ZERO ANAPHORA AND MEITHEI

Terence Cockerham

Thesis Prepared for the Degree of

MASTER OF ARTS

UNIVERSITY OF NORTH TEXAS

December 2012

APPROVED:

Shobhana L. Chelliah, Major Professor
Sadaf Munshi, Committee Member
Haj Ross, Committee Member
Brenda Sims, Chair of the Department of Linguistics and Technical Communication
Mark Wardell, Dean of the Toulouse Graduate School
Cockerham, Terence, *Zero Anaphora and Meithei*. Master of Arts (Linguistics), December 2012, 73 pp., 2 tables, 2 figures, 26 titles.

The focus of this thesis is to determine what factors predict zero anaphora in Meithei. The data for this thesis is derived from pear stories. Arguments were tabulated in spreadsheets counting nouns, pronouns, and zero anaphors; they were also examined for their semantic role and information status. The findings showed the agent role was typically represented by reduced forms of reference, the majority of the time by zero anaphora. Other semantic roles were typically represented by lexical full forms of reference. Agents were strongly correlated to previous subjects. Other semantic roles were typically found in the other information status categories. The conclusion drawn from these findings is semantic role and information status influence accessibility, and accessibility determines whether or not arguments are represented by zero anaphors in Meithei.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF EXAMPLES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 LITERATURE REVIEW</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Defining Zero Anaphora</td>
<td>3</td>
</tr>
<tr>
<td>2.2 Cognitive Accessibility</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Agreement and Participant Coding</td>
<td>11</td>
</tr>
<tr>
<td>2.4 Privileged Syntactic Arguments</td>
<td>16</td>
</tr>
<tr>
<td>2.5 Literature Review Summary</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER 3 TYPOLOGICAL OVERVIEW OF MEITHEI</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER 4 DATA SOURCE AND ANALYSIS</td>
<td>30</td>
</tr>
<tr>
<td>4.1 The Pear Story Data Collection</td>
<td>30</td>
</tr>
<tr>
<td>4.2 Interlinear Text to Tabular Format</td>
<td>32</td>
</tr>
<tr>
<td>4.3 Data Coding</td>
<td>33</td>
</tr>
<tr>
<td>4.4 Variance among Speakers</td>
<td>35</td>
</tr>
<tr>
<td>4.5 Data Analysis</td>
<td>36</td>
</tr>
<tr>
<td>4.6 Semantic Roles for Full and Reduced Forms</td>
<td>38</td>
</tr>
<tr>
<td>4.7 Full and Reduced Forms within Semantic Roles</td>
<td>40</td>
</tr>
<tr>
<td>CHAPTER 5 CONCLUSION</td>
<td>46</td>
</tr>
<tr>
<td>APPENDIX: PEAR STORY #1</td>
<td>48</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>71</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Representations of NPs from Least to Most Accessible</td>
</tr>
<tr>
<td>2.</td>
<td>Greek Agreement System with the Option to Omit NPs</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Possible Forms for Representing a Referent</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Snapshot of Coding Spreadsheet</td>
<td>33</td>
</tr>
<tr>
<td>3.</td>
<td>Forms of Representation for NPs by Speaker</td>
<td>36</td>
</tr>
<tr>
<td>4.</td>
<td>Reduced Forms of Reference by Semantic Role</td>
<td>39</td>
</tr>
<tr>
<td>5.</td>
<td>Pronouns by Semantic Role</td>
<td>39</td>
</tr>
<tr>
<td>6.</td>
<td>Zero Anaphora by Semantic Role</td>
<td>40</td>
</tr>
<tr>
<td>7.</td>
<td>Agents and Non-agents by Full and Reduced Forms of Reference</td>
<td>41</td>
</tr>
<tr>
<td>8.</td>
<td>Agents by Form of Reference</td>
<td>41</td>
</tr>
<tr>
<td>9.</td>
<td>Patients by Form of Reference</td>
<td>42</td>
</tr>
<tr>
<td>10.</td>
<td>Locations by Form of Reference</td>
<td>42</td>
</tr>
<tr>
<td>Example</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Spanish Agreement System with the Option to Omit NPs</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Russian Agreement System without the Option to Omit NPs</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>South Asian Agreement Systems with the Option to Omit NPs</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Chinese Absence of Agreement with the Option to Omit NPs</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Non-Nominative Case Blocks Agreement in Nepali</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Non-Nominative Case Blocks Agreement in Russian</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Frequent Zero Anaphora in Belhare</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Infrequent Zero Anaphor Production in Russian</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Infrequent Zero Anaphor Production in Maithili</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Absence of Case Marking and Agreement in Meithei</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Excerpt from a Meithei Pear Story in Interlinear Text Format</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

In Meithei clauses (Tibeto-Burman, Northeast India), it is possible to leave required arguments of a verb phonetically unexpressed. This thesis attempts to study one type of discourse, samples of Pear Story retellings, to determine what semantic and discourse factors determine zero anaphora in Meithei. In terms of semantics, I look at how the Agent role occurs more often as an unexpressed semantic role than other Noun Phrases, henceforth NPs, and I review literature which explains why this might be so. In terms of discourse, I look at the distance between mentions of the same entity and test if distance might have some connection to the possibility for unexpressed arguments. I find that the Meithei data provides good motivation for considering semantic role and discourse placement as factors which can determine the presence of unexpressed arguments. In previous studies these factors have been referenced but not thoroughly investigated in conjunction with each other.

This thesis is organized as follows: In section 2, I provide a review of previous research that defines what zero anaphors are and states how the following factors are predictive of where zero anaphors may occur: cognitive accessibility; the presence of subject–verb person, number, or gender agreement; and the presence of case marking on NPs. All of these enter into the prediction of the possibility of zero anaphora. In section 3, I review the data to be considered. I include background typological information about Meithei, the data source, data coding and analysis. In section 4, I compare the
Meithei data with the occurrence of zero anaphora in other languages and evaluate whether or not these factors play a role in zero anaphor production for Meithei.
CHAPTER 2
LITERATURE REVIEW

2.1 Defining Zero Anaphora

Referents can be represented by full (lexical) forms of reference or reduced forms of reference. Reduced forms of reference can be zero anaphors or pronouns. Pronouns can be either free or bound. These forms of an NP are listed in Figure 1.

Figure 1: Possible Forms for Representing a Referent

```
Referent
 / \
Full Reduced
 / \
Zero Pronoun
 / \
Free Bound
```

In this section, I provide a definition of zero anaphor, specifically distinguishing between obligatory omissions of NPs and optional ellipses.

Zero anaphors are NPs that are absent from the surface form of the sentence but are still needed to fulfill the argument structure of a verb. In the following Spanish sentences, for example, the subject noun phrases are not overtly expressed. However, there is no problem with interpretation.¹

¹ Abbreviations used in the following Spanish examples are as follows: ART ‘article’; DEF ‘definite’; M ‘masculine’; P ‘plural’; PRS ‘present’; S ‘singular.’
Example 1: Spanish Agreement System with the Option to Omit NPs

 Leo los libros.  
∅ leo los libros  
∅ le-o los libro-s  
[1s] read–1S.PRS M.P.DEF.ART book–P  
‘I read the books.’

Lees los libros.  
∅ lees los libros  
∅ le-es los libro-s  
‘You read the books.’

Lee los libros.  
∅ lee los libros  
∅ le-e los libro-s  
‘S/he reads the books.’

Free stranding words that stand in for nouns but cannot define them are sometimes called free pronouns. Verb affixes that carry the same kind of information as free pronouns such as number and person are sometimes described as bound pronouns. Many languages have both bound pronouns and free pronouns. This can occur in a language like Spanish, where nouns or free pronouns are optional with the presence of bound pronouns, or it can occur in a language like Russian, where nouns or free pronouns are normally required even though the bound pronoun may carry a great deal of information such as
person, number, or gender. Compare these examples from Russian to the previous Spanish ones.²

Example 2: Russian Agreement System without the Option to Omit NPs

Я читаю книги
¹a čita¹u knigi
¹a čita⁻¹u knig⁻i
1s.nom read–1sNPST.IPFV book–fp.ACC
‘I read the books.’

Ты читаешь книги
ti čita⁵eš knigi
ti čita⁻⁵eš knig⁻i
2s.nom read–2sNPST.IPFV book–fp.ACC
‘You read the books.’

Он читает книги.
on čita⁵et knigi
on čita⁻⁵et knig⁻i
3s.nom read–3sNPST.IPFV book–fp.ACC
‘He reads the books.’

Pronouns may also be free, as seen in the following examples from English. Free pronouns can stand in for, but cannot define referents, or occur without antecedents as in English I read the book.

In languages like Meithei, Japanese, and Chinese, logical subjects, direct objects, and indirect objects may occur as full NPs, pronouns, or zero anaphors.

² Abbreviations for Russian examples are: ACC ‘accusative’; F ‘feminine’; IPFV ‘imperfective’; NOM ‘nominative’; NPST ‘nonpast’; P ‘plural’; S ‘singular’.
The question is: What environments allow for the noun phrases to be full NPs, pronouns, or zero anaphors?

Note that this thesis deals with NP omissions that are optional, not those that are required. Optional omissions, often called ellipses, are characterized in Yamamoto (1999) as:

- Precisely recoverable
- Grammatically ‘defective’
- Constructions where the insertion of the missing expression results in a grammatical sentence with the same meaning as the original sentence
- Textually recoverable missing words are present elsewhere in the text in exactly the same form

Yamamoto further divides optional omission into three categories: textual, situational, and structural. Textual ellipses are the most prototypical kind of ellipsis. If a sentence has all of the features that characterize ellipses, then it is an example of textual ellipsis. Situational ellipses recover the omitted words from the situation outside of the text e.g. the deictic quality of conversation in a sentence like *(I am) Pleased to meet you*. Structural ellipses typically omit function words e.g. *I am happy (that) you are here*.

This is very different from obligatory omission such as in the subject of the infinitive in the English sentence *John wants ___ to avoid dysentery*. Here the syntax requires that the noun phrase be a zero anaphor. Although this sentence has an omission, it is not an ellipsis because the insertion of the omitted word produces the ungrammatical sentence *John wants John, to avoid*
dysentery. Additionally, English subjects of imperatives also have omissions, but they cannot be categorized as ellipses because the omitted NPs are not optional; the insertion of the omitted NP into an imperative sentence would raise a grammaticality question and change the meaning of the sentence slightly by elevating the level of rudeness. Subjects of infinitives and subjects of imperatives are omissions of the required type.

2.2 Cognitive Accessibility

In Talmy Givón's topicality continuum (see Arnold (1998) for a useful summary), it is argued that forms of reference from zero anaphors to referential NPs fall on a continuum. Full NPs are used when an entity is less cognitively accessible to the speaker. When NPs are more accessible to speakers, pronouns or zero anaphors are acceptable. Table 1 shows the scale from Ariel’s (1990) *Accessing Noun-phrase Antecedents*, supplemented by examples modeled after the ones given in Arnold (1998). The NPs earlier in the list would be used when the hearer knows less about the referent, and the NPs later in the list would be used when the speaker knows more about the referent. So, NP shape relates to the amount of information the hearer has.
<table>
<thead>
<tr>
<th>Marking Scale</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full name + modifier</td>
<td>Joan Smith, the president</td>
</tr>
<tr>
<td>Full name</td>
<td>Joan Smith</td>
</tr>
<tr>
<td>Long definite description</td>
<td>The tall and authoritative president</td>
</tr>
<tr>
<td>Short definite description</td>
<td>The president</td>
</tr>
<tr>
<td>Last name</td>
<td>Smith</td>
</tr>
<tr>
<td>First name</td>
<td>Joan</td>
</tr>
<tr>
<td>Distal demonstrative + modifier</td>
<td>That politician we elected last year</td>
</tr>
<tr>
<td>Proximal demonstrative + modifier</td>
<td>This politician we elected last year</td>
</tr>
<tr>
<td>Distal demonstrative + NP</td>
<td>That politician</td>
</tr>
<tr>
<td>Proximal demonstrative + NP</td>
<td>This politician</td>
</tr>
<tr>
<td>Distal demonstrative</td>
<td>That</td>
</tr>
<tr>
<td>Proximal demonstrative</td>
<td>This</td>
</tr>
<tr>
<td>Stressed pronoun + gesture</td>
<td>SHE (plus gesture)</td>
</tr>
<tr>
<td>Stressed pronoun</td>
<td>SHE</td>
</tr>
<tr>
<td>Unstressed pronoun</td>
<td>she</td>
</tr>
<tr>
<td>Cliticized pronoun</td>
<td>(no examples in English)</td>
</tr>
<tr>
<td>Extremely High Accessibility Markers</td>
<td>gaps, including pro, PRO and wh-traces, reflexives, and agreement</td>
</tr>
</tbody>
</table>

In Givón’s theory, a cognitively accessible noun phrase is one which is a continued topic in the discourse. NP topicality is thought of as a continuum. Topicality is defined as a combination of distance of mention (a more topical NP
is more recently mentioned than a less topical NP) and interference (a more
topical NP will not be preceded by a noun phrase with a potentially ambiguous
referent. For example, in Sally knew she was going to the party it is not clear if
she can refer back to Sally or to some other person. In this case she is less
topical because the antecedent is ambiguous). Speakers usually choose a
lexical form the first time a referent is mentioned in discourse but for more
topical forms where there is no possibility of ambiguity, reduced NPs pronouns
or zero anaphors will be used. This reflects the speaker’s belief that the listener
understands who or what the argument is.

Arguments that have the semantic role of agent are typically highly
topical. Thus, as pointed out by Arnold (1998), agents are likely to occur as full
forms in first mentions, but when their information status is old (they have
been previously mentioned in the discourse), they are much more likely to
occur as reduced forms rather than NPs in other semantic roles. The fact that
agents tend to be in reduced forms was also pointed out by DuBois (1987), who
observes a correlation between agents and old information. He states that
languages tend to introduce characters in intransitive clauses. Once the
character takes on an agent role, it will already have been mentioned, be highly
accessible, and thus there is a strong probability for that agent NP to be a zero
anaphor. He connects this to the idea that languages tend to prefer having
fewer arguments per clause, so they are motivated to omit one if possible.

In addition to semantic role, the distance of mention seems important.
Givón’s theory is based on this, but in Arnold it is more clearly laid out; she
differentiates between an NP mentioned in an immediately preceding clause and an NP mentioned before that. This can be tied together with the idea of semantic role. When an NP is the logical subject (agent or actor) of the previous clause, or to a somewhat lesser extent an object of the previous clause, it is very likely to be provided in a reduced form in the following clause. Consider the two following sentences: *John woke up early. He went for a jog in the park.* In the first sentence, the referent is a first mention, so it is represented by a full NP. In the second sentence, the referent is the subject of the previous clause. It has the same referent as the first sentence, but it is represented by a reduced NP.

If the NP is mentioned further back than the immediately previous clause it is more likely to be represented by a full NP. The reasoning here is that the hearer would no longer be able to access that NP and properly interpret a clause because of either forgetting what the reduced form refers to or having intervening nouns confuse its interpretation. Arnold (1998) also points out NPs that are old information have greater accessibility than first mentions. She categorizes inferable or assumable information as the same as old information because both types are kinds of shared knowledge.

In order to look at the correlations between information status, distance of mention, and thematic roles of NPs in Mapudungun (Araucanian, Chile & Argentina), Arnold (2003) divides NPs into four categories: New – the first mention of a referent in a text, Previous Subject – the subject of the previous clause, Old – the referent has been previously mentioned but not in the
previous clause, and Active – the object of the previous clause or part of the subject or object of the previous clause or both the subject and object of the previous clause. This coding scheme has been applied to the data from Meithei in order to examine it for explanations for the production of zero anaphora.

Other researchers have looked away from discourse and more towards syntax and morphology to find what factors seem to predict the possibility of zero anaphora. In two ways of looking at the distribution of reduced versus full noun phrases, inflectional morphology seems to have an influence. In one avenue of study, number, person, and gender in verbal agreement has been claimed to allow for zero anaphora. In another program of study, case marking on nouns has been argued to decrease the amount of zero anaphora possible.

2.3 Agreement and Participant Coding

There have been various explanations offered for the presence or absence of zero anaphors based on verb agreement. One explanation often given for the production of zero anaphors is that verb affixation that agrees with the subject provides the listener with enough information about the referent to allow it to be implicit rather than explicit. Alexiadou and Anagnostopoulou (1998) say this explanation views verbal agreement affixes that supply information such as person and number are seen as having the same status as pronouns. Alexiadou and Anagnostopoulou supply the following example from Greek.
Table 2: Greek Agreement System with the Option to Omit NPs

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>agapo ‘I love’</td>
<td>agapame ‘we love’</td>
</tr>
<tr>
<td>2nd</td>
<td>agapas ‘you love’</td>
<td>agapate ‘you love’</td>
</tr>
<tr>
<td>3rd</td>
<td>agapa ‘he loves’</td>
<td>agapane ‘they love’</td>
</tr>
</tbody>
</table>

The Greek verb forms given above have distinct suffixes for first, second, and third persons, singular and plural. Greek is a language that has the ability to optionally omit subjects of sentences. The verbal affixes were believed to allow for zero anaphora because they made overt nouns or pronouns representing subjects redundant. Those verbal affixes that supply listeners with information about referents such as number, person, or even gender are sometimes described as bound pronouns.

However, languages that generate zero anaphors at levels contrary to expectations based on this explanation can be easily found. For example, Russian has a subject agreement system that carries information about person and number, but it often avoids zero anaphors. They can be found in idiomatic phrases similar to English phrases such as gone fishing or back in five minutes. Nevertheless, even though implicit representations of arguments are permitted in Russian, Russian speakers explicitly represent arguments the majority of the time. We should expect to find more zero anaphors as subjects of Russian sentences, yet Russian does not produce the kind of high level zero anaphora
like that found in languages like Belhare. The explanation that subject agreement affixation with person and number information produces zero anaphora seems plausible for a languages like Greek, but fails to explain why other languages with the same kind of agreement such as Russian do not produce zero anaphora at the same level under the same circumstances.

In addition, based on examples from South Asian languages, Butt (2001) shows there is no correlation between agreement systems and the extent of zero anaphora. She observes that if agreement allows for representing arguments with zero anaphors, then different types of agreement systems should produce different results. Instead, she finds that the ability to represent arguments with zero anaphors is found in a variety of South Asian languages irrespective of the agreement system. The type of system does not change the extent of use of zero anaphors. Compare the following examples, the first three of which are all answers to the question *Did you give Nadya some food?* Notice that even when the agreement system is different (as characterized in parentheses) the same arguments can be omitted in the answer.

Example 3: South Asian Agreement Systems with the Option to Omit NPs

Urdu/Hindi (number, gender, and person agreement, rising intonation)

\[
tum=ne \quad nadya=ko \quad kʰana \quad di=ya
\]

\[
you=Erg \quad Nadya.F=Dat \quad food.M.Sg.Nom \quad give-Perf.M.Sg
\]

‘Did you give Nadya some food?’

\[
Ji \quad di=ya
\]

Yes.Polite \quad give.Perf.M.Sg

‘Yes, [I] gave [it].’
Punjabi (number, gender, and person)
ji dttta
yes.Polite give.Perf.M.Sg
‘Yes, [I] gave [it].’

Bengali (Person agreement only)
hæ di-e-tʃʰ-i
yes give-Part-Perf-3
‘Yes, [he] gave [it].’

Kashmiri (Pronominal clitics which indicate number, person, and case)³
raath vuch-n-ay
yesterday saw-3Sg-2Sg
‘He saw you yesterday.’

Additionally, there are languages with no supplementary morphology which can still delete pronouns. Li (1997) gives the following example from Late Archaic Chinese.

Example 4: Chinese Absence of Agreement with the Option to Omit NPs

<table>
<thead>
<tr>
<th>子</th>
<th>曰</th>
<th>見</th>
<th>賢</th>
</tr>
</thead>
<tbody>
<tr>
<td>zi</td>
<td>yue:</td>
<td>jian</td>
<td>xian</td>
</tr>
<tr>
<td>Confucius</td>
<td>say</td>
<td>see</td>
<td>wise</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>思</th>
<th>齊</th>
<th>焉</th>
</tr>
</thead>
<tbody>
<tr>
<td>si</td>
<td>qi</td>
<td>yan</td>
</tr>
<tr>
<td>think</td>
<td>equal</td>
<td>PART</td>
</tr>
</tbody>
</table>

³ A pronominal clitic is a morpheme that has the meaning of a pronoun but is phonologically bound, like an affix.
Confucius said: [When one] perceives wisdom, [one] thinks of equaling [it]. [When one] perceives the opposite of wisdom, [one][turns] inward [and] examines oneself.'

Argument marking on the verb as in the pronominal affixes in Kashmiri may be more predictive of the extent of zero anaphora in a language. Seifart (2012) examines several languages with different kinds of argument marking and measures this against the ratio of nouns and pronouns to verbs. In connection with this work, Seifart hypothesizes that the more extensive the system of argument marking, the less likely are nouns or pronouns used. His preliminary findings show a correlation between the level of argument marking on verbs and the level of overt representations of arguments. He examines data from two languages with both subject and object indexing on verbs Baure (Arawakan) and Chintang (Tibeto-Burman), one language Bora (Boran, northeastern Peru), which only has subject indexing on verbs, and two languages without subject or object indexing on verbs N|uu (Southern Khoisan) and Sri Lanka Malay (Malayan Creole). His preliminary findings showed for the languages he investigated that a high degree of argument marking correlated with the most zero anaphors; languages where a midlevel of argument marking
had a midlevel of zero anaphora, and the languages without any argument marking had the least zero anaphora.\(^4\)

2.4 Privileged Syntactic Arguments

In another approach to the study of where non-overt NPs can occur, Bickel (2003) suggests that languages with extensive direct case marking on NPs have more overt NPs. He hypothesized that case-marking causes a speaker to become habituated in choosing the appropriate case-marking affixation, which results in an increase in the awareness of the referents that they are choosing affixation for, thus resulting in the generation of more lexical forms of representation. For instance, agents and experiencers normally trigger agreement, while other arguments do not. In this sense, agents or experiencers are more “privileged” than arguments that are patients or stimuli. To use Bickel’s terminology, they are privileged syntactic arguments. For example, in Nepali, if a first-person agent of a clause is in the nominative case, it triggers first-person agreement with the verb. However, with a first-person experiencer, which takes the dative case, the verb defaults to a third-person singular masculine form. Example 5 shows Bickel’s (2003) example from Nepali. Example 6 show Stoll’s similar examples from Russian.\(^5\)

---

\(^4\) Surprisingly, Seifart also found that although fieldworkers have determined that Sri Lanka Malay has the ability to represent any argument with a zero anaphor in reality it seldom does.

\(^5\) The transcriptions of the Russian examples in this thesis have been modified in order to make them consistent with one another.
Example 5: Non-Nominative Case Blocks Agreement in Nepali

ma bhut sāṅga ḍarā-ē
1s.NOM ghost with fear-1s.PT
‘I was afraid of the ghost.’

ma-lāī bhut sāṅga ḍar lāg-yo (*lāg-ē)
1s-DAT ghost with fear feel-3s.M.PT feel-1s.PT
‘I was afraid of the ghost.’ (To me the ghost was frightening.)

Example 6: Non-Nominative Case Blocks Agreement in Russian

Я хочу идти домой.
y a xoč-u idti domoy
1sNOM want-1sNPST.IPFV go.INF.IPFV home
‘I want to go home.’

Мне хочется идти домой.
mn’ye xoč-yet-s’ya idti domoy
1sDAT want-3sNPST.IPFV-REFL go.INF.IPFV home
‘I want to go home.’

These examples of experiencers in the dative case illustrate how an argument can prevent the triggering of agreement. Similar patterns occur in other South Asian languages. In Maithili, the agent of a clause, which takes the nominative case, will trigger a nominative form of agreement on the verb. However, if the logical subject is in the ablative case (this would make it a source), the verb takes a non-nominative form on the verb. Case (or semantic role marking) thus becomes crucial in producing a grammatical sentence. Bickel hypothesizes that overt noun phrases and case marking on these nouns
are necessary for getting agreement right.

Stoll and Bickel (2009) propose an explanation for why some languages that are expected to have more zero anaphors seldom do. Stoll and Bickel collected ten Pear Stories for each language from language consultants in rural communities, and broke down their texts into single-predicate clauses, and counted the arguments for each predicate; then they examined each instantiation to determine whether it was a free-standing noun or pronoun, or a zero anaphor.

The results of their analysis showed Belhare represented arguments with zero anaphora 59% of the time, 53% for Nepali, 38% for Maithili, and 33% for Russian. It should be noted that Stoll and Bickel describe zero anaphora in languages in terms of referential density, so they would say Russian has 67% referential density rather than Russian has 33% zero anaphora. Stoll and Bickel (2009) observed that in Russian, all human referents were introduced with explicitly provided NPs throughout the texts. They point out that Russian speakers create identifying labels for each referent at the time they are introduced in the stories. In Russian Pear Stories, when a character is mentioned for the first time, the speaker names that character e.g. ‘farmer’, ‘man with a goat’, or ‘boy on bike’. This in itself is not remarkable, given Givón’s idea of cognitive accessibility. However, whenever any previously mentioned character reappears later on in the discourse, Russian speakers consistently identify those characters by referring back to their original labels e.g. ‘farmer’, ‘goat-man’, or ‘bike-boy’. Russian speakers normally supply
lexical noun phrases for human referents and avoid zero anaphora.

In contrast, Belhare speakers did not typically do this. They either introduce human referents with a number like ‘one’ or even an ellipsis. Bickel (2003) provides the following Belhare example.6

Example 7: Frequent Zero Anaphora in Belhare

pəila ... a: ... ambibu
first PTCL mango[ABS]

phig–he kinahungo
[3s.A–]pick.from.above–PT[3O] SEQ

otutuiʔ=na jhola–e ukt–he
quite.big=ART[s] bag–LOC [3s.A–]take.down–PT[3O]

inetnahungo dhaki–e lenṣ–e
then closely.weaved.basket–LOC [3s.A–]put–PT[3O]

il–lam il–lam sas–sa–ba lenṣ–e

∧ni ... riksa, e: saikil–lamma, saikil–lamma
and.then rikshaw PTCL bicycle–MED bicycle–MED

ta–he kinahungo ...
[3s.S–]come–PT SEQ

6 Bickel’s Belhare Abbreviations are: A ‘most prominent argument of a transitive’; ABS ‘absolutive’; CONV ‘converb’; DEM ‘demonstrative’; DIST ‘distal’; LOC ‘locative’; M ‘mediative’; O ‘least prominent argument of a transitive’; PT ‘past’; PTCL ‘particle’; S ‘sole argument of an intransitive’; SEQ ‘sequential’. Square brackets in interlinear glosses are zero–marked features lacking an overt morpheme.
‘First, ... uh ... [someone] picked mangos and took [them] down in a big bag. Then [s/he] put [them] into a basket. [Someone] moved over [an animal] by pulling from over there, and then [someone] came on a rickshaw, uh ... on a bike, on a bike and then ....’

In the above translation, the segments in square brackets are zero anaphors. The example given above comes from the opening of a Pear Story. This example stretches the definition of a zero anaphor because the referent is provided in the form of an ellipsis in its first mention, so it does not even have a previously supplied noun or pronoun to refer back to. Bickel found Belhare speakers are often quite implicit in their representation of arguments. Listeners must infer the identity of these implicit arguments. Bickel demonstrates how Belhare speakers tend to not use lexical labels until there is some interaction between similar characters that requires disambiguation.

Stoll and Bickel (2009) found that Russian speakers created narratives with both higher overall overt NP representation and higher lexical NP representation than the Himalayan languages they examined. Russian speakers not only represented arguments with zero anaphors less often than speakers of Belhare, Nepali, and Maithili, but they also demonstrated a preference for nouns over pronouns. Although Russian has the same kinds of verb agreement that allow Spanish speakers to omit nouns and pronouns, Russian speakers do not typically do this. They usually refer explicitly to the arguments of predicates, even though the verb has affixation that agrees with the referent. The
following Russian example from Stoll and Bickel (2009) has been modified to conform to the other interlinear texts.

Example 8: Infrequent Zero Anaphor Production in Russian

Сад на реке видно деревня
sad na рекe vidno деревня

garden.nom on river.loc visible village.nom

хозяин собирает яблоки
хозяин собирать яблоки
owner.nom collect.3s npst.ipv apples

груши в саду
груши в саду

появляется
появляется

другой мужчина с козой
другой мужчина с козой

коza блеет они прошли
коza блеет они прошли

мужчина с козой прошли мимо
мужчина с козой прошли мимо

man.nom with goat instr go by.p past.ipv close
'A garden. Near the river a village is visible. The owner is collecting apples in the garden. He collected one basket. Another man with a goat appears. The goat is baa-ing. They went by. The man with the goat went by. And the owner of the garden went to collect the second basket. Here, a boy came towards them on a bicycle, probably his son.'

Notice that the Belhare example is full of zero anaphors. In the example, all humans are referred to by zero anaphors. The first and following mention of the farmer, the second and following mention of the fruit, the first and only mention of the goat-man, the goat, and the bike-boy are all zero anaphors. No human was referred to by a lexical item. By contrast, the Russian example is full of lexical NPs. The garden, river, village, farmer, pears, basket, goat-man, goat, bike-boy, and bike are all explicitly referred to by lexical forms for all
first mentions, and all but one subsequent mentions of humans were overtly referred to.

Bickel’s (2003) study of zero anaphor production in Himalayan languages depicts Belhare, Nepali, and Maithili rather like points on a scale. Belhare has the least affixation and overtly represents arguments approximately four out of every ten times. Nepali has more affixation and overtly expresses arguments roughly five out of every ten times. Maithili, also an Indo-European language, has the most affixation and overtly provides arguments about six out of every ten times. Consider the following example of Maithili from Bickel (2003).7

Example 9: Infrequent Zero Anaphor Production in Maithili

ek-ṭa ām-ke gāch rah-ai. ā...a...a...
one-CL mango-GEN tree[NOM] be-3NH.NOM[PR] PTCL

ām me ek e-gota chaurā ām tor-ait mango in one one-CL boy[NOM] mango pluck–IP rah-ai ā... u AUX–3NH.NOM[–3NH.NONNOM.PR] PTCL 3NH.NOM

ām toir–ke tokari me rakh–ne mango[NOM] pluck–CONV basket in keep–INF

jāi che–l–ai omaharse e–gotā chaura AUX AUX–PT–3NH.NOM[–3NH.NONNOM] and.then one–CL boy[NOM]

__________________________

7 Bickel’s Maithili abbreviations are: AUX ‘auxiliary’; CL ‘classifier’; CONV ‘converb’; GEN ‘genitive’; H ‘honorific’; INF ‘infinitive’; IP ‘imperfective participle’; NH ‘nonhonorific’; NOM ‘nominative’; PR ‘present’; PT ‘past’; PTCL ‘particle’.
'There is a mango tree and ...uh...uh in the mangos, one, a boy is picking mangos. And when picking mangos, he put them into a basket. Then a boy came, a young man riding on a bike, and he stole one basket of mangos, and took off...'

This example demonstrates Maithili’s tendency to avoid zero anaphora. Notice that the fruit identified as mangos have subsequent mentions also using the lexical form of reference. Also, human referents are provided by lexical forms in first mentions and pronouns in subsequent mentions, but not by zero anaphors. Of the three Himalayan languages investigated by Bickel, Maithili is the one with the lowest level of zero anaphor production and the one that models Russian most closely. It is considerably different from Belhare, where speakers often use zero anaphora and avoid the use of lexical items for referents until it becomes absolutely necessary in order to disambiguate two similar referents.

Stoll and Bickel (2009) suggest that the differences in zero anaphor production between Belhare and Russian are not matters of speaker style, but are general patterns that result from the structures of the languages that they speak. Russian is a language with a rich case system for its nouns. Bickel
suggests that frequent inflection of NPs may have a priming effect on speakers. Russian speakers may perceive a need to assign case for arguments of predicates in order to generate grammatically correct sentences, and this may lead to a higher use of lexical NPs. By contrast, Belhare speakers may use more reduced forms because their NPs do not carry case. So, structural differences produce differences in discourse style.

Bickel suggests that the use of case-marking may cause speakers to unconsciously think about agreement causing them to fill the argument instantiations for each predicate with explicitly provided NPs that agree with the verb. That is, the presence of noun cases in a language may cause speakers to focus more on the arguments of predicates and thus cause speakers to provide more explicitly stated NPs. Bickel concludes that zero anaphor production is not a matter of the choices of the individual speakers, but is determined by the typological profile of a language, and the absence of noun cases in a language may result in a more implicit representation of arguments as zero reference forms. Although Bickel's hypothesis involves some speculation on what the speaker is unconsciously doing, it seems quite plausible. A great deal of linguistic activity in any language occurs outside of the speakers' awareness.

However, the absence of affixation on nouns does not guarantee the usage of zero anaphora. English has almost no case marking, yet it is unusual to find English verbs without nouns or pronouns accompanying them. Although the presence of nominal inflection seems to contribute to an increase
in the overt representation of arguments, the absence of nominal inflection does not produce a decrease in overt representation of arguments.

2.5 Literature Review Summary

I reviewed three ways of thinking about the distribution of zero anaphors.

- Cognitive accessibility: the less information needed, the more is known about the NP
- Number, person, and gender agreement: features in agreement make an overt NP redundant
- Case marking: case marking rules that distinguish privileged from non-privileged NPs require overt NPs to make that distinction

Cognitive accessibility is the most useful in studying Meithei with respect to where zero anaphor occur as there is no number, person, gender in this language and also no other changes in the verb relative to the types of NPs in the claus
CHAPTER 3

TYPOLOGICAL OVERVIEW OF MEITHEI

Meithei is classified as a Tibeto-Burman language within the Sino-Tibetan language family. It is a language spoken in Northeastern India, primarily in Manipur. However, Meithei-speaking communities can also be found in nearby countries such as Bangladesh and Myanmar. Although Meithei speakers do not account for a large percentage of the population of these countries nationally, Meithei is a very important language regionally; it has well over a million speakers and functions as a language of broader communication.

Meithei is more agglutinating than fusional, with each of its morphemes marking a single grammatical category. Although its morphemes are usually easily segmented, Meithei is not an isolating language, because Meithei words are typically formed from a stem with the addition of prefixes and/or suffixes with words formed from morphemes strung together sequentially rather than as separate monomorphemic words.

Meithei is an agent–patient–verb ordered language. It has no obligatory case marking. Consider the child mentioned in the two following examples. In the first example, the child is the subject of the sentence and an agent. In the second example the three boys are the subject, and the child is the direct object and a beneficiary. Note that in neither of the sentences does the word ə́ŋáŋ meaning ‘child’ have any case marking. Also, as can be seen in the same examples, Meithei has no number, person, or gender agreement.
Example 10: Absence of Case Marking and Agreement in Meithei

Story 9, line 12

\[ \text{áŋáŋ} \ldots \text{naspati-du} \text{mitar-mmí} \]
\[ \text{áŋáŋ} \text{naspati-tu} \text{mita-lam-í} \]

child  pear-DDET  eye-EVD-NHYP

‘The child eyed those pears.’

Story 2, line 25

\[ \text{má–khóy} \text{á–hun-khak} \ldots \text{áŋáŋsi} \ldots \text{háw-gat-pi-l-e} \]

3P-hpl  ATT-three-UPTO  child-PDET  begin-UP-REC-PROX-ASRT

they  three  child-this  helped up

‘The three of them helped this child to get up.’

Story 6, line 11

\[ \text{núpá–máčá} \text{á–má} \text{saykəl} \]

nú  pá  mə–čá  saykəl

person  female  NM–child  bicycle
girl  bicycle

tháw-dúŋə  lakləmmí
tháw–dúŋə  lak–ləm–í
drive–ING  come–EVD–NHYP
driving  came

‘A girl came driving a bicycle.’

---

8 The abbreviations used in the Meithei examples are as follows: AGN ‘agentive’; ATT ‘attributive’; DDET ‘distal ‘determiner’; EVD ‘evidential’; GPL ‘general plural’; HPL ‘human plural’; LOC ‘locative’; NHYP ‘nonhypothetical’; PDET ‘proximal determiner’.
Many languages, Navajo (Native American, Athabascan), for instance, do not obligatorily represent arguments in the form of free-standing nouns or pronouns; rather, they have bound pronouns attached to verbs. If there is a free noun or pronoun, these bound pronouns will agree with them in number, person, and/or gender. However, Meithei does not have this type of pronominal agreement, so this type of information to assist listeners with the identification of referents is not available.
CHAPTER 4
DATA SOURCE AND ANALYSIS

In this section, I review the data used for this study. In order to study zero anaphor production in Meithei there were a few decisions that needed to be made about how to approach the research. Data from both Bickel’s (2003) and Stoll and Bickel’s (2009) previous research were both derived from pear stories. Meithei pear stories would thus be an ideal source of data for comparison to this previous research. I describe the coding and analysis procedure and end with an analysis of the distribution of zero anaphora based on semantic roles.

4.1 The Pear Story Data Collection

The data used for this analysis consist of Pear Stories recorded and transcribed by Dr. Harimohon Thounaojam and Dr. Shobhana Chelliah in 2007. They can be found in Chelliah (in press), which is a Meithei Text Collection with interlinear analysis, a grammatical sketch and cultural notes. The Pear Story film can be viewed at http://www.pearstories.org/.

The Pear Story film begins with a man high up in a tree plucking pears. He climbs down a ladder and empties the pears from his apron into one of three baskets he is filling. He removes a handkerchief from around his neck and wipes off a pear. He climbs up again and continues plucking pears. A different man leading a goat walks by. Then a boy on a bicycle arrives at the base of the tree. He looks at the pear-plucking man, picks up a basket, puts it
on his bicycle, and leaves. The pear-plucking man continues plucking. As the boy is riding down the road, a girl on a bicycle comes from the opposite direction. The boy looks at the girl, his hat comes off, and his bike hits a rock in the road. The bike falls over and the pears pour out. Three other boys, one playing with a paddleball, approach the bike boy. They help pick up the fallen pears, put them in the basket, and help him up. The boys start to leave when the paddleball boy notices the bike boy’s fallen hat. The paddleball boy whistles to the bike boy and returns his hat. The bike boy gives three pears to the paddleball boy. The paddleball boy returns to his friends and gives each one of them a pear. The bike boy continues on his way, and the other boys continue in the opposite direction while eating their pears. The pear-plucking man climbs down the ladder, and is surprised to see only two baskets. He does not know what happened to the third basket. The three boys approach the pear-plucking man eating pears as they continue walking. The pear plucking man watches them as they walk by.

Pear story retellings are an especially useful type of data because they allow linguists to compare relatively natural language samples elicited from different language consultants describing the same experience. The common experience being described is the short six-minute film. It is in color and has sound, but no dialogue, so there is nothing to be translated. This helps linguists study verbal output in the target language without influence from any another language.
4.2 Interlinear Text to Tabular Format

The Meithei texts analyzed for this thesis were six Pear Stories. Interlinear glosses were created using the Summer Institute of Linguistics developed program Fieldworks Language Explorer (or FLEx). Additional information about FLEx can be found at http://fieldworks.sil.org/flex/. I, along with two other student researchers at University of North Texas, prepared interlinear texts for each of the Pear Stories.

Example 11: Excerpt from a Meithei Pear Story in Interlinear Text Format

Story 1, line 1
famər əmənə pers pəmə kəkhəttunə
famər ə-mə-nə pers pəmə kə-kət-tunə
famər ATT-one-AGN pears plant climb-UP-ING
farmer one pears plant climbing up
uhəysshə ədə həkləmmi.
u həy-čhiŋ ə-tu hək-lam-i
tree fruit-GPL ATT-DDET. pluck-EVD-NHYP
fruits that plucking

‘A farmer climbing up a pear tree plucked those fruits.’

Tyler Utt had worked extensively with Meithei texts prior to helping with this project; he reviewed interlinear texts to check the transcriptions and translations. This was an important step because some of the texts were phonetically transcribed while others were orthographically represented. The differences between orthography and transcription introduced the possibility of misreading the texts.
4.3 Data Coding

Josh Yamane and Shobhana Chelliah divided the Pear Story sentences into clauses containing one predicate each. We entered the data from each clause into the spreadsheet columns for data analysis. Then, we checked each other’s spreadsheets for uniformity. Josh rechecked the spreadsheets again and then he reviewed the results with the team. The data were coded for the following parameters shown in figure two which is a sample of our coding spreadsheet:

Figure 2: Snapshot of Coding Spreadsheet

- The Story# column identifies which Pear Story the clause being analyzed comes from. The stories were named by the sequential order in which they were worked on. Thus, every clause from the first Pear Story has ‘Story 1’ entered in the Story# column, any clause from the second Pear Story will have ‘Story 2’ entered into this column, and each clause from the third Pear Story will have ‘Story 3’ entered in that column, and so on.
• The Clause Number column gives the sentence number, followed by a decimal point and the clause number. Consequently, the sixth clause of the seventh sentence in a Pear Story would be identified as 7.6 in the Clause Number column.

• The NP column provides the English name of the referent e.g. the farmer, the pears, a basket, bike-boy, the girl, or paddleball-boy.

• The morphology column gives the surface form of the NP in Meithei e.g phamer farmer, naspati pears, polañ basket, ṣañ child, nupimäča girl, or T.T. nupamäča paddleball-boy.

• The Form column shows whether the NP is a lexical item, pronoun, or zero anaphor.

• The Semantic role column states the semantic role of the NP e.g. agent, characterized i.e. the subject of a copular verb, experiencer, patient, recipient, stimulus, or theme.

• The Mention column codes for information status i.e. first mention, previous subject, active for referents mentioned within the clause, and old for anything that was not mentioned in the previous clause, but has been mentioned somewhere earlier in the text.

• The Animacy column indicates if the NP being coded is animate for example, (farmer, goat, goat–man, bike–boy, girl, and three boys) or inanimate (pears, ladder, baskets, handkerchief, or paddle–ball).

The data were also coded for the morphology which accompanied the NPs being coded; for example, we also noted if and what type of pronominal
modifiers were being used, e.g. an adjective, possessive, relative clause, or nothing at all. We also coded for instantiation of other morphological categories: plural, numerals, determiners, semantic role markers, focus markers, and quantifiers. This coding is not relevant to the current thesis and therefore I do not discuss it further here.

4.4 Variance among Speakers

One point of interest found within the data is the variation in the levels of zero anaphor production among speakers. A considerable amount of variation was found among the speakers in our sample. Speakers varied not only in their levels of zero anaphor production, but also in the length of their stories resulting in widely different total numbers of NPs. The stories ranged from 57 NPs in story #17 to 183 NPs in story #6. However, it is the variation among speaker in their usage of zero anaphor that interests us most. Zero anaphora among Meithei speakers ranged from 59% in story #14 to 32% in story #9. If we compare these values with the representation of arguments with zero anaphors from Bickel (2003) and Stoll and Bickel (2009), we find that Meithei speakers from our sample extend all the way across the entire range of zero anaphor usage for all of the languages investigated: Belhare 59%, Nepali 53%, Maithili 38%, and Russian 33%. However, there were some trends that could be observed. A feature common for all of the Meithei speakers was that pronouns were used less often than nouns or zero anaphors. Additionally, all of the stories represented NPs overtly the majority of the time except in story #14. It
should be noted that although story #14 was not the shortest, it was unusually short and had the second smallest number of NPs of all the stories, so it was anomalous in more than one way. Similar research, e.g. Seifart (2012), suggests that although different languages actually do have distinctive overall patterns of representation, they sometimes have a few outlying examples in their data. However, the range of variation among speakers suggests that the different levels of overt representation may be to some extent be matters of individual choice. The total number of NPs for each speaker, divided into nouns, pronouns, and zero anaphors are given below.

Figure 3: Forms of Representation for NPs by Speaker

<table>
<thead>
<tr>
<th>Story</th>
<th>Noun</th>
<th>Pronoun</th>
<th>Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story 1</td>
<td>54</td>
<td>21</td>
<td>41</td>
</tr>
<tr>
<td>Story 2</td>
<td>43</td>
<td>22</td>
<td>50</td>
</tr>
<tr>
<td>Story 6</td>
<td>67</td>
<td>47</td>
<td>69</td>
</tr>
<tr>
<td>Story 9</td>
<td>39</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>Story 14</td>
<td>26</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>Story 17</td>
<td>28</td>
<td>10</td>
<td>19</td>
</tr>
</tbody>
</table>

4.5 Data Analysis

In order to effectively analyze the reasons for and extent of zero anaphor
usage in Meithei, it was necessary to count the number of nouns, pronouns, and ellipses and compare these numbers to the total possible number of nouns (phrases) and study the environment were each occurred. The total possible number is based on the expected argument structure of a verb: e.g., a one-place predicate like walk should take one argument while a two-place predicate like hit would take two arguments.

By comparing languages across the same discourse context, Bickel (2003), and Stoll and Bickel (2009), and the study in this thesis, somewhat level the playing field with respect to cognitive accessibility and NP representation, since, in the pear stories, the cognitive accessibility of the noun phrases will be similar in the linear order of the story regardless of the language being used to tell the story.

To see if cognitive accessibility had implications internal to the text on which type of NP was used, I examined each category to see if any patterns of representation emerged that were related to cognitive accessibility. I could not use Bickel’s notion of case sensitive privileged syntactic arguments as an explanation because there is no agreement in Meithei and thus no case sensitive privileged arguments. To examine cognitive accessibility, I followed Arnold’s (1998) perspective that semantic role and information status often can predict the form of reference.

There was one major methodological issue that complicated data analysis. In Bickel (2003) and Stoll (2009), zero anaphors are counted as reduced forms of reference while nouns and pronouns are counted as full noun
phrase instantiations. Kibrik (2011), however, observes that all pronouns should be considered reduced forms of reference because they do not carry enough information to uniquely identify their referents except through coreference with a full noun. Thus in his typology, zero anaphors, bound pronouns and free pronouns are all reduced forms of reference. I had to decide whether I was going to follow Bickel and Stoll in their counting process or Kibrik’s typology of reduced and full NPs. In Bickel and Stoll (2009), whether they count pronouns with the zero anaphors or with the full NPs it is not as significant. However, the arguments of predicates in this study of zero anaphor production in Meithei have been counted by breaking them out into three categories: nouns, pronouns, and zero anaphors. This will allow future researchers to combine pronouns and zero anaphors from this research into one category and count nouns as the other if they wish to follow Kibrik’s typology. Additionally, it allows us to combine nouns and pronouns into a single category and count zero anaphors as the other in order to compare our data with the previous studies by Bickel and Stoll.

4.6 Semantic Roles for Full and Reduced Forms

If we examine full forms of reference by semantic role, we find no semantic role dominates full forms of reference. However, among the reduced forms of reference, agents were much more common than any other semantic role; in fact, reduced form agents outnumbered all the other semantic roles
combined: of reduced forms, agents were 64% of the total, patients 16%, locations 3%, and all other semantic roles 17%.

Figure 4: Reduced Forms of Reference by Semantic Role

Among pronouns, agents formed 65% of the total, followed by patients 11%, locations 0%, and all other semantic roles 24%.

Figure 5: Pronouns by Semantic Role
Among zero anaphors, agents formed 64% of the total, followed by patients 18%, locations 5%, and all other semantic roles 13%.

Figure 6: Zero Anaphora by Semantic Role

If we examine pronouns and zero anaphors, we find that they modeled similarly to one another as well as reduced forms overall. This suggests that the factors that predict the occurrence of zero anaphors are likely to be the same or at least similar to the factors that predict the occurrence of pronouns.

4.7 Full and Reduced Forms within Semantic Roles

When we compare full and reduced forms for agents and non-agents, we find results that are consistent with Arnold’s prediction that agents will be more likely to be represented by reduced forms of reference, and non-agents will be more likely to be represented by full forms of reference.
Agents were represented by zero anaphors 53% of the time, pronouns 25%, and full forms of reference 22% of the time.

Patients were more likely to be represented by full forms of reference. They
were represented by full forms 63% of the time. Patients were represented by pronouns 8%, and zero references 29% of the time.

Figure 9: Patients by Form of Reference

Locations were always expressed by either nouns or zero anaphors never by pronouns. Locations were represented by full forms 82% and zero anaphor forms 18% of the time.

Figure 10: Locations by Form of Reference
These patterns of representation influence the level of overtly supplied referents for Meithei in that the overall number of nouns and pronouns for the language is lowered because agents are the most common semantic role, and more than half of all agents are represented by zero reference forms.

Since agents are normally the most topical semantic role, we should expect agents to be highly accessible, so we should expect them to be the semantic role most frequently represented by zero anaphora in particular and reduced forms of reference in general. This offers a reasonable explanation of where Meithei allows for zero anaphor production.

Accessibility of referents can be shown to have an influence where zero anaphora occurs. The more accessible a referent is, the less explicit the form of reference needs to be. Sentential topicality is defined as the quality of being what the sentence is about, and topicality directly relates to accessibility. To some extent it is possible to view every segment of a sentence as topical, but some segments are clearly more topical than others. Since topicality causes arguments to be more accessible, and agents are highly topical, we should expect agents to be the most accessible semantic role. They are the most likely semantic role to be represented by a reduced form of reference, and highly accessible arguments are often expressed by zero anaphors.

In addition to semantic role information, the occurrence of zero anaphor in Meithei can be predicted by information status. New arguments are inaccessible because they have never been mentioned before, so they cannot be represented by zero anaphors. Old arguments are accessible because they
have been mentioned before, but the distance between mentions reduces their accessibility. “Active” arguments are NPs that are either the object of the previous clause, part of the subject or object of the previous clause. They are accessible because they were mentioned in the previous clause (very recently), so they can be represented by reduced forms of reference e.g. pronouns. Subjects of previous clauses are the most accessible arguments. They are highly accessible because there is no distance between mentions and there are no other competing referents in between interfering with their accessibility. The high accessibility of previous subjects makes them the arguments that are most naturally represented by zero anaphors.

If we examine Meithei agents by information status we find a disproportionate number of agents are previous subjects. Out of 298 agents, 219 are previous subjects, 25 are new (first mentions), 14 are active, and 40 are old information. If we examine non-agents by information status we find the opposite relationship. Out of 322 non-agents, only 39 are previous subjects, 49 are new, 100 are active, and 134 are old. Additionally, if we approach the problem from the opposite perspective and examine information status by semantic role, we still find a strong correlation between agents and previous subjects. Out of 620 NPs 258 are previous subjects. As we saw earlier, 219 previous subjects are agents and 39 previous subjects make up all other semantic roles combined. So, not only are previous subjects the largest subcategory of agents (73%), but agents are the largest subcategory of previous subjects (85%). Importantly for our study of zero anaphora, previous subjects
are most likely to be zero anaphors, active (NPs that occur in the immediately preceding clause) the next likely and old and new the least likely to be zero anaphors.

In order to communicate more efficiently there is a tendency to restrict the number of new arguments to one per clause. This often results in new arguments being introduced as the subjects of intransitive verbs as non-agents. When a referent is a new argument it is typically introduced as a non-agent using a lexical form of reference for its first mention. In the following clause the same referent would now be the previous subject and have high accessibility. Because it is highly accessible, it can be represented by a zero anaphor. As a result, agents are much more likely to be previous subjects than first mentions, so we should expect agents to be frequently represented by zero anaphors, and this pattern is actually what we find in Meithei.
CHAPTER 5

CONCLUSION

Since Meithei does not have agreement or required case marking, other treatments of overt representations versus zero anaphors cannot be used to understand zero anaphors in Meithei. Only the discourse based explanation (accessibility based on semantic role and information status) is predictive of whether the argument will be overtly or non-overtly expressed.

In the Meithei data examined here, agents are the arguments most likely to be represented by zero anaphors, and in fact, zero anaphors occurred 158 times out of 298 agents. Zero anaphors account for 53% i.e. over half of all agents. In terms of information structure, previous subjects are most likely to be zero anaphors. This study has shown that for languages with neither agreement nor pronominal marking in the verb for subjects and objects, one must turn to cognitive accessibility as determined by semantic role and information status to predict zero anaphor production in languages like Meithei. Accessibility determines whether or not Meithei speakers represent arguments with zero anaphors. Speakers represent arguments with lexical forms of representation when referents are inaccessible to the listeners. When arguments are accessible, speakers choose pronouns to represent referents. However, speakers represent referents with zero anaphors only when they are certain the arguments are accessible to listeners. I predict that with a larger database of analyzed texts the patterns we see here would be stronger and the
correspondence between cognitive accessibility and zero anaphors would be indisputable.
APPENDIX

PEAR STORY #1
The following Meithei pear story was recorded at UNT in the summer of 2007. The first line of this interlinear gloss is a phonetic transcription giving the pronunciation of the text. The second line is a morphemic transcription showing the breaks for all of the morphemes. The third line is a gloss supplying the meanings of the morphemes. The fourth line is an English word gloss. Finally the bold face line is a free translation. The words inside square brackets are elements that are not actually present in the original text, but are required arguments of the predicates.

1 famə fə́m ə́ mə العربية pers pers ə́ pəmbi pámbi kəkhəttūnə farmer farmer ə́ mə́=mə one one =AGN pears pears plant plant kəkhə kəkhə tūnə tūnə climb clap -UP -ING

2 uhəyshin ədu ádu hékləmmi hékləmmi u həy həy -čhiŋ -čhiŋ á- á- =du hék hék -ləm -ləm -li -li wood wood fruit fruit -GPL -GPL ATT- ATT- =ddet pluck pluck -EVD -EVD -PROG -PROG fruits fruits that thatplucking plucking

‘A farmer was climbing up a pear tree, [he] plucked the fruit.’

3 uhəyshin ədu ádu hékləba hékləba máttūnda máttūnda u həy həy -čhiŋ -čhiŋ á- á- =du hék hék -ləbə -ləbə mə- mə- túŋ =tə túŋ =tə later later =LOC =LOC

fruits fruits that that plucked plucked later later
4 məháknə yallibə khow ádu
mə– hák =nə yaŋ –li –pə khow á– =du
3P– here =AGN hang –PROG –NOM bag ATT– =ddet
he hanging bag that

thəllábəgi mátuŋda
thən –lába –ki má– túŋ =tə
fill –HAVING –POS NM– later =LOC
having filled later

5 kəyrak ádudagi kúmithəaktúnə
kəyrak á– =du =dagi kúm –thə –lak –túnə
ladder ATT– =ddet =ABL climb –DOWN –DISTAL –ING
ladder then climbing down

6 láybák mákhadə ah ləyrábə
ləy pák má– kha =tə ah ləy –lábə
ground broad NM– under =LOC (hest) exist –HAVING
ground under hest existing

basketšiŋ
basket –čhiŋ
basket –GPL
baskets
‘After he had filled the hanging bag, [he] was climbing down from that ladder, [he] put [pears] in the baskets that were on the ground.’
‘After [he] finished putting in the [basket] the [pears] from the bag, after [he] finished picking up the pears that had fallen, after [he] was rubbing the [pears] with the handkerchief that was around his neck, [he] once again began putting the [pears] in the [baskets].’
Then he was climbing up the ladder once again, and [he] plucked pears.

While he was plucking [the fruit], a farmer came by dragging a goat.
‘...and he, only glancing at [the man] plucking [the fruit], [he] left the scene.’
‘Afterward, the goat bringing farmer left; a child riding a bicycle was coming.’

21 máháknə ah pers póláŋšiŋdu
má- hák =nə ah pers pół láŋ -čhiŋ =tu
3P- here =AGN (hest) pears thing arrange -GPL =DDET
he hest pears those baskets

ubədə
u -pə =tə

wood -NOM =LOC
upon seeing

22.1 saykəl ádu khámləmmi .
saykəl á- =du khám -ləm -li
bicycle ATT- =ddet stop -EVD -PROG
bicycle that stop

22.2 áməsuŋ ah
á- mə čhuŋ ah
ATT- one also (hest)
one hest
'He, seeing the basket of pears, [he] stopped his bicycle and [he] stood near that basket.'

...'and after [he] was looking at the man plucking [the fruit] he tried to take a pear.'
26 hékliba mía ádugi
hék -li -pə mía á- =du =gi
pluck -PROG -NOM man ATT- =dDET =POSS
pluck man that
máʔóŋdu máhákna ηayhák yenlágə
má- óŋ =tu má- hák =nə ηay hák yeŋ -láɡə
NM- shape =DDET 3P- here =AGN wait here look -AFTER
the manner he awhile after looking

27 máhák ləybəgi máʔóŋdu ádu
má- hák lə -pə =ki má- óŋ =tu á- =du
3P- here exist -NOM =POSS NM- shape =DDET ATT- =dDET
his of being the manner that
kərimətə kháŋbə mandəbədə
kət á- mə =tə kháŋ -pə man =tə -pə =tə
what ATT- one =-NEG know -NOM alike =-NEG -NOM =LOC
nothing knowing did not seem

28 máhákna pólāŋ ádumə
má- hák =nə pót láŋ á- =tu á- mə
3P- here =AGN thing arrange ATT- =DDET ATT- one
he basket that one
tháŋdúŋə purəkləmmi
tháŋ -túŋə pu -lək -ləm -li
transfer -ING bring -DISTAL -EVD -PROG
transfer bringing
‘After he observed for a while the stance of the man plucking [the fruit], [he] seeing that from his stance looked like [he] was not aware of anything, so he took the away basket.’

29 āməšun ámádu məhákki saykəlgi
á=du mə- hák =ki saykəl =ki
ATT- one also NM- =d det 3P- here =POSS bicycle =POSS
also one that his on the bicycle
məmanda həpləgə thəwrəkləmmi .
mə- maŋə =tə həp -ləgə thəw -λak -ləm -í
NM- front =LOC put -AFTER drive -DISTAL -EVD -NHYP
in front of after putting drove off

‘[He] placed that the front of his bicycle and [he] drove off.’

30 məhákna āsum thəwrəkəpədə
mə- hák =nə á= chum thəw -λak -pə =tə
3P- here =AGN ATT- thus drive -DISTAL -NOM =LOC
he thus upon driving

31 məhák saykəl təŋbə āməsú
mə- hák saykəl tə =pə á= ma -čhú
3P- here bicycle ride -NOM ATT- one -ALSO
she bicycle riding also one
məhákki máyyóknəbədə
mə- hák =ki máy yók -nəbə =tə
3P- here =POSS direction cover -IN.ORDER.TO =LOC
his in order to get by
'He was driving in this way, she was riding a bicycle too, [she] came riding opposite from [him].'

‘[He] was looking at [her], and his hat fell off.’
máhák pers pámbí ah pers póláŋ ádu
má- hák pers pámbí ah pers pó́ t láŋ á- =du
3P- here pears plant (hest) pears thing arrange ATT- =ddet
his pears plant hest pears basket that

saykəl máthāktəgi tadúnə
saykəl má- thák =təgi ta -túnə
bicycle NM- up =ABL fall -ING
bicycle on top falling

35 ah pers póláŋ ádu ta ...
ah pers pó́ t láŋ á- =du ta
(hest) pears thing arrange ATT- =ddet fall
hest pears basket that fall

36 saykəl máthāktəgi tabadəgi póláŋ
saykəl má- thák =təgi ta -pə =təgi pó́ t láŋ
bicycle NM- up =ABL fall -NOM =ABL thing arrange
bicycle on top from falling basket

ádunə saykəl ....
á- =du -nə saykəl
ATT- =ddet -ADV bicycle
then bicycle
'In a moment [he] was not anticipating, the basket fell from above and the front wheel of the bicycle ran over the basket and then the bicycle fell down.'

'and then the basket with the pears emptied out.'
‘[They], seeing that [he] had fallen, the children that were hanging around nearby came running towards [him].’
máphi- márólšiŋ ádu
má- phi- má- lón -čhíŋ á- =du
NM- cloth NM- weave -GPL ATT- =ddet
his clothes that

kənthokpibagə
kən -thok -pi -pə =kə
straighten -OUT -REC -NOM =ASS
when helping straighten out

44 lóynəna pólǎŋdu persšiŋ
lóy -nə -nə pó́t láŋ =tu pers -čhíŋ
companion -RECIP -ADV thing arrange =DDET pears -GPL
in addition to that thing pears

ádu hápsinbirágə
á- =du háp -čhin -pi -lágə
ATT- =ddet put -IN -REC -AFTER
that put

hápsinbirəmmí.
háp -čhin -pi -ləm -í
put -IN -REC -EVD -NHYP
put in

‘And then [he] was helping [him] to straighten out his clothes, [he] also was helping [him] put the pears in the basket, and [he] helped [him] put [them] in.’
Then he was putting that basket on the bicycle once again, [he] pushed it.

Then he was putting that basket on the bicycle once again, [he] pushed it.

[...]
‘[He] was going forward a little bit, the boys picked up his fallen hat.’

‘Also one them came back once again, and [he] returned giving the hat to [him].’
51 tupi ádu pírábə mátuŋda
    tupi ə- =du pí -lābə má- tūŋ =tə
    hat ATT- =dDET give -HAVING NM- later =LOC
    hat that having given later

52 tupi khunlaklibə əŋaŋ əduŋə
    tupi khun -lək -li -pə əŋaŋ ə- =du -nə
    hat pick -DISTAL -PROG -NOM child ATT- =dDET -ADV
    hat first picking up child then

53 pers áhum páydnə lakiñmí.
    pers ə- hum páy -tūnə lak -ləm -í
    pears ATT- three carry -ING come -EVD -NHYP
    pears three carrying came

‘After [he] had given the hat, the hat picking up child came carrying three pears.’

54 ádugə mákhøy másën áməmmən
    ə- =tu -kə má- khøy má- čhên ə- mə mə
    ATT- = DDET =ASS 3P- hpl NM- self ATT- one one
    and then they self one

yēnnáduŋə čàrəmmi.
    yén -nə -tūnə čá -ləm -li
    divide -RECIP -ING eat -EVD -PROG
    dividing evenly eating

‘Then they divided the pears among themselves, and [they] were eating them.'
They came towards the [man] plucking the pears.'

‘They came towards the [man] plucking the pears.’

‘They came towards the [man] plucking the pears.’

‘They came towards the [man] plucking the pears.’

‘They came towards the [man] plucking the pears.’
‘When the boys arrived, the farmer came down from his ladder.’

‘And he saw that the boys were each walking and eating a pear.’
59 ádubu máhákna karimatég
á- du -bu má- hák =nə kəri mə =tə
ATT- ddet =ADVR 3P- here =AGN what one =NEG those he not one
háydánə
háy -tána
say -BY
saying that

60 máhákna pers ádu maŋ ... pólánə
má- hák =nə pers á- =du maŋ pótlánə
3P- here =AGN pears at the front thing arrange
he pears that front basket
layrəmdəbəduda
ləy -ləm =tə -pə =tu =tə
exist -EVD =NEG -NOM =DDET =LOC
when not being

61 hayphétəŋ ah thoydoknə
hay phét táŋ ah thoy -thok -nə
approach by amount (hest) increase -OUT -ADV slightly hest being more
nuŋaytə áməsəŋ
nuŋay =tə -pə á- mə čhuŋ
happy =NEG -NOM ATT- one also
not being happy one
‘But he not saying anything, he was slightly unhappy seeing that there was a basket missing, and [he] was sad too, but he did not say a word.’
BIBLIOGRAPHY


