Title: A Comparison of Grammaticalization in Shan and Thai
Researcher: Daniel Peter Loss
Degree: Master of Arts in Linguistics
Advisor: Larin Adams, Ph.D.
Approval Date: 27 March 2017
Institution: Payap University, Chiang Mai, Thailand

The members of the thesis examination committee:

1. ___________________________ Committee Chair
   (Assistant Professor Pittayawat Pittayaporn, Ph.D.)

2. ___________________________ Committee Member
   (Larin Adams, Ph.D.)

3. ___________________________ Committee Member
   (Assistant Professor Phinnarat Akharawatthanakun, Ph.D.)
ACKNOWLEDGEMENTS

The knowledge, experience and various backgrounds of the staff and students of the Linguistics Department at Payap University creates an exciting environment for discussions about language. I am very appreciative to have had the opportunity to study and learn here. There are many people who deserve acknowledgement for their help and contribution to this thesis. Phinnarat Akharawatthanakun introduced me to the Shan language and provided extensive feedback to my work. As well as being my thesis advisor, Larin Adams’s approach to grammar is what originally attracted me to a grammar thesis, and it was in his grammar class where I first learned of the concept of grammaticalization. A very special thanks is needed for William Hanna who assisted me in the beginning and as I worked through the entire thesis. His knowledge and experience with Tai languages were an invaluable resource for conversation and exploration of the ideas within this thesis. Mark Post not only sparked an idea for the research methodology used in this thesis he was also gracious enough to share his Thai data with me. My journey into the Shan language was made much easier with the assistance of Kham Aung, Kham Myat, and Ajarn Sompong all who not only helped me find and interview Shan participants, but were also very patient and kind teachers. In writing this thesis I also was very fortunate to receive feedback on drafts from Patrick Fitzsimmons, Jacob Watson, Tyler Davis, and Thomas Robin. A final thanks must go to Nittaya Intapong whose help extends into both life and language.

Daniel Peter Loss
ABSTRACT

Grammaticalization, the process by which lexemes come to serve grammatical functions, provides many insights into the historical origins and processes of language. Among Southeast Asian languages, the Thai language has been the focus of much grammaticalization research. Yet, to better frame historical developments of grammar within Tai languages and of the nature of Proto-Tai, more information about grammaticalization in languages other than Standard Thai should also be considered. Inspired by previous research into Tai syntactic change and grammaticalization (Diller 2001) this study looked at two closely related Tai languages – Shan and Thai – with the aim of better understanding grammaticalization in each language. The starting point for this research was the extensive literature already existing on Thai grammaticalization from which the researcher compiled several grammaticalized morphemes, or “grams”, in various grammatical contexts for the purpose of comparison with Shan. This study also adopted an approach based on Post (2007) which entailed the use of two corpora, one for each language, for the purpose of comparing the relative frequencies of grams in each language. Use of this approach required the creation of a small topic-specific Shan corpus, made up of texts obtained through a video elicitation method. These texts provided data and several contexts for comparison of the two languages, first in generating many comparable examples and uses of language and second in also allowing frequency measurements of analogous grams to be done in each language’s corpus.
The findings of the study showed many similarities between the two languages. Grams coming from grammaticalization paths such as ‘to take’ verb > co-verb/manipulated object marker; ‘be.at’, verb > progressive aspectual marker; ‘place’ noun > relative clause marker; and ‘to communicate’ verb > verbal complement marker were used in both languages. Despite many similarities, the comparison also revealed many differences between Shan and Thai. Some of these differences included introduction of unique lexical material into common grammaticalization paths such as verbs ‘to give’ into benefactive and causative markers. More generally there were several differences in pre-verbal TAM markers, resulting from influence of areal factors (i.e. Burmese and Khmer). Functional expansion of verbs was also found to differ in the two languages, such as developments of verbs into prepositions, and verbs used in comparative constructions. General use of directional verbs as aspectual markers or “success markers” was similar in both languages, except for the Shan use of a verb ‘to pierce’ for aspectual marking of cognitive states. Other unique Shan grammaticalizations included cases such as verb ‘to meet’ > passive marker and the verb ‘to be diligent’ > adverb ‘often, frequent’. Additionally, differences in several nominalizers and other grams used in Shan noun-phrase syntax were found to maintain stronger lexical ties than the analogous Thai forms which are more “semantically bleached”. Taking into account all the individual similarities and differences reviewed above, this study discusses the possibility of Standard Thai being more deeply grammaticalized than Shan, and considers factors relating to areal influences and sociolinguistics that may have affected grammaticalization in Shan and Thai. It ends with some ideas and suggestions for more comparative grammatical research into other Tai languages.
บทคัดย่อ

การกลายเป็นคำไวยากรณ์เป็นกระบวนการที่ซับซ้อนและเป็นส่วนหนึ่งของกระบวนการที่ทำให้เราเข้าใจถึงลักษณะหรือด้านก่อกนิษฐาและกระบวนการทางภาษา ในบรรดาภาษาต่างๆ ในภูมิภาคเอเชีย-อาเซียน ภาษาไทยเป็นภาษาที่มีผู้สนใจที่มีสูงที่สุด อย่างไรก็ตาม การวิจารณ์กระบวนการทางไวยากรณ์ของกลุ่มภาษาไทยและธรรมชาติของภาษาที่มีกิจกรรมการกลายเป็นคำไวยากรณ์เป็นกระบวนการที่ซับซ้อน การกลายเป็นคำไวยากรณ์ของภาษาต่างๆ นอกเหนือจากภาษาไทยมาตรฐาน งานวิจัยเกี่ยวกับการเปลี่ยนแปลงทางไวยากรณ์และการกลายเป็นคำไวยากรณ์ในภาษาไทย (Diller 2001) เป็นแรงบันดาลใจให้ทั้งการวิจัยชิ้นนี้จูงใจให้ศึกษาภาษาไทยที่มีความใกล้ชิดกันของภาษา คือ ภาษาไทยใหญ่ และภาษาไทย โดยผู้วิจัยจึงมุ่งที่จะทำการวิจัยได้ผลในภาษาที่มีประโยชน์อย่างแพร่หลาย ซึ่งผู้วิจัยได้รวบรวมหน่วยคำไวยากรณ์ (grams) จำนวนหนึ่งที่พบในหลากหลายบริบททางไวยากรณ์ เพื่อนำมาเปรียบเทียบกับภาษาไทยผู้ศึกษานี้ได้ใช้วิธีการของ Post (2007) โดยมีขั้นตอนที่ชัดเจน แบ่งเป็นสองขั้นตอนที่ละเอียดอ่อน เพื่อจะเห็นถึงความสัมพันธ์กับภาษาที่มีประโยชน์อย่างแพร่หลาย ซึ่งผู้วิจัยได้รวบรวมหน่วยคำไวยากรณ์ (grams) ในแต่ละภาษา การใช้วิธีการนี้จำเป็นต้องใช้ข้อมูลหน่วยคำไวยากรณ์ที่ตรงกันได้ ซึ่งจะสร้างขึ้นจากตัวบัตรที่ได้จากการวิเคราะห์ที่ได้บันทึกไว้ ซึ่งเป็นแหล่งข้อมูลและทรัพยากรสำหรับการวิเคราะห์ภาษาทางไวยากรณ์ การวิจัยชิ้นนี้มุ่งที่จะนำเสนอการวิเคราะห์ของภาษาที่สามารถนำมาเปรียบเทียบกันได้ ประสบการณ์ที่สอน คือทำให้สามารถวิเคราะห์ความถี่ของหน่วยคำไวยากรณ์ (grams) ที่คล้ายคลึงกันในขั้นตอนที่ผลิตขึ้นได้
ผลการศึกษาพบว่าทั้งสองภาษามีลักษณะที่คล้ายคลึงกันในหลายประการ เช่น คำศัพท์ ‘เอา’ กลายเป็นคำกริยาช่วย/ภาคแสดงกรรมที่ถูกเปลี่ยนไป; คำศัพท์ ‘อยู่’ กลายเป็นภาคช่วยเป็นวงลักษณ์แบบกำหนดน้ำ; คำศัพท์ ‘ที่’ กลายเป็นภาคแสดงคุณนามประโยชน์; คำศัพท์ ‘ว่า’ กลายเป็นภาคแสดงส่วนขยายกริยา พบในทั้งสองภาษา นอกจากความคล้ายคลึงกันของทั้งสองภาษามั่นการศึกษาแสดงความแตกต่างบางประการของภาษาไทยและภาษไทด้วยซึ่งความแตกต่างดังกล่าวรวมถึงการเริ่มต้นด้วยคำ และนำไปสู่การกล่าวเป็นคำไวยากรณ์ที่อ้างเช่น คำศัพท์ ‘ใต้’ กลายเป็นภาคแสดงที่ให้ประโยชน์และเป็นต้นเหตุโดยเปรียบเทียบ ภาคแสดงกริยาสามารถมีความแตกต่างอยู่หลายประการซึ่งเป็นผลมาจากการเปลี่ยนแปลงของยุคสมัย การขยายหน้าที่ของคำศัพท์ในทั้งสองภาษามีความแตกต่างกันด้วย เช่นในการเปลี่ยนจากคำศัพท์เป็นคำสีบุพท์และคำกริยาที่ใช้ในรูปประโยคเรียกเหยื่อ การใช้กริยาบอกเหตุเพื่อเป็นภาคแสดงกรรมลักษณะหรือภาคแสดงการสิ้นสุด (success markers) โดยทั่วไปมีความคล้ายคลึงกันทั้งสองภาษา นอกจากการใช้คำกริยา ‘เหลว’ (to pierce) ในภาษาไทใหญ่ที่ถูกใช้เป็นภาคแสดงการเปลี่ยนแปลงในรูปประโยคแบบบริบท การเปลี่ยนเป็นคำไวยากรณ์ยังมีเอกลักษณ์ในภาษาไทยด้วยเช่น คำศัพท์ ‘พบ’ (to meet) ในภาษาไทใหญ่กลายเป็นข้อความและคำกริยาข้อความ ‘ขยัน’ (to be diligent)กลายเป็นกริยาช่วย ‘อยู่’ ข้อแตกต่างบางประการในทั้งสองภาษามีความแตกต่างกันของการเปลี่ยนแปลงของคำศัพท์ในภาษาไทใหญ่ที่มีความหลากหลายกว่าภาษาไทย ซึ่งเป็นการเปลี่ยน ‘การที่ความหมายจางลง’ (semantically bleached) มากกว่า หลังจากที่ได้รวบรวมมาเป็นข้อมูลแล้วพบว่าการเปลี่ยนแปลงของยุคสมัยและตัวแทนภาษาที่กล่าวม่ามีผลต่อการเปลี่ยนเป็นคำไวยากรณ์ในภาษาไทใหญ่และภาษาไทย งานวิจัยนี้ได้เห็นความติดบวกและข้อเสนอแนะในการศึกษาวิจัยเพื่อเปรียบเทียบลักษณะทางไวยากรณ์ของภาษาต่างๆ ในการใช้ภาษาไทยต่อไป
# TABLE OF CONTENTS

Acknowledgements ............................................................................................................ i
Abstract ............................................................................................................................... ii
บทคัดย่อ ............................................................................................................................ iv
List of Tables ....................................................................................................................... ix
List of Figures ...................................................................................................................... x
List of Abbreviations and Symbols ..................................................................................... xi
Chapter 1 Introduction ...................................................................................................... 1
  1.1 Introduction .................................................................................................................. 1
  1.2 Rationale and Objectives ............................................................................................ 2
  1.3 Methodology ............................................................................................................... 4
    1.3.1 The Elicitations .................................................................................................... 4
    1.3.2 The Corpus ......................................................................................................... 5
    1.3.3 Analysis .............................................................................................................. 7
  1.4 Limitations .................................................................................................................. 7
  1.5 Outline of the Thesis .................................................................................................. 8
Chapter 2 Thai, Shan and Thai: Languages and Peoples .................................................. 11
  2.1 Tai ............................................................................................................................... 11
  2.2 Shan ........................................................................................................................... 14
  2.3 Thai ............................................................................................................................ 15
  2.4 Phonology .................................................................................................................. 16
    2.4.1 Consonants ....................................................................................................... 16
    2.4.2 Vowels ............................................................................................................. 17
    2.4.3 Lexical Cognates and Tone .............................................................................. 18
  2.5 Basic Grammar and Ellipsis ....................................................................................... 23
    2.5.1 Nominal Modification and Classifier Phrases ................................................. 24
    2.5.2 Basic Clausal Syntax and Serial Verbs .......................................................... 26
  2.6 Discourse Reference Management .......................................................................... 27
Chapter 3 Literature Review ............................................................................................. 31
  3.1 Introduction ............................................................................................................... 31
  3.2 Grammaticalization Basics ....................................................................................... 32
    3.2.1 Grammaticalization: Theory and Examples .................................................. 33
LIST OF TABLES

Table 1 Shan speakers from corpus ................................................................. 6
Table 2 Shan consonants (Hudak, 2008) .......................................................... 16
Table 3 Thai consonants (Iwasaki & Ingkaphirom, 2005) .................................. 17
Table 4 Shan final consonants (Hudak, 2008) .................................................. 17
Table 5 Thai final consonants (Iwasaki & Ingkaphirom, 2005) ......................... 17
Table 6 Shan Vowels (Hudak, 2008) ................................................................. 18
Table 7 Thai Vowels (Iwasaki & Ingkaphirom, 2005) ....................................... 18
Table 8 Thai dipthongs (Iwasaki & Ingkaphirom, 2005) ................................... 18
Table 9 Shan tone box (Hudak, 2008) .............................................................. 22
Table 10 Thai Tone Box (Hudak, 2008) ............................................................ 22
Table 11 Shan Tones (Hudak, 2008) ................................................................. 22
Table 12 Thai Tones (Hudak, 2008) ................................................................. 23
Table 13 Percentage of cà vs. càk in Thai inscriptions (Diller, 2001) ................. 54
Table 14 Increasing modal functions of cà/càk (Takahashi, 2011) ................. 55
Table 15 Decreasing verbal functions of cà/càk (Takahashi, 2011) .............. 55
Table 16 Occurrences of nine functions of /thîi/ in the four periods ............... 56
Table 17 Thai Grammaticalized Morphemes .................................................. 58
Table 18 Comparison of different Thai and Shan TAM markers ................... 78
Table 19 Comparison of different Thai and Shan directional verbs ............... 83
Table 20 Summary comparison of Shan and Thai verb-phrase grams ............ 88
Table 21 Shan and Thai Nominalizers ............................................................ 96
Table 22 Comparison of /thîi/tii functions in corpora ................................... 103
Table 23 Possession in Shan and Thai corpora ............................................. 105
Table 24 Shan and Thai prepositions grammaticalized from nouns .............. 108
Table 25 Lexical Source of Classifiers (Iwasaki & Ingkaphirom, 2005) ......... 113
Table 26 Functions of classifiers in different constructions (Bisang, 2008) .... 114
Table 27 Shan and Thai narrative markers .................................................... 121
Table 28 Summary comparison of Shan and Thai noun-phrase grams .......... 124
Table 29 Summary comparison of “Miscellaneous” Shan and Thai grams .... 125
Table 30 Functions of most frequent Thai grams ........................................ 132
Table 31 Comparison of Tai Grams ............................................................... 136
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Grammaticalization of a lexeme to a gram</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Rationale and research question</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Major Tai languages in Southeast Asia</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Tai-Kadai Family Tree</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Southwestern Tai family (Hartman, 1980)</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>Chamberlain’s PSWT grouping</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>Grammaticalization of going to</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td>Layers of grammatical development (Heine &amp; Kuteva, 2007)</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>Grammaticalization path of thîi (adapted from Kullavanijaya, 2008)</td>
<td>57</td>
</tr>
<tr>
<td>10</td>
<td>Lexical/grammatical range of Tai ACQUIRE</td>
<td>66</td>
</tr>
<tr>
<td>11</td>
<td>Functions of Shan and Thai verb BE.AT</td>
<td>71</td>
</tr>
<tr>
<td>12</td>
<td>Development of obligative function of Shan ACQUIRE pre-verbally</td>
<td>77</td>
</tr>
<tr>
<td>13</td>
<td>Shan and Thai passive markers</td>
<td>86</td>
</tr>
<tr>
<td>14</td>
<td>Thai agent nominalizers</td>
<td>92</td>
</tr>
<tr>
<td>15</td>
<td>Shan and Thai adjectivilizers</td>
<td>99</td>
</tr>
<tr>
<td>16</td>
<td>Functions of Shan and Thai PLACE grams</td>
<td>104</td>
</tr>
<tr>
<td>17</td>
<td>Shan and Thai khɔɔŋ</td>
<td>106</td>
</tr>
<tr>
<td>18</td>
<td>Shan and Thai waa/wāa</td>
<td>108</td>
</tr>
<tr>
<td>19</td>
<td>Functional expansion of classifiers (Bisang, 1998)</td>
<td>114</td>
</tr>
<tr>
<td>20</td>
<td>Bi-polar distribution of lêew and kɔɔ</td>
<td>119</td>
</tr>
<tr>
<td>21</td>
<td>Thai grammaticalization paths</td>
<td>133</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS AND SYMBOLS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>First person singular pronoun</td>
</tr>
<tr>
<td>1PL</td>
<td>First person plural pronoun</td>
</tr>
<tr>
<td>2SG</td>
<td>Second person singular pronoun</td>
</tr>
<tr>
<td>3SG</td>
<td>Third person singular pronoun</td>
</tr>
<tr>
<td>ADJVZR</td>
<td>Adjectivilizer</td>
</tr>
<tr>
<td>ADV</td>
<td>Adversative</td>
</tr>
<tr>
<td>ASP</td>
<td>Aspect</td>
</tr>
<tr>
<td>AUX</td>
<td>Auxiliary</td>
</tr>
<tr>
<td>BEN</td>
<td>Benefactive</td>
</tr>
<tr>
<td>CAUS</td>
<td>Causitive</td>
</tr>
<tr>
<td>CLF</td>
<td>Classifier</td>
</tr>
<tr>
<td>COMPAR</td>
<td>Comparative</td>
</tr>
<tr>
<td>COM</td>
<td>Complementizer</td>
</tr>
<tr>
<td>CONN</td>
<td>Connective</td>
</tr>
<tr>
<td>CONT</td>
<td>Continuous</td>
</tr>
<tr>
<td>COP</td>
<td>Copular</td>
</tr>
<tr>
<td>DEM</td>
<td>Demonstrative</td>
</tr>
<tr>
<td>DV</td>
<td>Directional Verb</td>
</tr>
<tr>
<td>EMSEA</td>
<td>East and Mainland Southeast Asia</td>
</tr>
<tr>
<td>FUT</td>
<td>Future</td>
</tr>
<tr>
<td>INF</td>
<td>Infinitive</td>
</tr>
<tr>
<td>INTS</td>
<td>Intensifier</td>
</tr>
<tr>
<td>IPA</td>
<td>International Phonetic Alphabet</td>
</tr>
<tr>
<td>IRR</td>
<td>Irrealis</td>
</tr>
<tr>
<td>LINK</td>
<td>Linker</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

1.1 Introduction
This thesis is the result of an interest in two things: Tai languages and research of grammaticalization. The convergence of these areas has resulted in a large body of work done on grammaticalization in Standard Thai – the most researched of the Tai languages. Inspired from insights gained into the rise and evolution of Thai grammatical markers, this current work seeks to extend the explanatory power of previous Thai grammaticalization research by bringing data from a related language – Shan – into the discussion. While in essence this thesis is a comparative grammar, there is theoretical motivation driving each area and item of comparison, namely that of grammaticalization. The first obvious benefit of this type of research is it brings more attention to the Shan language – a lesser known Tai language in regards to grammaticalization, but additionally this thesis aims to contribute to cross-linguistic research and discussion on Tai languages, by answering calls for more research in this area by Diller (2001).

Research in grammaticalization – “a diachronic process in which a form used for one purpose is recruited for another, more grammatical, purpose” (Hurford, 2012, p. 274) – is primarily concerned with explaining the origin and development of grammatical forms. Since the late 20th century this field of linguistic inquiry has brought about “a re-examination of the nature of grammar itself” (Bybee, 2011, p. 11) to become a major topic in linguistics. Participating in the growth of this field, Thai linguistics now has an extensive body of literature dealing with grammaticalization, and where discussions in Thai historical linguistics once focused mostly on phonological change – historical syntax can now be included. A significant factor contributing to the breadth of Thai grammaticalization research is the existence of written data dating back to the 13th century and methods of investigation provided by approaches in corpus linguistics. Findings from research using Thai historical corpora have not only explained the historical development of Thai grammatical morphemes, but they have the potential to “unmask pseudo-problems” encountered by earlier linguistic theories, when dealing with Tai
languages (Diller, 2001, p. 139). For this reason the discussion on grammaticalization in Thai is ripe for inclusion of other related languages and dialects.

This thesis used findings from the literature on grammaticalization in Thai to form contexts of comparison with a related language – Shan. The comparison process began through direct elicitation with Shan-Thai bilingual speakers of Shan forms which corresponded to Thai grammaticalized morphemes. This information was used for further investigation of Shan texts and naturally occurring speech. Additionally, the creation of a small Shan topic-specific corpus was used to measure and compare rates of frequency for the Shan and Thai grammatical morphemes in question. The results of the comparisons include some expected similarities with Thai, as well as supporting evidence for previous work on Shan and Tai languages in general, but the results here also provide new information about Shan syntax which are possibly unique grammaticalizations within the Tai family, or at least divergent from Standard Thai. Overall, through the comparisons of all grammatical morphemes, the results here detail a clearer picture of how Shan and Thai differ in respect to high-frequency grammatical markers, but also highlight factors pertinent to each language’s historical development, as well as provide a context and questions for further research.

1.2 Rationale and Objectives
Essentially, grammaticalization describes the well-established process by which lexemes (e.g. nouns and verbs) change into function words or “grams” and collectively construct grammatical systems. Figure 1 represents this core concept with a horizontal arrow representing the diachronic shift and change in classification of item A, from a lexeme at A1 to a gram at A2.

\[ \text{A1} \xrightarrow{} \text{A2} \]

\text{lexeme} \hspace{1cm} \text{gram}

\text{Figure 1 Grammaticalization of a lexeme to a gram}

\[\text{1 An abbreviated form of “grammatical morphemes” from Bybee & Dahl (1989, p. 51)}\]
Research in grammaticalization involves identifying cases where developments like in that Figure 1 have taken place, as well as explaining the factors which gave rise to the change. Grammaticalization can also include a gram at A2 shifting even further into a more grammatical role, with more limitations on its use (e.g. affix). In identifying several cases of grammaticalizations for a language, a better understanding of the origin of whole grammatical systems can be gained (e.g. English auxiliary verbs coming from full verbs). For the Thai language a considerable amount of grammaticalization has been identified, more so than any other Tai language\(^2\). This leads to questions on the nature of grammatical markers in other Tai languages.

The rationale for this study is illustrated in Figure 2 which depicts the current state of knowledge on Thai grammaticalizations with known cases represented by items A-C. This representation of Thai grammaticalization points out that grams A2, B2, and C2 have been identified as coming from earlier lexical source concepts A1, B1, and C1. Since other Tai languages share a large amount of lexical content other closely related languages should be explored for similarities and differences in grammaticalization. Also pictured in Figure 2 is the question of whether or not similar developments in grammaticalization have been made in Shan. This inference is inspired by grammaticalization theory and factors of language typology that supports the rationale of a comparison of Thai grammaticalizations with Shan grammatical morphemes.

\[\text{Figure 2 Rationale and research question}\]

The comparison of Shan and Thai begins with a comparison of similar grams or lexemes, such as a case of comparing Thai A1 to Shan A1 with the question of whether or not Shan A1 also fulfills the grammaticalized role of Thai A2. In some cases it may be possible that Shan and Thai differ even in regards to the lexical

\(^2\) See Appendix A Thai Grammaticalizations Table
source concepts such as the representation of Shan’s lack of C1 in Figure 2. Considering factors such as thousands of shared lexical items and the regularity of grammaticalization clines (see §3.2.1) we can expect similar developments in Shan as there were in Thai, and where things differ then there are other factors which must have been in play. While Figure 2 is an over-simplification to some extent, the underlying logic of the approach in this thesis is that a gram to gram comparison has the potential to be revealing for Shan, Thai, and the history of Tai languages in general.

From the rationale outline above the the working objectives of this research were defined as the following:

1. Compare Shan grammatical morphemes and constructions with Thai grammaticalizations.
2. Use analogous grammatical morphemes ascertained from Objective (1) for comparison of grams in Shan and Thai texts and corpora.

1.3 Methodology
To compare Shan and Thai data was compiled from several sources. The initial process of eliciting Shan grams is detailed in §1.3.1. The creation of the Shan and Thai corpora is covered in §1.3.2. Several examples of Shan grammaticalizations in this thesis are presented through examples from previous work on Shan. Though the researcher was able to gather most of the analogous grammatical morphemes through the elicitations, if naturally occurring uses of grams were available, presentation of them was preferred over the elicited forms. Thus in total the Shan data within in this study is made up of three sources: elicited data, spoken texts compiled into a corpus, and other textual resources. The researcher analyzed the grammatical aspects of Shan with regard to analyses of similar Thai grams, and with the theoretical framework of grammaticalization in mind.

1.3.1 The Elicitations
To address the needs of Objective (1), information about specific grammatical markers in Shan was needed. First, the researcher compiled information from the body of literature on Thai grammaticalizations, which resulted in a compendium of Thai grammaticalization research (see Appendix A: Thai Grammaticalizations Table). This provided a basis for which grams or analogous grammatical functions
would be elicited in Shan. Along with the researcher's own knowledge of Thai, examples and constructions displaying specific functions of grammaticalized morphemes in Thai were also collected to help with the elicitation process with the Shan consultants. Language elicitations were then conducted with 3 male Shan speakers (ages 30-60) who were also bilingual in Thai. During the elicitations the Shan speakers were asked to produce equivalent sentences in Shan based off of Thai examples. During this time the researcher gathered data and noted similarities and differences between the languages. This allowed for a preliminary comparison of Thai grammaticalizations with Shan grammatical markers which provided targets of focus in the exploration of the Shan corpus and texts as outlined in Objective (2).

1.3.2 The Corpus
To create Thai and Shan corpora, texts were collected using a video elicitation method. The methodology in this thesis is modeled after Post (2007) who used compared grammaticalization in Thai and Chinese. The process begins by showing a 6 minute video, the “The Pear Story” to participants. The viewer of the film then relates what happened in the film to another native-speaker of the language who has not seen the film. The speakers were told that this information would be used for matters of studying of language and were asked to use normal everyday speech. This process was repeated with eight different speakers for each language, and audio recordings were made for each of the sessions. Each recording was then transcribed and glossed for entry into the corpus. The Thai texts used in this thesis come from Mark Post who personally allowed the researcher to reuse his data from Post (2007). The Thai corpus from Post (2007) was the result of 8 Thai speakers who produced spoken texts on the Pear Story. The mean time of the Thai speaker’s responses was 4:34, and the Thai corpus totaled 6,847 words. For this current thesis, the researcher recorded texts from 8 Shan speakers (3 female, and 5 male, age 25-49) who were all currently residing in Thailand. Table 1 provides information about the speakers who contributed to the Shan corpus.

3 The consultants were “Tai-Yai” speakers. See §2.2 for discussion of Shan varieties.
4 This video and methodology were specifically developed for linguistic investigation. See Chafe, W. (1980). The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production. Norwood, N.J., Ablex. or view online at http://pearstories.org/
The resulting texts had a mean time of 3:03 minutes, and the Shan corpus totaled 3,063 words (See an example text in Appendix B: Shan Pear Story). With the texts compiled into a corpus, target grammatical morphemes related to the objectives of the study could be examined and comparative levels of frequency of Shan and Thai grams could be analyzed, in line with the aims of Objective (2). Some background and theory on the use of corpus linguistics in the investigation of grammaticalization is discussed in §3.2.1.

A main idea for the use of frequency measurements is the idea that high frequency is an indicator of a grammatical morpheme (Bybee, 2003). The use of frequency measurements in this thesis is mainly seen as an aid to the direct elicitation of equivalent forms (some of the grams in this study did not appear in either corpus). Since comparisons between Shan and Thai in this thesis target analogous grammatical morphemes, text-frequency of grammatical items can be used to further explore possible differences between Shan and Thai (see Appendix C: Top 100 wordlist). By using overall frequency of grams a useful comparison of Shan and Thai could be supported. This is especially helpful where there were non-cognate forms of a gram. Limitation to this study’s use of corpus linguistics and frequency measurements are discussed in §1.4.
1.3.3 Analysis

Analysis of Shan grams within this study is mostly in relation to their similarity to the analogous Thai form. In doing this terminology used to describe a grammatical function in Thai may be applied to a Shan gram (e.g. irrealis marker). This is only done when it appears that the Shan gram functions within the same grammatical space as the analogous Thai form. Despite questions as to what counts as proof of grammaticalization (cf §1.4), a background assumption of this study is that “grammaticalization” in its broadest sense explains the development and maturation of all grammatical systems.

1.4 Limitations

Though this study is presented as a comparison of Shan and Thai readers should keep in mind the scope of the findings here in regards to speakers of languages sometimes categorized under the heading – and exonym – “Shan” (See §2.2). As the main assistants and speakers of the corpus texts were “Tai Long” (and have also resided in Thailand), data provided by them may not reflect uses throughout other Shan varieties. There is also the possibility that Shan speakers may have been susceptible to influence from Thai. This also pertains to the questions as to appropriateness of direct elicitation of Shan grammatical information. At most then this study’s findings should be considered exploratory in nature. Findings here should not be taken as absolutely representative of a specific gram’s function or of the Shan language(s) as a whole.

One main limitation of this thesis is that the status of any supposed Shan grammaticalizations here will require further historical research that shows bridging contexts, like that which has been done for cases of grammaticalization in Thai. The range of possible factors that have affected Shan can not be deduced just from a comparison with Thai, though findings from this study could be used for exploration of historical data. Grammaticalization is a diachronic process and is best supported by historical evidence. Fortunately the existence of Shan historical texts makes this an interesting possibility for future research. With analysis of Shan historical texts, many of the Shan grams identified here could be researched further.

Another limitation in this study is the quality and comparability of the corpora used in the study which could be improved both in size and in controls (e.g. background of the speakers or speech register). That being said the strength of the corpora was that its topic-specific nature created highly comparable uses and examples of
language. With only a small sample of language data within the corpora, no significant claims are solely based upon analysis of the data there.

As it is with any science, any of the hypotheses made within this study stand to be falsified and should be studied further. That being said, the researcher takes ultimate responsibility for all of the content and claims made within this study.

1.5 Outline of the Thesis

Chapter 2 begins with an introduction to the Tai language family before covering more specific background information on the Shan and Thai languages and the people who speak these languages. Some linguistic background for comparison of these languages is then laid out by reviewing the phonological inventory of each language in §2.2. The phonology of each language is particularly relevant in that differences here can account for differences on apparent cognate lexical items, and is especially relevant for the realization of tone on these similar lexical items.

Comparative phonology is also significant in that previous research on comparative Tai and much of the success of Proto-Tai reconstruction has been done from a phonological perspective. Section 2.3 Basic Grammar covers basic aspects of Shan and Thai syntax that are common to both languages this includes: nominal modification and uses of classifiers, basic clausal syntax and serial verbs, and discourse reference management and ellipsis. In addition to introducing the languages dealt with in this study, this chapter should assist in better understanding data presented in Chapters 4 and 5 which describes in detail more specific grammatical features of Shan and Thai.

Chapter 3: Literature Review is an attempt to bring together the relevant aspects for discussion concerning the matters in this study. As the motivation and context for comparison in this study comes from the perspective of grammaticalization, §3.2 first explains the theory and parameters of grammaticalization in some detail. Important terminology and the process of grammaticalization are shown using the example of the [be going to] > gonna grammaticalization in the English language. The focus on grammaticalization theory in general is finished off with §3.2.2 which outlines frameworks and methods common to grammaticalization research, namely corpus linguistics and construction grammar. Pointing out some relevant theoretical issues for Shan and Thai in regards to grammaticalization §3.3 reviews work of Bisang (1996, 2004, 2008, 2009, 2011, 2015) and Enfield (2001, 2002, 2003, 2005, 2011) which focus on particular characteristics of language change in Southeast.
Asian languages. Finally this chapter reviews the historical corpus approach used in identifying two cases of grammaticalization in Thai; the irrealis marker cà and the multi-functional marker thîi for the purpose of pointing out the methodology used to support many claims of grammaticalization for Thai grams covered in this thesis. This chapter should serve to orient discussions of the findings presented in the following chapters.

Findings from the comparisons of Shan and Thai as outlined in the objectives (see §1.3) are presented in Chapter 4 Verbal Phrase Syntax and Chapter 5 Noun Phrase Syntax and Miscellaneous. Each chapter is made up of sections each with their own basis for comparison – either a source lexical concept or an area of grammatical marking. Each section provides some context for the comparison by first reviewing literature on the respective gram or area of grammatical function. This includes a simple grammatical analysis, as well as some information from the perspective of grammaticalization. As there is more research on Thai, background information most often comes from the perspective of Thai language research, but insights or notes from a broader Tai perspective may be included where relevant. After the identification of a gram and its grammatical function(s), and illustration through examples of Thai and Shan, similarities and differences for the two languages are noted and summarized at the end of each section. Full analysis on the possibilities for each gram could be discussed further but is outside the scope of this paper. Instead the goal here is to what extent Thai grammaticalizations are shared by Shan, and where they are not, then the goal is to identify grams which accomplish the same or similar grammatical functions. A final analysis and interpretation of overall results of comparing Shan and Thai grammaticalizations is left for discussion in Chapter 6. As the data here comes from several different sources, for the sake of consistency, changes have been made to conform transcriptions to the phonologies of the language outlined for each language in §2.2 or with standard pronunciations. Language data are numbered and marked either with (S) for Shan or (T) for Thai to clearly indicate the source language. Data from the corpora is also marked with a two-letter abbreviation which refers to the name of the speaker (see Table 1). The author also takes responsibility for all translations which in some cases may have been changed or adjusted with data taken from other sources.

5 For Shan the SEAlang Library Shan Dictionary was used. This is an online source based off of Moeng (1995). See http://sealang.net/shan/dictionary.htm
Chapter 6 briefly reviews the study before presenting discussion on the findings of the study. After reviewing new information on Shan, the discussion moves to consider some possible broader conclusions for both Shan and Thai in regards to grammaticalization, but also in terms of their comparative levels of “grammatical depth” for the features looked at in this study. Attention is also brought to other factors besides grammaticalization that can account for the differences between Shan and Thai (i.e. sociolinguistic factors). Before concluding, some suggestions for further research on Tai languages and comparative grammaticalization research along with suggestions for the corpus methodology are mentioned.

Contained within the Appendices are additional resources for readers. Appendix A: Thai Grammaticalizations Table lists several Standard Thai grams with citations for grammaticalization research on them. Readers interested in Thai grammaticalization research or particular grams can consult many of the sources referred to within this table. Appendix B provides the transcript of a Shan Pear Story obtained through the process as outlined within §1.3.2. Readers who have not seen the film the Pear Story film can perhaps gain a better understanding of this text (and several of the examples provided in this study) by giving the film a viewing. Also from the methods as outlined §1.3.2, Appendix C provides the a sided by side comparison of the 100 most frequent words in both the Shan and Thai corpora.
Chapter 2
Thai, Shan and Thai: Languages and Peoples

2.1 Tai
Tai languages are a group of languages within the Tai-Kadai family. Speakers of Tai languages are widespread throughout Southeast Asia. With a total of 70 million speakers, they can be found in the countries of Burma, Cambodia, China, India, Laos, Malaysia, Thailand and Vietnam (see Figure 3). Typologically Tai languages are isolating, analytic languages with a S-V-O word order. Noun phrases are head-initial, and extensive serial verb constructions make up for a lack of verbal inflection or morphology. They are largely monosyllabic and share many phonological characteristics — the most noteworthy being a system of somewhere between 4-7 tones in each member language.

Modern Tai languages are thought to have all originated from an earlier form called Proto-Tai. Estimated to be no older than 2,000 years (Chamberlain, 1975), the descendants of Proto-Tai have reached their present locations through migrations from an ancestral homeland in southern China. Within the larger group of Tai-Kadai languages, the Tai node splits into three branches named after their relative geographic location: Northern, Central and Southwestern. Thai and Lao – the national languages of Thailand and Lao are both included in the Southwestern branch. Serious work on reconstruction of Proto-Tai and of sub-groupings Tai languages in the 20th century can be found in the work of Fang-Kuei Li (Li, 1977), and William J. Gedney (Hudak, 2008), but ongoing investigation into Proto-Tai still continues today (see Pittayaporn, 2009; Hanbo, 2016). Despite some challenges in naming and organizing these languages and their dialects, Figure 4 is presented as a working figure of the classification of Tai-Kadai languages.
Figure 3 Major Tai languages in Southeast Asia

Figure 4 Tai-Kadai Family Tree  
(adapted from Diller et al., 2008)

Under the Tai node a three-way distinction is made between Northern, Central, and Southwestern branches of Tai languages. This thesis focuses on two languages within the Southwestern branch: Shan and Thai. Figure 5 shows the relationships of languages within the Southwestern branch. The diagram here uses “Siamese” to refer to Thai.
2.2 Shan
Shan is the language of Shan State, Myanmar. It is in the Southwestern branch of the Tai family with 3,295,000 speakers (Lewis, 2015) of which approximately 150,000 are located in Northern Thailand (Jirattikorn, 2008, p.18). Shan is one of the largest national ethnic groups in Myanmar and one of the largest Tai speaking groups (after Thai, Lao, and Zhuang). Under the name Shan there are at least three distinguishable groups Tai Khamti, Southern Shan, and Northern Shan (Edmondson, 2008, p.203). Though outsiders typically use the name Shan, they usually refer to themselves as “Tai” with a secondary qualifying name that may come from the color of their costumes, or the places where they live, e.g. Tai Dang “Red Tai”, Tai Nua “Northern Tai” (Tun, 2009). Dates for the Tai migration into Myanmar go back as far as the 12th century CE, when the Bamar called these people “Shan”. The name seeming to be a corruption of “Siam”, with the lists of “Syam” slaves occurring in inscriptions dating back to 1120 CE in the ancient city of Bagan (Mong, 2004).

During British colonial rule of Myanmar in the late 19th century, writings about Shan peoples started to be published for western audiences. The earliest work on the Shan language starts with Rev. J. N. Cushing who published Grammar of the Shan Language (Cushing, 1887) and Shan-English dictionary (Cushing, 1888). His work presents the Shan language in an older writing system. Early Shan scripts are
thought to have developed from Burmese, yet due to natural geographic barriers various Shan groups ended up developing different scripts. As recent as 1975, Shan scripts have continued to go through various reforms (Mong, 2004). But today the script based on the dialect of the “Tai Long” who inhabit southern Shan state and make up the majority of Shan people has become a standard vernacular for different Shan groups. It can be found used in popular media and music as well as for purposes of promoting a national Shan identity (Jirattikorn, 2008, p.9).

2.3 Thai
Thai is the national language of Thailand. In total Thai speakers number over 60 million (Lewis, 2015). It is spoken as a second language by many speakers of other Tai-Kadai languages. The terms “Standard Thai” or “Central Thai” (or even “Siamese”) refer to the dialect spoken in the central area of Thailand that centers around the capital Bangkok, and is used to distinguish the language used in education, mass media, and by the middle class from other regional dialects in Thailand. Through media, education, and contact with native speakers, Thai has a very significant influence on other Tai dialects and varieties.

The Thai language is said to have begun in the 13th century when Tai people gained control of the Chao Phraya basin, and established kingdoms in areas previously under Burmese and Khmer control (Baker & Phongpaichit, 2014, p.4). This is also around the time when the Thai writing system, which is based off an Indic alphabet, began to be used. Most of the original Thai lexicon is monosyllabic but it has a long history of lexical borrowing with a high percentage of polysyllabic words coming from Sanskrit and Pali (Smyth, 2002, p.1). These spread to Thailand largely due to the religious influence of Buddhism. Thai is also heavily influenced by the borrowing of Khmer during Thailand’s early history when Khmer religious, cultural, linguistic and literary sources were prominent. Other languages that have influenced Thai include English and Chinese (Premsrirat, 2006).

The sheer number of Thai speakers as well as Thailand’s own academic community are two contributing factors to Standard Thai being the most researched of all Tai languages. Commonly in investigating the development of modern Thai, language researchers have made use of historical documents and ancient inscriptions. Research in this area divides the language based on Thai historical periods, with the

---

6 “Tai Yai” as they are known within Thailand. The abundance and variety of exonyms for Tai peoples can cause some confusion for ethnic and linguistic classification. For discussion of these factors see Keyes (1995) *Who are the Tai? Reflections on the Invention of Identities.*
Sukhothai period lasting from the 13th to 15th centuries, Ayutthaya the 15th to 18th, Ratanakosin the 18th to early 20th and the Modern era the early 20th to present day. Each of these periods having a considerable corpus of texts based off of writing materials available for use in historical linguistic research.

2.4 Phonology

This section compares basic features of Shan and Thai phonology. Information on Shan phonology here is based off of William Gedney’s analysis from Hudak (2008), while information on Thai was taken from Iwasaki & Ingkaphirom’s (2005) Reference Grammar of Thai. This section is organized in order to allow readers to compare phonological features of each language. First an inventory of phonemes for both Shan and Thai is presented. Then cases of phonological variation between the languages is presented for two purposes, one to show examples of phonological differences of cognate lexical items, and two to introduce the implications these differences have for each language’s tonal system. This leads to an analysis of each language by way of Gedney’s “tone box”, a major tool used in the reconstruction of Proto-Tai. Finally a list of the contrasting tones in each language with their corresponding, contour and diacritic marking is also provided.

2.4.1 Consonants

In Shan, there are 18 consonant phonemes. Some consonant clusters are possible with a liquid /l/ or semi-vowels /w, j/.

Table 2 Shan consonants (Hudak, 2008)

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p, pʰ</td>
<td>t, tʰ</td>
<td>k, kʰ</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s, sʰ</td>
<td></td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ɲ</td>
<td>ɲ</td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glide</td>
<td>w</td>
<td></td>
<td></td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

In Thai there are 21 consonant phonemes. Some consonant clusters are possible with the liquids /l, r/ or semi-vowel /w/.
Table 3 Thai consonants (Iwasaki & Ingkaphirom, 2005)

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p, pʰ, b</td>
<td>t, tʰ, d</td>
<td>c, cʰ</td>
<td>k, kʰ</td>
<td>?</td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>s</td>
<td></td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>η</td>
</tr>
<tr>
<td>Trill</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glide</td>
<td>w</td>
<td></td>
<td></td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

Final consonants for both Shan and Thai are nearly identical. Though Lengtai (2009, p.42) interprets Shan final glottals as a predictable feature occurring with low tone and falling tone in open syllables.

Table 4 Shan final consonants (Hudak, 2008)

<table>
<thead>
<tr>
<th></th>
<th>Stop</th>
<th>Nasal</th>
<th>Glide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>-p</td>
<td>-m</td>
<td>-w</td>
</tr>
<tr>
<td>Nasal</td>
<td>-t</td>
<td>-n</td>
<td>-j</td>
</tr>
<tr>
<td>Glide</td>
<td>-k</td>
<td>-η</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Thai Final Consonants (Iwasaki & Ingkaphirom, 2005)

<table>
<thead>
<tr>
<th></th>
<th>Stop</th>
<th>Nasal</th>
<th>Glide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>-p</td>
<td>-m</td>
<td>-w</td>
</tr>
<tr>
<td>Nasal</td>
<td>-t</td>
<td>-n</td>
<td>-j</td>
</tr>
<tr>
<td>Glide</td>
<td>-k</td>
<td>-η</td>
<td></td>
</tr>
</tbody>
</table>

2.4.2 Vowels

Thai has a much bigger vowel inventory than Shan. For every Thai vowel there is a contrasting form for length. Thai also has more diphthongs than Shan’s singular diphthong [aw].
### Table 6 Shan Vowels (Hudak, 2008)

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td>ə</td>
<td>o</td>
</tr>
<tr>
<td>Low</td>
<td>e</td>
<td>a, aː</td>
<td>ɔ</td>
</tr>
<tr>
<td>Diphthong</td>
<td></td>
<td>au</td>
<td></td>
</tr>
</tbody>
</table>

### Table 7 Thai Vowels (Iwasaki & Ingkaphirom, 2005)

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i, iː</td>
<td>u, uː</td>
<td>u, uː</td>
</tr>
<tr>
<td>Mid</td>
<td>e, eː</td>
<td>ə, əː</td>
<td>o, oː</td>
</tr>
<tr>
<td>Low</td>
<td>e, eː</td>
<td>a, aː</td>
<td>ɔ, ɔː</td>
</tr>
</tbody>
</table>

### Table 8 Thai diphthongs (Iwasaki & Ingkaphirom, 2005)

<table>
<thead>
<tr>
<th>ia</th>
<th>ua</th>
<th>ua</th>
</tr>
</thead>
</table>

### 2.4.3 Lexical Cognates and Tone

From Li’s (1977) work we know that Tai languages share thousands of lexical cognates. These are words still shared by many Tai languages, yet they often differ in regards to at least one phonological feature — either tone or initial consonant. The following are some examples of regular initial consonant differences between Shan and Thai – which have important implications for each language’s tonal system. This is presented along with the reconstructed Proto-Tai forms from Pittayaporn (2009).

A characteristic of the wider Shan community is a change in initial consonants from /t/ → /pʰ/. Edmondson (2008, p.200) claims the change from original Tai /t/ to /pʰ/ occurred early in Shan history. This difference results in a predictable pattern of initial consonant differences between Shan and Thai.
Proto-Tai: hwan⁴ ‘to dream’ ɕ.wun⁴ ‘rain’  
Shan: pʰān ‘to dream’ pʰən ‘rain’  
Thai: fān ‘to dream’ fōn ‘rain’

A word with a voiced bilabial plosive /b/ in Central Thai may appear with an approximant /w/ in Shan (this appears to be the case with Southern Shan only with some other varieties using /m/ instead). With the word ‘leaf’ you can also see the difference in vowels between the two languages. The /au/ diphthong was once a part of Thai but all /au/ forms moved to [aj], a remnant of this can be seen in the Thai script still having two symbols for vowels /aj/ ā and ė.

Proto-Tai: bîl⁴ ‘to fly’, bauq⁴ ‘leaf’  
Shan: wîn ‘to fly’, wau ‘leaf’  
Thai: bin ‘to fly’, baj ‘leaf’

These next examples are important in that they not only show a difference in initial consonant but they also show how this affects tonal contour. The first example of this can be seen with a contrast between /l/ and /d/. Shan lacks voiced plosives so cognate words that have an initial /d/ in Thai can be expected to take /l/ as their initial consonant in Shan.

Proto-Tai: dîj⁴ ‘to be good’ ɕ.dam⁴ ‘black’  
Shan: lî ‘to be good’, lām ‘black’  
Thai: dî ‘to be good’ dam ‘black’

Another example of variation is Thai words that begin with a liquid /r/. Shan instead uses fricative /h/ which is also the case with other Tai languages such as Northern Thai. The difference in tonal contour should also be noted.

Proto-Tai: rym⁹ ‘shade’ ruu ‘to know’  
Shan: hom ‘shade’ hû ‘to know’  
Thai: rôm ‘shade’ rûu ‘to know’

While knowing these variations might be useful for speakers of one language learning another, it is also significant for the purpose of grouping Tai languages. An example of this can be seen in the work of Chamberlain (1975) who used the
devoicing shift of Proto-Tai to either aspirated and unaspirated stops to identify a branch within Southwestern Tai languages, as shown in Figure 6.

![Diagram showing the PSWT grouping of Tai languages](image)

**Figure 6 Chamberlain’s PSWT grouping**
(adapted from Chamberlain, 1975)

These differences of initial consonants begin to play a larger role for the study and classification of Tai when we consider their impact on tone. This is not to say differences in tone always co-occur with a different initial consonant as shown in the examples of ‘rice’ and ‘eye’ below.

- Proto-Tai: \( \text{C}.qaw^{e} \) ‘rice’, \( \text{p.ta}a^{A} \) ‘eye’
- Shan: \( k^{h}a^{w} \) ‘rice’, \( t^{a}a \) ‘eye’
- Thai: \( k^{h}d^{a}w \) ‘rice’, \( t^{a}a \) ‘eye’

Yet the differences relate more to how phonological features are applied throughout each Tai language’s phonological system. This is related to the process of tonogenesis where changes in phonemes lead to the expansion or development of tones and tonal systems (Matisoff, 1973). The variety and range of tones in modern Tai languages had their origin in an earlier system. The earliest system that can be reconstructed had three tones, but as time and distance separated each Tai speaking community aspects of their phonology shifted and split more and more. Current research on the development of each phonological system, makes use of shared lexical items, as the basis for the Proto-Tai reconstruction. A major tool in this
research is Gedney’s “tone box” which is used in analyzing each language’s system of tones as well as grouping or classifying dialects and languages. The underlying logic of the tone box is explained by Gedney,

“When we come to consider initial consonants and tones (of Tai languages), the picture becomes more complex. Comparison of any two or more Tai dialects shows a complicated correlation between initials and tones from which scholars have been able to make the following inferences. First, the parent language had a system of three tones (sometimes called A, B, and C) on smooth syllables, and no tonal distinction (sometimes marked on D) on checked syllables. Second at some time after the break-up of the parent speech each daughter language or dialect underwent a number of sound changes involving tonal splits conditioned by the phonetic nature of the initial consonants at the time of the splitting... Reconstruction of these changes and of the system that existed before is possible only because each language or dialect made different tonal splits, and is aided by the fact that where subsequent changes (for example, in initial consonants) took place, the changes differed from one dialect to another” Gedney (1967, pp.17-19)

With “tone boxes” the tonal system of Tai languages or even smaller dialects or individual speech communities can be classified by how tone is realized on groups of Proto-Tai lexical cognates. Tone boxes form one of the main tools for describing Tai languages as presented in Li (1977) or Hudak (2008). Tone boxes for Shan and Thai in Table 9 and Table 10 show the pattern of tonal splits for each language. Each row corresponds to a class of initial consonants as they were in Proto-Tai. The columns A-C represent words with open syllables, and the last two D columns are for checked syllables ending with either /p/, /t/, /k/, or /ʔ/ with an additional distinction made for vowel length either short or long. Inside the tone box, the contour of the tone is given in a numerical form (e.g. 332, 41, etc.) and where it occurs is noted. This leads to areas of tonal split or merge and determines the tonal inventory of the language.
Table 9 Shan tone box (Hudak, 2008)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D-short</th>
<th>D-long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless fricatives</td>
<td>35</td>
<td>21</td>
<td>33</td>
<td>55</td>
<td>21</td>
</tr>
<tr>
<td>(Tone 1)</td>
<td>(Tone 2)</td>
<td>(Tone 3)</td>
<td>(Tone 4)</td>
<td>(Tone 2)</td>
<td></td>
</tr>
<tr>
<td>Voiceless unaspirated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stops</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glottal sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced sounds</td>
<td>55</td>
<td>33</td>
<td>53</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>(Tone 4)</td>
<td>(Tone 3)</td>
<td>(Tone 5)</td>
<td>(Tone 3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10 Thai Tone Box (Hudak, 2008)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D-short</th>
<th>D-long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless fricatives</td>
<td>24</td>
<td>22</td>
<td>41</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>(Tone 5)</td>
<td>(Tone 2)</td>
<td>(Tone 3)</td>
<td>(Tone 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless unaspirated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stops</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glottal sounds</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Tone 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced sounds</td>
<td></td>
<td>41</td>
<td>55</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>(Tone 3)</td>
<td>(Tone 4)</td>
<td>(Tone 3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Seeing the differences in how each language maps on to the tone box provides a basis for contrast, for example the lack of a tonal split for vowel length in Thai for checked syllables (columns DS and DL). The differences between Shan and Thai can also be seen by looking at the contour of individual tones, where other tones merge and split, along with the number of tones in total. From these analyses we see that Shan and Thai have different systems of five contrasting tones, as listed in Tables 11 and Table 12. Transcriptions of Shan and Thai data within this study make use of the system of diacritic marking as outlined in these two tables.

Table 11 Shan Tones (Hudak, 2008)

<table>
<thead>
<tr>
<th>Name/contour</th>
<th>Diacritic</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st tone: Mid-Rising</td>
<td>`</td>
<td>pû ‘crab’</td>
</tr>
<tr>
<td>2nd tone: Low-Falling</td>
<td>`</td>
<td>pù ‘grandfather’</td>
</tr>
<tr>
<td>3rd tone: Mid Level</td>
<td>no marking</td>
<td>pu ‘carpenter bee’</td>
</tr>
<tr>
<td>4th tone: High Level</td>
<td>`</td>
<td>pû ‘white mouse’</td>
</tr>
<tr>
<td>5th tone: High-Falling</td>
<td>`</td>
<td>pû ‘betel leaf’</td>
</tr>
</tbody>
</table>
### Table 12 Thai Tones (Hudak, 2008)

<table>
<thead>
<tr>
<th>Name/contour</th>
<th>Diacritic</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st tone: Mid Level</td>
<td>no marking</td>
<td><em>khaa</em> ‘to be unsettled’</td>
</tr>
<tr>
<td>2nd tone: Low Level</td>
<td>ʼ</td>
<td><em>khàa</em> ‘galangal root’</td>
</tr>
<tr>
<td>3rd tone: High-Falling</td>
<td>̂</td>
<td><em>khâa</em> ‘to kill’</td>
</tr>
<tr>
<td>4th tone: High Level</td>
<td>́</td>
<td><em>kháa</em> ‘to trade’</td>
</tr>
<tr>
<td>5th tone: Low-Rising</td>
<td>̌</td>
<td><em>khǎa</em> ‘leg’</td>
</tr>
</tbody>
</table>

#### 2.5 Basic Grammar and Ellipsis

Shan and Thai share much in terms of grammatical structure. Word order is SVO, noun phrases are head-initial, and both use chains of serial verbs. Like other Tai languages a system of pronouns in each language covers a large web of social relationships. Both languages construct numeral classifier phrases in the same order, but also make use of classifiers for several other functions. Other features of these languages include the reduplication of morphemes for semantic effect, and four-syllable forms known as elaborate expressions used for poetic effect.

As their proximity on the Tai family tree would suggest, it is not surprising to see the many ways in which Shan and Thai are similar. Those who are knowledgeable with either of these languages would find the other one’s grammatical structure quite familiar. One regular feature is the omission of arguments when pragmatically unnecessary. This is called ‘ellipsis’ or ‘zero anaphora’. This lack of overt marking may present a challenge to those first encountering this phenomena. The goal of this section then is to provide a basic overview of basic grammatical structures while also pointing out the significance of “abbreviated” or ellipsised constructions throughout Shan and Thai discourse. The bulk of data within this study comes from informal spoken texts. Language examples shown throughout this thesis may place ellipsed arguments into parentheses in the translation line. It is hoped that by covering this feature of both Shan and Thai, it will help readers to understand the data shown in later sections. The rest of this thesis is very much a comparative grammar and many more aspects of Shan and Thai grammar are shown in more detail within Chapter 4.
2.5.1 Nominal Modification and Classifier Phrases

Noun phrases (NP) have a head-initial structure following the pattern NOUN + MODIFIER. A head noun modified by an adjective phrase will follow the schema \([N_{\text{head}} \text{AP}]_{\text{NP}}\), such as in (1) where the ‘house’ is followed by an adjective phrase.

1. (S) ɯ́n jàu na
   house big INTS

   (T) bāan jāj mãak
   house big INTS
   ‘very big house’

Other types of nominal modification follow the head-initial pattern. This includes modification signifying possession, determiners, quantification, and relativization. Classifiers are words that can be used to numerate nouns. Prototypical Shan and Thai numeral classifier phrases are structured \([N + \text{[NUM+CLF]}]\), like in (2).

2. (S) mãa sām tó
   dog three CLF

   (T) mãa sām tua
   dog three CLF
   ‘three dogs’

Notice that the classifier in (2) above is different from the following example (3). Different nouns require specific classifiers based on categories such as animacy, human or non-human, or shape and function. Also, a numeral classifier phrase can come after other modifications of the head-noun like the adjectives in (3) where it is ‘beautiful women’ who are numerated.

---

3.  
(S) jíŋ haaŋ-lǐ sām kş
    woman beautiful three CLF

(T) jíŋ sǔaj sāam khon
    woman beautiful three CLF

‘three beautiful women’

Besides indicating number, classifiers can fulfill other grammatical functions (see Morev, 2000). Classifiers can be used along with demonstratives in the form [(N) CLF DEM], for anaphoric reference. The head-noun of a noun-phrase is optional if the referent head-noun is already understood from the context. An example situation could be a discussion of houses and then one speaker points out a specific house among a group of others. In that case the head-noun ‘house’ here could be omitted, and the typical classifier lǎŋ acts as the head.

4.  
(S) (hón)  lǎŋ nân
    house CLF that

(T) (bâan) lǎŋ nân
    house CLF that

‘that house’

It is also possible for the classifier in this construction to act as a head and receive the modification of the ellipsed head-noun. Continuing from the context in example (4) above it the intended house could be distinguished further with the accompanying adjective phrase indicating its color, like in (5).

5.  
(S) (hón)  lǎŋ sǐ lèŋ nân
    house CLF color red that

(T) (bâan) lǎŋ sǐ dê̂ŋ nân
    house CLF color red that

‘that red house’
2.5.2 Basic Clausal Syntax and Serial Verbs

Technically classified as SVO, an example like the following shows that whole grammatical arguments can be left out when they are contextually retrievable. In actual speech, even just a single verb can suffice as a complete sentence, such as in the response to a question like in (6) meaning approximately “Do you want to eat a banana?”. The answer provided for both Shan and Thai in (6) show that the omitted arguments within parentheses are not grammatically (or pragmatically) necessary.

6.

(S) kǐn koj kɔ̃ (háw) kǐn (koj)

| eat | banana TAG | 1SG | eat | banana |

(T) kin klûaj mǎj (chǎn) kin (klûaj)

| eat | banana TAG | 1SG | eat | banana |

Literally: ‘Eat banana? Eat.’

Main verbs can receive both pre-modification (e.g. negation, TAM) and post-modification by a small group of auxiliary verbs. An interesting phenomenon in both of these languages, however is that of serial-verb constructions: sequences of verbs with shared semantic arguments. These multi-verbs constructions are grammatically allowed in part by Tai languages’ ability to leave out whole syntactic arguments. This grammatical possibility creates Thai’s appearance as a verb loving language. Illustrating this fact by way of contrast with English, Diller (2006, p.175) says

“From the English point of view, informal Thai discourse often seems to provide an overload of verb-based detail with a corresponding dearth of nominal identification. This verb-dominant impression is quantitatively substantiated in a controlled analysis of parallel Thai and English narrations of a constant set of events (Iwasaki, 1989). The study shows that, using a comparable definition of utterance unit, in informal oral narrations Thai speakers use multiple verbs per unit with about twice (the frequency) as English narrators.”

This phenomena can be extended to Tai languages in general. The Shan example (7) shows a fairly typical serial verb construction featuring 3-5 multiple verbs within a sentence, bolded in the example below.
7. (S) [Pham, 2006, p.125]
   hários kɔ̀ kój lèn khun má há kón czj
1PL two CLF alone run ascend come seek people help
   ‘We (the two of us) ran up looking for someone to help’

This next extraordinary Thai example in (8) from Ratanakul (2007) shows an impressive 11 verbs.

8. (T) [Ratanakul, 2007, p.220]
   khários täŋcaj dèn paj cât háa súuu maa kēp wáj cháj
3SG intend walk go arrange search buy come keep put.in.place use
   háj sanùk
give fun
   ‘He intended to go searching for (that item) in order to keep (it) for usage that will bring enjoyment (to him).’

Isolated sentences with such extensive chaining of verbs, as in this previous example, may seem to leave listeners with quite a lot of interpretive work but as Smyth (2002) points out, the meaning of an a sentence such as example (9) becomes much more apparent when we restore the omitted arguments to which these verbs refer, like in (10).

9. (T) [Smyth, 2002, p.82]
   tɔ̀ŋ rîp klàp paj rîak háj maa bòk
must hurry return go summon cause come tell

10. (T) [Smyth, 2002, p.82]
    khùn tɔ̀ŋ rîp klàp paj rîak khários hâj maa bòk cháj
2SG must hurry return go summon 3SG cause come tell 1SG
    ‘You must hurry back and summon him to come tell me.’

Several common verbs within Shan and Thai serial-verb constructions are of particular interest when discussing grammaticalization many of which are discussed more fully in §4.2.

2.6 Discourse Reference Management
As the previous section pointed out with ellipsis, a lack of overt grammatical marking is a common feature in Shan and Thai. The lack of marking here is
accomplished in part by discourse-level management of information. These languages can be considered ‘topic-prominent’ languages (Iwasaki & Ingkaphirom, 2005, p.359), and a common way of introducing new items into discourse is a “presentational verb-first construction” (Enfield, 2007 p.157). These are essentially existential clauses using a verb ‘to have’. In more informal contexts, after an introduction, arguments can be referred to by demonstratives, pronouns, classifiers or even omitted as long as they remain accessible from the context. In the following example a Shan speaker begins reporting on the content of the Pear film (see §1.3). After their explicit introduction into discourse, whole arguments (e.g. the man and the fruit) are omitted from following sentences. Several other grammatical features from this section can also be seen here. In (11) the speaker begins setting the scene with the presentational verb-first construction. A null sign ‘ø’ is used to mark the dropped arguments.

11.  (S) [Corpus:OM]

\[\begin{array}{llllll}
\text{CONN} & \text{IRR} & \text{mis} & \text{lung} & \text{kon} & \text{num} \ 
\text{kow} \ ... \ kow & \text{ø} (\text{He}) & \text{pit} \\
\text{CONN} & \text{IRR} & \text{hav} & \text{uncle} & \text{people} & \text{1} & \text{PRT} & \text{CONN} & \text{pluck} \\
\text{maakmaj} & \text{waj} & \text{taj} & \text{njo} & \text{ø} (\text{tree}) \\
\text{fruit} & \text{POT} & \text{side} & \text{on} \\
\end{array}\]

‘There was an older man … and (He) was picking fruit on top (of a tree)’

The speaker first introduces an agent lung ‘uncle’, with a noun-phrase structured [N+CLF+NUM]. While indicating a quantity of one, the effect here also points out the definiteness of this particular agent who continues to act in the following sentence. This is a different from the typical numerical classifier phrase [N+NUM+CLF] given earlier in example (3) which mainly indicates numerical quantity. Next in example (12) the agent appears again through anaphoric reference with the pronoun mán ‘3SG’ in the clausal object position, which maintains the reference for the next sentence with an omitted subject.

12.  (S) [Corpus:OM]

\[\begin{array}{llllllllll}
\text{CONWK} & \text{IRR} & \text{mi} & \text{sam} & \text{pet} & \text{tam-waj} & \text{taj} & \text{mán} \ ... \ \text{maakmaj} \\
\text{CONN} & \text{IRR} & \text{have} & \text{3} & \text{basket} & \text{below} & \text{side} & \text{beneath} & \text{3SG} & \text{fruit} \\
\text{naj} & \text{ø} (\text{He}) & \text{saj} & \text{ø} (\text{fruit}) & \text{waj} & \text{pet} \ ... \ \text{pet} & \text{paj} & \text{ʔaw} & \text{kha} \ ... \\
\text{this} & \text{put} & \text{keep} & \text{basket} & \text{basket} & \text{remain} & \text{take} & \text{PRT} \\
\end{array}\]

‘and there were 3 baskets below him … this fruit (He’s) putting (fruit) in the baskets. The baskets haven’t been taken yet’
Continuing in example (13) the speaker is able to omit three arguments ‘the man’, ‘the tree’ and ‘the fruit’. Then a new entity, ‘a man leading a goat’ is introduced through the presentational first construction. Again the noun phrase with the modification, with the classifier $kɔ$ being for ‘person’, again with the number final construction [$N + CLF + NUM$].

13. (S) [Corpus:OM]

\[
\begin{align*}
kámnâj & \quad k\circ \quad \emptyset (He) \quad khun \quad khun \quad \emptyset (tree) \quad kwà \quad hâp \quad \emptyset (fruit) \\
\text{at this time CONN ascend ascend go grab} \\
\text{then} & \quad kám \quad nuŋ \quad k\circ \quad tê \quad mí \quad kón \quad cuŋ \quad pê \quad kɔ \\
\text{another/again moment one CONN IRR have people pull goat CLF} \\
uuŋ & \quad pʰât \quad kwà \\
\text{one pass go} \\
\text{‘... at this time (He’s) going up (the tree) to grab (fruit) again one more time} \\
\text{... then there’s a man leading a goat who passes by...’}
\end{align*}
\]

Similarities in discourse reference management can be seen in the next examples taken from another Thai Pear Story elicitation. The Thai speaker also began their report on the film with the presentational verb-first construction. Notice here too after already introducing the agent ‘the gardener’ with the same construction [$N + CLF + NUM$], it is omitted in a following sentence after the clear break of the linking word $lɛ́ɛ\ wkɔ̂\ ɔ$. Because the reference here is still retrievable and it does not need to be marked.

14. (T) [Corpus:TG]

\[
\begin{align*}
ton\ cháaw & \quad wan \quad nuŋ \quad mii \quad \ldots \quad khon \quad súan \quad khon \quad nuŋ \quad , \quad kamlaŋ \\
\text{time morning day one have person garden person one CONT} \\
kèp & \quad phõnlâmdâj \quad jiuu \quad lécwkɔ \quad \emptyset (He) \quad tɔŋ \quad \ldots \quad pûn \quad khùn \quad paj \\
\text{gather fruit PROG LINK must climb ascend go} \\
bôn & \quad tônmâj \quad \ldots \\
\text{on tree}
\end{align*}
\]

‘One morning there was ... one gardener gathering fruit and (He’s) going up on a tree’
Example (15) continues by first repeating the previous sentence *piin khǔn paj bon tōnmāaj* to frame the context for new information, which this time refers to the agent ‘the man’ who is now gathering fruit. Then both two arguments the man and the fruit are ellipised, but the basket which has yet to be introduce must be explicitly mentioned.

15. (T) [Corpus: TG]

\[
\begin{align*}
\text{piin} & \quad \text{khǔn} & \quad \text{paj} & \quad \text{bon} & \quad \text{tōnmāaj} & \quad \text{...} & \quad \text{khǎw} & \quad \text{kɔ̂} & \quad \text{kèp} & \quad \text{kèp} & \quad \text{phǒnlámāaj} \\
\text{climb} & \quad \text{ascend} & \quad \text{go} & \quad \text{on} & \quad \text{tree} & \quad 3SG & \quad \text{CONN} & \quad \text{gather} & \quad \text{gather} & \quad \text{fruit} \\
\text{lékewkɔ̂} & \quad \ldots & \quad \text{ô} & \quad (\text{He}) & \quad \text{ʔaw} & \quad \text{ô} & \quad (\text{fruit}) & \quad \text{loŋ} & \quad \text{maa} & \quad \text{sàj} & \quad \text{tákraa} & \quad \ldots & \quad \text{phɔɔ} \\
\text{LINK} & \quad \text{take} & \quad \text{descend} & \quad \text{come} & \quad \text{put} & \quad \text{basket} & \quad \text{as.soon.as} \\
\text{sèt} & \quad \text{léew} \\
\text{finish} & \quad \text{ASP} \\
\end{align*}
\]

‘(He’s) going up the tree. He’s gathering the fruit and...taking (it) down and putting (it) in a basket ... after that’s done well...’

Example (16) now formally brings in the recently mentioned basket with the presentational first construction which also features a classifier phrase (the classifier is the same as the noun, “repeater classifier” (Enfield 2007, p.122).

16. (T) [Corpus: TG]

\[
\begin{align*}
\ldots & \quad \text{kɔ̂} & \quad \ldots & \quad \text{mii} & \quad \ldots & \quad \text{mii} & \quad \text{tákraa} & \quad \text{jùu} & \quad \text{sɔŋ} & \quad \text{tákraa} & \quad \text{thîi} & \quad \ldots & \quad \text{mii} \\
\text{CONN} & \quad \text{have} & \quad \text{have} & \quad \text{basket} & \quad \text{PROG} & \quad \text{two} & \quad \text{basket} & \quad \text{REL} & \quad \text{have} \\
\text{phǒnlámāaj} & \quad \text{tem} & \quad \ldots & \quad \text{tem} & \quad \ldots \\
\text{fruit} & \quad \text{full} & \quad \text{full} \\
\end{align*}
\]

‘... there are two baskets ... two baskets that are full of fruit.’

Examples (11)-(16) provide a look at the strategy of introducing information into discourse, through a “presentational verb-first construction”, topic-prominence and ellipsis throughout Shan and Thai noun phrases and basic clauses. More specific aspects of Shan and Thai grammar are presented throughout Chapter 4.
Chapter 3

Literature Review

3.1 Introduction

This chapter is a literature review covering three topics: grammaticalization, southeast Asian languages and grammaticalization theory, and grammaticalization research on the Thai language. With a great deal of introductory literature on grammaticalization now available, §3.2 Grammaticalization Basics was written for those with little or no familiarity with grammaticalization. After a brief introduction to grammaticalization, in §3.2.1 a basic overview of the theory is laid out by going through example cases of grammaticalization and introducing key concepts along the way. This is followed by §3.2.2 Researching Grammaticalization: Framework and Methods, where important terminology related to the linguistic framework and research of grammaticalization, is defined. Section 3.3 Southeast Asian Languages and Grammaticalization Theory draws attention to important issues when it comes to linguistic research with languages within Southeast Asia, especially in regard to how grammaticalization occurs in these languages. This section is largely a review of the work of Bisang (1996; 2004; 2008; 2009; 2011; 2015) on grammaticalization theory in East and Mainland Southeast Asia, and Enfield (2001, 2002, 2003, 2005, 2011) on issues of areal and genetic inheritance in Mainland Southeast Asia. Section 3.4 Grammaticalization in Thai introduces this large body of research, and provides two example cases of Thai grammaticalizations for the purpose of showing approaches to historical research on the Thai language. Section 3.4.2 Thai Grammaticalized Morphemes outlines more on Thai grammaticalization by presenting a list of Thai grammaticalized morphemes. The importance of this list is that it features the grammaticalized morphemes that were used in this research, and which served as the starting point for a comparison with Shan. The ideas and discussion presented in this chapter should help readers to better understand the theoretical motivation which grammaticalization has provided.
3.2 Grammaticalization Basics

A basic distinction in any linguistic description is the one between the lexicon and the grammar of a language. The grammar refers to the systematic rules and structure of syntax and morphology, and the lexicon includes all of the meaningful words and morphemes that fill up the grammatical substrate. Within the lexicon a further division is made between lexical words — those that are rich or unambiguous in meaningful content— and function words — those with little meaning or no direct reference in the world. Words put in the categories of noun and verbs would be examples of lexical words (often termed open-classed in that they are always ready to accept new members), while function words would be prepositions, conjunctions, adjectives etc. (seen as closed-class — a small group that is difficult to get into). In a simple model of language the distinctions between the lexicon and grammar, or lexical words and function words, may provide a practical analysis of language at the synchronic level for modern usages. But when one begins to look more closely at language and these divisions from a diachronic perspective, the distinctions of lexicon and grammar can begin to blur. It may be here, then, between the simple distinction of “lexicon” and “grammar”, that some context for approaching the theory of grammaticalization is provided.

Grammaticalization, while raising questions at the core of linguistic categories, essentially provides a theoretical tool that explains processes by which complex linguistic structures arise. Usually, grammaticalization is defined as “the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions” (Hopper & Traugott, 1993, p.1). Though earlier ideas contributed to the idea of grammaticalization, the coining of the term is credited to Antoine Meillet (1866-1936) who used the term to describe the development of grammatical morphemes. As a theory of language change and as a branch of research, grammaticalization re-emerged in the late 1970s and has since continued to grow, making full use of modern technological innovations in linguistic research (i.e. computer based corpora). Ideas stemming from grammaticalization have come to give linguistic researchers a better picture of the cognitive basis of language and language change. And now modern instances of grammaticalization theory are closely aligned with usage-based and cognitive-functional approaches to language while also making use of ideas in constructionist grammars (See §3.2.2).
As mentioned earlier a great deal of introductory material for grammaticalization is
now available. Section 3.2.1 is greatly influenced by Hopper & Traugott (1993) and
Heine & Kuteva (2004, 2007) and borrows some of their examples and explanations
for the purpose of introducing grammaticalization theory. This is first done with the
commonly used example of the grammaticalization of English's [be going to] future
tense construction in order to explain the steps in parameters and principles in the
grammaticalization process. The aim here is to detail the development of the [be
going to] future constructions so that readers new to the topic will be introduced to
theoretical landscape of grammaticalization, important terminology and some of the
larger questions and frameworks behind work in this field.

3.2.1 Grammaticalization: Theory and Examples
What is the meaning and function of the word go? At first one might say that it is a
lexical item, a verb of motion used to indicate ‘movement from one point to
another’. While as an initial response this definition is satisfactory for many, if not
most usages, how does it fare for a sentence like (17)?

17. Tim is going to eat pizza.

Although one interpretation of (17) is one of movement from one point to another
— a present progressive one where Tim is currently “on his way” to a location
where he will eat pizza— it is also possible in this context to understand go as not
actually referring to literal movement but instead to an intention of ‘to eat pizza’
occurring at a future time. In actual speech this second interpretation of the [be
going to] construction in (17) would likely be produced in the form gonna shown in
(18a).

18. a. Tim’s gonna eat pizza.

b. Tim will eat pizza.

While there are some pragmatic considerations for the differences of (18a) and
(18b), in many cases gonna is perfectly viable alternative to will, as a future tense
marker. Going back to the original question on the meaning and function of go, with
consideration of this [be going to] future construction, we can also assign go a
grammatical function as a verbal auxiliary that indicates future tense. This now
becomes an interesting case because the definition of go is a case of polysemy with
meanings crossing over the boundary between lexical and grammatical categories.
If we look into the historical usages of each meaning of *go* we find that its use as a lexical verb dates back to Old English, while its grammatical use dates around the time of the late Middle English period (Hogg & Denison, 2006, p.296). This later appearance of auxiliary *go* implies that the grammatical function arose sometime after the verbal one. In work on grammaticalization, transformations like this are commonly written in the following form. \[ \text{go}\ V \ \text{‘to move location’} \rightarrow \text{gonna} \ \text{AUX} \] or they pictured in diagrams like Figure (7).

![Figure 7 Grammaticalization of going to](adapted from From Hoefler, 2009)

Knowing that the grammatical function of *go* came after the lexical form we can ask questions like “How did lexical *go* come to serve in this grammatical function?” and “What were the real processes of language use that made this new interpretation possible?” These are essential questions of grammaticalization, and are ones that grammaticalization research tries to answer. The underlying logic of investigating grammaticalization is the same as that of procedures within historical linguistics and reconstruction, outlined by Heine & Kuteva (2007, p.22) as following steps a-e.

a. X and Y are phenomena that are related in some way.

b. Hypothesis 1: X existed prior to Y.

c. Hypothesis 2: There was a change X > Y
   (but X continues to exist parallel to Y).

d. There is evidence in support of (c).

e. There are specific factors that explain (c).
From this perspective it is up to grammaticalization researchers to propose the specific developments or environments that allowed lexical item X to become grammatical item Y, and support claims that explain steps a-e for actual cases of grammaticalization. For example taking the apparent diachronic development of grammatical go through aforementioned dating of both the lexical and grammatical form, we must look for an explanation as to what happend in step c, as well as provide the evidence for steps d and e.

Tackling go’s grammaticalization begins by looking at the environments in which it occurs. The first thing to note is that grammatical go appears in the [be going to] construction, and is often reduced as gonna. Also we can see that the [be going to] usage only occurs in very specific sentential contexts and not others. One type of appearance is in purposive directional constructions with non-finite complements as in (17) or (18a) and not contexts where a locative adverb is present like (19a) or (19b), which instead use a verbal go.

19.  a. I am going to London.
     b.*I’m gonna London.

Looking at go in (19a) we can notice its inflected form indicating an action occurring in the present tense, while (17) and (18a) use it to mark the occurrence of a future event. This explains why the reduced form gonna is not acceptable in (19b). To see how these two forms are related, the lexical go in (19) and the grammatical go’s later appearance, a better understanding of specific processes of grammaticalization is needed. Heine & Kuteva’s (2007) parameters of grammaticalization is proposed as a model of the grammaticalization process with the following four main steps I-IV

(I) extension where a linguistic expression is extended to a new context
(“context-induced reinterpretation”)

(II) desemanticization (also known as “semantic bleaching” or
“generalization”) in meaning content

(III) decategorialization, loss in morphosyntactic properties

(IV) erosion (“phonetic reduction”).

Consider how a sentence with a future tense “I will eat pizza” points towards a future event while a sentence like (19a) has the literal sense of movement to a location “London”. Yet, in (19a) though one interpretation is one of a present continuous
tense indicating present movement (towards London), a listener could very likely expect that the future holds an arrival at the destination (London). The [be going to] construction in (19), thus lends itself even if ever so slightly to other interpretations. This possible future-ness (the arrival in London) implied in (19a) is what allows step (I) extension to occur. This first parameter of grammaticalization refers to where semantic ambiguity allows listeners to extend meaning into contexts other than the literally signified or intended one.

What exactly causes the [be going to] construction to become extended to this context? Hopper & Traugott (1993) say that two mechanisms that allow extension or “context-induced reinterpretation” are reanalysis and analogy. If listeners can infer extra information that was not explicitly marked – like the future arrival at a destination – then their reanalysis has given new meaning to the utterance. Abductive reasoning drives these two mechanisms and thus can lead to novel grammatical creations. The [be going to] construction at one time was limited strictly to contexts of movement of location, yet it semantically lends itself to a purposive intention about the future. Once a [be going to] construction was used in a context without any obvious movement the listener then interpreted the structure as not one of movement over space but as a movement through time. With this possible interpretation occurring enough, sentence (17) beginning with “Tim is going to ...” was ready to receive non-locative complements like “to eat pizza”.

Once the [be going to] construction was reanalyzed often enough, with its new sense, as a marker of ‘intention/future’, then parameter (II) desemanticization characterizes the gradual loss of the original meaning of ‘movement through space’. This process of losing meaning is referred to as “semantic bleaching”. Take an example such as “I’m going to be thinking of you” which does not have any literal sense of movement. This is a case where it can be said that the motion meaning of go has been “bleached” out. It should be mentioned that despite this bleaching, a grammaticalized form can still co-exist along with the lexical parent from which it came. It is important to point out here that this “semantic bleaching” does not have to apply to every use of ‘go’, instead bleaching can be seen as taking place only within certain contexts. The more and more frequently go was reanalyzed, within the [be going to] construction the more likely it is to become separated from its original sense. Consequently, after the verbal sense of go has gone through

---

8 Abductive reasoning is drawing a conclusion from many cases and applying it as a rule. Another place where this comes into discussions of language is in pointing the reasons for common mistakes of children learning English with over use of rules instead of irregular forms such as, “He go-ed”, or “gooses”.
desemanticization, in the [be going to] construction, it is free to take on a new grammatical function, that of an auxiliary. Look at example (20) where [be going to] is now used to mark the lexical verb go.

20. I am going to go London.

With the [be going to] construction now acting as an auxiliary on its lexical parent verb ‘go’, the contrast in function and the move from lexical word to grammatical morpheme can clearly be seen. This development brings us to parameter (III) decategorisation; the moving of a morpheme from one function to that of another. Verbs which have gone through decategorisation are no longer able to be inflected, morphologically derived or modified by adverbs — essentially they lose most features of verbs by moving from a open-class item to a closed-class one (Heine and Kuteva, 2007 p.41). This stage is also characterized by a lack of syntactic independence (see §3.3.1)

So far we have followed go through the first three parameters of grammaticalization extension, desemanticalization, and decategorization. Continuing to parameter (IV) erosion, Bybee’s (2003) ideas about the underlying cognitive and linguistic motivations for grammaticalization are important to consider. In the context of grammaticalization, Bybee (2003) discusses the role of frequency when looking at “grams” (grammatical morphemes). Retracing some of steps we have covered so far an important thing to point out is that as a lexical item goes through the process of extension or reanalysis, the frequency of its appearance in certain constructions may increase. As grammaticalization begins to occur, grams become more general or abstract when extended into new contexts. If this new form is grammatically productive then that gram is likely to be seen more frequently. Due to the cognitive basis for linguistic processing, eventually the role of the gram, and the constructions in which it appears, begin to become automated as a single processing unit (e.g. [be going to] > [gonna]). As the items within the construction tend to be associated together more often, the phonological boundaries between the items will begin to blend. This is done out of efficiency of expression when listeners are bound to expect the occurrence of items especially when they must occur together. This explains the reduction pattern of [be going to] where going to [gʊŋ tʰuː] becomes gonna [ɡənə] and even more so where I’m gonna becomes [aiməɾə]. Greater efficiency of expression and increasing frequency create a feedback loop where a gram and its constructions can lead to even further grammaticalization. The role of frequency
here then seems to be both the cause and effect of grammaticalization, for frequency leads to semantic bleaching and phonological changes in parameter (IV) erosion.

Again it should be stressed that after parameters (I)-(III), the final stage erosion or phonetic reduction is done out of efficiency of expression. Langacker (1977, p.106) portrays this aspect of language with the following metaphor, “It would not be entirely inappropriate to regard languages in their diachronic aspects as gigantic expression-compacting machines.” This information compacting happens at many levels, from pragmatic to phonological, and high levels of frequency motivate the phonological reduction in parameter (IV). A high level of frequency can be better understood as affecting grammaticalization when we see it as essentially being repetition for language users. Bybee frames this mind-grammar interface as follows,

“Repetition is universal to the grammaticalization process. Repetition and its consequences for cognitive representation are major factors in the creation of grammar. The conventionalized aspects of language provide the framework for manipulation of our thoughts into objects of communication” (Bybee, 2003, p.622).

So while frequency or repetition can most easily be seen as acting in parameter (IV) erosion, its role in every stage of grammaticalization should not be underestimated. Interestingly, as we look back at the parameters of grammaticalization (I)-(IV), each step corresponds to a level linguistic structure: (I) extension relates to matters of pragmatics, (II) desemanticalization to semantics, and (III) decategorialization to morphosyntax, and finally (IV) erosion to phonetics.

Seeing how each of the parameters contributed to the grammaticalization of the lexical verb go, that this happened in a certain construction is also of high importance. As Dahl (2004) puts it, “A lexical item cannot by itself come to serve a grammatical function; it must do so by virtue of becoming a fixed part of a larger pattern — a grammatical construction”(p.119) A construction allows a lexical item to become “trapped” in a pattern. He then says that this pattern can go through stages until it eventually “matures”. The major components of the maturation of grammatical patterns consist in the spread, regulation and adaptation of new constructions. He sees these patterns as finding “niches”, where there is a grammatical need. While a new pattern may be used at the same time as an older, more similar one, as soon as a new meaning or sense is given to one, they will begin to diverge in meaning and use. This is what has happened to a French noun pas
‘step’ in the negation construction. This example of grammaticalization is described in Hopper & Traugott (1993, p.65) as having occurred in the following way I-VI.

I. Negation was accomplished by placing the negative particle ne before the verb.

II. A verb of motion negated by ne could optionally be reinforced by the pseudo-object noun pas ‘step’ in the context of verbs of movement:

\[
\begin{align*}
  \text{Il ne va (pas)} \\
  \text{he not goes step} \\
  \text{‘He doesn’t go a step.’}
\end{align*}
\]

III. The word pas was reanalyzed as a negator particle in a structure of the type ne V movement (pas).

IV. Pas was extended analogically to new verbs having nothing to do with movement; i.e., the structure was now ne V (pas):

\[
\begin{align*}
  \text{Il ne sait pas} \\
  \text{he not know not} \\
  \text{‘He doesn’t know.’}
\end{align*}
\]

V. The particle pas was reanalyzed as an obligatory concomitant of ne for general negation: ne V pas.

VI. In the spoken vernacular pas came to replace ne via two stages: (ne) V pas (reanalysis of ne as optional), V pas (reanalysis by loss of ne), resulting in:

\[
\begin{align*}
  \text{Il sait pas} \\
  \text{he know not} \\
  \text{‘He doesn’t know.’}
\end{align*}
\]

Seeing how the negator ne could fall away after pas filled this grammatical niche is one example of different markers competing and serving a grammatical function. In English the [be going to] construction still coexists with another future tense marker will. If someone was to hold will as the prototypical grammatical marker for future tense, and place the [be going to] construction as a newcomer, best left for certain informal contexts, then they would miss the larger picture of grammaticalization.

From the point of grammaticalization, and in its deepest reaches, every grammatical marker can be thought of as having its own history in which it once competed for a grammatical niche. The future tense will was originally a lexical verb meaning ‘to want or desire, intend’ (Bybee, Pagliuca, & Perkins, 1994, p.255). Traces of this
meaning of intention are still seen in some modern usages such as “Do what you will.” From Dahl’s perspective we can see that gonna and will were both once lexical items but are now trapped in a grammatical space where it is possible for either of them to become fixtures or fade away, as is the case with the gradual disappearance of shall as a verbal auxiliary (shall was originally a verb ‘to owe’ Bybee, Pagliuca, & Perkins, 1991, p.26). In fact, the closer we look at any of English’s pre-verbal auxiliaries we can find a lexical past which may still have some semantic residue in current usages. In the case of English modals, all modern modals may, might, can, could, will, would, should, must have been grammaticalized from Old English verbs (Singh 2005, p.30). Since the Old English period they have gone through processes that have distinguished them from other usages as evidenced by limitations on which contexts they can appear in (i.e. main verbs).

The previous example of [be going to] showed how a verbal auxiliary came out of the usage of a lexical verb. The gonna form now occupies a grammatical space along with other English auxiliaries that are also products of grammaticalization. Yet, the true explanatory power of grammaticalization becomes apparent when we see that the case of something like will -the grammaticalization of a verb ‘to want’ to a future tense marker- is not limited to just to the English language. For when we look at other languages we find that this particular type of grammaticalization appears to occur across many languages. This brings us to the idea of the cline. Hopper & Traugott (1993) say, “The term "cline" is a metaphor for the empirical observation that cross-linguistically, forms tend to undergo the same kinds of changes or have similar sets of relationships, in similar orders” (p.6). A cline is a “pathway” of universal grammatical developments that occur cross-linguistically. Essentially showing that change in grammatical structures happen in predictable or regular ways, and which leads to eventually raises questions about all grammatical structures themselves and the larger cognitive principles at work.

While there are many quite specific clines, as a general rule adjectives, adpositions, and agreement markers are likely to come from nouns, and that TAM markers, passives, adverbs, and case-marking are likely to be grammaticalized from verbs (Heine & Kuteva 2007, p.114). Bringing our [be going to] example back into the discussion, from a modern perspective go can be seen to have at least two functions, that of a verb and a tense marker. With the knowledge of established clines, and even without any historical data we could still venture a hypothesis that there was lexical meaning to go before a grammatical function existed, based on the existence of clines based on similar cases of grammaticalization occurring in other languages.
In (21-24) other cases of grammaticalization in the cline of VERB ‘want’, ‘wish’, ‘desire’ > Future (Heine & Kuteva 2004, p.310), as mentioned earlier with English will we can see that other non-related languages have made similar developments such as in (21) > (22) and (23) > (24) below.

Latin volere ‘want’, verb > Romanian future marker.

21. *volo* cantare
   want:1SG sing:INF
   ‘I want to sing’

22. *voi* cinta
   want:1SG sing:INF
   ‘I will sing’


23. *a- taka ku- ja*
   3SG:PRES want INF come
   ‘She wants to come.’

24. *a- ta ku- ja*
   3SG:PRES FUT INF come
   ‘She will come.’

And as investigation into grammaticalization has increased the easier it is to see more examples and more languages fitting into known clines. Since studies in grammaticalization show that structure is in some sense predictable, forms and constructions can be accounted for in a principled way. These accounts then rest upon the principle of unidirectionality which holds that in general, less grammatical items move towards becoming more grammatical items. The unidirecionality of grammaticalization also serves as the foundation for the structure of clines. This means that we can expect certain features to arise out of others and not the other way around. Unidirectionality in language change has parallels in phonology (e.g. stop > affricate > fricative) and morphology (e.g. lexical item > clitic > affix) (Hopper & Traugott, 1993, p.17). Understanding the direction in which change happens gives us grounds to hypothesize about the earlier conditions of a language. And so by giving generic principles by which grammars evolve, principles from grammaticalization theory can aid in historical reconstruction.
That there are broad paths or clines which happen so similarly cross-linguistically points to underlying cognitive principles that help drive grammaticalization. Nouns and verbs as the main open lexical classes now must be ultimately held as the source for all grammatical developments. Given this view we can see that a strong distinction between lexicon and grammar as simple and at best only functioning as an aid in understanding synchronic snapshots of current usages. Grammaticalization ultimately frames grammars as emerging from a lexicon through natural processes. Outlining this process Heine & Kuteva (2007, p.111) provide the diagram in Figure 8 showing the layers of grammatical development.

![Figure 8 Layers of grammatical development (Heine & Kuteva, 2007)](image)

Abbreviations: I, II, et. = layers; AGR = agreement marker; ADP = adposition; ASP = (verbal) aspect; CAS = case marker; CPL = complementizer; DEF = marker of definiteness (“definite article”); DEM = demonstrative; NEG = negation marker; PASS = passive; PRN = pronoun; REL = relative clause marker; SBR = subordinating marker of adverbial clauses; TNS = tense marker.

Once taken all together these ideas have interesting implications outside of just historical reconstruction of languages. In terms of the evolution of human languages; grammaticalization provides a mechanism to explain the origin of linguistic complexity. Since we can see that the lexicon is a prerequisite for grammatical categories, we can infer that the earliest units of language were likely lexical in form (Heine & Kuteva 2007, p.119). As a process of linguistic change, grammaticalization has the potential to give enormous insights into the origin, historical development
and current processes of language change, and should be considered a significant linguistic phenomenon. The importance of grammaticalization for the study of language can be seen in Givón’s (2002, p.39) remarks, “Like other biological phenomena, language cannot be fully understood without reference to its evolution, whether proven or hypothesized.” Grammaticalization as a mechanism of language change and as a field of linguistic inquiry provides insight to some of the complexity of language.

3.2.2 Researching Grammaticalization

Current perspectives on grammaticalization occur in the usage-based or cognitive-functional view of language. This perspective on language focuses on the processes that create language. By not focusing on just surface linguistic structure, usage-based theories seek to derive linguistic structure from domain-general processes of human cognition (Bybee 2011, p.7). The linguistic unit studied here is not just individual morphemes but often the construction – a form and meaning pairing of linguistic information – which provides revealing generalizations of actual language use, such as the [be going to] construction discussed in §3.2.1. Constructions are schematic blocks of language that can occur in several levels of language. Hoffmann and Trousdale (2013, p.1) give the following examples of constructions.

(1) word construction: apple [æpl]—‘apple’
(2) idiom construction: e.g., X take Y for granted [X TAKE Y fə ɡənt’d]—‘X doesn’t value Y’
(3) comparative construction: e.g., John is taller than you [X BE Adj comparative ðən Y]—‘X is more Adj than Y’
(4) resultative construction: e.g., She rocks the baby to sleep [X V Y Z]—‘X causes Y to become Z by V-ing’

Each particular construction within a language is a potential resource of expression for speakers of that language. From the viewpoint of Construction Grammar, a “grammar” is built up from the vast and various constructions that are based on schemas of actual language use. In the words of Tomasello (2003, p.5) “In the usage-based approach, competence with a natural language consists of the mastery of all its items and structures, and these constitute a much more complex and diverse set of linguistic representations than the “core grammar” of formal approaches.” From this approach it is possible to treat rules of grammar as “bleached-out” idioms, as opposed to deep generative algorithms which define the grammatical possibilities. (Culicover & Jackendoff, 2005, p.43)
One of the methods by which usage-based approaches and construction grammar have been supported is through a wealth of authentic language data provided by corpus linguistics. Corpora were originally used as tools for lexicography and pedagogy, but with advancements in the ease of compiling and using corpora, the field of corpus linguistics has become accepted as a productive field of linguistic inquiry (Baker 2010, p.1). The logic of using corpora is that by carefully constructing a collection of texts a corpus can give language researchers easy access to language data numbering in thousands if not millions of words. Thus allowing for the large amounts of empirical data to be included in linguistic analysis.

Within the area of corpus linguistics there are many corpus which differ in their construction and ultimately their aims. A corpus that is made to be representative of a language over a long period of time is called a diachronic corpus. These can be used to track language shift or change within a particular language or variety up to the latest texts within the corpus. This type of corpus should be distinguished from a historical corpus which contains a body of texts that are seen as representative of a certain time period in a language’s history, though many times for the purposes of comparison with modern uses. A synchronic corpus contain texts which have all been produced around the same time period to capture a ‘snapshot’ of a language at a particular time. The construction of a corpus often depends on what it will be used for and careful considerations must be taken to ensure the validity of using it to make broader claims regarding a language. Another use of corpora is in comparative linguistics, where frequencies of certain aspects or can be compared between two or more languages. The claims within usage-based approaches and construction grammar have made use of corpus linguistics by being able to give quantitative measurements to the actual use of linguistic data. This measurement is often one of frequency. Two types of frequency are possible here the first, token frequency concerns the appearance of a word or morpheme in a text. The second type is type frequency, which is the frequency of a pattern like a stress pattern, affix or context within a construction. With the measurement of frequency in corpora researchers are able to make form hypotheses on phenomena in language.

In modern grammaticalization research the usage-based and construction grammar frameworks along with methods from corpus linguistics have come together as a productive way to explore grammaticalization. One example is Germanic Future Constructions by Martin Hilpert (2008). Investigating grammaticalization through a

---

9 See Baker (2006) for more on corpus linguistics
usage-based approach, Hilpert (2008) says that quantitative corpus linguistics “aims to address questions of theoretical relevance on the basis of frequency patterns in naturally occurring data, while maintaining methodological rigour to ensure the replicability of obtained results” (p.28). More importantly Hilpert sees his approach as an integration of the several linguistic fields, “The combination of Corpus Linguistics, Grammaticalization Theory, and Construction Grammar makes it possible to discover and describe phenomena that earlier research programs, and each component framework on its own, were bound to miss.” He concludes that this approach “may prove useful for exploratory studies and the evaluation of existing theories.” The integration of corpus linguistics and grammaticalization is not surprising since one of the most notable characteristics of grams and the constructions in which they occur is extremely high text frequency (Bybee & Dahl, 1989). Within grammaticalization theory, frequency can be seen as both a sign and cause of grammaticalization. (Bybee, 2007, p.336). Another study Post’s (2007) *Grammaticalization and compounding in Thai and Chinese*, compiled two synchronic corpora in order to compare relative grammaticalization levels in two languages. Post frames the role of frequency as being able to show obligatoriness in certain constructions or functions, and point towards a grammatical words as opposed to lexical ones. (Post 2007, p.125). Yet, Post cautions that there is no frequency “threshold” for grammatical status, and that value of frequency is in its comparative value, of “relative depth of grammaticalization” (2007, p.125). Both of these studies serve as examples of the use and validity of corpus linguistics for investigating grammaticalization, as well as their context within construction grammar and usage-based or cognitive-functional views on language.

### 3.3 Grammaticalization and Southeast Asian Languages

#### 3.3.1 The lack of the coevolution of meaning and form

In classic models of grammaticalization by C. Lehmann (1995) and Bybee (1985), it was thought that as grams moved further away from their lexical origins, they would become more abstract in meaning, have less syntactical autonomy, and undergo phonetic reduction (See. “parameters” in §3.2). This process often resulted in a change in both the form and meaning of a word. For example, the phonetic reduction of grams in *be going to* > *gonna* or affixation of *lic ‘body’* > *-ly ‘adverbial suffix’* (e.g. *quickly*) are cases were the form has changed along with a new grammatical function. This classic conception of grammaticalization can also be seen
in the "cline of grammaticality" [content item > grammatical word > clitic > inflectional affix] which outlines the changes in form a gram takes as it reaches higher levels of grammaticality (Hopper and Tragoutt, 1993 p.5). Another name for this aspect of grammaticalized items is the “coevolution of meaning and form”. But, as pointed out in the work of Bisang (1996; 2004; 2008; 2009; 2011; 2015) when it comes to languages in East and Mainland Southeast Asia (EMSEA), there are unique differences when it comes to “the coevolution of meaning and form” and other theoretical issues in grammaticalization studies.

One main typological features that explains the absence of the coevolution of meaning and form in EMSEA languages is that languages here display a lack of obligatory grammatical categories, thereby making it difficult to categorize many morphemes used in these languages. The indeterminateness of grammatical morphemes in EMSEA languages, as many also still serve largely lexical functions as well not only presents challenges for grammaticalization theory, but for general grammatical analysis as well. As in Enfield (2002, p.14),

“Grammarians perhaps experience some psychological difficulty in viewing certain items (in Southeast Asian languages) as morphosyntactic markers, since they also function as full lexical items (mostly verbs) elsewhere. However, there is good reason to view them as grammatical morphemes, given their often strict syntactic position and restricted morphosyntactic behaviour when performing more structural or grammatical functions.”

This categorical ambiguity in EMSEA explains why some “classic” features hypothesized to appear along with grammaticalization are absent, suggesting that previous conceptions of the theory did not consider factors which are pertinent to languages in this region. Bisang (2008, p.15) maintains that the lack of coevolution of meaning and form is due to the following characteristics in EMSEA languages (i) the relevance of pragmatic interpretation (ii) the non-existence of morphological paradigms (iii) the nonconvergent development of form and meaning. As an example Bisang provides case of Khmer ãn shown in (25), which without context in discourse can actually have more than one possible meaning, thus ãn in (25) is simply marked as a tense, aspect, and modality marker (TAM).
25. *khnom ban* *tɔŋu phsaː(r)*
   I TAM go market

a. I was able/allowed to go to the market.
b. I went to the market
c. I was at the market. [Against the presupposition that I was not]

Without the surrounding context from discourse the meaning or function of *ban* is indeterminable. Through possible translations a-c of (25), we can see how a lack of context allows *ban* to be open to different interpretations. Bisang (2008) thinks the possible meanings in a-c as come from one source concept while exact meanings are ones that must be pragmatically inferred from context. This indeterminateness is part of the reason for classification of *ban* as a TAM marker, though it may be criticized for a lack of precision. Not only is *ban* a multifunctional TAM marker, but it can also function as a main verb, and the pre-verbal position shown in (25) only one of other possible syntactic positions for *ban*.

Bisang’s purpose for presenting *ban* is to contrast this case with grammaticalizations in other languages which are explicitly marked through a change in form. Example (25) above also illustrates how EMSEA language’s strong reliance on pragmatic inference leaves sentences and clauses without the obligatory grammatical marking normally seen in inflecting languages which would distinguish meaning. The non-existence of morphological paradigms in EMSEA or the “weak correlation between lexicon and morphosyntax” consequently leaves many grams dependent on specific context for classification. This factor not only makes it more difficult to categorize clause-less or context-less morphemes, but also difficult to determine possible paths of grammaticalization. Thus, deciphering the grammatical development of *ban* becomes troublesome in that the nonconvergent development of form and meaning leaves *ban* without clear clues to possible grammaticalization paths in the same way as the grammaticalization of [be going to] > ‘gonna’ resulted in a gram with a distinct change in phonological form, which also corresponds with a highly specific grammatical role. Previous theories of grammaticalization hypothesized that high grammaticality would correspond to less inferential freedom, yet the opposite seems to be the case with EMSEA languages (Bisang 2004, p.109).

To address these issues Bisang recommends that researchers look beyond just individual grams and include constructions —form and meaning pairing of linguistic information into their investigations (See. §3.2.1). By looking to constructions, a better understanding grammaticalization in EMSEA languages is possible, and a
coevolution of meaning and form can be discovered. For example Takahashi’s (2008) corpus-based study examined the grammaticalization path of Thai dâaj\(^{10}\) by looking at its appearance in constructions. From a main verb meaning ‘emerge’ to a marker denoting ‘possibility’ the shift in function can be seen in the constructions in which it appears. An early use shown in (26) is a verbal construction [dâaj quantity-NP] which dates back to the 14\(^{th}\) century.

26. [Takahashi, 2008, p.124)]
\[
\text{sàkkaràat dâaj cèt ríj sàam sip pèct}
\]
\begin{array}{ll}
\text{era} & \text{DAJ} 738 \\
\end{array}
\]
‘As for the era, 738 years emerged. (The period of the era amounted to 738 years)’

A shift in form can be seen in the possibility construction [VP dâaj] shown in example (27), which is much more frequent in the 20\(^{th}\) century.

27. [Takahashi, 2008, p.129)]
\[
paj dâaj
\]
\begin{array}{ll}
\text{go} & \text{DAJ} \\
\end{array}
\]
‘It is possible to go.’

Notice in Takahashi’s examples (26) and (27) the glosses for the individual dâaj morpheme is left as DAJ. That is because the change in meaning and form here happens at the level of a construction as opposed to a word (e.g. phonological reduction). Through constructions grams can be understood as occupying syntactic slots within patterns which can be valuable for understanding the process of grammaticalization. Bisang thinks this is especially the case with EMSEA languages. On grammaticalization in Chinese, Bisang (2010, p.275) remarks,

“constructions with their syntactic slot-structure are the driving force of reanalysis and analogy. This may be different in other languages. For that reason, the construction grammar approach must integrate data from typologically divergent languages for a more deeply grounded empirical basis of its generalizations... constructions seem to be even more important for grammaticalization than in English and other languages summarized under the heading of Standard Average European.”

\(^{10}\) This particular ACQUIRE morpheme is covered extensively in the work of Enfield (2003) as an extensive feature of Southeast Asian languages. The gram ba:n in (25) here is similar to the Thai dâj and is similar in many respects, both in the relevance of pragmatic interpretation and the multiple syntactic slots it can occupy, which also mirrors Shan laj (See Section 4.2.2).
This “new” way of looking at grammaticalization is summed up in Bisang’s adaption of the definition of grammaticalization from “lexical item > grammatical item” to “use of lexical item in discourse > grammatical item” (Bisang 2004, p.116). With words crossing both lexical and grammatical boundaries it now becomes necessary to look at constructions when exploring grammaticalization — an idea that Bisang admits goes back to the model of Bybee et al. (1994), and concerns other languages as well. The typological characteristics of EMSEA languages namely the lack of coevolution of meaning and form and high reliance on pragmatic inference in discourse force researchers to reconsider classic grammaticalization theory. Hopefully, as grammaticalization theory expands its scope to include more disparate languages researchers will have a more detailed picture of how processes of grammatical evolution occur. Bisang highlights the relevance of constructions for grammaticalization theory in EMSEA in the following way.

“The way in which processes of grammaticalization develop in individual languages and in larger areas such as the one covered by the EMSEA languages is determined by specific factors that are responsible for the degree of (in)dependence with which the domains of pragmatics and phonology interact. This opens a new typological perspective in grammaticalization research, a perspective that may be able to account for different manifestations of grammaticalization. From such a typological perspective, research on grammaticalization is anything but irrelevant.”
(Bisang 2011, p.9)

Bisang’s work on grammaticalization in EMSEA languages brings attention to important factors and addresses the need for new approaches of research for grammaticalization. Previous models of grammaticalization held that there would be a coevolution of meaning and form as grams reached higher levels of grammaticality. Besides providing researchers with important theoretical considerations when investigating individual EMSEA languages, Bisang also broadens the discussion on grammaticalization by considering unique typological factors given by EMSEA languages. Research on grammaticalization in EMSEA languages must take into account the multi-functional nature of grams in these languages. Grams that still maintain their lexical roles should be investigated at the level of multi-word constructions. And more broadly EMSEA as a linguistic area opens up possibilities where grams and linguistic items across languages have influenced grammaticalization, testing the distinction between areal and genetic inheritance.
3.3.2 Areal versus genetic inheritance

When researching any languages of Mainland Southeast Asia (MSEA) it is important to recognize the sociolinguistic factors that have shaped the languages that currently inhabit this area. In MSEA prolonged language contact between languages may be seen as responsible for structural reanalysis and other processes of language shift (Diller 2004, p.94). These factors and other issues surrounding the topic of MSEA languages have been covered extensively in the work of Enfield (2001, 2002, 2003, 2005, 2011). Typologically languages in MSEA not only share many broad linguistic features but they also share some quite specific grammatical items and constructions (e.g. the syntax of classifier constructions). Finding grammatical similarities across the boundaries of unrelated languages, makes the job of historical reconstruction more complicated than in more homogeneous language areas, and reinforces the need for awareness of areal factors. Enfield points out that for linguists working with these languages “the puzzle in MSEA, like in every region, is often seen to be that of distinguishing between two kinds of cause for the existence of common structure in languages: internal vs. external change, vertical vs. horizontal transmission, descent vs. diffusion” (Enfield 2011, p.74). Essentially these questions are ones of determining whether some linguistic (grammatical or lexical) item is the result of areal or genetic inheritance.

The background to this issue is the unique social, linguistic, and cultural histories that have brought languages in the Austronesian, Hmong-Mien, Mon-Khmer, Sino-Tibetan, and Tai-Kadai families together in close proximity and created the linguistic environment in MSEA. Researchers must keep in mind how the history and movement of people who speak these languages may have shaped various aspects of each language’s current form. Along with how current sociolinguistic factors like diglossia and bilingual speech communities which also brings more variables into play.

An example of how these factors can affect linguistic research are cases where one encounters parallel forms in two or more languages. In his book Linguistic Epidemiology (2003), Enfield considers the case of the ACQUIRE words which appear throughout the region’s languages. ACQUIRE words function in similar grammatical constructions across MSEA languages: as main verbs meaning ‘come to have’, a post-verbal modal marker or introducer of an adverbial phrase, and also a preverbal marker of ‘attainment’. Examples of ACQUIRE words include Khmer ban and Thai dâaj (examples (25)-(27)in §3.3.1) and Shan laj within this study (See. §4.2.2), other languages that have this gram include Hmong and Vietnamese. The
widespread existence of such a particular grammatical marker raises questions on
the origin of these words and their appearance in each language. Enfield proposes
the following hypotheses (a-e) as possible explanations for the ACQUIRE pattern in
MSEA languages.

(a) the patterning has been inherited by both x and y from a single common
ancestor which displayed the patterning;
(b) the patterning has been copied from x to y, or from y to x, or from a third
language into both;
(c) x and y have each independently and internally innovated the patterning,
by natural cognitive principles (which may be contingent upon existing
semantic/grammatical structures);
(d) x and y have each independently and randomly innovated the pattern,
with any parallelism purely coincidental;
(e) any combination of the above (pg. 367).

In confronting cases of parallelism of grams in MSE
A languages, Enfield stresses how
the case of MSEA requires us to see past a simple model of areal versus genetic
inheritance, even so much as to call the distinction “illusory” (2011, p.74). Cases
like the ACQUIRE words show how “unit-based diffusion” by the calquing of function
words in between languages could explain some cases of parallelism found in MSEA.
New forms in languages here may be the result of using existing lexical items in
novel grammatical ways based off grammatical patterns in other surrounding
languages. Borrowing in this sense is not of an actual word, but instead it is the use
of a language’s own lexical material to perform a new and borrowed grammatical
function. This possibility erodes the distinction between strict areal or genetic
inheritance. This type of borrowing is also mentioned in Matisoff (1991) as having
an impact for grammaticalization studies in MSEA, such as in the case where new
functors were introduced to Chinese minority languages by administrators. But he
also mentions that “the borrowing or calquing of functors from more prestigious
languages is something that has gone on for millennia even without official
encouragement” (Matisoff, 1991, p.445). Borrowing in this way is based on
conceptual grammatical units that can be used for new applications in a borrower’s
language.

One of the factors that allows this type of unit-based diffusion to take place is
“typological poise” — the system internal aspects of a language that allow it to
accept new innovations. If the use of a lexical item in a grammatical construction is
natural within a language then it may develop a parallel grammatical construction in another language.

"Typological structure causes a grammar to be 'poised' for particular developments (and not for others), determining the readiness or susceptibility of speakers to make or adopt a given extension. Typological poise is less a determinant of how likely a language is to borrow a given structure than of how likely it is to develop it independently. The idea of typological poise can provide an account for why or how a particular innovation occurs in one case, but not in another, especially with respect to the novel use of existing lexical items in more grammatical ways." (Enfield, 2003, p.5)

It may also be the case that the typological poise of a language is not genetic relatedness, but the result of earlier contact between two languages. With the right typological poise some language’s may be more prone to develop or borrow new grammatical functions from the surrounding languages. The possible calquing of functors in MSEA languages means that modern grammars may not always be the result of solely language internal developments or the wholesale borrowing of a foreign words. It is these possible factors of language change that blur the line between areal and genetic inheritance. The linguistic situation in MSEA should give researchers pause when conducting research which touches upon historical reconstruction or grammaticalization. Enfield likens linguistic description here to epidemiology\textsuperscript{11} with the behavior of grams analogous to memes or viruses which can spread to other linguistic environments. This situation then puts the burden on researchers to be careful with cross-linguistic descriptions and requires extra methodological rigor in making claims here.

### 3.4 Grammaticalization Research in Thai

The Thai language is the most studied Southeast Asian language in regard to grammaticalization – due in part to the existence of written texts and inscriptions dating back over 700 years. This deep historical record of language provides researchers the diachronic resources to compile corpora and investigate grammaticalization in the Thai language. Plenty of this research is published in English, but there is also a large portion of theses and papers on the topic that are

\textsuperscript{11} This is the branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health.
published in Thai. The body of literature here is so large that reviewing all of it is outside the scope of this paper. For anyone approaching grammaticalization in the Tai language family for the first time, Diller (2001) is an excellent starting place that shows many examples of grammaticalization in Thai and the broader Tai language family as well. Some of Diller’s (2001) examples are from an earlier and relevant paper — Matisoff (1991) — which not only provides several examples of grammaticalization in Thai but many examples from other Southeast Asian languages as well. The growing interest of grammaticalization studies for Thai can be seen in Diller et al. (2008) which features an entire chapter composed of articles focused on aspects of grammaticalization in Thai.

The body of literature on grammaticalization in Thai has now grown to provide an interesting picture of the development of Thai grammar. Although reviewing the grammaticalization paths and details of each Thai grammaticalization can not be done within the space of this paper, it is important to at least understand the research methods by which many of these grammaticalization paths are evidenced. The purposes of this section then is to provide examples of how diachronic corpora have been used to support grammaticalization research in Thai. Understanding the logic of this type of research and how evidence was obtained and subsequently used to support the cases of grammaticalization presented next within §3.4.1, should allow readers to imagine the analogous methods by which many other Thai grams have been investigated. This section does not present examples of actual language use but instead features evidence in the form of frequency measurements from diachronic corpora from Diller (2001), Takahashi (2011), and Kullavanijaya (2008). Basically this research is done from the idea that quantifying lexical and grammatical usages over an extended period time can allow one to trace the pattern of a word’s grammaticalization.

To introduce the research methodology used in Thai grammaticalization research Section 3.4.1 provides two example cases of grammaticalization in Thai. The first one, the irrealis marker cà, is a gram used in VP syntax, and the next one, thûi, is a multi-functional gram used in NP syntax. Both of these grams have been put forth as grammaticalized and are supported through evidence from analyses of diachronic corpus data (Diller, 2001) (Takahashi, 2011) (Kullavanijaya, 2008). The goal of this section is not to explain the specific grammatical function of both cà and thûi but instead to emphasize the historical evidence used to support claims of each morpheme’s grammaticalization. After the example cases, a table is presented in
§3.4.2 that shows many other Thai grams that were of interest in this study. Many of the grams in the table were investigated by similar methods from corpus linguistics used for cà and thîi. For those interested in more detailed information on specific grams, a fuller version of the table can be consulted in the appendix which contains citations for literature about each morpheme.

3.4.1 The cases of cà and thîi

The word cà is a pre-verbal irrealis (IRR) marker often used to mark future events in addition to a range of modal concepts\. The grammaticalization of cà was studied by both Diller (2001) and Takahashi (2011). Diller (2001) established the grammaticalization of cà 'IRR' from an earlier verb câk ‘to desire, intend’, by tracking uses of it through a diachronic corpus. The grammaticalization of cà shows many of the parameters discussed in §3.2.1, such as decategorialization (full verb to auxiliary), phonological reduction (loss of a final consonant), and the coexistence of the two forms (câk is still found in some formal styles). Table 13, shows a gradual shift over the 13th to 16th centuries from use of the lexical verb câk, to the irrealis marker cà.

Table 13 Percentage of cà vs. câk in Thai inscriptions (Diller, 2001)

<table>
<thead>
<tr>
<th>period (A.D.)</th>
<th>cà ‘IRR’</th>
<th>câk ‘to desire, intend’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1292-1350</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1351-1400</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>1401-1450</td>
<td>37</td>
<td>63</td>
</tr>
<tr>
<td>1451-1550</td>
<td>91</td>
<td>9</td>
</tr>
</tbody>
</table>

As discussed in §3.2.1 the existence of clines is be reliant upon the existence of similar grammaticalization pathways cross-linguistically. In this case it should be noted that grammaticalization of cà also conforms to a cline V (‘want’, ‘wish’, ‘desire’) > FUTURE (Heine and Kuteva, 2004, p.310).

Continuing the investigation of cà, Takahashi (2011) looked more closely at the development of modern modal constructions that make use of cà. Takahashi discovered that over the 13th to 20th centuries there was an increase in the use of cà

---

\(^{12}\) Uses of cà can be seen in §4.2.5
in the marking of modality. Table 14 shows only a selection of Takahashi’s (2011) but they show a gradual emergence of gà marked modal concepts still commonly used today. The increasing percentages in Table 14 comes from measurements of corpus data and represent a tracking of the grammaticalization of gà modal concepts.

**Table 14 Increasing modal functions of gà/càk (Takahashi, 2011)**

<table>
<thead>
<tr>
<th></th>
<th>1292-1438</th>
<th>1438-1782</th>
<th>1782-1925</th>
<th>1925-1978</th>
</tr>
</thead>
<tbody>
<tr>
<td>inevitability</td>
<td>14.00%</td>
<td>9.00%</td>
<td>24.00%</td>
<td>24.00%</td>
</tr>
<tr>
<td>prediction</td>
<td>11.00%</td>
<td>9.00%</td>
<td>13.00%</td>
<td>14.00%</td>
</tr>
<tr>
<td>appropriateness</td>
<td>0.00%</td>
<td>5.00%</td>
<td>7.00%</td>
<td>14.00%</td>
</tr>
</tbody>
</table>

The increasing of gà for modal concepts should also be compared with the decreasing use of the lexical verbal concept over the same time span. Takahashi (2011) also shows that there was also decreasing use of certain functions of gà/càk, the most sharply decreasing being the verbal ‘desire’ use which is consistent with the findings of Diller (2001). The decreasing verbal usages of gà/càk is shown in Table 15 below.

**Table 15 Decreasing verbal functions of gà/càk (Takahashi, 2011)**

<table>
<thead>
<tr>
<th></th>
<th>1292-1438</th>
<th>1438-1782</th>
<th>1782-1925</th>
<th>1925-1978</th>
</tr>
</thead>
<tbody>
<tr>
<td>volition</td>
<td>38.00%</td>
<td>54.00%</td>
<td>32.00%</td>
<td>24.00%</td>
</tr>
<tr>
<td>desire</td>
<td>21.00%</td>
<td>12.00%</td>
<td>7.00%</td>
<td>8.00%</td>
</tr>
</tbody>
</table>

While Table 14 and 15 are just selection of Takahashi’s (2011) findings, the contrast of increasing and decreasing frequency of both lexical and grammatical uses of gà shown displays the use of measuring frequency and also corroborates a key idea in grammaticalization theory. This idea is that certain uses of morphemes may decrease in frequency if new, more productive usages come into existence. As a gram is locked into a certain construction it is possible that its frequency for that function if it serves a function of high grammaticality, such as in the gà marked modality constructions. From Takahashi (2011), we can see how measurements of function from a diachronic corpus provides strong evidence for the grammaticalization of gà. Analyses like these suggests that gà grammaticalized along a path of ‘desire’ > intention > future > irrealis > (irrealis-complement form)
(Diller, 2001, p.159). And from Takahashi’s (2011) research, the further developments of cà in modal constructions can be added. The use of cà in epistemic modality constructions also corresponds to a cline, Future > Epistemic Modality. (Heine and Kuteva 2004, p.156).

Another example of Thai grammaticalization research that makes use of a diachronic corpus is Kullavanijaya’s (2008) study on the grammaticalization of the noun thîi ‘place’13. Bisang (1996) had already noted that the noun thîi ‘place’ has been grammaticalized into many functions, but Kullavanijaya (2008) investigated the actual development of thîi more closely over four historical periods to better delineate its grammaticalization pathway. Tracking the expanding functions of thîi through a diachronic corpus, Kullavanijaya (2008) found that thîi has grammaticalized into at least nine functions. The findings of Kullavanijaya (2008) are repeated in Table 16.

**Table 16 Occurrences of nine functions of /thîi/ in the four periods**

<table>
<thead>
<tr>
<th>Function/periods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sukhothai</td>
<td>✓</td>
<td>✓-</td>
<td>✓-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ayutthaya</td>
<td>✓</td>
<td>✓</td>
<td>✓-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-Ratanakosin</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Modern Thai</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

1 = a simple noun; 2 = a class noun; 3 = a compound; 4 = a preposition, 5 = a relative clause marker; 6 = a complementizer; 7 = a nominalizer; 8 = a classifier; 9 = a numeral marker; ✓- rarely occurs

Kullavanijaya hypothesizes that as thîi gradually lost its concrete meaning it was able to serve more functions, and while once it may have been a rare lexical item it can now be found in almost any modern text serving grammatical functions. These findings are consistent with the claims of Bybee (2003) (cf. §3.2) that as grams become more general they tend to become more abstract and are thus able to appear in new contexts, and are thus likely to be seen more frequently. From this analysis the grammaticalization path in Figure 9 below is proposed.

---

13 Uses of thîi can be seen in §4.3.3
Figure 9 Grammaticalization path of thîi (adapted from Kullavanijaya, 2008)

The present multi-functionality of thîi may help to explain why thîi is one of the most frequently found words in modern Thai corpora. As grammatical usages of thîi in Thai have developed in schematic constructions they have coexisted along with the lexical form which is found in other Tai languages. Kullavanijaya’s (2008) findings are also interesting in that the grammaticalization of thîi seems to provide an interesting basis for comparison with other Tai languages.

“The historical study above reveals that /thîi/ ‘land, a piece of land, place’ has a unique and long history in Thai compared to some other Tai dialects. The historical development does not differ much from the grammatical pathway given by Bisang except, significantly, in the area of a nominalizer and a complementizer. The historical study clearly and interestingly shows the gradual semantic development over time. The abstract meaning of /thîi/ has become more frequent in the last two periods. Besides, it is evident that the derivation of /thîi/ into syntactic function words has been late, that is, only in the last two periods. These syntactic function words reveal Thai sentences which have become more complex with embedding and modifying constructions. This is evident in comparison to the Tai dialects such as Shangshi Zhuang.” (Kullavanijaya, 2008, p.466)

Earlier, grammaticalization was defined as the change of lexical items and constructions in certain linguistic contexts to serve grammatical functions. Common research tools used in investigating grammaticalization are those made available through corpus linguistics as is often the case with Thai grammaticalization research. This section reviewed diachronic corpus data from Thai grammaticalization studies. In Diller (2001) and Takahashi (2011) historical and quantitative evidence was used to show the grammaticalization of cà and thîi. The case of cà fit with typical aspects of grammaticalization (e.g. phonetic reduction) and corresponds to well-attested clines. Kullavanijaya (2008) showed that the lexical noun thîi ‘place’ has developed several grammatical functions over a period of 700 years, with interesting implications for comparative Tai studies. This was done by

14 Based on Thai corpus data from Chulalongkorn University (2013) Top 5,000 word list http://ling.arts.chula.ac.th/TNC/contents/File/freq-5000.xls
measuring the use of grams used for certain functions. The type of research conducted on cà and thîi has been done for many other Thai grammatical morphemes which are presented in the next section.

3.4.2 Thai Grammaticalized Morphemes

The examples of cà and thîi presented in the previous section are but only two of many grams in the overall picture of grammaticalization in the Thai language. While Diller (2001) provides a good overview of this field of inquiry, since its publication there has been continued investigation regarding grammaticalization, and a current compilation of Thai grammaticalization includes many new findings. In preparation for examining grammatical aspects of Shan, a list of all grammaticalized Thai words other researchers presented was compiled. Besides helping to better understand the function of many Thai grams, these findings also have the benefit of serving as the basis for comparison with other Tai dialects, as outlined in the rationale for this thesis. Table 17 presents many of Thai grammaticalized morphemes that were investigated along with some information on their grammaticalization.

<table>
<thead>
<tr>
<th>Gram</th>
<th>Thai</th>
<th>Grammaticalization Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>?aw</td>
<td>เอา</td>
<td>V &gt; verb &gt; MOM</td>
</tr>
<tr>
<td>dâaj</td>
<td>ได้</td>
<td>V &gt; ability, success</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V &gt; past tense marker</td>
</tr>
<tr>
<td>hâj</td>
<td>ให้</td>
<td>V &gt; CAUS, BEN, PURP</td>
</tr>
<tr>
<td>cà</td>
<td>จะ</td>
<td>V &gt; IRR</td>
</tr>
<tr>
<td>jùu</td>
<td>อยู่</td>
<td>V &gt; Aspect ‘progressive marker’</td>
</tr>
<tr>
<td>kamlaŋ</td>
<td>กำลัง</td>
<td>N &gt; Aspect ‘progressive marker’</td>
</tr>
<tr>
<td>?àat</td>
<td>อาจ</td>
<td>V &gt; Modal ‘may’</td>
</tr>
<tr>
<td>khuan</td>
<td>ควร</td>
<td>V &gt; Modal ‘should’</td>
</tr>
<tr>
<td>tɔŋŋ</td>
<td>ต้อง</td>
<td>V &gt; Modal ‘must’</td>
</tr>
<tr>
<td>kwàa</td>
<td>กว่า</td>
<td>V &gt; Comparative ADV, Temporal ADV</td>
</tr>
</tbody>
</table>
In compiling Table 17 dozens of papers and theses covering several specific cases of grammaticalization in Thai were consulted\(^\text{15}\). Many of these cases show clear diachronic changes and developments into towards grammatical categories, with historical evidence similar to that which was presented in §3.4.1. Though there may still be concerns on how we think about grammaticalization within Tai languages, it was not the aim of the author to reexamine previous Thai grammaticalization research, but instead use it as the basis for comparison with a closely related language. The logic of the approach is that if each of these grams represent a element of change within the Thai language, then these grams make natural points of comparison with other Tai languages. The main focus of this studying is to find similarities and differences in Shan and Thai, by looking to grams that have gone through a diachronic change either from grammaticalization or through other factors.

### 3.5 Summary

A main goal of this chapter was to introduce the concept of grammaticalization and the parameters or steps in which it works. In introducing this area of linguistic inquiry, it is hoped enough attention has been brought to the role of comparative linguistic research in grammaticalization studies. The concept of *clines*, which are based the cross-linguistic evidence, provide an interesting context for comparative studies. Though grammaticalization is a powerful mechanism of language change it

<table>
<thead>
<tr>
<th>Gram</th>
<th>IPA</th>
<th>Thai</th>
<th>Grammaticalization Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>thùuk</td>
<td>ถูก</td>
<td>V &gt; Passive marking</td>
<td></td>
</tr>
<tr>
<td>nák-</td>
<td>นัก-</td>
<td>N &gt; NOM</td>
<td></td>
</tr>
<tr>
<td>khwaam-</td>
<td>ความ-</td>
<td>N &gt; NOM</td>
<td></td>
</tr>
<tr>
<td>nâa</td>
<td>น่า</td>
<td>V &gt; ADJVZR</td>
<td></td>
</tr>
<tr>
<td>thîi</td>
<td>ที่</td>
<td>N &gt; REL</td>
<td></td>
</tr>
<tr>
<td>khɔ̌ɔŋ</td>
<td>ของ</td>
<td>N &gt; POSS</td>
<td></td>
</tr>
<tr>
<td>wâa</td>
<td>ว่า</td>
<td>V &gt; Complement marker</td>
<td></td>
</tr>
</tbody>
</table>

\(^{15}\) See Appendix A for a fuller version of this table which provides more cases of grammaticalization as well as a list of citations for literature on each of the grams.
is of course not the only way in which languages change. The ways in which certain words or grams find their way into a language can be varied and due to several factors (See §3.3.2). Despite these concerns, it is the viewpoint of the author that the findings from grammaticalization studies on Thai are broad enough to raise several interesting questions, first on other related languages and on the very nature of grammaticalization for languages in this study (See §3.3.1). With these questions in mind, it is in the spirit of an exploratory cross-linguistic comparison that the following findings were ascertained.
Chapter 4
Verbal Phrase Syntax

4.1 Introduction
Verb phrases in Tai languages include at least a main verb along with other verbs or one of many possible pre-verbal or post-verbal auxiliaries. In general both pre- and post-verbal slots can include markers of aspect or modality, but with limitations on the placement of certain grams (e.g. negation is exclusively pre-verbal, and directional verbs as success markers are usually post-verbal see §4.2.7). The following sub-sections should give a basic idea of significant grams and grammatical areas within both Thai and Shan verb-phrase syntax.

Sections 4.1.1-4.1.4 are organized around the lexical source concept of a grammaticalized morpheme. The idea of a source concept refers to clines which are used in mapping universal or general grammaticalizations paths found across the languages of the world (see §3.2.1). These sub-sections show similarities in how source concepts (i.e. lexical material) have been grammaticalized in similar ways for both Shan and Thai. The findings in these sub-sections mostly show singular cognate grams shared by both languages, with the likely possibility that these grammaticalizations occurred in Proto-Tai and are still consistent across the larger Tai language family. Sections 4.2.5-4.2.8 cover specific areas of verb-phrase syntax. Including a higher number of comparable grams involved in areas of verb phrase syntax, these subsections show more differences between Shan and Thai than §4.1.1-4.1.4.

4.1.1 TAKE
In both Shan and Thai, verbs ʔāw/?aw (reflexes of Proto-Tai ʔaw° Pittayaporn, 2009) function in similar ways. As a main verb it can mean ‘take’, ‘accept’, or ‘want’, as shown in examples (28) and (29) where it is used to mean ‘take’.
28. (S) [Corpus:SW]

\[
\text{taŋ mot tě mí sāam kōj jāwko lūkí?n ?āw kōj núŋ}
\]
every all IRR have three basket CONN child take basket one
In total there were 3 baskets, and the child took only one.'

29. (T) [Jagacinski, 1992a, p.118]

\[
?áw kɛɛŋk àj chìkɛn̥¼k sì ?arɔ̀ j kwàa kɛɛn\dot{u}a
\]
take chicken curry PRT delicious COMPAR beef curry
'Take the chicken curry. (It) is more delicious than beef curry.'

The verb \(?áw\) 'take' is seen to move into a more grammatical function when it appears in certain constructions. Focusing on Thai, Jagacinski's (1992a) coverb construction [NP1 ?áw NP2 V X] or Post's (2007) similar "manipulated object marker" (MOM) highlight cases where ?áw gains a prepositional quality. This happens when ?áw and a second verb share the same noun phrase such as in (30). This more grammatical usage of ?áw is possible in both languages.

30. (T) [Corpus:NK]

\[
khâw kɔ ?áw tàkrâa wàj khâaŋ-nâa
\]
3SG CONN MOM basket put side-front
'He placed the basket on the front (of the bike).'</n30>

In (30) the manipulation of the object tàkrâa 'basket' arguably refers mostly to the second verb wàj 'to put'. An example from Shan (31) shows an object màakmâj 'fruit' also within the MOM construction. The use of the gloss 'MOM' in these constructions emphasizes the sense that these objects are already within the agent's possession and focus on the integrated activity conveyed by ?áw and the second verb.

31. (S) [Corpus:SL]

\[
?âw màakmâj pân khâw nân sāam kɔ̃
\]
take fruit give 2SG/PL that three CLF
'(He) gave the three of them some fruit.'

Both Jagacinski and Post agree on the grammatical nature of ?áw marked sentences like examples (30) and (31) above. "In the ?áw construction, ?áw appears to have lost its original verbal quality and instead has gained a prepositional quality of introducing a noun phrase that has a definite reference." (Jagacinski, 1992a, p.135),
and Post a “higher degree of VP-integration in ʔaw marked clauses, implying reanalysis of ʔaw as a grammatical marker” (Post 2007, p.130). The more grammatical usage is interpreted from the way ʔaw functions along with another main verb in the clause. In both Shan and Thai similarities are also seen in the ellipsis of the ʔaw-modified NP, shown in (32) and (33) where the use of ʔaw implies a manipulated object.

32. (T) [Corpus:MN]

ʔaw  spep  ma  che  luik  peeCLF

‘(He) wiped the pear (with his handkerchief).’

33. (S) [Corpus:SL]

ʔǎw  spep  sàu  lót  thip  mán

‘(He) put (the basket of fruit) on the bicycle.’

From the corpora, the usages of Shan ʔǎw and Thai ʔaw are nearly identical, with almost all usages occurring within the MOM construction. Usages as a lone main verb were very rare within the corpora, with ʔaw usually accompanied by a second verb. This finding shows ʔǎw/ʔaw as a significant grammatical marker for both Shan and Thai. Overall the frequency of ʔaw was 1.6% in the Shan corpus and 0.77% in the Thai one. At first the difference here seemed to be alarming as most grams in this study had very similar frequencies across the two languages. Yet ʔǎw/ʔaw were of the few analogous grams that appeared to have a large discrepancy in frequency between the two corpora. After some consideration, this difference may be due to the imbalance of overall size between the two corpora, but the average mentions per speaker where similar (Shan = 6.5 and Thai = 7). The similarity in average uses of speakers, could support a view that though ʔaw may interpreted as having the grammaticalized MOM function in Shan and Thai, it still is closely aligned to its lexical source as a verb ‘to take’, and was more likely to only be used in describe the events in the video that corresponding to ‘taking’. Further exploration of ʔǎw/ʔaw as a gram is required to better understand how grammaticalized it may in fact actually be.
4.1.2 ACQUIRE

As a main verb Thai dâaj means something like ‘obtain, get’. Enfield (2001; 2003) refers to words like dâaj and others similar words throughout Mainland Southeast Asian languages as ACQUIRE words. Their widespread existence in MSEA languages, and their multi-functional nature make them important in discussions on grammaticalization or areal and genetic inheritance (see § 3.3.2). Even as a main verb dâaj has a range of possible meanings when taken without context (see “indeterminateness” from Bisang see § 3.3.1) shown in glosses a-c in (34).

34.  (T) [Takahashi, 2008, p.121]

   (khāw) dāaj ŋən
   3SG ACQUIRE money

   a. ‘(He) got money.’
   b. ‘(He) came to have money.’
   c. ‘Money emerged (for him, and he got it)’

Takahashi (2008) constructed a chronological picture of the grammaticalization of dâaj through data from a historical corpus. Takahashi (2008) says the original meaning was something more akin to ‘emerge’ (like in gloss c above). From this original verbal concept dâaj grammaticalized pre- and post-verbal auxiliary functions in the Sukothai period, with these auxiliary usages becoming more frequent by the time of Rama III (Meesat, 1997). In Shan “dâaj” is realized as laj (the phonological pattern of a Thai initial /d/ being an initial /l/ in Shan is found across other items, see §2.2.3)16. Enfield’s work on ACQUIRE words Tai languages mainly draws from Lao, yet he notes, "The Tai language family is the only one for which the exponents of ACQUIRE are plausibly phonologically related in every case. Historical linguists have been able to uncontroversially reconstruct dâaj for Proto-Tai (Li, 1977, p.108). Every modern Tai language has a reflex of this word, and it generally shows the range of basic functions described for Lao.” (Enfield, 2003, p.321). Enfield (2004) holds that ACQUIRE words in Tai languages have at least the following three functions: a main-verb, a pre-verbal marker, and a post-verbal marker all of which were found within both the Thai and Shan corpora. Examples (35)-(37) are show uses of the Shan laj ACQUIRE.

16 Proto-Tai dâajC (Pittayaporn, 2009)
a transitive verb ‘come to have’

35. (S) [Corpus:SL]

ʔàn laj Ḛkwàŋ pən he
NEG ACQUIRE permission people PRT
‘(He) didn’t get their permission.’

a preverbal modal/aspectual marker (typically ‘get to’: or ‘have to’)

36. (S) [Corpus:KG]

ʔàn laj cɔj luk
NEG ACQUIRE help CLF.child
‘(She) didn’t help him.’

a postverbal modal/aspectual marker (typically ‘potential’ or ‘completive’)

37. (S) [Corpus:JY]

kɔ pɛn kɔn lak kwà laj
CONN be people steal go ACQUIRE
‘He was the person who stole it.’

The findings from the corpora are in line with Enfield’s analysis of Tai ACQUIRE with Shan laj and Thai dâaj fulfilling similar grammatical roles. Overall the frequency of Shan and Thai ACQUIRE verbs were similar within the corpora with Shan laj at 0.64% and Thai dâaj at 0.66%. Both languages used their respective ACQUIRE verb to fulfill similar grammatical functions with the majority of usages for both languages being the preverbal marker. The findings here confirms that presently Shan laj ACQUIRE covers roles also fulfilled by Thai dâaj, other Tai languages, and other MSEA ACQUIRE grams, all of which is covered extensively in Enfield (2003). The syntactic position and meanings of dâaj / laj ACQUIRE grams are pictured in Figure 9 adapted from a diagram on Lao dâaj in Enfield (2004, p.275).
Figure 10 Lexical/grammatical range of Tai ACQUIRE
(adapted from Enfield, 2004)

Figure 10 shows that Shan and Thai ACQUIRE words share a similar lexical and grammatical range, with the grammaticalized usages being those in the pre-verbal and post-verbal slots. While this section has established the overall similarities of both Thai and Shan ACQUIRE, there seems to be a unique function of Shan laŋ for marking obligative modality meaning ‘must’. This will be covered in section §4.2.6 which looks more closely at Thai and Shan pre-verbal TAM markers.

4.1.3 GIVE
As a main verb Thai ḃāj means ‘to give’, but in other contexts it has come to serve other grammatical functions (e.g. preposition see Matisoff (1991, p.437)). Writing on Lao, Enfield (2002) sees ḃāj as a predicate often used for valency-increasing operations that add extra arguments into a clause. Some identified grammaticalized uses of Thai ḃāj include a marker of benefactive, causative or malefactive clauses like in (39) – (41)\(^{17}\).

Verb ‘to give’

38.  (T) [Takahashi, 2012, p.127]
    khāw ḃāj tūkkataː dēk dēk
    3SG give doll children
    ‘He gave the children a doll.’

\(^{17}\) The grammaticalization of ḃāj is one of the most written about Thai grams. The benefactive, causative and malefactive functions are but three general descriptions of grammaticalized ḃāj. Check the Appendix A for a full list of works on the use and development of this word.
Causative
39. (T) [Iwasaki, 2008, p.475]
\[\text{hâj châŋ kháw thamgn} \]
CAUS worker 3SG work
‘(I will) have workers work.’

Benefactive
40. (T) [Iwasaki, 2008, p.472]
\[\text{dîaw cà bôk hâj ná} \]
soon IRR tell BEN PRT
‘(I’ll) tell you in a moment.’

Malefactive
41. (T) [Iwasaki, 2008, p.473]
\[\text{tôp náa hâj} \]
hit face MAL
‘(I) will slap your face.’

Using a historical corpus Takahashi (2012) looked at the historical semantics of hâj by looking at over 1,000 usages in Thai texts covering a period of 700 years. From this analysis, Takahashi sees the grammaticalization of hâj arising from the meaning of verbal hâj as not giving in the sense of representing ‘transfer of control/possession of a thing’ (as it is in English), but as meaning more of something like ‘transfer of position of something’. Iwasaki (2008) puts forth a model of how the grammaticalization of hâj is a result of clause serialization and reanalysis across sentences boundaries. Grammaticalized uses of hâj ‘to give’ for causative and benefactive clauses occur within certain constructions which still reveal the position of clausal reanalysis (Iwasaki & Yap, 1998b, p.372).

Benefactive/Malefactive: [clause] GIVE NP
Causative: NP GIVE [clause]

The grammaticalization of Thai hâj ‘to give’ is paralleled in developments of the Shan verb pǎn ‘to give’. Jenny (2012) notes the use of pǎn in some Shan varieties for causative and benefactive clauses. Examples (42)-(45) show findings from the corpora and and texts which confirm Jenny’s (2012) analysis for speakers of Southern Shan dialects, additionally example (45) adds a syntactically similar but semantically differentiated use of pǎn for malefactive clauses.
Verb

42. (S) [Corpus:SL]

ʔǎn ʔǎw màakmâj pǎn khǎw nǎn sǎam kǐ

CLF take fruit give 3SG/PL that three CLF

‘He took the fruit and gave it to those three people.’

Causative

43. (S) [Corpus:SG]

mǎn kɔ pǎn lukʔn nǎn jǔk màakmâj jǎwkɔ sǎu lôt thîp

3SG CONN CAUS child that lift fruit CONN put bicycle

‘He got the child to lift up the fruit and put it on the bike.’

Benefactive

44. (S) [Field Notes]

hǎw kɔ tè bɔ̀k law pǎn

1SG CONN IRR tell tell BEN

‘I will tell (a story) “for you”.’

Malefactive

45. (S) [Pham, 2006, p.136]

mǔ kɔ mà khóp pǎn kǐ nǎn kámlǒw

pig CONN come bite MAL CLF that immediately

‘The wild pig then came to bite that person immediately.’

Jenny (2012) points out that the original sense of pǎn being not ‘to give’, but something more akin to ‘to share’. This use and meaning of pǎn is still held by speakers of Yong a small but closely related Tai dialect in Northern Thailand18. Also mentioned in Jenny (2012) is the fact that Shan speakers use pǎn interchangeably with hâu, the Shan cognate of Thai hâj. Within the Shan corpus there were indeed some grammatical uses of both hâj and hâu by the Shan speakers19. An example of a Shan speaker used hâu ‘to give’ as a causative is shown in (46).

---

18 The name Yong coming from a city within Shan State. This use of pǎn is from my field notes. For more on Yong see Davies (1979); Tuwakham (2005).

19 Thai originally had both of these vowels, but a vowel merge of aɯ>aj has left Standard Thai with only /aj/. This fact is still reflected in the Thai script having two characters for /aj/. The use of hâj by a Shan speaker is not surprising when it serves the same function as pǎn, and the influence of Thai could effect the pronunciation of the vowel of hâu a vowel or be a direct influence of Thai on the speaker. The Proto-Tai form has been reconstructed as hauʔ in Pittayaporn (2009).
Interestingly both Shan speakers in the corpus who used hâj and háu also made use of pǎn. An excellent example of the possibility of this case can be seen in Young (1985) where the speaker use of both háu and pǎn in the context of a single sentence, shown in (47).

47. (S) [Young, 1985, p. 41]

khíŋ món-sāj waa mán paj mì háu jän pǎn thëŋ
Mr. soft.inner.clear say he yet have let that give another
lôn nuŋ month one
‘Mr. Môn-sai said that he didn’t have the money yet, to be patient, and to
give him another month.’

Young’s (1985) glossing of háu in (47) as ‘let’ captures the semantic shift away from ‘give’. More specifically this usage may fall into what Takahashi (2012, p.135) calls a “desiderative complementizer” and preceded the grammaticalization of the more typical benefactive usage. The sense of pǎn in (47) is definitely verbal but the possibility of a clear change of meaning and function of pǎn is seen when it is then used by the same speaker as in similar fashion as háu in (47) above in the following example (48).

48. (S) [Young, 1985, p. 43]

tsan pǎn khíŋ món-sāj khi nāj hón mán
indeed allow Mr. soft.inner.clear shit in house he
‘It certainly gives Mr. Môn-sai the right to shit in my house.’

Within the corpus Shan pǎn had a 0.64% frequency while Thai hâj had a 0.58% frequency. Additionally there were 3 token uses of háu as a causative marker by Shan speakers. Along with the similar overall frequency both corpora showed that non-verbal uses made up about 25% of the uses. Research using Thai historical corpora shows that there is a clear diachronic development of hâj grammatical functions (Takahashi 2012, Iwasaki 2008 and Chamniyom 2003; 2006). These findings provide strong evidence for a language internal grammaticalization of Thai hâj ‘to give’ into a semantically varied clausal marker. The variation we find in Shan
of pān, and hâj / háu on the other hand could be used as evidence for the grammaticalization of pān being an areal development due to typological poise Shan would have for constructions using a verb ‘to give’. Without historical evidence to speak on the development of Shan pān, the findings here can be concluded as only showing that Shan and Thai both use lexical verbs ‘to give’ to function in other similar grammatical constructions.

4.1.4 BE.AT

The word jù(u) in Shan and Thai both act as a main verb meaning ‘dwell; be on/at a place’ conceptualized by the author as BE.AT and as a post-verbal auxiliary indicating progressive aspect\(^\text{20}\). Sriprasit (2003) and Chamniyom (2006) showed that historically Thai jìu only had the verbal meaning in the Sukhothai period while the post-verbal aspectual meaning grammaticalized later in the Ayutthaya period. Writing on the Thai form, Tansiri (2007) considers the jìu to be a static imperfective marker in its grammaticized form. This cline of grammaticalization a verb ‘dwell/be at’ > continuative/progressive is a common feature of other SEA languages (Matisoff, 1991). Shan use of jù as a main verb in (49) and as a post-verbal auxiliary in (50) and (51) parallels aspectual usages in Thai.

49. (S) [Corpus:JY]
   mán jù náu tonmâj
   3SG be.at in tree
   ‘He is (up) in the tree’

50. (S) [Corpus:JY]
   mí kòn kép luk koko jù
   have people gather CLF cocoa PROG
   ‘There are people gathering cocoa.’

51. (S) [Corpus:SG]
   kɔ pôn kwâ kwe-na pê nâj kai kîn jù
   CONN pass go because goat this want eat PROG
   ‘He just passes by because his goat was wanting to eat the fruit’

\(^\text{20}\) The different forms here of Thai jìu and Shan jù come from vowel length in Shan for /u/ is non-contrastive. The Proto-Tai form has been reconstructed as "juu" in Pittayapornt (2009).
Analysis from the corpora revealed Thai จู appeared with a 0.79% frequency, and Shan จุ with a 0.77% frequency. Of these uses Thai จู appeared as a post-verbal modifier 37% of the time in the Thai corpus, and 29% of the time in the Shan corpus. The lexical uses of จุ(u) as a lexical main verb and as a marker of progressive aspect are the same for Shan and Thai. Figure 11 shows a conception of the lexical and grammatical functions of จุ(u) for both Shan and Thai.

Figure 11 Functions of Shan and Thai verb BE.AT

4.1.5 Irrealis

The Thai irrealis marker ค่ะ has grammaticalized from a verb meaning ค่ะ ‘to intend, consider’ (Diller 2001; Takahashi 2011 covered in §3.4.1) to now occur in constructions marking future events and other aspectual/modal concepts. The Shan irrealis marker 低迷 functions in similar ways and occurs in similar constructions as the Thai irrealis marker ค่ะ. Examples (52) and (53) show both languages’ use of irrealis marking for future frames of reference.

52. (T) ชาย ค่ะ คิน กล้วย
   1SG IRR eat banana
   ‘I will eat banana.’

53. (S) [Elicitation]
   ชาย 低迷 คิน ก้อย
   1SG IRR eat banana
   ‘I will eat banana.’

The similarities of Shan and Thai irrealis marking are not limited to just indicating future tense, but also in how they are used along with pre-verbal auxiliary elements (several of which are discussed in §4.1.6). In Thai /ʔat ค่ะ/ is an epistemic modal auxiliary that conveys a low-level of confidence in attainment/accomplishment of a
state/activity usually glossed as ‘probably, may, could’. In examples (54) and (55) a similar construction of [‘may’ + IRR] is used by both Shan and Thai for the same modal concept.

54. (T) [Matisoff, 1991, p.393]
    ŋàat cà māj paj
    may IRR NEG go
    ‘probably won’t go’

55. (S) [Corpus:SG]
    ʔaap të pēn màak màak-kam-khoŋ hu  man ko màak chomphu
    may IRR be fruit apple or 3SG CONN fruit rose.apple
    ‘It might be apples or rose apples’

Another instance where të and cà also participate in similar pre-verbal marking is in an inceptive aspectual construction [[CONT+IRR]V], like in (56) and (57)

56. (S)
    hàw tūk të kīn koj
    1SG CONT IRR eat banana
    ‘I am about to eat a banana.’

57. (T) [Iwasaki & Ingaphirom, 2005, p.159]
    mɛɛ kamlàŋ cà nɔɔn
    mother CONT IRR sleep
    ‘Mom is about to sleep’

The corpora revealed similar frequency of both irrealis morphemes in each language. In the Shan corpus të ‘IRR’ had a frequency of 0.93% while cà ‘IRR’ had a 0.87% frequency in the Thai corpus. It should be mentioned that there were 7 cà tokens that appeared in the Shan corpus 3 of which were used with other pre-verbal elements like the examples above. The appearance of cà in the Shan corpus could possibly be due to influence of Thai on individual Shan speakers, but more investigation could be done into a possible coexistence of two irrealis markers in Shan, as is the case with Lao which has both cà and siø (Enfield 2007, p.215).

Comparison of Shan and Thai irrealis marking showed that të and cà are analogous grammatical markers, and overall the grammaticalized functions of Thai cà were mirrored by Shan të. This was shown in marking of future frames of reference, pre-verbal auxiliary constructions and additionally with similar frequencies in the corpora.
4.1.6 Pre-verbal TAM markers

Pre-verbal auxiliaries for both Shan and Thai include several possible markers of aspect and modality. Previous research into many Thai auxiliaries has revealed that they are the result of grammaticalization from full verbs. In this section Thai grammaticalized auxiliaries are compared with corresponding Shan grams which perform the same or similar function. One result of this comparison shows that though Thai and Shan share many analogous pre-verbal markers, Shan has not been subject to the same areal influences as Thai which has been influenced from Khmer. Other findings in this section bring out differences in Shan pre-verbal TAM markers.

The Thai CONT marker kamlaŋ is grammaticalized from a Khmer noun meaning ‘strength’ (Huffman, 1986, p.200). Previously in §4.1.5 examples with a CONTINUOUS gram, Shan tɯ̂k and Thai kamlaŋ were shown to occur in similar constructions as in the case of the inceptive construction [[CONT+IRR]V] in (56) and (57). Further supporting the analogous status of both of these grams is that they can mark continuous aspect on their own simply in the pre-verbal position as in (58) and (59).

58.  (T)
    chǎn kamlaŋ kǐn klûaj
    1SG CONT eat banana
    ‘I am eating a banana.’

59.  (S) [Elicitation]
    hǎw tɯ̂k kǐn koj
    1SG CONT eat banana
    ‘I am eating a banana.’

From their similar functions as a pre-verbal marker and within constructions with BE.AT and irrealis markers kamlaŋ and tɯ̂k both appear to be analogous grammatical morphemes. Measurements of frequency from the corpora also support the equivalence of these grams with both of them appearing at similar rates, kamlaŋ with 0.31% and tɯ̂k with 0.35%. Knowing the lexical origin of kamlaŋ and seeing the functional similarities of the equivalent Shan gram raise questions on the possible lexical origins of tɯ̂k21 along with the distribution of other CONT grams in other Tai languages.

---

21 Moeng (1995) provides an adverbial meaning of “denoting that a condition is adequately satisfactory”
The gram “ʔàat originally a verb ‘to be capable’, ‘to have power’” (Enfield 2007, p.212) comes from Khmer (Huffman, 1986). From historical corpus data, this word has been shown to be grammaticalized from a full verb into an auxiliary verb marking epistemic modality (Meevat, 1997). As shown earlier in §4.1.5, ʔàat was in a modal construction along with the Thai IRR marker cà in (54). A Shan example (55) showed similar construction of ['may'+IRR] with the word for ‘may’ ʔàap which seems to be most likely from the same the Khmer loan ʔàat which has been grammaticalized in Thai. Another pre-verbal marker in Thai which is based off of a Khmer loan is khuan which originally was a verb ‘to be appropriate’ but is now a deontic modality marker meaning ‘should’, also shown to be grammaticalized through the use of historical corpora (Meevat, 1997). Notice that this is often accompanied by cà IRR, like in (60)

60. (T) [Iwasaki & Ingaphirom, 2005, p.136]

khuan  cà  chuan khâw maa  dûaj
should IRR invite 3SG come also
'We should also invite him to come.'

For a modal meaning ‘should’ in Shan we find similar use of irrealis marker in the construction ['should'+IRR]MOD but again this is a case of Shan using a different morpheme thûk which does not come from Khmer, as in (61).

61. (S) [Elicitation]

kâw thûk  tê  kwâ waa mân
1SG should IRR go say 3SG
'I should go tell him.'

An interesting pattern is revealed when we see that Shan has been shown to have similar grammatical constructions as Thai; an inceptive construction [[CONT + IRR]V], epistemic modality ['may' + IRR], and a deontic modality construction ['should' + IRR]. Yet the Thai constructions are all done with Khmer loanwords, while Shan makes use of apparently non-cognate forms.

Differences in Shan and Thai pre-verbal TAM markers are not only limited to just areal influences. One example of where Shan differs in pre-verbal TAM marking constructions is in deontic modality markers signifying a strong obligation. Thai uses tɔŋ a pre-verbal auxiliary, as in (62), which is grammaticalized from a verb ‘to touch’. This differs from the previous pre-verbs in that it is not a Khmer loan and is a language internal development.
For Shan, one way of marking strong obligation like ‘must’ is through with two morphemes lò-laj ‘must’, like in (63). This example is from a field recording but I also received this form through direct elicitation from Shan speakers.

The morpheme laj was covered in §4.2.2 and falls under Enfield’s classification of Southeast Asian ACQUIRE words. This morpheme lò is originally a Burmese verb ‘to need’, it can function as a verb ‘to need’ in Dai Lue\(^{22}\). Examples (64) and (65) from Pham (2006) shows lò as a possible pre-verbal marker or lexical verb within a serial verb construction without the use of laj. The grammaticalization of most auxiliaries probably arises out of reanalysis within serial verb constructions.

The lò-laj construction is interesting when we consider that this is a possible construction within Lao [‘must’ + ACQUIRE] (“tòŋ often co-occurs with the

\(^{22}\) Hanna (2012)
achievement marker *daj*” (Enfield, 2007, p.223)). While it seems Burmese has affected Shan with the introduction of the morpheme *lò* another factor here is the effect Burmese may have had on the meaning/function of Shan *laj* ACQUIRE. Jenny (2009) holds that Burmese ACQUIRE\(^{23}/yá/\) has gone through grammaticalization to become mark obligatory modality. Jenny sees this development in Burmese as having influenced Shan *laj* in modal constructions like in (66). Jenny (2009) says that this use is often done along with the irrealis marking as well which is consistent with many of the other examples of pre-verbal markers mentioned above.

66. (S) [Jenny, 2009, p.123]
\[
\text{tě laj hau kôn pĕn náŋ wắj}
\]
IRR must CAUS person be (.sick) lie down keep
‘You must let the sick person lie down.’

The influence of Burmese on Shan is well known and the additional grammaticalization of Burmese ACQUIRE gives grounds to suppose a parallel shift in Shan ACQUIRE. These developments are also consistent with Bisang (2008, p.19) who holds that a modal interpretation: ‘must, have to’ is a possible inference of ACQUIRE words. Interestingly this interpretation of Shan *laj* is specifically addressed in Enfield (2003, p.325) a footnote on examples (67)-(68) from Young (1985) says, “Young’s text gives *tĭ-lắj*, glossed as ‘must’. However, since this is morphologically transparent (cf. the negated version in (67), I prefer to gloss it as given here.” The examples Enfield speaks of are given below.

67. (S) [Enfield, 2003, p.324]
\[
\text{khì lê kṓj tĕ am laj páa jeu}
\]
shît PRT but IRR NEG ACQUIRE accompany piss
‘Go ahead and shit. But you must not piss at the same time.’

68. (S) [Enfield, 2003, p.325]
\[
\text{waa kắw tĕ laj kāa ti lum kắmlaw ņắj}
\]
say 1SG IRR ACQUIRE go place govt. office time sole PRT this
‘(It) says I must go to the government office right away.’

Yet, from the influence from Burmese discussed in Jenny (2009) and evidence provided from comparisons in this study it seems probable that semantics of Shan ACQUIRE has expanded to included a modal concept ‘must’. This seems even more

\(^{23}\) Jenny (2009) refers to these words with GET, I have changed this to Enfield’s ACQUIRE for consistency within this paper.
probable with the lack of the 遹 or any other marker meaning 'must'. The similar grammaticalization of ACQUIRE in Burmese would also explain areal influence for this development in Shan especially when a Burmese lexical item lô 'to need' had already found its way into Shan obligatory marking. Whether the lô-laj construction or the tê-laj, or just laj is more common or other factors that might affect usage here is not known. Grammaticalization of Shan laj would expand Shan ACQUIRE into areas which do not correspond to Thai. This development is shown in Figure 12 should be compared with Figure 10 in §4.1.2.

Figure 12 Development of obligative function of Shan ACQUIRE pre-verbally

So far findings in this section show that Shan and Thai differ in pre-verbal markers for the marking of aspect and modality. A key difference here was the differing areal influences on each of these languages. It is well-known that Thai is influenced from Khmer as we saw with kamlaŋ, ʔàat and khuan. The influence of Burmese on Shan was seen in the case of modality indicating strong obligation meaning 'must'. Compared to the similarities presented in previous pre-verbal markers the pre-verbal TAM markers shows an area of greater differences between these two languages. Though many Thai grams here are the result of an areal influence the case of 遹 showed the difference a language internal grammaticalization in Thai not appearing in Shan. Another example like this would be Thai the preverbal marker khoŋ, shown in (69), 'probably' which grammaticalized from a full verb in Thai (Meesat, 1997) is also absent in Shan.

69. (T) [Corpus:MN]
khāw khoŋ cà jàak hâj tàkrâa thūi sǎam man tem
3SG MOD IRR want CAUS basket NUM three 3SG full
‘He most likely wanted to make the third basket full as well.’
Shan seems to have its own unique pre-verbals and a different construction. In Young (1985) kɯŋ-li is used as a preverbal marker meaning 'ought to', as in (70) and (71). Moeng (1995) lists an entry for kɯŋ as 'be suitable, fit, proper', but the use in (70) and (71) it appears with the word lǐ meaning 'good'. There is not an obvious Thai equivalent to this Shan construction.

70. (S) [Young, 1985, p.132]

\[
\text{háw am kɯŋ-li jù ti cam kón}
\]

1SG/PL NEG ought be.at place near people

'We ought to not live near people.'

71. (S) [Young, 1985, p.132]

\[
\text{háw kɯŋ-li kaa jù ti kaj kaj}
\]

1SG/PL ought go be.at place far-far

'We ought to go live far away.'

In Table 18 the pre-verbal markers covered in this section are shown. The equivalency of the modal markers could benefit more investigation. Further research would have better understand and delineate the semantics of the Shan modal markers, yet it is clear that pre-verbal auxiliaries which are grammaticalized from verbs differ in these languages, and are a rich area of difference in regards to grammaticalization. The MOD marker nāa will be covered in §5.1.2.

Table 18 Comparison of different Thai and Shan TAM markers

<table>
<thead>
<tr>
<th>Thai</th>
<th>Shan</th>
<th>Function/ Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>cà</td>
<td>tě</td>
<td>IRR/FUT</td>
</tr>
<tr>
<td>kamlaŋ*</td>
<td>tûk</td>
<td>CONT</td>
</tr>
<tr>
<td>?àat*/khoŋ</td>
<td>?àap</td>
<td>MOD 'may', 'probably'</td>
</tr>
<tr>
<td>khuan*</td>
<td>thûk / kɯŋ-li</td>
<td>MOD 'should, ought'</td>
</tr>
<tr>
<td>tŋ</td>
<td>(lò) laj</td>
<td>MOD 'must, have to'</td>
</tr>
<tr>
<td>nāa</td>
<td>---</td>
<td>MOD 'should'</td>
</tr>
</tbody>
</table>

*Khmer loans

4.1.7 Directional Verbs

Within Tai languages the class of lexical verbs known as “directional verbs” has been noted to also serve grammatical functions. Thepkanjana & Uehara define directional
verbs as “verbs which denote movements described in terms of their directionality with respect to a landmark, being directed towards or away from it” (2008, p.484). Common directional verbs are those that refer to movements go, come, ascend, descend, enter, and exit. And while all of these verbs can act as main predicates, in their grammaticalized form they can be seen as resultatives or success markers, or as markers of ablative or allative case (Enfield 2007, p.232). Thepkanjana & Uehara (2008) see this grammaticalized usage as normally occurring in the construction [NP1 V NP2 DV]. A typical construction includes a transitive verb, a direct object and then a directional verb, as in example (72).

72. (T) [Thepkanjana & Uehara, 2008, p.490]
   chān ?aːn laajmuu kʰɔŋ khun mâj ʔɔ̀ɔk
   1SG read handwriting POSS 2SG NEG exit
   ‘I tried to read your handwriting but was not successful.’

In (72) ʔɔ̀ɔk ‘exit’ is an example of a directional verb being used as a success marker. These success markers are used to mark new information, that modify sentences when needed pragmatically. Verb phrases that have success markers imply that the agent’s intention has resulted in a (non)successful completion of an action. Thepkanjana & Uehara (2008) characterize this grammaticalization path as directional verbs > directional markers > success markers, but admit that these words continue to still have all of these functions. To help fully appreciate the reanalysis of directional verbs one must first grasp the prolific use of multi-verb or serial verbs constructions and the polysemy of these verbs in Tai languages. While directional verbs post-negation could be taken quite literally as in (73), their appearance in varied and metaphorical contexts shows their expanded functional ability, such as in (74).

73. (T) [Thepkanjana & Uehara, 2008, p.496]
   khǎw jók kʰɔŋ mâj khûn
   3SG lift box NEG ascend
   ‘He tried to lift the box but was not successful.’

74. (T) [Elicitation]
   chān mâj khǎw caj
   1SG NEG enter-heart ‘understand’
   ‘I don’t understand.’

With (73) the lifting of the box, the normal expectation that a box is lifted up is negated with mây khûn ‘not ascend’. The semantic factors that motivate these types
of usages rely on our cognitive conception of spatial orientation, and thus get expressed in our language (often these examples occur in the negative, negating the expectation of the transitive verb). This use of directional verbs has been covered for Thai, but from this research we can see that Shan directional verbs also fulfill similar grammatical functions, as in (75) where someone has “come” to a state of bigness and tallness.

75. (S) [Field Notes]
   pēn ṭān somsā tē tē se laj lēŋ jāu mā ka sūŋ mā
   be CLF grateful really PRT able raise big come value tall come
   ‘I am really grateful to have raised my son to be big and tall.’

Shan and Thai both use these directional verbs in the same manner: most of their uses are in serial verb constructions where the movement is very literal (e.g. go + up + pick fruit). Yet examples like (76) and (77) show cases where use of directional verbs in serial verb constructions show the context for potential reanalysis as a success marker.

76. (S) [Corpus:JY]
   mān pīn khuun pāj ṭāw
   3SG climb ascend go take
   ‘He climbed up (the tree) to take (fruit).’

77. (T) [Corpus:MN]
   pīn tōnmdāj kẖūn paj kēp
   climb tree ascend go gather
   ‘He climbed the tree to gather (fruit).’

Besides the overall similarity in general directional verb usage, one interesting thing to point out here is that the Proto-Tai ‘to go’ kwā is grammaticalized as a comparative marker and temporal conjunction in Thai (Diller 2001, p.154), yet remains completely lexical in Shan. In (78) the height of one boy is compared to another.

78. (T) [Corpus:MJ]
   dēk phūuchaaj khon nuŋ sūŋ kwāa ṭīk khon nuŋ
   child male person one tall COMPAR another person one
   ‘One boy was taller than the other.’
This usage is not possible in Shan which like many other Tai dialects uses other words used for marking comparison (Morev, 1998, p.97). In examples of Shan (79) and (80) lə̌ a verb 'to exceed' is used to make comparisons.

79. (S) [Pham, 2006, p.121]
   kɔ̀ jëwə̌ lə̌ hàw
   CLF big COMPAR 3PL
   ‘... a person bigger than them...’

80. (S) [Pham, 2006, p.138]
   pò pèn pə́n tè ka lə̌ nàn cikkú waa nàn
   if be other.people IRR value COMPAR that think COM that
   ‘If (the car) belonged to other people it would be more expensive...’

Another comparative marker can be found in Morev (1998) which puts forth se as a the Southern comparative marker, in (81). This is grammaticalized from a verb 'to separate'. An example of this usage can also be found in Young (1985) shown in (82).

81. (S) [Morev, 1998, p.98]
   man lì-gaam se saaw taj
   3SG beautiful 'separated' girl Tai
   ‘She is more beautiful than Tai girls.’

82. (S) [Young, 1985, p.94]
   kàw ə́ catj pen ?àn mì kaa khàn ?àn lə̌m ə́ŋ se
   1SG PRT indeed be CLF have value-rate CLF important COMPAR
   sù tày sìŋ
   2SG all
   ‘I am the one who is certainly more valuable and important than any of you.’

Both of these forms were elicited from Shan speakers and neither are the result of a grammaticalization of a directional verb. This is different from Thai which features a completely grammaticalized Proto-Tai kwàa in that its lexical content has been bleached with most Thai speakers not knowing its lexical origins - a case which is different from many other directional verbs which still can be seen as lexically transparent.

Another difference between Shan and Thai with respect to directional verbs is that Thai uses khàw 'to enter' and òɔk 'to exit' for to describe some cognitive states such
as 'understanding, thinking, legibility' where Shan uses a verb pəŋ 'to pierce', as shown in (83) and (84). In Moeng (1995) pəŋ is listed as having the following meanings “reach a place; have a use; mean (as a meaning); understand, comprehend; negotiate understanding” and is found in several idiomatic constructions.

83. (T) [Elicitation]
   chân mājkhāwcaj
   1SG NEG enter-heart 'understand'
   ‘I don't understand.’

84. (S) [Pham, 2006, p.123]
   hāw kōpəŋcāuçaŋnāj
   3PL CONN pierce.heart 'understand' like.this
   ‘We understood (it) like this.’

Though Shan does not use a directional verbs for these constructions, the underlying metaphor seems to be similar with Thai with cognitive states being compared to an enclosed space, understanding or comprehension is achieved by movement through “piercing” the barrier. Example (85) should be compared with the Thai example (72) at the beginning of this section, and a Thai version of (86) would similarly use ʔɔ̀ɔ 'exit'.

85. (S) [Elicitation]
   hāw làanlajmu su ʔām pəŋ
   1SG read handwriting 2SG NEG pierce
   ‘I can't read your handwriting.’

86. (S) [Elicitation]
   hāw won ʔām pəŋ
   1SG think NEG pierce
   ‘I can't think of anything.’

The findings in this section show that some Shan and Thai directional verbs can be said to function in similar ways as success markers. The Proto-Tai verb kwāa is a lexical verb 'to go' in Shan, but has been completely grammaticalized in as a comparative marker in Thai. For Thai uses of directional verbs which refer to a metaphorical cognitive space using enter and exit, Shan makes use of a verb pəŋ 'to

---

24 Uses of directional verbs like in examples (83) and (84) should also be understood as cases of “lexicalization” see Brinton & Traugott (2005).
pierce’. The directional verbs for each language along with notes on their grammaticalize functions is shown in the Table 19.

**Table 19 Comparison of different Thai and Shan directional verbs**

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Shan</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>go</td>
<td>kwâa*</td>
<td>paj</td>
</tr>
<tr>
<td>come</td>
<td>mä</td>
<td>maa</td>
</tr>
<tr>
<td>ascend</td>
<td>khun</td>
<td>khûn</td>
</tr>
<tr>
<td>descend</td>
<td>lôŋ</td>
<td>lôŋ</td>
</tr>
<tr>
<td>enter</td>
<td>khaw**</td>
<td>khâw</td>
</tr>
<tr>
<td>exit</td>
<td>?ôk**</td>
<td>?ôk</td>
</tr>
</tbody>
</table>

*Comparative marker in Thai
** Not used as success markers in reference to metaphorical cognitive space

**4.1.8 Passives**

Prasithrathsint (2006) outlines the gradual development of the Thai passive marker thûuk. Originally thûuk was a lexical transitive verb meaning ‘to touch’, ‘to hit off exactly’ or ‘to hit on the point’, but Prasithrathsint has tracked its grammaticalization from a lexical verb to a full neutral passive marker in the present. A pair of active and passive sentences are shown in (87) and (88) to illustrate the interpretation of thûuk as a passive marker.

87. (T)
   mää kät  dëk
   dog   bite   child
   ‘The dog bit the child.’

88. (T)
   dëk  thûuk  mää kät
   child  PSSV  dog   bite
   ‘The child was bitten by the dog.’

From elicitations it was found that Shan passives are possible with the verb ndâa ‘to meet, encounter’, like in (89) which is similar to (88).

25 See more in Prasithrathsint (2006) for a review on the controversy as to whether the category of “passives” are appropriate to some Asian languages.
89. (S) [Elicitation]

\[\text{lukʔn } \text{náa } \text{māa khóp}\]

child PSSV dog bite

‘The child was bitten by the dog.’

More investigation revealed though that while th̀uk is can be considered a neutral passive which can occur in contexts with favorable or unfavorable meanings, like in (90) the Shan passives marker náa is only an adversative passive marker, which is used only with sentences with unfavorable meanings or difficult contexts, like in (89) above, or in (91).

90. (T) [Prasithratsint, 2006, p.118]

\[\text{bān } \text{lāŋ nī th̀uk } \text{sāŋ pī } \text{thī lɛ́ɛw}\]

house CLF this PSSV build year last

‘This house was built last year.’

91. (S) [Elic]

\[\text{mān } \text{náa } \text{pō } \text{mān } \text{shā-la}\]

3SG PSSV parents 3SG scold

‘He was scolded by his parents.’

The Shan náa passive marked sentences examples from above are the result of direct elicitation of Thai forms into Shan. A naturally occurring instances of náa as a passive marker can be seen in (92).

92. (S) [Panglong]²⁶

\[\text{kōn } \text{waan } \text{kĪ } \text{náa } \text{cāŋ } \text{hēt } \text{hāaj } \text{tō } \text{wàat } \text{jēp}\]

people village CONN PSSV elephant do very wound wound tread

\[\text{tāaj } \text{tāj } \text{mōt } \text{hāa } \text{kĪ}\]

die all altogether five CLF

‘People of the village were gravely injured by an elephant, all together five people died.

Interestingly in its active verbal sense náa was often found as a verb ‘to meet, encounter’ within the context of an adversarial situation such as in (93) and (94).

93. (S) [Pham, 2006, p.132]

pig that CONN turn.around return enter come in clearing all

meet into group group that

‘The wild pig turned around and ran back [returned] into the clearing, bumped into that group of people.’

94. (S) [Corpus:ON]

at.this.time turn.look return turn.around look hat CLF REL fall

go crash meet rock

‘At this time (they) looked back at the hat of the boy who crashed into the rock.’

Prasithrathsint (2006, p.123) holds that a stage in the development of the Thai neutral passive was that of ฤๅษี occurring within serial verb constructions where the verb that followed negatively affected an animate agent. Prasithrathsint believes that passive markers in other Southeast Asian languages may have developed in same way. An important takeaway from Prasithrathsint's analysis here is the role that serial-verb constructions play. Chains of serial verbs provide a context for reanalysis and decategorization of lexical verbs can lead to grammaticalization of passive markers. While it is likely that ฤๅษี is not a full passive Prasithrathsint predicts that in southeast Asian languages lexical verbs with closely related meanings to ‘to contact, to undergo (an unfavorable experience)’ may fulfill this role. This seemingly cross-linguistic possibility suggests a possible cline for passive markers, along with indicators as to their level of grammaticalization.

“Evidence from some other Southeast Asian languages that do not have passive constructions also supports the grammaticalization process. Since items at later stages imply those at earlier stages, we can formulate a theory for Southeast Asian languages that if a language has a neutral passive marker, it is likely to have an adversative passive marker and a lexical verb meaning 'to incur' and one meaning 'to hit off exactly.' On the other hand, if it does not have a passive, it may have a lexical verb with related meaning” (Prasithrathsint 2004, p.596)
Findings here on the Shan ɲáa support Prasithrathsint’s (2006) comments on passives. Thai thùuk was absent in the Thai corpus, and while the occurrence of ɲáa in the Shan corpus was limited to 9 tokens these were not within a passive-type construction. Figure 13 below illustrates the grammaticalization of Thai thùuk from a lexical verb to a full neutral passive, and shows that status of ɲáa as a passive as still questionable.

\[\text{Figure 13 Shan and Thai passive markers}\]

4.2 Summary
Comparison of grams in verb-phrase syntax revealed similarities and differences in Shan and Thai. Similar in form and function, a group of lexical cognates ʔǎw/ʔaw ‘take’, dâaj/laj ‘get’, and ju/jùu ‘be.at’ performed the same grammaticalized functions in both languages. The verbs ‘take’ function as a “manipulated object-marker” where a noun-phrase with a definite reference is modified by a second verb. In a post-verbal position, aspectual markers ju/jùu modify main verbs indicating a progressive aspect. Besides an extra modal concept, the functions of Shan laj were consistent with those identified for Thai dâaj and other ACQUIRE grams.

Several Shan grams were found to be non-cognate forms of the corresponding Thai grams. Grammatical uses of Shan pǎn ‘to give, to share’ corresponded to grammaticalized uses of Thai hāj ‘to give’, despite Shan speakers already using cognate hâu ‘to give’ for similar grammatical functions. The Shan irrealis marker tě marks similar modal concepts as Thai’s grammaticalized irrealis marker cà. A clear factor in the differences of pre-verbal TAM markers was the effect of areal influences. Uses of Shan tûk are analogous to functions and constructions of the Thai pre-verbal continuous marker kamlaj – a Khmer loan. Also, Thai modal markers derived from Khmer like ʔàat and khuan are not used in Shan. The relevance of areal factors was illustrated further by expansion of Shan laj into a modal marker meaning ‘must’, a development which parallels grammaticalization of the ACQUIRE gram in Burmese. The Burmese influence here was evidenced further
by the existence of a direct Burmese loan lò ‘to need’ also used in Shan modal marking indicating obligation. The corresponding Thai gram th̀ 'must' like other Thai pre-verbal TAM markers không and nāa – all grammaticalized from full verbs – were not shared by Shan.

While the general use of directional verbs as success markers seem to be similar in both languages, in some areas Shan makes use of other lexical material where Thai uses grammaticalized directional verbs. In the metaphorical marking of cognitive space, Shan uses a verb pò ‘to pierce’ instead of khâw ‘enter’ and ṭɔk ‘exit’. Another difference in directional verbs was found in their use as a comparative markers. Where Thai kwàa, has completely grammaticalized from Proto-Tai ‘go’ (still lexical in Shan) into a comparative marker, Shan uses verbs lò ‘to exceed’ or se ‘to separate' instead. A final difference in Shan and Thai verb-phrase syntax was in the use of passive markers. Evidence seems to suggest that the Shan verb júa ‘to meet’ can serve as an adversative passive marker, whereas Thai thìuk has grammaticalized into a neutral passive. On overview of the comparison of Shan and Thai verb-phrase grams are presented in Table 20.
Table 20 Summary comparison of Shan and Thai verb-phrase grams

<table>
<thead>
<tr>
<th>Thai</th>
<th>Shan</th>
<th>Function</th>
<th>Frequency</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>?aw</td>
<td>?āw</td>
<td>MOM</td>
<td>0.83%</td>
<td>1.60%</td>
</tr>
<tr>
<td>dāaj</td>
<td>laj</td>
<td>ACQUIRE</td>
<td>0.66%</td>
<td>0.64%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shan laj has a modal function.</td>
</tr>
<tr>
<td>hâj</td>
<td>pān</td>
<td>CAUS/BEN</td>
<td>0.64%</td>
<td>0.58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>About 25% of uses in both corpus were non-verbal (e.g. BEN/ CAUS).</td>
</tr>
<tr>
<td>cà</td>
<td>tē</td>
<td>IRR</td>
<td>0.87%</td>
<td>0.93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Some use of cà by Shan speakers.</td>
</tr>
<tr>
<td>jîu</td>
<td>jù</td>
<td>ASP</td>
<td>0.79%</td>
<td>0.77%</td>
</tr>
<tr>
<td>kamlaj</td>
<td>tûk</td>
<td>ASP</td>
<td>0.31%</td>
<td>0.35%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thai form from Khmer</td>
</tr>
<tr>
<td>?àat</td>
<td>?àap</td>
<td>MOD</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thai form from Khmer</td>
</tr>
<tr>
<td>khuan</td>
<td>thûk</td>
<td>MOD</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thai form from Khmer</td>
</tr>
<tr>
<td>tôŋ</td>
<td>lô-laj</td>
<td>MOD</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Influence from Burmese on Shan laj and in loan of the verb lô 'to need'.</td>
</tr>
<tr>
<td>khâw</td>
<td>pûŋ</td>
<td>ASP</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shan uses a verb pûŋ 'to pierce' instead of directional verbs.</td>
</tr>
<tr>
<td>?ôk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kwàa</td>
<td>lô</td>
<td>COMPAR</td>
<td>0.08%</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>se</td>
<td></td>
<td></td>
<td>Thai comparative kwàa is lexical verb 'to go' in Shan.</td>
</tr>
<tr>
<td>thùuk</td>
<td>ñáa</td>
<td>PASS</td>
<td>---</td>
<td>0.28%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shan ñáa is an adversative passive.</td>
</tr>
</tbody>
</table>
Chapter 5  
Noun Phrase Syntax and Miscellaneous

5.1 Introduction
This chapter shows the results of comparing grammaticalized markers in Thai with corresponding Shan grams in areas of noun-phrase syntax. In §5.1.1 affix-like morphology used for nominalization in Thai is compared with nominalization strategies and grams used in Shan. In §5.1.2 two different “adjectivelizers” – productive grams used to create adjectives, are presented. Section 5.1.3 looks to the grammaticalized functions of Thai thûi – a noun meaning PLACE – and compares uses of Shan tî, the lexical cognate. Lastly, §5.1.4 investigates the marking of possession in Shan and Thai with a focus on khû señ a genitive marker originally a noun meaning 'thing'.

5.1.1 Nominalizers
Nominalization “the process of forming a noun from some other word class” (Crystal, 2008, p.328), can be accomplished through affixation, where affixes – bound-morphemes – attach to a root word to add lexical or grammatical information. An affix which creates a noun is called a nominalizer. The Thai nominal prefix nák- originates from the Old Khmer anak ‘person’ (Diller, 2001, p.142), but is now a classifying prefix transforming verbs into agents of the action that their verb signifies, such as in (95) where it transforms the verb rian 'to study' into a noun 'student'.

95. (T) [Corpus:NK]  

sâj kràprooŋ sâm mûān kràprooŋ nák-rian jìipûn
wear skirt short same skirt NOM-study 'student' Japanese

‘(She) wore a short skirt, similar to the skirts of Japanese students.’

Like all Tai languages, Thai is typically classified as an analytic language with instances of affixation rarely occurring. Yet, the nominalizing affix nák- is one of the few cases of bound morphology within the Thai language. The prefix nák- is both productive when fixed to verbs (e.g. ‘fly’ in /nák-bin/ 'pilot', and some nouns (e.g.
‘politics’ in /nák-kaanmuaj/ ‘politician’). Through comparison of Thai words which are produced through the use of nák-, Shan was found to accomplish some similar nominalizations with the use of the word mɔ ‘expert’, as in example (96).

96. 
(S) mɔ tem  
NOM write  
‘writer’  
(T) nák khìan  
NOM write  
‘writer’  
(S) mɔ lɔŋ  
NOM sing  
‘singer’  
(T) nák rɔŋ  
NOM sing  
‘singer’

Despite being commonly used in compounds, the Shan mɔ and the Thai cognate mɔ also are still semantically transparent and can function as a free morpheme in both languages. Thai mɔ also has the sense of just ‘doctor’, and is found in several nominalized compounds such as mɔ-fan, ‘dentist’, mɔ-duu ‘fortune-teller’ or even in the lexicalized form mɔlam – a genre of music which features “expert singers”. This is not the case with nák- which exhibits a higher level of grammaticalization through its properties as an affix. This difference raises semantic issues for the nominalizations in (96) whereby the Shan and Thai forms could be seen as not equivalent. For Shan mɔ nominalizations the sense seems to be that the person is skilled at what they do, while it could be possible for a nák- nominalization to not be an “expert”, like that of a nák-rian ‘student’. Further dissimilarity between the languages is also seen where many common nák- nominalizations in Thai are accomplished with other types of compounding strategies in Shan like the examples in (97) where Shan nominalized compounds featuring kôn ‘person’ and phu ‘person’ correspond to a Thai nák- form.

97. 
(S) kôn khê héŋ  
person compete force  
‘athlete’ or ‘person who competes (in sport)’  
(T) nák kii-laa  
NOM sports  
‘athlete’  
(S) phu tem khàaw  
person write news  
‘journalist’ or ‘person who writes the news’  
(T) nák khàaw  
NOM news  
‘journalist’

27 Uses of Shan mɔ also a verb/pre-verbal marker have also been found, with the sense of “be able”, such as su mɔ lat kwam sàŋ pɔŋ (‘you able speak language which languages’) or “Which languages can you speak?”.

90
Nominalizations with person nouns like the Shan examples (99) are also possible in Thai where lexical cognates khon ‘person’, or phûu ‘person’ also serve as productive sources of nominalized compounds. Iwasaki & Ingkaphirom (2005, p.43) consider Thai phûu ‘person’ to be ‘quasi-prefix’ as it can stand alone as a free morpheme. This is also the case with Thai khon ‘person’ and Shan kón ‘person’. These words along other personal morphemes can be used to show an array of social distinctions. When used for nominalizing compounds there is some semantic and social distinction with khon being for ordinary people and phûu being a step up to higher-status people, although often the use of either form is somewhat pragmatically equivalent. (Singnoi, 2000, p.293). These factors also contrast the Shan and Thai nominalized compounds in (98) with both semantic and grammatical differences.

Grammatically the basic process of nominalizing compounds referring to agents or actors in Shan and Thai are similar. But the morpheme of nák- was most likely borrowed as a bound-prefix from Khmer and thus explains its productivity, while mɔ̌/ mɔ̌ɔ are more limited by its semantic content. Just as in the case of Thai mɛ̀ ‘mother’ which would be associated to nominalizations of female agents such as mɛ̀-khrua ‘(female) chef’28. In terms of grammaticalization nák- is semantically bleached and thus more capable of nominalizing a wider variety of source concepts. Figure 14 illustrates the difference in the function of nák- nominalizing strategies versus compounding used both in Shan and Thai29. The nák- morpheme though originally a loan now shows a higher level of grammaticalization in its productivity and limitation as a bound morpheme. The gram mɔ̌ɔ’s position on the lexical side of the spectrum takes into account its more specific meaning, while the general sense of nouns meaning ‘person’ like khon or phûu and able to nominalize a wider range of verbs.

28 There are exceptions, such as that a mɛ̀-tháp ‘general, commander’ can be a male.
29 Figure 14 uses the Thai forms for consistency in comparison with Thai nák-.
Another Thai nominalizer *khwaam* has grammaticalized from a noun meaning ‘sense or substance of a matter’ (Prasithratthasint, 1997). It is found mostly on verbs dealing with emotive or mental processes forming abstract nominalizations. Prasithratthasint (1997) found *khwaam* nominalizations going as far back as the 13th century. In Shan the lexical noun *táaŋ* ‘path/road’ functions as a nominal prefix, cover similar nominalizations as *khwaam* does in Thai. The noun *táaŋ* still holds its lexical sense in Shan as does the Thai cognate *thaŋ* ‘path’, but the Shan form has grammaticalized further into a productive prefix for abstract nominalizations. Examples of *táaŋ* nominalizations in Shan, which correspond to analogous *khwaam* nominalizations in Thai are shown in (98).

98. 

(S) *táaŋ* hû          (T) *khwaam* rûu
NOM know            NOM know
‘knowledge’          ‘knowledge’

(S) *táaŋ* hak          (T) *khwaam* râk
NOM love            NOM love
‘love’              ‘love’

Like abstract nominalizations in Thai, with *táaŋ* as a nominalizer Shan verbs become nominal concepts which can then be modified by adjectives, like in (99) where mental concepts like ‘knowledge’ and ‘sight’ are described in physical terms ‘big’ and ‘wide’.

---

30 Shan has a lexical noun *kwâam* ‘word, speech’. The author is not clear as to any relation to Thai.
99. (S) [Young, 1985, p.68]

pèn kón mí táaj hà jaj táaj hǎn kwaaj cím lik
be people have NOM-know big NOM-see wide follow letter
cím laaj máa cím nəŋ náaj jət fāa sīn hàj
follow method come because.the.fact lady peak-sky teach BEN

‘It wasn’t long before all the village children became literate and greatly knowledgeable with a broad outlook due to the education that Lady Yot-faa had given them.”

Only a few instances of abstract nominalization occurred within both corpora, but Moeng (1995) lists over a hundred entries with táaj as an abstract nominalizer. An example from a Shan Pear Story text in (100) shows the verb wôn ‘to think’ nominalized to become a noun ‘idea’ which is then possessed by a person lúŋ ‘uncle’.

100. (S) [Corpus:ON]

táaj wôn khɔ̀ ɭùŋ kón nàn wôn waa lukʔən cḍaj
NOM-think ‘idea’ POSS uncle CLF that think COM child male
sàam kəʔ əw kwà
three CLF take go
‘The idea of the man was that the three boys took (the fruit).’

An example from a Thai speaker (101), shows a compound khɔ̀ɔp-khun literally ‘thank-you’ is nominalized to be listed as a type of expression/demeanor that one could send.

101. (T) [Corpus:CT]

māj dáaj sòŋ thâa-thaaj wâa bɛ̀ɛp sàdɛŋ khwaam
NEG ACQUIRE send demeanor COM type express NOM

khɔ̀ɔp-khun ʔàraj jaŋ ʔii
thank-you what thing.this
‘He didn’t give any expression or expresses thanks or anything like that.’

In (101), thaaj is used within a lexical compound meaning ‘demeanor’ (literally ‘pose-path’), this is the Thai cognate of Shan táaj. There are several other lexical compounds in Thai which feature thaaj but most still maintain a sense of “path” (e.g. thaaj-dson [path-walk] ‘trail, footpath’). Though there are Shan uses of táaj for
similar “path” compounds only Shan features abstract nominalized concepts. There seems to be some overlap with Shan nominalized uses of táŋ and Thai uses of thaaŋ as a preposition meaning ‘about, concerning’ (see Prasithrathsint, 2010, p.75).

Another Thai nominalizer kaan- acts as a prefix, to form an activity noun from a verbal root. Prasithrathsint (1997) uses the term “gerundive nominalizations” to describe the function of kaan- nominalizations, in (102).

102. (T)

\[
\begin{array}{ll}
\text{kaan} & \text{rian} \\
\text{NOM} & \text{studying} \\
\text{kaan} & \text{khít} \\
\text{NOM} & \text{thinking}
\end{array}
\]

Like the nominalizer nák-, kaan also has come from the Khmer language (originally from Pali). Interestingly it seems to have grammaticalized further in Thai. “Thai kaan is the same word as Cambodian ka:(r) which is together with sëckdyy ‘subject, matter’ the most productive class noun. Unlike Thai kaan, the range of occurrence of Cambodian ka:(r) still seems to be tied to some lexical constraints which do not allow it to be fully productive syntactically” (Bisang, 1996, p.559). Prasithrathsint (1997) studied the historical development of kaan nominalizations and though they are first found in the Ayutthaya Period, it is not until the Bangkok Period that kaan nominalizations regularly occur, sometime after abstract khwaam nominalizations developed. Shan does not use kaan yet similar nominalizations are done with a productive prefix lɔŋ which denotes a topic or case, such as example (103).

103. (S) [Elicitation]

\[
\begin{array}{ll}
lɔng & \text{tem} \\
\text{NOM} & \text{write} \\
\text{‘writing’}
\end{array}
\]

There is a lexical sense of lɔŋ as a either a noun meaning ‘hole’ or a verb meaning ‘float’ which is cognate with Thai lɔŋ. Moeng (1995) also gives an entry for lɔŋ as a noun meaning “circumstance, fact what pertains to; (as a reason, cause)” which seems to be where the function as a nominalizing prefix comes from. Examples of lɔŋ nominalizations from Moeng (1995) are shown in (104)-(106).

31 There are also examples like táŋ-kin ‘food’ Young (1985:133) which is an a nominalization of a verb ‘to eat’, there is no corresponding nominalized Thai form like *khwaam-kin.
104. (S) [Moeng, 1995]

\[ \text{lông khwaak món} \]

NOM widen city
‘colonization’

105. (S)

\[ \text{lông maak mí} \]

NOM be.abundant have
‘economy, what pertains to wealth’

106. (S)

\[ \text{lông jù lông kinh} \]

NOM live NOM eat
‘livelihood, way of living’

While not all of the Shan examples have directly equivalent forms, an example (107) in Thai seems to match the sense used in Shan in (106) above. More data is needed to see how commonly lông is used as “activity” or “gerundive” nominalizations like kaan is used in Thai.

107. (T) 32

\[ \text{pát-caj thàaŋ sèet-thà-kit mąaj-thùnn̂̄} \text{ kaan-bò̂-rì-phòok rùw} \]

factors of economics refer NOM-consume or
\text{kaan-jùu-kaan-kin}

NOM-reside-NOM-eat

‘Factors of economics refer to consumption or the cost of living.’

Overall instances of nominalization in the corpora were rare, but elicitation and examples from texts showed clear differences in Shan and Thai nominalizers. Nominal compounds using nouns meaning ‘person’ are shared between Shan and Thai, but the nák- nominalization is unique to Thai (and Lao). Abstract nominalizations are formed through the use of different grams with Shan tāaŋ corresponding to Thai khwaam. Though Shan gerundive nominalizations were commonly ascertained through elicitations more investigation is needed here to better determine the use and function of Shan lông. A final Thai nominalization strategy, clausal nominalizations with thū ‘place’ is left for discussion in §4.3.3, but the findings there point towards Shan lacking this nominalization strategy.

32 From Ngamkham, Arirat (1993) pàa rák nàam (อารีรัตน์ งามขำ พ.ศ.2537 ป่ารักษ์น้ำ)
Table 21 shows the nominalizers used in Shan and Thai. The main difference shown here is the lack of both strong agent and clausal nominalizers in Shan. It should also be reiterated that two of the Thai nominalizers nák and kaan are of Khmer origin.

**Table 21 Shan and Thai Nominalizers**

<table>
<thead>
<tr>
<th></th>
<th>Shan</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>-</td>
<td>nák</td>
</tr>
<tr>
<td>Abstract</td>
<td>táañg</td>
<td>khwaam</td>
</tr>
<tr>
<td>Gerundive</td>
<td>郵ŋ</td>
<td>kaan</td>
</tr>
<tr>
<td>Clausal</td>
<td>-</td>
<td>thîî*</td>
</tr>
</tbody>
</table>

* thîî nominalizations are covered in §4.3.3

The grammaticalization of nominalization markers in Thai are hypothesized by Prasithrathsint (1997) to be the consequence of periods of modernization in Thai society.

> “With regard to their frequencies of occurrence, generally speaking, the use of all of the three (khwaam, kaan, and thîî) patterns of abstract nominalization today is more frequent than during the times of their first use; however, quantitative analysis shows that there are fluctuations throughout the period. The most important peaks are those extreme increases in the period of modernization of Thai society. Many socio-cultural changes in this period were likely to bring about change in the Thai language as well. Nominalization furnishes good evidence to support that hypothesis.”

(Prasithrathsint, 1997, p.189)

With Prasithrathsint's (1997) hypothesis on the role sociolinguistics factors have played here, we can perhaps understand some of the differences. But it is perhaps enough to conclude that findings here show that there are distinct differences between Shan and Thai in regards to grammaticalization of nominalizers.

### 5.1.2 Adjectivilizers

Adjective-forming prefixes or “adjectivilizers” modify verbs and adjectives to produce adjectival words which are then used to modify nouns (Iwasaki & Ingkaphirom, 2005, p.30). Examples in (108) show the Thai morpheme nāā functioning as an adjectivilizer.
108. (T)

nâa  
ADJVZR eat
‘edible/tasty looking’

nâa  
ADJVZR love
‘cute’

nâa  
ADJVZR look
‘attractive/beautiful’

Meesat (1997) holds that nâa grammaticalized from an intransitive verb meaning “to be good, to be special” into an adjectivilizer. Meesat hypothesizes that this development might have begun before the Sukhothai period. Later on around the time of Rama V nâa grammaticalized further to also have a function as an epistemic modal marker. Epistemic nâa ‘should’ be distinguished from the modal gram khuan ‘should’ in that “the quality of the conjecture is more subjective, and often expresses the speaker’s hope or expectation concerning the event” (Iwasaki & Ingkaphirom, 2005, p.139). Like khuan, nâa is followed by the cà IRR, and expresses a higher level of confidence than the modal marker ?àat cà. (See §4.2.6 for discussion of Pre-verbal TAM markers). In (109) we see the speaker use nâa cà to express their increased confidence on which type of fruit was seen after getting another look at it.

109. (T) [Corpus: MJ]

toon rêck khít wâa pen fârâŋ phôo maa duu ?îik thîi khít
time first think COM be guava when come look another time think
wâa  nâa  cà  pen lûuk phee
COM MOD IRR be CLF pear

‘At first (I) thought it was guava but when I looked again I thought it’s probably pears.’

Both functions of nâa can be seen in (110) from the Thai corpus where it modifies râk ‘to love’ into adjective meaning ‘cute’, and is used in the modal construction showing the speaker's confidence on the girl’s appearance.

---

33 Meesat states this as "เป็นการดีที่ พิเศษตรงที่ " (1997, p.132)
110. (T) [Corpus: MJ]

pen dèk  phùuŋŋ nāa-rāk  lɔŋ  khút  wāa  nāa  cā  pen …  nāŋ
be  child  female  cute  INTS  think  COM  MOD  IRR  be  wear
krāpɽroŋ  phǒm  jaaw
skirt  hair  long

‘It was a cute girl who passed by. I think (a girl) wearing a skirt with long hair.’

In Shan instead of the nāa adjectivilizer, the morpheme lǐ ‘good’ is used to accomplish the same function of creating adjectives from verbs. Shan examples in (111) correspond directly to the Thai examples given in (108).

111. (S) [Elicitation]

lǐ  kǐn
ADJVZR  eat
‘edible/tasty looking’

lǐ  hak
ADJVZR  love
‘cute’

lǐ  tój
ADJVZR  look
‘attractive/beautiful’

Examples in (111) were elicited from Shan speakers, but there are at least a dozen other examples within Moeng (1995). In example (112) lǐ is used along with mon ‘fun’ and kiaw ‘fun’. The usage in (112) looks very similar to an ‘elaborate expression’ – “four-syllable compound expressions that sometimes have idiomatic and colourful meanings” (Hanna, 2013, p.33) found throughout Tai languages.

112. (S) [Field Notes]

nāj  kɔ  pēn  ?ān  lǐ  mon  lǐ  kiaw  waj  se  jâw
this  CONN  be  CLF  good  fun  good  fun  quite  PRT  PRT
‘this (having a family) is a quite good and fun thing’

In (113) we can see lǐ is used to modify kɔ ‘scary’. The use in (113) strongly suggest some semantic bleaching of lǐ in that being “scared” (especially when being confronted by head-hunters) does not lend itself to any positive interpretation. This is similar to the bleaching found in Thai nāa in adjectives like nāa-kliat ‘gross’.
‘...and they looked really scary, (the two children) believed[understand] that (they) were head-hunters.’

Overall nāa had a 0.45% frequency within the Thai corpus. Most occurrences of nāa were for the modal concept within the [nāa cà] construction, with only 3 tokens of the adjectivilizer function. There were no lǐ adjectivilizer examples in the Shan corpus. It should also be pointed out that lǐ is cognate with Thai dii ‘good’, and that Thai dii does not have the adjectivilizer function further distinguishing the Shan form as a unique development. Interestingly both Shan and Thai adjectivilizers come from a source concept meaning ‘to be good’ which suggests a possible grammaticalization cline. Figure 15 shows Shan and Thai adjectivilizers along with their lexical source concepts.

![Figure 15 Shan and Thai adjectivilizers](image)

### 5.1.3 PLACE

Shan ti and Thai thîi are cognate nouns meaning ‘place’, which also serve several other grammatical functions. Kullavanijaya’s (2008) study on the grammaticalization of Thai thîi, used historical corpus data to show its increased number of grammatical functions and overall increased frequency. Originally a simple noun meaning ‘land, a piece of land, place’ Kullavanijaya holds that thîi now serves up to eight different grammatical functions (see §3.4.1 for a review of this study). The Shan cognate ti, also a noun meaning ‘place’, serves several of the grammaticalized functions that are found in Thai. (114) shows Shan ti used as a noun meaning ‘place’.

34 thîi is borrowed from the Chinese noun dì ‘earth, land, soil, place, locality’ (Bisang, 1996, p.557), reconstructed for Proto-Tai as dii (Pittayaporn, 2009).
114. (S) [Field Notes]

khàw nân mân tê phuk laj sêŋ ti

rice that 3SG IRR plant able two place
‘that rice, it can be planted in two places.’

Like Thai thī, Shan ti is a preposition in certain contexts, with post-nominal modification adding details of the location like in (115) and (116).

115. (S) [Corpus:SG]
kôn kêŋ ?àn kép jù ti nõ tonmâj

person CLF REL gather be.at at on tree
The person who was gathering (fruit) on top of the tree.

116. (T) [Corpus:MN]
?âj dëk sàam khon nû kë ñôn klâp maa phàan ... thûŋ thī
male child three people this CONN walk return come pass arrive at
?âj taa phûuchaaj khon nân kèp láuk phëe
male grandpa male person that gather CLF pear
‘The three boys walked back to where that old man was gathering the pears.’

From a simple noun one step in Thai thī’s grammaticalization was the development of its function as a class noun. Iwasaki & Ingkaphirom (2005, p.43) refer to this type of class noun as a “quasi-prefix” in that thîi/ti heads many compounds but are able to stand also as free morpheme. As a class noun thîi/ti heads a noun phrase indicating a specific – often lexicalized – location like in (117) and (118). Both Shan and Thai respective PLACE words are highly productive as class nouns.

117. (T) [Kullavanijaya, 2008, p.446]
thī khàaj tûa
place-sell-ticket
‘ticket booth’

118. (S) [Moeng, 1995]
ti tâ law
place-distill-alcohol
‘distillery’

Uses like those in (117) and (118) should be contrasted with other Thai thī compounds where there is no sense of ‘land’ or ‘place’ such as in (119), which have become possible through semantic bleaching of the original source concept PLACE. The researcher has not found any Shan ti compounds without some sense of ‘place’ or ‘location’.

100
Around the Mid-Ratanakosin period, usages of thîi to nominalize entire clauses and make them subject to a verb first occurred. In the Thai corpus there were instances of nominalized clauses, however it was used to nominalize verbs in (121) and (122). In example (121) the speaker uses thîi lûa ‘a remaining portion’ which is a fairly common usage nominalized compound, but in (122) the speaker uses thîi and the verb làn, to talk about a toy which they seem to be unfamiliar with.

121. (T) [Corpus:MN]
   thîi lûa màj mii ?àraj
   NOM remaining NEG have what
   ‘What’s remaining/left?’ There isn’t anything left.’

122. (T) [Corpus:CT]
   thîi làn râo…? khâw kô khêc tok-tok-tok
   NOM play TAG 3SG CONN just "sound of striking"
   ‘The thing he was playing with? He was just (hitting it) “ток-ток-ток”.’

A significant stage in the grammaticalization of Thai thîi was its development into a relative marker by connecting a clause to a head noun (Bisang, 1996, p.558). Kullavanijaya’s (2008) data places this function at the highest frequency compared to other functions. Both Shan and Thai PLACE words can be used as a relativizer as shown in (123) and (124) where relative clauses describe baskets.

123. (S) [Corpus:JY]
   mî jù sêŋ takaa ti kép jâw
   have be.at two basket REL gather already
   ‘There were two baskets that he gathered already.’
124. (T) [Corpus: TG]

.mií tàkrâa jìu sèŋ tàkrâa thîi mìi phǒnlámãaj tem

have basket be.at two basket REL have fruit full

‘There are two baskets that are full of fruit.’

Similar to the relativizer function is that of a complementizer where thîi appears between two nouns like in example (125) shown in Thai.

125. (T) [Corpus: GP]

dō̂n phàn maa phɔ̌ dìi naj khànà? thîi phûuchaaj kamlâŋ

walk pass come when good naj moment COM male CONT

"ej!" phǒnlámãaj chân hāaj paj nùŋ tàkrâa

"exclaim" fruit 1SG missing go one basket

‘(The three boys) passed just at the moment that man was exclaiming “Hey!?, one of my baskets of fruit is missing.’

Shan examples of ti as a complementizer, like (126), often includes a CO-NOUN or or genral classifier ʔǎn which is discussed more in depth in §5.2.3.

126. (S) [Corpus: ON]

cáj nùm sām kɔ̂ nân khaam pon kwa tonmâj ʔǎn ti láŋ pít

boy young three CLF that cross pass go tree CLF COM uncle pick

‘Those three young boys passed the tree where the man was picking (fruit).’

An additional use of thîi/ti can be found in the use of indicating ordinal numbers, which Kullavanijaya (2008) hold first appeared for Thai in the Ratanakosin period.35

127. (T) [Corpus: MN]

khâw khoŋ cà jâak hâj tàkrâa thîi sāam man tem

3SG MOD IRR want CAUS basket NUM three 3SG full

‘He most likely wanted to make the third basket full as well.’

While there was no ordinal ti in the Shan corpus usages as an ordinal number marker have been found in other sources, such as in (128).

128. (S) [Pham, 2006, p.135]

pɔ̂ pǎaj māj nǎn hâk kɔ̂ ti nùŋ nā tōk kwà

when branch tree that break CLF NUM one PRT fall go

‘When the top of the tree broke falling off, the first person fell off.’

35 Kullavanijaya (2008, p.463) writes of an interesting possibility here, on ordinal number marking, and it potentially being a re-borrowing of the Chinese numeral marking function.
The final function that seems to have grammaticalized in Thai is as a classifier, this usage like in (129), only appeared within Kullavanijaya’s (2008) corpora in the Modern Thai era within spoken data.

129. (T) [Kullavanijaya, 2008, p.447]

\[cɔɔŋ \text{ thîi wâj hâj sɔɔŋ thîi dûaj}\]
reserve seat keep BEN two CLF also

‘Reserve two seats for me.’

Overall frequency in the corpora for Thai thîi and Shan ti were both 1.3%. Despite the similarity in overall frequency further analysis revealed Shan and Thai uses of their respective PLACE gram differed. Table 22 gives a percentage of the function of PLACE words in each corpus. From this table we can see that Thai uses of thîi varied more than Shan ti, as well as that the percentage of uses as an Complimentizer or Relativizer were fairly similar. That both Shan and Thai both most commonly used PLACE words for relative clauses points to the level of grammaticalization these words have gone through and is consistent with findings from Kullavanijaya’s (2008) data.

**Table 22 Comparison of thîi/ti functions in corpora**

<table>
<thead>
<tr>
<th>% in Shan</th>
<th>% in Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Noun</td>
<td>3%</td>
</tr>
<tr>
<td>Ordinal Number</td>
<td>2%</td>
</tr>
<tr>
<td>Nominalizer</td>
<td>7%</td>
</tr>
<tr>
<td>Preposition</td>
<td>31%</td>
</tr>
<tr>
<td>Complimentizer</td>
<td>13%</td>
</tr>
<tr>
<td>Relativizer</td>
<td>55%</td>
</tr>
<tr>
<td>Classifier</td>
<td>1%</td>
</tr>
</tbody>
</table>

Kullavanijaya (2008) compared findings on Thai thîi with other Tai dialects (Mueang Hiam dialect of Lao, and Shangsi Zhuang, and concluded that Thai thîi has had a unique development with Thai sentences becoming “more complex with embedding and modifying constructions” (p.466). The results from the functions of ti in Shan support the claim that Thai thîi has made unique developments. Clearly the increase of the abstract meaning of thîi in the last two periods are part of the further grammaticalization of Thai thîi. This is supported by the lack of grammatical
marking done by Shan ti. Figure 16 illustrates the functions for Shan and Thai PLACE grams. The horizontal axis represents the grammaticalization path of thîi. These figures show that Shan ti lacks some of the grammaticalized functions found in Thai: Compound Nouns (for non-place senses), nominalizations, and classifiers.

![Diagram of grammaticalization paths for Shan and Thai PLACE grams](image)

**Figure 16 Functions of Shan and Thai PLACE grams**

### 5.1.4 Possession

In Thai the noun khɔ̌ɔŋ ‘thing’ has been grammaticalized as a genitive marker indicating possession (Matisoff, 1991, p.391). Though a grammaticalized marker khɔ̌ɔŋ is still lexical in both Thai and Shan. Example (130) shows usage of khɔ̌ɔŋ as a noun in Thai.

130. (T) [Corpus:CT]

```
ko chûaj kêp khɔ̌ɔŋ
```

CONN help gather thing

‘(They) helped gather his things.’

The Thai genitive khɔ̌ɔŋ construction is \([N_{\text{possessed}} [(khɔ̌ɔŋ) N_{\text{possessor}}]]_{\text{NP}}\) like in (131) were a hat’s owner is indicated.

---

36 Proto-Tai χɔɔŋ (Pittayaporn, 2009)
131. (T) [Corpus: TG]

\[ \text{ʔaj dèk sàam khon nán kòo hèn mùak khọɲ dèk phùuchaaj khon} \]

boy child three person that CONN see hat POSS child male person

PRT

Those three boys they saw the hat of that boy.’

However, the khọɲ possessive construction is largely optional in Thai like in (132) which shows possession just based contiguous word order \( \{N_{\text{POSS}} \} \).\( \{N_{\text{POSS}} \} \).

132. (T)

\[ mùak dèk \]

hat child

‘The child’s hat’

This same contiguous possession construction is the one used by Shan speakers as shown by (133).

133. (S) [Corpus: SL]

\[ \text{kam-nɔ ka hàn mòkhọʔaj tiʔànʔàw màakmâj} \]

short.moment PRT see hat boy REL CLF take fruit

‘a short moment later (they) saw the hat of the boy who took the fruit.’

Though the use of khọɲ for possession in Thai is optional, a look at its use in the Thai and Shan corpora shows a marked difference in how often it is used. There where a total of 20 Thai possessional NPs 12 of which were marked with khọɲ. The overall frequency of khọɲ in Thai was 0.22% which includes 2 lexical ‘thing’ usages. In Shan khọɲ only appeared 3 times in the corpus with only one use marking possession. Of the 24 possessional NPs in Shan, only one was marked with khọɲ the other two appearances of khọɲ were lexical uses. These differences in the marking of possession are shown in Table 23.

**Table 23 Possession in Shan and Thai corpora**

<table>
<thead>
<tr>
<th></th>
<th>Shan</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genitive khọɲ construction</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Contiguous word order</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>Total NP\text{POSS}</td>
<td>24</td>
<td>20</td>
</tr>
</tbody>
</table>

The findings here suggests that possession phrases are more likely to marked by khọɲ in Thai , while if Shan khọɲ has grammaticalized the possessive marking
function, it is not used as often and its use may be the result of influence from Thai. Figure 17 shows the functions of Shan and Thai \(khɔ̌ɔŋ\). The Shan diagram displays the status of \(khɔ̌ɔŋ\) as possibly not as grammaticalized, or not as frequent as the corresponding Thai form.

![Shan and Thai khɔ̌ɔŋ](image)

**Figure 17 Shan and Thai khɔ̌ɔŋ**

### 5.2 Miscellaneous

This section includes other grams and areas of grammaticalization which were also compared. In §5.2.1 the similarity of Shan \(waa\) and Thai \(wâa\) both in their lexical function as verbs ‘to say’ and their grammaticalized complementizer function is shown. Section 5.2.2 discusses differences between Shan and Thai prepositions some are grammaticalized from verbal concepts, and have continued grammaticalization past the preposition stage. The inclusion of a discussion of the general classifier \(ʔǎn\), in §5.2.3 was motivated by the large difference in frequency in the corpora which was much higher in Shan. In §5.2.4, comparison is done of Shan and Thai clause-final aspectual markers, and their extended function as narrative linkers along with the use of the “sequential indicator” morpheme \(kɔ\). These narrative discourse grams were of high-frequency in both Shan and Thai corpora. The final section here §5.2.5, looks at the grammaticalization path of verbs ‘to like’ into adverbs of frequency meaning ‘often’, and a possible unique development in Shan.

#### 5.2.1 SAY > COM

The Thai word \(wâa\) has grammaticalized from a simple verb meaning ‘to say’ into a complementizer which introduces a quotative or complement clause after verbs of speech or cognition (Matisoff, 1991, p.398; Iwasaki & Ingkaphirom, 2005, p.259). The verbal use still exists in both Shan and Thai as shown in (134) and (135).

Though the verbal use is glossed as ‘say’ here, there is a sense that it does not just refer to verbal communication, but has meaning somewhat closer to ‘to communicate’ whereas Thai \(phûut\) and Shan \(laat\) more explicitly refer to the concept ‘to speak’.
134. (S) [Corpus:SG]
\[
mán kɔ ʔâm \texttt{waa sāŋ}
\]
3SG CONN NEG say what
‘He didn’t say anything.’

135. (T)
\[
khâw kɔ mȧj \texttt{wāa ʔărəj}
\]
3SG CONN NEG say what
‘He didn’t say anything.’

As a complementizer, \texttt{wāa} is often used with verbs of speech for quotation of a speaker, but the grammaticalized use in (136) shows a more metaphorical extension in Shan \texttt{waa} complementing a verb of cognition \texttt{wōn} ‘to think’, adding the content of thought (or the lack thereof in this specific example).

136. (S) [Corpus:SG]
\[
mán kɔ ʔâm \texttt{wōn phāj ʔāw kwā}
\]
3SG CONN NEG think COM who take go
‘He wasn’t thinking “who took (the fruit)”.

The example (137) shows the possibilities of Thai \texttt{wāa} with it embedded within another \texttt{wāa} complement clause both which both complement Thai verbs of cognition \texttt{khūt} ‘to think’ and \texttt{khāw-caj} ‘to understand’.

137. (T) [Corpus:MJ]
\[
\texttt{khūt wāa khāw khāw-caj wāa sāam khon nīi tōg pen khon}
\texttt{khāmooj phōnlāmāaj khāw paj}
\]
think COM 3SG understand COM three CLF this must be people
steal fruit 3SG go
‘(I) think he understood that those three people must have been the ones who stole his fruit.’

Analysis of the corpora showed relatively similar frequencies of Shan and Thai \texttt{waa/wāa} (Thai 1.24 % and Shan 1.18%) Usages as a main-verb were extremely rare, with only three tokens in the Shan corpus and none in the Thai corpus. Most usages of \texttt{waa/wāa} were as a verb complement like the examples (136) and (137) above. Some other uses of \texttt{waa/wāa} included both Shan and Thai forms complementing linking words or conjunctions meaning ‘but’, ‘and’, or ‘because’. These findings suggest that Shan \texttt{waa} and Thai \texttt{wāa} serve the same grammaticalized functions overall, as shown in Figure 18.
5.2.2 Prepositions

Adpositions can grammaticalize from nouns and verbs (See fourth layer of grammatical development in Figure 8 in Section 3.2). Thai prepositions which grammaticalize from nouns often come from ‘relational nouns’ a noun whose meaning is “a location or direction potentially in relation to some other noun” (Hopper & Traugott, 1993, p.110). Prasithrathsint (2010) lists several prepositions in Thai which have grammaticalized from nouns, most of which still maintain the lexical noun usage. Many of these developments are also shared by Shan. Table 24 shows some similarities of Shan and Thai prepositions which have grammaticalized from nouns.

Table 24 Shan and Thai prepositions grammaticalized from nouns

<table>
<thead>
<tr>
<th>Shan</th>
<th>Thai</th>
<th>Noun</th>
<th>Preposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>láŋ</td>
<td>láŋ</td>
<td>back</td>
<td>behind</td>
</tr>
<tr>
<td>naa</td>
<td>nâa</td>
<td>face</td>
<td>front</td>
</tr>
<tr>
<td>nok</td>
<td>nɔɔk</td>
<td>outer part</td>
<td>outside</td>
</tr>
<tr>
<td>kāŋ</td>
<td>klaŋ</td>
<td>middle</td>
<td>amidst, in the middle</td>
</tr>
<tr>
<td>hım</td>
<td>rim</td>
<td>edge</td>
<td>by, beside</td>
</tr>
<tr>
<td>nŏ</td>
<td>mûa</td>
<td>north</td>
<td>above</td>
</tr>
</tbody>
</table>

Looking to the other possibility – prepositions which have grammaticalized from verbs – more differences between the two languages were revealed. Typologically, verbs tend to supply different types of adpositions than those that come from nouns, as Diller points out “Verbs have tended to supply allative and ablative case marking, while nouns, more static locatives” (2001, p.163). In contrast to the similarities of
prepositions which come from nouns, many Thai prepositions grammaticalizing out of originally verbal concepts were not found within Shan. Though there are several examples that could be used to contrast differences in Shan and Thai, this section will only introduce a couple of cases.

Intratat (1996) traced the grammaticalization of 23 Thai prepositions from verbs such as, <TKey>àak ‘to leave’ > ‘from’, khàam ‘to cross over’ > opposite, tɔ̀ɔ ‘to connect’ > ‘next’ etc. Charatdao points out that many of these words are now rarely used in their verbal sense, and instead are used only as prepositions. One example is in the use of Thai <TKey>àak ‘from’, which developed out a verb ‘to leave’ and begin to be used for spatial ablative purposes sometime after the 16th century (Diller, 2001, p.164). In (138) a question and answer pair shows typical usage of <TKey>àak in its more prepositional sense, with maa ‘to come’ acting as the main verb.

138. (T)
Q: maa <TKey>àak thûi  náj
    come from place   where

    ‘Where are (you) from?’

A: phôm  maa <TKey>àak  mueŋ thaj
    1SG(M) come from  state  Thai

    ‘I am from Thailand.’

The Shan equivalents in (139) show a difference between the two languages. In lûk a verb meaning ‘arise, get up’ is used instead of the verb <TKey>àak -- surprisingly since <TKey>àak is a Proto-Tai verb. In (139) the verb lûk is used along with ti ‘place’, for the sense of “coming from” a place. In the interrogative sentence lûk could be interpreted as a preposition, but the answer to the question reveals the possibilities for lûk as a main verb, in that the verb maa ‘to come’ is left out.

139. (S) [Elicitation]
Q: lûk  læuu  maa
    from which(place) come

    ‘Where are (you) from?’

A: háw lûk- ti  mûŋ táj
    1SG come.from state  Tai

    ‘I am from Shan state.’
The current use of *lûk* as a verb and as preposition (139) is a good example of what is called cross-class heterosemy – different meanings of word being associated with distinct grammatical categories. The possibilities for verb/preposition alternations in Tai languages and cross-class heterosemy show the importance of discussing grammaticalization even within descriptive grammars (See Enfield, 2006, p.18). And in cases where the lexical sense of a word begins to fade from use, like in the case of *càak*, understanding its grammaticalization can help to understand its grammatical behavior. But for Shan *lûk* it appears to still maintain either use as a verb or preposition.

Another preposition included in Intratat (1996) is the case of *laej* which has followed the path of a verb ‘to pass’ > PREP ‘beyond’. In contrast to some other cases, *laej* still retains its verbal meaning. In (140) *laej* describes a bus that has passed a specific point -- a sign, probably for a bus stop.

140. (T) [Longdo Dictionary]

```file
rôt mee laej pāaj pəaj tāŋ klaj con phûu-dooj-sāan tûŋ dəən klàp maa
bus pass sign go INTS far until passenger must walk return come
ʔeeŋ
oneself
‘The bus passed the sign so far that the passengers had to walk back on their own.’
```

In (141) with *jiu* as a main verb here, *laej* functions as a preposition indicating that the orbit of a star is beyond our solar system.

141. (T) [Charatdao, 1996, p.145]

```file
daaw duaj nān khooccon jiū laej râbòp sùrîjā?
star CLF that orbit be.at beyond system solar
That star. Its orbit is beyond the solar system.
```

An absence of *leej* in Shan is not surprising in that it is Mon-Khmer loan (Diller, 2001, p.166), but other differences can be counted with consideration that the path of ‘verb > preposition’ is not always the stopping point for grammaticalization. Further grammaticalizations after the preposition stage draws attention to some other differences between the languages. These developments are potentially uses which seem to be unique to Thai. Beginning with Thai *leej*, as well as a use as a prepositon it has also gained a function as post-verbal emphasis marker (Kessakul & Ohori, 1998, p.101), as shown in (142) and (143).
142. (T) [Kessakul & Ohori, 1998, p.102]

\[ \text{pâa bòòk pingu wàa } \text{kin laaj } \text{câ} \]

\[ \text{aunt tell Pingu COM eat immediately PRT} \]

‘His aunt told Pingu that “Please have it (at once)”.’

143. (T) [Kessakul & Ohori, 1998, p.102]

\[ \text{Hûh! jaŋ cháaw jùu } \text{laaj} \]

\[ \text{PRT still morning be.at EMPH} \]

‘Huh! It’s too early (to get up).’

First off, Shan does not even have the verb laaj ‘to pass’, so it’s not surprising to see it without the more specialized uses. For the Shan verb pôn ‘to pass’, Moeng (1995) does list an adverbial meaning ‘in excess, exceedingly, excessively’ suggesting some possible grammaticalization, but this is still far off from a function like the prepositional use of laaj. Developments like these are consistent with the general definition of grammaticalization which besides pointing out the movement of a lexical item to a grammatical item also includes the movement of ‘a grammatical item’ to ‘a more grammatical item’. Another example of further grammaticalization of prepositions is in Thai use of two prepositions lâŋ and càak together for the sense of ‘after’. Deictic time markers of this type (lâŋ-càak-nûn ‘after-that’, or lâŋ-càak-nû ‘after-this’) are found only in Modern Thai data (Kullavanijaya, 2003). Example (144) from the Thai corpus shows use of both lâŋ-càak as a deictic time marker to point to a time after the group of boys finished helping the boy who stole the pears.

144. (T) [Corpus:NK]

\[ \text{lâŋ-càak chûaj sêt } \text{man kôo mây phût mây khâp-khun } \text{ʔàraj kan} \]

\[ \text{after help finish 3SG CONN NEG speak NEG thanks what each.other} \]

\[ \text{sàk } \text{kham} \]

\[ \text{single word} \]

‘After (they) finished helping him, he didn’t say thanks or anything.’

The development of lâŋ-càak ‘after’ is but one example how Thai has added to the overall lexicon of time markers. Kullavanijaya (2003) holds that Modern Thai has developed more time markers than any previous periods in Thai language, many of which have come from prepositions such as thûn V ‘to arrive’ > PREP. These developments show the important role that grammaticalization plays in the further grammaticalization and specialization of Thai prepositions. Looking to Shan deictic time markers some general differences were noted. As discussed earlier Shan lacked the prepositional càak, yet it has a cognate form of lâŋ ‘back’. Moeng (1995) lists
wáaj a verb ‘be over, past the time or season’ as also having the sense of ‘after’, which also can be used along with lǎŋ in wáaj-lǎŋ ‘after’ but wáaj is also able to be used on its own like in (145).

145. (S) [Young, 1985, p.77]
   
   wáaj nán khíŋ ɲúm hóm ʔaan tój lik
   after that Mr. money-collect read look letter
   “Afterwards”, Mr. ɲúm-hóm read the letter.

Within the corpus the most common deictic time-markers for Shan made use of the noun kám ‘moment’ and a determiner nân that or nâj for ‘this moment’ and ‘that moment’. Moeng (1995) also lists kám lǎŋ meaning ‘afterwards’, but there were no uses of this in the corpus. The Thai corpus had 6 lǎŋ-càak uses and notably Shan had 2 tokens lǎŋ-càak (1 each from two different speakers). The appearance of it in the Shan corpus could be due to the influence of Thai on individual speakers, but more research would have to be done here to better understand to what extent Thai, lǎŋ-càak ‘after’ is actually used by Shan speech communities. An interesting possibility here is that Shan speakers are borrowing the grammaticalized deitic time marker. This would seem to be the case especially if Shan lacks càak.

In summary, Shan and Thai static locative prepositions are similarly grammaticalized from lexical nouns, while prepositions which grammaticalized from verbs showed more differences as illustrated in the examples of càak and bəj. Many other differences in verbal prepositions may exist, but are outside of the scope of this paper.

5.2.3 Classifier ʔan/?ăn

Classifiers are an important feature of Tai languages (see §2.3.1 for a basic overview). The term classifier is a general term used often used to describe “lexical units used with numerals for counting homogeneous discrete items” (Morev, 2000, p.75). Classifiers are usually grammaticalized from nouns and are open class linguistic items. The semantic core of classifiers revolve around categories like animacy vs. inanimacy, human vs. non-human, but also include shape and other physical parameters or functions related to the lexical source noun. Shan and Thai share many cognate nouns along with their grammaticalized classifier functions, as shown in Table 25.
The function of classifiers in Tai languages can both be to numerate a noun and to “classify” it as well by adding information. Take (146) where the classifier lûuk deriving from ‘child’ is used to indicate individual pieces of fruit. Through the use of lûuk the semantic scope of the head noun ‘banana’ is defined to individual fruits. Contrast this with (147) where the classifier wǐi, a lexical noun ‘comb’, is used to refer to bunches of bananas (apparently from the physical similarity bananas in a bunch being similar to the teeth of a comb).

146. (T) [Morev, 2000, p.75]
klûaj sāam lûuk
banana three CLF
‘Three bananas’

147. (T) [Morev, 2000, p.75]
klûaj sāam wǐi
banana three CLF
‘Three bunches of bananas’

Classifiers become grammaticalized from nouns through reanalysis or metaphorical extension of using one concept in another domain. The cognitive background to this move is explained by Bisang as follows,

“It (classification) can be employed to compare one particular sensory perception and its properties to the properties of other sensory perceptions in order to identify that particular perception by subsuming it under a certain concept or it can be employed to establish a sensory perception as an individuum by actualizing the inherent properties which constitute its conceptual unity.” (Bisang, 1998, p.48)

This first step of NOUN > CLF can be seen as a first step of grammaticalization of classifiers. But, the continued grammaticalization of classifiers extends further for
purposes of ‘identification’ or ‘individualization’. The “functional expansion of classifiers” put forth by Bisang follows the path in Figure 19.

![Figure 19 Functional expansion of classifiers (Bisang, 1998)](image)

A discussion of classifiers and grammaticalization must also account for how these items gain other specialized functions. This is best done by seeing the constructions where classifiers have these other grammaticalized functions. The different constructional contexts in which classifiers come to fulfill different functions are presented in Table 26. Though these constructions were written for Thai, they could very well apply to other Tai languages, like Shan.

### Table 26 Functions of classifiers in different constructions (Bisang, 2008)

<table>
<thead>
<tr>
<th>Basic function</th>
<th>Grammatical function</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuation</td>
<td>Countability</td>
<td>Numerative construction [N NUM [CLF]]</td>
</tr>
<tr>
<td>Identification</td>
<td>Singulative</td>
<td>Demonstrative [N DEM] vs. [N CLF DEM]</td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td>Adjective construction: [N ADJ] vs. [N CLF ADJ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(definiteness/specificity: the object is accessible to the hearer)</td>
</tr>
<tr>
<td>Contrastive focus</td>
<td></td>
<td>Adjective and demonstrative constructions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(presupposed items are in contrast)</td>
</tr>
</tbody>
</table>

The constructions within Table 26 could be understood as semantically neutral in that whichever classifiers fill them depends on the characteristics of the head-noun or the pragmatic context (e.g. individual bananas or bunches). And though many classifiers are determined by shape or form/function, an important classifier for Southwestern Tai languages is the general classifier ?an. Burusphat (2007) found ?an as a genuine general classifier, for Northern Zhuang, Nung, Dai, Shan, Lao and Central Thai languages. The classification of ‘general classifier’ comes from its ability to classify a range of objects despite their physical characteristics unlike the classifiers shown in Table 22. Example (148) shows a question in which ?an could refer to a range of possible items across a range of semantic boundaries.
148. (T) [Iwasaki & Ingkaphirom, 2005, p.77]

ʔaw ʔan nǎŋ

take CLF which

‘Which one do you want?’

While presently ʔan might be classified as a ‘general classifier’ Wilaiwan (1976) (in Burusphat, 2007, p.149) characterized the original function of ʔan as a noun substitute, functioning in a manner similar to a relative clause marker.

Kullavanijaya’s (2008) data also confirms that for Thai ʔan was commonly used as a relative clause marker in the Sukhothai period. This usage was eventually replaced by the Khmer loan sùng in the Ayutthaya period (Diller, 2001, p.161) before the now much more common use of thăi for relativization and complementation (see §4.3.3.).

The main reason for including ʔǎn /ʔan within the comparative data here was the discrepancy in the two corpora. In the Shan corpus ʔǎn had a 1.92% frequency – the 11th most frequent word – while the Thai corpus had a very low comparative frequency of 0.07%. Looking at the Shan data, ʔǎn is used as a general classifier but we can see that it still is used for many of the functions associated with older forms of Central Thai understood as ‘noun-substitute’. In Young (1985) ʔǎn is listed an relativizer, but this does not seem to cover all of the functions or the type of constructions in which it appears.

Inglis’s (2014) investigation of Khamti Shan – a northern Shan language with approximately 14,000 speakers – concludes that ʔǎn has several grammaticalized functions as a conoun. Accounting for high frequency in his corpus Inglis states “(ʔǎn) makes up 4% of the overall corpus. In an isolating language such as Khamti, this kind of percentage is suggestive of a high degree of heterosemy—or polyfunctional polysemy. The frequency of ʔǎn is due to its grammaticalizing into a variety of conoun functions in many nominal constructions” (Inglis, 2014, p.57).

From Inglis’s conoun construction [NOUN][ʔǎn +... ] some similarities with data in this study can be seen. For example the use of ʔǎn by speakers in the Shan corpus were sometimes used to form relative clauses such as in (149), where ʔaj ‘boy’ is modified by a the clause headed with ʔǎn.

149. (S) [Corpus:SG]

ʔaj kʒ ʔǎn lak màakmǎj kôn nǎŋ

boy CLF REL steal fruit person that

‘The boy the one who stole that person’s fruit.’
This construction in (149) above features the classifier which would fit within the Reference construction in Table 26, but there are some questions left as to how much ʔǎn can be considered a relative marker in Shan. In §4.3.3 the use of ʔ for relative clauses was covered. In the corpus Shan speakers often used ʔ within an ʔǎn headed clause. In (150) ｌｕｇ ‘uncle’ is modified by clause headed with ʔǎn and that makes use of ʔ to connect the rest of the clause.

150. (S) [SN]

lúŋ ʔǎn ti ʔǔŋ tǔŋ kěp màakmāj lōŋ mà hē jāw

uncle CLF REL ascend CONT gather fruit descend come PRT already

‘The man, the one who was going up and gathering fruit, (he) came down.’

An interpretation of Shan ʔǎn usages here is that these are not integrated into one clause but instead two separate clauses, the use of ʔǎn here then is more of just a noun substitute that serves as a head of a separate modifying clauses in the form [ S ][ ʔǎn REL S ] like in (151).

151. (S) [Corpus:ON]

mā hǎa ti màakmāj ʔǎn ti mī nāu pet

come search at fruit CLF REL have in basket

‘(He) came and looked at his fruit. The fruit that’s in the baskets.’

This should be contrasted with Thai relative clause strategy of embedding the relative clause within the larger clause in the form [ S ][ ʔǎn S ] such as in (152).

152. (T) [Corpus:TG]

ʔaw phǒnlāmāj thū kęp dāaj maa kęp sāj tàkrāa

take fruit REL gather able come gather put basket

‘(He) takes the fruit that he can gather and put it in the basket.’

The use of ʔǎn is understood much better when we take into account the pauses in the production of speech. In example (153) ʔǎn makes several appearances, the thing to notice in this example is ʔǎn occurring after clear pauses in speech (marked with “...”). An interpretation of ʔǎn either as a noun substitute or relativizer then seems to depend on how clearly there is a grammatical break between the two clauses – the head-clause and the modifying one. In this example the first appearance of ʔǎn could be a noun substitute, but then as the speaker uses the relative marker ʔ, he pauses and reintroduces the head-noun ʔɔn ‘time’ followed by ʔǎn which introduces the relative clause. The last appearance of ʔǎn in (153) shows ʔǎn reappearing after relative clause has begun but, with a pause in speech the speaker reintroduces a nominal head by using ʔǎn as a noun-substitute.
This discussion of ʔăn above highlights some the context in which ʔăn occurs and the constructions in which it can be categorized as either a noun-substitute or relativizer. Examples like these may provide some possible context for grammaticalization of thîi as a relative clause marker. Where at first topicalization in the form of a [NOUN][ʔăn ti S] eventually becomes integrated with the loss of the general classifier acting like a head in the modifying clause so relativization now can happen in the form [N thîi S]. This is speculative and for the purposes of this study it is enough to note that the general classifier ʔăn had a much higher frequency in the Shan corpus than in the Thai corpus. This difference highlighted a difference in relative clause integration with Thai more often embedding the modifying clause next to the head noun.

5.2.4 ASP ‘finish’ > LINK

The Thai verb ล่วง ‘finish’ has grammaticalized to have both the functions of a temporal-aspectual marker meaning completion and an interclausal linking function (Diller, 2001, p.166). The use of ล่วง as a function word – a post-verbal auxiliary and conjunction – goes back to the Sukhothai period (Sriprasit, 2003). Matisoff (1991) holds that Thai ล่วง was originally a Chinese loan – a verb ‘to finish’, and that in its clause-final position it has “grammaticized into one of the language’s most important aspectual postpositions, signifying ‘completed action; change of state’” (1991, p.436), such as in (156). The Shan cognate จาว, also acts as a clause-final aspectual marker indicating ‘completion of an action’ as shown in (154) and (155).

154. (T) [Matisoff, 1991, p.436]
kláp paj bāan ล่วง khráp
return go home ASP PRT
‘He’s gone off home already sir.’
155. (S) [Corpus:SG]

mán lak kwà jâw
3SG steal go ASP

‘He stole (the fruit).’

From a clause-final aspectual marker léew and jâw have gained a function as a clause linker marker. This development arises out of reanalysis across clause boundaries. Iwasaki characterizes this move as follows,

“Being a verb meaning ‘finish,’ it naturally developed into an aspect marker, then into a pragmatic particle. In this process, léew appears toward the end of a sentence. The same word with the discourse conjunction, on the other hand, appears at the beginning of a sentence. It as proposed that this development is a consequence of a reanalysis applied across two sentences.”

In the beginning of (156) linking léew’s main function is to push the narrative forward from the previous sentence, hence the translation of “And then..”, this function is contrasted with by the additional léew in the clausal final position which indicates that the action within the clause – chêt ‘to wipe’ – is complete. This auxiliary function of léew is made even more clearer with contrast to the verb sêt ‘to finish’ also being within the clause.

156. (T) [Corpus:MI]

léew kɔ̂ɔ maa phɔɔ chêt sêt léew
LINK take come as.soon.as wipe finish ASP

‘And then he took (the fruit) finished wiping it …’

As mentioned earlier léew grammaticalized from a verb ‘to finish’ to an aspectual marker, this further grammaticalization results from changing positions across layers or boundaries of sentences. In (156) above the linking léew was accompanied with the morpheme kɔ̂ɔ and is glossed along with léew for the linking function. This morpheme – kɔ̂ – serves a sequential indicator even on its own, and represents an important narrative marker not just in Thai but in other Southeast Asian languages. Burusphat (2008) describes the spread of this morpheme throughout the region and into Tai languages,
“Having the Angkor kingdom as the focal area and the Siamese as the newcomers, it can be concluded that kɔ has spread from Khmer to Thai. The interpersonal contacts seem to play an important role on the diffusion. The conclusion is further supported by the comparison of forms of the sequential indicator. Only the Southwestern Tai group who moved from China to the present locations have the cognates of kɔ whereas the form is absent in the Northern Tai group who are still residing in China and present in all languages of Mon-Khmer groups except Mon.” (Burusphat, 2008, p.439)

The use of kɔ and the reanalysis of lécw happen at the edge of the clause boundary. In an idea similar to Bisang’s (1996) attractor positions, Iwasaki (2008) presents a model of “bi-polar distribution”, whereby reanalysis occurs outside of the bound of one sentence. Figure 20 shows the possibilities for reanalysis here using Thai lécw and kɔ.

The use of Shan jâw, kɔ́ and jâwkɔ́ is the same as the role of the Thai equivalents kɔ and lécw as discourse conjunctions. Example (157) shows the use of jâwkɔ́. An interpretation of jâw and kɔ́ here using the either of bracketing possibilities in Figure 20 shows the context for grammaticalization, described by Iwasaki (2008). While an interpretation of jâw here as the aspectual function might be possible, pragmatically the context suggests a linking function. Also featured within (157) is use of kɔ́ as a lone morpheme.

157. (S) [Corpus:PM]

\[
\begin{array}{l}
\text{phuítỳ khì sephin}^{37} \text{má jâwkɔ́ mán kɔ́ ku ku} \\
\text{female ride bicycle come LINK 3SG CONN laugh}
\end{array}
\]

‘The girl rode her bike and then she kɔ... laughed.’

Perhaps more clearly showing the linking function is jâwkɔ́ in (158) where the presentation-first construction introduces the boy, then jâwkɔ́ is the used to continue the narrative and the boy’s movement toward the tree which is marked with the aspectual jâw.

---

37 Burmese loan
As it is in Thai, kľ also serves the narrative linking function in Shan. In example (162) a Shan speaker uses it three times, here kľ is glossed as a connective as it occurs at almost any part of a clause. In actual speech it often comes off as a “filler” word, giving the speaker time to think of what they will say next while signaling to the listener that they have yet to finish speaking. For Shan, one should be aware of the similarity of kľ CONN and kľ CLF ‘person’ which are distinguishable by tonal contour. In (162) the uses of kľ are good examples of the filler role kľ can play. The translation line for (162) reproduces kľ to highlight the difference from the linking function of jâwkľ which is often glossed as ‘and then’. Thai kľ often functions as a filler in the same way as in the Shan example below.

A main reason for including analysis of these important narrative markers is that they were the most frequent words in both the Thai and Shan corpus. This is not surprising since the texts that make up the corpora come from natural spoken narratives on the event of the Pear Story. Table 27 shows a comparison of both Shan and Thai discourse conjunctions. Of these kľ/kľ was the most frequent for both languages. It was the most frequent word in the Thai corpus by a wide margin, and was the second most frequent word in the Shan corpus. Overall the Table also shows a total frequency for all of the three morphemes together, and though the Thai corpus had a higher frequency, it was not drastically different from the Shan total.

38 In §4.4.2 prepositions that have developed into deictic time markers was discussed. In this example the Shan speaker uses ship ‘next’ this is different from Thai tło ‘next’ but interestingly they both seem to have grammaticalized from verbs meaning ‘to connect’.

158. (S) [Corpus:KG]

mí  lukʔn  kón nuŋ kha  mán jâwkľ  phan má  jâw

have child  CLF one  PRT 3SG  LINK  pass  come  ASP

‘There was a boy he came (to the tree).’

162. (S) [Corpus:WN]38

mán kľ  khùn  ship  päj  kwà  ship  khi  cakajun  kwà  kón  sâam
3SG  CONN  return  next  walk  go  next  ride  bicycle  go  people  three

kľ  nân  kľ  khùn  phê  màakmâj  nân  kľ  nuŋ  luk  pên
CLF  that  CONN  give.back  share  fruit  that  CONN  one  CLF  be

sâam  luk
three  CLF

‘He kľ... walked back and then rode his bicycle back to those three boys kľ... (he) gave them some of the fruit kľ... each of them one, a total of three.’
### Table 27 Shan and Thai narrative markers

<table>
<thead>
<tr>
<th>Shan</th>
<th>Rank</th>
<th>Frequency</th>
<th>Thai</th>
<th>Order</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>jâw</td>
<td>#32</td>
<td>0.83</td>
<td>lêw</td>
<td>#8</td>
<td>1.88</td>
</tr>
<tr>
<td>kɔ́</td>
<td>#2</td>
<td>4.04</td>
<td>kɔ́ɔ</td>
<td>#1</td>
<td>5.29</td>
</tr>
<tr>
<td>jâwkɔ́</td>
<td>#4</td>
<td>3.18</td>
<td>lêwkɔ́</td>
<td>#7</td>
<td>1.89</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>8.05</td>
<td>Total</td>
<td></td>
<td>9.06</td>
</tr>
</tbody>
</table>

One consideration here is analyzing the conjoining of lêwkɔ́ and jâwkɔ́. The transcription of both the Thai data coming from Post (2007) and the Shan data here being transcribed by the researcher, can leave questions on these matters. Better phonological methods of analysis could perhaps shed light on the joining of the each language’s discourse markers.

#### 5.2.5 LIKE > OFTEN

A grammaticalization path in Tai languages is the movement of a verb 'to like' into an adverb of frequency meaning 'often'. As covered in Jaroenkitboworn (2009) the pragmatic motivation for reanalysis of chɔ́ɔp 'to like' a sentence like (163) is quite clear, in that if the speaker likes to eat bananas, a likely interpretation then is that the speaker eats bananas often.

163. (T) 
chân chɔ́ɔp kin klûaj
1SG V 'to like' or ADV 'often' eat banana
'I like to eat bananas.' or 'I often eat bananas.'

The grammaticalization of the verb chɔ́ɔp becomes clear though when we see it used in a sentence like (164) where there is not a favorable interpretation.

164. (T) [Jaroenkitboworn, 2009, p.81] 
phôm chɔ́ɔp tîaun kuôn klaaŋ dûk phrɔ́ fǎnráaj
1SG 'often' wake ascend middle night because nightmare
'I often wake up at night because of nightmares.'
The grammaticalization of Thai ช่อง parallels a similar development in Lao with the verb มาก 'to like' which also can be used as an adverb of frequency meaning 'often' (See Enfield, 2007, p.221). Interestingly the adverbial meaning of มาก is commonly used in Thai while the lexical verb meaning 'to like' is not. Such as in example (165) which also shows the common pairing of มาก along with the irrealis marker  cará.

165. (T) 39
   บ้านคน太多 ถ่ม cará พิษ ถ้าจาวบ่อย บ่อย
   some person too often IRR change deportment frequently
   ‘Some people tend to change their deportment frequently.’

Though Shan does not share ช่อง with Thai, it seems to share both the verbal and adverbial meanings of มาก. Moeng (1995) lists มาก (notice the tone shift) as a verb meaning 'to like, enjoy' , the adverbial meaning seems likely to be possible with examples like (166).

166. (S) [Moeng, 1995]
   คน มาก แอลกอฮอล์
   person like alcohol
   ‘person who likes alcohol’

Example (166) is a sign of Shan sharing the similar use of Lao in the LIKE > OFTEN grammaticalization. Interestingly Shan appears to have its own grammaticalization of an adverb of frequency arising out of the verb ทำ 'to be diligent, hardworking'. In (167) ทำ describes the subject as a 'hardworking'. But in (168) ทำ moves to a position before the copular เป็น to act as an adverb meaning 'frequently, often'.

167. (S) [Elicitation]
   มันเป็นคนทำ
   3SG be person hardworking
   ‘He is hardworking.’

168. (S) [Elicitation]
   มันทำเป็นวัต
   3SG often be have.a.cold
   ‘He often catches colds.’

39 From Phukkanasut, Naphat (2008) ได้พิษ สาเหตุเป็นความทุกข์ไม่ยาก (นภัทร พุกกะณะสุต พ.ศ.2552
‘ได้พิษ สาเหตุเป็นความทุกข์ไม่ยาก’)
This development in Shan is most likely caused by the same sort of pragmatic motivation that caused Thai ชื่อ and Lao _md to grammaticalize. With the sense of someone being hardworking or repeatedly doing something an extension of meaning into frequency explains how kâj could be reinterpreted as an adverb meaning 'often'.

5.3 Summary
Comparison of grams in noun-phrase syntax revealed similarities and differences in Shan and Thai. In the area of nominalization Shan and Thai differed in respect to specific grams. The person nominalizer nāk- was not found to be used in Shan. Other types of compounding accomplishing person nominalizations was found to be very similar between the two languages. Shan ตาญ 'path' was found to function in for similar abstract nominalizations as the Thai abstract nominalizer khwaam. The gerundive nominalizer kaaŋ used in Thai was not found to be used in Shan which instead had some similar types of nominalization or topicalization, though more research is needed here. Shan and Thai were both shown to have separate grams which function as adjectivilizers. Though these grams differed interestingly they both seem to originate from verbs meaning 'to be good'. The Thai adjectivilizer has grammaticalized further into a modal marker, a function not held by the Shan equivalent. The multi-functional PLACE grams in Shan and Thai shared some of the same grammaticalized functions but through analysis of the corpus showed similar frequencies for both languages, Shan was found to lack the abstract nominalizer and classifier function. In line with Kullavanijaya (2008) these findings support the argument that Thai ที่ has grammaticalized multiple functions not held by other Tai languages. Finally in comparing uses of noun-phrases of possession in the corpora the genitive marker khaŋ was shown to not be used nearly as often in Shan as it is in Thai, suggesting a disparity in the levels of grammaticalization of this gram for each language. Table 28 presents an overview of the findings of the comparison of Thai and Shan noun phrase grams.
The last section included some final comparisons of Shan and Thai grams many of which were of high frequency in the corpora. One similarity here was in the grammaticalization of verbs 'to say' > COM. The section on prepositions addressed the difference between prepositions which grammaticalize from nouns and those which come from verbs. The examples of câak and ləəj were used to show the differences between Shan and Thai in prepositions which come from verbs. The noun-substitute ʔǎn was found to be much more frequent in Shan, and uses showed a difference in the level of clausal integration in Thai and Shan relative clauses. The grammaticalization of aspectual markers meaning 'to finish' in Shan and Thai were both shown to have grammaticalized across clause boundaries to become important linkers in narrative discourse. Both languages also included the sequential indicator/filler morpheme kɔ̂ɔ/kɔ́ in the linking construction. The last section covered looked at an interesting grammaticalization path of verbs 'to like'> ADV 'often, frequent', though Thai and Lao have grams which have made this move Shan differs with the gram kâj which has grammaticalized from a verb 'to be diligent' into an ADV meaning 'often, frequent'. Table 29 presents an overview of the findings of the comparison of Thai and Shan miscellaneous grams.

<p>| Table 28 Summary comparison of Shan and Thai noun-phrase grams |
|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Thai</th>
<th>Shan</th>
<th>Function</th>
<th>Frequency</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Thai</td>
<td>Shan</td>
</tr>
<tr>
<td>nák</td>
<td>-</td>
<td>NOM</td>
<td>1 token</td>
<td>-</td>
</tr>
<tr>
<td>khwaam</td>
<td>táaj</td>
<td>NOM</td>
<td>5 tokens</td>
<td>2 tokens</td>
</tr>
<tr>
<td>kaan</td>
<td>lī</td>
<td>ADJVZR</td>
<td>0.41%</td>
<td>---</td>
</tr>
<tr>
<td>thūi</td>
<td>ti</td>
<td>REL/COM</td>
<td>1.30%</td>
<td>1.31%</td>
</tr>
<tr>
<td>khɔɔŋ</td>
<td>---</td>
<td>POSS</td>
<td>0.22%</td>
<td>1 token</td>
</tr>
</tbody>
</table>

The last section included some final comparisons of Shan and Thai grams many of which were of high frequency in the corpora. One similarity here was in the grammaticalization of verbs 'to say' > COM. The section on prepositions addressed the difference between prepositions which grammaticalize from nouns and those which come from verbs. The examples of câak and ləəj were used to show the differences between Shan and Thai in prepositions which come from verbs. The noun-substitute ʔǎn was found to be much more frequent in Shan, and uses showed a difference in the level of clausal integration in Thai and Shan relative clauses. The grammaticalization of aspectual markers meaning 'to finish' in Shan and Thai were both shown to have grammaticalized across clause boundaries to become important linkers in narrative discourse. Both languages also included the sequential indicator/filler morpheme kɔ̂ɔ/kɔ́ in the linking construction. The last section covered looked at an interesting grammaticalization path of verbs 'to like'> ADV 'often, frequent', though Thai and Lao have grams which have made this move Shan differs with the gram kâj which has grammaticalized from a verb 'to be diligent' into an ADV meaning 'often, frequent'. Table 29 presents an overview of the findings of the comparison of Thai and Shan miscellaneous grams.
<table>
<thead>
<tr>
<th>Thai</th>
<th>Shan</th>
<th>Function</th>
<th>Frequency</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Thai</td>
<td>Shan</td>
</tr>
<tr>
<td>wâa</td>
<td>waa</td>
<td>COM</td>
<td>1.24%</td>
<td>1.18% Complements verbs of speech/cognition and conjunctions</td>
</tr>
<tr>
<td>càak</td>
<td>---</td>
<td>PREP</td>
<td>0.18%</td>
<td>0.09% Both examples of V &gt; PREP in Thai which were not found in Shan, along with the the post PREP grammaticalizations.</td>
</tr>
<tr>
<td>ləaj</td>
<td>---</td>
<td>PREP</td>
<td>1.01%</td>
<td>--- Similar functions uses as ASP/LINK (See Table 27 for full frequency breakdown)</td>
</tr>
<tr>
<td>ʔan</td>
<td>ʔǎn</td>
<td>'thing'</td>
<td>0.07%</td>
<td>1.92% High Shan use as a noun-substitute or conoun</td>
</tr>
<tr>
<td>lézw</td>
<td>jâw</td>
<td>ASP/LINK</td>
<td>---</td>
<td>--- Similar functions uses as ASP/LINK (See Table 27 for full frequency breakdown)</td>
</tr>
<tr>
<td>(kɔɔ)</td>
<td>(kɔ)</td>
<td></td>
<td>---</td>
<td>--- Thai also uses Lao mák which also has followed the verb ‘to like’ &gt; ADV ‘often’ path.</td>
</tr>
<tr>
<td>chɔɔp</td>
<td>kâj</td>
<td>ADV</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
Chapter 6
Discussion and Conclusion

6.1 Review of the Study
Inspired by previous research into Tai syntactic change and grammaticalization this study looked at two closely related Tai languages – Shan and Thai – with the aim of better understanding grammaticalization in each language. The logic of the study was that by comparing modern synchronic data, a clearer picture of specific grammatical developments and of the divergence of Shan and Thai might be gained. The starting point for this research was the extensive literature on Thai grammaticalization from which the researcher compiled several grams and grammatical contexts for the purpose of a comparison with Shan. After eliciting analogous Shan grams from Shan-Thai bilingual speakers, the researcher was then able to analyze Shan texts and explore aspects of Shan grammar further. This study also adopted an approach from Post (2007) which entailed use of two corpora for the purpose of comparing frequencies of grammatical morphemes or “grams” in each language. Use of this approach required the creation of a small topic-specific Shan corpus, made up of texts obtained through a video elicitation method, as was done for the Thai texts from Post (2007) which were used again in this study. These texts provided data and several contexts for comparison of the two languages, first in generating many comparable examples and uses of language and second in also allowing frequency measurements of analogous grams to be done in each language's corpus. The findings from the comparison were presented in three sections: Noun-phrase syntax, Verb-phrase syntax, and a final section for other miscellaneous grams. Although the findings cover many similarities between the two languages, there were also some differences between the languages, with several unique developments due to factors of grammaticalization as well as sociolinguistic factors. With these findings the discussion of grammaticalization in Tai languages – dominated by Standard Thai – can now include several new Shan examples. Discussion of the findings are presented in the following sections within this chapter along with some suggestions for further research.
6.2 Grammaticalization

6.2.1 Shan

Grammaticalization theory proved to be useful for investigating Shan in two ways. First in how grammaticalization research in Thai pointed to areas for exploration in Shan, and allowed the identification of key Shan grams — many of which differed from an analogous Thai form. And though much of the findings in this study can speak to Shan and Thai similarities, at the extremities of each language's grammatical system certain differences can be understood as arising from differences in grammaticalization. Thus the second way grammaticalization theory was useful was in describing unique Shan developments within theoretical frameworks from grammaticalization theory (e.g. “the parameters of grammaticalization” see §3.2.1). These frameworks have the potential to outline the process or steps by which apparent shifts from lexical to functional word classes in the Shan data are likely to have happened. As a consequence, discussion of grammaticalization in Shan contributes not only to linguistic description of Tai syntactic change but also to broader field of grammaticalization also.

Several similarities in Shan and Thai were seen in cognate grams which shared the same grammaticalized roles in both languages. Grams coming from grammaticalization paths such as **TAKE > MOM**, **BE.AT > ASP**, **PLACE > REL** and **SAY > COM** were used in both languages. That these grams are obviously cognate and functioned so similarly suggests that these represent possible Proto-Tai developments or at least early Southwestern Tai developments which have continued to be used in each language. Existence of these grams in other Tai languages could support this interpretation as well. An interesting difference in Shan was an addition to the grammaticalization cline of verbs ‘to give’ (see §4.1.3). Even though the Shan gram **hāu** shared the grammaticalized function of its Thai cognate **hâj**, Shan differed with introduction of new lexical material into this grammaticalization path in uses of the verb **pâñ** ‘to share’. Grammaticalized uses of **pâñ** are noted by Jenny (2012) but this study found additional evidence of benefactive and malefactive constructions which further support the overall similarity with grammaticalized Thai **hâj**. The extension of **pâñ** here also serves as a good example of the Enfield’s (2003) “typological poise” (see §3.3.2) in that Shan lexical material has come to mirror Thai **hâj** grammaticalizations. This additionally raises the interesting question of how old grammaticalized usages of **pâñ** are, as well as to what extent other similarities in Shan may have followed Thai, some of these possibilities are discussed in §5.4.
The grammaticalized use of directional verbs as aspectual markers or “success markers” were similar in both languages (see §4.1.7). One interesting exception was the use of the verb pɔ̀ŋ 'to pierce' in Shan for aspectual marking of cognitive states where Thai uses the verbs khâw 'to enter' or ʔɔ̀ɔk 'to exit'. Though the form of the Shan gram differed here, pɔ̀ŋ’s functional use can be understood as arising out of the same grammaticalization processes as the Thai forms. Following the process of grammaticalization the verb pɔ̀ŋ has gone through the steps of reanalysis or metaphorical extension, and desemanticalization. The underlying metaphor for reanalysis is of a barrier of conscious awareness where things can 'enter', 'exit' or 'pierce' through. This sets up a context to where the words are extended into functional roles as aspectual markers. Consideration of differences like these allow some interpretations or hypotheses concerning comparative levels of grammaticalization in the two languages. While both the Thai and Shan grams still function as lexical verbs, Thai khâw and ʔɔ̀ɔk may seem to be lighter in semantic content and more general than the Shan pɔ̀ŋ which not only refers to a space but also a manner of penetration of the metaphorical cognitive barrier. But does Thai having two grams here distinguishing between receptive (khâw) and productive (ʔɔ̀ɔk) cognitive activities represent a more complex grammaticality? Or does Shan have a more deeply grammaticalized gram that encapsulates both of the functions mentioned for the Thai grams? From grammaticalization theory it seems that semantic bleaching could be a helpful parameter to consider. Since Shan pɔ̀ŋ functions more broadly, while having a more specific lexical sense it may be closer to a verb, as opposed to the more general verbs of movement like 'to enter' or 'to exit' used in Thai. Further questions to explore here would be use of directional verbs in lexical compounds, or other idiomatic usages in both languages, which could indicate levels of grammaticalization. Ultimately the answer to which is “more grammaticalized” here may not be significant or answerable in this context, but questions of grammatical depth do seem to gain more traction when comparing some other grammatical contexts of Shan and Thai.

A better example of the full effects of semantic bleaching and their implications for levels of grammaticalization can be seen better with the example of the Thai comparative marker. Thai makes use of kwàa a Proto-Tai lexical verb 'to go', whereas Shan uses verbs lɔ̀ 'to exceed' and se 'to separate' in comparative constructions (see §4.1.7). For Thai speakers kwàa no longer holds its lexical sense and has been semantically bleached, yet it is still a lexical verb 'to go' in Shan. Compare this with the Shan comparative markers lɔ̀ and se which still hold their verbal senses and can act as main predicates. In terms of decategorialization the
Thai grams have completely shifted to a functional comparative marker, while the Shan comparative markers still maintain links to the lexical word class. With Thai kwāː showing many of the features of classic grammaticalization, it could be given as one counter-example to Bisang’s concerns on the indeterminateness of grammatical morphemes in Southeast Asian languages (see §3.3.1). The differences of Shan and Thai comparative markers more sharply delineates the line between fully functional words and lexical ones, and shows some disparity in Shan and Thai in regards to levels of grammaticalization.

Looking further at Shan grams which blur the line between functional and lexical classes are words used in noun-phrase syntax. In comparing Shan and Thai nominalizers, several differences were revealed (see §5.1.1). The abstract nominalizer in Shan was fulfilled with tāːaj a lexical noun ‘path’, while the analogous Thai khwaam- is grammaticalized enough to be classified as a prefix in some grammars. The adjectivilizer in Shan is ˈlǐ shows good evidence for its grammaticalization, but it still holds its sense as ‘to be good’, whereas Thai nāː which no longer maintains any lexical usages and has even grammaticalized further into a modal marker. Another difference in nominalization was in agent nominalizers where Shan lacked a productive agent nominalizer like Thai nāːk-. The Khmer origin of nāːk, – along with the gerundive nominalizer kaan – seems to account for much of the productivity of these Thai functional prefixes, and explains their higher grammatical functionality in contrast with other more common Shan and Thai nominalizing/compounding strategies. That the Shan nominalizers still have concrete lexical senses while the Thai nominalizers have a more abstract sense also explains the Thai nominalizers affix-like properties which are normally associated with deeper levels of grammaticalization.

Differences in Shan and Thai also extended beyond the form of grams to also how often they were used. The cases of khɔ̌ɔŋ and ʔǎn brought attention to areas of difference in explicit grammatical marking or integration. Though the lexical form of khɔ̌ɔŋ 'things' exists in Shan, findings from the corpora showed a stark difference in its use as a genitive marker (see §5.1.4). Shan speakers almost always used possessive constructions based off of contiguous word order while the Thai speakers made use of the khɔ̌ɔŋ as a genitive marker in possessive constructions at least half of the time. The possessive construction based on contiguous word order is also possible in Thai, but the common marking of possession in Thai with khɔ̌ɔŋ shows a higher preference by speakers for the grammaticalized function. The high-frequency of ʔǎn in the Shan corpus also highlighted a difference with Thai. Focus on uses of
Shan ʔǎn as a noun-substitute outlined a difference in Thai having more fully integrated relative clauses (see §5.2.3). Within both languages the PLACE grams of both languages thîi/tii were both found to fulfill the main grammaticalized functions as a relativizer, but the frequent use of ʔǎn in Shan contrasted with the fully integrated Thai relative clauses which were more likely to be proximate to the head-noun referent. Analysis of PLACE gram frequencies in each corpus highlighted the other later developments of Thai thîi which were not found in Shan, such as uses as a clausal nominalizer or classifier functions. This difference further contrasted the limitations on Shan tii as opposed to the more bleached and extended uses possible for Thai thîi.

Two general areas which featured many differences were pre-verbal TAM markers (§4.1.6) and the prepositions grammaticalized from verbs (§5.2.2). The different TAM markers in Shan and Thai mainly reflect the differences in areal influences on the two languages, more of which is discussed in §6.3. Though there was a difference in grams, the base constructions were similar – especially in regards to irrealis marking in modal constructions. With the origin of many of the Thai markers known, questions as to the possible origins of Shan pre-verbal TAM and whether they are genetic (i.e. language internal grammaticalizations) or areal developments, must be answered elsewhere. From grammaticalization theory we can expect aspectual markers to come from verbs, and despite the expansion of Shan laj into a modal marker and the introduction of Burmese lo 'to need', the lexical concepts from which Shan pre-verbal TAM grammaticalized are unknown. Differences in prepositions – specifically those derived from verbs – highlights an important feature of grammaticalization in Shan and Thai – a divergence in the functional or grammaticalized uses of verbal concepts. Section 5.2.2 gave càak and laaj, as examples of differences in prepositions, but other differences such as the lack of other Thai forms (e.g. tɔ̀ 'next', lɔ̀t 'under', pràcam 'about', taam 'follow/next’) show the potential for different developments arising from verbs in the two languages.

Also important here are the more specialized functions that these prepositions continue to take on as they grammaticalize further. Taking this into account along with unique grammaticalizations of Shan verbs (e.g. pɔ̀ŋ 'to pierce', se 'to separate', lɔ̂ 'to exceed', jnáa 'to meet', kaj 'to be dilligent'), we can see the possibilities for distinct and divergent grammaticalizations of verbs in both Shan and Thai.

Taking into account all the individual similarities and differences reviewed above, one possible take away from an overall comparison is that Standard Thai has more grams which are more deeply grammaticalized than Shan. What is meant by more
deeply grammaticalized is that Standard Thai has more grams which are more strongly aligned with a purely functional role, or has grammatical aspects which are more likely to be explicitly marked through the uses of grammaticalized words. The use of the analogy of “depth” here is from Post (2007) who compared Thai to Chinese grammaticalizations and ultimately concluded that Thai was “shallower” than Chinese, based off of an overall comparison of analogous grams and processes of lexical compounding. A conclusion like Post’s is perhaps easier to come to when the languages are more distant (both figuratively and literally). The differences in grammaticalization for Shan and Thai listed in this study are definitely not as extreme as those between Thai and Chinese, or what the difference would be if we compared any Tai language with any analytic language, which by definition would leaves traces of its grammaticalizations through its inflections and other purely functional morphemes (§3.3.1). But with consideration of several of the grammatical aspects and grams that were compared in this study we can see that several grammaticalized features that appear with Thai grams were lacking in the analogous Shan grams. An interpretation of grammaticalized depth in each category is based on the comparison of desemaneticalization and decategoricalization of the analogous grams or the amount of explicit marking or integration of certain grammatical constructions.

Ultimately claims of greater or lesser grammatical depth here should be taken lightly, as they speak to issues of measurements of linguistic complexity and even to nature of grammaticalization itself. A claim of a lack of grammatical depth should be responded to with more investigation into idiosyncratic features of Shan grams and their grammaticalization, and with more research into Shan these claims could be falsified. It should be stressed again that this study looked for grammaticalizations from the perspective of Thai, and despite the evidence provided in this study Shan deserves further research into the possible idiomatic constructions that Shan has developed genetically from within. The method of comparing Shan with Thai was useful in the exploring grammaticalization, but more investigation into Shan could reveal more grammaticalizations without direct parallels or analogous grams in Thai.

6.2.2 Thai
Thai grammaticalization studies have given insight into the history and development of some of the most significant grams in Standard Thai. That grammaticalization has played a crucial role in the state of modern Thai grammar can be seen in the fact
that 20 of the Thai grams discussed in this study appear within the top 50 of the most frequent words of Thai (See Table 30). One might wonder as to how much of their frequency is due to lexical uses, but it is more likely that their high-frequency is consistent with high-functionality predicted by theories of grammaticalization (Bybee, 2007, p.336).

Table 30 Functions of most frequent Thai grams

<table>
<thead>
<tr>
<th>Gram</th>
<th>#</th>
<th>Gram</th>
<th>#</th>
<th>Gram</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>thîi REL</td>
<td>1</td>
<td>khûn ASP</td>
<td>16</td>
<td>kĵo CONN</td>
<td>29</td>
</tr>
<tr>
<td>dâaj ACQUIRE</td>
<td>2</td>
<td>càak PREP</td>
<td>18</td>
<td>khwaam NOM</td>
<td>32</td>
</tr>
<tr>
<td>cà IRR</td>
<td>3</td>
<td>kaan NOM</td>
<td>19</td>
<td>lćew ASP/LINK</td>
<td>35</td>
</tr>
<tr>
<td>wâa COM</td>
<td>4</td>
<td>paj ASP</td>
<td>20</td>
<td>thûuk PASS</td>
<td>36</td>
</tr>
<tr>
<td>hâj BEN</td>
<td>8</td>
<td>maa ASP</td>
<td>21</td>
<td>khâw ASP</td>
<td>44</td>
</tr>
<tr>
<td>tôŋ MOD</td>
<td>9</td>
<td>kwâa COMPAR</td>
<td>23</td>
<td>loŋ ASP</td>
<td>47</td>
</tr>
<tr>
<td>khôŋ POSS</td>
<td>10</td>
<td>jûu ASP</td>
<td>28</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

# = Ranking

One thing that stands out in Table 30 and in this study is the high number of grams developing from verbal concepts. Characterizations of Thai as just a “verb-loving language” (see §2.5.2), may be underplaying the role grammaticalization has had in shifting specific lexical verbs into more functional roles in Tai languages. Two broad clines common in Thai were those of V > TAM or V > PREP, but as pointed out in §3.2, an important aspect of grammaticalization is that as things become grammatical they tend to move into more grammatical roles. Further research in Thai grammaticalization should look to identify these. Some examples of Thai grammaticalization which show further and continued grammaticalization are shown in Figure 21.

---

40 Based on Thai corpus data from Chulalongkorn University (2013) Top 5,000 word list
http://ling.arts.chula.ac.th/TNC/contents/File/freq-5000.xls
When we look at the later levels of Thai grammaticalization like the examples in Figure 21 we find more variation and idiomatic developments (such as the expanded and multiple roles of thîi). Knowledge of these specific developments in Thai help to motivate claims that Thai grams are more deeply grammaticalized than Shan. These types of developments are what is meant in the the second half of Hopper & Traugott's definition of grammaticalization as “the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions” (1993, p.1). And though some hedging as to the applicability of this definition to Tai languages were discussed with Bisang (2008) raising the issue of “the lack of coevolution of meaning and form” for grammaticalization studies on Southeast Asian languages (see §3.3.1), a closer look at Thai reveals several cases which fall within the more typical parameters and classical conceptions of grammaticalization. Examples of these include grams such as the completely functional shift of comparative kwà, the change in form of the irreallis marker càak > cà, the conjoining and phonetic reduction in linking words lɛ́ɛ̂w+kɔ̂ɔ > lɛ́ɛ̂wkɔ̂ɔ, as well as the change of two prepositions into a single deictic time marker lāŋ+càak = lāŋcàak. Additionally we see productive nominal affixes in Thai which do not maintain strong lexical senses or uses – a factor associated with later levels of grammaticalization. An additional example against the idea that Thai grammaticalizations do not change forms, is the example of phonetic reduction or loss of syllable/morpheme boundary in the Thai male polite particle examples khɔ̀ rõp > khráp. This coming from two verbs khɔ̀ 'to beg' and rõp 'to receive' (Matisoff, 1991, p.443). Interestingly if we compare the gendered polite particles of Shan and Thai we find that Shan kha, serves as the both the male and female polite particle while Thai has a gendered distinction (Thai khâ for females and khráp for males). All of these examples especially stand out when they are contrasted with analogous Shan forms which lack many of classic grammaticalization features as mentioned by Bisang (1996, 2004, 2008, 2009, 2011, 2015).
The amount of grammaticalization studies done on Thai is quite impressive, not only for Tai languages but for Southeast Asian languages in general. While this study's main goal was to learn more about Shan, in compiling the Thai grams a more current compendium of Thai grams and recent literature in this field was assembled (See Appendix A: Thai Grammaticalizations Table). This research only gives us an account of how markers in Standard Thai have changed, but when we include comparative data like that from Shan a fuller story of developments and grammaticalization in the South-Western Tai branch is given. The usefulness of these findings are not limited to just providing more information on a lesser known language, but also in how they can be used to better frame the developments made through grammaticalization in Thai. In essence the extent to which Thai has grammaticalized can be better understood when it is compared to other closely related languages. With a more detailed picture of the broader situation, grammaticalization in Thai can be more accurately linked to grammaticalization theory and other types of historical linguistic study.

6.3 Sociolinguistic Factors

Despite this study's focus on grammaticalization, factors relating to areal influences and sociolinguistics can not be ignored when studying Shan or Thai. Common to both Shan and Thai – like all Tai languages – is the early connection to Chinese (Pittayaporn, 2014), also there is the influence of Indic languages used in Buddhism – the majority religion of both Shan and Thai. Yet one difference that was brought out from the comparison was the existence of several Khmer loans used in Thai (e.g. kamlaŋ, tàat, khuan, nák-kaan-), which were not in Shan. For Shan, some areal influence comes from Burmese both in the functional expansion of Shan laj into a modal marker along with the use of lò, a Burmese loan, appearing in these constructions. The cause of these differences in functional words – along with other lexical loans both in Shan and Thai – can easily be understood as being due to geographic proximity to the source language. In this case it seems that the extensive borrowing of Khmer did not reach the more northerly Shan language. As covered by Enfield (see §3.3.2), the linguistic environment of mainland Southeast Asia creates the possibility for extensive borrowing and the spread of grams throughout the region, this possibility adds extra factors for consideration when researching grammaticalization. Through the calquing of functors and typological poise, languages can quickly appropriate grammatical developments from other languages, such as the increase in passives, relative clauses, adverbial clauses of time and nominalization in Thai after contact English, as claimed by Prasithrathsint (2006,
Another thing to considered are the effects of Standard Thai on Shan. As a major language in the area, the influence of Thai on neighboring languages through media and economic pressures means the dispersion of Thai units of language, especially to speakers of Shan is highly probable. Within the Shan texts there where several indications of Thai influence on the Shan speakers, this was expected since the Shan participants used in this study living in Thailand at the time of the study. It is very well likely that the rate at which Thai exports more linguistic into Shan speaking communities will continue to grow, and with Shan’s typological similarity more Thai grams could easily be adopted.

6.4 Suggestions for further research
The video-elicitation and corpus-based approach used for this study proved to be useful, but there are several ways upon which it could be improved. Besides just an overall increase of size, for this type of focused and topic-specific approach used here, there are some things that could improve results. One thing would be better comparability of participants, essentially better controlling for the background of the speakers (e.g. gender, age, etc.) in each language group, this along with an increase in size would better improve the empirical validity of the corpus data. A more specific suggestion for the type of video-elicitation methodology would be construction of formatted questions for the interviewers. These could be questions which target particular grams, grammatical areas, constructions or content within the video. With some preparation, questions could be translated into each language and research assistants could be trained to implement the questions in the interviews of the participants. These considerations would help to improve the overall comparability of the texts and each language’s corpus.

The rationale of this study was based off the usefulness of grammaticalization studies for comparative studies of related languages. This study compared two languages within the South-Western Tai branch but more comparative grammatical research could be expanded to include other languages. An idea of what this would look like is given in Table 31 which compares several grammaticalized features across Tai languages. For the South-Western Tai languages here (Shan, Dai-Lue, Thai, and Lao), several of these grams already known, yet for a Central Tai language like Zhuang the status of other grams and/or grammaticalizations would be interesting.
This suggestion for further research points to a sort of Pan-Tai grammatical analysis, like has been done for dialects of Thai. Though there are big questions and many factors to consider here, hopefully this study has demonstrated the usefulness of grammaticalization as a point of comparison for Tai languages. Using well-established Tai grammaticalization clines, future researchers already have several contexts in which to investigate (either general developments e.g. V > PREP, or quite specific ones e.g. V 'to like' > ADV 'often'). The extent to which these developments represent clines which exist in other languages could be explored further and linked to backed grammaticalization theory (See Heive & Kuteva, 2004). This type of comparative research would not have to be limited to just synchronic data, but could use historical resources as well. Fortunately for a Shan, there are historical texts which could used for further investigation into the grams and grammaticalization reviewed within this study. All of this research would be taking place within broader questions of Proto-Tai research. In a sense we could look for evidence for “Proto-Tai grams”, but there may serious limitations on how far researchers are able to look back. Also for reconstruction of Proto-Tai the majority of previous research has used comparative phonology as the basis of comparison, so how to use Tai grammaticalization research along with the hypothesis based on phonological reconstruction of Tai languages would also be of interest.


### Table 31 Comparison of Tai Grams

<table>
<thead>
<tr>
<th></th>
<th>Shan</th>
<th>Dai Lue</th>
<th>Thai</th>
<th>Lao</th>
<th>Zhuang</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRR</td>
<td>tē</td>
<td>cak/di</td>
<td>cà</td>
<td>ca/si</td>
<td>yaak</td>
</tr>
<tr>
<td>PROG</td>
<td>tūk</td>
<td>tuk</td>
<td>kamlaj</td>
<td>kamlaj/phuam</td>
<td>čič/túk</td>
</tr>
<tr>
<td>OFTEN</td>
<td>kāj</td>
<td>caŋ</td>
<td>chɔ̃p/māk</td>
<td>mak</td>
<td>?</td>
</tr>
<tr>
<td>CAUS</td>
<td>hau/panic</td>
<td>huuu</td>
<td>hāj</td>
<td>ħāj</td>
<td>háu</td>
</tr>
<tr>
<td>NOM 'abstract'</td>
<td>tāŋ</td>
<td>khwaam</td>
<td>khwaam</td>
<td>khaam</td>
<td>?</td>
</tr>
<tr>
<td>NOM 'activity'</td>
<td>lŋŋ</td>
<td>kaaŋ</td>
<td>kaaŋ</td>
<td>kaaŋ</td>
<td>?</td>
</tr>
<tr>
<td>ADJVZR</td>
<td>lî</td>
<td>caʔ</td>
<td>nāa</td>
<td>tàa</td>
<td>?</td>
</tr>
<tr>
<td>COMPAR</td>
<td>Ṽ/se</td>
<td>lɔ̃</td>
<td>kwàa</td>
<td>kua</td>
<td>kwa/tō</td>
</tr>
<tr>
<td>PASS</td>
<td>nāa</td>
<td>?</td>
<td>thūuk</td>
<td>thūuk</td>
<td>nāai/tūk</td>
</tr>
</tbody>
</table>

---

6.5 Conclusion

This thesis aimed to answer questions on grammaticalization in two Tai languages. Using a combination of methods (comparison, video-based elicitations, and corpus linguistics) differences between Shan and Thai were outlined. While many significant functional words were similar, enough differences existed to propose some hypotheses about Shan and Thai as a whole. The furthest reaching of these was that some Standard Thai grams have developed a overall deeper level of grammaticalization than Shan. This conclusion was motivated in part by a lack of desemanticalization, decategoricalization and grammatical integration in Shan grams and constructions when compared with analogous Thai forms. Though Shan seemed to lack some of the grammaticalized features in Thai, there was also evidence that Shan has developed some of its own grammaticalizations – uniquely using lexical material in ways not shared with other Tai languages. Still, these Shan grammaticalizations were shifts of lexical material into functional roles, which were consistent within frameworks of grammaticalization theory. This study also exposed differences in due to the influence of other languages. Several Thai grams were shown to have Khmer origins, such as several pre-verbal TAM markers and the nominalizers. Shan, on the other hand, lacked the Khmer loans, but instead showed some effects of a Burmese influence with the expansion of the ACQUIRE gram into a modal marker, along with direct lexical borrowing of lò 'to need' in obligative modal constructions. Besides different forms of grams – either in phonological realization or base lexeme – there is a general similarity in each language's grammatical system, which provides the context for grammaticalization.

Ideas from grammaticalization have the power to explain several features of language. Successful use of these ideas in the Thai historical linguistics has been revealing, but because much of what has been said here has been from the perspective of Thai, it leaves questions on related languages. The aim of this study was primarily to help to expand discussion of grammaticalization in Tai languages by contributing data from a closely related language – Shan. The presentation of these findings, hopefully has shown the power of juxtaposing Shan and Thai grammaticalization to add insight into the history of both languages and the larger Tai family. As Shan is a lesser known language (linguistically), this research also had the added benefit of identifying current synchronic usages of several functional words in Shan. It should be restated again here that much of this study approached Shan from the perspective of the Thai language. This was due from in part to the researcher's familiarity with Thai, but also the fact that there is more research and
literature on Thai. Though this study is titled as a comparison of two languages, it is in one sense a comparison of Shan with Thai. Therefore, Shan – and its varieties – deserve more attention and exploration for their own sake, and the possible unique grammaticalizations that they have developed.
BIBLIOGRAPHY


## APPENDIX A

### THAI GRAMMATICALIZATIONS TABLE

<table>
<thead>
<tr>
<th>#</th>
<th>IPA/Lexical Gloss</th>
<th>Thai</th>
<th>Grammaticalization Path</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>?àat ‘be able to’</td>
<td>อาจ</td>
<td>V &gt; epistemic modal marker</td>
<td>Diller (2001); Matisoff (1991); Meesat (1997)</td>
</tr>
<tr>
<td>2</td>
<td>?aw ‘take’</td>
<td>เอา</td>
<td>V &gt; coverb (object marker) or “manipulated-object marker”</td>
<td>Diller (2001); Enfield (2002); Jagacinski (1992a); Matisoff (1991); Post (2007)</td>
</tr>
<tr>
<td>4</td>
<td>gà ‘will’</td>
<td>ฉี่</td>
<td>V &gt; IRR</td>
<td>Diller (2001); Takahashi (2011)</td>
</tr>
<tr>
<td>5</td>
<td>càak ‘to leave’</td>
<td>จำก</td>
<td>V &gt; PREP ‘from’</td>
<td>Intratat (1996); Diller (2001); Kessakul &amp; Ohori (1998); Kullavani jaya (2003)</td>
</tr>
<tr>
<td>6</td>
<td>chɔɔp ‘to like’</td>
<td>ชอบ</td>
<td>V &gt; ADV ‘often’</td>
<td>Jaroenkitboworn (2009)</td>
</tr>
<tr>
<td>No</td>
<td>Verb</td>
<td>Meaning</td>
<td>Analysis</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>10</td>
<td>จิ้ว ‘to reside’</td>
<td>อยู่</td>
<td>V &gt; PROG</td>
<td>Chamniyom (2003, 2006); Diller (2001); Intratat (1996); Matisoff (1991); Sriprasit (2003)</td>
</tr>
<tr>
<td>11</td>
<td>เย้ ‘yet’</td>
<td>ยัง</td>
<td>V &gt; PREP</td>
<td>Diller (2001); Intratat (1996)</td>
</tr>
<tr>
<td>13</td>
<td>กาล่าญ ‘energy’</td>
<td>กำลัง</td>
<td>N &gt; AUX</td>
<td>Diller (2001); Huffman (1986)</td>
</tr>
<tr>
<td>14</td>
<td>ข้าม ‘cross over’</td>
<td>ข้าม</td>
<td>V &gt; PREP</td>
<td>Chamniyom (2003, 2006); Intratat (1996); Jagacinski (1992); Matisoff (1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>16</td>
<td>khāaŋ ‘towards’</td>
<td>ข้าง</td>
<td>V &gt; PREP</td>
<td>Diller (2001)</td>
</tr>
<tr>
<td>17</td>
<td>khāw ‘to enter’</td>
<td>เข้า</td>
<td>V &gt; PREP</td>
<td>Chamniyom (2003, 2006); Diller (2001); Intratat (1996); Thepkanjana &amp; Uehara (2008)</td>
</tr>
<tr>
<td>18</td>
<td>khoŋ ‘to remain’</td>
<td>คง</td>
<td>V &gt; MOD ‘probably’</td>
<td>Meesat (1997)</td>
</tr>
<tr>
<td>19</td>
<td>khɔŋ ‘thing’</td>
<td>ของ</td>
<td>N &gt; POSS</td>
<td>Diller (2001); Matisoff (1991)</td>
</tr>
<tr>
<td>20</td>
<td>khuan ‘to be appropriate’</td>
<td>ควร</td>
<td>V &gt; MOD ‘should’</td>
<td>Diller (2001); Meesat (1997)</td>
</tr>
<tr>
<td>21</td>
<td>khūn ‘to ascend’</td>
<td>ขึ้น</td>
<td>V &gt; PREP</td>
<td>Chamniyom (2003, 2006); Diller (2001); Intratat (1996); Thepkanjana &amp; Uehara (2008)</td>
</tr>
<tr>
<td>22</td>
<td>khlāaj ‘to be similar’</td>
<td>คล้าย</td>
<td>V &gt; PREP</td>
<td>Chamniyom (2003, 2006); Intratat (1996)</td>
</tr>
<tr>
<td>23</td>
<td>kwāa ‘to pass beyond’</td>
<td>กว่า</td>
<td>V &gt; comparative adverb, temporal conjunctive, ADV</td>
<td>Diller (2001); Morev (1998)</td>
</tr>
<tr>
<td>No.</td>
<td>Verb</td>
<td>Meaning</td>
<td>Example Sentences</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>เลย ‘to pass’</td>
<td>V &gt; PREP &gt; EMPH</td>
<td>Intratat (1996); Iwasaki (2004); Kessakul &amp; Ohori (1998)</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>ลง ‘to descend’</td>
<td>V &gt; PREP</td>
<td>Diller (2001); Intratat (1996); Thepkanjana &amp; Uehara (2008); Chamniyom (2003, 2006)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>เหมือน ‘to be similar’</td>
<td>V &gt; PREP</td>
<td>Intratat (1996)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>หน้า ‘face’</td>
<td>N &gt; PREP ‘front’</td>
<td>Diller (2001)</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>น้า ‘ought to’</td>
<td>V &gt; ADJVZR</td>
<td>Meesat (1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>33</td>
<td><strong>nǔ: ‘rat’</strong></td>
<td>หนู</td>
<td>N &gt; PN</td>
<td>Diller (2001)</td>
</tr>
<tr>
<td>34</td>
<td><strong>nák- ‘NOM’</strong></td>
<td>นัก</td>
<td>N &gt; NOM ‘agent’</td>
<td>Diller (2001)</td>
</tr>
<tr>
<td>35</td>
<td><strong>nǐ &amp; nǐ` nǐa</strong></td>
<td>นี่, นี่, เนี่ย</td>
<td>DET &gt; pronoun, topic marker</td>
<td>Prasitrathsin (2000); Iwasaki (2004)</td>
</tr>
<tr>
<td>37</td>
<td><strong>pen ‘to be’</strong></td>
<td>เป็น</td>
<td>V &gt; PREP</td>
<td>Intratat (1996)</td>
</tr>
<tr>
<td>38</td>
<td><strong>phôn ‘to pass’</strong></td>
<td>ผ่าน</td>
<td>V &gt; PREP/CON</td>
<td>Matisoff (1991)</td>
</tr>
<tr>
<td>39</td>
<td><strong>phàan ‘to pass’</strong></td>
<td>ผ่าน</td>
<td>V &gt; PREP</td>
<td>Intratat (1996); Chamniyom (2003, 2006)</td>
</tr>
<tr>
<td>No.</td>
<td>Verb</td>
<td>Followed by</td>
<td>Type of Movement</td>
<td>Source</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>-------------</td>
<td>-----------------</td>
<td>--------</td>
</tr>
<tr>
<td>40</td>
<td>phūa ‘for’</td>
<td>เพื่อ</td>
<td>PREP &gt; BEN</td>
<td>Diller (2001); Jenny (2010)</td>
</tr>
<tr>
<td>41</td>
<td>prācam ‘to estimate’</td>
<td>ประจำ</td>
<td>V &gt; PREP</td>
<td>Intratat (1996); Chamniyom (2003, 2006)</td>
</tr>
<tr>
<td>42</td>
<td>wāa ‘to say’</td>
<td>ว่า</td>
<td>V &gt; COM</td>
<td>Diller (2001); Matisoff (1991); Jagacinski (1992b)</td>
</tr>
<tr>
<td>43</td>
<td>sū ‘to share’</td>
<td>สู่</td>
<td>V &gt; PREP ‘to, towards’</td>
<td>Diller (2001); Intratat (1996)</td>
</tr>
<tr>
<td>44</td>
<td>taam ‘to follow’</td>
<td>ตาม</td>
<td>V &gt; PREP</td>
<td>Intratat (1996); Chamniyom (2003, 2006)</td>
</tr>
<tr>
<td>45</td>
<td>thaany ‘path’</td>
<td>ทาง</td>
<td>N &gt; locative particle</td>
<td>Matisoff (1991); Prasithrathsint (2010)</td>
</tr>
<tr>
<td>46</td>
<td>thūŋ ‘to reach’</td>
<td>ถึง</td>
<td>V &gt; PREP-like coverb ‘to think/speak about’</td>
<td>Intratat (1996); Diller (2001); Chamniyom (2003, 2006); Kessakul &amp; Ohori (1998); Kullavanijaya (2003)</td>
</tr>
<tr>
<td>47</td>
<td>tɔɔ ‘to connect’</td>
<td>ต่อ</td>
<td>V &gt; QUANT, TAM</td>
<td>Intratat (1996); Diller (2001)</td>
</tr>
<tr>
<td>48</td>
<td>thūuk ‘to touch, strike/be struck’</td>
<td>ถูก</td>
<td>V &gt; PASS</td>
<td>Diller (2001); Prasithrathsint (2004, 2006)</td>
</tr>
<tr>
<td>49</td>
<td>ต่อง ‘touch’</td>
<td>V &gt; MOD ‘must’</td>
<td>Chancharu (2009); Diller (2001); Meesat (1997); Post (2007)</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>ที่ ‘place’</td>
<td>N &gt; REL&lt;br&gt;relative marker &gt; complementizer (COM, PREP, NOM, CLF)</td>
<td>Bisang (1998); Diller (2001); Kullavanijaya (2008); Post (2007)</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>ถ้า ‘if’</td>
<td>V ‘wait’ &gt; COND. ‘if’</td>
<td>Diller (2001)</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>ตรง</td>
<td>V &gt; PREP</td>
<td>Intratat (1996)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classifiers</td>
<td>CLF &gt; Deictic</td>
<td>Diller (2001); Post (2007); Morev (2000); Bisang (1998, 2008)</td>
<td></td>
</tr>
</tbody>
</table>
This story begins with one person who is gathering fruit up in a tree and then putting it in a basket.'

‘He’s gathered enough fruit to fill two baskets, and another basket that's about to be full.’ ‘then there's a man who's leading a goat and comes over to the left of him (the man picking the fruit).’

‘He just passes (to the left) by because his goat was wanting to eat the fruit, but he doesn't let the goat eat any of it. He pulls at the goat, he's making the goat pass by the baskets of fruit.’

1. ʔǎ.pǔm nāj kó  mán mì  kón kə nuŋ kép  tūk kép
   story  this CONN 3SG have people CLF one gather CONT gather

   mākmāj khun  ton mán kép  mākmāj sāj koj  nɔ  ...
   fruit  ascend tree 3SG gather fruit  put basket PRT

   ‘This story begins with one person who is gathering fruit up in a tree and then putting it in a basket.’

2. kép  mākmāj sāj koj  laj  sāŋ koj  jāw
   gather fruit  put basket ACQUIRE two basket ASP

   tʰɛŋ  koj  tɛ  tɛm jāw  ... jāwkɔ  mĩ  kũn  kə cuŋ
   another/again basket IRR full ASP  CONN have people CLF pull

   pɛ  ko  mā  mā  mā  jāa  sāj  mān  ...
   goat  CONN come come come meet left 3SG

   ‘He's gathered enough fruit to fill two baskets, and another basket that's about to be full.’ ‘then there's a man who's leading a goat and comes over to the left of him (the man picking the fruit).’

3. mān  jāa  sāj  kó  mān  kɔ  pɔŋ  kwɔ mān  waa  sāŋ  mān
   3SG meet left CONN 3SG CONN pass go 3SG say what 3SG

   kɔ  pɔŋ  kwɔ  kwe-na  pɛ  nāj  kaj  kĩn  jũ  mān  ŋũm  pũŋ  kĩn
   CONN pass go because goat this want eat PROG 3SG NEG give eat

   ‘He just passes (to the left) by because his goat was wanting to eat the fruit, but he doesn't let the goat eat any of it. He pulls at the goat, he's making the goat pass by the baskets of fruit.’
4. mán tut pě mán kwà haj pôn ?àn haj mán pôn ... pě mán
  3SG pull goat 3SG go CAUS pass CLF CAUS 3SG pass goat 3SG
  pě soŋsaj màakmâj nàn pě kaj kǐn tik-tik cuŋ pôn kwà ...
goose wonder fruit that goose want eat repeatedly pull pass go

‘He pulls at the goat, he's making the goat pass by the baskets of fruit. He's making the goat go, probably because that goose wants to eat the fruit, so he leads the goat away.’

5. kámnâj mí mí lukʔn t'eq̃ kɔ nʊʔ kʰi lôt thîp phan
at.this.time have have child another/again CLF one ride bicycle pass

lôt thîp lôŋ mā .. na jə wkɔ mā mā mā mā ..
bicycle descend come PRT LINK come come come come

‘At this time there’s another person – a child – who is riding a bicycle down towards the tree.’

6. mā hǎn koj koj màakmâj nàn mān súk mān hâŋ wǎn
  come see basket basket fruit that 3SG ripe 3SG beautiful sweet

mān sì mān kɔ hâŋ lǐ na
  3SG color 3SG CONN beautiful good PRT

‘The boy sees the baskets of fruit. The fruit is ripe and beautiful, it looks sweet and it has a beautiful color.’

7. pɔ tɔj nadj ?àap tɛ pěn màak màak-kam-khɔŋ huu mān kɔ'
  when look this may IRR be fruit apple or 3SG CONN

màak chomphu nɔ jə wkɔ ...
fruit rose apple PRT LINK

‘Looking at this fruit it's most likely apples or rose apples.’
8. jâwkó ?ăn luk?ìn nân kò mán kò ?ám jēn laj mán
LINK CLF child that CONN 3SG CONN NEG patient ACQUIRE 3SG
kō jēk āw koj nân sàj tāj sàj lōttʰip he
CONN lift take basket that put on put bicycle PRT

‘And so that boy he wasn't patient, he picked the basket up and put it on his bicycle.’

9. măn măn măn cà kōn kō ?ān kép jē ti nō tonmâj măn
same same same IRR people CLF CLF gather stay at on fruit same
IRR may IRR see stay PRT CLF child that lift CLF basket fruit
sāj lōttʰip mān kō ?ām wā sāj ...
put bicycle 3SG CONN NEG say what

‘It was kind of like the person who was gathering fruit on top of the tree might of seen him – the boy, the one who took the fruit, but he didn't say anything.’

10. jâwkó mān kō pān luk?ìn nān jēk màakmâj jâwkó sāj lōttʰip
LINK 3SG CONN give child that lift fruit LINK put bicycle

‘And then... he (might of) had that boy put the fruit on his bicycle’

11. jâwkó pān kwà ... kāmnañ kwà kō mí jēj ?ōn
LINK pass go at.this.time go CONN have girl young
tʰeŋ kā pan lōttʰip mā mōn kān sa mā ...
another/again CLF pedal bicycle come same together and come

‘And then he passes by, at this time there's also a girl who is riding a bicycle too’

12. phâa niā mān it nuŋ pʰâa niā ?aj kō ?ān lāk màakmâj
crash meet 3SG little one crash meet boy CONN CLF steal fruit
kōn nān nā ... jâwkó kāmnañ lōt ?aj kō màakmâj pīn
people that PRT LINK at.this.time vehicle boy CONN fruit fall
They crash into each other a little bit, and the boy who stole the fruit, his bicycle falls down.

At this time when the boy who stole the fruit falls, his bicycle falls too and the basket of fruit spills with all the fruit spills out all over the place.

Then there are these 2 or 3 boys who come and help him.

They help him gather the fruit and put it back in the basket and put it back up on the bicycle.
'The boy goes off, but at this time that he's leaving, over by the 3 boys, they see that
the boy forgot his hat, so the boys say “Hey, Hey, Hey, you forgot your hat!”.'

'So then the boy runs and gets his hat to give back to him'

'He then gives them some of the fruit, three pieces for the group of boys that came
and helped him. The boy who was playing (paddleboard), he gave the boy the hat
and then the boy gave him 3 pieces of fruit.'

'And then they go back to that place again (the tree), (Unfortunately I forget what
happens next’)

20. lúŋ kón k₃ nán caŋ mán caj-lum kʰa lukʔn kón nán uncle people CLF that together 3SG angry PRT child people that
ʔǎw màak sỳŋ kwà sỳŋ koj màakmâj mán kwà take fruit two go two basket fruit 3SG go

‘That man he was angry. The boy took two baskets of his baskets of fruit.’

21. àm tan hǎn mán waa hû nə àm tan hǎn waa hû ... NEG in.time see 3SG say know PRT NEG in.time see say know
jâwkɔ mán tẹ tẹ caj-lum kwe-na ... LINK 3SG IRR IRR angry because

‘He didn't see what happened or know what happened, he's angry because…’

22. mán lóŋ má kaj ton jâwkɔ tój caj-lum wâj jù 3SG descend come want time LINK look angry keep PROG
caj-lum tẹ caj-lum wâj màakmâj mán angry IRR angry keep fruit 3SG

‘He comes down from the tree, and then he looks angry, he's angry about his fruit.’

23. mán tój mán hǎa tój màakmâj màakmâj kâw nâj kwâ koj 3SG look 3SG search look fruit fruit 1SG this go basket
nuŋ nâj nân ?ǎn nâj kɔ́ mán tẹ pên ?ǎ,pûm kàj waa kón one this that CLF this CONNECT 3SG IRR be story near say people
ti ʔàm caj khɔŋ haw ya pe ʔǎw nâj lanaw kàj waa lâk REL NEG yes thing 1SG don't take take this how near say steal
khɔŋ pən ʔàm lî thing people NEG good
‘He's looking and searching for his fruit, and where the basket of fruit had gone. This is a story that's showing people who take things that that don't belong to them and how stealing things that don't belong to you is not good.’

24. jâwkọ to mán läk kwà jâw pâj tan pôn ka hu mán pó
   LINK next 3SG steal go ASP yet in.time pass PRT or 3SG when
   kăn jâa lôt têŋ kăn nuŋ mà con mán lôm
   together meet vehicle another/again together one come crash 3SG fall
   kwà
   go

‘For after he stole the fruit he crashed into the another bicycle, and it fell down.’

25. nâj mân kâj waa haw hét pit ?i.saj ko jâw ?ám huŋ ?ám
   this 3SG near say 1SG do mistake what CONN ASP NEG long NEG
   naan tak të pân haw maj hét pʰít nân
   long for IRR CAUS 1SG 1SG do mistake that

‘This is trying to show how if we do something wrong, it won't be very long until, something will happen that will makes us not do like that again.’
APPENDIX C

TOP 100 WORDS IN SHAN AND THAI CORPORAS

<table>
<thead>
<tr>
<th>#</th>
<th>Shan</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mán</td>
<td>‘it’</td>
</tr>
<tr>
<td>2</td>
<td>kɔ̀</td>
<td>‘CONN’</td>
</tr>
<tr>
<td>3</td>
<td>má</td>
<td>‘to come’</td>
</tr>
<tr>
<td>4</td>
<td>jâwkɔ̀</td>
<td>‘LINK’</td>
</tr>
<tr>
<td>5</td>
<td>kwà</td>
<td>‘to go’</td>
</tr>
<tr>
<td>6</td>
<td>nân</td>
<td>‘that’</td>
</tr>
<tr>
<td>7</td>
<td>kón</td>
<td>‘person’</td>
</tr>
<tr>
<td>8</td>
<td>mî</td>
<td>‘to have’</td>
</tr>
<tr>
<td>9</td>
<td>sâm</td>
<td>‘three’</td>
</tr>
<tr>
<td>10</td>
<td>kǭ</td>
<td>‘CLF. person’</td>
</tr>
<tr>
<td>11</td>
<td>ʔǎn</td>
<td>‘CLF’</td>
</tr>
<tr>
<td>12</td>
<td>ʔǎw</td>
<td>'to take'</td>
</tr>
<tr>
<td>13</td>
<td>màkmâj</td>
<td>‘fruit’</td>
</tr>
<tr>
<td>14</td>
<td>nâj</td>
<td>‘this’</td>
</tr>
<tr>
<td>15</td>
<td>ti</td>
<td>‘place’</td>
</tr>
<tr>
<td>16</td>
<td>mûj</td>
<td>‘one’</td>
</tr>
<tr>
<td>17</td>
<td>wa</td>
<td>‘say’</td>
</tr>
<tr>
<td>18</td>
<td>ʔàm</td>
<td>‘NEG’</td>
</tr>
<tr>
<td>19</td>
<td>hân</td>
<td>‘to see’</td>
</tr>
<tr>
<td>20</td>
<td>kép</td>
<td>‘to collect’</td>
</tr>
<tr>
<td>21</td>
<td>lukʔɔ̀n</td>
<td>‘child’</td>
</tr>
<tr>
<td>22</td>
<td>no</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>23</td>
<td>na</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>#</td>
<td>Word</td>
<td>Gloss</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>24</td>
<td>ka</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>25</td>
<td>té</td>
<td>‘IRR’</td>
</tr>
<tr>
<td>26</td>
<td>khun</td>
<td>‘to ascend’</td>
</tr>
<tr>
<td>27</td>
<td>kámną́j</td>
<td>‘this moment’</td>
</tr>
<tr>
<td>28</td>
<td>kha</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>29</td>
<td>he</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>30</td>
<td>koj</td>
<td>‘CLF’</td>
</tr>
<tr>
<td>31</td>
<td>hâm</td>
<td>‘again’</td>
</tr>
<tr>
<td>32</td>
<td>jąw</td>
<td>‘to finish’</td>
</tr>
<tr>
<td>33</td>
<td>khâw</td>
<td>‘3.SG/PL’</td>
</tr>
<tr>
<td>34</td>
<td>cą́j</td>
<td>‘to help’</td>
</tr>
<tr>
<td>35</td>
<td>jù</td>
<td>‘BE.AT’</td>
</tr>
<tr>
<td>36</td>
<td>theŋ</td>
<td>‘another’</td>
</tr>
<tr>
<td>37</td>
<td>lâthhip</td>
<td>‘bicycle’</td>
</tr>
<tr>
<td>38</td>
<td>tój</td>
<td>‘to look’</td>
</tr>
<tr>
<td>39</td>
<td>pęn</td>
<td>‘to be’</td>
</tr>
<tr>
<td>40</td>
<td>sőŋ</td>
<td>‘two’</td>
</tr>
<tr>
<td>41</td>
<td>màakhố</td>
<td>‘hat’</td>
</tr>
<tr>
<td>42</td>
<td>cą́j</td>
<td>‘male’</td>
</tr>
<tr>
<td>43</td>
<td>laj</td>
<td>‘able’</td>
</tr>
<tr>
<td>44</td>
<td>luk</td>
<td>‘CLF’</td>
</tr>
<tr>
<td>45</td>
<td>pęn</td>
<td>‘to give’</td>
</tr>
<tr>
<td>46</td>
<td>khun</td>
<td>‘to return’</td>
</tr>
<tr>
<td>47</td>
<td>mǎn</td>
<td>‘to be similar’</td>
</tr>
<tr>
<td>48</td>
<td>cuŋ</td>
<td>‘to pull’</td>
</tr>
<tr>
<td>49</td>
<td>lōm</td>
<td>‘to fall’</td>
</tr>
<tr>
<td></td>
<td>Shan</td>
<td>Thai</td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>50</td>
<td>lóŋ</td>
<td>loŋ</td>
</tr>
<tr>
<td></td>
<td>'to descend'</td>
<td>'to descend'</td>
</tr>
<tr>
<td>51</td>
<td>wâj</td>
<td>tua</td>
</tr>
<tr>
<td></td>
<td>'to place'</td>
<td>'CLF'</td>
</tr>
<tr>
<td>52</td>
<td>pôn</td>
<td>ná</td>
</tr>
<tr>
<td></td>
<td>'to pass'</td>
<td>'PRT'</td>
</tr>
<tr>
<td>53</td>
<td>?₃n</td>
<td>klâj</td>
</tr>
<tr>
<td></td>
<td>'to be young'</td>
<td>'to be close'</td>
</tr>
<tr>
<td>54</td>
<td>khi</td>
<td>naj</td>
</tr>
<tr>
<td></td>
<td>'to ride'</td>
<td>'in'</td>
</tr>
<tr>
<td>55</td>
<td>təŋ</td>
<td>kàp</td>
</tr>
<tr>
<td></td>
<td>'all'</td>
<td>'with'</td>
</tr>
<tr>
<td>56</td>
<td>pʰɛ</td>
<td>khuuu</td>
</tr>
<tr>
<td></td>
<td>'pear'</td>
<td>'to be'</td>
</tr>
<tr>
<td>57</td>
<td>lâk</td>
<td>hëń</td>
</tr>
<tr>
<td></td>
<td>'to steal'</td>
<td>'to see'</td>
</tr>
<tr>
<td>58</td>
<td>sâj</td>
<td>phûuŋ</td>
</tr>
<tr>
<td></td>
<td>'to put in'</td>
<td>'female'</td>
</tr>
<tr>
<td>59</td>
<td>tə</td>
<td>rúu</td>
</tr>
<tr>
<td></td>
<td>'but'</td>
<td>to know</td>
</tr>
<tr>
<td>60</td>
<td>màːk</td>
<td>rûuu</td>
</tr>
<tr>
<td></td>
<td>'CLF'</td>
<td>'or'</td>
</tr>
<tr>
<td>61</td>
<td>hët</td>
<td>khamooj</td>
</tr>
<tr>
<td></td>
<td>'to do'</td>
<td>'to steal'</td>
</tr>
<tr>
<td>62</td>
<td>ton</td>
<td>pàn</td>
</tr>
<tr>
<td></td>
<td>'tree'</td>
<td>'to pedal'</td>
</tr>
<tr>
<td>63</td>
<td>khûn</td>
<td>maŋ</td>
</tr>
<tr>
<td></td>
<td>'to give back'</td>
<td>'EMPH'</td>
</tr>
<tr>
<td>64</td>
<td>tûk</td>
<td>sîi</td>
</tr>
<tr>
<td></td>
<td>'CONT'</td>
<td>'color'</td>
</tr>
<tr>
<td>65</td>
<td>lûŋ</td>
<td>lóm</td>
</tr>
<tr>
<td></td>
<td>'uncle'</td>
<td>'to fall down'</td>
</tr>
<tr>
<td>66</td>
<td>pó</td>
<td>nîk</td>
</tr>
<tr>
<td></td>
<td>'enough'</td>
<td>'more'</td>
</tr>
<tr>
<td>67</td>
<td>ne</td>
<td>wâj</td>
</tr>
<tr>
<td></td>
<td>'PRT'</td>
<td>'to place'</td>
</tr>
<tr>
<td>68</td>
<td>thûŋ</td>
<td>kamläŋ</td>
</tr>
<tr>
<td></td>
<td>'to arrive'</td>
<td>'PROG'</td>
</tr>
<tr>
<td>69</td>
<td>pê</td>
<td>khêc</td>
</tr>
<tr>
<td></td>
<td>'goat'</td>
<td>'just'</td>
</tr>
<tr>
<td>70</td>
<td>tôk</td>
<td>chaaj</td>
</tr>
<tr>
<td></td>
<td>'to fall'</td>
<td>'male'</td>
</tr>
<tr>
<td>71</td>
<td>kîn</td>
<td>rûaj</td>
</tr>
<tr>
<td></td>
<td>'to eat'</td>
<td>'continually'</td>
</tr>
<tr>
<td>72</td>
<td>se</td>
<td>waâŋ</td>
</tr>
<tr>
<td></td>
<td>'PRT'</td>
<td>'to place'</td>
</tr>
<tr>
<td>73</td>
<td>ŋâa</td>
<td>khâaŋ</td>
</tr>
<tr>
<td></td>
<td>'to meet'</td>
<td>'side'</td>
</tr>
<tr>
<td>74</td>
<td>caj</td>
<td>tônmâaj</td>
</tr>
<tr>
<td></td>
<td>'male'</td>
<td>'tree'</td>
</tr>
<tr>
<td>75</td>
<td>pàn</td>
<td>bôkk</td>
</tr>
<tr>
<td></td>
<td>'to pedal'</td>
<td>'to tell'</td>
</tr>
</tbody>
</table>

166
<table>
<thead>
<tr>
<th>Shan</th>
<th></th>
<th>Shan</th>
<th></th>
<th>Thai</th>
<th></th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word</strong></td>
<td><strong>Gloss</strong></td>
<td><strong>#</strong></td>
<td><strong>Word</strong></td>
<td><strong>Gloss</strong></td>
<td><strong>#</strong></td>
<td><strong>Word</strong></td>
</tr>
<tr>
<td>76</td>
<td>kám</td>
<td>‘moment’</td>
<td>9</td>
<td>0.29</td>
<td>piin</td>
<td>‘to climb’</td>
</tr>
<tr>
<td>77</td>
<td>wôn</td>
<td>‘to think’</td>
<td>9</td>
<td>0.29</td>
<td>klàp</td>
<td>‘to return’</td>
</tr>
<tr>
<td>78</td>
<td>pît</td>
<td>‘to pluck’</td>
<td>9</td>
<td>0.29</td>
<td>chûaj</td>
<td>‘to help’</td>
</tr>
<tr>
<td>79</td>
<td>?aj</td>
<td>‘male’</td>
<td>8</td>
<td>0.26</td>
<td>khèŋ</td>
<td>‘basket’</td>
</tr>
<tr>
<td>80</td>
<td>hû</td>
<td>‘to think’</td>
<td>8</td>
<td>0.26</td>
<td>lâ?</td>
<td>‘each’</td>
</tr>
<tr>
<td>81</td>
<td>le</td>
<td>‘each’</td>
<td>8</td>
<td>0.26</td>
<td>thaanj</td>
<td>‘path’</td>
</tr>
<tr>
<td>82</td>
<td>se</td>
<td>‘PRT/to separate’</td>
<td>8</td>
<td>0.26</td>
<td>pheec</td>
<td>‘pear’</td>
</tr>
<tr>
<td>83</td>
<td>tàkaa</td>
<td>‘basket’</td>
<td>8</td>
<td>0.26</td>
<td>khôññ</td>
<td>‘thing/POSS’</td>
</tr>
<tr>
<td>84</td>
<td>sàu</td>
<td>‘to put in’</td>
<td>8</td>
<td>0.26</td>
<td>sianj</td>
<td>‘sound’</td>
</tr>
<tr>
<td>85</td>
<td>tàaaj</td>
<td>‘path’</td>
<td>8</td>
<td>0.26</td>
<td>phàa</td>
<td>‘cloth’</td>
</tr>
<tr>
<td>86</td>
<td>cakajan</td>
<td>‘bicycle’</td>
<td>8</td>
<td>0.26</td>
<td>chêt</td>
<td>‘to wipe’</td>
</tr>
<tr>
<td>87</td>
<td>màakhìn</td>
<td>‘rock’</td>
<td>8</td>
<td>0.26</td>
<td>khoñ</td>
<td>should</td>
</tr>
<tr>
<td>88</td>
<td>ca</td>
<td>‘IRR’</td>
<td>7</td>
<td>0.22</td>
<td>jàññjì</td>
<td>‘like this’</td>
</tr>
<tr>
<td>89</td>
<td>le</td>
<td>‘and’</td>
<td>7</td>
<td>0.22</td>
<td>tɔɔ</td>
<td>‘next’</td>
</tr>
<tr>
<td>90</td>
<td>háw</td>
<td>‘1SG’</td>
<td>7</td>
<td>0.22</td>
<td>tɔɔnj</td>
<td>‘at that time’</td>
</tr>
<tr>
<td>91</td>
<td>nàu</td>
<td>‘in’</td>
<td>7</td>
<td>0.22</td>
<td>sɔɔññ</td>
<td>‘two’</td>
</tr>
<tr>
<td>92</td>
<td>lam</td>
<td>‘to begin’</td>
<td>7</td>
<td>0.22</td>
<td>thàññ</td>
<td>‘all’</td>
</tr>
<tr>
<td>93</td>
<td>nọ</td>
<td>‘on, above’</td>
<td>7</td>
<td>0.22</td>
<td>khràp</td>
<td>‘PRT’</td>
</tr>
<tr>
<td>94</td>
<td>lì</td>
<td>‘to be good’</td>
<td>6</td>
<td>0.19</td>
<td>luññ</td>
<td>‘uncle’</td>
</tr>
<tr>
<td>95</td>
<td>mɔ</td>
<td>‘when’</td>
<td>6</td>
<td>0.19</td>
<td>hàaj</td>
<td>‘missing’</td>
</tr>
<tr>
<td>96</td>
<td>wàj</td>
<td>‘to turn’</td>
<td>6</td>
<td>0.19</td>
<td>plàaw</td>
<td>‘empty’</td>
</tr>
<tr>
<td>97</td>
<td>tonmàj</td>
<td>‘tree’</td>
<td>6</td>
<td>0.19</td>
<td>tem</td>
<td>‘full’</td>
</tr>
<tr>
<td>98</td>
<td>kàmnàñ</td>
<td>‘at that time’</td>
<td>6</td>
<td>0.19</td>
<td>càak</td>
<td>‘from’</td>
</tr>
<tr>
<td>99</td>
<td>khit</td>
<td>‘think’</td>
<td>6</td>
<td>0.19</td>
<td>tham</td>
<td>‘to do’</td>
</tr>
<tr>
<td>100</td>
<td>sam</td>
<td>‘and, then’</td>
<td>6</td>
<td>0.19</td>
<td>phûññ</td>
<td>‘elder’</td>
</tr>
</tbody>
</table>
RESUME

Name: Daniel Peter Loss

Date of Birth: 14th February 1984

Place of Birth: United States of America

Institutions Attended:
2006, A.A. General Education, Pensacola State College
2008, B.A. Philosophy, University of West Florida
2013, M.A. Education (TESOL), Chulalongkorn University