Two Types of Variable Elements in Hmong Anaphora

David Mortensen
University of California, Berkeley
dmort@socrates.berkeley.edu

September 17, 2003

1 Introduction

- Hmong\(^1\) is like Thai (Lasnik 1989; Narahara 1995), Vietnamese (Lasnik 1989; Narahara 1995), San Luis Quiaviní Zapotec (SLQZ, Lee 2003), and Quegolani Zapotec (QZ, Black 2000) in that it apparently allows pronouns and R-expressions to be bound as reflexives:

(1) a. Nwg\(i\) yeej qhuas nwg\(i\).
   3SG always praise 3SG
   ‘He always praises himself.’

b. Puab\(i\) xav xaav txug puab\(i\).
   3PL only is think about 3PL
   ‘They only think about themselves.’

c. Txiv\(i\) yeej qhuas txiv\(i\).
   Father always praise father
   ‘Father always praises himself.’

d. Pov\(i\) yeej qhuas Pov\(i\).
   Pao always praise Pao
   ‘Pao always praises himself.’

e. Puab, xaav has ta Maiv\(j\) tua Maiv\(j\).
   3PL think say that May kill May
   ‘They think that May killed herself.’

1.1 Some Past Analyses

These appear to be violations of the generalizations described in classical binding theory by Conditions B and C.

Earlier investigations (couched in classical binding theory) have used similar data in Thai (Lasnik 1989) and Quegolani Zapotec (Black 2000) to argue that Condition C (and perhaps Condition B) were parametric.

More recently, Lee (2003) has argued that such bound “pronouns” and “R-expressions” are actually variable copies.

1.2 Current Proposal

The current research modifies and extends Lee’s account, arguing that these variable copies may be treated as instances of a single, phonologically empty lexical item and that its properties may be described in terms of Safir’s (2002a; 2002b) Form to Interpretation Principle (FTIP), a competitive principle governing dependence relations.

This anaphoric element is only “variable” in the following senses:
- Its referent is determined by its antecedent (it cannot independently refer).
- Its phonological form is determined by its antecedent.

Such elements are to be distinguished in Hmong, from a more prototypical class of variables. These are bound from A’-positions (typically topic phrases) in order to allow distributive interpretations.
2. Bound Pronouns and R-Expressions: The Variable Anaphor

2.1 Parametric Condition C accounts are inadequate

- Earlier approaches to bound pronouns and names have held that Condition C is parametric (Lasnik 1989; Black 2000).
- Condition C, in these accounts, is not set in languages that allow bound names.

2.1.1 Pronouns cannot bind names

- However, in languages that allow bound names, only names can bind names (pronouns cannot bind names).
- This can be seen in Hmong:

  (3) a. Nwg \text{ yeej} qhuas Pov\text{\textsubscript{vij}.} \\
      3SG always praise Pao \\
      ‘He always praises Pao.’ \\
      *‘Pao always praises himself.’

  b. Nwg \text{ yeej} qhuas Txiv\text{\textsubscript{vij}.} \\
      3SG always praise Father \\
      ‘He always praises Father.’ \\
      *‘Father always praises himself.’

2.1.2 Hierarchies of referentiality

- To account for such phenomena, Lasnik (1989) proposed that anaphoric elements can vary in their referentiality, and that referential elements (names) cannot be bound as referential elements (e.g. pronouns).

2.1.3 Names cannot bind pronouns locally

- This is false:

  (4) a. Pov\text{\textsubscript{\textit{\textit{i}}/j}} qhuas nwg\text{\textsubscript{\textit{\textit{i}}/j}.} \\
      Pao 3SG always praise \\
      ‘Pao always praises him.’ \\
      *‘Pao always praises himself.’

  b. Txiv\text{\textsubscript{i}} tua nwg\text{\textsubscript{\textit{\textit{i}}/j}.} \\
      Father kill 3SG \\
      ‘Father killed him.’ \\
      *‘Father killed himself.’

2.1.4 Names cannot bind non-identical names

- Furthermore, such an account would seem to predict that R-expressions can freely bind R-expressions (since they do not differ in “referentiality.” This is also false:

  (5) a. Tug xibfw\textsubscript{b} qhuas tug xibfw\textsubscript{\textit{\textit{b}}/j}. \\
      CLF teacher praise CLF teacher. \\
      ‘The teacher praised the (other) teacher.’ \\
      *‘The teacher praised himself.’

  b. Tug xibfw\textsubscript{b} qhuas Pov\textsubscript{vij} \\
      CLF teacher praise Pao \\
      ‘The teacher praised Pao.’ \\
      *‘Pao, the teacher, praised himself.’

  c. Pov\text{\textsubscript{\textit{\textit{i}}/j}} qhuas tug xibfw\textsubscript{\textit{\textit{b}}/j}. \\
      Pao CLF teacher praise \\
      ‘Pao, the teacher, praised the teacher.’ \\
      *‘Pao, always praised himself, the teacher.’

Actually, it will later be seen that classified nouns can neither bind nor be bound in this construction. However, this does not seem to be a prediction of the parameterized binding theory being discussed here, so these examples are still useful.

- This notion of referentiality was extended into a kind of hierarchy by Narahara (1995).
- However, such a proposal does not make the right predictions for Hmong (or Thai, SLQZ, and Quegolani Zapotec).
- This hypothesis seems to predict that if pronouns can bind pronouns in Domain D, names should also be able to bind pronouns in Domain D.
• There are even more damning data on this front.
• In Hmong society, an adult male has two names, one given at birth (*npe hluas ‘young-name’) and one given by his father-in-law after he has a few children (*npe laug ‘old-name’).
• This two names can refer to exactly the same person at exactly the same time.
• However, they cannot be bound by one another:

   ‘Kub’s young-name is Tshuv-Ntxaij.’

b. Kub *yeej thuam Tshuv-Ntxaij eyj.
   ‘Kub always criticizes Tshuv-Ntxaij.’

   *‘Kub always criticizes himself.’

2.1.5 Pronouns cannot bind non-identical pronouns
• Furthermore, the exhaustive pronoun *suavdlawg ‘everybody’ may bind *suavdlawg, but not the plural pronoun *puab ‘3PL’.

(7) a. *Suavdlawg yeej *quas suavdlawg.
   ‘Everyone always praise everyone.’

   *‘Everyone (as a group) always praises themselves (as a group).’

b. *Suavdlawg yeej *quas *puab eyj
   ‘Everyone always praise 3PL.

   *‘Everyone (as a group) praises them (a different group).’

   *‘Everyone praises themselves.’

• These are instances of what Lee (2003) has called the identical antecedent requirement.
• In languages that allow bound pronouns and R-expressions of this type (as is also shown by data from Thai, Vietnamese, and SLQZ and QZ), the bound name or R-expression must be identical to its antecedent.
• This important generalization is missed by the parameterized account framed in classical binding theory.
• It also seems redundant to maintain Conditions A, B, C and a hierarchy of referentiality, especially since, as Narahara (1995) argues, Condition C effects can basically be reduced to epiphenomenal effects of such a hierarchy.

2.2 Bound names/pronouns and Safir’s FTIP

2.2.1 Introducing the FTIP

• There have been many subsequent attempts to relate generalizations about “referentiality” to the binding properties of pronouns and other anaphoric elements.
• One recent attempt is the Form to Interpretation Principle (Safir 2002a,b):

(8) Form to Interpretation Principle (FTIP)
If *x c-commands *y and *z is not the most dependent for available in position *y with respect to *x, then *y cannot be directly dependent on *x (Safir 2002b:16)

• For Safir, dependence is a lexical property apparently separate from the semantics of the item under examination.
• The FTIP acts as a replacement for Conditions B and C of binding theory.
• Condition A is replaced by a requirement called Local Antecedent Licensing (LAL), which requires that anaphors (lexical items specified as anaphors) have a licencing antecedent in Domain D.

2.2.2 Dependent reference versus independent coreference

• Safir also draws a distinction between independent coreference and dependent reference (and it is dependent reference that is regulated by the FTIP).
• Thus, Safir is able to justify the grammaticality of sentences like (9):

(9) a. Even Alfred says that Alfred is crazy.
   b. Even Alfred says that Alfred is crazy, and more surprising, even Irv does.

• The FTIP only regulates dependent reference.
• In (9b), only a strict (no sloppy) reading is available. That is to say, the elided clause cannot be interpreted as meaning that Irv thinks that Irv is crazy, but only that he thinks that Alfred is crazy.
• Safir uses this type of elision as a diagnostic for identifying dependent reference.
• “Insofar as sloppy readings are dependent ones, the distribution of sloppy readings for reflexives in ellipsis contexts can be used as evidence that reflexivity is a dependency relation” (Safir 2002a).
2.2.3 Hmong bound pronouns and numerals do not independently refer

- We cannot treat Hmong pronoun and name reflexives as instances of independent coreference, not just because they have reflexive interpretations, but because they allow sloppy readings:

(10)

a. Koj yeej qhuan koj; nwg los kuj ua le hab.  
   2sg always praise 2sg; 3SG TOP also do as too  
   ‘You always praise yourself, and so does he.’

b. Pov yeej qhuan Pov; Maij los kuj ua le hab.  
   Pao always praise Pao May TOP also do as too  
   ‘Pao always praises himself, and so does May.’

- The same is also true of SLQZ and Thai, where (according to Lee) only sloppy readings are licensed (Lee 2003):

(11)

a. B-gwi’ih Gye’eihlly lohoh Gye’eihlly zë’cy cahgza’ Li’eb.  
   PERF-look Mike at Mike likewise Felipe  
   ‘Mike looked at himself, and Felipe did too.’  
   *[Mike looked at himself, and Felipe looked at Mike’]  
   [SLQZ]

b. John koonnuat khong John lae Peter ko muankan.  
   John shave of John and Peter the same  
   ‘John shaved himself, and Peter did too.’  
   *[John shaved himself and Peter shaved John.]  
   [Thai]

- In contrast, when a name is not bound, only a strict reading is allowed, illustrating a clear contrast in the properties of bound and unbound names:

(12)

a. Pov nyam Maij; Maij los kuj ua le.  
   Pao like May May TOP also do as  
   ‘Pao likes May and May does too (likes May).’

b. Pov hab Tub yeej ib-txwm ua yeebncuab. Pov ntsub Tub; Tub los kuj  
   Pao and Tou always always do enemy Pao hate Tou Tou TOP also  
   ua le hab, do as too  
   ‘Pao and Tou have always been enemies. Pao hates Tou and so does Tou (hate Tou).’

- This being the case, Safir’s FTIP seems unable to explain the distribution of dependent readings in Hmong.

- Only the most dependent form available in a particular syntactic context should be able to receive a dependent reading.

- This would require that the name Pov is the most dependent form available when Pov is a c-commanding antecedent, but that nwg is the most dependent form available when nwg is a c-commanding antecedent.

2.3 Proposed solution: variable anaphors as instances of a lexical item

- There is a way to accommodate the Hmong data while still retaining the considerable advantages of a competitive algorithm like Safir’s.

- We may do this by borrowing a card from Felicia Lee, and proposing that what appear to be bound names and pronouns in languages like Hmong, Thai, Vietnamese, QZ, and SLQZ are not names and pronouns at all, but variable elements of some type.

2.3.1 Properties of \( \alpha \)

- I propose that these elements are not syntactic variables of some kind, but base-generated instances of a universally available anaphoric lexical item (on par with PRO and pro, as treated by Safir 2002a).

- This item, which I will call \( \alpha \), is not specified for phonological form. Rather than surfacing as null (like PRO and, pro) its form is spelled out as that of its antecedent (thus accounting for the identical antecedent requirement identified by Lee).

- We can then see the source for two different readings in sentences like (13):

(13)

a. Pov yeej qhuan Pov.  
   Pao always praise \( \alpha \)  
   ‘Pao always praises himself.’

b. Pov yeej qhuan Pov.  
   Pao always praise Pao  
   ‘Pao always praises Pao.’ (there are two Paos)

- In (13a), the second Pov is actually \( \alpha \). Since it is the most dependent form available here, it receives a dependent reading.

- In (13b), the second Pov is base-generated Pov. Since it is not the most dependent form available here, it does not receive a dependent reading.
2.4 Some properties of $\alpha$

2.4.1 Subject to Local Antecedent Licensing

- Lee (2003) suggests that “bound variable copies” in SLQZ can act as long-distance anaphors.
- However, in Hmong, equivalents to her example sentences, while grammatical, seem best explained as examples of independent coreference.

(14) a. Pov has tas Maiv nyam Pov.
   Pao say that May like Pao
   ‘Pao said that May likes him.’

   b. Pov has tas Maiv nyam Pov; Tub los kuj ua le hab.
      Pao say that May like Pao Tou TOP also do as too
      ‘Pao said that May likes him, and so did Tou (say that May likes Pao).’
      ‘*Pao said that May likes him, and so did Tou (say that May likes Tou).’

- The unavailability of the sloppy reading in (14b) suggests that the relationship between the two instances of Pao in (14a) is one of independent coreference rather than dependence. That is, the second Pao is actually Pao rather than $\alpha$.
- This suggests that $\alpha$ requires a local, c-commanding antecedent. In Safir’s terminology, it is subject to Local Antecedent Licencing.

2.4.2 No lexical noun antecedents

- $\alpha$ cannot depend directly on lexical nouns unless they are interpreted as names.

(15) dlev tum dlev.
     dog bite dog.
     ‘Dogs, bite dogs.’
     ‘*Dogs, bite themselves.’

- Unclassified (bare) nouns in Hmong are interpreted as properties (as suggested by the first reading) rather than entities Bisang (1993).
- As properties, nouns may not act as direct antecedents to $\alpha$.
- If nouns are interpreted as names, as in the second interpretation, then the dependent reading is possible, but somewhat strained.
- If markers are added that specifically identify the noun as an epithet or name, the reflexive reading is improved:

2.4.3 No classified or quantified antecedents

- The presence of a noun classifier causes a noun to be interpreted as an entity rather than a property.
- In almost all cases, nouns must take classifiers in order to be quantified.
- As in SLQZ, quantified nouns in Hmong may not participate in the bound R-expression reflexive construction. However, classified nouns may not participate either:

(17) a. tug dlev tum dlev.
     CLF dog bite CLF dog
     ‘The dog bit the dog.’
     ‘*The dog bit himself.’

   b. tug tuabneeg thum tug tuabneeg.
      CLF person criticize CLF person
      ‘The person criticized the person.’
      ‘*The person criticized himself.’

- For Hmong, it seems that these prohibitions against certain R-expressions acting as antecedents for $\alpha$ can be reduced to a single, simple, descriptive generalization:

(18) Generalization of the Distribution of Variable Anaphors

The variable anaphor $\alpha$ may be licenced by pronouns and names, but not by R-expressions that are not names.

2.4.4 Distributed Reflexives in Hmong

- Of course, it is possible to have reflexive meanings distributed over properties or quantified groups of entities.
- Lee (2003) mentions two strategies for doing this, one which she associates with Thai, and another which she associates with SLQZ.
In the Thai-like strategy, another reflexive anaphor is used in place of \( \alpha \) (in Hmong, an anaphor constructed of a pronoun + \textit{tug kheej} “CLF self”).

\begin{enumerate}
  \item In the Thai-like strategy, another reflexive anaphor is used in place of \( \alpha \) (in Hmong, an anaphor constructed of a pronoun + \textit{tug kheej} “CLF self”).
  \item In the second (SLQZ-like) strategy, the quantifier phrase or bare noun is topicalized, and binds the distributive pronoun \( \textit{nyas} \), which in turn binds \( \gamma \). That is to say, \( \alpha \) depends upon \( \textit{nyas} \), which depends in turn upon the topic.
\end{enumerate}

\subsection*{2.5 Summary}
\begin{itemize}
  \item We can analyze bound pronouns and R-expressions, in Hmong, as instances of the lexical item we have called \( \alpha \).
  \item In Safir’s terminology, \( \alpha \) is the “most dependent form” available in the types of constructions we have discussed here.
  \item These variable anaphors cannot be bound from \( \alpha' \)-positions (e.g., topics and quantified expressions). They also cannot be bound by R-expressions that are not names.
  \item In contexts where \( \alpha' \)-binders are involved, other types of variable elements are involved. It is possible for such variables, which are dependent upon \( \alpha' \) elements, to bind \( \alpha \), thus allowing distributed reflexives.
\end{itemize}

\section*{3 Topic-bound variables}
\begin{itemize}
  \item Hmong pronouns, with the exception of \( \alpha \), can depend upon topics for reference.
  \item Take the following example (dependent forms are in boldface):
\end{itemize}

\begin{enumerate}
  \item In these sentences, \( \textit{nyas} \) is acting as a variable in a sense that \( \alpha \) cannot: it is receiving a distributive reading.
  \item The pronoun \( \textit{nyas} \) is special in that it always receives a distributive interpretation (distributes the topic over some proposition). The other pronouns may also act as variables in this sense.
  \item \( \alpha \) is unique in that it is not this kind of variable. That is, it never receives a distributive reading directly.
  \item It is to this other type of variable that we will turn next.
\end{enumerate}
3.2 Pro as a variable.

- The null pronominal can replace \textit{nwg} in either subject or object positions: \(^3\)

\begin{enumerate}
\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textbf{pro} nyam \textit{nwg\textsubscript{3SG}},
\textit{CLF which} \textit{TOP} \textit{like} \textit{3SG}
\end{enumerate}

‘Anybody would like him.’

\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textit{nwg\textsubscript{3SG}} nyam \textbf{pro},
\textit{CLF which} \textit{TOP} \textit{3SG} \textit{like}
\end{enumerate}

‘He would like anybody.’
\end{enumerate}

- In sentences of this type where \textit{nwg ‘3SG’} and \textit{pro} are present, \textit{pro} will always depend on the topic and \textit{nwg} will always have disjoint reference.

- If both arguments of the main-clause verb are instances of \textit{pro}, there are three possible readings:

\begin{enumerate}
\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textbf{pro}, nyam \textit{pro},
\textit{CLF which} \textit{TOP} \textit{like} \textit{α}
\end{enumerate}

‘Anybody would like themselves.’

\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textit{nwg\textsubscript{3SG}} nyam \textbf{pro},
\textit{CLF which} \textit{TOP} \textit{like}
\end{enumerate}

‘Anybody would like him.’

\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textit{nwg\textsubscript{3SG}} nyam \textbf{pro},
\textit{CLF which} \textit{TOP} \textit{like}
\end{enumerate}

‘He would like anybody.’
\end{enumerate}

- The first interpretation is marginal. The second interpretation is the default, but the third is perfectly acceptable as long as an appropriate context is provided.

3.3 The ability to act as a variable is a scalar lexical property

- If both arguments of the main-clause verb are identical in their properties, either of them (but not both of them) can depend directly upon the topic.

- If the arguments differ in their properties, one of them will depend upon the topic for reference and the other will be unable to do so.

\footnote{The astute reader will, at this point, ask why I have chosen to treat these empty elements as \textit{pro}, rather than as traces left by the movement of the topic out of the argument positions (perhaps explaining—depending on one’s theory of dependent reference—the coreference between this element and the topic). One reason for the \textit{pro} example should be clear: in (21) above, there are pronouns where are putative traces should be. It seems contrived to posit that the topic is base generated in topic position in (21), but moves to topic position in (23). For arguments regarding a similar analysis of Thai topics, see Hoonchamlong (1991) and Aroonmanakun (1999).}

3.4 Some Complications: Donkey Sentences

- In fact, this pattern is even more striking add a couple more types of elements to the mix: \textit{yawg ‘grandfather; man; Mister’} and \textit{Pov ‘Pao’}.

- A summary of the resulting paradigm can be seen in Table 1.

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|l|}
\hline
\textbf{SUBJECT} & \textbf{OBJEKT} & \textit{nwg ‘3SG’} & \textit{yawg ‘Mister’} & \textit{Pov ‘Pao’} \\
\hline
\textit{pro} & S/O/R? & S & S & S \\
\textit{nwg ‘3SG’} & O & S/O/R & S & S \\
\textit{yawg ‘Mister’} & O & O & S?/O/R? & S \\
\textit{Pov ‘Pao’} & O & O & O & — \\
\hline
\end{tabular}
\caption{Distribution of dependent readings of \textit{Tug twg los SUBJECT nyam OBJECT}}
\end{table}

- The ability to act as a variable, in this construction, seems to be primarily a lexical property (not a result of structural configuration).

- The elements seem to fit into a competitive hierarchy, which could be compared in some ways to Safir’s hierarchies of dependence.

- In general, the most “dependent” elements are the ones with the least idiosyncratic information and the least featural content.

- This raises an important question: why does \textit{α} never occur as an \textit{A’}-bound variable? If the relevant lexical characteristic targeted by this phenomenon is the same as Safir’s dependence, and if \textit{α} is the most dependent form, then why, in examples like (21a), is it \textit{nwg} rather than \textit{α} that depends directly upon the topic?

- Possible answer: \textit{α} has a licensing requirement that precludes it from depending upon topics (and perhaps all other things in \textit{A’} positions).

- Of course, there are additional complications. One of the most interesting is presented by a type of donkey sentences that occur in this construction:

\begin{enumerate}
\item \begin{enumerate}[label=(\arabic*)]
\item [Tug twg ] los \textit{nwg\textsubscript{3SG}}, nyam \textbf{pro},
\textit{CLF which} \textit{TOP} \textit{like}
\end{enumerate}

‘Whoever sees him will surely like him.’
\end{enumerate}
b. [ Tug twg ], pun pro los nwg pro yeej yam pro pro,
   CLF which see LOS 3SG always like
   ‘Whoever sees him surely like him.’

c. [ Tug twg ], pun pro los pro yeej yam pro pro,
   CLF which see TOP always like
   ‘Whoever sees him will surely like him.’

- In these sentences, the dependency relationships seem to be based upon a direct mapping of the argument positions of the topic clause and the main clause.
- These are similar in some respects to donkey sentences in Mandarin Chinese:

(26) Shei, xian lai shei, (jiu) then xian first come who then first eat
   ‘Whoever comes first eats first.’ (Cheng and Huang 1996)

- Interestingly, it is not possible to translate sentence (25a) directly into Chinese:

(27) a. * Shei kan jian pro, shei, (jiu) xihan pro.
    who look see who then like
    Intended: ‘Whoever sees (him), likes (him).’

b. Shei, kan jian ta, shei/pro, (jiu) xihan ta.
    who look see 3SG who/pro then like 3SG
    ‘Whoever sees him, likes him.’ (Patrick Chew p.c.)

- However, it is likely that a similar principle governs the interpretation of these sentences in Hmong.
- These sentences are unlike the previous pattern in that dependent reference cannot be described in terms of lexical properties. A description of the phenomenon must refer to structural position or argument structure.

4 Conclusions and Directions for Further Research

- Bound names and pronouns in Hmong may be profitably analyzed as instances of a lexical item α, an anaphor underspecified for phonological form which therefore copies its form from that of its antecedent.

(28) Question: What would this phenomenon look like if reanalyzed in terms of Hornstein’s (2001) representational view of anaphora, in which all coreference results from the existences of multiple copies of the same element?

- The “variable” α should be distinguished from variables bound from topics and other A’-bar positions, which often receive distributive dependent readings.
- α can only receive a distributive reading when this reading is mediated by a variable of the distributive (A’-bound) type.
- Dependence upon topics and other A’-elements seems to involve competition among potential variables within the numeration. The outcome of this competition may be determined by a property like or identical to Safir’s notion of dependence.
- This generalization, however, does not explain some Hmong donkey sentences, for which a different principle would have to be invoked.

References


