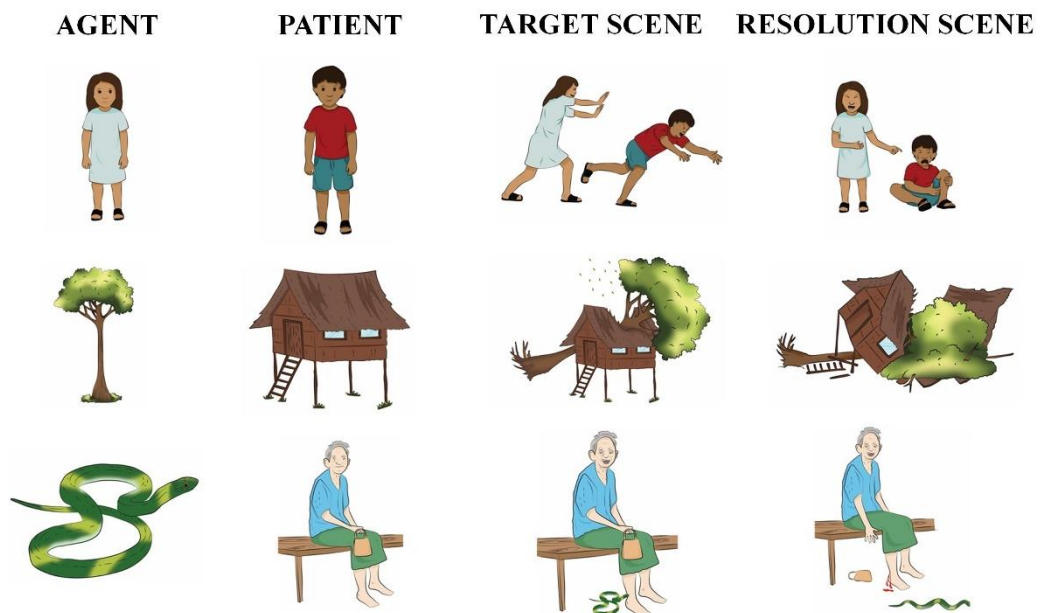


Figure 21 Illustrations for three Transitive Event Picture Sequences

As seen in Figure 21, target scene illustrations have the potential to elicit transitive clauses and two discourse referents. For example the target scene with the event ‘Girl pushes boy’ could elicit referring expressions for both ‘girl’ or ‘boy’ as

well as a predicate with a verb ‘to push’. However, with agent-initial and patient-initial versions of each sequence, participants can encounter target scenes under two possible conditions, each of which is distinguished by contrasting information statuses of the referents. The underlying experimental design of a sequence of TEPS is provided in Figure 22.

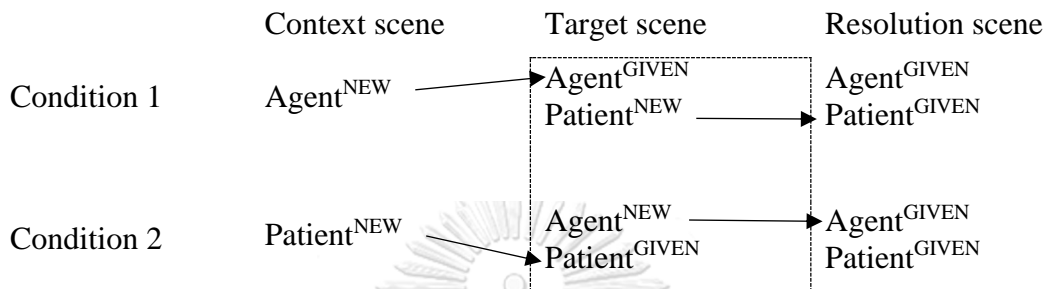


Figure 22 TEPS Sequence experimental design

The target scene is the primary controlled variable and the first context of interest in investigating argument structure, as our understanding of the speech context is framed by knowledge of the preceding cue. The resolution scene, however, was also thought to be an additional context of interest. First off, it serves as an important additional signaling of the completion of the transitive event. But since by the time of the resolution scene both agent and patient would be given, it provides an additional opportunity to capture potential givenness effects on Moklen word-form. So, in total, each individual sequence potentially offers two new to given transitions of discourse referents, one for each referent involved in the transitive event. Therefore, the experimental design of the TEPS instrument not only provided a context to elicit variations in argument structure, (Objective 1), but it also could potentially draw out informationally salient changes in word-form (Objective 2).

Careful consideration was taken in deciding which referents and events were to be portrayed within the sequences. In the interest of eliciting a range of lexical items, a variety of depicted discourse referents were portrayed. Also, as Hypothesis 3 specifically concerns monosyllabic alternants, referents with canonically monosyllabic Moklen forms e.g., ʔɣj ‘dog’ were excluded. On the whole, the following guidelines were used in designing sequences of TEPS:

- Canonical forms of potential lexemes are disyllabic.
- Intended isolated referents and events can be easily recognized in a graphical illustration.
- Scenes should depict culturally salient activities, while also avoiding areas likely to elicit potential loanwords (e.g., modern technology, domains of the majority culture).
- Overall, a general variety of referent types and events are presented throughout the entire task.

Half of the depicted agents across all sequences are human referents. The first reason for this is that many if not most transitive verbs not only require an animate agent, but belong exclusively to the human domain (e.g., cutting, washing, fishing). Secondly, besides many verbs just being naturally associated with human agents, there was also a range of categorical distinctions in the Moklen lexicon for human referents (e.g., number, gender, age). Additionally, depicting a variety of unique human characters was in line with a need to avoid potential givenness effects across sequences. Therefore, deciding on a balance of human referents required not only considering what would be natural arguments for the verbs—both in terms of agency and cultural saliency (e.g., men do fishing, women do laundry)—but it also required providing coverage of lexemes used for human referents in the Moklen lexicon. In consideration of these factors, the following balance of referents, as shown in Table 13 was achieved.

Table 13 Breakdown of referent types in the 24 TEP sequences.

TYPE	AGENT	PATIENT			
		Human	Animal	Inanimate	
Human 12	Boy				
	Boy				
	Girl				
	Girl				
	Man				
	Man	3	4	5	
	Female				
	Female				
	Grandfather				
	Grandmother				
	Group				
	Group				
Animal 6	Bird				
	Crab				
	Scorpion	2	2	2	
	Bird				
	Snake				
	Chicken(s)				
Inanimate 6	Rock				
	Coconut				
	Fishhook	2	1	3	
	Tree				
	Stump/Root				
	Fruit(s)				
<i>PATIENT</i> SUB-TOTAL		7	7	10	=24
<i>AGENT</i> SUB-TOTAL		12	6	6	=24
Referent type TOTAL		19	13	16	=48

From Table 13 we can see that 12 sequences had human agents and 12 had non-human agents (6 animal and 6 inanimate agents). Numbers detailing how many patients from each category individual agents are paired with shows the further consideration given to maintaining a balance of referents across the entire TEPS stimulus. Pairing of agents and patients was done with reference to transitive verbs appropriate for this task—the main criteria for selection of verbs being canonical disyllabicity and easily comprehensible in a graphical depiction. Referent type totals at the bottom of Table 13 shows there were a total of 19 human referents (not including multiple members a groups), and 29 non-human referents that are depicted in TEPS. Regarding transitive events with human agents, the uneven distribution of patient referent types (human=3, animal=4, inanimate=5) is due mostly to the verbs selected for this task. Because these verbs must be depicted through a graphical illustration, and have appropriate arguments, a greater number of patients naturally suggested inanimate objects e.g., patients of verbs like *wash*, *cover*, *peel*, *to fell (a tree)*, and *chop*. It also seemed that depictions of transitive events with two human referents might elicit responses focused on the intentions or relations between paired human referents, all of which might detract from the emphasis on basic transitive actions. All in all, sequences that had simple transitive verbs along with clear and contextually appropriate referents were favored. In the end, the sequences in Table 14 were the ones selected for inclusion in the TEPS stimulus.

Table 14 Referents and events of Transitive Event Picture Sequences

Item#	AGENT	PATIENT	VERB	Target Scene description
1	boy	woman	hug	Boy hugs mother.
2	boy	coconut	open/cut	Boy opens coconut with machete.
3	girl	water	get/ladle/pours	Girl gets water.
4	girl	boy	push	Girl pushes boy.
5	man	fish	fish	Man fishes fish with fishing pole.
6	man	fish	fish/cast	Man casts net catches fish.
7	woman	baby	cover	Woman covers child with blanket.
8	woman	clothes	launder	Woman washes clothes.
9	grandfather	chicken	kill	Grandpa kills chicken.
10	grandmother	banana	peel	Grandma peels banana.
11	group	tree	fell/chop	Father, son, and daughter fell a tree.
12	group	pig	carry	Men and women carry pig on stick.
13	bird	caterpillar	bite/pecks/eat	Bird pecks caterpillar.
14	crab	fish	open/eat/grab	Crab grabs fish.
15	chickens	branch	fly/perch/break	Chickens break tree branch.
16	bird	window	crashes/breaks	Bird breaks window.
17	snake	grandmother	bite	Snake bites grandma.
18	scorpion	girl	sting	Scorpion stings girl.
19	rock	boat	crash/go	Rock breaks boat.
20	coconut	bottle	fall/smash/break	Coconut breaks bottle.
21	mango	crab	fall/smash/break	Mango smashes crab.
22	tree	house	fall/smash/break	Tree smashes house.
23	stump	grandfather	trip/go/walk	Tree stump trips grandpa.
24	fishhook	man	scratch/cut	Fishhook scratches man.

HUMAN

ANIMAL

INANIMATE

The final design of TEPS was but one of several potential possibilities. In total, 25 individual human referents are depicted, each of which was illustrated so as to distinguish them as unique individuals. Assuming different words for human types e.g., boy, girl etc. (see Table 14) and general words for ‘human’, human referents were expected to generate at least eight distinct lexemes. Additionally, there were 22 non-human referents (11 animal and 11 inanimate) with the following items being used twice: *coconut*, *fish*, *chicken*, *tree*, *pig*, *rock*, and *crab*. Since potential verbs are limited for the inanimate agents, three sequences made use of a *fall/break* type of transitive events, creating some potential redundancy with regards to verb variety.¹⁷ However, supposing the 24 sequences could generate 24 verbal lexemes, taken with the 22 non-human referents and the eight human referent lexemes, at least 54 different Moklen lexemes were expected to be elicited in an idealized form of the task.

Each of the 24 sequences of TEPS could be presented in its agent-initial or patient-initial version. For actual presentation to participants, the sequences and their different versions were compiled into four different sets: A1, B1, A2, B2. Each TEPS set contained 12 agent-initial and 12 patient-initial sequences appearing in alternating orders. Each set with the same number A1 and B1, and A2 and B2 have the same sequences presented in the same exact chronological order but they are opposite with regards to which context scene they begin with (i.e., *agent* or *patient*). Therefore, they contrast in regard to which referents are given by the time of the target scene. A side-by-side description and comparison of the first three sequences of versions A1 and B1 are presented in Figure 23, note the different context scenes and contrasting information statuses of referents in the target scenes.

¹⁷ As well as questions about their “transitivity”. While previous research on Moklen as well as knowledge of MSEA typology already informed likely outcomes of these sequence types, examples of Moklen constructions here were the result of direct elicitation. It was though best not to assume this and go ahead include these events in order to increase referent diversity.

	CONTEXT	TARGET	RESOLUTION
VERSION A			
1	Man	Man fishes fish.	Man has fish .
2	Crab	Fruit smashes crab .	Crab is dead. Fruit on ground.
3	Chickens	Chickens break branch.	Branch, chickens have fallen.
VERSION B			
1	Fish	Man fishes fish .	Man has fish .
2	Fruit	Fruit smashes crab.	Crab is dead. Fruit on ground.
3	Branch	Chickens break branch .	Branch, chickens have fallen.

(Bold=given)

Figure 23 Comparison of TEPS sets: A1's and B1's first three sequences

The order of sequences in each set were arranged with an aim to increase the distance between similar referents in order to avoid any potential givenness effects carrying across the different sequences, especially avoiding back-to-back human referents in contiguous sequences. Table 15 provides descriptions of the Target Scene for A1 and B1 versions with a side-by-side comparison of the *given* discourse referent. Also, to aid with reference to specific sequences, starting in Table 14 above, a reference number (#1-24) was assigned to each sequence. This allows for easy cross-reference of sequences across the entire TEPS stimulus. It also allows for descriptive labels, such as *agent-initial* sequence #5 or *patient-initial* #5, which can also be abbreviated with coded tags like A5 or P5. Furthermore, coded tags can indicate which scene through an additional decimal point and number from 1-3, to refer to any or all of the three pictures of the sequence e.g., 'A5.2' refers to the target scene of sequence #5 'Man fishes fish' (see §4.5 Data presentation).

Table 15 TEPS Target Scenes Versions A1 and B2

ORDER	VERSION A1 (Blue)	VERSION B1 (Tan)	Ref #
1	Man fishes fish with fishing pole.	Man fishes fish with fishing pole.	5
2	Fruit falls and smashes crab .	Fruit falls and smashes crab.	21
3	Chickens break branch.	Chickens break branch .	15
4	Woman covers child with blanket.	Woman covers child with blanket.	7
5	Tree falls and smashes house.	Tree falls and smashes house .	22

6	Crab grabs fish .	Crab grabs fish.	14
7	Boy hugs mother.	Boy hugs mother .	1
8	Men and women carry pig on stick.	Men and women carry pig on stick.	12
9	Bird pecks caterpillar.	Bird pecks caterpillar .	13
10	Coconut falls and breaks bottle .	Coconut falls and breaks bottle.	20
11	Girl gets water.	Girl gets water .	3
12	Grandpa kills chicken .	Grandpa kills chicken.	9
13	Snake bites grandma.	Snake bites grandma .	17
14	Bird breaks window .	Bird breaks window.	16
15	Rock smashes boat.	Rock smashes boat .	19
16	Boy opens coconut with machete.	Boy opens coconut with machete.	2
17	Fishhook scratches man.	Fishhook scratches man .	24
18	Girl pushes boy .	Girl pushes boy.	4
19	Woman washes clothes.	Woman washes clothes .	8
20	Root of a tree stump trips grandpa .	Root of a tree stump trips grandpa.	23
21	Father and kids fell a tree.	Father and kids fell a tree .	11
22	Scorpion stings girl .	Scorpion stings girl.	18
23	Grandma peels banana.	Grandma peels banana .	10
24	Man casts net catches fish .	Man casts net catches fish.	6

(Bold=given)

TEPS sets A2 and B2 preserve the same controls for givenness as sets A1 and B1 however they were arranged in a different presentation order (see Table 16). Therefore, the TEPS stimulus is composed of four different versions, all of which differ in terms of givenness and presentation order. To reiterate, the difference between A and B are givenness conditions in the target scene (i.e., same context scenes), and the difference between 1 and 2 is the order in which sequences are presented. Each set was also given a different color binding (Blue, Tan, Green, Purple) to aid in identification in the field.

Table 16 TEPS Target Scenes Versions A2 and B2

ORDER	VERSION A2 (Green)	VERSION B2 (Purple)	Ref #
1	Fruit falls and smashes crab .	Fruit falls and smashes crab.	21
2	Boy hugs mother.	Boy hugs mother .	1
3	Men and women carry pig on stick.	Men and women carry pig on stick.	12

4	Bird pecks caterpillar.	Bird pecks caterpillar .	13
5	Coconut falls and breaks bottle .	Coconut falls and breaks bottle.	20
6	Girl gets water.	Girl gets water .	3
7	Bird breaks window .	Bird breaks window.	16
8	Woman washes clothes.	Woman washes clothes .	8
9	Root of a tree stump trips grandpa .	Root of a tree stump trips grandpa.	23
10	Father and kids fell a tree.	Father and kids fell a tree .	11
11	Scorpion stings girl .	Scorpion stings girl.	18
12	Rock smashes boat.	Rock smashes boat .	19
13	Man casts net catches fish .	Man casts net catches fish.	6
14	Snake bites grandma.	Snake bites grandma .	17
15	Girl pushes boy .	Girl pushes boy.	4
16	Chickens break branch.	Chickens break branch .	15
17	Woman covers child with blanket.	Woman covers child with blanket.	7
18	Tree falls and smashes house.	Tree falls and smashes house .	22
19	Crab grabs fish .	Crab grabs fish.	14
20	Grandma peels banana.	Grandma peels banana .	10
21	Grandpa kills chicken .	Grandpa kills chicken.	9
22	Fishhook scratches man.	Fishhook scratches man .	24
23	Boy opens coconut with machete.	Boy opens coconut with machete.	2
24	Man fishes fish with fishing pole.	Man fishes fish with fishing pole.	5

(Bold=given)

Each illustration was printed in color in a landscape orientation onto a single side of a 250gm A4 sheet of paper (210 x 297 mm). All illustrations were then bound into books preserving each versions' determined order. Numbered filler pages were also inserted to break up each three-page sequence and to aid in sequence identification. With three scenes per sequence a TEPS book contained 72 illustration pages and 24 filler pages for a total of 96 pages. See a side-by-side example of the sixteenth sequence (sequence Ref. #2) from TEPS sets A1 and B1 in Figure 24 below.



Figure 24 Photograph of the 16th sequence of TEPS sets A1 and B1

Accompanying the TEPS stimulus, a training booklet was created which contained six example three-picture sequences. Besides explaining the nature of the task, it was also a tool for assessment of participants' comfort with the task and a preliminary step in gaining their verbal consent for further participation. The first three sequences of the training booklet presented all three scenes on the same page to emphasize the idea that three-picture sequences were part of one singular story. The last three sequences shifted to the one picture per page format of the actual TEPS stimulus to help prepare participants for the pacing of the actual task. Illustrations for the training booklet were taken from other available materials which were not created by the researcher. Briefly describing these sequences' target scene, they were: (1) Dog chases cat (2) Person fries egg (3) Woman washes hair (4) Boy blows out candle (5) Person peels garlic (6) Man gets on elephant.

4.4.2 Procedure

The ideal set up for this task in the field was only a single participant and the research/interviewer, in an environment of minimal distractions. Before beginning the

task, the training booklet was presented to the participant. The interviewer explained that it contained short stories that were depicted through three illustrations. If needed, with the training booklet, the interviewer could demonstrate the aim of the task by providing an example narration using the target language. After the interviewer provided an example description of the first training sequence the interviewer asked the participant how they would narrate the sequence using the target language. Once they provided a response, the interviewer turned the page to reveal another three-picture sequence and asked the participant how they would describe it. In total, the training booklet provided six practice sequences for the participant. The training session also provided an opportunity for the interviewer to clarify any aspect of the task, as well as allow the participant to familiarize themselves with the task's pacing and flow. Once the training booklet was completed, the participant was asked if they would like to continue with the rest of the task.

Each participant completed one of the 4 versions of the TEPS stimulus. Participants were asked to provide narration and description of the 24 three-picture sequences. The stimulus book was placed in front of the participant (see Figure 17). During the session the interviewer maintained physical control of the stimulus, ensuring not only that the participant did not preview following pages, but that progression through the book was kept to a brisk yet natural pace. However, participants were free to point and make physical contact with the page. Once the participant completed description of an illustration, the interviewer turned the page, allowing the participant to freely describe each scene. If a participant had any questions during the task the interviewer responded but aimed to redirect the participants to providing their own interpretation. For example, if they sought clarification about the content of a picture, the interviewer would respond, "I don't know, what do *you* think?". If participants had any difficulty during a sequence description, guidance could be provided by the interviewer but was ultimately aimed at preserving the conditions of the design and the genre of text (i.e., a monologic picture description). Once the participant completed all 24 sequences the task was considered complete.

4.4.3 Data

In total 16 TEPS sessions were processed for analysis. The participants of these sessions consisted of 8 females and 8 males from 10 different communities (age range=47-77 and approximate mean 62). 2 females and 2 males completed one of the four versions (A1, B1, A2, B2) of the stimulus, as shown in Table 17. As the TEPS stimulus is composed of 4 versions, a full round of implementation consisted of 4 unique participants completing one of the 4 versions: A1, B1, A2, and B2.

Table 17 TEPS Sessions selected for analysis

Contributor	Gender	Community	Version	Length
TG	F	Thap Tawan	A1	15:38
TW	F	Hin Lat	A1	07:28
NK	M	Tha Yai	A1	09:16
NN	M	Thung Dap	A1	07:41
LP	F	Thap Tawan	A2	09:24
EW	F	Hin Luk Diaw	A2	08:05
LW	M	Theppharat	A2	07:18
NJ	M	Bon Rai	A2	13:52
YN	F	Thap Tawan	B1	12:41
PB	F	Laem La	B1	14:23
LN	M	Thap Tawan	B1	10:27
YG	M	Thap Plaa	B1	11:26
WN	F	Thung Waa	B2	08:38
DW	F	Bon Rai	B2	11:00
LI	M	Theppharat	B2	05:26
CU	M	Laem La	B2	07:51

The length of the session indicated in Table 17 reflects the length of the primary session and not the training session. Participants could spend anywhere from 20-90 seconds on an individual sequence, but on average an entire TEPS session, excluding the training portion, lasted around 10 minutes. In total the 16 TEPS sessions selected for analysis amounts to 2 hours and 40 minutes of speech. With 16 TEPS sessions

selected for analysis, the resulting TEPS data had 384 sequence descriptions, or in terms of individual scenes 1,152 picture descriptions.

4.4.4 Annotation

Two aspects of TEPS data were annotated for this study: the argument structure of transitive clauses in responses to target scenes (Objective 1) and the information status of monosyllabic alternants of disyllabic word-forms (Objective 2). Figure 25 is a screenshot of annotation of TEPS data in ELAN.

S-Vt-E	Vi	2 Vi	S	E	A	r-given	l-given	1/1	Vt
			bu:m pot'aw	kate:	na? 7ola:n				
			r-given	r-given	r-new				
			l-given	l-new	l-new				
			1/1	1/1	0/0				
					Vt				

Figure 25 Screenshot of TEPS annotation in ELAN

A valid response, with regards to Objective 1, was one in which the speaker related the two discourse referents through a singular predicate during description of the target scene. Speakers' initial responses were the target of analysis. Non-content vocatives and regulatory or fragmentary IUs were excluded from analysis. Out of 384 target scene descriptions, 354 target scenes were deemed to have valid responses. Valid responses to target scenes were then annotated for their grammatical properties, namely the linear order of verbs and core arguments of transitive clauses. Variant argument structures also needed to fit the criteria of being within a single intonation unit (see the full GU condition in §3.5.2). Responses which utilized detachment constructions, ellipsis, and passivization were noted as discourse strategies with which participants manage the conditions of the task. Linear coding of verbal predicates and

core arguments along with accounting for prosodic disjunction resulted in various response types.

In presenting findings clausal elements are described using the Dixon's (2010) convention of A=agent-like argument of transitive clause, O=patient-like argument of transitive clause, and V=Verb. And while all valid responses involve bivalent predicates, some core arguments could be better described with S=subject of an intransitive clause, and E=extension to the core (see §3.4.1), as the interpretation of the target scene did not reflect prototypical transitivity. But, since in most cases A and O reflect the semantic roles of most arguments (and correspond directly to the semantics of the experimental conditions) sometimes AGENTS and OBJECTS are used to speak generally of all the arguments within the TEPS data set.

Several issues concerning analysis of argument structure in TEPS data need to be addressed. A response was considered valid if the speaker used a singular predicate to relate the agent and patient referents during description of the target scene. Therefore the minimal criterion for a valid response was that the predicate be bivalent, that is it selected for two arguments. Consequently responses vary in terms of exhibiting prototypical transitivity (which can be regarded as a scalar notion, see Hopper & Thompson 1980). Additionally, there are some instances where a target scene elicited an interpretation that inverted the intended semantic roles of the depicted discourse referents. For example, the illustration in the target scene of sequence #21 'Fruit smashes crab' could be interpreted as 'Crab eating fruit'. Rather than discarding such responses, which were valid in their own right, they were instead tallied in accordance with the participant's interpretation: crab=AGENT and fruit=PATIENT, as they still exhibited a transitive clause. In such cases, however, TEPS scene reference codes are still in accordance with the intended design (see Table 14).

For investigation into the relationship between changes in information status and Moklen word-form (Objective 2), all monosyllabic alternants within the TEPS data received information status tags (see RefLex Scheme §3.3.2). Additionally, whether the disyllabic form appeared within the same sequence, as well as the referential distance (i.e., the number of intonation units) between the two word-forms was also noted. Information status was annotated at two levels, a referential level (r-level) and

a lexical level (l-level). The r-level is used as an analysis of the tracking of discourse referents through the use of referring expressions. Tags at the l-level are attached at the word level. Nouns and classifiers, however, which can act as the heads of referring expressions (i.e., noun phrases) receive tags at both levels. Individual verbs, prepositions, and numbers, however, do not receive r-level tags and are therefore only evaluated at the l-level, as shown in Table 18.

Table 18 Word Classes and levels of information status analysis

	referential level	lexical level
nouns, classifiers	✓	✓
verbs, prepositions, numbers	—	✓

The criteria for given and new differ at each level. At the r-level once a discourse referent has been introduced it can no longer be *r-new*. At the lexical level words are considered *l-new* if they are not related to another word within the last 5 IUs. Within the RefLex Scheme, Riester and Baumann also use a distance of 5 intonation units to distinguish between the categories of *r-given* and *r-given-displaced*, and to determine a renewal of the category *l-new*, after being *l-given*. They admit that a distance of five units is arbitrary to a certain degree but point out that for annotation tools with the ability to automatically process the distance of anaphoric links removes the need for certain categorical distinctions. However, as the TEPS data elicited brief narrative descriptions, long-distance links beyond 5 IUs was not an issue. Moreover, several of the distinctions within the taxonomy of tags were not relevant to the discourse context nor the types of lexical process of a language of this typological profile and therefore were not used within the TEPS data. A very small proportion of other tags (e.g., *r-bridging*, *r-given-displaced*, *l-accessible*, see Table 9) did make appearances in the data, mainly in the Stolen Fish texts. These are regarded as functionally not “new” and were therefore considered “given” for the purpose of analyses. However, when relevant, other tags and features of the Reflex Scheme are discussed. “Given” usually corresponds to the categories *r-given* and/or *l-given*. Ultimately, given the narrow discourse context of the picture stimuli the full breadth of potential information status tags is not reflected in this study. Simplifying the large

majority of tags used within this study, core definitions for “given” and “new” for both dimensions are provided in Table 19.

Table 19 Given and new in the RefLex Scheme

	<i>new</i>	<i>given</i>
r-level	discourse-new entities	previously mentioned referents
l-level	inactive lexical concept	active lexical concept

A limitation of picture-stimuli is a certain type of paradox presented by certain information statuses. For example, tags such as *r-given-sit* or *r-environment* are for referents immediately present in the text-external context. As participants are describing a static illustration at the time of speech, one might say that all referring expressions in this genre might correspond to these tags. However, a speaker’s narrative prose generally takes a stance that posits distinct characters and objects as discourse entities. Still, there is a lingering question as to what extent speakers’ acts of reference were more directed at a *depiction* of a referent. For example, the contrast would be evident with sentences like “The man is getting a fish” versus “This is a picture of man getting a fish.” For simplicity, I assumed that speakers are usually maintaining a narrative stance with respect to depictions within the stimulus. But, this aspect must be considered a part of a picture-description genre.

4.4.5 Analysis จุฬาลงกรณ์มหาวิทยาลัย

To assess Hypothesis 2: Variations in argument structure and associated grammatical constructions can be accounted for by the given-before-new principle, valid responses to target scenes of the Transitive Event Picture Sequences (TEPS) stimulus were examined. There were 185 valid responses for agent-initial sequences and 168 for patient-initial ones, for a total of 353 valid responses. Patterns from the linear coding of clausal constituents were used to categorize the different response types. Descriptive statistics were used to determine whether variant argument structure patterns and discourse strategies were influenced by the experimental conditions of the TEPS stimulus (agent-initial or patient-initial), and whether this was in line with the given-before-new principle. Reference to given and new here refers solely to the *r-new* and *r-given* tags of the referential level. Further assessment of the

given-before-new principle also draws upon clausal intonation units from the Stolen Fish corpus.

For assessment of Hypothesis 3: Use of monosyllabic alternants will correspond to given information statuses, all monosyllabic alternants within the TEPS data were examined. In total, the 16 TEPS sessions selected for analysis amounted to 2 hours and 40 minutes of speech. From this, a total of 530 monosyllabic alternant tokens made up of 64 different lexemes were elicited. These included 28 different nouns, 31 verbs, and 5 closed-class items. Descriptive statistics were used to determine the extent to which monosyllabic alternants corresponded to a “given” information status or particular information status profiles. Further analysis looked to the alignment of the information statuses of disyllabic word-forms and monosyllabic alternants within instances of word-form shifts: the appearance of both the disyllabic and monosyllabic word-form within a single sequence description.

4.5 Data presentation

As the primary data of this study is derived from the use of picture-based stimuli in staged communicative events, a certain level of transparency of the discourse context is available. To facilitate readers’ interpretation of shared examples, every example taken from staged communicative events indicates the picture being viewed at the time of speech. The coding for this is detailed below.

At the end of the translation line the source of the data is coded within square brackets. For example a source tag may appear as follows:

[AB.SF16.50-51]

The first two-letter code refers to individual speakers, which can be referenced to Tables 11 and 17. After this, the illustration that the speaker was viewing at the time of speech will be indicated. In the example above, SF16 indicates that scene #16 of the Stolen Fish picture book was being viewed. The last part of the tag here is an indication of which intonation unit(s) from that particular text are included within the example.

Tags for examples drawn from the Transitive Event Picture Sequences differ. For example an utterance from here will follow this format:

[AB.A7.1-3]

Again the first two-letter code refers to the speaker. The second part of tags for TEPS examples indicates whether the sequence was either the agent-initial version (A) or the patient-initial version (P). The numerical value after either A or P then refers to the sequence reference #, which runs from 1-24 and can be referenced to Table 14. The last part of the tag here then indicates which of the sequence's three scenes were viewed at the time of speech: 1=context scene, 2=target scene, 3=resolution scene. All illustrations used within this study are presented as appendices.

Within singular examples I use ellipsis [...] to distinguish between separate intonation units. Also as consideration of Moklen word-form is a main interest of this study, I wanted to clearly indicate lexical material of Thai origin in order to clearly distinguish Thai monosyllables from Moklen ones. Therefore, words that are of Thai origin within Moklen speech are indicated by the use of numerical labels (1-5) which correspond to the lexical tone of Standard Thai (Pittayaporn, 2018). This is not to say that these words have these corresponding tonal characteristics, but instead it serves to merely indicate a Thai source of lexical material (in actuality Southern Thai is the most likely source for Moklen). Therefore, words with diacritics will either be Moklen words which are canonically monosyllables (e.g., *klá:n* 'bone', *háh* 'NEG', *cǎj* '1.sg'); an iambic disyllable, with heavy stress on the second syllable (e.g., *baǵò:ʔ* 'frog', *cʰədiŋ* 'to be loud', *dalá:n* 'path'), or a variant monosyllable of a disyllabic form where the minor syllable is elided.

4.6 Summary

The methodology of this study is based on the use of two custom-designed field stimuli with Moklen collaborators. The Stolen Fish picture book was used to elicit monologic narrative texts from which the informational and syntactic properties of Moklen intonation units could be examined. The Transitive Event Picture Stimulus offered a narrower speech context from which givenness effects on Moklen argument structure and word-form could be examined. Data from both staged communicative events was annotated and analyzed using ELAN and EXCEL. The study's hypotheses are evaluated with descriptive statistics and linguistic analysis.

5 Informational properties of intonation units.

5.1 Introduction

Description of the informational properties of intonation units offers a detailed account of the shape and contents of Moklen speech. To study the informational properties of Moklen intonation units, the following hypothesis was tested: clausal intonation units will conform to the one-new-idea constraint. Findings from the Stolen Fish corpus confirm this hypothesis by showing that when clausal intonation units had more than one overt referring expression, there was almost always a single new referring expression. More generally, clausal intonation units within the corpus can be characterized as displaying a low-information load. Typically, clausal intonation units under-specify discourse referents, with 54% having only a single referring expression. As for a few marked violators of the one-new-idea constraint, further examination revealed they contained idiomatic verb-object constructions or were likely the result of factors stemming from the stimulus procedure. Included along with these findings is a description of constructions for introducing new referring expressions and discussion of the role of ellipsis and disjointed syntactic dependencies within Moklen discourse.

5.2. The one-new-idea constraint

Clausal intonation units at minimum contain a verb.¹⁸ However, overt expression of core and/or non-core arguments can vary. Based on the one-new-idea constraint a prediction was made that clausal intonation units (IUs) would contain no more than one new referring expression, a noun phrase which picks up an identifiable entity as a discourse referent (Baumann and Riester, 2012; Gundel et al., 1993). The number of referring expressions (R-EXPs) in a clausal IU is the result of verbal semantics (e.g., transitivity and multi-verb constructions) minus the amount of backgrounding of information through ellipsis—which effectively represents a reduction in the amount of overtly encoded information. Within the Stolen Fish corpus, the number of R-EXPs within clausal IUs ranged from zero to four, as shown in (5.1) - (5.5).

¹⁸ An exception is verbless copular constructions, see §5.2.1.

Zero R-EXPs

- (5.1) *ʔɿ: ʔón be*
 AFFR.VOC give PRT
 ‘Yeah, (I’ll) just give (him) (it).’ [NK.SF24.185]

One R-EXP

- (5.2) *ʔón [ʔadá:] nám*
 give duck eat
 ‘(He) lets a duck eat (it).’ [DW.SF15.75]

Two R-EXPs

- (5.3) [*caná:t tálj lùj*] *ná:ʔ [ʔeká:n ʔa=làt]*
 child three CLF have fish one=CLF
 ‘The three children have one fish.’ [WN.SF16.62]

Three R-EXPs

- (5.4) [*bá:p pʰɔ:3tʰaw3 ni:4 tak2 [pukát] bʲj [ʔeká:n tálj pʰóh]*]
 elder.male old.person this scoop gillnet acquire fish three CLF
 ‘The old man is pulling up the gillnet, having gotten three fish.’ [YT.SF04.8]

Four R-EXPs

- (5.5) [*tʰi:3 nʲj [ʔaná:t nʲj] bá:ʔ [ʔeká:n] káw héʔ [cʰuwiák]*]
 at DEM.PROX child DEM.PROX carry fish go at fishhook
 ‘Here, the child is carrying the fish and going to the fishing line.’ [CN.SF18.66]

Of the corpus’s 1,517 clausal IUs it was found that 54% ($n=822$) contained only R-EXP, as shown in Figure 26. The remaining 46% consisted mostly of IUs with zero R-EXPs ($n=282$) or two R-EXPs ($n=378$).

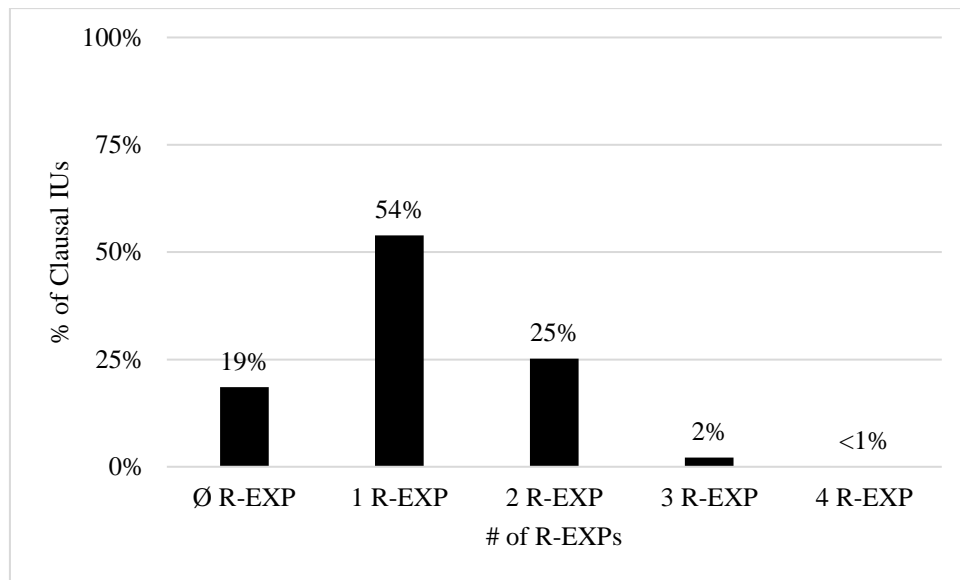


Figure 26 Percentage of clausal IUs by # of R-EXPs

Clausal IUs were seen as conforming to the one-new-idea constraint (ONIC) if they had no more than one new R-EXP. Given that 73% of clausal IUs had either one or zero R-EXPs, a general conformity to the constraint was immediately on display. For clausal IUs with *more than one* R-EXP, evaluation of their conformity to the ONIC required assessing the information status of multiple overt R-EXPs. For example, since the clausal IU in (5.6) has one given R-EXP and one new R-EXP, it can be framed as having a composite structure of “given + new” (G + N). And as the grammatical object *bag̀:ʔ* ‘frog’ is the only new one, the IU in (5.6) is deemed as not violating the ONIC.

(5.6)	<i>ʔɨj</i>	<i>nəŋé:n</i>	<i>bag̀:ʔ</i>
	dog	chase	frog
	GIVEN		NEW

‘The dog is chasing a frog.’[YN.SF27.95]

Evaluation of all clausal IUs in terms of their compositions of new and given R-EXPs, revealed the extent to which all clausal IUs abided by the ONIC. As shown in Table 20, only eight clausal IUs failed to meet the criteria of having one new R-EXP (see bolded rows). The main finding from this analysis is that clausal intonation units overwhelmingly conform to the one-new-idea constraint.

Table 20 Clausal IUs' given and new R-EXPs

Clausal IU information structure type		<i>n</i>	%
∅ R-EXP	-	282	19%
1 R-EXP	N	111	7%
	G	711	47%
2 R-EXP	G + N	53	4%
	G + G	305	20%
	N + G	13	1%
	N + N	7	< 1%
3 R-EXP	G + G + G	27	2%
	G + G + N	5	< 1%
	G + N + N	1	< 1%
4 R-EXP	G + G + G + G	2	< 1%
Total		1,517	100%

G=given, N=new, **bold**=one-new-idea constraint violators

5.2.1 New referring expressions in clausal intonation units

It is important to point out the large asymmetry in the amount of new and given R-EXPs within the corpus. Compiling just clausal IUs with *any* new R-EXPs amounts to 190 clausal IUs (see Table 21), only 12.5% of all clausal IUs. Clausal IUs with new R-EXPs were therefore already a minority of IU tokens, a fact revealing the extent to which speech in the corpus is about given information. Still, as predicted by Chafe's (1994) description of the informational properties of IUs (§3.5.3), for clausal IUs a focus is placed of new information, as 111 out of 190 of the clausal IUs even having a new R-EXP had it as the *sole* overt argument.

Table 21 Clausal IUs with new R-EXPs

Information Structure		<i>n</i>
1 R-EXP	N	111
	G+N	53
2 R-EXP	N+G	13
	N+N	7
3 R-EXP	G+G+N	4
	G+N+N	1
Total		190

As a new R-EXP within a clausal IUs represents an informationally marked type, it is worth identifying and describing syntactic constructions used to introduce them. Such as an existential construction with the verb *ná:ʔ* ‘exist’ that is used to introduce new discourse referents. In (5.7) and (5.8), we see the typical pattern is for the verb *ná:ʔ* ‘exist’ to be followed by a single new R-EXP.

(5.7) *ná:ʔ* [*kanáj bujáj*]_{r-new}
 exist man young.male
 ‘There are young men.’ [NG.SF26.181]

(5.8) *ná:ʔ tɛ:2* [*kiǎk*]_{r-new} *káʔ le:w4*
 exist only fishbone PRF PRF
 ‘There are only fishbones now.’ [NK.SF27.98]

Another possibility is fronting the posited entity before the verb *ná:ʔ*, like in (5.9).

(5.9) [*ʔáj*]_{r-new} *kɔ:3 ná:ʔ*
 dog CONN exist
 ‘There’s a dog.’ [WN.SF27.113]

Usage of *ná:ʔ* ‘exist’ in an existential construction is distinct from another sense of *na:ʔ* ‘have’ which is a bivalent verb selecting for two core arguments, as shown in (5.10).

(5.10) *caná:t tǎlj lùj ná:ʔ ʔeká:n ʔa=làt*
 child three CLF have fish one=CLF
 ‘The three children have one fish.’ [WN.SF16.62]

Another means of introducing new R-EXPs particular to reference to physical objects—in this case the picture stimuli—is a verbless copular construction (Dixon 2009). Typically, this construction begins with a copular subject, often the proximate demonstrative *nəj*, accompanied by a deictic gesture towards a depicted entity (see Figure 27). Uses of *nəj* as a copular subject were marked as given as they were “situationally-evoked”, that is they were items evident to both speaker and hearer from the external discourse context (Riester & Baumann, 2017). The verbless copular construction then equates the copular subject, in this case the depicted item on the page, with a copular complement in the form of a full lexical expression, which in turn could constitute a new R-EXP. Verbless copular complements with discourse-new

entities were analyzed as having an information structure of G+N such as (5.11), where the speaker is identifying a new entity of the scene, *ʔapúj* ‘fire’.

- (5.11) *nɔ̀j* *[ʔapúj]*_{r-new}
 DEM.PROX fire
 ‘This (is) fire.’ [TG.SF17.76]



Figure 27 Deictic gesture accompanying verbless copular construction

Outside of this meta-narrative use of describing the stimuli, the verbless copular construction was also found within narrative prose. In (5.12) the speaker provides an utterance given as speech from one character when he is presenting the fish to the picnicking girls.

- (5.12) *nɔ̀j* *ká:n kʰɔ:ŋ5 bɔ̀j*
 DEM.PROX fish POSS 2.SG
 ‘These (are) your fish.’ [LN.SF16.61]

For clausal IUs, in most cases a new R-EXP was the sole overt argument (see Table 21). For these IUs a new R-EXP might be the grammatical subject of a clause with a simple predicate, that is the verb did not select for further grammatical arguments. Like in (5.13) where the common O argument *ʔeká:n* ‘fish’ was not

included or (5.14) with just the intransitive verb *nenán* ‘stand’. For clausal IUs only having the single new R-EXP, it was an A or S argument only 26 out of 111 times.

- (5.13) [*caná:t*]_{r-new} *dó:k mán*
 child sit fish.with.rod
 ‘A child’s sitting fishing.’ [KG.SF01.1]

- (5.14) [*təbǝj lùj ni:4*]_{r-new} *nenán*
 three CLF TOP stand
 ‘These three people are standing.’ [DW.SF10.24]

More often new R-EXPs were arguments occurring later in the clause. For example, they were new grammatical objects like in (5.15) and (5.16), where ellipsed grammatical subjects corresponded to previously mentioned discourse referents.

- (5.15) \emptyset *nəbáj* [*chəpʰlǝw*]_{r-new}
 wear pants
 ‘(He’s) wearing shorts.’ [NG.SF01.002]

- (5.16) \emptyset *dúk* [*matá:ʔ təmán*]_{r-new}
 put fishhook
 ‘(He’s) putting (it) on a fishhook.’ [DW.SF23.67]

This also included other argument types like an E argument for intransitive verbs (5.17) or a lexical non-core argument in a clause final prepositional phrase like in (5.18).

- (5.17) \emptyset *mɛjǝk dəbút sa2ʔdut2* [*batǝj*]_{r-new}
 walk run trip rock
 ‘(He’s) walks, runs, and trips on a rock.’ [NG.SF12.67]

- (5.18) \emptyset *namé:n nɛjǝ:ʔ tʰi:3* [*batǝj niʔú:n*]_{r-new} *káʔ*
 hide look.at at tree.trunk coconut PRF
 ‘(He’s) hidden, watching from the trunk of a coconut tree.’ [HJ.SF04.016]

As shown in Table 22, clausal IUs with a single new R-EXPs most often had it in slots for O, E, or non-core arguments. In conforming to the one-new-idea constraint, clausal IUs which introduce new R-EXPs often omitted grammatical subjects that corresponded to previously mentioned discourse referents. Assuming the preferred argument structures of AVO and SVE, the ellipsis of given A or S arguments (despite

there being no overt argument) is in principle also consistent with the given-before-new principle (see §6.2.3).

Table 22 Argument type: Clausal IUs with only a single new R-EXP

new R-EXP argument type	<i>n</i>	%
A, S	26	23.4
O, E, lexical non-core	85	76.6
Total	111	100

5.2.2 New referring expressions in phrasal intonation units

Another site for new R-EXPs was phrasal IUs. A phrasal IU is an intonation unit that contains a disjointed syntactic dependency, such as an isolated noun phrase that conveys only a fragment of a proposition. In the Stolen Fish corpus, phrasal IUs were typically noun phrases appearing as distinct IUs, that is they were not *prosodically* integrated into a clause. Semantically, however, phrasal IUs could still contain new R-EXPs that were appropriate arguments of a nearby clausal IU. For example, the S or A arguments in initially detached phrasal IUs shown in (5.19) and (5.20).

(5.19) [ʔenɔŋ kʰu:3 ʔaná:t]_{r-new} ... dɔ:k hɛʔ batán kaʔé:w dɔ:k
 mother and child sit at base.of.tree wood sit
 ‘A mother and child... are sitting by a tree sitting.’ [NG.SF07.34-35]

(5.20) [ʔaká:ʔ təlɔ́j lúj]_{r-new} ... nenán neŋɔ:ʔ
 older.sibling three CLF stand look.at
 ‘Three older siblings...standing, looking.’ [YN.SF10.26-27]

Finally-detached phrasal IUs, however, with new R-EXPs consisted only of noun phrases that were either O or lexical non-core arguments like in (5.21) or (5.22).

(5.21) lát mɔ:ŋl háh lɔ:jl ... [tamán ʔe:ŋl]_{r-new}
 to.steal look NEG INTS fishing.basket REFL
 ‘As (it’s) stolen, (he’s) not looking at all!...(at) his own fishing basket.’
 [NJ.SF33.117-118]

(5.22) bá:ʔ ʔeká:n dín hɛʔ ... [kʰanl nəmán pɔʔ]_{r-new}
 carry fish to.come at fishing.rod fish.with.rod 3SG
 ‘(He’s) carrying the fish coming to...his fishing rod.’ [KG.SF18.60-61]

In total there were 348 phrasal IUs in the corpus, 55 of which had a new R-EXP. During discourse, a referent within a disjointed phrasal IU might later be incorporated into a clausal IU. For example, in (5.23.1) we can see how a topicalized noun phrase introduces a brand-new discourse referent ‘the dog’, which would be the appropriate A argument for the following verb *neḡó*: ‘look’. Note that in the following utterance in (5.23.2), when the ‘dog’ is now given it was integrated into a singular clausal IU.

(5.23)

(5.23.1) [*ʔáʝ nɔ́ʝ*]_{r-new} ... *dó:k neḡó:ʔ*

dog DEM.PROX sit look.at

‘The dog here...it’s sitting, looking.’ [YN.SF17.56-57]

(5.23.2) *ʔáʝ dó:k neḡó:ʔ*

dog sit look.at

‘The dog’s sitting, looking.’ [YN.SF17.58]

With the design of *Stolen Fish*, a potential lack of new R-EXPs was anticipated. Therefore, scenes with secondary characters, like the dog in scenes #16-17 or the frog in #27 (Figure 28) were included as opportunities to elicit the introduction of new discourse entities.

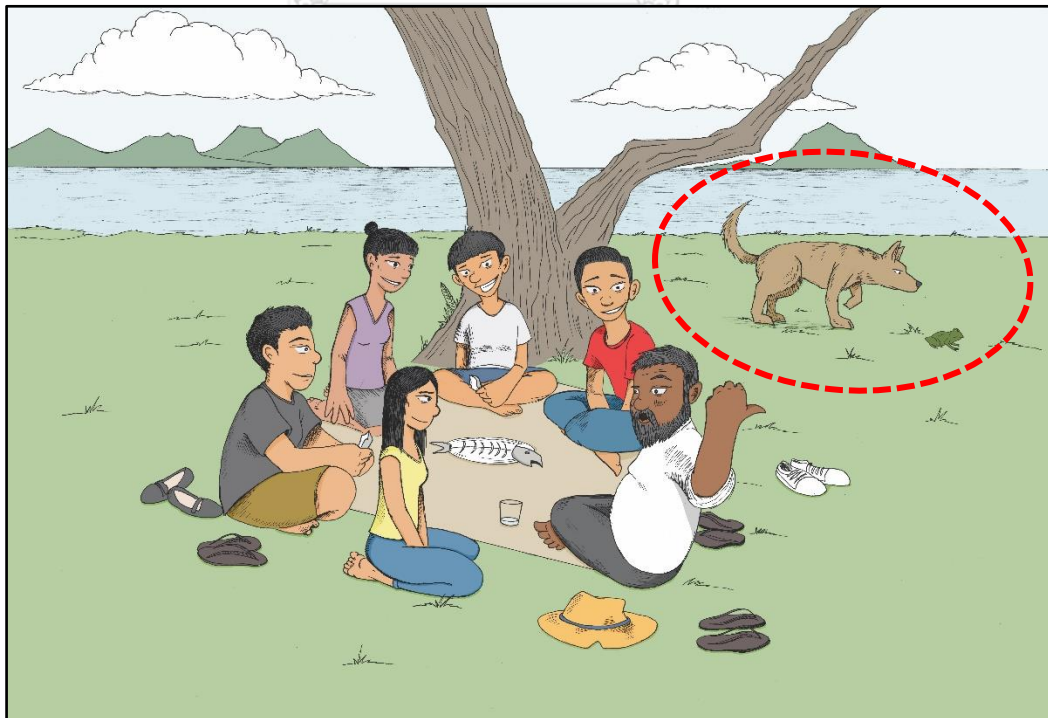


Figure 28 Secondary characters in Scene #27 ‘Eating together.’

However, when (and if) speakers referred to these discourse referents, they still followed the ONIC. For example, in (5.24.1), a description of scene #27, we can see how a speaker first used two separate IUs for the two new discourse referents before integrating them into a singular clausal IU, in (5.24.2). Note also with (5.24.2) the assemblage of constituents into an AVO order, as well as the reduction of the verb to a monosyllabic alternant (see §7).

(5.24)

(5.24.1) *nɛŋɔːʔ [baŋɔːʔ]_{r-new} ... kɔːʔ nɔːj [ʔɔːj]_{r-new}*
 look.at frog CONN DEM.PROX dog
 ‘(It’s) looking at a frog...here, the dog.’ [DW.SF27.76-77]

(5.24.2) *ʔɔːj ŋɔːʔ baŋɔːʔ*
 dog look.at frog
 ‘The dog’s looking at the frog.’ [DW.SF27.78]

5.2.3 Violating the one-new-idea constraint

Investigation into the one-new-idea constraint (ONIC) was operationalized by limiting the amount of new referring expressions within a clausal IU to one. As shown in §5.2 the vast majority of clausal IUs in the Stolen Fish corpus abided by this criterion, with only eight clausal IUs marked as being in apparent violation. All marked violators, however, were found to likely be the result of priming effects or contain particular verb-object collocations. To better get at what this study’s framing of the ONIC distinguishes, in this section each of these eight “violators” is examined more closely. Overall, these violators are a rarity and therefore do not pose a problem for the main hypothesis. But, closer attention highlights the role of verb-object constructions as well as issues surrounding both the procedure and use of a picture-based stimuli. Furthermore, discussion of these instances connects back to Chafe’s (1994) original framing and conception of the one-new-idea constraint.

As part of the session procedure, before beginning primary narration, speakers first previewed and discussed the entire Stolen Fish picture book. Two ONIC violators were initial utterances describing scene #1 ‘Line in water’. Crucially, these utterances occurred only moments after each speaker finished viewing scene #33 ‘Just deserts’ during the preview phase. Therefore, speakers were likely beginning the primary session with content from the picture-stimuli already in mind. For example,

in (5.25) Speaker YT's use of *caná:t* 'child' and *ʔeká:n* 'fish' were both tagged new as they were discourse-new entities with regards to the text of the primary session.

- (5.25) [*caná:t*]_{r-new} *nəmán* [*ʔeká:n*]_{r-new}
 child fish.with.rod fish
 'A child is fishing.' [YT.SF01.1]

Similarly, in (5.26) Speaker HJ also had two new R-EXPs in their very first IU.

- (5.26) [*bó:ʔ nɿj*]_{r-new} *nuwiák* [*ká:n*]_{r-new} *lá:j nú:*
 person DEM.PROX fish fish PROG DEM.MED
 'This person is fishing, there.' [HJ.SF01.1]

Although the R-EXPs in (5.25) and (5.26) were new with respect to the text of the primary session, an interesting contrast is provided by each of the speakers' utterances from the preview phase—that is when they actually viewed the stimulus for the very first time. Consider example (5.27), where YT simply identified the depicted activity using just the verb *nəmán* 'fish.with.rod' without any overt arguments.

- (5.27) *nəmán*
 fish.with.rod
 'Fishing.' [YT.*SF01.001]

Again, compare this utterance to (5.25) when YT was providing their primary narration and see how only on their second pass did YT supply overt core arguments for the same verb. Similarly, during the preview phase HJ also first abided by the ONIC, as shown in (5.28). Here you can see that while they did supply overt R-EXPs, all constituents were not integrated into a singular clause until the third IU—which conforms to the ONIC as *ná:t* 'child' and *ká:n* 'fish' were given. The comparison between speakers' actual first interactions with these scenes and their utterances during the primary recording, therefore highlights the potential for priming effects in early scenes of the primary narration session.

- (5.28) [*caná:t*]_{r-new} *nuwiák* ... [*ká:n*]_{r-new} *nɛ ʔɿ:* ... *ná:t wiák ká:n*
 child fish fish PRT VOC child fish fish
 'A child is fishing...for fish, uh...a child's fishing.' [HJ.*SF01.1-3]

Another important aspect of scene #1 is that although utterances describing it, often posit (*ʔe*)*ká:n* 'fish' as a grammatical object, in scene #1 there is actually no

identifiable entity corresponding to this referent—that is there is no depiction of a fish (see Figure 29). This brings us to an additional feature of the RefLex Scheme, the *+generic* tag used to append information status annotations. This tag is used along with all r-categories for generic, non-specific, or hypothetical entities, meaning that R-EXP *(ʔe)ká:n* ‘fish’ in (5.25) and (5.26) actually had an information status of *r-new+generic*.



Figure 29 Stolen Fish scene #1 ‘Line in water’

Given both the primed discourse context ONIC violators in (5.25) and (5.26) and the indefinite nature of the grammatical object in both clauses, it seems that instead of these being separately “activated” ideas, something else is going on. What seems more likely is that for both verbs in these examples, *nəmán* ‘fish.with.rod’ and *nuwiák* ‘fish’, the grammatical object *ʔeká:n* ‘fish’ is part of a conventionalized verb-object construction. Chafe’s (1994) broader framing of ONIC is that “there are no constructions internal to an intonation unit with two items that independently express new information” (p.116). From this, Chafe allows for verb-object constructions or more general lexicalized phrases, idiom, and collocations to be subsumed as *functionally unitary* within his model. This means that where we do find what appears on the surface to be separately activated new referents in a single IU, they are likely to be conventionalized constructions within that language. This aspect of Chafe’s

framing of constructions within ONIC is significant because it seems to account for the other six clausal IUs marked as ONIC violators.

Five ONIC violators, also occurred relatively early in the texts with scene #3 ‘Today’s catch’ where the ‘fisherman’ character first appears. For example, in both (5.29) and (5.30) the agent *(ʔe)ba:p* ‘elder.male’ is new and so is their grammatical objects *ʔuan1* ‘gillnet’.

(5.29) [*ʔe:ba:p nɔ̃j*]_{r-new} *dín wa:ŋ1* [*ʔuan1*]_{r-new}
 elder.male DEM.PROX come lay gillnet
 ‘This man is laying out a gill net.’ [CN.SF03.5]

(5.30) [*ba:p nɔ̃j*]_{r-new} *tak2=na:* [*ʔuan1*]_{r-new}
 elder.male DEM.PROX scoop=3.SG gillnet
 ‘This man is pulling up a gillnet.’ [TG.SF03.8]

Disregarding potential priming effects, it seems likely that these examples also feature verbs and objects that are closely associated. The difference, for verbs in the two examples, just reflects different interpretations as to whether the fisherman was either setting up or collecting his gillnet. A more definite case for a verb-object collocation can be made for all the other ONIC violators, which feature the collocation *tʰɔ:t3 he:5* ‘cast a cast net’, where *he:5* ‘cast net’ was tagged as *r-new*.¹⁹ In (5.31) and (5.32), there are two new referring expressions, as the collocation appeared along with the very first mention of the fisherman. Note that for scene #3 there is a depiction of a net (variably interpreted as a gillnet or cast net), and so there is an observable definite referent for these grammatical objects.

(5.31) [*pʰɔ:3tʰaw3*]_{r-new} *tʰɔ:t3* [*he:5*]_{r-new}
 old.person cast cast.net
 ‘An old person is casting a net.’ [WN.SF03.10]

(5.32) [*ʔe:ba:p*]_{r-new} *kɔ:3 tʰɔ:t3* [*he:5*]_{r-new}
 elder.male CONN cast cast.net
 ‘A man is casting a net.’ [YN.SF03.3]

¹⁹ This collocation has an obvious Thai origin, as indicated by each lexeme’s tonal numerical labels. The use of Thai forms in Moklen is discussed more in §8.5.

The point is that if we accept that *hɛ:5* ‘cast net’ in this verb-object pair is part of a functionally unitary construction, the identified ONIC violators fall more in line with Chafe’s conception of the informational limits of IUs. Taking on this view, we could then also account for the last two apparent ONIC violators. One in (5.33), which would now only have *təʔá:w* ‘sea’ as a “new” referring expression.

- (5.33) *ʔebá:p naʔék táŋ [təʔá:w]_{r-new} tʰɔ:tʰ hɛ:5*
 male.elder ascend from sea cast cast.net
 ‘The man is coming from the sea, having cast his net.’ [NK.SF19.60]

And the other in (5.34) where *tʰɔ:tʰ hɛ:5* is integrated as part of a serial verb construction along with *bɿj* ‘acquire’, leaving the clause final grammatical object of ‘three fish’ as now the sole “new” discourse referent.

- (5.34) \emptyset *tʰɔ:tʰ hɛ:5 bɿj [ʔeká:n tálɿj pʰɔʰ]_{r-new}*
 cast cast.net acquire fish three CLF
 ‘(He’s) cast a net and gotten three fish.’ [KG.SF03.6]

During annotation, an interpretation of these verb-object collocations as functionally unitary could have potentially been made purely on semantic grounds. However, the decision not to reflect the study’s focus on referring expressions and the prioritization of informational matters in framing linguistic structures. And consequently, that attention to these constructions was drawn out by the *r-new* information status tag served to highlight a functional unitary role. Therefore, having remained agnostic towards collocations during analysis served to point out these verb-object constructions later and allow acknowledgement to further aspects of Chafe’s model. While there are still some questions surrounding the use of the picture stimulus, conformity to one-new-idea constraint still stands as one informational aspect of Moklen clausal IUs.

5.3 Discussion

Speech from the Stolen Fish corpus was determined to be in line with the one-new-idea constraint. In other words, intonation units rarely had more than one new referring expression. This finding points to the low information load of most IUs and helps us to think more about the role of IUs across discourse. Chafe (1994) proposed that each substantive IU can be taken as a verbalization of a speaker’s singular

focusing of consciousness. Further, his one-new-idea constraint held that within the flow of naturally occurring speech, IUs will proceed in a manner whereby there are few cases in which an IU has two or more “separately activated new ideas”. Key for Chafe (1994) was seeing how IUs outlined an incremental flow of information. With this in mind, consider an excerpt from one Stolen Fish text. Here, the speaker is describing the scenes in which the old man first notices the fish are gone and then encounters the bird. Note, IU boundaries are indicated either by ‘...’ or by inclusion as a separately labeled example.

Stolen Fish NG text excerpt²⁰

(NG.81) *ʔebá:p p^hɔ:3t^haw3 t^hi:4 wa:4 nəlén pukát ... báj ʔeká:n*

elder.male old.person REL COM descend gillnet acquire fish
‘The old man who was gillnetting, and got the fish.’

(NG.82) *naʔék dín neŋjɔ:ʔ ... ná:ʔ lu:a5 dagà: pla:w2*

ascend come look exist remaining basket empty
‘(He) comes up and looks. All that’s remaining is an empty basket.’

(NG.83) *ʔeká:n ʔót ... caná:t nəlát*

fish depleted child steal
‘The fish are gone. The child stole (them).’

(NG.84) *dán já:j ʔacáw lát háh ... ná: ná:ʔ dán háh*

to.know COM who steal NEG 3SG yet know NEG
‘(He) doesn’t know who stole (them). He still doesn’t know.’

(NG.85) *p^hɔ:l di:l ... ticúm ... nám ʔeká:n*

at.same.time bird eat fish
‘At the same time, a bird, (it’s) eating fish.’

(NG.86) *k^ha:p4 ʔeká:n bá:ʔ dín*

to.hold.in.mouth fish carry come
‘(It’s) holding a fish carrying it.’

(NG.87) *p^hɔ:l: nan3 ... p^hú:ŋ wajlrun4 buján ʔón ʔa=bulát*

when then group teenager young.male give one=CLF
‘For that time, the group of young men gave (it) one of them.’

(NG.88) *t^hi:3 nỳj ... ticúm jú: bá:ʔ dín*

at DEM.PROX bird DEM.MED carry come
‘Right here, that bird is bringing (it) over.’

²⁰ The entire text is included as an appendix. Concerning the topic of changes in word-form (see §7), also note the lexeme and head noun *ʔebá:p* was *l-new&r-given-displaced*, and that subsequent uses of the lexeme in this excerpt are in the form of the monosyllabic alternant *bá:p*. Also note the verbal word-form shift for *nəlát* ‘steal’ across (NG.83)-(NG.84).

(NG.89) *dín tʰi:3 nɿj bá:p pʰɔ:3tʰaw3 pəná:ʔ*

come at DEM.PROX elder.male old.person see
 ‘(It) comes right here the old man sees.’

(NG.90) *já:j ticúm nɿj nəlát ʔeká:n pɔʔ*

think bird DEM.PROX steal fish 3SG
 ‘(He) thinks this bird stole his fish.’

(NG.91) *nəlát ʔeká:n bá:p pʰɔ:3tʰaw3 nə:4 nə:4*

to.steal fish elder.male old.person PRT PRT
 ‘(It’s) stolen the fish of the old man for sure!’

There are several things to note about information flow in this excerpt. First, we can acknowledge that clausal IUs need not match with sheer syntactic entailment. From the Chafean perspective, such disjunctures are likely a reflection of speakers activating “separate” ideas. For example, the utterance in (NG.81) marks the return of the old man after an extended absence within this text. The speaker last mentioned the old man 95 IUs prior; in terms of information status this referential expression was technically *r-given-displaced*, an intermediate category between *r-new* and *r-given* (see §4.4.4). Therefore, it is not surprising that reactivation of the old man required a separate phrasal IU. Note, then, in the following clausal IUs, omission of any overt referring expression for this now “given” discourse referent. Similarly, in (NG.85), an utterance spread across 3 IUs, we again see a discourse referent activated in a disjointed phrasal IU prior to its role in a following clausal IU. Here, *ticúm* ‘bird’ was also *r-given-displaced*, having last been mentioned 35 IUs prior. Although these detached phrasal IUs contain corresponding overt arguments for following clauses, in Chafean terms they were “inactive” discourse referents and, therefore, establishing reference to them was mentally distinct from relaying events in which they were involved. Again, the logic here is that activating “new” ideas takes up attention, and this in turn corresponds to space within intonation units. As argued by Chafe, intonation units offer insight into a relationship between the mind and language in that they are limited in the amount of “inactive” discourse referents and events that can be integrated within their bounds.

In this study, the one-new-idea constraint was operationalized as a specification on the number of “new” referring expressions (R-EXPs). Such a framing was similar to Matsumoto’s (2003) focus on NPs, the difference being phrasing with a

functional term rather than grammatical one. While adherence to the constraint was backed by findings on clausal IUs, perhaps more significant was that regardless of information status, the number of R-EXPs was normally limited to one. This first reveals the extent to which speech within the Stolen Fish corpus was about “given” referents, but a limit on overt arguments is also reminiscent of *Preferred Argument Structure*’s constraint limiting lexical arguments to one (see §3.4.2). For example, in the excerpt above, only two IUs have two overt R-EXPs. First, in (NG.87) for the trivalent predicate *ʔón* ‘give’ we see an argument for the agent ‘the young men’ realized as a full lexical expression, while the direct object, corresponding to ‘one of them’, is referred to through a classifier phrase *ʔa=bulàt*. The other IU (NG.90), however, did have two lexical arguments, though, it was an utterance explicitly describing what the old man was thinking about two salient discourse referents.

The general low information of most clausal IUs leads us to ask: What should we make of clausal IUs with more than two R-EXPs? Since these are a rarity, it is worth considering their properties and the contexts of their appearance. Within the entire Stolen Fish corpus, there were two clausal IUs marked as having 4 R-EXPs. These instances are shown in examples (5.35) and (5.36).

(5.35) *tʰi:3 nɔ̀j ʔaná:t nɔ̀j bá:ʔ ʔeká:n káw héʔ cʰuwíə̀k*
 at DEM.PROX child DEM.PROX carry fish go at fishhook
 ‘Here, the child is carrying the fish going to the fishing line.’ [CN.SF18.67]

(5.36) *tʰi:3 nɔ̀j wajlrun4 jú: bɔ̀j ʔeká:n lán ná:t*
 at DEM.PROX teenager DEM.MED acquire fish from child
 ‘Here, those teenagers got fish from the kid.’ [NG.SF15.91]

Both were initial utterances at the beginning of a brand-new scene, functioning as a sort of broad scene setting statement. They both start with *tʰi:3 nɔ̀j* literally ‘place here’, a referring expression indicating a point in the narrative. Crucial for understanding use of *tʰi:3 nɔ̀j* here is that the picture stimulus also anchors the narrative to an actual physical object speakers are describing. Both IUs have definite agents indicated with the determiner *nɔ̀j* ‘DEM.PROX’. And, both IUs have verbs selecting one core object argument, which happens to be *ʔeká:n* ‘fish’ in both examples. Finally, each IU ends with a non-core argument marked with a preposition.

No R-EXPs were marked *r-new*, and therefore neither instance is in violation of the one-new-idea constraint, but still they are outliers in terms of overtly encoded R-EXPs.

Clausal IUs with 3 R-EXPs were a bit more common but still comparatively rare (less than 2% of the corpus). One factor accounting for these is the role of clausal embedding, especially with reported speech/thought, as shown in (5.37) and (5.38) where use of *já:j* ‘to speak, think’ takes a clausal complement.

(5.37) *tɛ:2 caná:t nɛ̀j já:j ʔeká:n pɔ̀ʔ tit2 təmán*
 but child DEM.PROX say fish 3SG attach fishing.rod
 ‘But, the child says his fish is attached to the fishing rod.’ [NJ.SF22.88]

(5.38) *já:j pɔ̀ʔ káw nɔ̀léŋ ʔū:an hɛʔ ʔé:n nɛ̀j*
 say 3SG go descend gillnet at water DEM.PROX
 ‘(He) says he went gillnetting at the water here.’ [NG.SF27.184]

With clausal IUs with 2 R-EXPs (25% of the Stolen Fish corpus) we can begin to underline that the norm for IUs is *implicit* information. Intonation units with several overt arguments are attested, but IUs with fewer R-EXPs are more typical of the blocks with which Stolen Fish text are built. Furthermore, even with these IUs, omission of arguments in clausal IUs with 2 R-EXPs is still a common informational property. Like in (5.39) where a grammatical object was omitted or, in (5.40) where an agent was omitted.

(5.39) *ná:t tit2 təmán pɔ̀ʔ nɛ̀j*
 child attach fishing.rod 3SG PRT
 ‘The kid’s attached (it) to his fishing rod.’ [YG.SF24.76]

(5.40) *tʰɔ:t3 tá:ʔ mán ʔeká:n káʔ*
 remove fishhook fish PRF
 ‘(He’s) removed the fishhook from the fish.’ [UN.SF28.118]

These previous examples, (5.35)-(5.40), while notable for their quantity of overtly encoded information, are a minority, as 54% of clausal IUs were limited to one overt R-EXP and 19% did not have any R-EXPs at all. Therefore, when we consider that a significant portion of Stolen Fish texts unfolded through IUs with either 1 R-EXP (5.41) or none (5.42), we see that IUs normally underspecify discourse referents and

are likely shaped around other pressures for informational brevity, such as word-form reductions (see §7).

(5.41) *ká:n matáj ká?*

fish die PRF

‘The fish was dead.’ [NG.SF27.193]

(5.42) *kʰəná:ʔ lát*

undergo steal

‘(They) were stolen.’ [NK.SF19.62]

5.4 Summary

Findings from the Stolen Fish corpus confirmed Hypothesis 1 that clausal intonation units would conform to the one-new-idea constraint. Adherence to the constraint was operationalized as a limit on the amount of “new” referring expressions to one (Matsumoto, 2003), where “new” corresponds to the *r-new* ‘discourse new entities’ tag of the RefLex Scheme (Riester & Baumann, 2017). A few clausal intonation units were marked as containing more than one “new” referring expression; however, these were found to be the result of likely verb-object constructions, an aspect predicted by Chafe (1994). Accompanying findings here is a description of constructions used for the introduction of new referring expressions, along with a look at prosodic disjuncture and ellipsis. In discussing these findings, it was shown that although long syntactic strings of overt elements are attestable, they are rather infrequent and rely on certain informational conditions. Instead, the informational properties of clausal intonation units, and therefore the majority of discourse consists mostly of brief incremental communicative moves in small clausal structures.

6 Syntactic properties of intonation units

6.1 Introduction

Description of the syntactic properties of intonation units offers an account of clausal grammar within Moklen speech. To study the syntactic properties of Moklen intonation units, the following hypothesis was tested: variation in argument structure and associated grammatical constructions could be accounted for by the given-before-new principle. Data obtained through the Transitive Event Picture Sequences, however, reliably demonstrated that an AVO pattern was the preferred argument structure of transitive clauses in both givenness conditions. This finding first challenged the premise that there would be variant argument structures. And moreover, in opposition to the hypothesis, alternant word orders were not motivated by the given-before-new principle. Therefore, to further explore the potential for variant argument structures, clausal intonation units from the Stolen Fish corpus were also examined. Overall findings show that despite some minor variations, AVO is at the core of transitive clausal syntax, which again undercuts a premise of the hypothesis. To this point, aspects of Moklen grammar previously put forward as alternant orders (Larish, 1999; Swastham, 1982) are reexamined and found to be the byproducts of ellipsis and use of coreferential elements. Finally, additional examination argument structures shows that the given-before-new principle is still an attestable pattern within discourse.

6.2 Argument structure

To address Hypothesis 2: Variations in argument structure and associated grammatical constructions can be accounted for by the given-before-new principle, responses to target scenes of the Transitive Event Picture Sequences stimulus were examined. Findings here overwhelmingly point to AVO—with a variably overt or elided A argument—was the preferred argument structure in both givenness conditions. In terms of variation rather than there being alternate argument structures, apparent “non-AVO” responses reflect various discourse strategies used by participants in managing the task conditions, such as topicalization, ellipsis, and passivization. Findings from the Transitive Event Picture Sequences (TEPS) stimulus therefore show that while some deviation from a typical AVO clausal intonation unit

was elicited, variation in argument structure is likely rare and moreover not motivated simply by matters of givenness.

Patterns from the linear coding of clausal constituents were used to categorize the different types of responses. Agent-initial sequences elicited 185 valid responses, while patient-initial ones elicited 168.²¹ As shown in Table 23, the two most common response types displayed either an AVO or VO argument structure, which together constitute 70% of responses overall. The remaining 30% of responses are made up mostly of tokens from the patient-initial condition.

Table 23 Valid responses to TEPS target scenes

Description	Response type	A-initial	P-initial	Total	%
Preferred argument structure	AVO	90	55	145	41%
	VO	66	36	102	29%
Initial detachment	O...AV(O)	18	17	35	10%
	A...(A)VO	3	25	28	8%
Object ellipsis	AV	2	17	19	5%
Passive	(O)VAV	3	12	15	4%
Solitary verb	V	1	4	5	1%
Final detachment	VO...A	2	2	4	1%
Total		185	168	353	100%

AVO was the most common response type in both conditions. However, responses to the same target scene could still reflect semantic contrasts arising from their differential framing. For example, the speaker in (6.1) had already designated kinship relations among depicted human referents during the context scene, and therefore used *klák* ‘husband’ for the given agents of the target scene. In contrast, with the patient-initial version, the speaker of (6.2) used the general term *ma4nut4* ‘human’ as the depicted agents were new.

- (6.1) [klák]_{r-given} nɔc^hɔŋ babú:j
 husband carry.with.partner pig
 ‘The husbands are carrying a pig.’ [LI.A12.2]

²¹ Agent-initial sequences elicited 7 invalid responses. Patient-initial sequences elicited 24 invalid responses, a result suggesting some difficulties with a patient-initial discourse context. Limitations of the stimulus are discussed in §8.5.

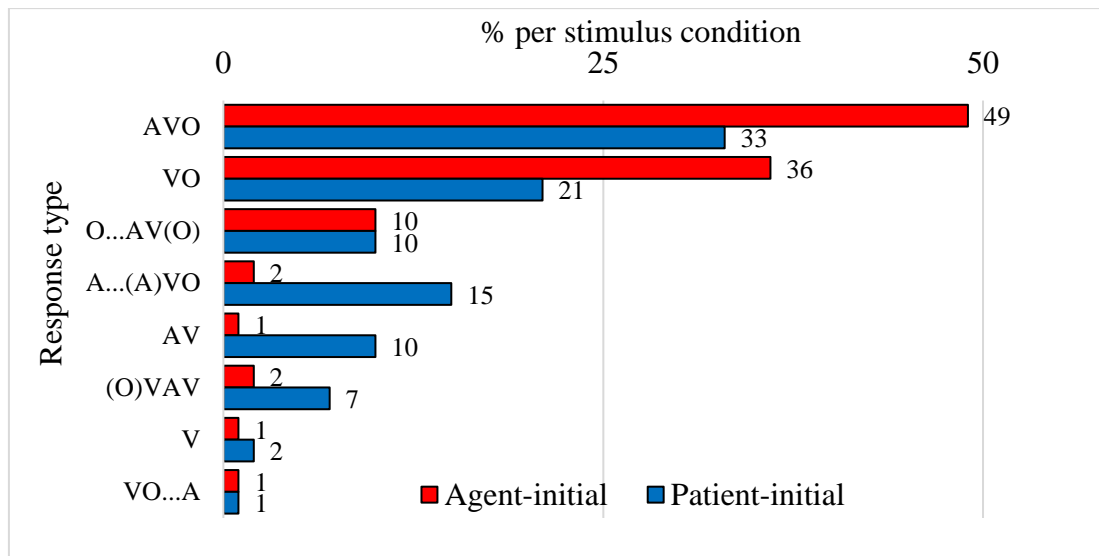


Figure 30 Percentage of response types per target scene condition

Also evident in Figure 30, is the fact that most of the other response types were elicited under the patient-initial condition. In looking at these responses, two things should be kept in mind. Firstly, responses here represent an overall minority of valid responses to the target scenes (30%) (see Table 23). Secondly, rather than providing reliably attested patterns of clausal grammar—like AVO—these responses reflect more general discourse strategies used by speakers in managing the task conditions. Below each of the response types are discussed in order of most to least attested, i.e., descending through the categories within Table 23 and Figure 30.

A strategy of initial detachment or topicalization with phrasal IUs characterizes an aspect of the response type O...AV(O). Use of phrasal IUs for establishment of referents like this is common in other languages (see §3.5.2). The core feature of this category is the speaker's activation of the patient before a validating clausal predicate. This pattern was 10% of responses under both conditions. In the patient-initial condition the utterance has a sense of being about the preposed object, such as in (6.5) where *má:k* 'house' is the grammatical object of the verb *nama:ʔ* 'enter'.

- (6.5) [má:k ni:3]_{r-given} ... ticúm mén nama:ʔ
 house this bird IRR enter
 'The house, a bird's going to enter.' [NN.P16.2]

In contrast, in an agent-initial condition initially detached objects are new and therefore show the speaker first identifying the patient before uttering the transitive

clause. Important to note for these patients is that while they are appropriate arguments for the following verb they were not within the bounds of the same clausal IU. Such as in (6.6) where *p^hɔ:3t^haw3* ‘elder’, the patient of the verb *nəkó:t* ‘bite’, is clearly a distinct phrasal IU with its own intonational contour and a .6 sec pause before the following clause, see Figure 31.

- (6.6) [*p^hɔ:3t^haw3*]_{r-new} ... *ʔólá:n nəkó:t*
 old.person snake bite
 ‘An old person...the snake bites (them).’ [NJ.A17.2]

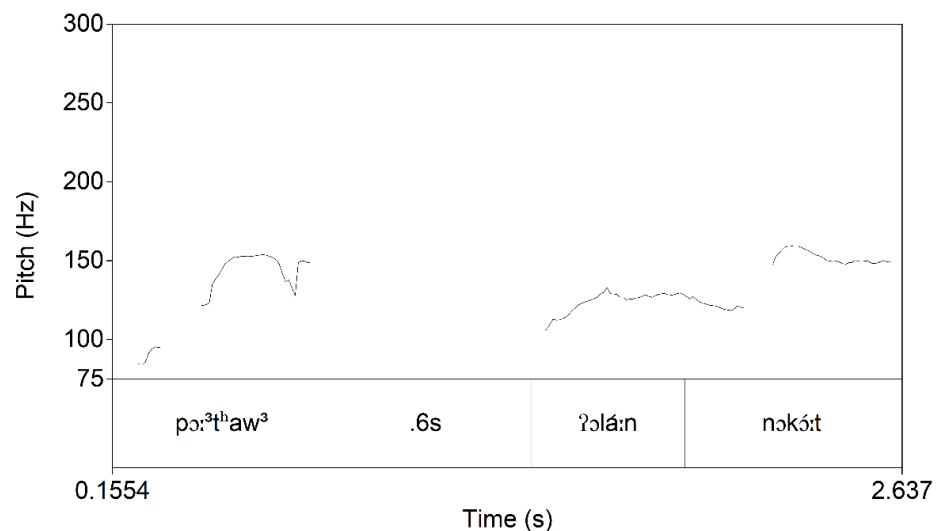


Figure 31 Initially-detached patient in preposed phrasal IU

Grammatically, responses categorized as initial detachment are not a homogenous type. Take example (6.7) where the new patient *ká:n* ‘fish’ was activated in a separate clausal IU, before the speaker uttered a second IU with the qualifying predicate (verbs *k^hi:p3* ‘grab’ and *ʔá:k* ‘place’) for which *ká:n* was the appropriate grammatical argument. Therefore, the qualifying clausal IU actually had an AV pattern.

- (6.7) *ká:n ni:4 p^hlèt ká?* ... *kətá:m k^hi:p3=ná: ʔá:k ne*
 fish this fall PRF crab grab=3.SG place PRT
 ‘This fish has fallen; the crab grabs and places (it).’ [YG.A14.2]

Or consider (6.8) where a patient *p^hɔ:3t^haw3* ‘old person’ again was activated first, however the qualifying predicate is a grammatically embedded complement clause

and object of the main verb *dán* ‘know’. So, while in relation to framing of the stimulus the “patient” came first, grammatically the overall argument structure is AVO.

- (6.8) *p^hɔ:3t^haw3 dɔ:k dán háh ʔólá:n bətək Ø*
 old.person sit know NEG snake strike

‘The old person is sitting not knowing a snake’s bit (her).’ [WN.P17.2]

Further emphasizing that object fronting is an artifact of the discourse context is that the fronted patients were sometimes repeated within the qualifying clause, like in (6.9) or (6.10). All of which again serves to underscore a preference for AVO.

- (6.9) *ká:n ... ʔebá:p káw nəmán b́j ʔeká:n*
 fish elder.male go fish.with.rod acquire fish

‘The fish. A man’s gone fishing and got the fish.’ [YN.P5.2]

- (6.10) *ʔəmá:k ... kaʔé:w nəkí:t ʔəmá:k baliək*
 home wood fall.down home break

‘A house. The tree falls down smashing the house.’ [LW.A5.2]

Initial-detachment of agents, a A...(A)VO response type, occurred more with patient-initial sequences. These instances represent the topicalization or pre-posing of agents. With these instances the speaker first identifies the new agent referent before uttering a transitive clause, however subsequent overt agents in the qualifying clause were variable. For example, in (6.11) the agent *máʔ* ‘mother’ was repeated but the agent in (6.12) *ʔólá:n* ‘snake’ was not. Also, note for (6.11) the choice of a kinship term as for the new agent *máʔ* ‘mother’ is an inference based on the patient-initial context scene depicting a sleeping baby.

- (6.11) *máʔ ... máʔ napót cəwát p^ha:3 hom2*
 mother mother cover cloth blanket

‘Its mother. Its mother is covering (it) with a blanket.’ [EW.P7.2]

- (6.12) *ʔólá:n ... Ø nəkó:t bú:m p^hɔ:3t^haw3*
 snake bite elder.female old.person

‘A snake. (It) bites the old woman.’ [CU.P17.2]

Several factors may explain the prevalence of initial-detachment of agents in a patient-initial condition over the agent-initial (15% and 2% respectively). Perhaps as

AVO is the preferred argument structure speakers feel the need to activate agents before incorporating them into a clausal IU. Agents' newness under the patient-initial context therefore results in a slight lag before speakers produce a validating clause IU. Another factor to consider is potential influence stemming from the visual composition of the target scenes, such as the size, salience and relative positioning of new human referents. Given the nature of the picture-based discourse context it is essential that we continually acknowledge such factors, however, since none of this was rigorously accounted for in the design of the stimulus it is beyond further speculation. The main point is that responses featuring initial-detachment, whether A or O arguments do not constitute any alternate "word orders". And moreover, they offer further examples of AVO acting as the chief syntactic pattern of clausal IUs.

Other informational salient effects arising from the task conditions are visible with less attested response types. For example, the cause for patient ellipsis (AV responses) is easier to interpret, given the relative disparity between agent-initial and patient-initial conditions, $n=2$ and $n=17$ respectively. Clearly, the previous activation of patients in the context scene drives their ellipsis in the following the target scene, like in (6.13) and (6.14).

(6.13) *ʔɔ́lɑːn tək Ø lɔːj1*
 snake strike INTS
 'A snake's struck (her)!' [LIP14.2]

(6.14) *təmán tit2 =nɑː Ø*
 fishhook stick=3.SG
 'A fishhook's stuck (him).' [DW.P22.2]

For comparison consider responses from agent-initial versions of the same sequences in (6.15) and (6.16) which did elicit overt new patients.

(6.15) *ʔɔ́lɑːn bətək ʔibúːm niː4 lɛːw4*
 snake strike elder.female DEM.PROX PRF
 'The snake has struck this woman.' [TG.A17-2]

(6.16) *matáːʔ cʰuwiák tábáːʔ=nɑː ɲán*
 fishhook pierce=3.SG hand
 'The fishhook is piercing (his) hand.' [NJ.A24-2]

One response type where we do see an inversion of core arguments is with adversative passive constructions. The prevalence of these in the patient-agent condition is also more easily accounted for. Adversative passives use the verb *kʰəná:ʔ* ‘undergo’ and are common in contexts where an experiencer undergoes an unfavorable action.²² For example in (6.17), *caná:t* ‘child’ is the grammatical object of the verb *bətək* ‘strike’ but has moved to a clause initial position. With the adversative passive construction the sense is that the patient has undergone the whole event of being stung by a scorpion.

- (6.17) *caná:t kʰəná:ʔ ... tʰuəjɪmanɪ bətək*
 child undergo scorpion strike
 ‘The child’s getting... stung by a scorpion.’ [LP.P18.2]

A similar example in (6.18) again shows how the framing of patient-initial sequences prompts the adversative passive interpretations, where *kanáj* ‘male’ occupies the position of grammatical subject. Compare this example to the agent-initial framing in (6.16), above.

- (6.18) *kanáj cʰəná:ʔ matá:ʔ cʰuwiək tábá:k*
 male undergo fishhook pierce
 ‘The man’s gotten stuck by a fishhook.’ [WN.P24.2]

An interesting aspect of the passivizer is that it also has a monosyllabic alternant form *ná:ʔ* ‘undergo’, like in (6.19). Note here the appearance of monosyllabic alternant appears at an IU-initial position (see §7). Another informational aspect of passive constructions is that they also may have ellipsed grammatical subjects when the patient is “given”, as shown in (6.19) and (6.20).

- (6.19) \emptyset *ná:ʔ tʰuəjɪmanɪ bətək*
 undergo scorpion strike
 ‘(She’s) getting stung by a scorpion.’ [TW.P18.2]

²² Adversative passives are a feature of many Mainland Southeast Asian languages (Prasithratsint 2004). As pointed out by Larish (1999), *kʰəná:ʔ* is likely a loan from Malay (See also Nomoto and Wahab, 2012). Note, also the variant initial consonant in the form *cʰəná:ʔ*, an issue across the lexicon (see §2.4.1). Larish (1999) also reports attestations with either initial consonants.

(6.20) Ø *kʰəná:ʔ cʰuwiáʔ kiaw2 ɲán*
 undergo fishhook to.hook hand

‘(He’s) gotten hooked by the fishhook (on his) hand.’ [CU.P24.2]

Clarifying the underlying semantics of the passivizer we can also acknowledge its transitive sense meaning ‘hit, come into contact’ (Larish, 1999), like in (6.21) and (6.22) where speakers used it to describe a falling coconut impacting a grammatical object *kəcók* ‘bottle’ thereby breaking it. Such usages were not counted as passives.

(6.21) *niʔú:n kəpʰlâ:t kʰəná:ʔ kəcók*
 coconut fall hit bottle

‘The coconut falls hitting a bottle.’ [CU.A20.2]

(6.22) *nɲ̩ niʔú:n bəluáŋ ... kəpʰlâ:t kʰəná:ʔ kəcók bəkáh*
 DEM.PROX coconut be.fresh fall hit bottle break

‘Here, a fresh coconut. (It) falls and hits the bottle breaking (it).’ [NK.P20.2]

While the adversative passive construction promotes the patient to a grammatical subject the clear semantics of suffering a negative experience suggest that the best analysis is for clauses with passive *kʰəná:ʔ* to be viewed as an intransitive verb which can potentially take a complement clause as an argument (S-Vi-CoCl:E) (Dixon 2009). This in turn better accounts for how *kʰəná:ʔ* was used within the agent-initial condition, where the interpretation salient to the speaker was the adversative experience of a new human referent, such as in (6.23) or (6.24).

(6.23) *caná:t dɔ:k kʰəná:ʔ tʰuəj1man1 bətək*
 child sit undergo scorpion strike

‘A child sitting and getting stung by the scorpion.’ [WN.A18.2]

(6.24) *caná:t kʰəná:ʔ cʰuwiáʔ ... héʔ ɲán*
 child undergo fishhook at hand

‘A child’s got (stuck) by the fishhook, at (his) hand.’ [EW.A24.2]

Some valid responses while bivalent were low on a transitivity scale (Thompson, 1997). As these were not distinguished in the overall analysis (see §4.4.4), it should be pointed out that the preferred argument structure for these in terms of Dixon’s (2010) marking of core arguments is S-Vi-E, where E serves as an extended argument to the core of an intransitive verb. Examples from the TEPS data

include (6.25) where although *təkát* ‘tree.stump’ is the cause of the designated patient’s tripping, the verb *saʔ2dut2* ‘to trip’ only has an intransitive sense.

- (6.25) *bá:p* *p^hɔ:3t^haw3* *saʔ2dut2=na:* *təkát*
 elder.male elder trip=3.SG tree.stump
 ‘An old man trips on the stump.’ [YN.A23.2]

Another example of an SVE clauses is shown in (6.26) where the verb in *napóʔ* ‘to perch’ selects for an argument of location, in this instance *kón kaʔé:w* ‘a tree’ but there is no sense of affectedness on the E argument.

- (6.26) *manók* *napóʔ* *kón* *kaʔé:w*
 chicken perch tree wood
 ‘The chickens perch on a tree.’ [EW.A15.2]

Arguably, there are limitations to the framing of discourse and argument structure within the context of TEPS stimulus. Still that (A)VO was 70% of valid responses clearly establishes it as a preferred argument structure. However, there was some slight variation between the two conditions. For one, conformity to an overt AVO was higher in the agent-initial condition than in the patient-initial one, 49% and 33% respectively. Also, the higher number of adversative passives within the patient-initial condition provide some evidence for the patient-initial condition influencing the semantics and syntax of a response. Despite these differences, the preferred argument structure for clausal IUs within the TEPS data clearly points to an AVO (or SVE) order. Importantly, no clear instances of variant argument structures were elicited. Instead with other response types, we see that discourse strategies such as initial detachment, ellipsis, or passivization were used to manage the task conditions, and furthermore even within these strategies preferred argument structures are the main syntactic patterns.

6.2.1. Variation in argument structure

To further study the syntactic properties of Moklen clausal IUs, in this section I draw upon the *Stolen Fish* corpus to examine the potential for variant argument structures. Results of this examination show that, like the TEPS data, clausal IUs largely follow preferred argument structure patterns of AVO and SVE. And while there is some minor variation in the patterning of clausal constituents, such as number

of verbs and argument relations, there is virtually no attestation for alternant orders of core arguments within the Stolen Fish corpus.

Clausal IUs with at least two overt referring expressions are needed to display the orderings of multiple grammatical arguments. Presented within Figure 32 are categories reflecting the ordering of clausal constituents in clausal IUs with two overt referring expressions (R-EXPs). These constitute 25% of clausal IUs within the corpus, as 73% of clausal IUs within the Stolen Fish corpus either have only one overt referring expression or none (see Table 20). Corroborating findings about preferred argument structure in §6.2 note that AVO (and SVE) is more attested than other orderings.

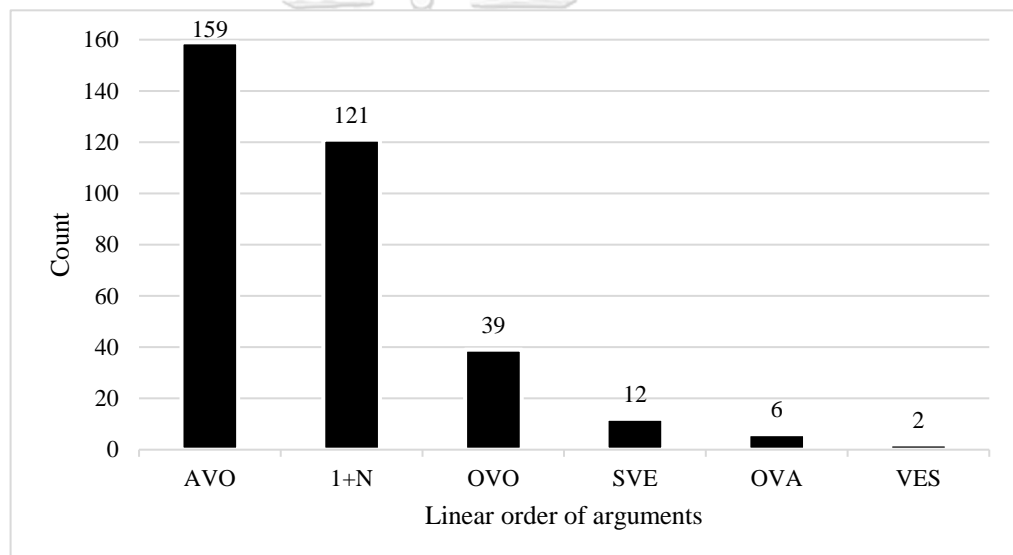


Figure 32 Linear order of arguments in clausal IUs with two R-EXPS

Nearly all argument structure patterns in Figure 32 have at least one verb interspersed between each of the two arguments. However, tokens within each category also display variable positioning and number of verbs across the intonation unit. For example, in (6.27) the verb *lát* ‘steal’ cements an AVO structure, but the verbs *bá:ʔ* ‘carry’ and *dəbút* ‘run’ also appear after an O argument *ʔeká:n* ‘fish’ as part of a serial verb construction.

(6.27) *caná:t lát ʔeká:n bá:ʔ dəbút*
 child steal fish carry run

‘The child steals the fish, carries it and runs.’ [NK.SF06.13]

Categories in Figure 32, therefore, are mainly descriptive in terms of the ordering of overt R-EXPs but are under-specifications of the occurrence of all clausal elements. So, even though most tokens attest the same preferred argument structures there is still room for some grammatical and semantic variation. Consider (6.28) where an AVO clause is a grammatical object and complement clause of an initial verb *já:j* ‘say, think’.

- (6.28) *já:j ticúm nỳj nəlát ʔeká:n pǝʔ*
 think bird DEM.PROX steal fish 3SG
 ‘(He) thinks the bird stole his fish.’ [NG.SF20.140]

Along with core arguments A,S,O,E (Dixon, 2010), there are also non-core arguments. The difference between core and non-core (N) can sometimes be subtle (Thompson, 1997), but within the Stolen Fish corpus, non-core arguments typically refer to locations and so use of a *preposition* was a key criterion. A prototypical example is provided in (6.29) where the single referring expression *ʔəmá:k* ‘house’ is preceded by the preposition *hɛʔ* ‘at’ and moreover is a non-core argument not selected for semantically by the verbal predicate *nəmán* ‘fish with rod’.

- (6.29) *nəmán hɛʔ ʔəmá:k*
 fish.with.rod at house
 ‘(He’s) fishing at home.’ [YN.SF01.2]

In (6.30) the argument *tamán* ‘fishing basket’ is more closely connected to the verbal predicate, but as it was preceded by the preposition *dalám* ‘in’ and therefore was classified as a non-core argument.

- (6.30) *ʔá:k dalám tamán*
 to.place in fishing.basket
 ‘(He) places (it) inside the fishing basket.’ [NJ-SF29.109]

The second most frequent type of clausal IUs with two R-EXPs within Figure 32, was the category “1+N”. This category served as a catch all for any clausal IUs with a single core argument plus one non-core argument. For example, in (6.31) the non-core argument precedes a core O argument, while in (6.32) the non-core argument comes after an initial S argument. As non-core arguments are outside the core layer of clausal syntax, their position either before or after a core argument was

not discriminated. Also shown in these examples is how demonstratives often took on the role of non-core arguments through pronominal reference to a location.

- (6.31) *t^huŋ1 dè:ʔ paʔ2 kanáj təlǰj lùj lá:j*
 arrive DEM.DIST encounter male three CLF again
 ‘Once (he) gets **there** then (he) encounters three men.’ [LN.SF10.32]

- (6.32) *ná:t didú:n latá: nǰj*
 child sleep on DEM.PROX
 ‘A kid is laying up **here**.’ [TG.SF17.77]

It is worth noting that based on a criterion of preposition use some verbs variably select for either core or non-core arguments, such as the verb *lut2* ‘slip out’. In (6.33) we see it takes a non-core argument *ŋán* ‘hand’ as it is marked by the preposition *láj* ‘from’. However, in the slightly different context of (6.34), *ŋan* ‘hand’ was considered a core argument E as there was no preposition and therefore was regarded as being directly selected for by *lut2* ‘slip out’.

- (6.33) *ká:n lut2 láj ŋán*
 fish slip.out from hand
 ‘The fish fell from his hand.’ [NK.SF12.36]

- (6.34) *nəmóh ʔeká:n lut2 ŋán*
 fall.down fish slip.out hand
 ‘(He) falls over and the fish slips out of his hand.’ [NG.SF12.68]

Given such a possibility, there is potential for some intransitive verbs to generate either an SVE argument structure or one of SV+N. Semantically, one might argue that when a referring expression is a core argument it is more salient or more closely integrated with the verbal concept and non-core ones less so. However, such a contrast may reflect more general discourse factors rather than semantic/syntactic features intrinsic to the verb. For example, consider differential argument selection with the verb *ʔé:m* ‘be located’. In (6.35) *ʔé:m* takes a core argument E *pəkón niʔú:n* ‘coconut tree’ to demarcate the spatial boundaries of which the grammatical subject *caná:t* ‘child’ is residing. In other words, the grammatical subject is emphasized as actively enacting the verbal concept in relation to the E argument.

- (6.35) *caná:t ʔé:m ɸəkón niʔú:n*
 child be.located tree coconut
 ‘The child’s located at a coconut tree.’ [WN.SF04.16]

But in (6.36) *ʔé:m* takes on the preposition *tá:* ‘on’, indicating the incidental relative positioning of the subject to the non-core argument *kón kaʔéw* ‘tree’. So, while there is no major difference in meaning for the verbal concept *ʔé:m*, differential focus at a discourse level may drive selection for core arguments.

- (6.36) *ʔá=bó:ʔ ɸóʔ ʔé:m tá: kón kaʔéw*
 one=person 3SG be.located on tree wood
 ‘One of them, they are on the tree.’ [YN.SF17.54]

All and all semantic differences of these verbs in clauses, if any, are extremely subtle, however, they still illustrate some variable relations within clausal syntax. Additional examples of variability is available with the verb *baléh* ‘return’. In (6.37), *baléh* ‘return’ selects for a core S argument, an ellipsed grammatical subject returning to a particular location, but grammatically the clause also features a non-core argument ‘his fishing rod’.

- (6.37) *baléh dín héʔ ... tʰi:3=mán ɸóʔ lá:j*
 return come at NOM=fish.with.rod 3SG again
 ‘(He’s) returning back to...his fishing rod again.’ [NA.SF18.52-53]

Or like in (6.38) it can be a subordinate part of a transitive predicate *báj* ‘acquire’ selecting for the O argument ‘three fish’.

- (6.38) *tʰi:3 nɿj kə:3 ɸóʔ báj ʔeká:n təlɿj ɸʰóh baléh*
 at DEM.PROX CONN 3SG acquire fish three CLF return
 ‘Here, well, he gets the three fish back.’ [CN.SF14.46]

But acting as its own predicate *baléh* ‘return’ can also select for a specific core E argument *ʔomá:k* ‘home’, in an idiomatic construction meaning to ‘return home’, like in (6.39). Again, the core verbal semantics is not drastically different, but its relation to clausal constituents have changed, all while preferred argument structures of either (A)VO or (S)VE are maintained.

- (6.39) *nɔ̀j bá:ʔ ʔəká:n ʔá=bulàt baléh má:k*
 DEM.PROX carry fish one=CLF return home
 ‘Here, (he’s) carrying the one fish going home.’ [GP.SF18.39]

Double object constructions OVO (see Figure 32) were a clausal IU type with two O arguments. For these clausal IUs the omitted A argument is a previously mentioned discourse referent, and therefore they represent an underlying (A)VOVO structure. Semantically double object constructions come about when two consecutive and closely related events are conveyed as one broader event within a single IU. Built upon serial verbs, each object is closely related to the overall event in a stepwise fashion, like in (6.40) and (6.41).²³

- (6.40) \emptyset *mén ʔəká:n kiaw2 cʰuwiák*
 take fish hook fishhook
 ‘(He’s) taking the fish, hooking it to the fishhook.’ [YT.SF28.81]

- (6.41) \emptyset *pók cʰuwiák ɲó:ʔ pá:n*
 lift.up fishhook look.at bait
 ‘(He’s) lifting up the fishhook, looking at the bait.’ [NK.SF02.006]

Despite all their potential variations, clausal IUs with two overt referring expressions largely reflect AVO and SVE patterns. Semantically, while predicates and argument relations may shift into various constructions, again preferred argument structures remain at the core of clausal syntax.

6.2.2 Co-referring expressions

Figure 32 indicated there were 8 token orders of OVA and EVS. At first, in terms of direct surface coding of referring expressions’ relationship to the clause, these instances might seem like variant argument structures. A closer look, however, reveals the role of remnant co-referring expressions, namely: verbal enclitics, a reciprocal pronoun, and clause-final classifier phrases. However, rather than constituting variant order of clausal constituents or “word orders” (cf. Larish, 1999; 2005), these instances are better understood as a byproduct of the omission of lexical arguments.

²³ Grammatically, Moklen serial verb constructions share many similarities with Thai ones (See Diller 2006; Iwasaki and Ingkaphirom 2005)

Larish (1999) describes Moklen clitic pronouns as occurring to the right of verbs and being co-referential to grammatical subjects. Important for the analysis here is differences with full lexical arguments. With the verbal enclitic *ɲá:* ‘3.SG’ in (6.42) we can see how this IU generated an OVA coding of clausal arguments, as *ɲá:* was the only overt information for an A argument. And since the O argument *təmán* ‘fishing rod’ preceded it, a strict interpretation would suggest this constitutes an OVA order.

(6.42) *dúk təmán kɪ:aw2=ɲá:*
 put fishing.rod to.hook=3.SG

‘(He’s) putting (it) on the fishing rod, hooking (it).’ [TG.SF23.114]

Calling verbal enclitics like *ɲá:* ‘3.SG’ “co-referential elements” is based on the fact that they often appears alongside antecedent lexical arguments, like in (6.43) and (6.44) where *ɲá:* is coreferential with the grammatical subject of each clause.

(6.43) *caná:t bá:ʔ=ɲá: ʔeká:n káw má:k pɔʔ*
 child carry=3.SG fish go home 3SG

‘The child’s carrying the fish going to his home.’ [YN.SF18.60]

(6.44) *ʔeká:n ɲám=ɲá: pá:n pɔʔ met2*
 fish eat=3.SG bait 3SG be.depleted

‘The fish have eaten all his bait.’ [YG.SF02.3]

Instances of *ɲá:* as an independent pronoun were rare and are seemingly reserved only for grammatical subjects, that is A and S core arguments, like in (6.45) and (6.46).

(6.45) *ɲá: ná:ʔ dán háh*
 3SG yet know NEG

‘He still doesn’t know.’ [NG.SF19.130]

(6.46) *tɛ:2 ɲá: bɣj mán háh*
 but 3SG to.acquire fish.with.rod NEG

‘But he didn’t catch (it).’ [TG.SF23.120]

In contrast, the third person singular pronoun *pɔʔ* regularly occupies various grammatical positions, such as the A argument of a ditransitive clause in (6.47) or an indirect object in (6.48). Additionally, *pɔʔ* ‘3.SG’ is typically the pronoun of choice in possessive constructions, see (6.43) and (6.44) above.

- (6.47) *pɔʔ ʔɔn ʔɔlá:ŋ ní: təlɔ́j lùj*
 3SG give person DEM.PROX three CLF
 ‘He gave (it) to these people, the three of them.’ [TG.SF27.149]
- (6.48) *ʔɔn pɔʔ bé*
 give 3SG PRT
 ‘Give (it) to him.’ [LN.SF15.56]

During the coding of arguments (§4.3.4), a rule was adopted of not counting co-referring expressions as additional arguments when a corresponding lexical argument was present in the same IU. So, for a clause like (6.49), *na:* ‘3.SG’ was not counted as an additional A argument as the corresponding lexical argument *cúm* ‘bird’ was also present. Instead, this clause was regarded as exhibiting an overt AV structure with an ellipsed O argument.

- (6.49) *cúm lát=na: Ø met2*
 bird steal=3.SG be.depleted
 ‘The bird’s stolen (it) all.’ [DW.SF19.59]

However, if a corresponding lexical argument was *not* present within the same IU, then a verbal enclitic would be counted as an “argument” as it still signaled overt information within the clausal IU. This informational caveat to coding co-referring expressions, therefore, sometimes meant that if lexical arguments were omitted, remnant overt elements could generate a “variant order” of clausal constituents. Such as in (6.50), where the S argument is only overtly represented by *na:* ‘3SG’, or in (6.51) where the A argument is referred to by *lá:ŋ* 3.PL. Consequently, in terms of the coding of overt elements, these verbal enclitics produced variant VSE or VAO orders.

- (6.50) *dut2=na: batɔ́j*
 trip=3.SG rock
 ‘He tripped on the rock.’ [YG.SF12.30]
- (6.51) *mók=lá:ŋ ka:n set2*
 cook=3.PL fish be.finished
 ‘They’ve finished cooking the fish.’ [HJ.SF26.117]

Verbal enclitics are precisely what led Larish (1999, p.212) to put forth a “VSO word order” for Moklen. However, an alternant interpretation is that such instances are

more of a result of ellipsis of lexical arguments, rather than the basis of a major syntactic operation. Consider that when it comes to clausal IUs with full lexical arguments, the pattern is overwhelmingly AVO such as in (6.52).

- (6.52) *nəŋé:n ... ticúm k^ha:p3 ʔeká:n*
 chase bird hold.in.mouth fish
 ‘(He’s) chasing...the bird biting onto the fish.’ [KG.SF22.71-72]

A co-referring expression like *na:* ‘3.SG’, however, has a unique position post-verbally and variably co-occurs in clauses with full lexical arguments, like in (6.53).

- (6.53) *ticúm k^ha:p3=na: ʔeká:n*
 bird hold.in.mouth=3.SG fish
 ‘The bird’s biting onto the fish.’ [YN.SF21.69]

But while lexical arguments are often omitted, a verbal enclitic may variably still occur in the post-verbal position, like in (6.54). In other words, clauses where the only overt information for a core argument is a verbal enclitic does not indicate a shifting position of core argument slots within the grammar but instead they are just remnants after the ellipsis of lexical arguments.

- (6.54) \emptyset *k^ha:p3=na:* \emptyset *káw nɛ:*
 hold.in.mouth=3.SG go PRT
 ‘(It’s) biting onto (it) and going.’ [TG.SF22.111]

The role of pronominals is featured in both Larish’s (1999) and Swastham’s (1982) claims for variant word orders in Moklen. For Swastham pronouns appear frequently in examples, while Larish explicitly acknowledges the use of clitics. The key question then is whether these pronominal clitics are substantive or not for variant argument structures.²⁴ Furthermore, there do not seem to be not any noticeable discourse factors that can help account for them. One anecdotal exception, however, was (6.55) where the speaker was giving voice to the character of the thief as he *emphasized*—in an effort to fool the fisherman—that the fish on his line was a fish

²⁴ Additionally, there is the question of clitics as a linguistic category (See Haspelmath (2015), or the question of whether clitics can be considered a part of syntax or morphology (Gaglia and Schwarze 2015). Influencing the choice to not view these enclitics as full arguments comes from arguments by Himmelmann (2014) who points out that postposed function words are more likely to be grammaticized function words. However, because of the focus on information structure they were viewed as still representing overt clausal information.

that he himself had just caught (see Figure 33). What is interesting to note here is that this is reported speech provided in the first-person, and so there is not the same case for supposed ellipsis of a prior lexical argument.

- (6.55) *caná:t já:j ʔebá:p ... ʔeká:n nəmán cǎj já:*
 child say old.man fish fish.with.rod 1SG just
 ‘The child says, “Sir...This fish, I JUST caught (it).” ’ [YN.SF22.76-77]



Figure 33 Stolen Fish scene #24 ‘Nothing to see here.’

Notably, the second IU in (6.55) features an OVA order and all constituents were deemed to be within the bounds of a single intonation unit (see Figure 34). Additionally, the pronoun *cǎj* ‘1SG’ also appears to be more of an independent morpheme and not a verbal enclitic.

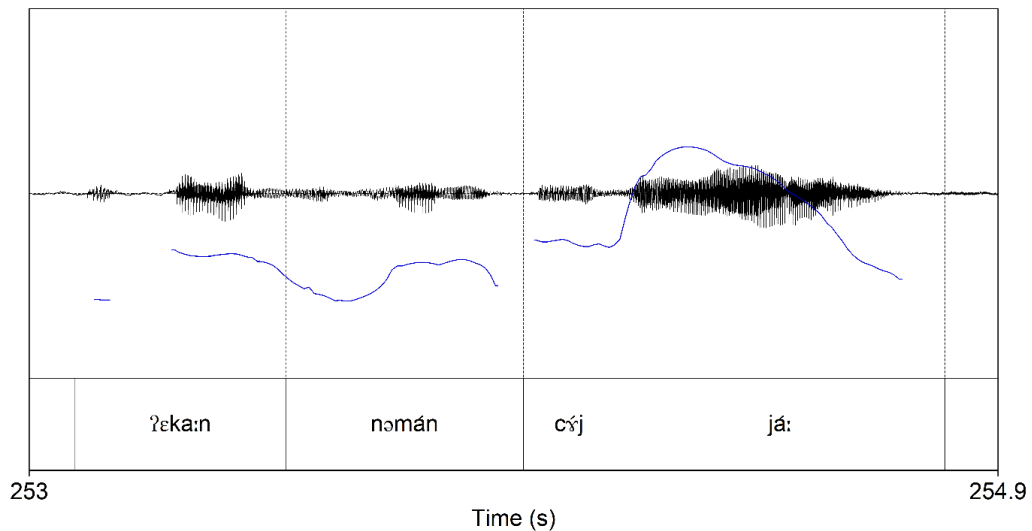


Figure 34 Waveform attested OVA order.

Considering this IU’s OVA order and its role within the narrative—that is the clear emphasis being placed upon the O argument *ʔeka:n* ‘fish’—it seems that (6.55) could possibly stand as a lone attestation of a variant argument structure. A contrast to the OVA order is also provided by the same speaker elsewhere in the same text when they provided the same constituents in canonical AVO order, as shown in (6.56).

- (6.56) *hm kɔ:3 cʔj nəmán ʔeká:n bʔj lá:j*
 VOC CONN 1SG fish.with.rod fish acquire again
 ‘Hm, well, I am fishing again.’ [YN.SF28.96]

While in §6.2 it was shown that variant argument structures were not elicited from marked givenness conditions, perhaps variant argument structures are used for other forms of emphasis, such as the clear emphasis on the fish in (6.55). This at least would fall in line with Swastham’s (1982) citation of “emphasis” as the cause for variant word orders. Still though, these orders are extremely rare, and pronominal verbal enclitics most often behave in a manner distinct from what we find with full lexical arguments with respect to clausal grammar.

Other co-referring expressions provide analogous cases of the same type of phenomenon discussed for verbal enclitics. These were also responsible for other tokens coded alternant orders. For example, there is the reciprocal pronoun *ʔabɔ:ʔ*

‘together’.²⁵ Co-referential with plural subjects *ʔabɔːʔ* appears post-verbally as an adverbial indicating a joined action of multiple participants, such as speaking together (6.57) or roasting a fish together (6.58).

(6.57) *bináj duwàːʔ lúj dɔːk kláːw ʔabɔːʔ*
 woman two CLF to.sit speak together
 ‘Two women are sitting speaking together.’ [NK.SF07.17]

(6.58) *pʰúːŋ náj kɔː3 məʔiáŋ ʔekáːn bɔːʔ pʰə*
 group TOP CONN roast fish together PRT
 ‘This group, well, (they’re) roasting fish together.’ [NG.SF17.117]

As a co-referring expression, again the rule was to not count *ʔabɔːʔ* ‘together’ as an additional referring expression unless it was the only overt core argument information within the clausal IU. This again led to a situation wherein if the lexical argument was omitted or not integrated into the same IU, the order of clausal elements could appear as OVA. Such as (6.59) with no lexical A argument, while for (6.60) the corresponding lexical arguments were not prosodically integrated into the same clausal IU.

(6.59) *Ø báj ʔekáːn ʔa=bulát báːʔ pəniáŋ ʔabɔːʔ láːj*
 acquire fish one=CLF carry roast together PROG
 ‘(They’ve) gotten one fish and are bringing (it) to roast together.’ [YT.SF17.44]

(6.60) *mináj ...kanáj ... kamllaŋl jám ʔekáːn ʔabɔːʔ*
 female male PROG eat fish together
 ‘The women, the men. (They’re) eating fish together.’ [KG.SF26.89]

Therefore, while in a strict literal sense of overt information these intonations units could be cast as having “A-final orders” but given the function of *ʔabɔːʔ* and its specific syntactic position, it easy to see such instances do not actually constitute a distinct argument structure, but instead are byproducts of omitted lexical arguments and/or a lack of prosodic integration.

Another instance that produces similar variant coding of clausal IU argument orders is classifier phrases. Classifier phrases offer a definite quantity of a previous

²⁵ There is another sense for *ʔabɔːʔ* that means ‘one person’. wherein *ʔa-* is a proclitic meaning ‘one’ and *bɔːʔ* likely derives from a pronominal. This is a distinct usage reserved with human nominals (see Larish, 1999).

argument and so while coreferential they also provide additional information. Often, classifier phrases form part of a noun phrase when contiguous with the nominal head like in (6.57). But they may also appear separately at the end of an IU yet still co-referential with a corresponding lexical head, like in (6.61) and (6.62).²⁶

(6.61) *kanáj nenán tǎlj lùj*

male stand three CLF

‘Men are standing, three of them.’ [DW.SF10.25]

(6.62) *ʔeká:n p^hlà:t ʔa=làt*

fish fall one=CLF

‘A fish fell, a single one.’ [GP.SF08.19]

Like other co-referring expressions, if a clausal IU’s slot for the lexical argument is left empty then a clause-final classifier phrases may be the only overt marking of the core argument, and therefore result in a variant coding of clausal constituents, such as a VES order in (6.63) or VS in (6.64).

(6.63) *dəp^hlà:t dunl ʔa=bulàt*

fall ground one=CLF

‘(It) fell to the ground, one of them.’ [CN.SF13.45]

(6.64) *dín tǎlj p^hǎh*

come three CLF

‘(He) came with three of them.’ [YT.SF08.19]

Viewing these coreferential elements as leftover artifacts of the ellipsis of lexical arguments discounts the extent to which they can be held as evidence for variant argument structures. This argument hinges on the idea that while co-referring expressions can be counted as contributors of overt information, their place in clausal syntax is not equivalent to the slots and roles of core arguments.

6.2.3 The given-before-new principle

Hypothesis 2 held that variation in argument structure and associated grammatical constructions could be accounted for by the given-before-new (G>N) principle. However, no variant argument structures were elicited through the manipulated givenness conditions in the Transitive Event Picture Sequences stimulus

²⁶ See classifier ‘float’ Larish (2005:526).

nor were they attested to any noticeable degree in the Stolen Fish corpus. Still, the G>N principle as a discourse strategy is an observable informational property of intonation units within the Stolen Fish corpus. This is first demonstrable with a given before new pattern for overt referring expressions. Additionally, ellipsis of A and S arguments when “given” with overt realization of *new* O, E, and non-core arguments was regarded in principle reflecting a discourse adherence to the G>N principle. Finally, and further underscoring the lack of givenness effects on argument structure, a few violators of the G>N principle are shown to still exhibit preferred argument structures.

Within the Stolen Fish corpus clausal IUs with new R-EXPs are an overall minority. But, where new R-EXPs do appear the given-before-new (G>N) principle characterizes one informational aspect of clausal IUs. This is first evident in that a G+N structure ($n=53$) was the most common information structure type with more than one R-EXP as shown in Table 24.

Table 24 Clausal IUs with new R-EXPs

Information structure		<i>n</i>
1 R-EXP	N	111
	G+N	53
2 R-EXP	N+G	13
	N+N	7
3 R-EXP	G+G+N	4
	G+N+N	1
Total		190

While rare overall, there were a few clausal IUs with three overt R-EXPs that exhibited adherence to the G>N principle. For example, both (6.65) and (6.66) have two given arguments preceding one final new argument, thereby producing a G+G+N structure.

(6.65) *caná:t ʔá:k ʔeká:n lám [dagá:]_{r-new}*

child to.place fish in basket

‘The child puts the fish in a basket.’ [KG.SF29.99]

(6.66) *bá:p p^hɔ:3t^haw3 ni:4 tak2 pukát bʔj [ʔeká:n təlʔj p^hɔ́h]_{r-new}*

elder.male old.person TOP scoop gillnet acquire fish three CLF

‘This old man is pulling up the gillnet and has gotten three fish.’ [YT.SF04.8]

One reason IUs with multiple overt R-EXPs are infrequent is because argument ellipsis is so prevalent. Consider that while clauses of a G+N information structure do provide two overt R-EXPs, often these are only two-thirds of the arguments required by a trivalent predicate. For example, the G+N structure in (6.67) has the ditransitive verb *ʔón* ‘give’, but only has the given agent *ná:t* ‘child’ and the new indirect object *ticúm* ‘bird’, while the direct object argument is omitted.

- (6.67) [*ná:t nɣj*]_{r-given} *kɔ:ʔ* *ʔón* [*ticúm*]_{r-new}
 child TOP CONN give bird
 ‘This kid, well, he gives (it) to a bird.’ [GP.SF15.33]

Or in (6.68) a ditransitive verb *dúk* ‘put’ (which variably selects for core or non-core arguments) has a given object *ʔeká:n* and a final new non-core argument *dagà:* ‘basket’, but the agent is omitted.

- (6.68) *mekén* [*ʔeká:n*]_{r-given} *dúk lám* [*dagà:*]_{r-new}
 gather fish put in basket
 ‘(He’s) gathering the fish putting (them) in a basket.’ [YN.SF04.6]

For such clauses, the orders of overt referring expressions’ information statuses are consistent with the G>N principle, but co-occurring argument ellipsis is also an informational aspect in play. Take (6.69), another ditransitive clause. The unmarked order for core arguments for ditransitives is A-DO-IO.²⁷ Here, the indirect object *ticúm* ‘bird’ is new and stands as the sole overt R-EXP.

- (6.69) *káw ʔón* [*ticúm*]_{r-new}
 go give bird
 ‘(He) goes and gives (it) to a bird. [LN.SF15.58]

What is worth noting is that omitted arguments in these clauses correspond to *previously mentioned discourse referents*. This means that although there are no actual R-EXPs to take information status annotations, the appropriate referents for these slots are still in essence “given”. Furthermore, considering that preferred argument structures outline strict grammatical slots, the G>N principle can be extended to include constructions with ellipsed arguments. In Table 24, it is shown

²⁷ Technically, Dixon (2010) core argument marking system does not make use of direct object and indirect object. Instead, O and E are used where DO=O and IO=E.

that for the 111 clausal IUs with only single new R-EXPs, most often the R-EXP corresponded to arguments occurring later in the clause ($n=85$). Namely, these were O and E core arguments and lexical non-core arguments in clause final prepositional phrases. Assuming preferred argument structures of AVO and SVE, we can infer that preceding these new R-EXPs were empty because they corresponded to “given” referents. Conceiving of information status tags for non-existent grammatical arguments presents its own theoretical challenges, but based on both overt patterns and constructions with ellipsed arguments clausal IUs of the Stolen Fish corpus are in line with the G>N principle.

Table 25 Argument type: Clausal IUs with only a single new R-EXP

new R-EXP argument type	<i>n</i>	%
A, S	26	23.4
O, E, lexical <u>non-core</u>	85	76.6
Total	111	100

Finally, as represented in Hypothesis 2, there is the question as to whether a N+G can produce an alternant argument structure. As indicated in Table 25, there were 13 IUs with a N+G structure, tokens which appear to be violators of the G>N principle. One aspect of these violators, however, again is the combination of argument ellipsis and co-referring expressions (see §6.2.2). For example, in (6.70), in terms of the information status of R-EXPs, we have an N+G structure as *ticúm* ‘bird’ was a new R-EXP, while the final classifier phrase is a given referring expression corresponding to an omitted lexical argument *ʔeká:n* ‘fish’. For contrast, note the position of a lexical argument in (6.71) and use of a co-referring classifier phrase.

(6.70) *ʔón [ticúm]_{r-new} [ʔa=làt]_{r-given} lá:j*
 give bird one=CLF PROG
 ‘(He’s) giving a bird one.’ [KG.SF15.45]

(6.71) *mén ʔeká:n ʔón ticúm ʔa=bulát*
 take fish give bird one=CLF
 ‘(They) take the fish and give the bird one.’ [NG.SF15.93]

Some clausal IUs with a N+G information structure were due to given non-core arguments following the introduction of a new referring expression. This could

happen when a new R-EXP precedes a demonstrative referring to a location within the narrative or on the page of the stimulus. Such as with (6.72) where *nu*: ‘DEM.MED’ corresponds to a non-core argument indicating a location within the narrative. This followed the new R-EXP *c^halióη* ‘basket’.

- (6.72) *bá:ʔ dín dúk [c^halióη]_{r-new} lá:j [nú:]_{r-given}*
 carry come put basket PROG DEM.MED
 ‘(He’s) carrying it to come put (it) in a basket, there.’ [HJ.SF28.129]

Example (6.73), offers a verbless copular construction but this time the new copular subject appears first as a lexical expression, while the copular complement is pronominal referring to a part of the illustration.

- (6.73) *[təmán póʔ]_{r-new} [nú:]_{r-given}*
 fishhook 3SG DEM.MED
 ‘That’s his fishing rod.’ [YG.SF22.67]

For some N+G tokens there is a sense in which a following given R-EXP serves to ground the new R-EXP within the text, as if the new information needs linking to the discourse context. Consider two more N+G examples with non-core arguments in prepositional phrases. For example, in the scene #14 ‘Let’s help’ where the ‘bird’ first appears, speakers could introduce the bird with spatial relation to given characters, like in (6.74) and (6.75).

- (6.74) *tɛ:2 wa:3 [ʔadá:]_{r-new} ʔé:m waj4 kót [póʔ]_{r-given} lá:j*
 but COM duck be.located to.be.placed behind 3SG again
 ‘But, a duck is located behind him, as well.’ [NJ.SF14.51]

- (6.75) *[ticúm]_{r-new} ʔé:m lekót [caná:t]_{r-given}*
 bird be.located behind child
 ‘A bird is behind the child.’ [WN.SF14.56]

Another possibility for instances of N+G were double object constructions where it was the first object that was new. Like in (6.76) with a new object *ʔapúj* ‘fire’ and given second object *ʔeká:n*.

- (6.76) *c^hún [ʔapúj]_{r-new} nióη [ʔeká:n]_{r-given}*
 ignite fire roast fish
 ‘(They’re) starting a fire, roasting the fish’ [YN.SF17.53]

Double object constructions are ostensibly an AVOVO structure so again the N+G structure in (6.76) is not indicative of givenness driving a variant syntactic structure. It is also likely that many double object constructions are also conventionalized verb-object constructions (see §5.2.3). In (6.77), we see the same syntactic-semantic pattern of (6.76) the only difference here was contrasting informational conditions, in that the R-EXPs reflect a G+N structure as “roasting fish” was newer information.²⁸ Note the contrast in word-form for the verb *məʔiəŋ* ‘roast’.

(6.77) *cʰún ʔapúj məʔiəŋ ʔeká:n káʔ*
 ignite fire roast fish PRF
 ‘(He’s) started the fire to roast the fish.’ [YG.SF17.49]

Important to note with all these N+G examples is that each clausal IU still maintains preferred argument structures. And contrary to the prediction of Hypothesis 2, an informationally marked order, that is a violation of the G>N principle, does not result in variant argument structures. Still, the G>N principle characterizes a general informational aspect of clausal IUs within the Stolen Fish corpus that happens within the syntactic scheme of preferred argument structures.

6.3 Discussion

Responses to the *Transitive Event Picture Sequences* (TEPS) demonstrated that AVO was the preferred argument structure of transitive clauses regardless of givenness conditions. According to the literature, if givenness was marked using word order alternations then we could have expected something like OVA. But instead, variant responses to TEPS target scenes were characterized more by argument ellipsis and prosodic juncture. Findings on argument structure, therefore, mainly just confirm the “SVO canonical word order” of transitive clauses (Larish, 1999; Swastham, 1982). And although, the absence of evidence is not evidence of absence, due to a general lack of variant orders,²⁹ findings here cast some doubt on the veracity of previous reports, such as Swastham’s (1982) positing of VSO, OSV, and OVS.

²⁸ Here, I am appealing to referential distance (RD). Technically, *ʔeká:n* ‘fish’ in (6.77) was *r-given*, but it had a RD of 5 IUs from its previous mention, one more IU and it would have qualified for the category of *r-given-displaced*. In (6.77) *ʔapúj* ‘fire’ is given with an RD of 1, highlighting the potential and need for more qualitative rather than categorical distinctions in information status marking.

²⁹ See one anecdotal example in §6.2.2

Larish's (1999) VSO word order was based on the appearance of post-verbal clitics, an attested pattern throughout this study. However, an alternant analysis was put forth in which these instances are more of a by-product of argument ellipsis of full lexical expressions. To recap, consider (6.78). The first IU has a full lexical argument *matá:ʔ cʰuwiáʔk* 'fishhook' as an A argument in the preferred argument structure of AVO. Note, however, that the verb here, *tábá:k* 'to pierce, prick', also has the *ɲá:* clitic, which is co-referential with the A argument. Then in the next IU when the speaker specifies it was a finger that was pricked, the lexical argument is elided while the verbal clitic remains. Rather than supposing "movement" of core argument slots, a simpler interpretation is that these patterns represent an overt VO argument structure with a remnant verbal enclitic.

- (6.78) *matá:ʔ cʰuwiáʔk tábá:k =ɲá: ɲán ... Ø bá:k =ɲa: niw4*
 fishhook to.pierce=3.SG hand to.pierce=3.SG finger
 'The fishhook pricks his hand. (It) pricked his finger.' [NJ.A24.2]

One candidate for an alternant argument structure is the OVAV order of the adversative passive construction. However, given the semantics of *kʰána:ʔ* 'to undergo', an alternative analysis is that grammatical subjects have the role of EXPERIENCER. So, for the first argument in (6.79) 'the man' it would be better to describe it as being a subject (S) of an intransitive verb *cʰána:ʔ* 'to undergo', which takes a complement clause (CoCl) to fill an extended intransitive argument slot (E). This interpretation avoids the need for variant word orders and recasts the issue in terms of semantics and moreover is in line with the general idea of Moklen being SVO.

- (6.79) [*bó:ʔ kanáj*]S *cʰána:ʔ* [*matá:ʔ cʰuwiáʔk tábá:k*]CoCl:E
 person man to.undergo fishhook to.pierce
 '[The man]s got [pricked by a fishing hook]E.' [WN.P24.2]

A key aspect for understanding the relation of information structure to syntactic properties is the role of prosodic segmentation in delimiting basic units of grammaticality. Discourse-oriented approaches starting from IUs (or other "basic units of speech", see Izre'el et al., 2020) foreground the place of actual communicative utterances into discussions of syntax. Here, there is more to word order than mere

sequential order of clausal constituents, as both prosodic unity and syntactic contiguity contribute to a notion of grammaticality (Croft, 2001). Furthermore, contemporary approaches to information structure acknowledge the need to accurately represent *oral grammar*, especially in lesser-described oral languages (Fernandez-Vest, 2015). For example, within the TEPS data initial detachment of patients, i.e., an O...AV(O) structure, in patient-given contexts was identified as a discourse strategy. Perhaps it was structures like these that corresponded to previous claims for a Moklen “OSV” word order, but without any prosodic information we simply do not know. For example in (6.80) a given direct object appears in an initially detached phrasal IU, preceding a clause with a corresponding predicate. Based on syntactic entailment, one might say this is an object-initial sentence. On the other hand, instead of claiming an object-initial word order, a more accurate account could acknowledge the prosodic disjuncture and contrast it to the *highly attested* AVO argument structure occurring within the bounds of singular IUs, like in (6.81).

(6.80) *ká:n kʰɔ:ŋ5 cɛj ... Ø bá:ʔ dín ʔón bɛj mók*
 fish POSS 1SG carry come give 2SG cook
 ‘This fish of mine, (I’ve) brought (it) to you to cook.’ [LN.SF16.64-65]

(6.81) *ʔa=bɔ:ʔ bɛj bá:ʔ ʔeká:n ʔón pəníəŋ*
 one=person to.acquire carry fish give roast
 ‘One of them brings the fish for (them) to roast.’ [NK.SF16.53]

Similarly, instead of verb-initial word orders, we can see that a phrasal IU sometimes appears as a final-detachment containing an argument not included within the clausal IU. For example, in (6.82) we have an A argument post-posed, or in (6.83), with a final-detached O argument. This is not to say that these detachment constructions cannot be considered part of the grammar at some level, but the disjunction underscores the importance of prosodic information when discussing orders of clausal constituents as constituting “word orders”.

(6.82) *nuwá:j pəsáŋ lá:j ... bá:p nú:*
 to.fell banana PROG elder.male DEM.MED
 ‘(He’s) felling the banana (tree), that man.’ [LI.P11.2]

(6.83) *bá:ʔ dúk cʰalíəŋ káʔ ... ká:n nú:*
 carry put basket PRF fish DEM.MED

‘(He’s) carried (it) and put (it) in a basket, that fish.’ [HJ.SF29.131-132]

A better model of oral grammar would be to explicitly acknowledge multi-IU grammatical constructions and showcase their informational properties. For example, in (6.84) an initial detachment construction may be framed as a topicalization of core argument. Note the use of full lexical expression, rather than a pronominal like in (6.81). There is also a demonstrative topic marker *ní:4* (Iwasaki & Ingkaphirom, 2005), as well as the monosyllabic alternant for *ɽuɽúj* ‘younger.sibling’ (see §7). This construction is basically characteristic of “topic-prominent” languages of Mainland Southeast Asian languages (Vittrant & Watkins, 2019). But, the point is that rather than overlooking prosodic and informational properties and misattributing alternant word orders, such properties inform a discussion of syntax.

- (6.84) *ɽuj* *leʔ4 ɽaka:ʔ* *ní:4* ... *pəná:ʔ* *caná:t* *dəlát* *ʔeká:n*
 younger.sibling and older.sibling TOP see child steal fish
 ‘These younger and older siblings, (they) see the child stealing the fish.’
 [YT.SF08.17-18]

The information structure approach to Moklen syntax contextualizes several of the language’s syntactic properties and simplifies the matter of variant “word orders”. Preferred argument structure of clausal IUs is either AVO or SVE and where both overt arguments are overtly realized in one IU they follow a given-before-new order. However, arguments are often ellipsed or appear in detachment constructions outside of clausal IUs. Findings here, therefore, do not radically change our picture of the core of Moklen syntax, but instead places it within its discourse contexts.

6.4 Summary

Hypothesis 2 was undermined as findings from the Transitive Event Picture Sequences demonstrated a failure of givenness effects to elicit any variant argument structures. Therefore, rather than there being any alternant word orders (cf. Larish, 1999; Swastham, 1982), Moklen syntax is shown to largely adhere to an AVO argument structure. Overall, the available variation of Moklen clausal syntax is better described as exhibiting argument ellipsis, verb-object constructions, remnant co-referring expressions, and prosodic disjuncture.

7 Monosyllabic alternants

7.1 Introduction

Moklen disyllabic lexemes had been noted for their reduced monosyllabic forms. Investigation into Moklen word-form changes offers a better picture of lexical variability within connected speech. To study the relationship between Moklen word-form and information status, the following hypothesis was tested: use of monosyllabic alternants will correspond to “given” information statuses. Data used to directly test this hypothesis comes from the Transitive Event Picture Sequences. Further analysis of monosyllabic alternants comes from the Stolen Fish corpus and the researcher’s field data.

To first determine whether “givenness” was underlying the reduction of disyllables, monosyllabic alternants were analyzed in accordance with the RefLex Scheme. Initial results pointed to mere “givenness” as not indicative of monosyllabic alternants overall. Further examination of the monosyllabic alternants, however, in the context of *word-form shifts* showed that a change from a disyllabic to monosyllabic word-form corresponded to a change from “new” to “given”. Given this pattern, the elision of minor syllables and resulting monosyllables were interpreted as reflecting an informational shift towards *topics*: mutual knowledge established as shared conceptual grounding” (Masia 2022). While there is a role for “givenness” here, other aspects surrounding the word-form shifts led to this more encompassing view. These aspects include: nominal alternants’ position at the head of compounds, predicate complexity, and occurrence of the reduced form at the left-edge of intonation units—where they serve as starting points for additional information.

To capture this discourse phenomenon a process of *minor-syllable elision* is put forward as a morphological consequence of discourse-conditioned reductions in prominence. Ultimately, it and the topical interpretation is presented as an interesting tendency impacting the shape of Moklen words within discourse. This findings clarifies Moklen’s monosyllabic alternants beyond characterizations as “colloquial forms” or the minor syllables as “optional” and contributes to other areas for further discussion.

7.2 Word-form shifts³⁰

In total, the 16 TEPS sessions selected for analysis amounted to 2 hours and 40 minutes of speech. Within this, a total of 530 monosyllabic alternant tokens made up of 64 lexemes were elicited. These included 28 different nouns, 31 verbs, and 5 closed-class items. Table 26 provides these figures along with a comparative figure for the overall frequency of the corresponding disyllables of all monosyllabic alternants in that lexical class.

Table 26 TEPS Monosyllabic alternants by lexical class.

Class	# of Lexemes	Tokens	
		Monosyllable	Disyllable
Nouns	28	288	1,139
Verbs	31	163	352
Closed class	5	79	19
Total=	64	530	1,510

Most monosyllabic alternant tokens were nouns. Instances of verbs were made up of a few more distinct lexemes but had fewer corresponding disyllables than nouns. Disparities here are due in large part to the nature of the TEPS task. Context scenes trigger nominal descriptions of discourse referents, which are then depicted across a total of three pictures. Verbs, on the other hand, were typically not elicited until the target scene and could potentially have two overt nominal referring expressions filling core argument slots. Resolution scenes also portray the agents and patients of each sequence; but the transitive event is completed, and a new activity is depicted, thereby triggering elicitation of a new verb. For the few elicited closed-class items, the monosyllabic word-form appears to be the more common form. Because staged-communicative events induced by the TEPS stimulus are limited to a set of predetermined storylines (see Table 14 in §4.4), absolute frequency of vocabulary items cannot be included as a factor in analysis. A list of all lexemes with identified

³⁰ Sections in this chapter are adapted from (Loss, Chanchaochai, Pittayaporn, and Enfield, forthcoming)

monosyllabic alternants elicited in the TEPS data along with a count of each lexemes' monosyllabic tokens and occurrence within word-from shifts is in Appendix B.

Information status tagging for the 530 monosyllabic alternants resulted in 6 different information status profiles—that is the possible combinations of “given” and “new” across the two levels of the RefLex scheme. To review, information status was annotated at two levels, a referential level (r-level) and a lexical level (l-level). The r-level is used as an analysis of the tracking of discourse referents through the use of referring expressions. Tags at the l-level are attached at the word level. Nouns and classifiers, however, which can act as the heads of referring expressions (i.e., noun phrases) receive tags at both levels. Individual verbs, prepositions, and numbers, however, do not receive r-level tags and are therefore only evaluated at the l-level. In Table 27 we can see that in terms of lexical givenness, overall, the monosyllabic alternants were evenly distributed across “given” and “new”, 51% and 49% respectively. That nearly half of all monosyllabic alternants were lexically new immediately challenged a purely givenness-based hypothesis.

Table 27 Information Status of monosyllabic alternants in TEPS

lexical	referential	<i>n</i>	Word-form shift	%
new	new	89 (16.79%)	5	
new	-	142 (26.79%)	4	49
new	given	29 (5.47%)	4	
given	new	3 (0.56%)	1	
given	-	85 (16.03%)	47	51
given	given	182 (34.33%)	117	
Total=		530	178	100

Overall, of all 530 occurrences of monosyllabic alternants in data, 178 also had the disyllabic form appear within the same discourse context—a description of a singular three-picture sequence in TEPS. These instances were comprised of 48 different lexemes (23 nouns, 23 verbs, 2 prepositions) and constituted a *word-form shift*—an alternation of the disyllabic and monosyllabic word-form during speech.

Instances of word-form shift were especially interesting because they offered a context for comparison of both disyllabic and monosyllabic word-forms within a singular discourse context. Examination of word-form shifts revealed that the overwhelming pattern of word-form shifts was disyllabic *before* monosyllabic, as shown in Figure 35.

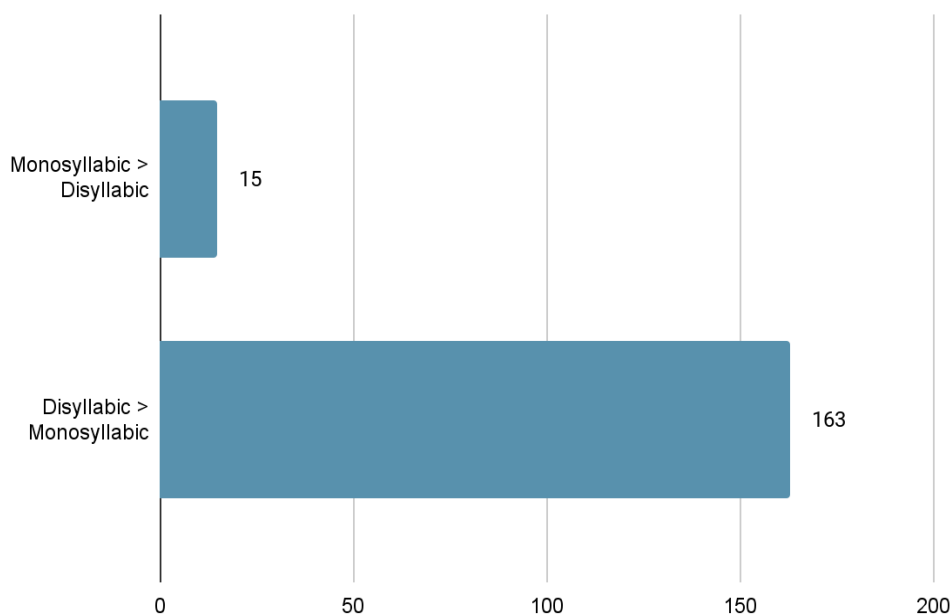


Figure 35 Patterns of word-form shifts

With regards to the original hypothesis, this pattern was significant as it meant changes in word-form mainly corresponded with a shift from “new” to “given”. Moreover, this pattern was consistent with both nouns and verbs (l&r-given and l-given profiles in Table 27). Therefore, while merely being “given”, lexically or referentially, was not indicative overall of use of a monosyllabic alternant, a salient informational aspect of the *elision* of disyllables’ minor syllables was a shift to “given”.

7.2.1 Nouns

There was a total of 288 nominal monosyllabic alternant tokens in the TEPS data. These came from 28 different lexemes. Nominal alternants were annotated at both the r-level and l-level, and therefore tokens fit into four information status profiles, see Table 28.

Table 28 Information status of nominal monosyllabic alternants.

lexical	referential	<i>n</i>	Word-form shift
new	new	87 (30.21%)	5/87 (5.75%)
new	given	22 (7.64%)	4/22 (18.18%)
given	new	3 (1.04%)	1/3 (33.33%)
given	given	176 (61.11%)	117/176 (66.48%)
Total =		288	127/288 (44.1%)

Many nominal monosyllabic alternants fit within the category l&r-new (30.21%) but most were l&r-given (61.11%). This result was in line with the original hypothesis. However, the alternation between the disyllabic and monosyllabic form were found to be a better context in which to examine changes in word-form. Further analysis within word-form shifts showed that nominal monosyllabic alternants' IU-initial position and role as the heads of compounds pointed to a more general backgrounding phenomenon, namely a shift to topical information.

While most monosyllabic nominal alternants, both by themselves within cases of word-form shifts align with l&r-given, that other tokens had new tags suggested mere givenness might not be adequate in accounting for nominal alternants overall. Alternants marked l&r-given indicate use of a previously used lexical concept for a referent with a coreferential antecedent. There were 176 tokens with this information status, and of these, 117 were a part of a word-form shift (see Figure 35), meaning that the corresponding disyllabic form was used in the same sequence description. That the monosyllabic alternants here are l-given points to the fact that all 117 instances were cases where the disyllabic form preceded the monosyllabic form. Put another way, it was the disyllables that were first *activating* the lexical concept.

There are several things to note about shifts from disyllabic to monosyllabic word-form. For one, a disyllable might first appear as a sort of citation form for an initial description of a referent before being immediately reduced. Like in (7.1) where *ʔlá:n* 'snake' is used first, but when the speaker qualifies their description with a

modifier *k^{hiaw5}* ‘to be green’, monosyllabic *lá:n* ‘snake’ is used in the typical compound structure.

- (7.1) *ʔólá:n ... lá:n k^{hiaw5}*
 snake snake green
 ‘A snake...a green snake.’ [NJ.A17.1]

The reduction of the disyllable may also happen when the lexeme becomes an l&r-given grammatical subject. For example, in (7.2), we first see initial use of the disyllabic word-form *ʔólá:n* ‘snake’ for reference in the context scene, but then in the target scene monosyllabic *lá:n* appears at the start of the clause in (7.2.2).

- (7.2)
 (7.2.1) *ʔólá:n xj*
 snake VOC
 ‘A snake, ah!’ [NN.A17.1]
 (7.2.2) *lá:n bətók kakáj=ná: siʔA ja:jl*
 snake strike foot=3.SG PRT grandmother
 ‘The snake’s striking your foot, grandma!’ [NN.A17.2]

The most common nominal lexeme displaying word-form shifts was *ʔaná:t* ‘child.offspring’. While one context for the appearance for its monosyllabic alternant is in compounds, such as *ná:t mináj* ‘girl’, *ná:t dalà:ʔ* ‘young girl’ or *ná:t kanáj* ‘boy’, these do not account for all uses of its monosyllabic alternant. For example in (7.3) and (7.4) the word-form shift has monosyllabic *ná:t* appear as a lone nominal.

- (7.3) *nɣj ʔaná:t póʔ ... ná:t wà:ʔ lùj*
 DEM.PROX child.offspring 3P child two CLF
 ‘These are their children, two kids.’ [PB.A12.1]
 (7.4) *mináj bʔj ʔaná:t ... ná:t ká:ʔ dɛn*
 woman acquire child.offspring child QPLR Dan
 ‘The woman’s got a child. A kid, Dan?’ [DW.P1.2]

The overwhelming pattern for nominal word-form shifts was disyllables *before* monosyllables, yet the shift was not always immediate. For example, in (7.5), a description of Figure 36, the lexeme *kabá:ŋ* ‘boat’ appears three times as a grammatical subject, but it is not until the third instance in (7.5.3) that it is reduced to its monosyllabic form *bá:ŋ*.



Figure 36 Patient-initial sequence #19

(7.5)

(7.5.1) *kabá:ŋ büt*

boat run

‘A boat’s going.’ [LI.P19.1]

(7.5.2) *kabá:ŋ kǎj batǎj lɛ:j1*

boat to.ground rock INTS

‘The boat grounds onto a rock!’ [LI.P19.2]

(7.5.3) *bá:ŋ kalám lɛ:w4*

boat sink PRF

‘The boat has sunk.’ [LI.P19.3]

A similar pattern is shown within a single target scene in (7.6). Here disyllabic *kicú:m* ‘bird’ was used twice before shortening to the monosyllabic form. The first IU shows the speaker hesitated momentarily whilst identifying the sequence’s new patient, *lací:* ‘worm’. The speaker then completes the transitive clause in the second IU; but note that when the entire proposition is reiterated in a third IU, the lexeme for ‘bird’ is realized in its monosyllabic form.

(7.6) *kicúm nám=ná: ... kicúm nám=ná: lací: ... cúm nám lací:*

bird eat=3.SG bird eat=3.SG worm bird eat worm

‘The bird eat... The bird’s eating a worm. The bird’s eating a worm.’

[TG.13A.2]

Shown this way, the loss of a minor syllable within running speech might seem like just a probable eventuality. However, when we consider that for nominal word-form shifts, there was an overwhelming tendency for co-occurring disyllables to first be l&r-new and then monosyllabic alternants to be l&r-given, as shown in Figure

37, the alignment of minor syllable omission seems to correlate with an informational contrast.

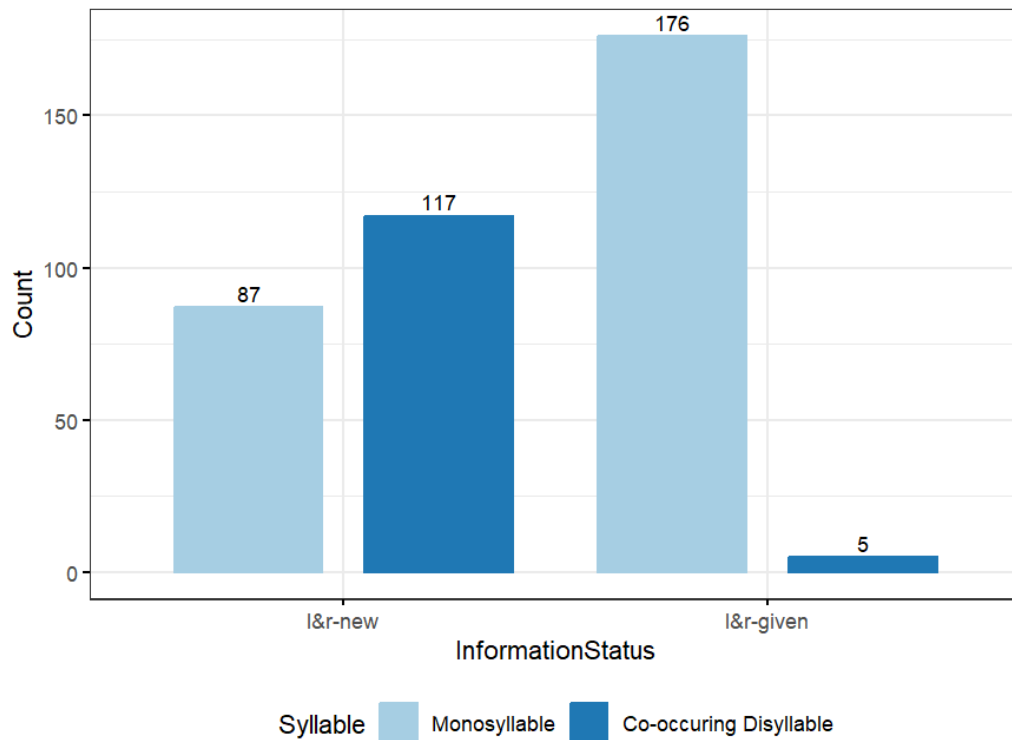


Figure 37 Information status of nominal alternants and co-occurring disyllables.

A closer look at word-form shifts offers several points for consideration. One significant factor is a resulting monosyllabic alternant's appearance at the left-edge of an IU. Within cases of word-form shifts, 95/127 nominal monosyllabic alternants appeared in an IU-initial position. For example, in the sequence description of Figure 38 in (7.7), we can see the change in word-form for both *ʔeká:n* 'fish' and *kətá:m* 'crab' occurs not only with a shift to l&r-given, but in both instances, the monosyllabic form begins its respective IU.

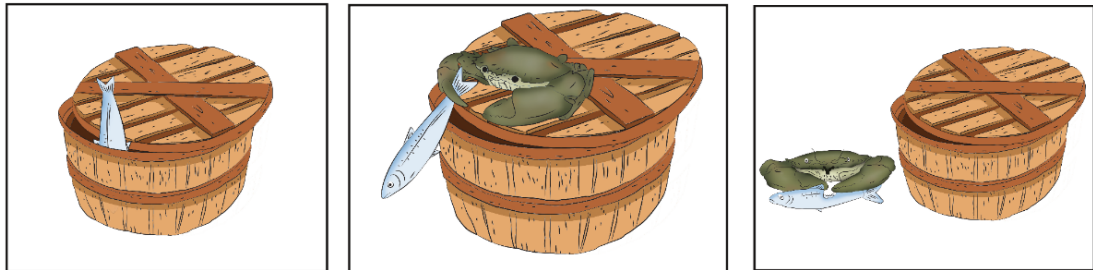


Figure 38 Patient-initial sequence #14

(7.7)

(7.7.1) *ʔeká:n namá:ʔ lám tʰaŋ5*

fish enter in bucket

‘A fish is going in a bucket.’ [LW.P14.1]

(7.7.2) *ká:n bé:t ca:k2 tʰaŋ5 ... kətá:m kʰi:p3 =ná:*

fish exit from bucket crab pinch=3.SG

‘The fish goes out of the bucket. A crab grabs (it).’ [LW.P14.2]

(7.7.3) *tá:m nám ʔeká:n*

crab eat fish

‘The crab eats the fish.’ [LW.P14.3]

Especially instructive in (7.7.3), however, is the reappearance of disyllabic *ʔeká:n* ‘fish’ within the predicate despite being l&r-given. Drawing from Cresti’s (2018) illocutionary model, these IU-initial monosyllabic alternants can be construed as marking the “field of application” of an utterance’s illocutionary force. That is, the monosyllabic alternants are appearing as starting points upon which more information is added. In other words, while givenness may be a frequent feature, the reduction into a monosyllable alternant may be indicative of just being sufficiently *topical*. So, while a shift to *given* characterizes one aspect of minor syllable loss, the position of *non-focus* at the start of utterances seems to offer a key clue. Specifically, it suggests that it is when an IU is beginning with information well within the common ground, that we can potentially get reduction of a disyllabic lexeme.

As shown in Table 28 many nominal alternants also appeared with *new* tags. Most of these were the 87 l&r-new nominal alternants. Alternants in this category signify first uses of a lexical concept for discourse referents with no coreferential

antecedent. Being maximally *new* these alternants first stood out as they were counter to the hypothesis that mere givenness was underlying the realization of reduced word-forms. For these l&r-new alternants, most were found functioning as the head of a compound, constructions previously identified by Larish (1999). The two most frequent within the data were reduced forms for *pəkón* ‘tree’ and *buwá:k* ‘fruit’, like in *kón pɛc^háj* ‘banana tree’ and *wá:k paʔó:k* ‘mango’. Typically, these compounds were used for initial descriptions of referents depicted in context scenes.

Highlighting their role as nominal heads, examples (7.8) and (7.9) show how *niʔú:n* ‘coconut’ appears in compounds indicating either the tree or its fruit. Note also in the second IU in (7.9), an additional shift to monosyllabic *ʔú:n* ‘coconut’ with the addition of a further modifier *bəluəŋ* ‘fresh’.

(7.8) *kón niʔú:n*

tree coconut

‘A coconut tree.’ [LN.P11.1]

(7.9) *nɿj wá:k niʔú:n ... wá:k ʔú:n bəluəŋ*

DEM.PROX fruit coconut fruit coconut fresh

‘This is a coconut, a fresh coconut.’ [NK.P2.1]

The high frequency of *wá:k* and *kón* as l&r-new alternants must first be understood as a direct result of the stimulus design; depictions of fruits and trees were used as non-human inanimate referents in multiple sequences (see §4.4.1). Still though, that these were l&r-new challenged the idea that mere givenness could account for the overall use of monosyllabic alternants. Consider also that the next two most common l&r-new alternants were monosyllabic forms for *ʔɛbá:p* ‘elder.male’, and *ʔibú:m* ‘elder.female’. These also functioned as nominal heads in compounds like *bá:p p^hɔ:3t^haw3* ‘an elder old man’ or *bú:m mináj* lit. ‘elder female woman’. More generally, *bá:p* and *bú:m* were known to serve as honorifics before individuals’ names, like *bá:p sampan*, ‘Elder Sampan’ or *bú:m là:p* ‘Elder Lap’, but within the TEPS data, they were used along with a modifier in the same manner as other compounds. Overall, because depictions of fruits, trees, and elderly people were chosen, alternants with an l&r-new information status were largely made up of these

four lexemes (71.5%). Still, that they and other similar nominal compounds occurred while *new* undermined a sheer givenness-based explanation.

Interestingly, if we take *topics* to represent a discourse strategy of broad evidentiality, it seems we can contextualize one aspect of use of l&r-new alternants in compounds within the TEPS task. Masia (2022), holds that topics encode information “which represents *mutual* knowledge established as shared conceptual grounding with both speaker and hearer as committed source” (p.83). If monosyllabic alternants are seen as reflecting topical information, this could account for some of the *new* alternants within the discourse context of the TEPS task. In taking this line of thinking, we need to zoom out to the broader context of the TEPS sessions. Consider that since the stimulus was visible to both speaker and hearer, when nominal compounds were first elicited in a context scene, the speaker was beginning from a point of *mutual* knowledge, that is the contents of the illustration were already “situationally evoked” for both participant and interviewer. Initial reference to the illustration might even occur along with a deictic gesture (see Figure 17). Descriptions of context scenes, therefore, typically begin without any question as to whether something was a tree, fruit, or fish. Instead, the speaker’s aim and the weight of their illocutionary force is in sharing their individual interpretation of the referent’s kind. For example, in (7.10), a speaker first commits to *ká:n* the monosyllabic form of *ʔeká:n* ‘fish’ before then indicating that it is an ‘Asian sea bass’ through use of a nominal compound.

(7.10) *ká:n* ... *ká:n kəpʰóŋ*
 fish fish asian.sea.bass
 ‘A fish...an Asian sea bass.’ [LN.P5.1]

For nominal compounds there are several reasons to believe that for many lexemes the monosyllabic alternants function as a sort of class noun (see §7.4 for more discussion). However, it is worth stressing that there is no semantic restriction on using the full disyllabic form, and a speaker can in fact use a disyllabic head if the item is sufficiently in focus (e.g., *ʔeká:n kəpʰóŋ* ‘Asian sea bass’). Take for example, a rare case of the monosyllabic form occurring before the disyllabic form in (7.11). Here, the lexeme *pəkón* ‘tree’ first appears as the monosyllabic head of a compound.

But when the referring expression for the same discourse referent moves into the predicate in (7.11.2), the disyllabic word-form was used.

(7.11)

(7.11.1) *kón pɛc^háj*

tree banana

‘A banana tree.’ [CU.P11.1]

(7.11.2) *nuwáj pəkón pɛc^háj*

to.fell tree banana

‘He’s felling the banana tree.’ [CU.P11.2]

As shown in Table 28 a small proportion of tokens also fit within the l-new&r-given category.³¹ This category represents use of a new lexical concept for a given discourse referent. Such as in (7.12) where the speaker first uses *mináj* ‘woman’, then later switches to the monosyllabic alternant form of *ɲenój* ‘mother’ when describing the same referent in a subsequent scene. Hence, in this instance, the monosyllabic alternant *nój* ‘mother’ is *lexically* new yet *referentially* given. But, as it is serving as the grammatical subject in an IU-initial position, it is still consistent with an interpretation of nominal alternants aligning with topical information, as the discourse referent is mutual knowledge. Most instances in this category are of this nature.

(7.12) *mináj (...) nój mé:ʔ ɲaná:t*

woman mother hip.carry child

‘A woman (...) The mom is carrying her child.’ [LN.P1.1-3]

In summary, analysis of nominal monosyllabic alternants’ information status showed that they most often appeared as l&r-given, a finding in line with the original hypothesis that givenness was connected to the use of monosyllabic alternants. However, taking into account nominal alternants’ IU-initial position and their more general function as the heads of compounds suggested a more encompassing view. Namely, that the main informational aspect of a shift to nominal monosyllabic alternants is not a matter of mere givenness in strict information status terms, but

³¹ The other category l-given&r-new ($n=3$) were cases where the same lexeme was used in two contiguous sequences. In arranging the order of sequences of TEPS, the design aimed to avoid eliciting the same lexical material in contiguous sequences. Nevertheless, there were 3 instances where speakers interpreted a context scene as related to the prior sequence. These alternants, therefore, were regarded as invalid as they bypassed a control of the stimulus design.

rather an alignment with topical information—syntactic regions on which speakers present weaker individual epistemic commitment (Masia 2022).

7.2.2 Verbs

A total of 163 monosyllabic alternant tokens from 31 verbal lexemes appeared within the TEPS data. As the information status of verbs was classified solely at the lexical level, the difference between new and given represents a contrast between an unused lexical concept and a previously used lexical concept. As shown in Table 29, there were 92 *l-new* and 71 *l-given* verbal monosyllabic alternants.

Table 29 Verbal monosyllabic alternants

lexical	referential	<i>n</i>	Word-form shifts
new	-	92 (56.44%)	3/92 (3.26%)
given	-	71 (43.56%)	44/71 (61.97%)
Total=		163	47

The initial take-away, therefore, was that lexical newness *did not* preclude use of a verbal monosyllabic alternant, a result counter to a purely givenness-based explanation. However, looked at in terms of word-form shifts the prevailing pattern again was disyllables preceding monosyllables. Further examination of these instances provided several indications that the loss of the minor syllable was accompanying a backgrounding of the verbal concept. First, there was integration of more elements into the predicate, such as the addition of verbal modifiers or grammatical objects. But another key clue was reduced verbal alternants' proximity to positions of ellipsed arguments.

For verbal lexemes, on the surface there at first seems to be a general equivalence of the two word-forms in lexically new contexts. Moreover, there were many instances of both the disyllabic and monosyllabic forms being elicited in either stimulus condition. For example, the verb *nəpɔ:k* 'launder' was variably realized in target scenes of agent-initial and patient-initial versions of the same sequence (Figure 39), as shown in (7.13) and (7.14).

(7.13) Agent-initial sequence #19

(A) *nəpók cəwát*
 launder clothes
 ‘(She’s) washing clothes.’ [NJ.A19.2]

(B) *pók cəwát*
 launder clothes
 ‘(She’s) washing clothes.’ [TW.A19.2]

(7.14) Patient-initial sequence #19

(A) *p^hɔ:3t^haw3 nəpók cəwát*
 old.person launder clothes
 ‘An old person is washing the clothes.’ [WN.P19.2]

(B) *mináj pók bajáj*
 woman launder shirt
 ‘A woman is washing a shirt.’ [LN.P19.2]



Figure 39 Full sequence #19 illustrations

Since the verbal monosyllabic alternants’ appearance when lexically new ruled out givenness as the underlying cause, again word-form shifts were looked to as a context of comparison. Here it was found that within instances of word-form shifts, co-

occurring disyllables were usually *l-new*, as shown in Figure 40, and a corresponding reduced form was *l-given*.

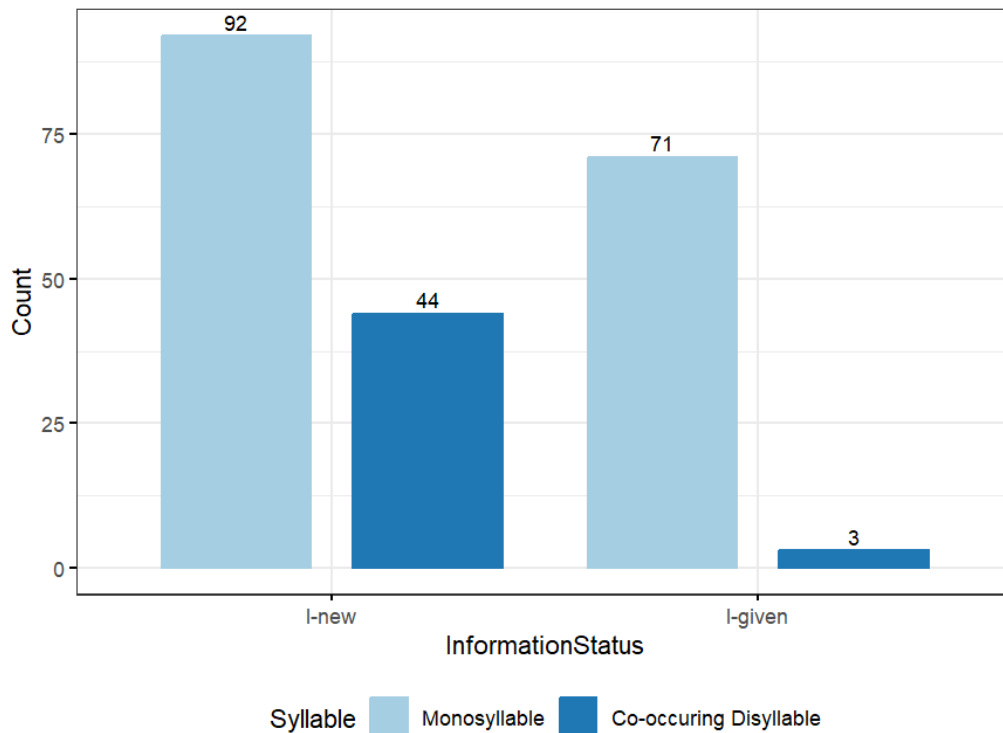


Figure 40 Information status of verbal alternants and co-occurring disyllables.

One aspect of shifts from disyllabic to monosyllabic is the integration of more components into the predicate. Consider the following two sequence descriptions of Figure 41 featuring the lexeme *didú:n* ‘to sleep.’ In (7.15.1), we first see the disyllabic form appear as a solitary verb. The sequence description continues through the target scene (7.15.2), but by the resolution scene in (7.15.3), monosyllabic *dú:n* ‘to sleep’ appears in a more complex verb phrase along with the auxiliary verb *bǎj* ‘to be able’ and an adverb *ʔabó:ʔ* ‘together.’

(7.15)

(7.15.1) *caná:t didú:n*

child sleep

‘A child is sleeping.’ [NK.P7.1]

(7.15.2) *máʔ póʔ mén cəwát pit2 ʔaná:t*

mother 3SG take cloth close child.offspring

‘Its mother takes a blanket and covers the child.’ [NK.P7.2]

(7.15.3) *máʔ kʰu:3 ʔaná:t bɣj dú:n ʔabó:ʔ*

mother and child.offspring able sleep together

‘The mother and child are sleeping together.’ [NK.P7.3]

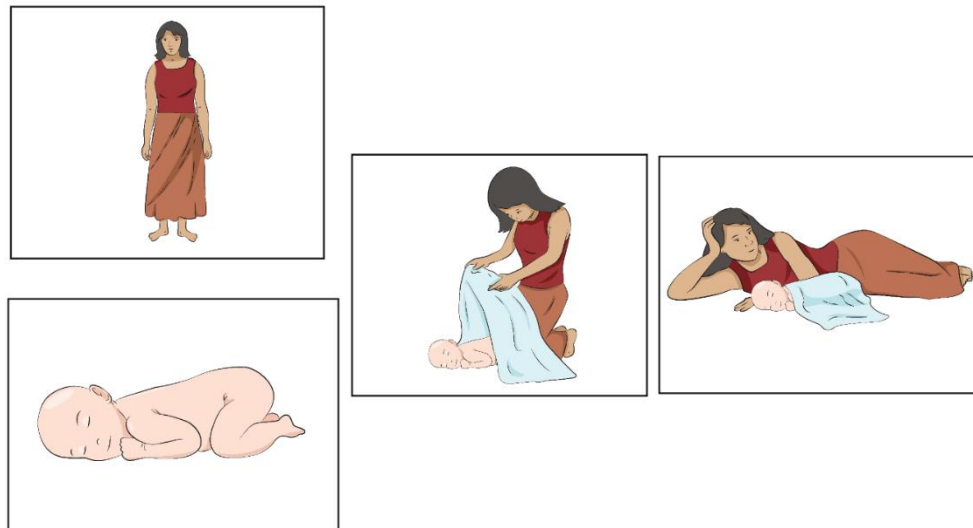


Figure 41 Full sequence #7 illustrations

Now, consider an agent-initial version of the same sequence in (7.16). In this example, the lexeme *didú:n* ‘to sleep’ appears during the target scene as the sole verb of a causative complement after the patient *ʔaná:t* ‘offspring’. For the resolution scene, the same verbal lexeme appears, but it is realized as monosyllabic *dú:n* ‘to sleep’ and is joined by the addition of a prepositional phrase.

(7.16)

(7.16.1) *mináj*

woman

‘A woman.’ [LN.A7.1]

(7.16.2) *bóh ʔaná:t didú:n*

make offspring sleep

‘She’s putting her child to sleep.’ [LN.A7.2]

(7.16.3) *ʔa: ʔenóŋ dú:n troŋl ná:t*

VOC mother sleep directly child

‘Ah, the mother sleeps by her kid.’ [LN.A7.3]

Examples of word-form shifts for verbs have elision of the minor syllable accompanying reuse of the lexeme. A preliminary explanation for the change in word-form is that it reflects a process of information integration. Framing this within the context of the TEPS task, we see that speakers are repeatedly presented with scenarios in which they must establish two referents and predicative and relational information concerning them. And, descriptions of resolution scenes like (7.15.3) and (7.16.3) show elision of a verb’s minor syllable as the speaker works towards linking all this information together. On a surface level, what we are capturing is a situation of a speaker first being confronted with a brand-new event and subsequently incorporating it in relation to a new discourse referent. But, from an information structure perspective, it is possible to view changes in word-form as signifying a general *backgrounding* of the verbal concept. This interpretation supports the idea, raised earlier for nouns, that the change to a monosyllable is indicative of a shift to topical information; that is, at the time of speech, monosyllabic alternants are aligned with *mutual knowledge* established as shared conceptual grounding between interlocutors. Or as put by Masia (2022) topics mark a syntactic region on which the speaker shows weaker *individual* epistemic commitment.

Additional evidence for a topical interpretation is provided by how verbal alternants are commonly used as starting points for incomplete predicates. In these instances, the speaker first produces the full disyllabic verb but without a grammatical object. Then in an immediately following IU, the speaker starts from the reduced form and adds more information. For example, in (7.17), monosyllabic *ŋé:n* ‘to chase’

begins an IU completing the proposition from the previous IU. Within cases of word-form shifts, 31/47 verbal monosyllabic alternants appeared in an IU-initial position.

- (7.17) *máʔ pɔʔ nəŋé:n ... ŋé:n ʔaná:t*
 mother 3SG chase to.chase child
 ‘Its mother is chasing. (She’s) chasing her child.’ [EW.P4.2]

One compelling reason for seeing some verbal alternants as aligning with topics is their proximity to the positions of ellipsed arguments. First, consider that a prevalent and the most extreme form of backgrounding information in Moklen is argument ellipsis (see §5.2). In (7.13) and (7.14), we saw examples of the variable realizations of new uses of *nəpɔ:k* ‘to launder,’ which on their own were not particularly revealing. However, if we contrast both word-forms in the context of a word-form shift, like in (7.18), we see that the monosyllabic form is a part of the topical portion of the utterance. Starting with (7.18.1), we see the disyllabic *nəpɔ:k* ‘to launder’ first appears clause-finally as part of a serial verb construction. Note that the agent is ellipsed and is thus topical, while a grammatical object *cəwát* ‘clothes’ is present. Then in (7.18.2), the speaker begins with a detached adverbial clause wherein not only are both agent and object ellipsed, but the word-form of the verb has shifted to monosyllabic *pɔ:k*, for the portion of the utterance which outlines *mutually shared information*.

- (7.18)
 จุฬาลงกรณ์มหาวิทยาลัย
 CHULALONGKORN UNIVERSITY
 (7.18.1) *∅ leʔ4 mén cəwát bá:ʔ nəpɔ:k*
 and take clothes carry launder
 ‘And (she) takes the clothes to wash.’ [YG.P8.2]

- (7.18.2) *p^hɔ:4 ∅ pɔ:k ∅ set2 káʔ ... bá:ʔ niʔù:n héʔ la:wI*
 when launder finished PRF carry dry.in.sunlight at clothesline
 ‘When (she’s) finished washing, she takes them to dry at a clothesline.’
 [YG.P8.3]

Examination of verbal monosyllabic alternants within instances of word-form shifts indicated that the loss of the minor syllable might be indicative of a backgrounding of the verbal concept. Supporting this interpretation, there are examples of verbal monosyllabic alternants appearing within more complex verb phrases and instances where they acted at starting points for additional information.

Key for this interpretation of the informational role of verbal alternants is that argument ellipsis represents the most extreme form of informational backgrounding, and hence a verbal alternant's proximity to the position of ellipsed arguments implies some alignment with topical information.

7.2.3 Closed-class items

Many lexemes from closed-class categories also display variability in word-form. However, in the TEPs data, only 79 monosyllabic alternants from 5 closed-class lexemes were elicited. Correspondingly, examples of word-form shifts were also rare as corresponding disyllabic forms were scarce, as shown in Table 30. While findings are limited here, that these monosyllabic alternants often appear integrated within larger grammatical units implies to some degree an affinity towards the informational background. Still, with some selected examples, aspects of closed-class monosyllabic alternants are consistent with the developing analysis.

Table 30 Closed class monosyllabic alternants

Lexeme	Gloss	l-new	l-given	<i>n</i>	Disyllabic Tokens	Word-form shift
<i>dalám</i>	'in'	30	10	40	7	3
<i>bulàt</i>	CLF	9	6	15	5	0
<i>duwà:ʔ</i>	'two'	13	2	15	0	0
<i>datá:</i>	'on'	6	2	8	4	1
<i>lemá:ʔ</i>	'five'	1	0	1	3	0
		59 (74.68%)	20 (25.32%)	79	19	4

Appearing in the data were alternants for numerical lexemes *lemá:ʔ* 'five' and *duwà:ʔ* 'two,' for which monosyllabic *wà:ʔ* was the only realization, despite having disyllabic attestations elsewhere. Disyllabic numerical lexemes were already known to be reduced in numerical compounds (Larish, 1999), such as *wà: plóh* 'twenty', which also features another monosyllabic alternant of *cʰəpʰlólh* 'ten'. But in the TEPS data numerical alternants only appeared within classifier phrases, like in (7.19) and (7.20), where they are used to specify an amount. Note for (7.19) that *pá:t* 'four' is canonically monosyllabic.

(7.19) *ni:4 ʔeká:n ... pá:t p^hʂh má:ʔ p^hʂh*
 this fish four CLF five CLF
 ‘These are fish...four...five (of them).’ [NN.P6.1-2]

(7.20) *paʔó:k ... ʔó:k wà:ʔ lât*
 mango mango two CLF
 ‘Mangoes. Two mangoes.’ [CU.A21.1]

In (7.20), we also see the monosyllabic alternant for *bulât*—a general classifier for non-human entities used exclusively for numerical values with either ‘one’ or ‘two’ in the ones place (see §2.4.2). Both word-forms were attested within the data, the disyllabic form to a lesser extent (see Table 30), and both could take the numerical proclitic *ʔa-* ‘one’ (e.g., *ʔabulât* and *ʔalât*). It is interesting to note that despite only having 5 tokens, disyllabic *bulât* was always in a clause-final focused position, like in (7.21) where it is used in emphasizing the wholeness of a pig on a spit (see Figure 42).



Figure 42 Resolution Scene sequence #12

(7.21) *mə4nut4 pá:t lùj ... pəniáŋ babú:j t^haŋ4 bulât*
 people four CLF roast pig all CLF
 ‘The four people...(they’re) roasting the whole pig.’ [WN.A12.3]

Information status marking for the classifier *lât* differed from other closed-class items as it could also serve as the head of a referring expression and, therefore, tagging at both the lexical and referential level were applicable. For example, in (7.22) *lât* is used to provide definite reference to individual chickens. One interpretation of these constructions would be that the nominal heads of these referring expressions are ellipsed and thus the monosyllabic alternant is the remaining overt topical

information. Note also in (7.22) the word-form shift for the verb *kəp^hlà:t* ‘to fall’ with the integration of more verbal elements.

- (7.22) *manók* (...) *Ø làt ni:4 kəp^hlà:t* ... *Ø làt ni:4 p^hlà:t bú:k lɔ:jl*
 chicken CLF this fall CLF this fall under INTS
 ‘Chickens. (...) ‘This one’s falling. This one’s falling all the way down.’
 [LP.A15.1-2]

The other two closed-class items were the prepositions *dalám* ‘in’ and *datá:* ‘on’, both of which were more common in their monosyllabic form. Again, given the relatively low occurrence of these alternants, we are limited in how much can be said. However, a pattern of word-form shifts, similar to what occurs with verbs, has also been noted for these prepositions. For example, in (7.23), *datá:* ‘on’ ends the first IU without a grammatical object, but in the next IU monosyllabic *tá:* occurs at the left-edge followed by the additional remaining information.

- (7.23) *ʔɔj nəmá:t ʔé:m datá: ... tá: caná:t*
 dog startled be.located on on child
 ‘The dog gets scared and is on...on the child.’ [Field data]

As tokens for closed-class items were less attested, no determination about givenness effects for the use of a monosyllabic alternant is put forth. But, that many closed-class lexemes typically function within larger grammatical units, like numerical compounds or classifier phrases, does at least suggest a relative alignment with backgrounded information. While this is tempting especially given the thrust of the topical interpretation, the paucity of data for closed-class items, as well as the need for more theoretical development of information status annotation of grammatical items, means we are limited in how much can be resolved here.

7.3 Topical shift

Identifying units of *topic* and *focus* is dissimilar to information status as it relies on interpreting the communicative intention of speakers. However, like information status, topic and focus deal with updates to common ground content. Basically, the topic-focus distinction has *utterances*³² bifurcated into two informational units: one based upon common ground and one that updates it. Framed within speech act theory, Cresti, (2018) says the core part of utterances is their illocutionary force—the effect of the speech act intended by the speaker. *Topics* serve as the conceptual domain or field of application for the illocutionary force, while *focus* achieves the illocution (e.g., asserting a proposition). Connecting these functional definitions to information structure, Masia (2022, p.83) presents an epistemic profile of topic and focus where they are regarded as encoding meanings of “broad evidentiality” within conversation. The two units are defined as follows:

Focus encodes information conveyed by the speaker as her communicative intention and as individual knowledge of which she is the only epistemic source.

Topic encodes information not conveyed the speaker’s communicative intention and which represents mutual knowledge established as shared conceptual grounding with both speaker and hearer as committed source.

Masia emphasizes in this view that topic and focus are not holders of content but a reflection of how speakers *present* content. Moreover, Masia distinguishes topic-focus as working at a broader level of “packaging”—one which is distinct from an “activation” level of *given* and *new*. Still, information presented as topical might coincide with “given” expressions; however, the key distinction is that a topical shift signifies a matter of *discourse commitment* from both speaker and hearer, regardless of any underlying “cognitive states”.

Word-form shifts from other discourse contexts offer further insight into the reduction of the disyllable during speech. For example, interactions between a main

³² Crystal (2008) defines utterance as “a stretch of speech about which no assumptions have been made in terms of linguistic theory” (p.505). I am mainly switching here to capture the possibility that stretches of speech that include more than one intonation unit, yet are semantically or syntactically connected (e.g., phrasal IUs) work in this way.

speaker and a Moklen hearer from within the Stolen Fish corpus attest to the same types of patterns outlined within the TEPS data. Beginning with (7.24), the speaker NA is describing the scene where the fisherman is invited to join the picnic. NA's utterance is a causative construction with *ʔón* 'to give' taking a complement clause, which contains the only overt referring expression *ʔeká:n* 'fish'. As NA's utterance ends, DW, who has been listening to the narration, interjects that this fish in fact belongs to the fisherman. Note though, DW began her utterance with monosyllabic *ká:n* 'fish'. In terms of topical information, it is clear that in this exchange *ká:n* is the portion of the utterance to which both speaker and hearer are epistemically committed. DW's communicative intention is to assert the fisherman's ownership of the fish, but this focus is built upon the shared conceptual grounding of monosyllabic *ká:n*.

(7.24)

- NA: *cʰu:an1 ʔón dín ɲám ʔeká:n*
 invite give come eat fish
 '(They) invite (him) to come eat fish.' [NA.SF26.87]
- DW: *ká:n pɔʔ káʔ né:*
 fish 3SG PRF PRT
 '(It's) his fish.'

Other examples of word-form shifts across dialogic interactions show the same pattern. For example, the sequence in (7.25) from a "Frog Story". In this exchange, CU, the speaker, is describing a scene and has placed *ticúm* 'bird' as a grammatical object. His wife, PN, is within earshot but does not have visual access to the illustration. Interestingly, she interprets a grammatical error within CN's utterance by mistaking the joint action sense of *ʔabó:ʔ* 'together', which has scope over the VP as a misuse of a classifier phrase *ʔa=bó:ʔ* 'one person'—a distinct usage reserved with human nominals (Larish, 1999). PN then interjects with the "correct" classifier phrase *ʔabulât*, as the corresponding nominal expression was a bird. Note, however, that she begins with monosyllabic *cúm* squarely as the topic of the utterance. CU responds by clarifying that there were many birds, not a single bird. Yet again,

monosyllabic *cúm* is the topical portion of the utterance—the information to which both speaker and hearer are committed.

(7.25)

- CU: *caná:t kʰu:3 ʔɣj cʰɔláʔ ticúm ʔabó:ʔ*
 child with dog shout bird together
 ‘The child and the dog are shouting at the birds together.’
- PN: *cúm ʔa=bulàt*
 bird one=CLF
 ‘A SINGLE bird’
- CU: *cúm lá:j pʰɔh bé ... ʔa=bulàt háh*
 bird many CLF PRT one=CLF NEG
 ‘There are many birds, not a single one.’

Dialogic examples like (7.24)-(7.25) are useful as they more directly map onto the speaker and hearer framing of Masia’s (2022) definitions of focus and topic. For the TEPS monologic examples, givenness correlations across word-form shifts may only be a proxy for the topical portion of the utterance. Consider that more than half of the verbal monosyllabic alternants in the TEPS data were lexically new (§7.2.2). Information status on its own—in this case lexical givenness—most likely is not capable of discriminating the degree to which speakers are actively backgrounding a verbal concept. But with word-form shifts, a verbal monosyllabic alternant’s relation to topical information is seen through positionings near ellipsed arguments and their role as starting points for further information.

Several other relevant aspects can also be seen within the Stolen Fish corpus. For example in (7.26.1), we first see disyllabic *mékén* ‘to gather’ appear after an overt grammatical subject *caná:t* ‘children’, while the direct and indirect objects of a second verb, ditransitive *ʔón* ‘to give’, have yet to be specified. In the next IU (7.26.2), however, the speaker omits the agent and, begins with the topical VP, featuring monosyllabic *kén*, before supplying the previously omitted arguments.³³

³³ There is also the interesting word-form shift here from *caná:t* to *ná:t*. This would be merely a lexical shift as the IO argument is not coreferential with the previous A argument. The most prosodically prominent argument here is definitely *ʔeká:n* ‘fish’

(7.26)

(7.26.1) *ʔaná:t nɤj mɛkén=ʔá: ʔón lá:j*
 child.offspring DEM.PROX gather=3SG give PROG
 ‘The child is gathering and giving (them).’

(7.26.2) *Ø kén=ʔá: ʔón ʔeká:n ná:t nɛ:*
 gather=3SG give fish child PRT
 ‘(He’s) gathering and giving the fish to the kid, see!’ [TG.SF13.53-54]

An additional line of evidence comes from examples of word-form shifts occurring across question-and-answer sequences, an important context for documenting topic-focus relations (Aissen, 2023). In (7.27), GP is narrating an interaction between characters of *Stolen Fish*. First, GP has the fisherman ask the picnicking group if they had stolen the fish. The question features disyllabic *nəlát* ‘to steal’ and overt reference to the object *ʔeká:n* ‘fish.’ GP then has the group respond “No”, but crucially, the answer begins with the monosyllabic form *lát* ‘to steal’ before then focusing on negating the proposition with the post-verbal negator *háh*.

(7.27) GP: *já:j nəlát ʔeká:n ká:ʔ*
 say/think steal fish QPLR
 ‘(They) say, “Did you steal these fish?”’

GP: *lát háh*
 steal NEG
 ‘No, (we) didn't steal (them).’ [GP.SF26.63-64]

A common question-answer sequence in (7.28) highlights another example of verbal word-form shift. Here, using the verb *bətáy* ‘to be satiated’, speaker A can ask B if they have had enough to eat. The proposition can be affirmed with the post-verbal perfective marker *káʔ*, but the topical portion of the utterance is conveyed through use of monosyllabic *táy*.

(7.28) A: *bətáy ká:ʔ*
 be.satiated QPLR
 ‘Are you full?’

B: *táy káʔ*
 be.satiated PRF
 ‘Yes, I am.’

Reviewing Moklen discourse with a view of minor-syllable elision and resulting monosyllables as aligning to a topical shift seems to clarify some of the observed variation during speech. Consider this excerpt from a telling of the Legend of Sampan, the progenitor of the Moklen people.³⁴ Starting from (7.29.1) the speaker is asserting that Sampan's home (*ʔɔmá:k ʔɛbá:p*) had moved to a new location, indicated with the demonstrative pronoun *nú:* 'there', and that it was no longer in Nakhon Si Thammarat, a location on the opposite side of the Thai peninsula. Then in (7.29.3), the speaker further clarifies that Sampan's home was in Bangsak, a modern location of a Moklen community. But this is done with monosyllabic *má:k* 'home' for the topical portion of the utterance, which also features omission of the previously used modifier *ʔɛbá:p* 'elder male'.

(7.29) [Arunrungsawat et al., 2018]

(7.29.1) *ʔɔmá:k ʔɛbá:p ʔé:m héʔ nú:*
 home male.elder be.located at DEM.MED
 'Elder's home was there.'

(7.29.2) *láʔ Ø ʔé:m háh káʔ tʰi:3 na4kʰɔ:n1*
 EMPH.NEG be.located NEG PRF at Nakhon
 'It wasn't in Nakhon.'

(7.29.3) *má:k pin4 ʔé:m héʔ tʰi: lusák*
 home COP located at at Bangsak
 '(His) home was in Bangsak.'

Findings from the TEPS data and continued textual analysis seem to offer a new understanding of Moklen monosyllabic alternants. Crucially, by bringing their discourse context into view, a subtle informational aspect of the monosyllabic alternants is brought into light. Further underscoring this new analysis is that even within the only previously published Moklen text, (Larish, 2005), there is a word-form shift that appears consistent with the topical interpretation. The text features a dialogic exchange that is shared in (7.30). The example begins with a question from A whether it was raining heavily at a specific moment and uses an adverbial clause headed by the Thai loan *wella:l* 'time.' The response by B starts by mirroring the adverbial clause, which can be taken as the topical element. But note that this entire

³⁴ See §2.2. Also see Ivanoff (2001) for Moklenic folktales.

constituent starts with the monosyllabic alternant *la:l* ‘time,’ and features omission of the argument *lot4k^hluan3* ‘motor vehicle.’

(7.30) [Larish 2005:531]

A: [*we:l*:1 *lot4k^hluan3* *nəmóh*] *kɔ́já:n* *nɔ́léŋ* *lahán* *háh* *ká:ʔ*
 time motor.vehicle fall.down rain descend many NEG QPLR
 ‘When the motorcycle fell, wasn't it raining heavily?’

B: [*la:l* Ø *nəmóh*] *ɲú:* *kɔ́já:n* *plɛʔ* *plɛʔ* *já:* *láʔ* *lahán* *háh*
 time fall.down DEM.DIST rain little only EMPH many NEG
 ‘When it fell, there was only a little rain, not much at all.’

From both the patterns of word-form shifts within the TEPS data and supporting evidence from other discourse contexts, the changes in word-form, or more specifically cases of minor-syllable elision during speech, appear to be driven at least in part by a shift towards topical information. Basically, the picture on offer here is one of Moklen speakers picking up others’ (or their own) ideas in following utterances and then, sometimes, presenting mutual knowledge with a reduced word-form. The elision of minor-syllables of previously disyllabic lexemes therefore tends to correspond with a topical shift. This account of changes in word-form does not entail categorical distinctions but, rather, characterizes an interesting tendency within discourse.

7.4 Discussion

Monosyllabic alternants are a reduced monosyllabic form of a Moklen lexeme for which there is still an attested corresponding disyllabic form e.g., *ɲú:n* from *niɲú:n* ‘coconut’. Prior to this study, Moklen’s monosyllabic alternants and changes in word-form had never been a direct target of study. Larish (1999) coined the term “non-ultimate syllabic aphaeresis” to describe a synchronic phenomenon of “optional” deletion of minor syllables, but based on this current study, simply viewing the two word-forms as equivalent entries within a static picture of the lexicon seems like a misgeneralization. To the extent that reduction of disyllabic words during speech needs a specific term, I prefer the more direct phrasing: *minor-syllable elision*.³⁵ The

³⁵ One main motivation for positing “minor-syllable elision” was to signify a reassessment of the phenomenon Larish (1999) describes as “non-ultimate syllable aphaeresis”. In using “minor-syllable”, I

main evidence for the phenomenon is the directionality of disyllabic > monosyllabic word-form shifts within discourse. Explicitly defined, minor-syllable elision would be a form of clipping during discourse whereby iambic disyllables become abbreviated through omission of the initial minor syllable. In these instances, when minor-syllable elision produces monosyllabic alternant forms, they can be seen as arising from dynamic discourse processes happening at a time frame of connected speech (Enfield, 2016), rather than just lexical retrieval of an alternant “colloquial form”. Positing minor-syllable elision as a broader tendency, however, does not rule out the possibility of specific lexemes having equivalent or functionally distinct word-forms. At this point, though, having at least documented and detailed minor-syllable elision as a general tendency, we seem to have some grounds for discussing narrower discourse factors leading to the reduction of canonically disyllabic lexemes.

The appearance of monosyllabic alternants in an IU-initial position was a key clue in understanding minor-syllable elision. One potential view of the reduction of disyllabic word-forms would be to see it as a morphological consequence of prosodic deaccentuation at the level of the word. An alternative view concerns the role of *prominence* as a dynamic principle shaping discourse (von Heusinger & Schumacher 2019) and the shape of intonation units. Working with Baumann and Cangemi's (2020) definition (see §3.3.3), with “prominence”, here, we are only concerned with relative differences in phonetic material and not any singular acoustic cue. Notably, for Moklen intonation units, the IU-initial position is reliably an area of lower prosodic prominence and, moreover this was also a frequent site of the reduced monosyllabic alternants. For example, let us revisit an example of minor-syllable elision in (7.32) where in two IUs monosyllabic alternants for *kətá:m* ‘crab’ and *ʔeká:n* ‘fish’ appeared in the IU-initial position.

have ignored Larish’s distinction between “pre-syllable” and “minor syllable” a phonetic distinction not discussed within this study. As for phonological processes, one option would have been “procope”, which is used by Matisoff (1990) to describe the change of sesquisyllables to monosyllables, but use of this term is seemingly rare. Since the term “apheresis” implies loss of sounds at the beginning of the word and minor syllables are at the beginning of Moklen disyllables by definition, use of it seemed somewhat redundant. “Elision”, on the other hand, is more general and is distinguished from a diachronic sense of “deletion” and is associated to losses in natural connected speech, which seemed to align with the correct time frame in which to consider this phenomenon.

(7.32)

(7.32.1) *ʔeká:n namá:ʔ lám tʰaŋ⁵*

fish enter in bucket

‘A fish is going in a bucket.’ [LW.P14.1]

(7.32.2) *ká:n bé:t ca:k² tʰaŋ⁵ ... kətá:m kʰi:p³ =ná:*

fish exit from bucket crab pinch=3.SG

‘The fish goes out of the bucket. A crab grabs (it).’ [LW.P14.2]

(7.32.3) *tá:m nám ʔeká:n*

crab eat fish

‘The crab eats the fish.’ [LW.P14.3]

Juxtaposing waveforms for these two IUs in Figure 43, we can see the place of prominences as it relates to the monosyllabic alternants *tá:m* and *ká:n*. Notice how the remaining major syllables of each monosyllabic alternant *tá:m* and *ká:n* appear less prominent within their respective IUs. Most visible are the relative differences in intensity and duration between elements at the beginning and end of each IU. Additionally, note the occurrence of both word-forms for *ʔeká:n* ‘fish’ while “given”. Hence, the key difference is their position within the intonation unit. Finally, underscoring the differences is the comparative contrast in prominence for the syllables *ká:n* in each word-form, where it is composed of much more phonetic material in the IU-final position.

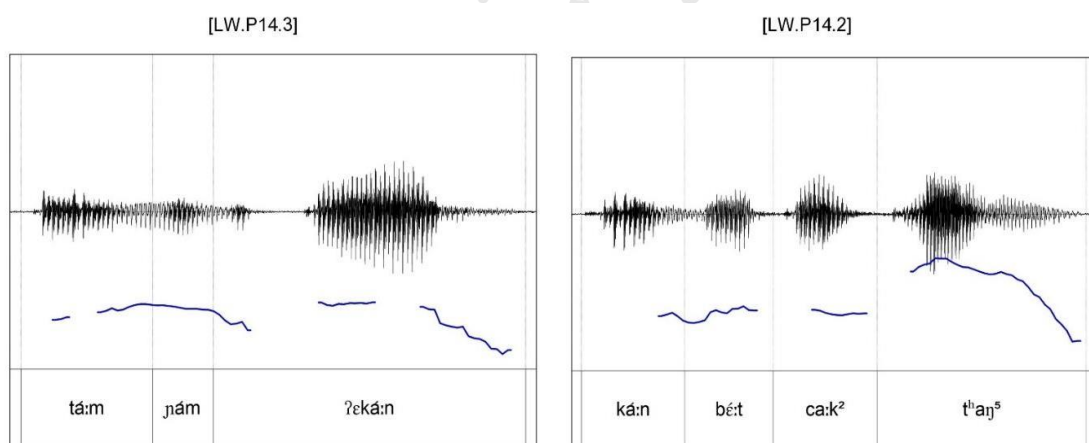


Figure 43 IU-initial low prominence with minor-syllable elision.

Low prominence has been put forth as a correlate of givenness (Lambrecht, 1994). Here, though, I am highlighting how characteristically topical portions of

clausal IUs – which may often contain “given” elements – may also be a frequent site of low prominence. A crucial context for minor-syllable elision therefore seems to be having an utterance’s point of departure *sufficiently* within the common ground. Only then will the attenuated pronunciation of a topical disyllable cross a morphological threshold causing the relatively weak and unstressed minor syllable to be lost. This in turn helps explain why sometimes the monosyllabic alternant might not appear until after a few repeated mentions. This is the case in (7.33.3) where the minor-syllable elision of *kabá:ŋ* ‘boat’ occurred in the third and final mention of the resolution scene. In other words, although there are reasons to believe words and referring expressions are susceptible to givenness effects, it also makes sense to think about the intonational contour of IUs as also reflecting types of informational profiles with respect to their prominences.

(7.33)

(7.33.1) *kabá:ŋ b̀̀t*

boat run

‘A boat’s going.’ [L.I.P19.1]

(7.33.2) *kabá:ŋ k̀̀j bat̀̀j l̀̀:jl*

boat to.ground rock INTS

‘The boat grounds onto a rock!’ [L.I.P19.2]

(7.33.3) *b̀̀:ŋ kalám l̀̀:w4*

boat sink PRF

‘The boat has sunk.’ [L.I.P19.3]

This study sought to frame changes in word-form from the perspective of information structure. In total, there were 64 disyllabic lexemes within the TEPS data alone that displayed monosyllabic alternant forms; data from other sources could supply dozens of additional lexemes to a total number of attested monosyllabic alternants for Moklen. And while it is true to say all Moklen contributors to this study use monosyllabic alternants and display word-form shifts, I neither want to overplay the extent of minor-syllable elision nor ignore other contributing factors to the appearance or persistence of monosyllabic alternants, as many disyllables across the lexicon appear to be relatively stable. Therefore, which disyllabic lexemes are

susceptible to minor-syllable elision is likely to include a myriad of phonetic and phonological factors, which is outside the scope of our discussion here.

One important issue, as it connects to conceptual matters for speakers, is the role of lexicalization. Lexicalization is a process of adding new “holistically processed linguistic units” (Hilpert, 2019) to the lexicon. A key aspect of lexicalization is that it entails the encoding of formerly separate conceptual components into one lexical unit (Levin & Hovav, 2019). As noted by Larish (1999), the monosyllabic form often appears as the head of a nominal compound. This can give the impression that certain monosyllabic nominal alternants serve either as a type of class noun or that compounds with monosyllabic heads are a more unified lexical unit, such as with classifications of types of *ʔeká:n* ‘fish’ (e.g., *ká:n dapú:ʔ* ‘grouper’, *ká:n bəláy* ‘mackerel’, etc.). Importantly, however, there appears to be no semantic restriction on the use of disyllables as nominal heads in these constructions. For example, in (7.34) we see both *pəkón* and *kón* serving as a nominal head for ‘coconut tree’.

- (7.34) *ni:4 pəkón ... pəkón niʔú:n ká:ʔ ... kón niʔú:n buwá:k ʔót*
 this tree tree coconut QPLR tree coconut fruit depleted
 ‘This is a tree. A coconut tree? A coconut tree with no fruit.’ [YN.P11.1]

Thus, there is a lingering question as to what degree monosyllabic nominal heads can be associated with a more holistically lexicalized unit. As brought up in §7.2.1, we might argue that in (7.35), the appearance of reduced *buwá:k* ‘fruit’ in *wá:k niʔú:n* ‘coconut’ whilst l-new&r-new might reflect a level of lexical integration. However, since so many of these compounds *do not* have non-compositional meanings, we are without one reliable diagnostic of lexicalization (Hilpert, 2020). Furthermore, given the underlying structure of nominal+predicative information within the compound, the broadest interpretation, despite any “newness” (referentially or lexically), is that the monosyllabic alternants still at some level represent *topical* information.

- (7.35) *nɿ wá:k niʔú:n ... wá:k ʔú:n bəluəŋ*
 DEM.PROX fruit coconut fruit coconut fresh
 ‘This is a coconut, a fresh coconut.’ [NK.P2.1]

In the long view, even if we did discover some compounds with monosyllabic heads are indeed lexicalized, it is still hard not to see the origin of such constructions

as arising from a process of minor-syllable elision. Capturing instances of minor-syllable elision therefore provides insight into the types of “bridging contexts” in which certain lexicalizations arose (Traugott, 2012). For example in (7.36), the three compounds featuring the modifier *melá:k* ‘red’ have less compositional meanings.

- (7.36) *ʔaná:t* ‘child.offspring’ *ná:t melá:k* ‘newborn baby’
 ʔaták ‘head’ *tá:k melá:k* ‘white person of European ancestry’
 ʔeka:n ‘fish’ *ká:n melá:k* ‘red-bellied fusilier’

Word-form reduction could also reflect a layer of integration within the frame of particular collocations: the co-occurrence of two words in some defined relationship (Yarowsky, 1993). For example, there appear to be several verb+object collocations which feature monosyllabic alternants such as these examples with *ʔuʔé:n* ‘water’ as an object, as in (7.37).

- (7.37) *məʔám* ‘to drink’ + *ʔuʔé:n* ‘water’ = *ʔám ʔé:n* ‘to drink water’
 məʔé:n ‘to bathe’ + *ʔuʔé:n* ‘water’ = *ʔé:n ʔé:n* ‘to bathe’

As there are attested disyllabic variants of these constructions, again there appears to be no semantic restriction on the use of full disyllabic forms for either lexeme here. However, given collocational effects can have an effect on word-shortening (Gregory et al., 1999), the appearance—or more accurately the *persistence*—of monosyllabic alternants within these constructions cannot only be attributed to discourse effects. For Moklen verbs, Larish (1999) points out that compared to synthetic insular-Austronesian languages, Moklenic languages could already be seen to be shifting towards the analytic norms of Mainland Southeast Asian languages. Part of this includes a diachronic loss of verbal morphology, like the missing nasal-initial minor-syllables of contemporary verb forms in *dín* ‘to come’, *káw* ‘to go’, and *dán* ‘to know’. Other verbal lexemes are more variable in the realization of the minor-syllable, but do often appear reduced in what are likely grammatical collocations. For example, verb+preposition constructions featuring *naʔék* ‘to ascend’ and *ləʔɛ:m* ‘to be located’ with prepositions *dalám* ‘in’ *datá:* ‘on’ often have reduced forms.

- (7.38) *naʔék* ‘to ascend’ + *data:* ‘on’ = *ʔék ta:* ‘to get up on’
 ləʔɛ:m ‘to be located’ + *dalam* ‘in’ = *ʔé:m lam* ‘to be inside of’

The original interest in the word-form shifts came from the idea that the reduction of disyllables within speech reflected a type of information integration. For

verbal lexemes, one premise of the topical interpretation was that the appearance of the monosyllabic alternant in more complex predicates indicated alignment to backgrounded information. The addition of a new argument to a previously bare predicate, such as the change from *nújú:k* > *jú:k* ‘to point’ in (7.39) with the addition of *ticúm* ‘bird’ is one example.

- (7.39) *nújú:k nɛ ... jú:k hɛʔ ticúm nɛ*
 point PRT point at bird PRT
 ‘(He’s) pointing. (He’s) pointing at the bird.’ [YG.SF20.060-061]

On the other hand, other types of grammatical integration can also be noted. For example, a common possessive construction like *má:k pɔʔ* ‘their house’ is heard more than *ʔómá:k pɔʔ*. Or even with coordination, like in (7.40) where monosyllabic forms for *ʔapɔ́ŋ* ‘father’, *ʔaká:ʔ* ‘older.sibling’, and *ʔuʔu:j* ‘younger.sibling’ appeared together. In terms of information status the whole referring expression was *r-given*, while lexically, *pɔ́ŋ* was *l-given*, and *ká:ʔ* and *ʔúj* were *l-new*. Grammatically integrated in an unmarked coordination structure, they appear together as a detached phrasal IU.

- (7.40) *təlɔ́j lùj nú: ... pɔ́ŋ ká:ʔ ʔúj*
 three CLF DEM.MED father older.sibling younger.sibling
 ‘Those three people... the dad, the older one, and the younger one.
 [HJ.SF.41-42]

Particular verb-object combinations, and serial verb constructions more generally, also seem like likely contexts wherein informational pressures lead to the reduction of disyllabic word-forms. For example, in what is conveyed as one integrated event, (7.41) has two verbs and an object in a VOV pattern, none of which are in their attested disyllabic forms (*nəbá:k*, *ʔuʔé:n*, and *məʔám*, respectively)

- (7.41) *bá:k ʔé:n ʔám*
 to.ladle water drink
 (She’s) ladling water to drink.’ [NN.A3.2]

The topical interpretation in its starkest terms represents the idea that along a spectrum of overt forms, the disyllabic word-forms correspond with *more focus* while the monosyllabic alternants are *less focused*. However, the big picture for monosyllabic alternants overall is that there are factors leading to the reduction are at

several levels, such as lexicalization, grammaticalization (Lehmann, 2002), and more generally, frequency effects stemming from the distributional properties of words (Arnon & Snider, 2010; Aylett & Turk, 2004). Therefore, there is much more to the use of monosyllabic alternant than a pure alignment to “topical information”. A fuller account of Moklen word-form would need to encompass several lines of research, but in principle the issue seems explainable through information theoretic and usage-based approaches, all which rest on the idea of efficiency-based motivations in shaping linguistic structures (Jaeger & Buz, 2017; Mahowald et al., 2013). Put simply, there is much more to consider beyond givenness and topichood that could contribute to the types of redundancy that might lead to minor-syllable elision. However, based on some attested tendencies in speech, it looks like information structure factors may have consequences for Moklen word-form.

7.5 Summary

To study the relationship between Moklen word-form and information status, Hypothesis 3 (“Use of monosyllabic alternants will correspond to “given” information statuses”) was tested. Initial results pointed to mere “givenness” as not indicative of overall usage of monosyllabic alternants. Further examination of word-form shifts led to the idea that many monosyllabic alternant forms were aligned with topical information. For nouns, the topical interpretation was evidenced with monosyllabic alternants being either a given grammatical subject or the head of a nominal compound. For verbs, alignment to backgrounded information was shown through the integration of more elements into the predicate and proximity to positions of ellipsed arguments. Findings for closed-class items were limited but still followed similar patterns found with other classes, such as the appearance of the monosyllabic alternant in an intonation-unit-initial position, where the reduced alternant can function as a starting point for additional information. Finally, framing the whole phenomenon as motivated by matters of information structure, I proposed that one way to understand minor-syllable elision was to see it as reflecting a *topical shift*—a move of information to a shared conceptual grounding. This explanation does not account for all uses of monosyllabic alternants, but instead proposes an informational aspect of the observed changes in word-form.

8 Discussion & Conclusion

8.1 Introduction

This study took on ideas from the field of information structure as heuristic for exploring Moklen. The study aimed to fill in the gap for discourse-based research on Moklen and investigate some of its variant linguistic forms. First, there were the alternant word orders besides SVO, but the use of reduced monosyllabic alternant word-forms was of special interest. Findings on informational and syntactic properties of intonation units showed that clausal intonation units are usually limited in number of referring expressions and have an AVO preferred argument structure. As for monosyllabic alternant forms, they appeared throughout discourse but were beyond a solely information-status based account. Instead, frequent instances of word-form shifts, specifically the change from a disyllable to a monosyllable, were interpreted as outlining a move towards topics. Based on this interpretation, a process of minor-syllable elision was put forth as arising from discourse-conditioned reductions in prominence. Overall, this study offers a picture of Moklen morpho-syntactic structures in their discourse context and how they relate to matters of information structure.

8.2 Informational properties

Informational properties of intonation units were studied to better understand information structure factors influencing Moklen speech. Findings here must first be understood as pertaining to the context of the Stolen Fish narration task. In line with Hypothesis 1, clausal intonation units conformed to the one-new-idea constraint as they rarely ever had more than one new referring expression. Further, in the very few cases where they did have more, they were found to contain functionally unitary verb-object constructions, a factor predicted in Chafe's original framing of the constraint.

Operationalization of the one-new-idea constraint as a limit on "new" R-EXPs drew further attention to a general limit of overt R-EXPs. This finding reveals how speakers, when constructing a Stolen Fish text, largely use chains of IUs, relying on the previous activation of discourse referent and events. One question raised by this picture regards the role informational properties of utterances should have in positing

grammatical structures. For example, long “sentences” with several overt arguments, although attestable, are informationally marked and less representative of the majority of speech. Use of such sentences for the basis of linguistic analyses has been the subject of critique from Chafe (1994) and within discussions on the “written language bias in linguistics” (Linell, 1982, 2019). Zero-anaphora is a well-known factor for languages of the region (Michaud & Brunelle, 2016), but this can easily be forgotten when linguists need to display overt grammatical strings for grammatical analysis. Often these examples are drawn from written sources, but then connected to cognitive implications, again exposing a failure to overcome the *written language bias in linguistics* (Linell 2019).

One interesting role for an informational perspective on grammar is that it appears to provide one criterion for capturing unitary grammatical constructions within speech. For example, marked violators of the one-new-idea constraint were shown to make use of certain types verb-object constructions that were determined to be functionally unitary predicates (e.g., *tʰɔ:tɔ hɛ:5* ‘cast a cast net’) and not demand an extra activation cost for the new R-EXP. That investigation into the one-new-idea constraint uncovered this, suggests an interesting cognitive basis for further lexicalization/grammaticalization research.

The study of Moklen intonation units described some of the informational pressures shaping their contents within a particular discourse context. In the broader context, intonation units (or similar units) seem poised for continued significance within contemporary discussions of linguistics. Here, they are framed as “basic units of speech” (Izre’el et al., 2020), a part of understanding morphosyntax (Croft, 2022), a key component at the bottom-most scale of linguistic structures (Enfield, 2023), and the linguistic unit that is closest to Zipf’s law of abbreviation (Linders & Louwerse, 2022). This study, therefore, serves as an example of the role of intonation units in framing a variety of linguistic questions.

8.3 Syntactic properties

The syntactic properties of Moklen intonation units were studied to better understand the place of alternant word orders in discourse. Hypothesis 2 held that variation in argument structure and associated grammatical constructions could be

accounted for by the given-before-new principle. However, a premise of the hypothesis was challenged when responses to the Transitive Event Picture Sequences reliably demonstrated a consistent AVO pattern under both givenness conditions. The search for the variant orders continued in the Stolen Fish corpus, but again AVO was shown to be the main pattern of transitive clauses. As put forth by Larish (1999) and acknowledged in this study, there are some uses of pronominal clitics, but at this point their connection to information structure and argument structures is anecdotal and not entirely clear.

One conclusion, however, is that alternant orders of clausal constituents, to whatever extent they are used by Moklen speakers, are not motivated simply by givenness. Crucial for understanding the previous reports of alternant word orders was that they were not based on naturally occurring speech. Instead these claims came from translation-based direct elicitation (Larish, 1999; Swastham, 1982). It seems, however, that with such a basic grammatical construction, which can easily be reduced to three-letter acronyms, claims for potential variation require a certain level of methodological/empirical clarity, especially when overt realization of all clausal arguments is not the norm. Given the disparity between findings in this study and previous reports, issues of methodology and their theoretical significance should be addressed.

A case in point is Baclawski Jr. & Jenks's (2016) study on *Moken* syntax. Here, they are quick to use various word order permutations (SOV, OSV, VSO) elicited out of context. Here, they rely on their consultants' "strong grammaticality judgements" as the basis for "word order facts" before arguing for abstract theoretical dependencies between clausal grammatical elements. To their credit, they do admit that data from narrative texts displayed "zero credible instances" of non-SVO utterances, but also they add, "we assert that our consultants provided strong grammaticality judgments in elicitation and that narration typically exhibits low information load per sentence, at least in this narrative tradition." (p.83). The current study can corroborate the remark about a low information load per "sentence", but just like the discrepancy in this study and previous reports of Moklen's variant word orders, the difference between Baclawski Jr. and Jenks's narrative data and their

consultant's grammaticality judgments leads one to question claims based solely upon speakers' metalinguistic judgements.³⁶

Given the pattern of such discrepancies, we should seriously question the use of positing variant word orders when they are not found within actual communicative contexts. The given-before-new principle was originally sought as a potential account for the reported variant word orders. But as the TEPS instrument failed to elicit alternant argument structures, one well-known factor influencing word order has seemingly been eliminated. At the same time, however, previous methods for researching word-order alternations (Prentice, 1967; Skopeteas & Fanselow, 2009) were the inspiration for investigating other phenomena (i.e., word-form). And, more practically, it offered an interesting way to discover other discourse strategies from Moklen speakers. Further research on "oral grammar" (Fernandez-Vest, 2015) should consider these methods along with analysis of clausal syntax's informational properties.

8.4 Moklen word-form

Most Moklen lexemes take the form of an iambic disyllable made up of an initial unstressed minor syllable and a stressed major syllable. Because this is the most prevalent word-form, monosyllabic alternants and changes in Moklen word-form were an object of interest. Previously, the monosyllabic alternants were described as "colloquial forms" (Court, 1971; Larish, 1999; Lewis, 1960; Wolff, 2010). After preliminary fieldwork, a hypothesis was put forward that use of monosyllabic alternants in speech would correspond to "given" information statuses. Findings from this study, however, revealed that mere givenness was not completely adequate for capturing the full range of contexts and constructions featuring monosyllabic alternants. Alternatively, a connection to information structure was argued to be observable within patterns of word-form shifts. Accounting for the pattern of disyllabic before monosyllabic, a process of *minor-syllable elision* was put forth. Clarifying the colloquial nature of monosyllabic alternants, this tendency for word-form reduction was cast as corresponding to a shift towards topical information. This

³⁶ For more on the critique on the methodology of grammaticality judgements in linguistics see Schütze (2016) in which it is basically argued that grammaticality judgements are not pure sources of data but instead "instances of metalinguistic performance".

finding contextualizes one aspect of Moklen word-form and offers an interesting example of the potential of discourse factors to impact word-form. The appearance and persistence of monosyllabic alternants overall, however, is a much larger question.

8.4.1 Minor-syllable elision and information structure

The main evidence for minor-syllable elision and monosyllabic alternants' topical alignment is the pattern of word-form shifts in the TEPS data. In terms of information status, though this pattern does correspond with a change from “new” to “given”, other features can be noted. For nouns, monosyllabic alternants often appeared as given grammatical subjects or as the heads of nominal compounds. For verbs, an alignment towards backgrounded information was inferred from the integration of more elements into the predicate and a proximity to positions of ellipsed arguments. Findings for closed-class items were limited but still outlined similar patterns as nouns and verbs, such as integration into grammatical constructions and an IU-initial position where the monosyllable functions as a starting point for additional information. Given the character of these word-form shifts, elision of the minor-syllable was seen as corresponding with shifts to topics: “mutual knowledge established as shared conceptual grounding” (Masia, 2022 p.83). Additional support for this view also came from dialogic interactions and question-answer sequences, in which word-form shifts between interlocutors were on display.

One lesson from this study is that models of information status may miss out on other informational distinctions, such as *topic* and *focus*. Put simply, mere tracking of anaphoric links and uses of a lexeme is likely to overlook other variables. However, the RefLex scheme and the TEPS data were a useful lens with which to explore this previously unstudied phenomenon. To better account for information theoretic factors leading to a process like minor-syllable elision, a much more rigorous experimental framework is needed (e.g., Kanwal, Smith, Culbertson, & Kirby, 2017). Supporting further research in this direction, findings from this study provide one picture of how information structure may impact word-form.

8.4.2 Diachronic implications

A broader implication of minor-syllable elision is its inclusion as a contributor to diachronic monosyllabization. A diachronic trend towards monosyllabicity is something Larish (1999, p.182) and Wolff (2010, p.526) had already connected to Moklenic loss of minor syllables (see §1.1.2). More generally, word-form changes, from disyllables to monosyllables in particular, is a prominent topic within Mainland Southeast Asian (MSEA) linguistics, wherein language contact is often framed as a catalyst of diachronic shifts (cf. Brunelle & Pittayaporn, 2012).

Larish originally saw diachronic changes, like a shift to monosyllabism, as analogous to changes in Chamic languages, fellow Austronesian languages of MSEA in contact with Austroasiatic languages. However, for Cham, Brunelle (2020) argues that contact-induced change probably happened in indirect ways and that monosyllabization was more likely driven by internal phonetic and phonological pressures.³⁷ Moreover, Chamic disyllabic and monosyllabic forms are said to currently co-exist in a sort of quasi-diglossia (Brunelle, 2009), but complete deletion of the “presyllable” (i.e., minor-syllable) is still held as the most common path for monosyllabization. Interestingly, there exist anecdotal reports of ongoing monosyllabization through loss of initial syllables in Cham (Alieva, 1994) and Ruc (Solntsev, 1996), an Austroasiatic language of Vietnam, reports which are reminiscent of the earlier framings of Moklenic monosyllabic alternants as “colloquial forms”. However, despite all indications that it could potentially be a more widespread phenomenon, throughout the literature there appears to be no discourse contextualization of alternant word-forms nor actual examples of the clipping of minor syllables within spontaneous speech except for the current study.

The broader diachronic point is that monosyllabic lexemes developing from historically polysyllabic roots has been cast as a “common evolutionary path” for languages throughout the region (Michaud, 2012), and loss of minor syllables is at least one part of what Matisoff, (1990) describes as “cyclic swings of expansion and

³⁷ For Moklen, it is very interesting to note that Southern Thai, the dialect of most contemporary influence on Moklen, has many monosyllabic variants of Standard Thai forms; forms that also have clipped initial syllables (e.g., *kʰà:nŭn* vs. *nŭn* ‘jackfruit’, *sà?pʰā:n* vs. *pʰā:n* ‘bridge’, *cà:mù:k* vs. *mù:k* ‘nose’ etc.).

contraction” in word formation for these languages. In this cycle, there are basically two possibilities for a diachronic change of a polysyllabic word into a monosyllable: either minor syllables merge into a new onset of the major syllable (e.g., Thach, 1999), or the minor syllables are completely omitted. Nevertheless, within the literature, such phenomena are most often only discussed in terms of long-term systemic phonological changes. Ultimately, a realistic model of language change must have monosyllabization in its initial stages occurring in individual speech acts (Croft, 2000). Minor-syllable elision, therefore, as an information structure process offers some cognitive motivations for how reduced variants within discourse could be generated within a population of utterances. Documentation of minor-syllable elision therefore captures how one path for word-form changes may initially start when phonologically weak minor syllables quickly give way through discourse-based deaccentuation and as a result generate a monosyllabic alternant.

Given that minor-syllable elision is potentially widely attested throughout MSEA, it would be interesting to see more discourse-based investigation into information structure effects on word-form within spoken corpora for languages of the region. For example, it is not surprising that within field data on *Moken*, Moklen’s sister language, there are instances of the same sort of word-form shifts discussed in this study, such as in (8.1).

(8.1) [Jenks, 2007]

(8.1.1) *ma4nut4 ni:4 mələn manək*

people this hit chicken

‘The person hits the chicken.’

(8.1.2) *məŋap manək ... ŋap manək bə soba:j*

grab chicken grab chicken make rice.side.dish

‘(He) grabs the chicken, grabs it to make food.’

Although, minor-syllable elision is initially reductive, diachronically it can also be cast as a creative force—first in the generation of a lexeme’s alternant word-form, but also in that these can eventually be integrated into collocations and compounds (§7.4). Interestingly, if we recall some background on the Moklenic people (see §2.2), there is one interesting anecdote related to word-form. White’s

(1922) *The Sea-Gypsies of Malaya*, was the first prolific and insistent use of “Mawken” in the literature. Regarding the term, White also put forward an apparent etymological root, sharing that disyllables *l'maw* means ‘to drown’ and *o'ken* ‘salt water’ coalesced into an endonym meaning ‘the sea-drowned’. It is entirely likely that this is merely a folk linguistic account from the researcher or even Moklenic peoples themselves. However, as to whether or not this is a linguistic possibility, given a process of minor-syllable elision, the answer appears to be “yes”.

8.5 Limitations

There are several limitations of the study to address. One issue is the broader sociolinguistic context. First, there is language vitality. The ages of the 24 Moklen speakers whose texts were selected for analysis ranged from 43 to 77, with an approximate average of 60. Throughout periods of fieldwork from 2019 to 2022 it was consistently found that across all communities it was only speakers in this age range who possessed the type of Moklen language fluency that could allow them to participate in the language elicitation tasks. This is of course indicative of Moklen’s endangered status, but it should also be considered as coming with additional constraints for selecting task participants (e.g., sufficient visual acuity).

An additional factor looming in the background is the role of the Thai language. All Moklen speakers speak Thai and/or a Southern dialect. Early on in other staged communicative events with premade materials, use of Thai lexical material was high, as these materials often portrayed contexts associated with the majority culture. This ended up leading to the creation of the *Stolen Fish* picture book and the Transitive Event Picture Sequences. However, as evident throughout examples in this study, modern Moklen speech is filled with Thai loans and constructions. Here, it should also be mentioned that the researcher, and therefore the primary “hearer” for all sessions, is not a native Moklen speaker, but prior to any of the session that made it into analysis the researcher put in great effort in preparing and practicing running the sessions in Moklen. Still, the imbalance in language ability between the interlocutors should be seen as another layer in which data here is unrepresentative of a more natural Moklen speech context.

One fundamental limitation of the study concerns the use of picture-based stimuli.³⁸ For example, the TEPS stimulus has obvious limitations for studying argument structure. The primary aim of the design was investigation of word-form changes, and therefore there was no control for other relevant factors such as animacy, humanness, and prototypical transitivity (van Nice & Dietrich, 2003). Another major limitation was the lack of ideal experimental controls and session conditions during implementation. Perhaps even more relevant for this study, though, is that use of a picture-based stimuli poses several puzzles, especially when deeper psychological principles are directly implicated within the theoretical framework of information structure. The issue being if one were to adopt the view that everything within a graphical stimulus is already in some sense “situationally evoked”, then categories of given and new in this genre would point to different mental states than what you get with an unprompted narrative genre of speech (or any other genre for that matter). There are critiques of picture-stimuli that mention other limitations (Klamer & Moro, 2020), but given the inter-related nature of information structure phenomena, inclusion of environmental entities such as a stimulus seemingly add another level of complexity. The appropriate response here seems to be to just acknowledge the centrality of the external object as a component of this discourse genre. Despite all these concerns, both stimuli ultimately served as a practical means for uncovering Moklen speakers’ management of linguistic information. A next step in a study of Moklen would be a broader corpus composed of observed communicative events (Himmelman, 1998).

There are also several theoretical limitations. For a study of Moklen, adoption of any model of information structure presents challenges as most theoretical groundwork is built upon highly studied and typologically different languages. Masia’s (2022) framing of topic/focus units as discourse strategies of broad evidentiality offered a useful perspective, but use of “topic” in this study is still mostly in line with commonly accepted conceptions (Krifka & Musan, 2012; Matić, 2015). Annotation of information status using the RefLex Scheme, while a practical operationalization of some information structure theory, also presented challenges.

³⁸ One interesting and related area of research outside of our discussion here concerns the “grammar” of visual language in comics (Cohn, 2018).

Although the scheme is presented as data-oriented, applying its annotation guidelines to Moklen required conceptual work. A more robust application of the scheme is needed to develop information status tagging for languages of this typological profile. More fundamentally, it must also be acknowledge that this study's attempt to align information structure notions (e.g., information status and topic/focus) with specific types of linguistic form falls short in overcoming many of the core challenges and critiques of information structure research (Ozerov, 2018; Matić, 2022). These issues, however, plague almost any conception of discourse as a process of information management and therefore must be left for discussion elsewhere. For further theoretical development, information structure research on spoken discourse in more easily available and typologically similar languages (e.g., Thai) is a research area brimming with potential.

8.6 Conclusion

This study of information structure presents aspects of Moklen morpho-syntax within discourse. Data came from collaboration with Moklen speakers in 32 semi-spontaneous language elicitation sessions (16 Stolen Fish texts and 16 Transitive Event Picture Sequences trials). This data represents a curated selection from sessions that the researcher thought were most representative of Moklen speech across Moklen communities in Phang Nga and Phuket provinces. Analysis in terms of information structure allowed for discussion of several aspects of Moklen grammar and discourse. Information flow was shown to occur mostly through short intonation units that abide by the one-new-idea constraint. Grammatically, clausal intonation units have an AVO argument structure in which arguments follow a given-before-new order. However, ellipsis and detachment-constructions are also prevalent informational strategies. One interesting finding was a pattern of word-form shifts during speech and its relation to information structure. Clarifying the appearance of Moklen's monosyllabic alternants, this study captured and described this interesting tendency. The study's underlying objective was to provide a description of information flow in Moklen discourse. In the end, any claim for why Moklen linguistic forms are the way they are must be taken lightly. As for that descriptive aim, my hope is that this is one of many useful and lasting pictures of Moklen.

Appendix A Stolen Fish Text

STOLEN FISH (NG)

[SF1]

- (1) *caná:t ... nɔ́báj cʰəpʰlɿw*
 child wear pants
 ‘A child, (he’s) wearing pants.’
- (2) *nəlɔ́h bajxj dín ... nəmán tá: saʔ2pʰan1*
 remove shirt come fish.with.rod on bridge
 ‘(He’s) removed (his) shirt and come to fish on a bridge.’

[SF2]

- (3) *nəmán káw*
 fish.with.rod go
 ‘(He’s) fishing.’
- (4) *pʰɔ:1 pacʰáʔ dín ... ʔəká:n ʔót*
 when encounter come fish depleted
 ‘When (there) he finds out, there’s no fish.’
- (5) *ná:ʔ tɛ̀: matá:ʔ cʰuwíə̀k pla:w2*
 have only fishhook empty
 ‘(He’s) only got an empty fishhook.’

[SF3]

- (5) *nɿj kɔ:3 ... bá:p pʰɔ:3tʰaw3 já:j ... káw nɔ́léŋ pukát*
 DEM.PROX CONN elder.male old.person think go descend gillnet
 ‘Here, well...an old man thinking, (I’ll) go gillnetting.’
- (6) *nɔ́léŋ ... ʔu:an1 pukát ...káw tak2 ʔəká:n*
 descend gillnet gillnet go scoop fish
 ‘(I’ll) go gillnetting. Go scoop up (some) fish.’

- (7) *tit2 ʔeká:n bɔ́j talɔ́j pʰɔ́h*
 stick fish acquire three CLF
 ‘(He’s) caught three fish.’

[SF4]

- (8) *tʰi:3 ni:3 plút ʔeká:n*
 at this take.off fish
 ‘Here, (he) takes the fish out.’
- (9) *bá:ʔ dín dúk dagà: ... tʰaŋ4 talɔ́j pʰɔ́h*
 carry come put basket all three CLF
 ‘(He) brings (them) and puts (them) in a basket, all three of them’
- (10) *caná:t ... namé:n nɔ́j batáŋ niʔú:n neŋó:ʔ*
 child hide DEM.PROX tree.trunk coconut look
 ‘The child, (he’s) hiding here at the trunk of a coconut (tree), watching.’
- (11) *já:j ʔebá:p pʰɔ:3tʰaw3 nɔ́j bɔ́j ʔeká:n*
 think elder.male old.person DEM.PROX acquire fish
 ‘(He’s) thinking this old man has got fish.’

[SF5]

- (12) *lɛʔ4 ʔebá:p pʰɔ:3tʰaw3 nɔ́j bɔ́j ʔeká:n talɔ́j pʰɔ́h*
 and elder.male old.person DEM.PROX acquire fish three CLF
 ‘So the old man has gotten three fish.’
- (13) *dúk lám lagà:*
 put in basket
 ‘Puts them in the basket.’

[SF6]

- (14) *lɛ4kɔ:3 nɔ́lɛŋ*
 CONN descend
 ‘And then (he) goes down.’

- (15) *bá:p p^hɔ:3t^haw3 nɔléŋ káw ...tàk ʔeká:n lɜ:j1*
 elder.male old.person descend go scoop fish INTS
 ‘The old man’s going down, to scoop up fish!’
- (16) *bá:ʔ pukát káw tàk ʔeká:n*
 carry gillnet go scoop fish
 ‘(He’s) bringing the gillnet to go scoop up fish.’
- (17) *caná:t ... namé:n héʔ batáŋ niʔú:n di:aw1*
 child hide at tree.trunk coconut alone
 ‘The child’s hiding at the trunk of a coconut tree, alone.’
- (18) *pəná:ʔ ʔeká:n lám dagà: k^hu:1*
 see fish in basket COP
 ‘(He) sees the fish in the basket and so’
- (19) *dín=ɲá: mén ... bá:ʔ dəbút*
 come=3.SG take carry run
 ‘He comes and takes (them) and runs off with (them).’

[SF7]

- (20) *lɛ4kɔ:3 ... ʔenɔŋ k^hu:3 ʔaná:t ... dɔ:k héʔ batáŋ kaʔé:w dɔ:k*
 CONN mother with child sit at tree.trunk wood sit
 ‘And, a mother with her child, are sitting by a tree.’
- (21) *já:j dɔ:k klá:w ʔabó:ʔ dɔ:k*
 say sit speak together sit
 ‘Sitting and talking together.’
- (22) *ɲám dɔ:k nin1t^ha:1 ʔabó:ʔ*
 eat sit gossip together
 ‘Eating, sitting, chatting together.’

[SF8]

- (23) *t^hi:3 nɲj ʔaná:t dalá:ʔ ... pəná:ʔ*
 at DEM.PROX child young.female see
 ‘Here, the young girl, (she) sees.’

- (24) *pəná:ʔ caná:t ... bá:ʔ ʔeká:n dəbút*
 see child carry fish run
 ‘(She) sees the child carrying the fish running.’
- (25) *p^ha:n2 ... p^ha:n2*
 pass pass
 ‘(He’s) passing by.’
- (26) *t^hi:3 nəj pəná:ʔ ʔeká:n kəp^hlà:t ʔa=bulət*
 at DEM.PROX see fish fall one=CLF
 ‘Right here, (she) sees one fish has fallen.’
- (27) *dujú:k já:j ʔeká:n kəp^hlà:t*
 point say fish fall
 ‘(She) points saying the fish has fallen.’
- (28) *ná:t nəj kə:3 ... ná:t dán*
 child DEM.PROX CONN child know
 ‘The kid, well...the kid knows...’
- (29) *ná:t dán háh ... bá:ʔ ʔeká:n dəbút p^hə*
 child know NEG carry fish run PRT
 ‘The kid doesn’t know. (He’s) carrying the fish running away!’
- [SF9]**
- (30) *tok2loŋ2*
 in.the.end
 ‘Ultimately,
- (31) *ʔenón ʔ ... mén ʔeká:n t^hi:3 kəp^hlà:t bá:ʔ dín ʔón*
 mother take fish REL fall carry come give
 ‘The mother takes the fish that fell and carries it over and hands (it) to (him).’
- (32) *p^hu:a3 caná:t*
 for child
 ‘For the sake of the child.’
- (33) *caná:t nəj dín ... dín mén baléh*
 child DEM.PROX come come take return
 ‘The child comes. Comes and takes (it) back.’

[SF10]

- (34) *lɛ4kɔ:3 ... datá: jalá:n nɔ̀j ... pʰú:ŋ caná:t*
 CONN on path DEM.PROX group child
 ‘And so, on this path, a group of children.’
- (35) *waj1run4 dín talɔ̀j luj*
 teenager come three CLF
 ‘Teenagers are coming, three of them.’
- (36) *dín pʰə nin1tʰa:l bó:ʔ pʰə klá:w bó:ʔ pʰə*
 come PRT gossip together PRT speak together PRT
 ‘(They’re) coming, chatting away together, talking away together.’
- (37) *mɛɲák ta:m1 jalá:n*
 walk follow path
 ‘(They’re) following the path.’

[SF11]

- (38) *pʰɔ̌1di:l ʔaná:t ... nəlát ʔeká:n nɔ̀j ta:m1*
 at.same.time child steal fish DEM.PROX follow
 ‘At the same time, the child, (who) stole the fish, is going.’
- (39) *dɔ:k nəmán saʔ2pʰan1 nəlát ʔeká:n*
 sit fish.with.rod bridge steal fish
 ‘to sit and fish at the bridge, having stolen the fish.’
- (40) *bá:ʔ ʔeká:n dəbút ta:m1 jalá:n mu:an5 pla:w2*
 carry fish run follow path similar empty
 ‘(He’s) carrying the fish following along the same path.’

[SF12]

- (41) *pʰɔ̌:1di:l kɔ:3 cʰəpʰáʔ ... mɛɲák dəbút saʔ1dut1 batɔ̀j*
 at.same.time CONN encounter walk run trip rock
 ‘Just then, well, (he) encounters, (he) walks, runs, and trips on a rock.’

- (42) *nəmɔ́h ʔeká:n lùt2 ɲán*
 fall.over fish slip.out hand
 ‘(He) falls over and the fish slip out of his hand.’
- (43) *kʰu:1 pʰú:ɲ waj1run4 ...já:j*
 COP group teenager think
 ‘So, the group of teenagers think’
- (44) *ow bʲj ʔeká:n láɲ tám ʔabɔ́:ʔ*
 VOC acquire fish from where together
 ‘Woah! where did (he) get these fish from?’
- (45) *tʰáɲ talʲj pʰɔ́h*
 all three CLF
 ‘All three of them.’

[SF13]

- (46) *pʰɔ́:1 metá:ʔ cʰá:ʔ làt ... wà:ʔ làt bʲj ká:ʔ*
 when request one CLF two CLF acquire PRT
 ‘When (they) request one, “Can (we) get two?”.’
- (47) *tʰi:3 nʲj ná:t ... nəmán ... já:j ... bʲj bʲj*
 at DEM.PROX child fish.with.rod say acquire acquire
 ‘Then, the kid who fishes says, “Sure, sure.”.’

[SF14]

- (48) *ʔɔ́n=ɲá: wà:ʔ làt*
 give=3.SG two CLF
 ‘(He) gives (them) two of them.’
- (49) *ʔɔ́n ʔeká:n wà:ʔ làt ... waj1run4 nʲj*
 give fish two CLF teenager DEM.PROX
 ‘(He) gives (them) two of the fish, these teenagers.’
- (50) *lɛ4kɔ́:3 ticúm*
 CONN bird
 ‘And, a bird.’

- (51) *ticúm ... nám ?eká:n*
bird eat fish
'The bird eats fish.'
- (52) *ticúm ... na:ηInu:an1 nε:3 na:3*
bird gull PRT PRT
'A seagull, it is.'
- (53) *cúm na:ηInu:an1*
bird gull
'A seagull.'
- (54) *ticúm nỳj nám ?eká:n nε:3 nε:3*
bird DEM.PROX eat fish PRT PRT
'This bird eats fish, for sure.'
- (55) *neηó:ʔ=já: ?eká:n*
look=3.SG fish
'It's looking at the fish.'
- (56) *pʰə:1 kə:3 jà:ʔ ...mén nám mot2 lε:w4 ?eká:n*
when CONN dispose take eat depleted PRF fish
'When (they) toss (it), (it) will eat the fish all gone.'
- [SF15]**
- (57) *tʰi:3 nỳj waj1run4 nú: bỳj ?eká:n lán ná:t*
at DEM.PROX teenager DEM.MED acquire fish from child
'Here, those teenagers got fish from the kid.'
- (58) *nəmán ?eká:n di:awl ni:3*
fish.with.rod fish alone this
'the one that fishes alone.'
- (59) *mén ?eká:n ʔón ticúm ʔa=bulàt*
take fish give bird one=CLF
'(They) take the fish and give the bird one.'

- (60) *tʰi:3 nɿj*
 at DEM.PROX
 ‘Right, here’
- (61) *pɔʔ təŋá:k héʔ nɿj ná:ʔ ʔeká:n ʔa=bulàt*
 3SG middle at DEM.PROX have fish one=CLF
 ‘Him in the middle, (he) has one fish.’
- (62) *bɿj wà:ʔ làt*
 acquire two CLF
 ‘(They) got two’
- (63) *ʔón ticúm ʔa=bulàt*
 give bird one=CLF
 ‘(They) gave the bird one.’
- [SF16]**
- (64) *lɛʔ4 kʰu:1 ... wajlrun4 dín cʰəpʰáʔ pʰú:ŋ*
 and COP teenager come encounter group
 ‘And so, the teenagers come and meet the group.’
- (65) *ʔenɔŋ ná:t dalá:ʔ*
 mother child young.female
 ‘The mother and the young girl.’
- (66) *tʰi:3 nɿj mén ʔeká:n bá:ʔ ʔón*
 at DEM.PROX take fish carry give
 ‘Right, here (they) bring the fish and give (it to them).’
- (67) *lɛʔ4 kʰu:1 ... ʔɿj dəbút ... pʰa:n2 dín ... láŋ ləkót kón kaʔé:w*
 and COP dog run pass come from behind tree wood
 ‘And, a dog is running, passing by from behind the tree.’

[SF17]

- (68) *lɛʔ4 wa:3 ... pʰú:ŋ wajlrun4 bujáj nɿj dɔ:k*
 and COM group teenager young.male DEM.PROX sit
 ‘And so, the group of young men sit down.’

- (69) *klá:w ʔabó:ʔ klá:w pajl klá:w ma:l kʰu:l*
 speak together speak go speak come COP
 ‘(They’re) talking together, speaking back and forth, and so
- (70) *tok2loŋ2 ná: mɛ́n ʔeká:n bá:ʔ məʔíəŋ nám*
 in.the.end 3SG take fish carry roast eat
 ‘in the end, he takes the fish to roast for (them) to eat.’
- (71) *bá:ʔ mók nám ʔa=bulàt*
 carry cook eat one=CLF
 ‘(He) brings the single one to cook for (them) to eat.’
- (72) *lɛʔ4 kʰu:l ... bóʔ ʔuʔúj*
 and COP person younger.sibling
 ‘And well, the younger one’
- (73) *ʔúj pʰú:ŋ bujáj ni:3 na:3*
 younger.sibling group young.male this PRT
 ‘This younger one of the group of young men, you see.’
- (74) *ʔék káw tá: pəkón kaʔé:w*
 ascend go on tree wood
 ‘(He) goes up on to the tree.’
- (75) *lɛ4kɔ:3 ʔɔj ... mə:ŋl nəŋó:ʔ naʔék nəj kɔ:3 bóh ʔanóŋ*
 CONN dog look look ascend DEM.PROX CONN do what
 ‘And the dog’s looking at him going up there to see what he’s doing.’”
- (76) *pʰú:ŋ nəj kɔ:3 məʔíəŋ ʔeká:n bó:ʔ pʰə*
 group DEM.PROX CONN roast fish together PRT
 ‘The group, well, (they’re) continuing to roast the fish together.’

[SF18]

- (77) *tʰi:3 nəj ʔaná:t nəmán*
 at DEM.PROX child fish.with.rod
 ‘Right here, the child who fishes

- (78) *ná:t nəlát ʔeká:n di:aw1*
 child steal fish alone
 ‘The kid who’s stolen the fish is alone.’
- (79) *bá:ʔ ʔeká:n ʔa=bulàt dín latá: saʔ2pʰan1*
 carry fish one=CLF come on bridge
 ‘(He’s) brought the one fish and comes onto the bridge.’
- (80) *dín neŋó:ʔ ... kʰan1kəmán póʔ*
 come look fishing.rod 3SG
 ‘(He’s) looking at his fishing rod.’

[SF19]

- (81) *ʔebá:p pʰɔ:3tʰaw3 tʰi:4 wa:4 nələŋ pukát ... báj ʔeká:n*
 elder.male old.person REL COM descend gillnet acquire fish
 ‘The old man who was gillnetting, and got the fish.’
- (82) *naʔék dín neŋó:ʔ ... ná:ʔ lu:a5 dagà: pla:w2*
 ascend come look exist remaining basket empty
 ‘(He) comes up and looks. All that’s remaining is an empty basket.’
- (83) *ʔeká:n ʔót ... caná:t nəlát*
 fish depleted child steal
 ‘The fish are gone. The child stole (them).’
- (84) *dán já:j ʔacáw lát háh ... ná: ná:ʔ dán háh*
 to.know COM who steal NEG 3SG yet know NEG
 ‘(He) doesn’t know who stole (them). He still doesn’t know.’

[SF20]

- (85) *pʰɔ:1di:1 ... ticúm ... nám ʔeká:n*
 at.same.time bird eat fish
 ‘At the same time, a bird, (it’s) eating fish.’
- (86) *kʰa:p4 ʔeká:n bá:ʔ dín*
 hold.in.mouth fish carry come
 ‘(It’s) holding a fish carrying it.’

- (87) *p^hɔ:1 nan3 ... p^hú:ŋ wajlrun4 bujájŋ ʔón ʔa=bulàt*
 when then group teenager young.male give one=CLF
 ‘For that time, the group of young men gave (it) one of them.’
- (88) *t^hi:3 nỳj ... ticúm jú: bá:ʔ dín*
 at DEM.PROX bird DEM.MED carry come
 ‘Right here, that bird is bringing (it) over.’
- (89) *dín t^hi:3 nỳj bá:p p^hɔ:3t^haw3 páná:ʔ*
 come at DEM.PROX elder.male old.person see
 ‘(It) comes right here the old man sees.’
- (90) *já:j ticúm nỳj nəlát ʔeká:n pɔʔ*
 think bird DEM.PROX steal fish 3SG
 ‘(He) thinks this bird stole his fish.’
- (91) *nəlát ʔeká:n bá:p p^hɔ:3t^haw3 nɛ:4 nɛ:4*
 to.steal fish elder.male old.person PRT PRT
 ‘(It’s) stolen the fish of the old man for sure!’
- (92) *ʔeká:n pɔʔ talíj p^hóh*
 fish 3SG three CLF
 ‘Of his three fish.’
- (93) *k^ha:p4 bá:ʔ dín ʔa=bulàt já:*
 hold.in.mouth carry come one=CLF only
 ‘(It’s) only come carrying one (in its mouth).’

[SF21]

- (94) *t^hi:3 nỳj bá:p p^hɔ:3t^haw3*
 at DEM.PROX elder.male old.person
 ‘Here, the old man
- (95) *ŋé:n nəŋé:n mén mén ʔeká:n baléh he*
 chase chase IRR take fish return VOC
 ‘(He’s) chasing and chasing. (He) wants to take the fish back, hey!’
- (96) *ticúm k^hu:1 nəpá:t baléh ... nəpá:t dəbút*
 bird COP fly return fly run
 ‘The bird, however, is flying away. It quickly flies away.’

[SF22]

- (97) *lɛ4kɔ:3 nəŋɛ:n dín*
 CONN chase come
 ‘And so, (he) chases (it).’
- (98) *ticúm nɔ̀j dín*
 bird DEM.PROX come
 ‘The birds comes’
- (99) *patʰáʔ saʔ2pʰan1 tʰi:3 na:ʔ ʔaná:t ... dɔ:k nəmán*
 encounter bridge at exist child sit fish.with.rod
 ‘(They) come upon the bridge that has the child, sitting fishing.’
- (100) *tʰi:3 nɔ̀j pəná:ʔ ... ná:t nəmán ʔeká:n ʔa=bulàt*
 at DEM.PROX see child fish.with.rod fish one=CLF
 ‘Right here, (he) sees, the kid’s fished a single fish.’

[SF23]

- (101) *ciŋ1ciŋ1 lɛ:w4 ʔaná:t nɔ̀j ... nəlát ʔeká:n bá:p pʰɔ:3tʰaw3*
 really PRF child DEM.PROX steal fish elder.male old.person
 ‘Really though, this child stole the old man’s fish.’
- (102) *bá:ʔ dín ... bóh lɔ:k2*
 carry come do to.trick
 ‘(He’s) brought it and does a trick.’
- (103) *dín ki:aw2 cʰuwíæk pɔ̀ʔ*
 come hook fishhook 3SG
 ‘(He’s) hooks it onto his fishing hook.’
- (104) *dín ki:aw2 kʰan1kəmán pɔ̀ʔ ... datá: saʔ2pʰan1*
 come hook fishing.rod 3SG on bridge
 ‘(He’s) hooking it onto his fishing rod, on the bridge.’

[SF24]

- (105) *já:j pɔ̀ʔ nuwíæk bɔ̀j*
 say 3SG to.fish acquire
 ‘(He) says he caught (it).’

- (106) *bá:p p^hɔ:3t^haw3 nɛ̃j k^hu:1*
 elder.male old.person DEM.PROX COP
 ‘The old man, well (he)...’
- (107) *pó? ... c^hənú:n mu:an5 pla:w2 já:j*
 3SG wonder similar or.not COM
 ‘He, wonders if it’s the same or not.’
- (108) *c^hənú:n já:j ... nuwíək c^həŋbán ʔeká:n kəc^hú:n háh ká?*
 wonder COM to.fish how fish move NEG PRF
 ‘(He) wonders, “How’d could he have fished it? The fish isn’t even moving.”.’

[SF25]

- (109) *lɛʔ4 k^hu:1 ... ticúm jú: k^hu:1 jám ʔeká:n jú:*
 and COP bird DEM.MED COP eat fish DEM.MED
 ‘And so, that bird, so it’s eaten that fish.’
- (110) *t^hi:3 ni:3 ... caná:t nɛ̃j já:j pó? nuwíək b́j ʔeká:n*
 at this child DEM.PROX say 3SG to.fish acquire fish
 ‘Right here, the child says that he caught the fish.’
- (111) *bá:p p^hɔ:3t^haw3 já:j pó?*
 elder.male old.person think 3SG
 ‘The old man thinks, he...’
- (112) *c^hənú:n já:j ʔeká:n pó? káw tám kət́j*
 wonder COM fish 3SG go where missing
 ‘(He) wonders, “Where have my fish gone, (they’re) missing!?”.’
- (113) *taĺj p^hɔ̃h*
 three CLF
 ‘The three of them’
- (114) *kɔ:3 meŋák baléh*
 CONN walk return
 ‘So, (he) walks back.’

[SF26]

- (115) *meɲák dín patʰáʔ ... dɔ:k ʔɔlá:ŋ cʰumInu:mI ʔabó:ʔ dɔ:k*
 walk come encounter sit person gather together sit
 ‘(He) walks and comes across a group of people sitting gathered together.’
- (116) *ninIʰa:1 dɔ:k ...ninIʰa:1 dɔ:k klá:w ʔabó:ʔ*
 gossip sit gossip sit speak together
 ‘(They’re) chatting, sitting. Chatting, sitting, speaking together.’
- (117) *pəná:ʔ ʔeká:n ʔa=bulàt*
 see fish one=CLF
 ‘(He) sees one fish.’
- (118) *ʔɔj kɔ:3 ná:ʔ héʔ nɔ̀j*
 dog CONN exist at DEM.PROX
 ‘A dog is right there.’
- (119) *ná:ʔ kanáj bujáj ... kʰu:1 ná:ʔ talɔ̀j lúj*
 exist male young.male COP exist three CLF
 ‘There are young men, that is there’s three of them.’
- (120) *kʰu:1 pɔʔ dín ʰa:m5 pʰú:ŋ nɔ̀j*
 COP 3SG come to.ask group DEM.PROX
 ‘So, he comes and asks this group.’

[SF27]

- (121) *jà:j pɔʔ káw nɔ̀léŋ ʔu:anI héʔ ʔé:n nɔ̀j ... pukát héʔ dé:*
 say 3SG go descend gillnet at water here gillnet at DEM.DIS
 ‘(He) says he went gillnetting at the water here, (he) was gillnetting there.’
- (122) *ʔeká:n talɔ̀j pʰɔ̀h cʰu:n5 kə̀tɔ̀j já:j*
 fish three CLF missing missing say
 ‘His three fish have gone missing, (he) says.’
- (123) *tʰi:3 nɔ̀j pɔʔ káw táw*
 at DEM.PROX 3P go where
 ‘Here, “Where’d they go?”’

- (124) *ʔaná:t jú: ... ná:t təmán héʔ saʔ2pʰan1 já:j*
 child DEM.MED child fishhook at bridge say
 ‘That child, the kid fishing at the bridge, (he) says.’
- (125) *póʔ nəmán bʰj ʔa=bulàt*
 3SG fish.with.rod acquire one=CLF
 ‘He caught a single one.’
- (126) *tè: ʔeká:n ... kəcʰú:n háh káʔ*
 but fish move NEG PRF
 ‘But, the fish, (it) wasn’t moving.’
- (127) *ká:n matáj káʔ*
 fish die PRF
 ‘The fish was dead.’
- (128) *tok2loŋ2 ... pəná:ʔ ʔeká:n ʔa=bulàt héʔ nʰj dɔ:k=lá:ŋ*
 in.the.end see fish one=CLF at DEM.PROX sit=3P
 ‘Ultimately, (he) saw the one fish here and sat together with them.’
- (129) *məʔiəŋ nám ʔabó:ʔ*
 roast eat together
 ‘(They) roast (it) and eat together.’
- (130) *bá:p pʰɔ:3tʰaw3 ... já:j cʰənú:n nekót ʔj nʰj*
 elder.male old.person say suspect behind dog DEM.PROX
 ‘The old man, says (he) suspects the dog behind (him)’
- (131) *mén nəlap nám*
 want snatch eat
 ‘(It) wants to snatch and eat (it)’

[SF28]

- (132) *pʰɔ:1di:1 ʔaná:t ... tʰi:3 nəlát ʔeká:n*
 at.same.time child REL steal fish
 ‘At that time the child, who stole the fish.’
- (133) *di:aw1 kɔ:3 jip2 plút ʔeká:n ... láŋ kʰan1kəmán pɔʔ*
 shortly CONN grab take.off fish from fishing.rod 3SG
 ‘Well (he’s) about to grab and take the fish off from his fishing rod.’

- (134) *bá:ʔ dúk dagà:*
 carry put basket
 ‘Bring (it) and put it in a basket.’

[SF29]

- (135) *bá:ʔ dúk dagà:*
 carry put basket
 ‘(He) brings (it) and puts (it) in the basket.’

- (136) *mén bá:ʔ baléh hə be*
 take carry return VOC PRT
 ‘(He) has brought (it) back!’

- (137) *cīŋ ʔeká:n matáj kaʔ*
 really fish to.die PRF
 ‘Really though, the fish is dead.’

[SF30]

- (138) *tʰi:3 nỳj ticúm ... pəná:ʔ ʔeká:n naŋá:t dín nələŋ*
 at DEM.PROX bird see fish fly come descend
 ‘Right here, the bird sees the fish and flies on down.’

- (139) *dín=ná: ... mén dín nəlát ʔeká:n tɔ:2 lá:j*
 come=3.SG IRR come steal fish next again
 ‘It’s come to continue stealing fish!’

[SF31]

- (140) *lɛʔ4 kʰu:1 ... nỳj kɔ:3 ba:3 nəmán pʰə*
 and COP DEM.PROX CONN obsessed fish.with.rod PRT
 ‘The thing is, here, well (he’s) focused only on fishing.’

- (141) *nəmán káw nəmán kʰu:1 ʔeká:n ... ɲám háh cʰuwiək*
 fish.with.rod go fish.with.rod COP fish eat NEG fishhook
 ‘(He’s) going fishing. The thing is the fish aren’t biting the fishhook.’

- (142) *ticúm nỳj kʰu:l ... neŋɔːʔ já:j kʰu:l*
 bird DEM.PROX COP look think COP
 ‘The bird, so, (it) looks and thinks
- (143) *neŋɔːʔ=já: ... loŋl loŋl lekɔ́t kaʔ ʔi*
 look=3.SG descend descend behind PRF PRT
 ‘It’s looking and has descended right down behind (him)!’
- (144) *lát ʔeká:n bá:ʔ káw lx:jl*
 steal fish carry go INTS
 ‘(It’s) stolen the fish and taken it away!’

[SF32]

- (145) *bá:ʔ napá:t leʔ4 bé:t káw lx:jl*
 carry fly and exit go INTS
 ‘(It’s) flown off and gone away!’
- (146) *nỳj caná:t dán lu:an3 háh*
 DEM.PROX child know story NEG
 ‘Here, the child doesn’t know what’s going on.’
- (147) *ná:t ba:3 nəmán ʔeká:n lx:jl*
 child obsessed fish.with.rod fish INTS
 ‘The kid’s focused only on fishing!’
- (148) *dán háh já:j ʔeká:n lám dagaː ʔót kaʔ*
 know NEG COM fish in basket depleted PRF
 ‘(He) doesn’t know that the fish in the basket is gone.’

[SF33]

- (149) *ná:t dán lu:an3 háh*
 child know story NEG
 ‘The kid doesn’t know what’s going on.’
- (150) *ná:t nəmán pʰə neŋɔːʔ ... cʰuwiək pɔʔ*
 child fish.with.rod continually look fishhook 3SG
 ‘The kid is continuing to fish and look at his fishing hook.’

- (151) *tʰi:3 ni:3 ... ŋó:ʔ ʔót kaʔ nɔ̀j*
 at this look depleted PRF DEM.PROX
 ‘Right here, (he) see (it’s) gone.’
- (152) *ʔeká:n dagá: kɔ:3 ʔót*
 fish basket CONN depleted
 ‘The fish in the basket, (it’s) gone.’
- (153) *nəmán kɔ:3 lɔ:j1 ʔeká:n tít háh*
 fish.with.rod CONN INTS fish to.stick NEG
 ‘(He’s) gone fishing, and didn’t catch any fish.’
- (154) *tok2loŋ2 kʰu:1 ... lát bɔ̀j ʔacáw háh*
 in.the.end COP CLF acquire who NEG
 ‘In the end, (he) couldn’t steal from anyone.’
- (155) *cop2*
 finish
 ‘Finished.’

Appendix B TEPS Monosyllabic alternants and word-form shifts

Nominal monosyllabic alternants			
lexeme	gloss	<i>n</i>	Word-form shifts
<i>pəkón</i>	tree	45	12
<i>buwá:k</i>	fruit	41	4
<i>ʔibú:m</i>	elder.female	28	9
<i>ʔebá:p</i>	elder.male	25	3
<i>ʔuʔé:n</i>	water	21	7
<i>ʔaná:t</i>	child.offspring	16	16
<i>caná:t</i>	child	16	9
<i>ʔomá:k</i>	house	14	12
<i>ʔeká:n</i>	fish	12	10
<i>niʔú:n</i>	coconut	11	9
<i>kaʔé:w</i>	wood	9	3
<i>ʔenón</i>	mother	9	7
<i>katá:m</i>	crab	7	4
<i>ʔlá:n</i>	snake	5	5
<i>ʔapón</i>	father	5	2
<i>ticúm</i>	bird	4	3
<i>c^huwíæk</i>	large fishhook	3	3
<i>matá:ʔ</i>	eye, fishhook	3	3
<i>təmán</i>	fishing.rod	3	1
<i>kecók</i>	bottle	2	2
<i>ʔaták</i>	head	2	-
<i>kabá:ŋ</i>	boat	1	1
<i>paʔó:k</i>	mango	1	1
<i>mə4nut4</i>	human	1	-
<i>kapáw</i>	bag	1	-
<i>ləkó:ŋ</i>	neck	1	-
<i>kəp^hló:</i>	can (container)	1	1
<i>ʔuʔúj</i>	younger.sibling	1	-

Verbal monosyllabic alternants			
lexeme	gloss	<i>n</i>	WS
<i>nəp^hlâ:t</i>	fall	34	5
<i>nəpó:k</i>	launder	21	4
<i>məʔám</i>	drink	13	1
<i>bətɔk</i>	strike	12	6
<i>nəbá:k</i>	to ladle	10	-
<i>k^həná:ʔ</i>	undergo	8	1
<i>dəbút</i>	run	8	1
<i>didú:n</i>	sleep	7	6
<i>nɔléŋ</i>	descend	7	-
<i>nuwíək</i>	to fish	7	3
<i>niŋít</i>	slice	5	3
<i>dəbúh</i>	boil	4	-
<i>bəkáh</i>	shatter	3	3
<i>neŋó:ʔ</i>	look	3	1
<i>nəŋé:n</i>	chase	2	2
<i>nəpó:ŋ</i>	chop	2	2
<i>bətáj</i>	be.satiated	2	-
<i>niʔù:n</i>	dry.in.sunlight	2	1
<i>nəmán</i>	to.rod.fish	1	-
<i>paták</i>	break	1	1
<i>matáj</i>	die	1	1
<i>nəmóh</i>	fall	1	1
<i>nuwá:j</i>	to fell	1	-
<i>bəlúəŋ</i>	be.unripe	1	1
<i>mətét</i>	cut	1	1
<i>nəmát</i>	dispose	1	1
<i>məʔíəŋ</i>	roast	1	1
<i>makét</i>	hurt	1	-
<i>mɔc^hóŋ</i>	carry(2 people)	1	-

<i>tábák</i>	pierce	1	1
<i>naʔék</i>	ascend	1	-
<i>sa2dut2</i>	trip	1	1

Closed-class alternants			
lexeme	gloss	<i>n</i>	WS
<i>dalám</i>	in	40	3
<i>bulàt</i>	CLF	15	-
<i>duwà:ʔ</i>	two	15	-
<i>datá:</i>	on	8	1
<i>lemá:ʔ</i>	five	1	-

Information status of TEPS monosyllabic alternants (See Tables above for lexeme gloss)						
Speaker	TEPS ID	Lexeme	Word-form shift	Lexical	Referential	Part of Speech
DW	21.1A	<i>bulàt</i>	0	new	new	CLF
WN	20.1A	<i>bulàt</i>	0	new	new	CLF
CU	6.1A	<i>ʔebá:p</i>	0	new	new	N
CU	24.1P	<i>ʔebá:p</i>	0	new	new	N
DW	9.1A	<i>ʔebá:p</i>	0	new	new	N
LI	11.2P	<i>ʔebá:p</i>	0	new	new	N
LW	9.2A	<i>ʔebá:p</i>	0	new	new	N
LW	24.2A	<i>ʔebá:p</i>	0	new	new	N
NN	23.1A	<i>ʔebá:p</i>	0	new	new	N
PB	23.2A	<i>ʔebá:p</i>	0	new	new	N
PB	11.2P	<i>ʔebá:p</i>	0	new	new	N
PB	6.1A	<i>ʔebá:p</i>	0	new	new	N
TG	9.2P	<i>ʔebá:p</i>	0	new	new	N
TG	24.2A	<i>ʔebá:p</i>	0	new	new	N
TG	23.1P	<i>ʔebá:p</i>	0	new	new	N
YN	23.2A	<i>ʔebá:p</i>	0	new	new	N
CU	17.1P	<i>ʔibú:m</i>	0	new	new	N
CU	7.1A	<i>ʔibú:m</i>	0	new	new	N
LI	17.1P	<i>ʔibú:m</i>	0	new	new	N

LI	7.1A	<i>ʔibú:m</i>	0	new	new	N
LI	10.2P	<i>ʔibú:m</i>	0	new	new	N
LN	17.1P	<i>ʔibú:m</i>	0	new	new	N
LW	17.2A	<i>ʔibú:m</i>	0	new	new	N
LW	10.1A	<i>ʔibú:m</i>	0	new	new	N
NN	8.1A	<i>ʔibú:m</i>	0	new	new	N
PB	17.1P	<i>ʔibú:m</i>	0	new	new	N
PB	8.2P	<i>ʔibú:m</i>	0	new	new	N
TG	10.1A	<i>ʔibú:m</i>	0	new	new	N
PB	10.2P	<i>ʔibú:m</i>	M>D	new	new	N
TW	13.1A	<i>ticúm</i>	M>D	new	new	N
LN	5.1P	<i>ʔeká:n</i>	0	new	new	N
NN	14.1P	<i>ʔeká:n</i>	M>D	new	new	N
LN	9.2A	<i>ləkó:ŋ</i>	0	new	new	N
CU	15.1P	<i>ʔəkón</i>	0	new	new	N
DW	11.1P	<i>ʔəkón</i>	0	new	new	N
EW	15.2A	<i>ʔəkón</i>	0	new	new	N
EW	22.1A	<i>ʔəkón</i>	0	new	new	N
LI	13.1P	<i>ʔəkón</i>	0	new	new	N
LI	15.1P	<i>ʔəkón</i>	0	new	new	N
LN	23.1A	<i>ʔəkón</i>	0	new	new	N
LN	11.1P	<i>ʔəkón</i>	0	new	new	N
LP	23.2P	<i>ʔəkón</i>	0	new	new	N
LP	15.2A	<i>ʔəkón</i>	0	new	new	N
LP	22.1A	<i>ʔəkón</i>	0	new	new	N
LW	22.1A	<i>ʔəkón</i>	0	new	new	N
NN	21.1A	<i>ʔəkón</i>	0	new	new	N
NN	13.2A	<i>ʔəkón</i>	0	new	new	N
TW	15.2A	<i>ʔəkón</i>	0	new	new	N
TW	22.1A	<i>ʔəkón</i>	0	new	new	N
YG	15.1P	<i>ʔəkón</i>	0	new	new	N
YG	13.2P	<i>ʔəkón</i>	0	new	new	N
YG	11.1P	<i>ʔəkón</i>	0	new	new	N
TG	22.1A	<i>ʔəkón</i>	M>D	new	new	N
WN	16.2A	<i>ʔəmá:k</i>	0	new	new	N
LP	24.1A	<i>təmán</i>	0	new	new	N
PB	18.2A	<i>caná:t</i>	0	new	new	N
WN	12.1A	<i>caná:t</i>	0	new	new	N
YN	24.1P	<i>caná:t</i>	0	new	new	N
YN	4.1A	<i>caná:t</i>	0	new	new	N
PB	5.2P	<i>mə4nut4</i>	0	new	new	N

NN	17.3A	<i>kapáw</i>	0	new	new	N
LW	11.1A	<i>ῥαπόη</i>	0	new	new	N
TW	11.1A	<i>ῥαπόη</i>	0	new	new	N
CU	14.1A	<i>κατά:m</i>	0	new	new	N
DW	21.1A	<i>buwá:k</i>	0	new	new	N
LI	21.1A	<i>buwá:k</i>	0	new	new	N
LN	21.1A	<i>buwá:k</i>	0	new	new	N
LP	20.2P	<i>buwá:k</i>	0	new	new	N
LW	2.1P	<i>buwá:k</i>	0	new	new	N
NJ	21.2P	<i>buwá:k</i>	0	new	new	N
NK	21.2P	<i>buwá:k</i>	0	new	new	N
NK	2.1P	<i>buwá:k</i>	0	new	new	N
NN	21.2P	<i>buwá:k</i>	0	new	new	N
NN	2.1P	<i>buwá:k</i>	0	new	new	N
PB	20.1A	<i>buwá:k</i>	0	new	new	N
TW	21.2P	<i>buwá:k</i>	0	new	new	N
WN	21.1A	<i>buwá:k</i>	0	new	new	N
YG	21.1A	<i>buwá:k</i>	0	new	new	N
YN	21.1A	<i>buwá:k</i>	0	new	new	N
TG	15.3A	<i>buwá:k</i>	M>D	new	new	N
NJ	19.1A	<i>ῥυῖέ:n</i>	0	new	new	N
NN	3.2A	<i>ῥυῖέ:n</i>	0	new	new	N
TW	19.1A	<i>ῥυῖέ:n</i>	0	new	new	N
WN	19.1P	<i>ῥυῖέ:n</i>	0	new	new	N
NK	15.2A	<i>καῖέ:w</i>	0	new	new	N
NN	23.2P	<i>καῖέ:w</i>	0	new	new	N
NN	11.2A	<i>καῖέ:w</i>	0	new	new	N
TW	2.1P	<i>niῖú:n</i>	0	new	new	N
NN	11.1A	<i>ῥυῖύj</i>	0	new	new	N
LI	5.3P	<i>bulàt</i>	0	new	given	CLF
LN	21.1A	<i>bulàt</i>	0	new	given	CLF
LP	15.2A	<i>bulàt</i>	0	new	given	CLF
TG	10.3A	<i>bulàt</i>	0	new	given	CLF
TG	6.3P	<i>bulàt</i>	0	new	given	CLF
WN	20.2A	<i>bulàt</i>	0	new	given	CLF
WN	20.2A	<i>bulàt</i>	0	new	given	CLF
CU	4.3A	<i>ῥεβά:p</i>	0	new	given	N
WN	4.3A	<i>ῥιβú:m</i>	0	new	given	N
EW	11.3A	<i>ῥοκόν</i>	0	new	given	N
NN	23.2P	<i>ῥοκόν</i>	0	new	given	N
TG	16.1P	<i>ῥομά:k</i>	M>D	new	given	N

LN	1.3P	<i>ρενόη</i>	0	new	given	N
TW	15.3A	<i>ρενόη</i>	0	new	given	N
LW	1.2A	<i>ρενόη</i>	M>D	new	given	N
LW	4.2P	<i>ρενόη</i>	M>D	new	given	N
TW	1.2A	<i>ρενόη</i>	M>D	new	given	N
TW	15.3A	<i>ραπόη</i>	0	new	given	N
LP	23.3P	<i>ρατάκ</i>	0	new	given	N
NJ	2.2P	<i>ρατάκ</i>	0	new	given	N
EW	2.1P	<i>buwá:k</i>	0	new	given	N
LP	21.2P	<i>buwá:k</i>	0	new	given	N
PB	21.1A	<i>buwá:k</i>	0	new	given	N
NK	3.3A	<i>ρuρé:n</i>	0	new	given	N
NK	2.2P	<i>ρuρé:n</i>	0	new	given	N
PB	2.2A	<i>ρuρé:n</i>	0	new	given	N
TW	3.3A	<i>ρuρé:n</i>	0	new	given	N
WN	2.3A	<i>ρuρé:n</i>	0	new	given	N
NN	22.2A	<i>kaρé:w</i>	0	new	given	N
EW	19.1A	<i>dalám</i>	0	new	0	PREP
EW	6.3P	<i>dalám</i>	0	new	0	PREP
EW	14.1P	<i>dalám</i>	0	new	0	PREP
EW	5.3A	<i>dalám</i>	0	new	0	PREP
LN	5.3P	<i>dalám</i>	0	new	0	PREP
LN	3.1P	<i>dalám</i>	0	new	0	PREP
LP	16.2P	<i>dalám</i>	0	new	0	PREP
LP	6.3P	<i>dalám</i>	0	new	0	PREP
LP	14.1P	<i>dalám</i>	0	new	0	PREP
LP	14.2P	<i>dalám</i>	0	new	0	PREP
LP	14.3P	<i>dalám</i>	0	new	0	PREP
LW	14.1P	<i>dalám</i>	0	new	0	PREP
NJ	19.1A	<i>dalám</i>	0	new	0	PREP
NJ	9.3P	<i>dalám</i>	0	new	0	PREP
NJ	5.3A	<i>dalám</i>	0	new	0	PREP
NK	14.1P	<i>dalám</i>	M>D	new	0	PREP
NK	19.1A	<i>dalám</i>	0	new	0	PREP
NN	5.3A	<i>dalám</i>	0	new	0	PREP
NN	14.2P	<i>dalám</i>	0	new	0	PREP
PB	3.2P	<i>dalám</i>	0	new	0	PREP
PB	6.3A	<i>dalám</i>	0	new	0	PREP
TG	14.1P	<i>dalám</i>	0	new	0	PREP
WN	19.1P	<i>dalám</i>	0	new	0	PREP
YG	5.3P	<i>dalám</i>	0	new	0	PREP

YG	3.2P	<i>dalám</i>	0	new	0	PREP
YG	9.3A	<i>dalám</i>	0	new	0	PREP
YG	19.3P	<i>dalám</i>	0	new	0	PREP
YG	19.3P	<i>dalám</i>	0	new	0	PREP
YG	8.1P	<i>dalám</i>	0	new	0	PREP
YG	6.3A	<i>dalám</i>	0	new	0	PREP
NN	6.2P	<i>lemá:ʔ</i>	0	new	0	NUM
LI	13.1P	<i>datá:</i>	0	new	0	PREP
NN	15.3A	<i>datá:</i>	0	new	0	PREP
NN	22.2A	<i>datá:</i>	0	new	0	PREP
NN	13.2A	<i>datá:</i>	0	new	0	PREP
WN	22.P	<i>datá:</i>	0	new	0	PREP
TW	15.2A	<i>datá:</i>	0	new	0	PREP
DW	21.1A	<i>duwà:ʔ</i>	0	new	0	NUM
DW	20.3A	<i>duwà:ʔ</i>	0	new	0	NUM
EW	11.1A	<i>duwà:ʔ</i>	0	new	0	NUM
LN	21.1A	<i>duwà:ʔ</i>	0	new	0	NUM
LN	12.1A	<i>duwà:ʔ</i>	0	new	0	NUM
NK	1.2A	<i>duwà:ʔ</i>	0	new	0	NUM
NK	12.2P	<i>duwà:ʔ</i>	0	new	0	NUM
NN	20.2P	<i>duwà:ʔ</i>	0	new	0	NUM
TG	20.3P	<i>duwà:ʔ</i>	0	new	0	NUM
TG	11.1A	<i>duwà:ʔ</i>	0	new	0	NUM
TG	10.3A	<i>duwà:ʔ</i>	0	new	0	NUM
WN	11.2P	<i>duwà:ʔ</i>	0	new	0	NUM
WN	11.3P	<i>duwà:ʔ</i>	0	new	0	NUM
DW	3.1P	<i>nábá:k</i>	0	new	0	V
LI	3.2A	<i>nábá:k</i>	0	new	0	V
LN	3.2P	<i>nábá:k</i>	0	new	0	V
LW	3.2A	<i>nábá:k</i>	0	new	0	V
NN	3.2A	<i>nábá:k</i>	0	new	0	V
TG	3.2A	<i>nábá:k</i>	0	new	0	V
WN	3.2P	<i>nábá:k</i>	0	new	0	V
YN	3.2P	<i>nábá:k</i>	0	new	0	V
LI	9.3A	<i>dábúh</i>	0	new	0	V
LW	9.3P	<i>dábúh</i>	0	new	0	V
NN	9.3P	<i>dábúh</i>	0	new	0	V
LI	17.3P	<i>dábút</i>	0	new	0	V
LW	17.3A	<i>dábút</i>	0	new	0	V
NK	19.2A	<i>dábút</i>	0	new	0	V
NN	17.3A	<i>dábút</i>	0	new	0	V

NN	23.2P	<i>dəbút</i>	0	new	0	V
PB	17.3P	<i>dəbút</i>	0	new	0	V
PB	19.2P	<i>dəbút</i>	0	new	0	V
NN	12.2P	<i>məc^hɔ̃ŋ</i>	0	new	0	V
CU	7.3A	<i>didú:n</i>	0	new	0	V
TG	4.3P	<i>makét</i>	0	new	0	V
LN	5.3P	<i>nəléh</i>	0	new	0	V
TG	22.2A	<i>nəléh</i>	0	new	0	V
TW	11.2A	<i>nəléh</i>	0	new	0	V
WN	6.2A	<i>nəléh</i>	0	new	0	V
YG	22.2P	<i>nəléh</i>	0	new	0	V
YN	21.2A	<i>nəléh</i>	0	new	0	V
LN	5.2P	<i>nəmán</i>	0	new	0	V
LN	17.2P	<i>k^həná:ʔ</i>	0	new	0	V
TG	20.2P	<i>k^həná:ʔ</i>	0	new	0	V
TW	18.2P	<i>k^həná:ʔ</i>	0	new	0	V
WN	20.2A	<i>k^həná:ʔ</i>	0	new	0	V
YG	22.2P	<i>k^həná:ʔ</i>	0	new	0	V
YG	20.2A	<i>k^həná:ʔ</i>	0	new	0	V
WN	7.2A	<i>nəhó:ʔ</i>	0	new	0	V
DW	9.2A	<i>nihít</i>	0	new	0	V
NN	9.2P	<i>nihít</i>	0	new	0	V
LI	8.2A	<i>nəpó:k</i>	0	new	0	V
LN	8.2P	<i>nəpó:k</i>	0	new	0	V
LP	8.2A	<i>nəpó:k</i>	0	new	0	V
LW	8.2A	<i>nəpó:k</i>	0	new	0	V
NN	8.2A	<i>nəpó:k</i>	0	new	0	V
PB	8.2P	<i>nəpó:k</i>	0	new	0	V
TG	8.2A	<i>nəpó:k</i>	0	new	0	V
TW	8.2A	<i>nəpó:k</i>	0	new	0	V
YN	8.2P	<i>nəpó:k</i>	0	new	0	V
DW	20.2A	<i>nəp^hlà:t</i>	0	new	0	V
DW	22.2P	<i>nəp^hlà:t</i>	0	new	0	V
LI	21.3A	<i>nəp^hlà:t</i>	0	new	0	V
LI	16.3A	<i>nəp^hlà:t</i>	0	new	0	V
LN	16.3A	<i>nəp^hlà:t</i>	0	new	0	V
LP	22.2A	<i>nəp^hlà:t</i>	0	new	0	V
LW	16.3P	<i>nəp^hlà:t</i>	0	new	0	V
LW	23.2P	<i>nəp^hlà:t</i>	0	new	0	V
LW	23.3P	<i>nəp^hlà:t</i>	0	new	0	V
NJ	4.2P	<i>nəp^hlà:t</i>	0	new	0	V

NN	17.3A	<i>nəp^hlɑ:t</i>	0	new	0	V
NN	16.3P	<i>nəp^hlɑ:t</i>	0	new	0	V
PB	20.3A	<i>nəp^hlɑ:t</i>	0	new	0	V
TW	15.2A	<i>nəp^hlɑ:t</i>	0	new	0	V
TW	22.2A	<i>nəp^hlɑ:t</i>	0	new	0	V
WN	20.2A	<i>nəp^hlɑ:t</i>	0	new	0	V
WN	17.3P	<i>nəp^hlɑ:t</i>	0	new	0	V
WN	15.3P	<i>nəp^hlɑ:t</i>	0	new	0	V
WN	22.2P	<i>nəp^hlɑ:t</i>	0	new	0	V
YG	14.2A	<i>nəp^hlɑ:t</i>	0	new	0	V
YG	20.2A	<i>nəp^hlɑ:t</i>	0	new	0	V
YN	16.3A	<i>nəp^hlɑ:t</i>	0	new	0	V
LW	10.3A	<i>bətán</i>	0	new	0	V
NN	10.2A	<i>bətán</i>	0	new	0	V
LI	17.2P	<i>bətək</i>	0	new	0	V
LW	18.2P	<i>bətək</i>	0	new	0	V
LW	17.2A	<i>bətək</i>	0	new	0	V
NN	11.2A	<i>nuwá:j</i>	0	new	0	V
LI	5.2P	<i>nuwíək</i>	0	new	0	V
LW	5.2A	<i>nuwíək</i>	0	new	0	V
NN	5.2A	<i>nuwíək</i>	0	new	0	V
CU	3.3P	<i>məʔám</i>	0	new	0	V
DW	3.3P	<i>məʔám</i>	0	new	0	V
EW	3.3A	<i>məʔám</i>	0	new	0	V
LI	3.3A	<i>məʔám</i>	0	new	0	V
LN	3.3P	<i>məʔám</i>	0	new	0	V
LN	2.3A	<i>məʔám</i>	0	new	0	V
LP	3.3A	<i>məʔám</i>	0	new	0	V
NN	3.2A	<i>məʔám</i>	0	new	0	V
YG	3.3P	<i>məʔám</i>	0	new	0	V
YN	24.2P	<i>məʔám</i>	0	new	0	V
LI	15.2P	<i>naʔék</i>	0	new	0	V
TW	8.3A	<i>niʔù:n</i>	0	new	0	V
YG	7.3A	<i>didú:n</i>	M>D	new	0	V
EW	2.2P	<i>bəlúəŋ</i>	M>D	new	0	V
WN	4.2A	<i>nəŋé:n</i>	M>D	new	0	V
CU	11.1P	<i>pəkón</i>	M>D	given	new	N
PB	24.1P	<i>caná:t</i>	0	given	new	N
WN	12.1A	<i>caná:t</i>	0	given	new	N
DW	21.2A	<i>bulàt</i>	0	given	given	CLF
DW	5.2P	<i>bulàt</i>	0	given	given	CLF

LP	15.2A	<i>bulàt</i>	0	given	given	CLF
LP	15.2A	<i>bulàt</i>	0	given	given	CLF
LP	15.2A	<i>bulàt</i>	0	given	given	CLF
LP	15.2A	<i>bulàt</i>	0	given	given	CLF
LI	19.3P	<i>kabá:η</i>	D>M	given	given	N
LW	9.3P	<i>ʔebá:p</i>	0	given	given	N
NN	23.2P	<i>ʔebá:p</i>	0	given	given	N
PB	23.2A	<i>ʔebá:p</i>	0	given	given	N
PB	6.2A	<i>ʔebá:p</i>	0	given	given	N
TG	8.1A	<i>ʔebá:p</i>	0	given	given	N
TG	8.1A	<i>ʔebá:p</i>	0	given	given	N
TG	11.1A	<i>ʔebá:p</i>	0	given	given	N
LW	23.3P	<i>ʔebá:p</i>	D>M	given	given	N
PB	9.2A	<i>ʔebá:p</i>	D>M	given	given	N
PB	9.2A	<i>ʔebá:p</i>	D>M	given	given	N
CU	17.2P	<i>ʔibú:m</i>	0	given	given	N
LI	17.3P	<i>ʔibú:m</i>	0	given	given	N
LN	17.2P	<i>ʔibú:m</i>	0	given	given	N
LN	17.3P	<i>ʔibú:m</i>	0	given	given	N
LW	17.3A	<i>ʔibú:m</i>	0	given	given	N
PB	17.2P	<i>ʔibú:m</i>	0	given	given	N
LW	10.1A	<i>ʔibú:m</i>	D>M	given	given	N
LW	10.1A	<i>ʔibú:m</i>	D>M	given	given	N
LW	10.2A	<i>ʔibú:m</i>	D>M	given	given	N
LW	10.3A	<i>ʔibú:m</i>	D>M	given	given	N
LW	5.1A	<i>ʔibú:m</i>	D>M	given	given	N
PB	10.2P	<i>ʔibú:m</i>	D>M	given	given	N
TG	17.3A	<i>ʔibú:m</i>	D>M	given	given	N
TG	16.2P	<i>ʔibú:m</i>	D>M	given	given	N
DW	1.2P	<i>caná:t</i>	D>M	given	given	N
LN	4.3A	<i>caná:t</i>	D>M	given	given	N
LW	4.3P	<i>caná:t</i>	D>M	given	given	N
NJ	4.2P	<i>caná:t</i>	D>M	given	given	N
PB	2.1A	<i>caná:t</i>	D>M	given	given	N
PB	2.2A	<i>caná:t</i>	D>M	given	given	N
PB	2.3A	<i>caná:t</i>	D>M	given	given	N
TW	4.3P	<i>caná:t</i>	D>M	given	given	N
WN	4.2A	<i>caná:t</i>	D>M	given	given	N
CU	20.3A	<i>kecók</i>	D>M	given	given	N
YG	20.3A	<i>kecók</i>	D>M	given	given	N
LI	16.1A	<i>ticúm</i>	0	given	given	N

LN	16.3A	<i>ticúm</i>	D>M	given	given	N
TG	13.2A	<i>ticúm</i>	D>M	given	given	N
LN	5.1P	<i>ʔeká:n</i>	0	given	given	N
LP	14.2P	<i>ʔeká:n</i>	D>M	given	given	N
LW	14.2P	<i>ʔeká:n</i>	D>M	given	given	N
NJ	6.2P	<i>ʔeká:n</i>	D>M	given	given	N
NN	5.3A	<i>ʔeká:n</i>	D>M	given	given	N
PB	14.3A	<i>ʔeká:n</i>	D>M	given	given	N
TG	5.2A	<i>ʔeká:n</i>	D>M	given	given	N
TG	6.1P	<i>ʔeká:n</i>	D>M	given	given	N
YG	14.2A	<i>ʔeká:n</i>	D>M	given	given	N
YN	5.2P	<i>ʔeká:n</i>	D>M	given	given	N
DW	11.1P	<i>ʔokón</i>	0	given	given	N
EW	22.2A	<i>ʔokón</i>	0	given	given	N
EW	22.3A	<i>ʔokón</i>	0	given	given	N
EW	22.3A	<i>ʔokón</i>	0	given	given	N
LN	11.2P	<i>ʔokón</i>	0	given	given	N
LN	11.3P	<i>ʔokón</i>	0	given	given	N
LP	23.3P	<i>ʔokón</i>	0	given	given	N
LP	22.2A	<i>ʔokón</i>	0	given	given	N
NN	23.2P	<i>ʔokón</i>	0	given	given	N
TW	15.2A	<i>ʔokón</i>	0	given	given	N
YG	15.2P	<i>ʔokón</i>	0	given	given	N
YG	11.1P	<i>ʔokón</i>	0	given	given	N
CU	23.2A	<i>ʔokón</i>	D>M	given	given	N
CU	23.3A	<i>ʔokón</i>	D>M	given	given	N
LP	11.2A	<i>ʔokón</i>	D>M	given	given	N
LP	11.3A	<i>ʔokón</i>	D>M	given	given	N
LP	11.3A	<i>ʔokón</i>	D>M	given	given	N
PB	15.2P	<i>ʔokón</i>	D>M	given	given	N
PB	11.2P	<i>ʔokón</i>	D>M	given	given	N
TG	22.2A	<i>ʔokón</i>	D>M	given	given	N
WN	11.3P	<i>ʔokón</i>	D>M	given	given	N
YN	11.1P	<i>ʔokón</i>	D>M	given	given	N
LI	17.3P	<i>ʔolá:n</i>	D>M	given	given	N
NJ	17.1A	<i>ʔolá:n</i>	D>M	given	given	N
NN	17.2A	<i>ʔolá:n</i>	D>M	given	given	N
NN	17.2A	<i>ʔolá:n</i>	D>M	given	given	N
TG	17.3A	<i>ʔolá:n</i>	D>M	given	given	N
WN	16.3A	<i>ʔomá:k</i>	0	given	given	N
DW	22.3P	<i>ʔomá:k</i>	D>M	given	given	N

LI	22.3P	<i>ʔamá:k</i>	D>M	given	given	N
LN	22.1P	<i>ʔamá:k</i>	D>M	given	given	N
NN	16.2P	<i>ʔamá:k</i>	D>M	given	given	N
TG	22.2A	<i>ʔamá:k</i>	D>M	given	given	N
TG	16.1P	<i>ʔamá:k</i>	D>M	given	given	N
TG	16.2P	<i>ʔamá:k</i>	D>M	given	given	N
TW	22.3A	<i>ʔamá:k</i>	D>M	given	given	N
TW	16.2P	<i>ʔamá:k</i>	D>M	given	given	N
TW	16.2P	<i>ʔamá:k</i>	D>M	given	given	N
WN	22.3P	<i>ʔamá:k</i>	D>M	given	given	N
LP	24.2A	<i>təmán</i>	0	given	given	N
LP	24.3A	<i>təmán</i>	D>M	given	given	N
PB	24.1P	<i>caná:t</i>	0	given	given	N
TG	1.3A	<i>ʔenón̄</i>	D>M	given	given	N
TW	4.3P	<i>ʔenón̄</i>	D>M	given	given	N
WN	12.2A	<i>ʔenón̄</i>	D>M	given	given	N
YG	7.3A	<i>ʔenón̄</i>	D>M	given	given	N
LW	11.3A	<i>ʔapón̄</i>	D>M	given	given	N
NN	11.2A	<i>ʔapón̄</i>	D>M	given	given	N
PB	3.2P	<i>kəp^hló:</i>	D>M	given	given	N
CU	14.1A	<i>katá:m</i>	0	given	given	N
DW	14.2A	<i>katá:m</i>	D>M	given	given	N
LW	14.3P	<i>katá:m</i>	D>M	given	given	N
NN	21.2P	<i>katá:m</i>	D>M	given	given	N
NN	21.3P	<i>katá:m</i>	D>M	given	given	N
CU	14.1A	<i>katá:m</i>	M>D	given	given	N
LW	24.1A	<i>matá:ʔ</i>	D>M	given	given	N
NJ	24.1A	<i>matá:ʔ</i>	D>M	given	given	N
TG	24.1A	<i>matá:ʔ</i>	D>M	given	given	N
DW	21.2A	<i>buwá:k</i>	0	given	given	N
LI	21.2A	<i>buwá:k</i>	0	given	given	N
LI	21.3A	<i>buwá:k</i>	0	given	given	N
LN	21.2A	<i>buwá:k</i>	0	given	given	N
LN	21.2A	<i>buwá:k</i>	0	given	given	N
LP	20.3P	<i>buwá:k</i>	0	given	given	N
NJ	21.3P	<i>buwá:k</i>	0	given	given	N
NJ	21.2P	<i>buwá:k</i>	0	given	given	N
NK	21.2P	<i>buwá:k</i>	0	given	given	N
NK	2.1P	<i>buwá:k</i>	0	given	given	N
NN	21.3P	<i>buwá:k</i>	0	given	given	N
TG	20.2P	<i>buwá:k</i>	0	given	given	N

TW	21.2P	<i>buwá:k</i>	0	given	given	N
WN	21.2A	<i>buwá:k</i>	0	given	given	N
YG	21.1A	<i>buwá:k</i>	0	given	given	N
YG	21.2A	<i>buwá:k</i>	0	given	given	N
YG	21.2A	<i>buwá:k</i>	0	given	given	N
YN	21.2A	<i>buwá:k</i>	0	given	given	N
YN	21.3A	<i>buwá:k</i>	0	given	given	N
LP	2.1P	<i>buwá:k</i>	D>M	given	given	N
TG	21.2P	<i>buwá:k</i>	D>M	given	given	N
TG	2.1P	<i>buwá:k</i>	D>M	given	given	N
LW	24.3A	<i>c^huwíək</i>	D>M	given	given	N
NK	24.2A	<i>c^huwíək</i>	D>M	given	given	N
NN	24.3A	<i>c^huwíək</i>	D>M	given	given	N
DW	7.2A	<i>ʔaná:t</i>	D>M	given	given	N
LN	7.3A	<i>ʔaná:t</i>	D>M	given	given	N
LW	7.3P	<i>ʔaná:t</i>	D>M	given	given	N
NJ	11.2A	<i>ʔaná:t</i>	D>M	given	given	N
PB	12.1A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.2A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
PB	4.3A	<i>ʔaná:t</i>	D>M	given	given	N
TG	11.1A	<i>ʔaná:t</i>	D>M	given	given	N
TW	1.2A	<i>ʔaná:t</i>	D>M	given	given	N
TW	11.1A	<i>ʔaná:t</i>	D>M	given	given	N
TW	11.1A	<i>ʔaná:t</i>	D>M	given	given	N
WN	7.3A	<i>ʔaná:t</i>	D>M	given	given	N
CU	21.1A	<i>paʔó:k</i>	D>M	given	given	N
NK	2.2P	<i>ʔuʔé:n</i>	0	given	given	N
NK	2.3P	<i>ʔuʔé:n</i>	0	given	given	N
NN	3.2A	<i>ʔuʔé:n</i>	0	given	given	N
PB	2.3A	<i>ʔuʔé:n</i>	0	given	given	N
WN	19.3P	<i>ʔuʔé:n</i>	0	given	given	N
EW	3.3A	<i>ʔuʔé:n</i>	D>M	given	given	N
LP	3.3A	<i>ʔuʔé:n</i>	D>M	given	given	N
NJ	2.2P	<i>ʔuʔé:n</i>	D>M	given	given	N
NK	20.3A	<i>ʔuʔé:n</i>	D>M	given	given	N
PB	3.3P	<i>ʔuʔé:n</i>	D>M	given	given	N
PB	3.3P	<i>ʔuʔé:n</i>	D>M	given	given	N

YG	3.3P	<i>ʔuʔé:n</i>	D>M	given	given	N
NK	15.3A	<i>kaʔé:w</i>	0	given	given	N
NN	23.2P	<i>kaʔé:w</i>	0	given	given	N
PB	23.1A	<i>kaʔé:w</i>	D>M	given	given	N
TG	21.2P	<i>kaʔé:w</i>	D>M	given	given	N
WN	15.3P	<i>kaʔé:w</i>	D>M	given	given	N
TW	2.2P	<i>niʔú:n</i>	0	given	given	N
DW	20.1A	<i>niʔú:n</i>	D>M	given	given	N
EW	2.1P	<i>niʔú:n</i>	D>M	given	given	N
EW	2.2P	<i>niʔú:n</i>	D>M	given	given	N
EW	2.3P	<i>niʔú:n</i>	D>M	given	given	N
LW	20.3P	<i>niʔú:n</i>	D>M	given	given	N
NJ	2.2P	<i>niʔú:n</i>	D>M	given	given	N
NK	2.1P	<i>niʔú:n</i>	D>M	given	given	N
NN	2.2P	<i>niʔú:n</i>	D>M	given	given	N
PB	2.2A	<i>niʔú:n</i>	D>M	given	given	N
LP	21.2P	<i>dalám</i>	D>M	given	0	PREP
NJ	9.3P	<i>dalám</i>	0	given	0	PREP
NK	19.1A	<i>dalám</i>	0	given	0	PREP
NN	5.3A	<i>dalám</i>	0	given	0	PREP
PB	3.2P	<i>dalám</i>	0	given	0	PREP
PB	3.2P	<i>dalám</i>	0	given	0	PREP
PB	6.3A	<i>dalám</i>	0	given	0	PREP
TG	14.1P	<i>dalám</i>	0	given	0	PREP
WN	19.3P	<i>dalám</i>	D>M	given	0	PREP
TW	21.2P	<i>dalám</i>	0	given	0	PREP
NK	21.2P	<i>datá:</i>	D>M	given	0	PREP
WN	22.3P	<i>datá:</i>	0	given	0	PREP
LN	12.1A	<i>duwà:ʔ</i>	0	given	0	NUM
TG	20.3P	<i>duwà:ʔ</i>	0	given	0	NUM
DW	3.2P	<i>nəbá:k</i>	0	given	0	V
NN	3.2A	<i>nəbá:k</i>	0	given	0	V
NN	9.3P	<i>dəbúh</i>	0	given	0	V
TG	22.2A	<i>nəlén</i>	0	given	0	V
TG	20.3P	<i>kʰəná:ʔ</i>	0	given	0	V
WN	7.3A	<i>neŋó:ʔ</i>	0	given	0	V
LN	8.2P	<i>nəpó:k</i>	0	given	0	V
LP	8.3A	<i>nəpó:k</i>	0	given	0	V
NN	8.3A	<i>nəpó:k</i>	0	given	0	V
PB	8.2P	<i>nəpó:k</i>	0	given	0	V
PB	8.2P	<i>nəpó:k</i>	0	given	0	V

PB	8.3P	<i>nəpók</i>	0	given	0	V
PB	8.3P	<i>nəpók</i>	0	given	0	V
TG	8.2A	<i>nəpók</i>	0	given	0	V
DW	20.3A	<i>nəp^hlà:t</i>	0	given	0	V
DW	20.3A	<i>nəp^hlà:t</i>	0	given	0	V
DW	22.3P	<i>nəp^hlà:t</i>	0	given	0	V
WN	21.2A	<i>nəp^hlà:t</i>	0	given	0	V
WN	20.3A	<i>nəp^hlà:t</i>	0	given	0	V
WN	22.3P	<i>nəp^hlà:t</i>	0	given	0	V
YG	14.2A	<i>nəp^hlà:t</i>	0	given	0	V
LW	18.3P	<i>bətək</i>	0	given	0	V
LW	18.3P	<i>bətək</i>	0	given	0	V
LW	17.3A	<i>bətək</i>	0	given	0	V
NN	5.3A	<i>nuwíək</i>	0	given	0	V
LP	3.3A	<i>məʔám</i>	0	given	0	V
NN	3.3A	<i>məʔám</i>	0	given	0	V
NJ	24.2A	<i>təbák</i>	D>M	given	0	V
NK	4.2P	<i>dəbút</i>	D>M	given	0	V
LN	7.3A	<i>didú:n</i>	D>M	given	0	V
LW	7.3P	<i>didú:n</i>	D>M	given	0	V
NJ	7.3P	<i>didú:n</i>	D>M	given	0	V
NK	7.3P	<i>didú:n</i>	D>M	given	0	V
TW	7.3P	<i>didú:n</i>	D>M	given	0	V
LW	20.3P	<i>bəkáh</i>	D>M	given	0	V
LW	20.3P	<i>bəkáh</i>	D>M	given	0	V
NN	20.3P	<i>bəkáh</i>	D>M	given	0	V
NK	10.3A	<i>nəmát</i>	D>M	given	0	V
LP	4.3P	<i>nəmóh</i>	D>M	given	0	V
NK	20.3A	<i>k^həná:ʔ</i>	D>M	given	0	V
LW	18.2P	<i>neŋó:ʔ</i>	D>M	given	0	V
EW	4.2P	<i>nəŋé:n</i>	D>M	given	0	V
TG	9.2P	<i>niŋít</i>	D>M	given	0	V
TG	9.2P	<i>niŋít</i>	D>M	given	0	V
YG	9.2A	<i>niŋít</i>	D>M	given	0	V
DW	8.2P	<i>nəpók</i>	D>M	given	0	V
DW	8.2P	<i>nəpók</i>	D>M	given	0	V
NJ	8.2A	<i>nəpók</i>	D>M	given	0	V
YG	8.3P	<i>nəpók</i>	D>M	given	0	V
TG	11.2A	<i>nəpó:ŋ</i>	D>M	given	0	V
WN	11.3P	<i>nəpó:ŋ</i>	D>M	given	0	V
LP	15.2A	<i>nəp^hlà:t</i>	D>M	given	0	V

LP	15.2A	<i>nəp^hlà:t</i>	D>M	given	0	V
NK	21.2P	<i>nəp^hlà:t</i>	D>M	given	0	V
NK	20.3A	<i>nəp^hlà:t</i>	D>M	given	0	V
YN	8.1P	<i>nəp^hlà:t</i>	D>M	given	0	V
CU	16.3A	<i>matáj</i>	D>M	given	0	V
TG	11.3A	<i>paták</i>	D>M	given	0	V
LP	2.3P	<i>mətét</i>	D>M	given	0	V
LN	18.3A	<i>bətək</i>	D>M	given	0	V
LN	18.3A	<i>bətək</i>	D>M	given	0	V
LP	18.3P	<i>bətək</i>	D>M	given	0	V
LP	18.3P	<i>bətək</i>	D>M	given	0	V
LP	17.3A	<i>bətək</i>	D>M	given	0	V
TG	17.3A	<i>bətək</i>	D>M	given	0	V
PB	5.3P	<i>nuwíək</i>	D>M	given	0	V
TG	24.3A	<i>nuwíək</i>	D>M	given	0	V
TG	24.3A	<i>nuwíək</i>	D>M	given	0	V
PB	3.3P	<i>məʔám</i>	D>M	given	0	V
NK	12.3P	<i>məʔáŋ</i>	D>M	given	0	V
TG	8.3A	<i>niʔù.n</i>	D>M	given	0	V

Appendix C Stolen Fish Illustrations

(Illustrations are not the same scale as in field instrument)





#1 Line in the water



#2 No Luck



#3 Today's catch



#4 I spy



#5 Full basket



#6 The theft



#7 Picnic spot



#8 Hey you!



#9 Returning the fish



#10 Three boys



#11 Down the path



#12 Have a nice trip



#13 Let's help



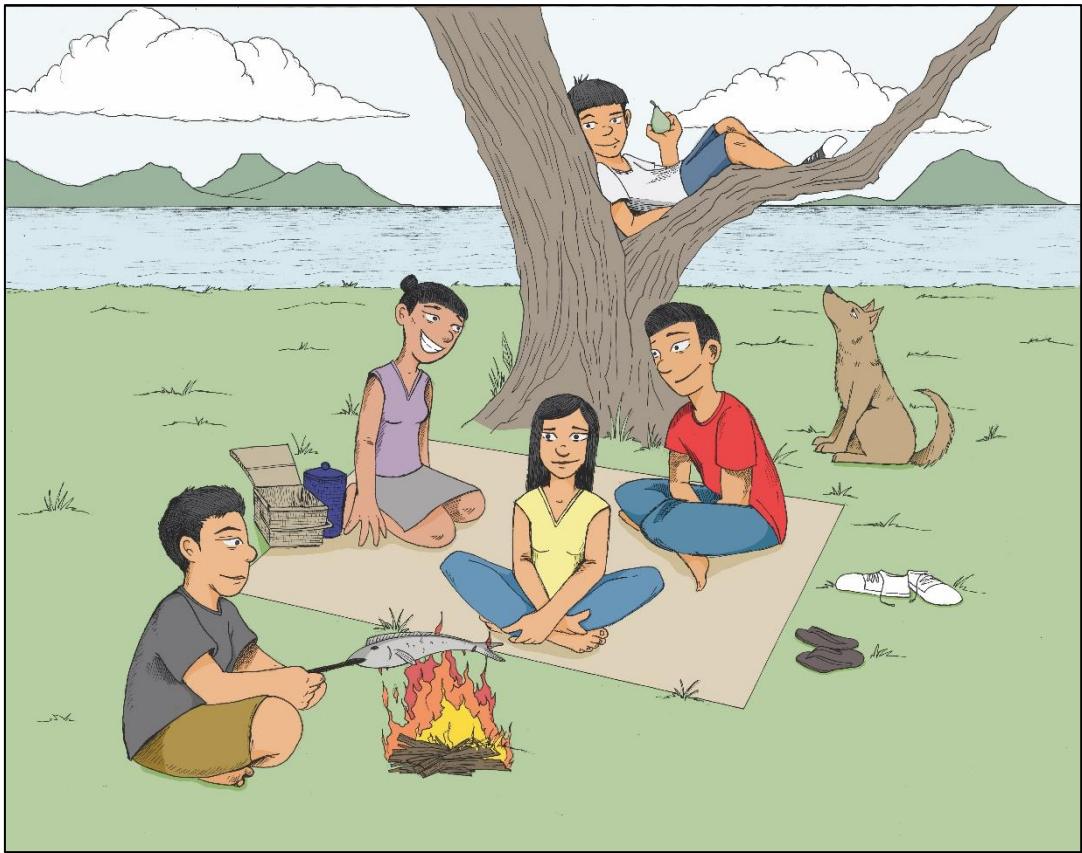
#14 Sharing is caring



#15 Feed the birds



#16 Look what we got



#17 Cookout



#18 Back at the bridge



#19 Where's my fish?



#20 Pesky Birds



#21 Bird Flees



#22 Look who's coming



#23 Trickery



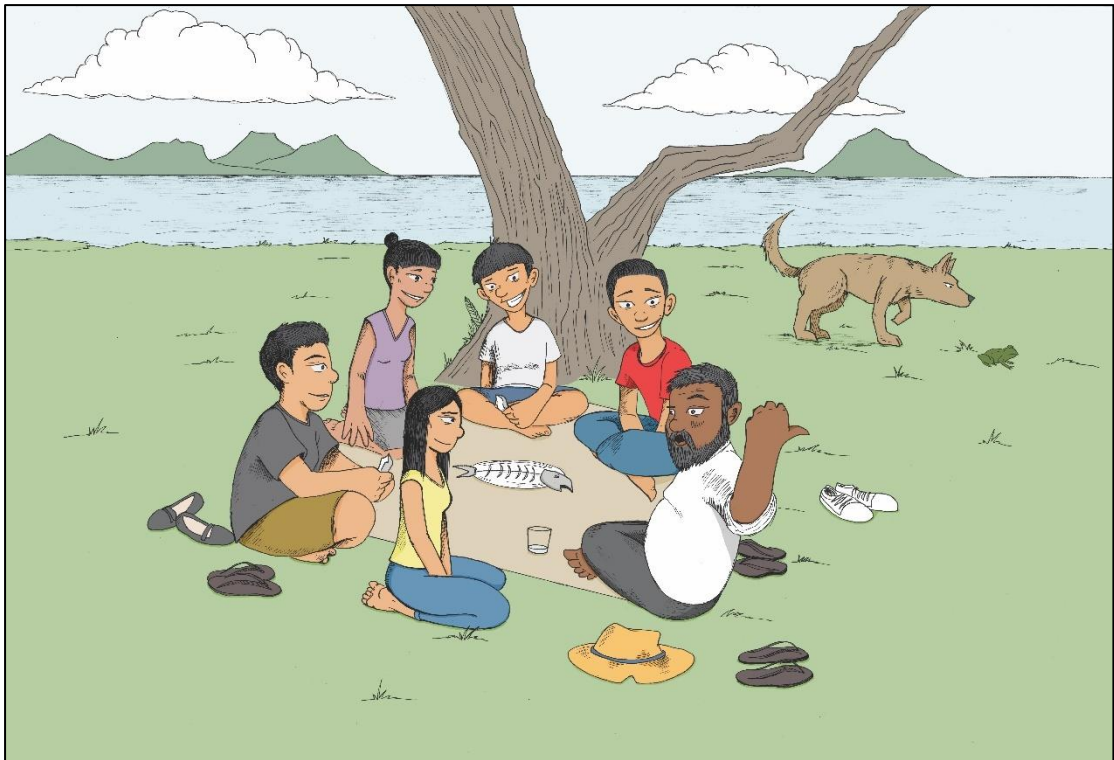
#24 Nothing to see here



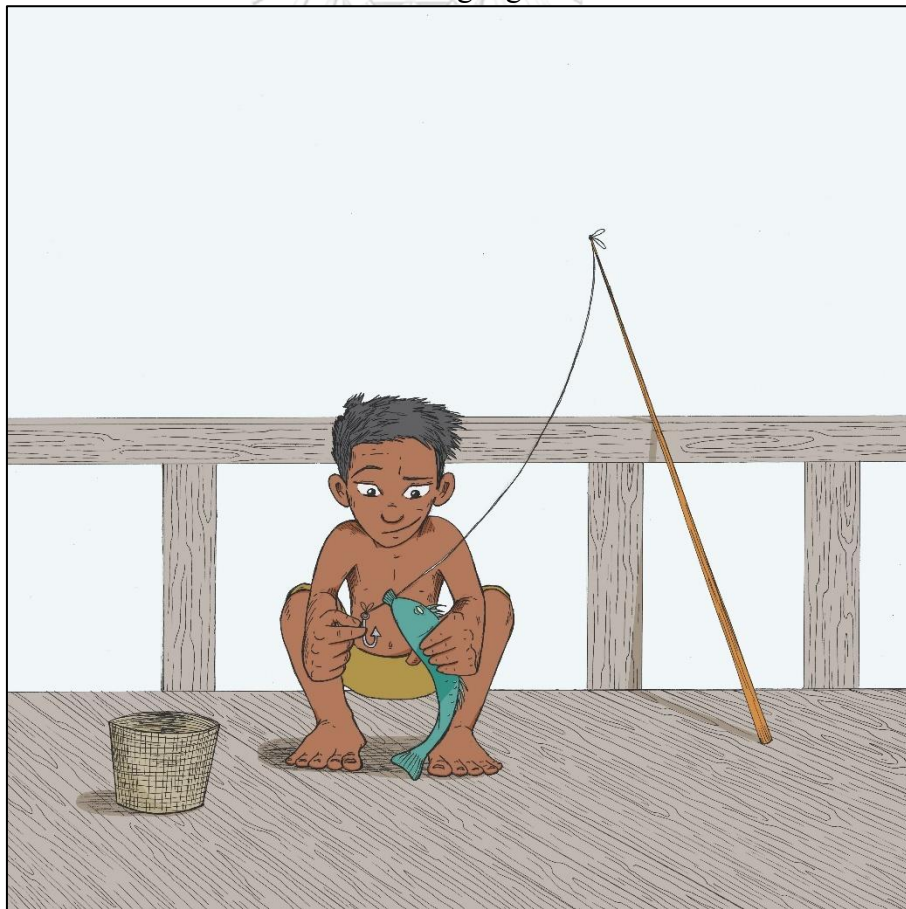
#25 Poor fisherman



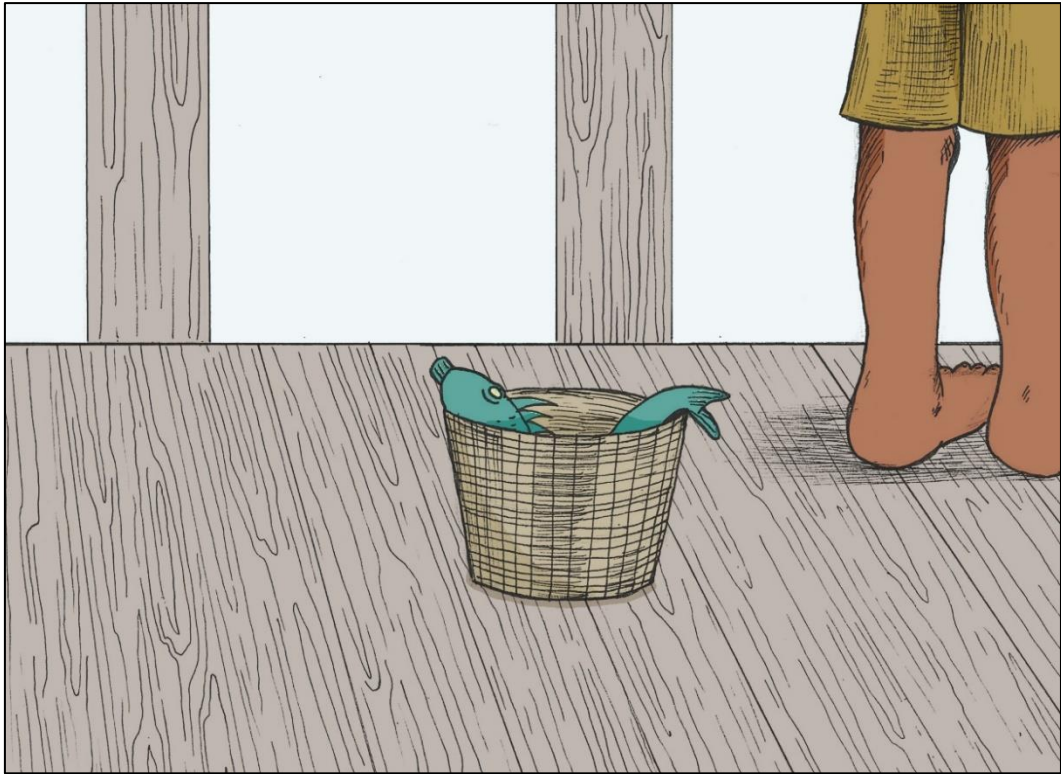
#26 Join us



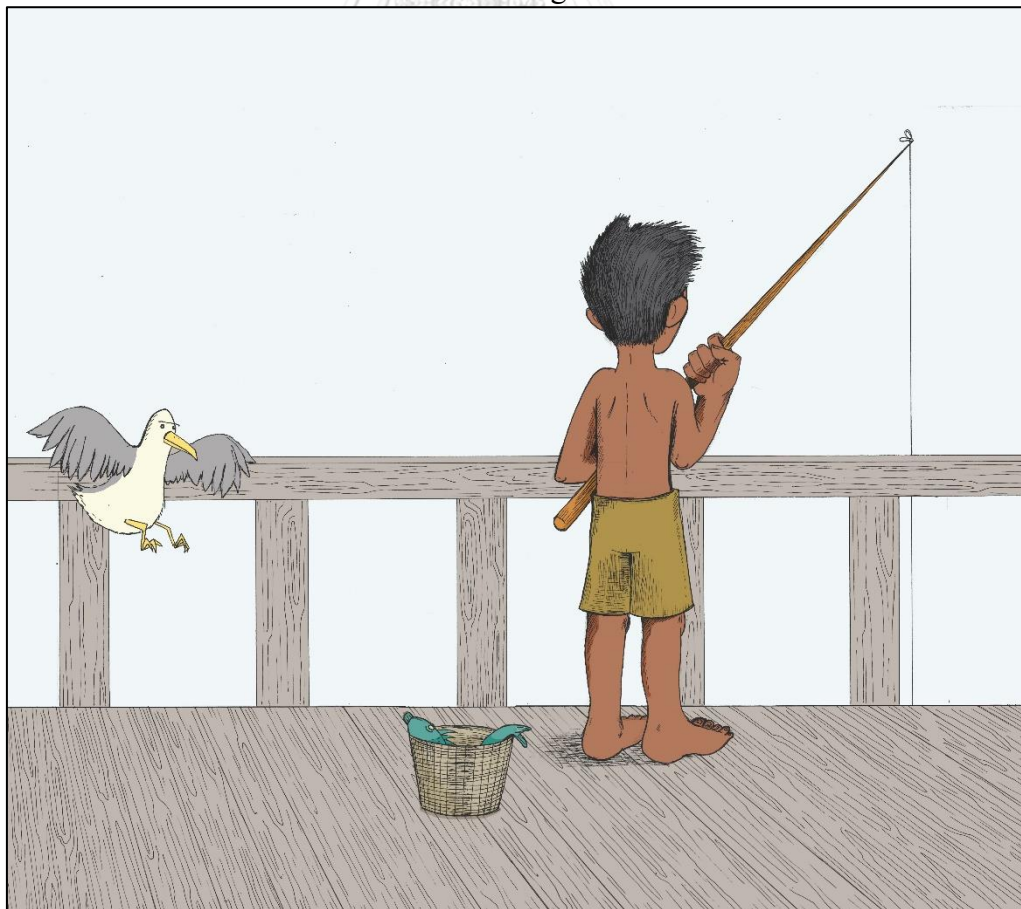
#27 Eating together



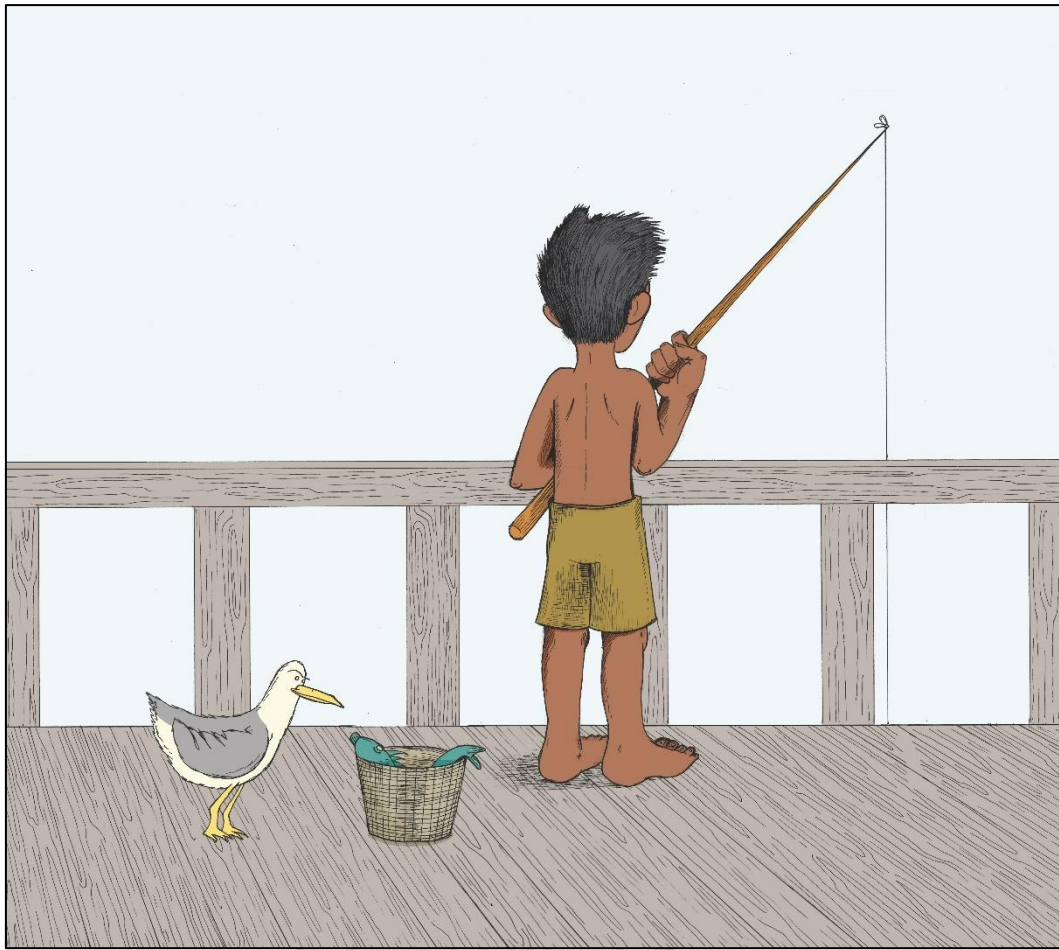
#28 Off the hook



#29 Stolen goods



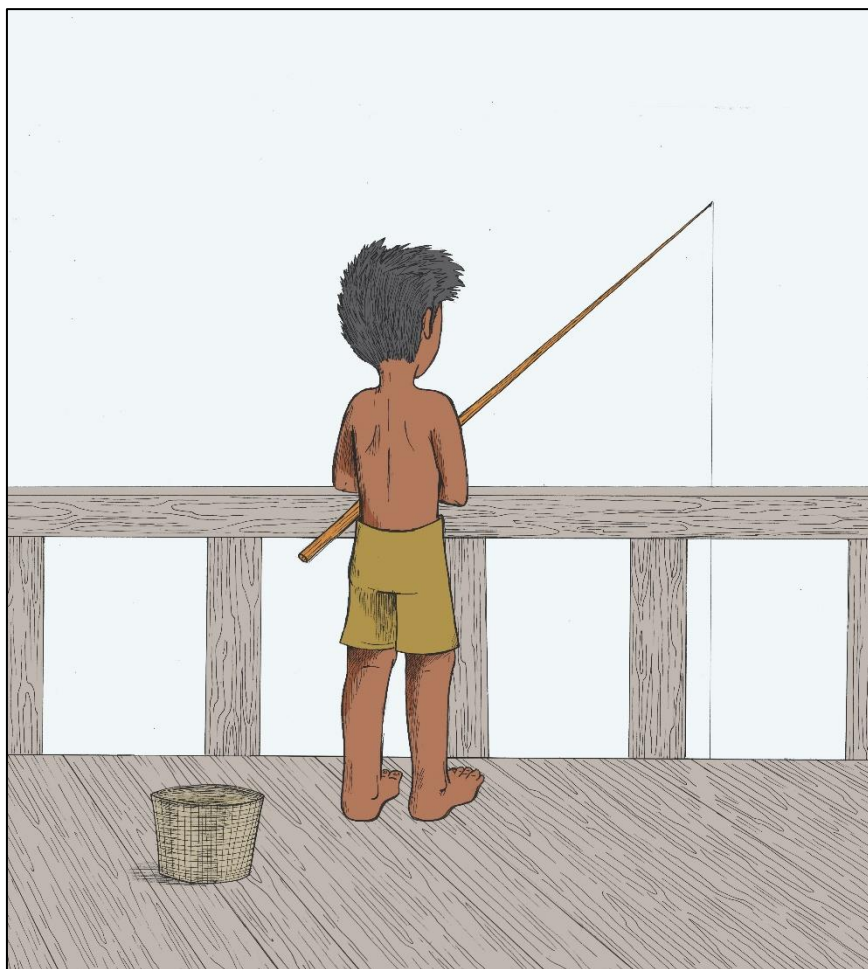
#30 Return of the bird



#31 Bird Creeping



#32 Bye bye bird



#33 Just deserts

Appendix D Transitive Event Picture Sequences

(Illustrations are not the same scale as in field instrument)



Sequence #1 'Boy hugs mother'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



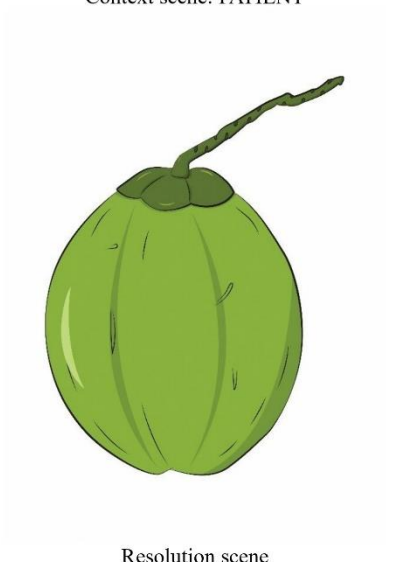
Sequence #2 'Boy opens coconut with machete'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



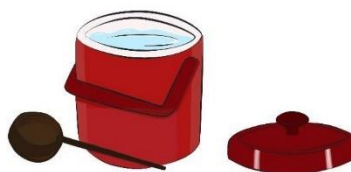
Sequence #3 'Girl gets water'

Context scene: AGENT

Context scene: PATIENT



Target scene



Resolution scene



Sequence #4 'Girl pushes boy'

Context scene: AGENT



Context scene: PATIENT



Target scene



Resolution scene

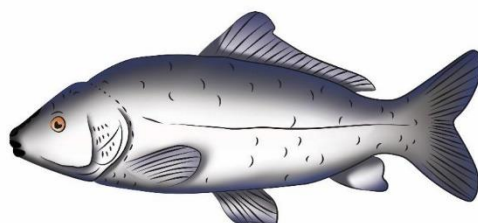


Sequence #5 'Man fishes fish with fishing pole'

Context scene: AGENT



Context scene: PATIENT



Target scene



Resolution scene



Sequence #6 'Man casts net catches fish'

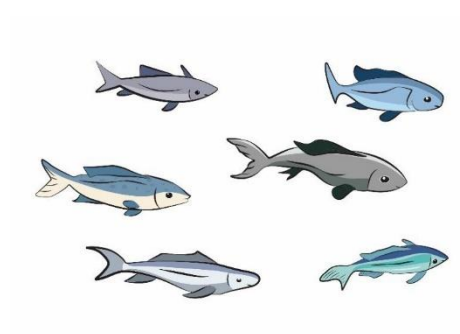
Context scene: AGENT



Target scene



Context scene: PATIENT



Resolution scene



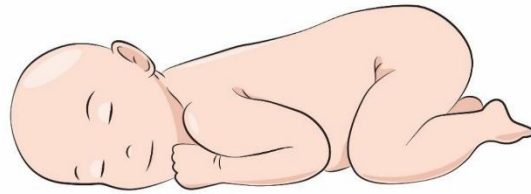
Sequence #7 'Woman covers child with blanket'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



Sequence #8 'Woman washes clothes'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



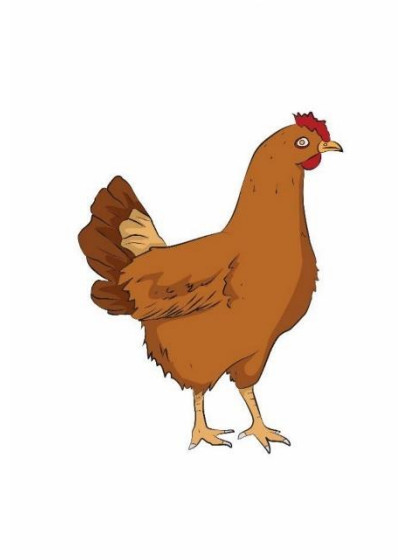
Sequence #9 'Grandpa kills chicken'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene

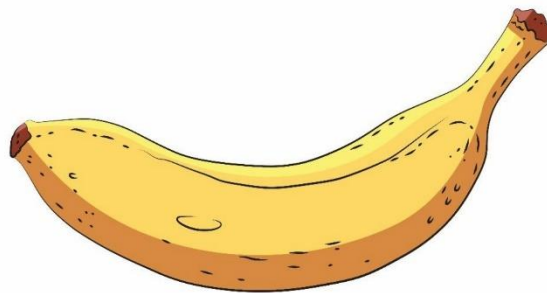


Sequence #10 'Grandma peels banana'

Context scene: AGENT



Context scene: PATIENT



Target scene

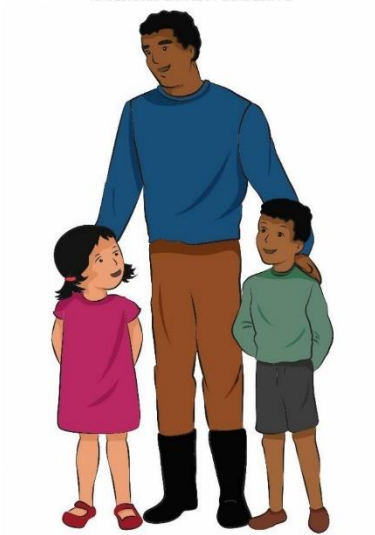


Resolution scene



Sequence #11 'Father, son, and daughter fell a tree'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



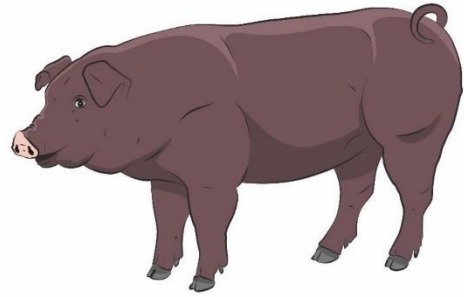
Sequence #12 'Men and women carry pig on stick'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene

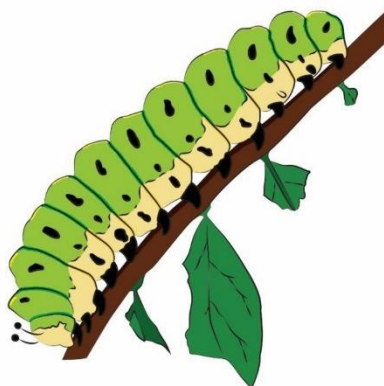


Sequence #13 'Bird pecks caterpillar'

Context scene: AGENT



Context scene: PATIENT



Target scene

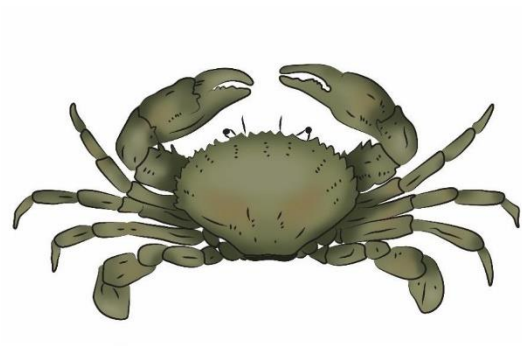


Resolution scene

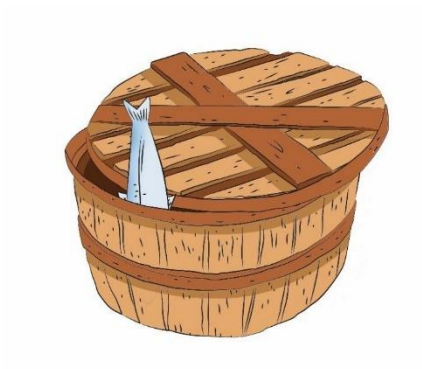


Sequence #14 'Crab grabs fish'

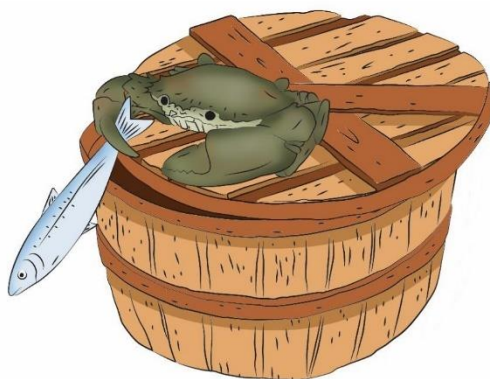
Context scene: AGENT



Context scene: PATIENT



Target scene

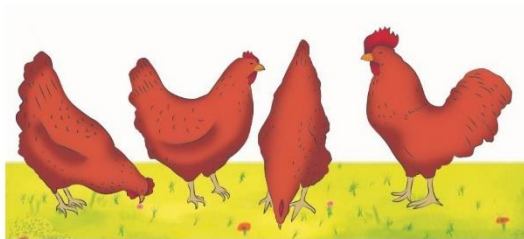


Resolution scene



Sequence #15 'Chickens break tree branch'

Context scene: AGENT

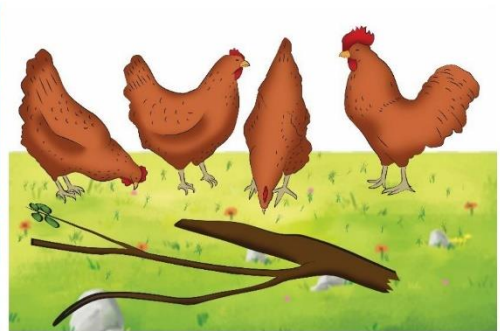


Context scene: PATIENT



Target scene

Resolution scene



Sequence #16 'Bird breaks window'

Context scene: AGENT



Context scene: PATIENT



Target scene

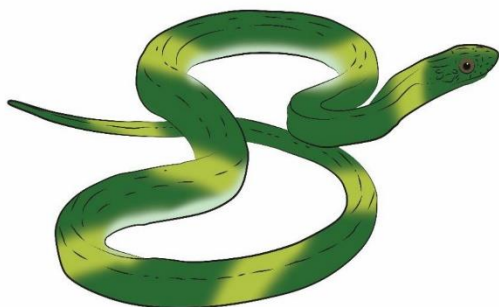


Resolution scene



Sequence #17 'Snake bites grandma'

Context scene: AGENT



Context scene: PATIENT



Target scene



Resolution scene



Sequence #18 'Scorpion stings girl'

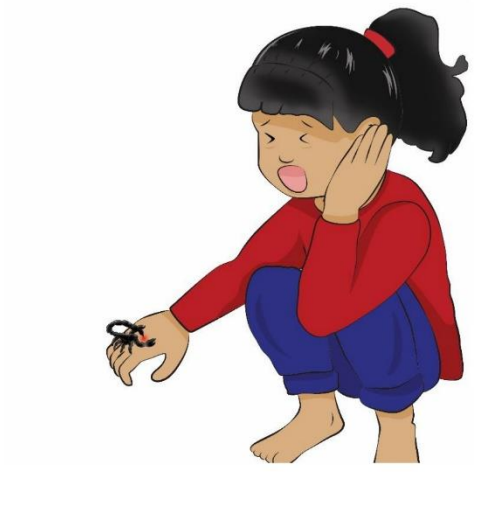
Context scene: AGENT



Context scene: PATIENT



Target scene



Resolution scene



Sequence #19 'Rock breaks boat'

Context scene: AGENT

Context scene: PATIENT



Target scene

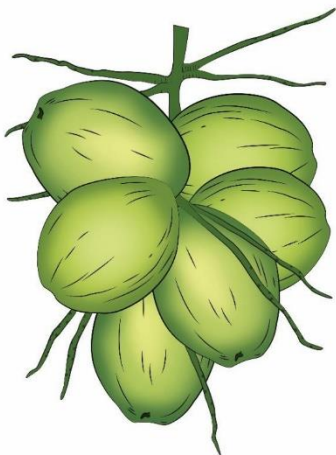


Resolution scene



Sequence #20 'Coconut breaks bottle'

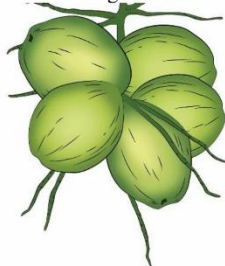
Context scene: AGENT



Context scene: PATIENT



Target scene

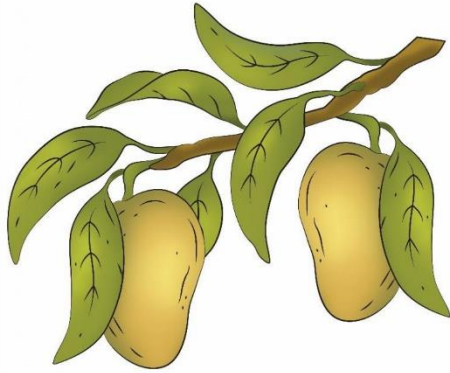


Resolution scene



Sequence #21 'Mango smashes crab'

Context scene: AGENT

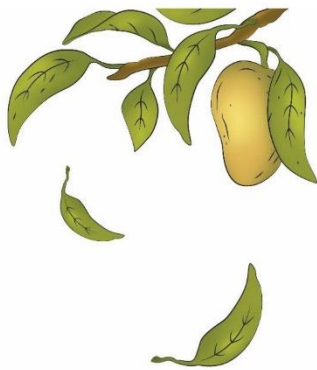


Context scene: PATIENT



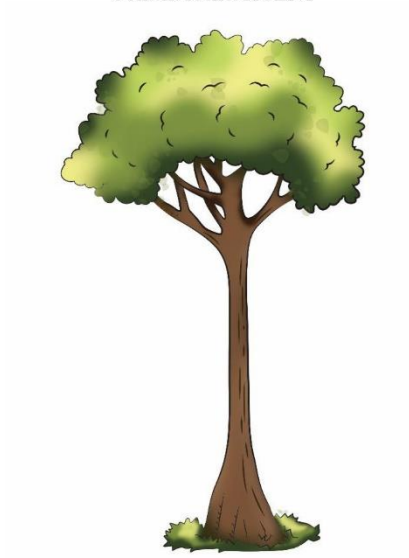
Target scene

Resolution scene



Sequence #22 'Tree smashes house'

Context scene: AGENT



Target scene

Context scene: PATIENT



Resolution scene



Sequence #23 'Tree stump trips grandpa'

Context scene: AGENT



Context scene: PATIENT



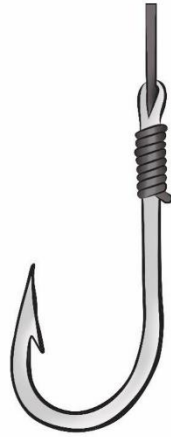
Target scene

Resolution scene



Sequence #24 'Fishhook scratches man'

Context scene: AGENT



Context scene: PATIENT



Target scene



Resolution scene



REFERENCES

- Adamou, E., Haude, K., & Vanhove, M. (2018). Investigating information structure in lesser-known and endangered languages. In E. Adamou, K. Haude, & M. Vanhove (Eds.), *Information Structure in Lesser-described Languages* (pp.1-13). John Benjamins Publishing Company.
- Adelaar, A. (2005). The Austronesian Languages of Asia and Madagascar: A Historical Perspective. In A. Adelaar & N. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 1–42). Routledge.
- Aikhenvald, A. Y., & Dixon, R. M. W. (2006). *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford University Press.
- Ainsworth, L. (1930). *A merchant venturer among the sea gipsies*. Nisbet & Co. Ltd.
- Aissen, J. (2023). Documenting topic and focus. *Language Documentation & Conservation Special Publication*, 26, 11–57.
- Alieva, N. (1994). The Progress of Monosyllabization in Cham as Testified by Field Materials. In C. Ode & W. Stokhof (Eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics (ICAL)* (pp. 541–549). Rodopi.
- Amir, N., Silber-Varod, V., & Izre'el, S. (2004). Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In B. Bel & I. Marlien (Eds.), *Speech Prosody 2004, Nara, Japan, March 23-26, 2004: Proceedings* (pp. 677–680).
- Anderson, J. (1890). *The Selungs of the Mergui Archipelago*. Trubner & Co.
- Ariel, M. (1988). Referring and Accessibility. *Journal of Linguistics*, 24(1), 65–87.

- Ariel, M. (1990). *Assessing Noun Phrase antecedents*. Routledge.
- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55. <https://doi.org/10.1353/lan.2000.0045>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. <https://doi.org/10.1016/j.jml.2009.09.005>
- Arunotai, N. (2008). Saved by an old legend and a keen observation: The case of Moken sea nomads in Thailand. In R. Shaw, N. Uy, & J. Baumwoll (Eds.), *Indigenous knowledge for disaster risk reduction: Good practices and lessons learnt from the Asia-Pacific region* (pp. 73–78). UN ISDR Asia and Pacific.
- Arunotai, N. (2017). “Hopeless at Sea, Landless on Shore”: Contextualising the Sea Nomads’ Dilemma in Thailand. *AAS Working Papers in Social Anthropology (Austrian Academy of Sciences)*, 31, 1–27.
- Arunrungsawat, P., Boonmak, K., Janadisornchai, C., Pinsuwan, T., & Thammanutham, P. (2018). *The Legend of Sampan* [Unpublished Spoken Moklen Text].
- Attavanich, M., & Kobayashi, H. (2014). The construction project of the Moklen ethnic house, Sea Gypsy architecture in Southern Thailand. In *Vernacular Architecture: Towards a Sustainable Future*. CRC Press.
- Attavanich, M., Neef, A., Kobayashi, H., & Tachakitkachorn, T. (2015). Change of Livelihoods and Living Conditions After the 2004 Indian Ocean Tsunami: The Case of the Post-Disaster Rehabilitation of the Moklen Community in Tungwa Village, Southern Thailand. In R. Shaw (Ed.), *Recovery from the Indian Ocean*

- Tsunami: A Ten-year Journey, Disaster Risk Reduction* (pp. 471-486). Springer.
- Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, 41(1), 31–56. <https://doi.org/10.1177/002383090404700102>
- Baclawski Jr., K., & Jenks, P. (2016). Clefts and anti-superiority in Moken. *Journal of the Southeast Asian Linguistics*, 9, 81–96.
- Bard, E. G. B., Anderson, A. H., Sotillo, C., & Aylett, M. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, 42, 1–22. <https://doi.org/10.1006/jmla.1999.2667>
- Barth-Weingarten, D. (2016). *Intonation Units Revisited: Cesuras in talk-in-interaction*. John Benjamins Publishing Company.
- Baumann, S., & Cangemi, F. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, 81, Article 100993. <https://doi.org/10.1016/j.wocn.2020.100993>
- Baumann, S., & Riester, A. (2012). Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 119–162). De Gruyter Mouton. <https://doi.org/10.1515/9783110261790.119>
- Baumann, S., & Schumacher, P. B. (2020). The incremental processing of focus, givenness and prosodic prominence. *Glossa: A Journal of General Linguistics*, 5(1), 6. <https://doi.org/10.5334/gjgl.914>
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability

effects on durations of content and function words in conversational English.

Journal of Memory and Language, 60, 92–111.

<https://doi.org/10.1016/j.jml.2008.06.003>

Benjamin, G. (forthcoming). Moken, Aslian and the Malay World: Ancient connections.

In M. Boutry & J. Ivanoff (Eds.), *Cultural Continuity within the Austronesian People of the Sea. Sojourn: Journal of Social Issues in Southeast Asia*.

Singapore: ISEAS.

Berman, R. A., Slobin, D. I., Aksu-Koç, A. A., Bamberg, M., Dasinger, L., Marchman, V., Neeman, Y., Rodkin, P. C., Sebastián, E., & et al. (1994). *Relating events in narrative: A crosslinguistic developmental study* (pp. xiv, 748). Lawrence Erlbaum Associates, Inc.

Bernatizik, H. A. (2005). *Moken and Semang: 1936-2004 Persistence and Change* (J. Ivanoff, Ed.). White Lotus Co., Ltd.

Bi, Y., Ganushchak, L., Konopka, A., Ren, G., Sui, X., & Chen, Y. (2016, May 31). *Prosodic Encoding of Information Structure in Mandarin Chinese: Evidence from Picture Description Task*. <https://doi.org/10.21437/SpeechProsody.2016-149>

Bishop, N., & Peterson, M. M. (1987). *A Preliminary Language Research Survey Report: The Moklen and Sakai Language Groups*. Summer Institute of Linguistics Library, Bangkok, Thailand.

Blust, R. (2010). The Greater North Borneo hypothesis. *Oceanic Linguistics*, 49, 44–118.

Brayton, D. L. (1846). *A primer of the Selong language*. American Mission Press.

- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. *Cognitive Foundations of Interpretation*, 69–94.
- Brunelle, M. (2009). Diglossia and monosyllabization in Eastern Cham: A sociolinguistic study. In J. Stanford & D. Preston (Eds.), *Variation in Indigenous Minority Languages* (pp. 49–75). John Benjamins Publishing.
<https://doi.org/10.1075/impact.25.04bru>
- Brunelle, M. (2020). The loss of affixation in Cham: Contact, internal drift and the limits of linguistic history. In D. Gil & A. Schapper (Eds.), *Austronesian Undressed: How and why languages become isolating* (pp. 97–118). John Benjamins Publishing. <https://doi.org/10.1075/tsl.129.02bru>
- Brunelle, M., & Pittayaporn, P. (2012). Phonologically constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. *Diachronica*, 29(4), 411–433. <https://doi.org/10.1075/dia.29.4.01bru>
- Carrapiett, W. J. S. (1909). *Ethnographical Survey of India: Burma—The Salons*. Office of the Superintendent, Government Printing, Burma.
- Carroll, A., Kelly, B., & Gawne, L. (2011). *The jackal and crow picture task*.
- Chafe, W. L. (1980). *The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production*.
- Chafe, W. L. (1984). How People Use Adverbial Clauses. *Annual Meeting of the Berkeley Linguistics Society*, 437–449. <https://doi.org/10.3765/bls.v10i0.1936>
- Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing* (pp. xiii, 327). University of Chicago Press.

- Chafe, W. L. (2018). *Thought-based Linguistics: How languages turn thoughts into sounds*. Cambridge University Press.
- Chafe, W. L. (2007). *Language and consciousness*. <https://philpapers.org/rec/CHALAC>
- Chantanakomes, V. (1980). *A description of Moken: A Malayo-Polynesian language* [MA Thesis. Mahidol University].
- Chou, C. (2020). On Being Orang Suku Laut in the Malay World. In J. Levin (Ed.), *Nomad-State Relationships in International Relations*. Palgrave Macmillan.
- Chulalongkorn University Social Research Institute. (2016). Ngan ruam yat chattiphan chao le pi thi 7 "Patthana phuhen thi khum krong thang wathanatham chattiphan chao le su kan phatthana thi yang yuen [The 7th Meeting of the Chao Le Ethnic groups "Development of cultural protected areas for the Chao Le ethnic groups for sustainable development].
- Clark, H. H., & Haviland, S. E. (1974). Psychological processes as linguistic explanation. In D. Cohen (Ed.), *Explaining Linguistic Phenomena*.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the Given-New Contrast. In R. O. Freedle (Ed.), *Discourse Production and Comprehension* (pp. 1–40). Lawrence Erlbaum Associates.
- Clifton, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition*, 32(6), 886–895. <https://doi.org/10.3758/BF03196867>
- Cohn, N. (2018). In defense of a “grammar” in the visual language of comics. *Journal of Pragmatics*, 127, 1–19. <https://doi.org/10.1016/j.pragma.2018.01.002>
- Court, C. (1971). A fleeting encounter with the Moken (the sea gypsies) in Southern Thailand: Some linguistic and general notes. *Journal of the Siam Society*, 59(1),

83–95.

- Cowan, H. K. J. (1948). Aanteekeningen betreffende de verhouding van het Atjehsch tot de Mon-Khmer talen. In *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indie.: Vol. Deel 104*. The Hague.
- Cresti, E. (2018). The illocution-prosody relationship and the Information Pattern in spontaneous speech according to the Language into Act Theory (L-AcT). *Linguistik Online*, 88(1). <https://doi.org/10.13092/lo.88.4189>
- Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33(5), 839–882. <https://doi.org/10.1515/ling.1995.33.5.839>
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Longman.
- Croft, W. (2001). *Radical construction grammar. Syntactic theory in typological perspective*. Oxford University Press.
- Croft, W. (2005). Logical and typological arguments for Radical Construction Grammar. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 273–314). John Benjamins Publishing Company. <https://doi.org/10.1075/cal.3.11cro>
- Croft, W. (2007). Intonation Units and Grammatical Structure in Wardaman and in Cross-linguistic Perspective. *Australian Journal of Linguistics*, 27(1), 1–39. <https://doi.org/10.1080/07268600601172934>
- Croft, W. (2022). *Morphosyntax: Constructions of the World's Languages*. Cambridge University Press. <https://doi.org/10.1017/9781316145289>
- Cruttenden, A. (1993). The De-accenting and Re-Accenting of Repeated Lexical Items. In D. House & P. Touati (Eds.), *Proceedings of the ESCA Workshop, Working*

- papers* (Vol. 41, pp. 16–19). Dept of Linguistics and Phonetics, Lund, Sweden.
- Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal? In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic Organization of Discourse in the Languages of Europe* (pp. 311–356). Mouton.
- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics, 6th Edition*. Wiley-Blackwell.
- Dancause, K. N., Chan, C. W., Arunotai, N. H., & Lum, J. K. (2009). Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *Journal of Human Genetics*, 54(2), 86–93.
<https://doi.org/10.1038/jhg.2008.12>
- Degand, L., & Simon, A. C. (2008). Minimal discourse units in spoken French: Uncovering genre-bound segmentation strategies. In W. Ramm & C. Fabricius-Hansen (Eds.), *Linearisation and Segmentation in Discourse. Multidisciplinary Approaches to Discourse 2008 (MAD 08)* (pp. 31-42). Oslo: Dept. of Literature, Area Studies and Europ. Languages, Univ. of Oslo
- Dixon, R. M. W. (2009). *Basic Linguistic Theory Volume 1: Methodology*. OUP Oxford.
- Dixon, R. M. W. (2010). *Basic Linguistic Theory Volume 2: Grammatical Topics*. OUP Oxford.
- Downing, P. A., & Noonan, M. (1995). Word Order in Discourse. In *Tsl.30*. John Benjamins Publishing Company.
- Dryer, M. S. (2005). Order of Subject, Object, and Verb. In B. Comrie, M. S. Dryer, D. Gil, & M. Haspelmath (Eds.), *World Atlas of Language Structures* (pp. 330–

- 333). Oxford University Press.
- Dryer, M. S. (2006). Descriptive theories, explanatory theories, and basic linguistic theory. In F. Ameka, A. Dench, & N. Evans (Eds.), *Catching Language: The Standing Challenge of Grammar Writing* (pp. 207–234). Mouton De Gruyter.
- Du Bois, J. W. (1987). The Discourse Basis of Ergativity. *Language*, 63(4), 805–855.
<https://doi.org/10.2307/415719>
- Du Bois, J. W. (2003). Argument Structure: Grammar in use. In J. W. Du Bois, L. E. Kumpf, & W. J. Ashby (Eds.), *Preferred Argument Structure: Grammar as architecture for function* (pp. 11-60). John Benjamins Publishing.
<https://doi.org/10.1075/sidag.14.04dub>
- Du Bois, J. W. (2008). *Intonation unit cues in context. Unpublished manuscript. University of California, Santa Barbara.*
- Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). Discourse Transcription. *Santa Barbara Papers in Linguistics*, 4, 237.
- Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of Discourse Transcription. In J. A. Edwards, M. D. Lampert (Eds.) *Talking Data: Transcription and coding in discourse research* (pp. 45-89). Lawrence Erlbaum Associates.
- ELAN (Version 6.4) [Computer software]* ((Version 6.4)). (2022). [Computer software]. Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- Enfield, N. J. (2016). *Natural causes of language: Frames, biases, and cultural transmission*. Language Science Press. <https://langsci-press.org/catalog/book/48>

- Enfield, N. J. (2021). *The Languages of Mainland Southeast Asia*. Cambridge University Press. <https://doi-org/10.1017/9781108605618>
- Enfield, N. J. (2023). Scale in Language. *Cognitive Science*, 47(10), e13341. <https://doi.org/10.1111/cogs.13341>
- Everett, C. (2009). A reconsideration of the motivations for preferred argument structure. *Studies in Language*, 33, 1–24.
- Fernandez-Vest, M. M. J. (2015). *Detachments for cohesion – Toward an information grammar of oral languages*. De Gruyter Mouton.
- Ferrari, O., Hinshirana, N., Utpuay, K., & Ivanoff, J. (2006). *Turbulence on Ko Phra Thong*. Amarin Printing and Publishing.
- Fowler, C. A. (1988). Differential Shortening of Repeated Content Words Produced in Various Communicative Contexts. *Language and Speech*, 31(4), 307–319. <https://doi.org/10.1177/002383098803100401>
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5), 489–504. [https://doi.org/10.1016/0749-596X\(87\)90136-7](https://doi.org/10.1016/0749-596X(87)90136-7)
- Gaglia, S., & Schwarze, C. (2015). The controversial status of Romance pronominal clitics – a new criterion. *Linguistische Berichte*, 2015.
- Givon, T. (1975). Focus and the scope of assertion: Some Bantu evidence. *Studies in African Linguistics*, 6, 185–205.
- Givon, T. (1983). Topic Continuity in Discourse. In *Tsl.3*. John Benjamins Publishing Company. <https://benjamins.com/catalog/tsl.3>
- Goldberg, A. E. (2006). *Constructions at work*. Oxford University Press.

- Grant, A., & Sidwell, P. (Eds.). (2005). *Chamic and beyond: Studies in mainland Austronesian languages / edited by Anthony Grant and Paul Sidwell*. Pacific Linguistics.
- Gregory, M., Raymond, W., Bell, A., Fosler-lussier, E., & Jurafsky, D. (1999). The Effects of Collocational Strength and Contextual Predictability in Lexical Production. *The Proceedings of the Institute of Medicine of Chicago*, pp. 151–166.
- Güldemann, T., Zerbian, S., & Zimmermann, M. (2015). Variation in Information Structure with Special Reference to Africa. *Annual Review of Linguistics*, 1(1), 155–178. <https://doi.org/10.1146/annurev-linguist-030514-125134>
- Gundel, J. K., Bassene, M., Gordon, B., & Humnick, L. (2010). Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics*, 42(7), 1770–1785.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69(2), 274–307. JSTOR. <https://doi.org/10.2307/416535>
- Gussenhoven, C., & Gussenhoven, P. of G. and E. P. C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.
- Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, 3(2), 199–244. JSTOR.
- Haspelmath, M. (2015). Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In *Defining vs. Diagnosing linguistic categories: A case study of clitic phenomena* (pp. 273–304). De Gruyter Mouton.

<https://doi.org/10.1515/9781614514510-009>

- Haspelmath, M. (2020). The structural uniqueness of languages and the value of comparison for language description. *Asian Languages and Linguistics*, 1(2), 346–366.
- Hilpert, M. (2019). Lexicalization in Morphology. In *Oxford Research Encyclopedia of Linguistics*. <https://doi.org/10.1093/acrefore/9780199384655.013.622>
- Himmelman, N. (2014). Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language*, 90, 927–960.
<https://doi.org/10.1353/lan.2014.0105>
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36, 161–195.
- Himmelman, N. P., & Ladd, D. R. (2008). Prosodic Description: An Introduction for Fieldworkers. *Language Documentation & Conservation*, 2(2), 244–274.
- Himmelman, N. P., Sandler, M., Strunk, J., & Unterladseter, V. (2018). On the universality of intonational phrases: A crosslinguistic interrater study. *Phonology*, 35, 207–245. <https://doi.org/10.1017/S0952675718000039>
- Hinshirana, N. (1996). *The analysis of Moken opportunistic foragers' intragroup and intergroup relations* [Ph.D. Dissertation, University of Hawai'i at Manoa].
- Hirst, D., & Di Cristo, A. (1998). *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press.
- Hogan, D. W. (1972). Men of the sea: Coastal tribes of south Thailand's west coast. *Journal of the Siam Society*, 59(1), 83–95.
- Hogan, D. W. (1983). *A Thai Orthography for the Moken Language*. Unpublished

manuscript. Melbourne, Australia: Pacific College of Graduate Studies.

Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-72739-4>

IRASEC. (2009). *Southern Ethnic Dynamism: The Andaman Littoral and Marine Populations*. Chulalongkorn University Social Research Institute.

Ivanoff, J. (1997). *Moken: Sea-Gypsies of the Andaman Sea Post-war Chronicles*. White Lotus Co., Ltd.

Ivanoff, J. (1999). *The Moken Boat: Symbolic Technology*. White Lotus Press.

Ivanoff, J. (2001). *Rings of Coral, Moken Folktales*. White Lotus Press.

Ivanoff, J. (2005). *Moken and Semang: 1936-2004 Persistence and Change*. White Lotus Press.

Iwasaki, S. (1993). The structure of the intonation unit in Japanese. *Japanese/Korean Linguistics*, 3, 39–53.

Iwasaki, S. (1996). The syntactic and functional structures of intonation unit in Thai. *Pan-Asiatic Linguistics*, 2, 750–762.

Iwasaki, S., & Ingkaphirom, P. (2005). *A Reference Grammar of Thai*. Cambridge University Press.

Iwasaki, S., & Tao, H. (1993). A Comparative Study of the Structure of the Intonation Unit in English, Japanese, and Mandarin Chinese. Paper presented at the Annual Meeting of the Linguistic Society of America.

Izre'el, S. (2020). The basic unit of spoken language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In S. Isre'el, H.,

- Mello, A. Panunzi, & T. Raso (Eds.) *In search of Basic Units of Spoken Language: A corpus-driven approach* (pp. 77–106). John Benjamins Publishing Company.
- Izre'el, S., Mello, H., Panunzi, A., & Raso, T. (2020). Introduction. In search of a basic unit of spoken language: Segmenting speech. In *In Search of Basic Units of Spoken Language. A corpus-driven approach*. John Benjamins Publishing Company (pp. 1-32). <https://doi.org/10.1075/scl.94.int>
- Izre'el, S., & Mettouchi, A. (2015). Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In A. Mettouchi, M., Vanhove, & D. Caubet (Eds.) *Corpus-Based Studies of Lesser-Described Languages: The CorpAfroAs corpus of Spoken AfroAsiatic languages* (pp.13–41). John Benjamins Publishing.
- Jaeger, T. F., & Buz, E. (2017). Signal Reduction and Linguistic Encoding. In E. M. Fernández & H. S. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 38–81). Wiley-Blackwell. <https://doi.org/10.1002/9781118829516.ch3>
- Jenks, P. (2007). *Unpublished Moken field data* [dataset].
- Jenks, P. (2010). *Incipient Tone in Moken* [Unpublished manuscript].
- Kaland, C., & Himmelmann, N. P. (2020). Repetition Reduction Revisited: The Prosody of Repeated Words in Papuan Malay. *Language and Speech*, 63(1), 31–55. <https://doi.org/10.1177/0023830918820044>
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52.

<https://doi.org/10.1016/j.cognition.2017.05.001>

- Kiser, M. (2014). *Analyzing Navajo Discourse: Investigating Form and Function of Intonational Units in Referential Discourse* [Ph.D. Dissertation, The University of New Mexico].
- Klamer, M., & Moro, F. R. (2020). What is “natural” speech? Comparing free narratives and Frog stories in Indonesia. *Language Documentation & Conservation*, 14, 238–313.
- Koh, A. (2016). *How the Moken Sea Gypsies got their book*. Tyros Global.
- Kraisame, S. (2016). *Process of documentation and preservation of an endangered language: A case study of Moken on Surin Islands* [Ph.D. Dissertation, Mahidol University].
- Krifka, M., & Musan, R. (2012). Information structure: Overview and linguistic issues. In *The Expression of Information Structure* (pp. 1–44). De Gruyter Mouton.
- Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, 38(8), 1137–1146.
<https://doi.org/10.3758/MC.38.8.1137>
- Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents* (pp. xvi, 388). Cambridge University Press. <https://doi.org/10.1017/CBO9780511620607>
- Larish, M. (1997). Moklen-Moken phonology: Mainland or insular Southeast Asian typology? In C. Ode & W. Stokhof (Eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics* (pp. 125–150). Leiden University.

- Larish, M. (1999). *The position of Moken and Moklen within the Austronesian language family*. [Ph.D. Dissertation, University of Hawaii at Manoa].
- Larish, M. (2005). Moken and Moklen. In A. Adellar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 513–533). Routledge.
- Lehmann, C. (2002). New reflections on grammaticalization and lexicalization. In I. Wischer & G. Diewald (Eds.), *New Reflections on Grammaticalization* (pp. 1–18). John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.49.03leh>
- Levin, B. (2018). Argument Structure. In M. Aronoff (Ed.), *Oxford Bibliographies in Linguistics*. Oxford University Press.
<https://www.oxfordbibliographies.com/display/document/obo-9780199772810/obo-9780199772810-0099.xml>
- Levin, B., & Hovav, M. R. (2019). Lexicalization Patterns. In R. Truswell (Ed.), *The Oxford Handbook of Event Structure* (pp. 395–425). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199685318.013.18>
- Lewis, M. B. (1960). *Moken texts and word-list: A provisional interpretation*. Museums Department, Federation of Malaya.
- Linders, G. M., & Louwrese, M. M. (2022). Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-022-02142-9>
- Linell, P. (1982). *The Written Language Bias in Linguistics*. Linköping University Electronic Press. <https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-111550>
- Linell, P. (2019). The Written Language Bias (WLB) in linguistics 40 years after. *Language Sciences*, 76, 101230. <https://doi.org/10.1016/j.langsci.2019.05.003>

- Loss, D., Chanchaochai, N., Pittayaporn, P., & Enfield, N. J. (forthcoming). Information structure and changes in Moklen word-form.
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. <https://doi.org/10.1016/j.cognition.2012.09.010>
- Masia, V. (2022). Remarks on information structure marking asymmetries: The epistemological view on the micropragmatic profile of utterances. In D. Garassino & D. Jacob (Eds.), *When Data Challenges Theory: Unexpected and paradoxical evidence in information structure* (pp. 57–90). John Benjamins Publishing Company. <https://doi.org/10.1075/la.273.02mas>
- Mason, F. (1860). *Burmah, its people and natural productions*. Thos Stowe Ranney.
- Matić, D. (2015). Information Structure in Linguistics. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 95–99). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.53013-X>
- Matić, D. (2022). Alternatives to information structure. In D. Garassino & D. Jacob (Eds.), *When Data Challenges Theory: Unexpected and paradoxical evidence in information structure* (pp. 91–112). John Benjamins Publishing. <https://doi.org/10.1075/la.273.03mat>
- Matisoff, J. (1990). Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony. *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, 543–559.
- Matsumoto, K. (2003). *Intonation Units in Japanese Conversation: Syntactic, Informational and Functional Structures*. John Benjamins Publishing.

- Matter, F. (2020). An Inspection of Preferred Argument Structure in Mapudungun Narratives. *International Journal of American Linguistics*, 86(1), 59–93.
<https://doi.org/10.1086/705754>
- Mayer, M. (1969). *Frog, where are You?* Dial Books for Young Readers.
- McElvenny, J. (n.d.). *Podcast episode 21: Karl Bühler's Organon model and the Prague Circle* (21). Retrieved January 1, 2022, from
<https://hiphilangsci.net/2022/01/01/podcast-episode-21/>
- Mettouchi, A., Lacheret-Dujor, A., Silber-Varod, V., & Izre'el, S. (2007). Only Prosody? Perception of speech segmentation. *Nouveaux Cahiers de Linguistique Française 28 : Interfaces Discours – Prosodie : Actes Du 2ème Symposium International and Colloque Charles Bally*, 207–218.
- Michaud, A. (2012). Monosyllabization: Patterns of evolution in Asian languages. In N. Nau, T. Stolz, & C. Stroh (Eds.), *Monosyllables: From phonology to typology* (pp. 115–130). Akademie Verlag.
- Michaud, A., & Brunelle, M. (2016). Information Structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic). In C. Fery & S. Ishihara (Eds.), *Oxford Handbook of Information Structure*. Oxford University Press.
- Mithun, M. (2020). Basic units of speech segmentation. In S. Izre'el, H. Mello, & T. Raso (Eds.), *In Search of Basic Units of Spoken Language: A corpus-driven approach* (pp. 349-358). John Benjamins Publishing.
- Mithun, M. (2021). The Multidimensional Organization of Speech: *Cadernos de Linguística*, 2(1), 01–23. <https://doi.org/10.25189/2675-4916.2021.v2.n1.id287>
- Mithun, M., & Chafe, W. (1999). What are S, A, and O? *Studies in Language*, 23, 569–

596. <https://doi.org/10.1075/sl.23.3.05mit>

Mueanhawong, K. (2017, February 4). “Home at last”: Six sea gypsy families cleared of land-encroachment charges in Phuket. *Phuket Gazette*.

Nakagawa, N. (2020). *Information structure in spoken Japanese: Particles, word order, and intonation*. Language Science Press. <https://langsci-press.org/catalog/book/178>

Naw Say Bay. (1995). The phonology of the Dung dialect of Moken. In D. Bradley (Ed.), *Paper in Southeast Asian Linguistics No.13: Studies in Burmese languages* (pp. 193–205). Pacific Linguistics, Series A - No. 83. Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Neeleman, A., & van de Koot, H. (2015). Word Order and Information Structure. In C. Fery & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure*. Oxford University Press, pp. 383-401.

Nomoto, H., & Wahab, K. Abd. (2012). Kena Adversative Passives in Malay, Funny Control, and Covert Voice Alternation. *Oceanic Linguistics*, 51(2), 360–386.

Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, 30(1–2). <https://doi.org/10.1080/01690965.2013.805795>

Ozerov, P. (2018). Tracing the sources of Information Structure: Towards the study of interactional management of information. *Journal of Pragmatics*, 138, 77–97. <https://doi.org/10.1016/j.pragma.2018.08.017>

Ozerov, P. (2021). Multifactorial Information Management (MIM): Summing up the

emerging alternative to Information Structure. *Linguistics Vanguard*, 7(1).

<https://doi.org/10.1515/lingvan-2020-0039>

- Pan, H., Huang, W., & Huang, Y.-H. (2005, June 15). Production and Perception of New, Given, and Contrastive Information in Taiwan Mandarin. *Proceedings of ISCA Workshop on Plasticity in Speech Perception*. PSP2005, London, UK.
- Pittayaporn, P. (2005). Moken as a Mainland Southeast Asian language. In A. Grant & P. Sidwell (Eds.), *Chamic and Beyond: Studies in Mainland Austronesian Languages* (pp. 189–209). Pacific Linguistics.
- Pittayaporn, P. (2006). When words erode: Moken trisyllabic syncopation and PAN stress. *Tenth International Conference on Austronesian Linguistics*.
- Pittayaporn, P., & Choemprayong, S. (to appear). A proposal for a Thai-based Moklen orthography. *Language Documentation & Conservation*.
- Pittayaporn, P., Pornpottanamas, W., & Loss, D. (2022). *Moklen-Thai-English Dictionary: A Pilot Version*. Academic Work Dissemination Project, Faculty of Arts, Chulalongkorn University.
- Prasithratsint, A. (2004). The adversative passive marker as a prominent areal feature in Southeast Asian languages. In S. Buruspat (Ed.), *Papers from the eleventh annual meeting of the Southeast Asian linguistics society 2001* (pp. 583–598). Monograph Series Press.
- Premisrat, S., Kraisame, S., Choosri, I., & Pengraphan, S. (2013). *A study of Moken language and Moken—Thai basic lexicography for Moken orthography development by community-based*. Department of National Parks, Wildlife and Plant Conservation.

- Prentice, J. L. (1967). Effects of cuing actor vs cuing object on word order in sentence production. *Psychonomic Science*, 8(4), 163–164.
<https://doi.org/10.3758/BF03331600>
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In *Radical Pragmatics*. /paper/Toward-a-taxonomy-of-given-new-information-Prince/6e8d16d8e320d0ebb3756080390a6aa778b5a1fe
- Ratitamkul, T. (2007). *Argument Realization in Thai* [Ph.D. Dissertation, University of Illinois at Urbana-Champaign].
- Riesberg, S., Kalbertodt, J., Baumann, S., & Himmelmann, N. P. (2018). On the perception of prosodic prominences and boundaries in Papuan Malay. In S. Riesberg, A. Shiohara, & A., Utsumi (Eds.) *Perspectives on information structure in Austronesian languages* (pp. 389-414). Language Science Press.
- Riesberg, S., Kalbertodt, J., Baumann, S., & Himmelmann, N. P. (2020). Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology*, 11(1), 1–35.
- Riester, A., & Baumann, S. (2017). *The RefLex scheme—Annotation guidelines* [workingPaper]. Stuttgart : Universität Stuttgart, SFB.
<https://doi.org/10.18419/opus-9011>
- Robinson, D., Drozdowski, D., & Enright, J. K. (2021). Land territorialisation, contestation and informal place-laws of Indigenous peoples in Phuket and Phang Nga, Thailand. In *Handbook on Space, Place and Law* (pp. 156–169). Edward Elgar Publishing.
- Röhr, C. T. (2016). *The Information Status of Nominal and Verbal Expressions:*

Intonational Evidence from Production and Perception in German. Universitäts zu Köln.

- Röhr, C. T., Thies, T., Baumann, S., & Grice, M. (2016). Prosodic Marking of Information Status in Task-Oriented Dialogues. *Proceedings of the 12. Tagung Phonetik Und Phonologie in Deutschsprachigen Taum*, 168–171.
- Ross, B., Fletcher, J., & Nordlinger, R. (2016). The Alignment of Prosody and Clausal Structure in Dalabon. *Australian Journal of Linguistics*, 36(1), 52–78.
<https://doi.org/10.1080/07268602.2016.1109434>
- San Roque, L., Gawne, L., Hoenigman, D., Miller, J. C., Rumsey, A., Spronck, S., Carroll, A., & Evans, N. (2012). Getting the Story Straight: Language Fieldwork Using a Narrative Problem-Solving Task. *Language Documentation & Conservation*, 6, 135–174.
- Schütze, C. T. (2016). The empirical base of linguistics. In *Language Science Press*. Language Science Press. <https://doi.org/10.17169/langsci.b89.100>
- Seetaraso, T., Kutanan, W., Kampuansai, J., Muisuk, K., & Sri, M. (2020). Unique Genetic Structure of Y-chromosomal Lineage of the Moken from the Andaman Sea of Thailand. *Chiang Mai University Journal of Natural Sciences*, 19, 1066-1079. <https://doi.org/10.12982/CMUJNS.2020.0066>
- Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140–155.
<https://doi.org/10.1016/j.cognition.2014.06.013>
- Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.

- Shields, L. W., & Balota, D. A. (1991). Repetition and Associative Context Effects in Speech Production. *Language and Speech*, 34(1), 47–55.
<https://doi.org/10.1177/002383099103400103>
- Sidwell, P. (2005). Acehnese and the Aceh-Chamic Language Family. In A. Grant & P. Sidwell (Eds.), *Chamic and beyond: Studies in mainland Austronesian languages* (pp. 211-246). Pacific Linguistics.
- Simard, C., & Schultze-Berndt, E. (2011). Chapter 7. Documentary linguistics and prosodic evidence for the syntax of spoken language. In G. Haig, N. Nau, S. Schnell, & C. Wegner (Eds.), *Documenting Endangered Languages* (pp. 151–176). De Gruyter Mouton. <https://doi.org/10.1515/9783110260021.151>
- Simon, A.-C., & Christodoulides, G. (2016). Perception of Prosodic Boundaire By Naïve Listeners in French. *Proceedings Speech Prosody 2016*.
- Simpson, H. E. (2016). *The Role of Intonation Units in Memory for Spoken English* [Ph.D.]. University of California Santa Barbara.
- Skopeteas, S., & Fanselow, G. (2009). Effects of givenness and constraints on free word order. In M. Zimmermann & C. Féry (Eds.), *Information Structure: Theoretical, Typological, and Experimental Perspectives* (pp. 307-331). Oxford University Press.
- Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., & Krifka, M. (2006). Questionnaire on Information Structure (QUIS): Reference manual. In *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)* (Vol. 4). Universitätsverlag Potsdam.
<http://opus.kobv.de/ubp/volltexte/2007/1241/>

- Smalley, W. A. (1994). *Linguistic Diversity and National Unity: Language Ecology in Thailand*. University of Chicago Press.
- Smith, A. D. (2017). The Western Malayo-Polynesian problem. *Oceanic Linguistics*, 56(2), 435–489.
- Solntsev, V. (1996). Some remarks on the Ruc Language. *Mon-Khmer Studies*, 26, 29–32.
- Sopher, D. E. (1965). *The Sea Nomads: A study based on the literature of the maritime boat people of Southeast Asia* (Vol. 5). Memoirs of the National Museum.
- Spearman, H. R. (1880). Ethnology. In *British Burma Gazetteer* (Vol. 1, pp. 188–192). Union of Burma Press.
- Stalnaker, R. (2002). *Common ground*. <https://philpapers.org/rec/STACG>
- Strömquist, S., & Verhoeven, L. (Eds.). (2004). *Relating Events in Narrative, Volume 2: Typological and Contextual Perspectives*. Routledge.
- Swastham, P. (1982). *A description of Moklen: A Malayo-Polynesian language*. [MA]. Mahidol University.
- Tao, H. (1992). NP Intonation Units and Referent Identification. *Proceedings of the 18th Annual Meeting of the Berkeley Linguistics Society*, 237–247.
- Tao, H. (1996). *Units in Mandarin Conversation: Prosody, Discourse and Grammar* (Vol. 5). John Benjamins Publishing Company.
- Thach, N. M. (1999). Monosyllabization in Kieng Khmer. *Mon-Khmer Studies*, 29, 81–95.
- Thavisak, A. (2004). Fo behavior of vowels influenced by aspirated and unaspirated initials in Southeast Asian languages: Implications for tonogenesis theories. In S.

- Burusphat (Ed.), *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society* (pp. 721–730). Program for Southeast Asian Studies, Arizona State University. <http://purl.org/sealang/amon2004fo.pdf>
- Thies, T., Tanja, C., Baumann, S., & Grice, M. (2018). Prosodic Marking of Information Status in Picture Story Descriptions. *Proceedings of the Conference on Phonetics & Phonology in German-Speaking Countries (P&P 13)*, 193–196.
- Thompson, S. A. (1997). Discourse Motivations for the Core-Oblique Distinction as a Language Universal. *Directions in Functional Linguistics*, 59.
- Thurgood, G. (1999). *From Ancient Cham to Modern Dialects: Two thousand years of language contact and change*. University of Hawai‘i Press.
- Tomlin, R. (1986). *Basic Word Order: Functional Principles*. Croom Helm.
- Traugott, E. C. (2012). On the persistence of ambiguous linguistic contexts over time: Implications for corpus research on micro-changes. In *Corpus Linguistics and Variation in English* (pp. 231–246). Brill.
https://doi.org/10.1163/9789401207713_019
- Trón, V. (2008, June 26). On the Durational Reduction of Repeated Mentions: Recency and Speaker Effects. *Proceedings of the International Conference on Language Resources and Evaluation*. LREC 2008, Marrakech, Morocco.
- Vallauri, E. L., & Masia, V. (2018). Facilitating Automation in Sentence Processing: The Emergence of Topic and Presupposition in Human Communication. *Topoi*, 37(2), 343–354. <https://doi.org/10.1007/s11245-016-9417-9>
- Vallduví, E. (1993). *The Informational Component* [Ph.D.]. University of Pennsylvania.
- Vallduví, E. (2016). Information structure. In M. Aloni & P. Dekker (Eds.), *The*

- Cambridge Handbook of Formal Semantics* (pp. 728–755). Cambridge University Press. <https://doi.org/10.1017/CBO9781139236157.024>
- Vallduví, E., & Engdahl, E. (2013). The linguistic realization of information packaging. *Linguistics*, 51(s1). <https://doi.org/10.1515/ling-2013-0041>
- van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, 41(5), 825–849. <https://doi.org/10.1515/ling.2003.027>
- Vittrant, A., & Watkins, J. (2019). Guidelines for the description of Mainland Southeast Asian languages. In *Guidelines for the description of Mainland Southeast Asian languages* (pp. 653–686). De Gruyter Mouton. <https://doi.org/10.1515/9783110401981-015>
- von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, 154, 117–127. <https://doi.org/10.1016/j.pragma.2019.07.025>
- Wahl, A. (2015). Intonation Unit boundaries and the storage of bigrams: Evidence from bidirectional and directional association measures. *Review of Cognitive Linguistics*, 13(1), 191–219.
- White, W. G. (1922). *The Sea Gypsies of Malaya*. Seeley, Service & Co. Limited.
- Wilson, H. H. (1827). *Documents illustrative of the Burmese War with an Introductory Sketch of the events of the war and an appendix*.
- Wolff, J. U. (2010). *Proto-Austronesian phonology with glossary* (1–2). Southeast Asia Program Publications.
- Yarowsky, D. (1993). One sense per collocation. *Proceedings of the ARPA Workshop*

on Human Language Technology, 266–271.

Yoder, B. (2020). *A grammar of Abawiri, a Lakes Plain language of Papua, Indonesia*

[Ph.D. Dissertation, Univeristy of California Santa Barbara].





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME Daniel Peter Loss

DATE OF BIRTH February 14, 1984



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY