Two kinds of variable elements in Hmong anaphora

Article · July 2003

CITATIONS
7

READS
55

1 author:

David Mortensen
Carnegie Mellon University
21 PUBLICATIONS **80** CITATIONS

SEE PROFILE

Two Kinds of Variable Elements in Hmong Anaphora*

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

June 2, 2003

Hmong[1] is like a number of other languages (including Thai, Vietnamese, and some Zapotec languages) in that it allows names and pronouns to be bound by antecedents in order to form reflexives. The ordinary way to form reflexives in Hmong is shown in example (1):

(1) a. Nwg$_i$ yeej qhuas nwg$_i$.
3SG always praise 3SG

'He always praises himself.'

b. Puab$_i$ tsuas yog 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

*This paper was made possible by Hmong consultants and teachers too numerous to name exhaustively. However, it would be ungracious not to mention a few individuals who provided data, judgments, and guidance without which this paper could not have come to be. A special debt is owed to Neeb Hawj (Leena Her), Nchais Laaj Hawj, Tshuv Ntxaij Yaaj (Shawn Yang), Xab Yaj (Sa Yang), Ntaub Muas (Dao Moua), and Iab Hawj (Ia Her).

[1]The Hmong data in this paper are from Mong Leng (also called Green Hmong, Blue Hmong, Blue Meo, Hmong Njua, etc.), a dialect of Hmong spoken by around 1,000,000 speakers in Southern China, Vietnam, Laos, Thailand, and various western countries (due to the Hmong diaspora from Laos following the "Secret War" there). Hmong is a member of the Far Western Hmongic group of the Western Hmongic branch of the Hmongic subfamily, which is, in turn, a member of the Hmong-Mien family. The wider genetic affiliations of Hmong-Mien are undetermined.

Like other members of the Hmong-Mien family, Mong Leng is tonal, having 7 contrastive tones. In the RPA orthography used in this paper (and by most Hmong in the West), tones are represented by letters at the end of syllables (-b, -j, -v, -∅, -g, -s, and -m). There are other aspects of this orthography that the naive linguist may not find completely transparent. For example, doubled vowels represent the presence of nasalization or a velar nasal coda. For a good introduction to the sounds and orthographic representation of the closely related Hmong Daw (White Hmong) dialect, see Ratliff (1992). Mong Leng has several sounds not present in Hmong Daw, and these are represented here in the conventional Mong Leng adaptation of the RPA orthography. The only important differences are as follows: <dl> [tɬ ∼ kɬ], <dlh> [$t^hɬ$ ∼ $k^hɬ$], <ndl> [ntɬ ∼ ŋkɬ], <ndlh> [$nt^hɬ$ ∼ ŋ$k^hɬ$], and <aa> [ãŋ ∼ ã].

'Father always praises himself.'

d. Pov_i yeej qhuas Pov_i.
Pao always praise Pao

'Pao always praises himself.'

e. Puab_i xaav has tas Maiv_j tua Maiv_j.
3PL think say that May kill May

'They think that May killed herself.'

On the face of it, these appear to be violations of Binding Conditions B and C. By all appearances, Hmong pronouns and names can be bound by antecedents. There is, in fact, a respectable tradition of trying to explain the very similar phenomena in Thai and Zapotec. Early proposals, such as those of Lasnik (1989) and Black (2000) argued that Condition C (and perhaps Condition B too) must be parametric, on the grounds that this type of Condition C and B violation was present in Thai (Lasnik), Quiegolani Zapotec (Black), and other languages.

However, more recent work by Lee (2003) suggests a different solution: namely, that the bound expressions in sentences like those in (1) are not actually pronouns and names at all, but rather are bound variables that are spelled out as copies of their antecedents. This explanation reflects the significant insight that these bound elements never differ in their exponence from their antecedents, a fact totally missed by the parametric-Condition C approach, and is generally compatible with the observation that the languages displaying this pattern of binding still display various Condition C effects.

Lee's (2003) account of anaphoric R-expressions and pronouns is consistent with the patterns of binding found in Hmong. However, Lee does not make explicit the relationship between the type of bound copy occurring in reflexive constructions and other types of bound variables. I argue that there are two types of bound variables in Hmong, which differ from one another in their properties. These I will call A-bound variables and A′-bound variables.

My A-bound variables are equivalent to Lee's (2003) bound copies. I argue that these are best represented in Hmong as a single lexical item (*ana*) lacking its own phonological form (which is necessarily copied from that of its antecedent)[2]. The occurrence of this form must always be licensed by an A-binding antecedent. Further, I propose that the distribution of this element can be described in terms of Safir's (2002a; 2002b) Form to Interpretation Principle (FTIP) and that, in Safir's terminology, this variable anaphor must be the "most dependent form."

A′-bound variables, in contrast, are not a distinct lexical item or set of lexical items. Rather, normal anaphoric elements are able to act as A′-bound variables in proportion to their "dependence". However, the distribution of bound variable readings cannot be described in terms of Safir's FTIP, but requires a different competitive principle making reference to this same lexical property (dependence).

[2]This should not be taken to imply that *ana* will always be phonologically identical to its antecedent (although this is the case in Hmong), but only that its phonological form will be deterministically related to that of its antecedent.

1 A Brief Introduction to Hmong Pronominals

1.1 Personal pronouns

Hmong is unlike some Southeast Asian languages (such as Vietnamese—see Thompson 1965) in that personal pronouns (as opposed to kinship terms) are the usual terms of personal reference. Hmong is like Vietnamese, however, in that both types of reference are available. The Hmong (Mong Leng) pronoun paradigm is given in Table 1.

	Singular	*Dual*	*Plural*
First Person	kuv	wb	peb
Second Person	koj	meb	mej
Third Person	nwg	ob-tug	puab
pro			
Impersonal	yug	—	luas
Distributive	nyas	—	—
Exhaustive	—	—	suavdlawg

Table 1: The personal pronouns of Hmong (Mong Leng)

1.2 Demonstrative pronouns

In addition to personal pronouns, Hmong has a system of demonstratives which may function as pronouns (but which are most often used adjectively). These exist along both person-based (see Ratliff 1997) and temporally based scales, as shown in Figure 1. While these elements do not enter directly into the discussion in this paper, they

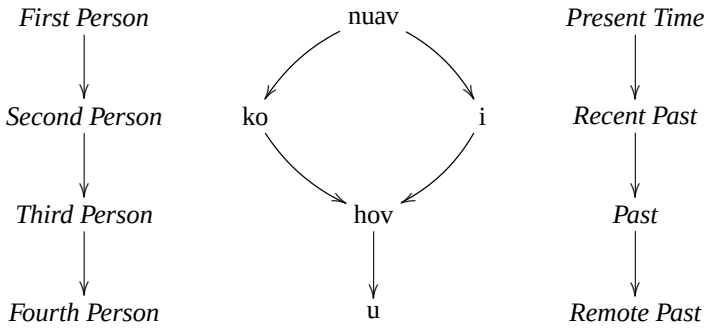

Figure 1: The demonstrative pronouns of Hmong (Mong Leng)

are frequently used along with one of the noun classifiers for humans, *tug*, as a means of personal reference when a third person pronoun would be ambiguous in the same

context[3]. As such, they do participate in anaphoric reference.

1.3 Zero Anaphora

Hmong also makes rather free use of zero anaphora, which will be treated in this paper as a function of the zero-pronoun *pro*. The zero pronoun can be licensed in most contexts where an overt pronoun can be licensed. Following Hoonchamlong's (1991) analysis of Thai, I assume that the empty category (EC) seen in topic/comment constructions that refers back to the topic is *pro* rather than *t*. The reasons for doing this are the same as those given by Hoonchamlong (1991) and Aroonmanakun (1999) for Thai. This specific detail will become important in the discussion of A′-bound variables that follows.

1.4 Indefinite pronouns

Finally, Hmong has a system of indefinite (WH-like) pronouns, which are like Chinese WH-indefinites in both in that they remain in situ in questions and can have purely indefinite (non-question) readings (see Li 1992):

***dlaabtsi* 'What? / What sort? / anything / of any sort'** This indefinite pronoun functions like a common noun and even takes a noun classifier (*tug*, 'CLF for persons, animals, long or cylindrical objects, and certain abstract concepts').

***caag* 'How? / Why? / in any manner / for any reason'** *Caag* replaces demonstrative pronouns in context where they point to the exemplar for a manner or reason.

***twg* 'Which? / any'** This pronoun replaces demonstrative pronouns when they pick out a specific entity. The usual means of saying 'Who?/anybody' and 'Where?/anywhere' are formed with this indefinite: *tug* + *twg* 'CLF (PERSON)' + which', *qhov* + *twg* 'CLF (PLACE) + which'

***(pis)tsawg* 'How many? / any number'** This WH-indefinite is actually a pro-numeral rather than a pronoun, but it is mentioned here because it shares the indefinite properties of the pronouns above. It replaces a numeral (but not other types of quantifier expressions).

1.5 Names and kinship terms

Hmong is like many Southeast Asian languages in that names and kinship terms can be used for both first person, second person, and third person address. In their properties, names and kinship terms behave quite similarly, in contrast to ordinary nouns on the one hand and pronouns on the other. For example, both names and kinship terms can refer to a concrete, real-work entity without being modified by a noun classifier, but this is not possible for common nouns.

[3]The same could be said for another set of spatial deictics in Hmong, which also may function like prepositions. See Heimbach (1969); Ratliff (1992, 1997).

2 A-Bound Variables

As seen in (1), names, kinship terms, and pronouns can all occur as bound anaphoric elements in Hmong. Felicia Lee's (2003) approach to bound pronouns and R-expression in SLQZ (San Luis Quiaviní Zapotec) can be adapted to the Hmong quite easily. What it lacks is a principled account of the relationship between this type of bound variable and the types of bound variables more generally known in the anaphora literature (e.g. *wh*-traces). Following Lee, then, I argue that what appear to be bound R-expressions and pronouns are not R-expressions and pronouns at all, but "variables." But I go beyond Lee in proposing that these "variables" are instances of a distinct anaphoric lexical item with its own DEPENDENCE[4] properties. As a result, it is "variable" in a fundamentally different way than the A′-bound variables discussed below.

But what would motivate us to adopt this relatively abstract representation for what appear to simply be bound anaphors and names? It is the fact that neither traditional binding theory, or more recent competitive versions of binding theory can straightforwardly account for the pattern seen in (1) if it is assumed that the bound elements are actually anaphors or names.

2.1 Parametric Condition C accounts are inadequate

The conventional account of bound anaphors (see Lasnik 1989; Black 2000) holds that Condition C is parametric and that it is not active in languages that allow bound names. Of course, earlier authors like Lasnik were aware that, even in such languages, pronouns could not bind names, as is also the case in Hmong:

(2) a. Nwg_{i} yeej qhuas $\text{Pov}_{*\text{i/j}}$.
3SG always praise Pao

'He always praises Pao.'
*'Pao always praises himself.'

b. Nwg_{i} yeej qhuas $\text{Txiv}_{*\text{i/j}}$.
3SG always praise Father

'He always praises Father.'
*'Father always praises himself.'

Lasnik (1989) accounts for this by proposing that there is a "referential hierarchy" such that names are more referential that pronouns, which are more referential than anaphors. He proposes a universal principle that a less referential element cannot bind a more referential element.

This proposal does not make the right predictions for Hmong. First of all, it predicts that, if pronouns can bind pronouns within Domain D, names should also be able to bind pronouns within Domain D (since names would thus be higher than pronouns on the referentiality hierarchy). But this is not so:

[4]The term *dependence* is used here in the sense of Safir (2002b), where it refers to a formal, scalar property of lexical items that gives them the ability to depend upon other syntactic elements.

(3) a. Pov$_{i}$ yeej qhuas nwg$_{*i/j}$.
Pao always praise 3SG

'Pao always praises him.'
*'Pao always praises himself.'

b. Txiv$_{i}$ tua nwg$_{*i/j}$.
Father kill 3SG

'Father killed him.'
*'Father killed himself.'

Furthermore, it seems to imply that all R-expressions should be able to bind R-expressions, but again, this is not the case. Certain R-expressions—notably bare nouns that are not names or kinship terms—are resistant to binding. And quantified nouns may not be bound at all:

(4) a. Dlev$_{i}$ yeej tum dlev$_{*i,j}$.
dog always bite dog.

'Dogs always bite dogs.
*'Dogs always bite themselves.'

b. ? [Tug dlev]$_{i}$ yeej tum [tug dlev]$_{*i/j}$.
CLF dog always bite CLF dog

'The dog always bites the dog.'
*'The dog always bites himself.'

c. [Ob tug dlev]$_{i}$ yeej tum [ob tug dlev]$_{*i/j}$.
two CLF dog always bite two CLF dog

'The two dogs bite the (other) two dogs.'
*'The two dogs bite themselves.'
*'The two dogs bite each other.'

Beyond the empirical inadequacy of this account, there is something conceptually suspicious about invoking both the binding conditions A, B, and (parameterized) C in addition to a referentiality hierarchy, since these two sets of conditions would seem to rule out some of the same structures: As Narahara (1995) argues, Condition C could be reduced almost entirely to the referentiality hierarchy (especially if this hierarchy is conceived of as scalar rather than discreet or binary) and there is some question whether the empirical data are consistent with the parameterization proposed by Lasnik (1989). Most importantly, such a proposal seems to be missing a vital generalization about referentiality and anaphora.

2.1.1 Safir's FTIP cannot account for bound names

Various attempts have been made to unite these generalizations about "referentiality" and the binding of pronouns and R-expressions. A recent one is the Form to Interpretation Principle (FTIP), a competitive algorithm proposed by Safir (2002a,b) as a replacement for binding Conditions B and C.

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

This principle of interpretation assumes the existence of an (apparently language specific) hierarchy of "dependence." It has the effect of predicting that pronouns can receive reflexive (dependent) readings if (and only if) there is no anaphoric form available in a given syntactic context for the same interpretation. This appears to be counter-exemplified by Hmong, since Hmong can form reflexive expressions by the addition of *tug kheej* 'CLF self' to pronouns and R-expressions. Thus, Hmong allows both of the following reflexive sentences:

(6) a. Nwg_i yeej qhuas nwg_i.
3SG always praise 3SG.

'He always praises himself.'

b. Nwg_i yeej qhuas [nwg [tug kheej]]$_i$.
3SG always praises 3SG CLF self

'He always praises himself.'

Under the FTIP, we might expect (6a) to have an obviative reading (where both pronouns would have to pick out separate referents from outside the sentence). While my consultants claimed that such a reading was possible, they asserted that it would require a highly specific discourse context. The reading given here is the preferred default reading. In (6b), only the reflexive reading is available. However, (6b) is pragmatically marked in that it has object focus, whereas (6a) is pragmatically unmarked. Since this may be seen as a difference in interpretation, this does not contradict the FTIP as proposed by Safir[5].

The real problem with the FTIP emerges when we look at names in relation to pronouns. According to the FTIP, if a form can receive a reflexive reading, it is because there is no more dependent form in the lexicon that would be available in that syntactic context. If we set up a dependency hierarchy like that in (7) it would explain the facts seen in the bound-pronoun examples:

(7) **Most Dependent Scale**

$$\text{Pron} \gg \left\{ \begin{array}{l} \text{[Pron [}\textit{tug kheej}\text{]]} \\ \text{[Name [}\textit{tug kheej}\text{]]} \end{array} \right\} \gg \text{Name} \qquad \text{(provisional)}$$

In this scheme, pronouns are the most dependent forms. The *tug kheej*-forms are less dependent, and thus can receive dependent readings in pragmatically marked contexts. But given such a hierarchy (and with naive notions about the lexical identity of elements in the sentence), it is impossible to explain the following examples:

(8) a. Pov_i yeej qhuas $\text{Pov}_{i/??j}$.
Pao always praise Pao

[5] Although, if this sort of argument is used to defend the FTIP, it is questionable whether it actually makes any interesting predictions, since it would be quite odd if the lexicons of many languages contain lexical items differing only in Safir's putative property of dependence.

‘Pao always praises himself.’
?‘Pao always praises (the other) Pao.’

b. Pov_{i} yeej qhuas [nwg [tug kheej]]$_{\text{i}/?\text{j}}$.
Pao always praise 3SG CLF self

‘Pao always praises himself.’

c. Pov_{i} yeej qhuas [Pov [tug kheej]]$_{\text{i}/??\text{j}}$.
Pao always praise Pao CLF self

‘Pao always praises himself.’

d. Pov_{i} yeej qhuas $\text{nwg}_{*\text{i}/\text{j}}$.
Pao always praise 3SG

‘Pao always praises him.’
*‘Pao always praises himself.’

If pronouns are more dependent than R-expressions, then the availability of *Pov* for a dependent reading in (8a) would imply the availability of *nwg* for a dependent reading in (8d). However, this is not what we observe. In fact, no permutation of these types along the dependency scale will give us the desired predictions. This is true due to the fact that pronouns may be bound by pronouns but not names and that names may be bound by names but not pronouns. This observation is a specific instance of a more general observation, namely that locally bound R-expressions and pronouns are always identical to their antecedents, a restriction Lee (2003) dubs the *Identical Antecedent Requirement*.

2.1.2 Locally bound R-expressions and pronouns are always identical to their antecedents

As Lee (2003) observed for Zapotec, Hmong bound R-expressions never differ in form from their antecedents. A mature male Hmong person, for example, has two names: a name given at birth (*npe hluas* ‘young-name’) and a name given after he has fathered children (*npe laug* ‘old-name’). Even if it is understood that these two names refer to the same person, one cannot be bound by the other:

(9) a. Kub npe laug hu ua Tshuv-Ntxaij.
Kou name old call do Chu-Ndzai.

‘Kou’s old-name is Chu-Ndzai.’

b. Kub_{i} yeej thuam $\text{Tshuv-Ntxaij}_{*\text{i}/\text{j}}$.
Kou always criticize Chu-Ndzai

‘Kou_{i} always criticizes $\text{Chu-Ndzai}_{\text{j}}$.’
*‘Kou_{i} always criticizes $\text{himself}_{\text{i}}$.’

Likewise, reflexive readings are not possible with bound pronouns unless the pronouns are identical:

(10) a. Suavdlawg$_i$ yeej qhuas suavdlawg$_i$.
everyone always praise everyone.

'Everyone (as a group) always praises themselves (as a group).'

b. Suavdlawg$_i$ yeej qhuas puab$_{*i/j}$
everyone always praise 3PL

'Everyone (as a group) praises them (a different group).'
*'Everyone praises themselves.'

This generalization is stable throughout all of the data I have examined and elicited, and is perhaps the most significant evidence that the bound element in what appear to be bound R-expression and bound pronoun constructions are simply copying their phonological form from their antecedents.

2.1.3 Locally bound R-expressions and pronouns are non-referring

The second important piece of evidence supporting this line of reasoning is the fact that these bound forms do not have the ability to pick out referents in the real world, even when the forms are names. This can be shown in VP ellipsis examples:

(11) a. Koj$_i$ yeej qhuas koj$_i$; nwg$_j$ 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$_i$ yeej qhuas Pov$_i$; Maiv$_j$ los kuj ua le hab.
Pao always praise Pao May TOP also do as too

'Pao always praises himself, and so does May.'

For both examples in (11), the sloppy reading is preferred over the strict reading, although some of my consultants report that a strict reading is possible, at least for (11b). Lee (2003) reports similar data for Thai and SLQZ, but asserts that they *only* allow sloppy readings:

(12) 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]

These examples suggest that in Hmong, Thai, and SLQZ, the thing that is copied in the ellipsis of the VP is not an R-expression with full referential potential, but simply the relationship between between the antecedent and the form depending upon it. In cases where a non-bound R-expression is elided, the copy is referential:

(13) Pov nyam Maiv; Maiv los kuj ua le.
Pao like May May TOP also do as

'Pao likes May and May does too (likes May).'

If a name is not referential—if it lacks deictic potential and idiosyncratic content, as appears to be the cases in (11), is it still a name?

2.2 The solution: *ana*

The solution to this problem I propose is slightly different than that given by Lee (2003), at least in its exposition. Lee states that these bound copies are elements of "type *e*," but does not suggest that they are base generated and does not suggest what the featural content of these elements may be. I argue, in mild contrast, that these bound copies are in fact a single anaphoric lexical item that is part of the numeration of the sentence from the beginning of the derivation. I will call this lexeme *ana* (by analogy with *pro*). This explanation is important in distinguishing this type of "variable" from the more tradition types of bound variables discussed below.

Languages which allow bound names and pronouns in the same manner as Hmong and SLQZ include *ana* in their lexicon, whereas languages which do not allow such constructions do not. Having no idiosyncratic content or phonological form of its own, *ana* gains its phonological exponence parasitically from its antecedent. We may thus re-imagine some of the examples given above in new terms. The fact that the sentence *Pov yeej qhuas Pov.* can have two different interpretations can be explained in terms of example (14):

(14) a. Pov_{i} yeej qhuas Pov_{i}.
Pao always praise *ana*

'Pao always praises himself.'

b. Pov_{i} yeej qhuas Pov_{j}.
Pao always praise Pao

'Pao always praises Pao.' (there are two Paos)

In (14a), the second *Pao* is really *ana*, whereas in (14b) it is really *Pao*. We should expect a reflexive reading for (14a), since the lower element is an anaphor, and therefore able to depend for its reference on the upper element. In (14b), both elements are fully referential, therefore one can not depend upon the other for reference.

2.3 *ana* must be A-bound

An interesting observation of (Black 2000:79–84) and Lee (2003) for Zapotec is that quantified R-expressions never serve as the antecedent of bound (reflexive) anaphors. This same generalization holds for Hmong.

In Hmong, it should be noted, common nouns are properties rather than entities. They do not have individual real world referents (and cannot be quantified) unless they are converted by a classifier, measure word, adjective (for which, see (Bisang

1993:8)), or (in Mong Leng) individuating prefix. As a result, reflexive readings with bare common nouns are semantically strained:

(15) dlev pum dlev.
dog see dog

?'Dogs see dogs.'
??'(A) dog sees itself.'

But if the noun is individuated with an adjective or prefix, the reflexive reading becomes more readily available:

(16) a. quas-dlev$_{i}$ pum quas-dlev$_{i/j}$.
IND-dog see IND-dog

'(The/a) dog sees itself.'
'(The/a) dog sees a dog.'

b. [nam dlev]$_{i}$ pum [nam dlev]$_{i/j}$.
great dog see great dog

'The Ol' Dog sees itself.'
'The Ol' Dog sees the (other) Ol' Dog.'

However, if the noun phrase is quantified (which usually entails the addition of the noun classifier) a reflexive reading becomes impossible:

(17) a. [ob tug dlev]$_{i}$ pum [ob tug dlev]$_{*i/j}$.
two CLF dog see two CLF dog

'Two dogs saw two (other) dogs.'
*'Two dogs saw themselves.'

b. [txhua tug dlev]$_{i}$ pum [txhua tug dlev]$_{*i/j}$.
every CLF dog see every CLF dog

'Every dog sees every (other) dog.'
*'All the dogs see themselves.'

Of course, it is possible to reflexivize quantified elements in Hmong. In fact, there are two strategies: one that Lee (2002) describes for SLQZ and one that she attributes to Thai. In the SLQZ-like strategy, the quantified element is topicalized and the reflexive relationship is shown with the distributive pronoun *nyas*:

(18) a. [Txhua tug dlev]$_{i}$ mas nyas$_{i}$ rov qaab pum nyas$_{i}$.
every CLF dog TOP DIST return back see *ana*

'Every dog saw itself.'

b. [Peb tug kwv-tij]$_{i}$ mas nyas$_{i}$ yeej qhuas nyas$_{i}$ has tas [nwg [tug
three CLF brother TOP DIST always praise *ana* say that 3SG CLF
kheej]]$_{i}$ zoo xwb.
self good only.

'Each of the three brothers praised himself saying that he alone was good.'

This strategy employs *ana*, but its binder is not the quantified element, but the pronoun *nyas*, which is bound as a variable by the topicalized QP.

The second (Thai-like) strategy is to use another type of anaphor—in the case of Hmong, a *tug kheej*-form—as the reflexive:

(19) a. [Txhua tug tub hluas]$_i$ yeej qhuas [nwg [tug kheej]]$_i$.
every CLF boy young always praise 3SG CLF self

'Every boy always praises himself.'

b. [Peb tug kwv-tij]$_i$ yeej qhuas [puab [tug kheej]]$_i$.
three CLF brothers always praise 3PL CLF self

'The three brothers always praised themselves.'

Lee attributes the analogous effect in Zapotec to the type of element, claiming that the impossibility of binding quantified expressions is due to a type mismatch between the bound variable and a quantifier phrase:

> Referential DPs and pronouns, which are elements of type e, may appear as bound copies, but quantified phrases, elements of type <<e,t>,t>, cannot appear as bound copies. (Lee 2002)

I propose an alternate explanation, namely that *ana* must be licensed by a c-commanding antecedent in an A-position:

(20) **Variable Anaphor Licensing Condition (VALC)**
The variable anaphor *ana* must be licensed at LF by a c-commanding antecedent in an A-position.

Quantifier phrases, and nouns in quantifier phrases cannot license *ana* because they move to A′-positions at LF. It is significant to note that this condition does not specify that the licensing antecedent must be local (within Domain D with respect to *ana*). This is by design, since *ana* can be licensed in Hmong (and apparently in SLQZ—see Lee 2003) non-locally:[6]

(21) a. Pov$_i$ xaav has tas Maiv$_j$ nyam Pov$_i$.
Pao think say that May like *ana*

'Pao thinks that May likes him.'

b. Pov$_i$ xaav has tas Pov$_i$ txawj lug-txaj.
Pao think say that Pao able word-money

'Pao thinks that he is good at poetry.'

Assuming that these are instances of *ana*—which seems likely—and not instances of accidental co-reference, *ana* is a long distance anaphor. That is, it can be anteceded from outside Domain D. In the terminology introduced by Safir (2002a), such forms are called Unbounded Dependency (UD) forms.

[6]Actually, Lee (2003) claims that bound copies in SLQZ do not have to be c-commanded by their antecedent. However, the data that she cites do not seem to demonstrate this point, and might better be explained as independent co-reference. In any case, binding without c-command would require a rather fundamental reformulation of binding theory.

2.4 Safir's FTIP and *ana*

Given *ana* and the VALC, in fact, it is now possible to return to Safir's FTIP and find in it a rather insightful account of the distribution of variable anaphors in Hmong (and probably the other languages displaying this type of variable anaphora, such as SLQZ, QZ, and Thai). Given Safir's competitive algorithm, we can see that the ability of *tug kheej*-forms to give reflexive readings in contexts where there is no A-binder, but the preference for *ana* in contexts where there is an A-binder can be derived wholly from the VALC and their respective degrees of dependence. At this point, we may set up a provisional dependence scale:

(22) **Most Dependent Scale**

$$\textit{ana} \gg \left\{ \begin{array}{l} [\textit{ana}\ [\textit{tug kheej}]] \\ [\text{Pron}\ [\textit{tug kheej}]] \end{array} \right\} \gg \left\{ \begin{array}{l} \textit{pro} \\ \text{Pron} \\ \text{Kin-term} \\ \text{Epithet} \\ \text{Name} \end{array} \right\} \qquad \text{(Provisional)}$$

This scale will be refined later. However, at this point it is sufficiently explicit to describe the distribution of the two types of anaphoric elements. From this hierarchy and the FTIP, the following set of data emerge naturally:

(23) a. Pov_{i} yeej qhuas $\textit{ana}_{\text{i}}^{\text{Pov}}$. (*ana*—Unmarked Reflexive)

b. Pov_{i} yeej qhuas $[\textit{ana}^{\text{Pov}}$ [tug kheej]$]_{\text{i}}$. ([*ana* [*tug kheej*]]—Marked Focus)

c. Pov_{i} yeej qhuas [nwg [tug kheej]]$_{\text{i}}$ ([Pron [*tug kheej*]]—Marked Focus)

d. Pov_{i} yeej qhuas $\text{pro}_{*\text{i/j}}$. (*pro*—Obviative)

e. Pov_{i} yeej qhuas $\text{nwg}_{*\text{i/j}}$. (Pron—Obviative)

f. Pov_{i} yeej qhuas $[\text{nam tsuv tum}]_{*\text{i/j}}$. (Epithet—Obviative)

g. Pov_{i} yeej qhuas $\text{Txawj-Pov}_{*\text{i/j}}$. (Name—Obviative)

h. Pov_{i} yeej qhuas $\text{yawg}_{*\text{i/j}}$. (Kin-term—Obviative)

3 A′-Bound Variables

Moving beyond *ana*, however, there is a point at which the FTIP breaks down and stops making the correct prediction for Hmong, or at very least, stops making interesting predictions. For Hmong, this threshold of usefulness lies at point where we attempt to predict the interpretation of A′-bound variables, a species not wholly unknown to the reader who has followed this paper thus far. The problem with the interpretation of these variables, vis-a-vis the FTIP, is that the availability of Hmong anaphoric elements for readings as elements dependent on elements in A′-positions (notably, on topics and QPs), cannot be determined with reference to the dependence of items within the

lexical inventory only. Instead, an algorithm that looks at the lexical dependence of other elements in the sentence is required.

This can be seen most clearly in the indefinite concessive construction created using the topic marker *los*[7]:

(24) Paab tuabneeg twg los puab xaav txug puab cov nkaus xwb.
group person which TOP 3PL think about 3PL group exactly only.

'Any group of people thinks only of their own.'

Take the following examples, where the item receiving the variable reading is highlighted with boldface:

(25) a. [Tug twg]$_i$ los **nwg**$_i$ nyam *ana*$_i^{\text{nwg}}$.
CLF which TOP 3SG like

'Anybody would like themselves.'

b. ? [Tug twg]$_i$ los **nwg**$_i$ nyam *nwg*$_{*i/j}$.
CLF which TOP 3SG like 3SG

'Anybody would like him.'

c. ? [Tug twg]$_i$ los *nwg*$_{*i/j}$ nyam **nwg**$_i$.
CLF which TOP 3SG like 3SG

'He would like anybody.'

(26) a. [Tug twg]$_i$ los **pro**$_i$ nyam nwg$_{*i/j}$.
CLF which TOP like 3SG

'Anybody would like him.'

b. [Tug twg]$_i$ los nwg$_{*i/j}$ nyam **pro**$_i$.
CLF which TOP 3SG like

'He would like anybody.'

(27) a. ? [Tug twg]$_i$ los **pro**$_i$ nyam *ana*$_i^{pro}$.
CLF which TOP like

'Anybody would like themselves.'

b. [Tug twg]$_i$ los **pro**$_i$ nyam *pro*$_{*i/j}$.
CLF which TOP like

'Anybody would like him.'

c. [Tug twg]$_i$ los *pro*$_{*i/j}$ nyam **pro**$_i$.
CLF which TOP like

'He would like anybody.'

[7]The marker *los* is one of at least three topic markers present in Hmong. Informally, an expression *X los Y* can be stated: X is a member of the set of entities or propositions having property Y (or of which Y is a true proposition).

It appears that both *pro* and *nwg* are sometimes available as A′-bound variables in the same syntactic context, as shown by examples (25a) and (27a)[8]. That is to say, the distribution between these two elements does not seem to be complementary. This could be discounted though, based upon the fact that (27a), (27b), and (27c) are all surface identical, and listeners can only know what grammatical structure to attribute to such an utterance from discourse context. (27a) is not pragmatically marked in the sense that it conveys specific pragmatic information different from (25a). However, it is the least preferred reading for this surface utterance, whereas (25a) is the most preferred reading for the surface string represented by the three examples in (26). This being the case, it is not a strong counter-example to the FTIP, but it is clear that something is happening here that the FTIP is failing to capture.

3.1 Dependence and the interpretation of A′-bound variables

Be that as it may, there is another very interesting feature of this paradigm: *nwg* can never have a variable reading if *pro* is available to act as a variable. Taken together with the previous observation that the distribution of *nwg* and *pro* is not complementary, this tendency suggests that dependence on A′-binders is not determined by an algorithm like the FTIP, which compares the dependence of the current lexical item with other items in the lexical inventory, but by an algorithm that compares the dependence of lexical items within a sentence. In other words, a bound variable is not variable because it is the most dependent form that can occupy its position, but because there is no more dependent form available in the sentence.

3.1.1 A revised dependence scale

This becomes even more apparent if we look at the whole gamut of anaphoric and referring expressions available in Hmong. This would consume huge amounts of paper, but the following examples should be sufficient:

[8]Note that I am making a non-trivial assumption about the identity of the ECs in this construction. The linguist unfamiliar with the syntactic phenomena of Hmong would likely assume that the topic was raised to the left edge of the sentence as a result of Move α, and that the empty category bound by the topic would be a trace. One could then attempt to account for the behavior of the EC by an argument about the properties of traces. However, such an account would not elegantly account for a number of the facts displayed in this data: significantly, why a pronoun cannot depend on the topic when an empty category is present as one of the arguments of the verb. Perhaps more importantly, assumption that the topic is raised to its position rather than being generated there is problematic in light of examples like those in (28):

(28) a. [Tug twg]$_i$ los Pov xaav has tas peb yuav-tau khu **nwg**$_i$ lub tsheb.
CLF which TOP Pao think say that 1PL should fix 3SG CLF car
'Pov thinks that we should fix anyones car.'

b. [Tug hluas nkauj twg]$_i$ los Pov xaav has tas peb yuav-tau paab *pro*$_i$.
CLF youth maiden which TOP Pao think say that 1PL should help
'Pao thinks we should help any young woman.'

If Hmong has the same bounding nodes of English, the topic construction demonstrated in these examples must be a base-generated left-dislocation structure rather than the result of movement. Otherwise, these sentences would display subjacency violations. For related arguments for Thai see Hoonchamlong (1991) and Aroonmanakun (1999).

(29) a. [Tug twg]$_i$ los **nwg**$_i$ tsi nyam [nam tsuv tum]$_{*i/j}$.
CLF which TOP 3SG NEG like great tiger bite

'Anybody would dislike the damn bastard.' (lit. 'the great tiger bite')

b. [Tug twg]$_i$ los [nam tsuv tum]$_{*i/j}$ tsi nyam **nwg**$_i$.
CLF which TOP great tiger bite NEG like 3SG

'The damn bastard would dislike anyone.'

c. ? [Tug twg]$_i$ los [**nam tsuv tum**]$_i$ tsi nyam [nam tsuv tum]$_i$
CLF which TOP great tiger bite NEG like *ana*

'No matter who he is, the damn bastard dislikes himself.'

(30) a. [Tug twg]$_i$ los **nwg**$_i$ tsi nyam Pov$_{*i/j}$.
CLF which TOP 3SG NEG like Pao

'Anybody would dislike Pao.'

b. [Tug twg]$_i$ los Pov$_{*i/j}$ tsi nyam **nwg**$_i$.
CLF which TOP Pao NEG like 3SG

'Pao would dislike anyone.'

c. * [Tug twg]$_i$ los Pov$_i$ tsi nyam *ana*$_i^{\text{Pov}}$
CLF which TOP Pao NEG like

Intended: 'No matter who he is, Pao dislikes himself.'

Just as *nwg* cannot receive a variable reading if *pro* is available to take the reading, an epithet or name cannot receive a variable reading if an appropriate pronoun is available to act as the variable. This allows us the further refine our dependence scale:

(31) **Most Dependent Scale**

$$\textit{ana} \gg \left\{ \begin{array}{l} \textit{ana}\ [\textit{tug kheej}]] \\ \text{Pron}\ [\textit{tug kheej}]] \end{array} \right\} \gg \textit{pro} \gg \text{Pron} \gg \text{Epithet} \gg \left\{ \begin{array}{l} \text{Kin-term} \\ \text{Name} \end{array} \right\}$$

It also appears that epithets are only marginally able to serve as variables and that names (and presumably less-dependent forms) are not able to function as variables at all. This is not surprising, since names have inherent deictic potential.

3.1.2 Other constraints on A′-variables

Of course, the variable reading is not simply assigned to the most dependent element in the sentence—otherwise, in (25a) and (27a), the variable reading would be assigned to *ana*. However, common sense would rule this out since *ana* must be directly dependent upon an element in an A-position, and *tug twg* is in an A′-position. However, there are additional examples that raise a similar quandary[9]:

[9]Of course, there is a similar resolution available for (32) if it is recognized that *tug kheej*-forms must be subject to some antecedent licensing condition as well.

(32) a. [Tug twg]$_i$ los **pro**$_i$ nyam [nwg [tug kheej]]$_i$.
CLF which TOP like 3SG CLF self

'Anybody likes themselves.'

b. [Tug twg]$_i$ los **nwg**$_i$ nyam [nwg [tug kheej]]$_i$.
CLF which TOP 3SG like 3SG CLF self

'Anybody likes themselves.'

If the only criteria for receiving a variable reading in a sentence that sets up a variable context was being the most dependent form in the sentence, we would expect [*nwg* [*tug kheej*]] to be directly dependent upon *tug twg* in both sentences in (32). Such a hypothesis would predict that the pronominal elements *pro* and *nwg* in (32) would not receive variable readings, would not depend upon *tug twg*, and therefore pick up their referents from outside the sentence. It is clear, then, that the principle that selects the variable within a sentence should only look at elements that do not depend upon an antecedent more local than the A′-binder.

3.1.3 The A′-bound variable interpretation principle

Based upon these facts, it is possible to formulate a provisional principle that can predict the binding facts for A′-bound variables:

(33) **Variable Interpretation Principle (VIP)** (Provisional)
An anaphoric element α may be interpreted as a variable dependent on the constituent at position x if α is the most dependent anaphoric element available in sentence y.

1. α is available in sentence y if α has no potential antecedent in sentence y more local than x.
2. α is the most dependent form available in sentence y if it is available in y and there is no form β such that it is available in y and is lexically marked as more dependent than α.

According to this principle, it appears likely that *ana* should never be bound from an A′-position—the desired effect. Since topics are at the left periphery of the sentence, since *ana* must be licensed by a potential antecedent in an A-position (which, due to the syntactic structure of Hmong, should not be on the left periphery), and since the word order of Hmong is relatively inflexible, the potential A-binder should always be more local to *ana* than the potential A′-binder.[10]

[10] But see the case of elided VPs which seem to involve vehicle change:

(34) Pov$_i$ xaav has tas Pov$_i$ nam hlub Pov$_i$; txhua tug tub hluas kuj xaav le hab.
Pao think say that *ana* mother love *ana* every CLF son young also think as too

'Pao thinks that his mother loves him, and so does every boy.'

This sentence gets a sloppy reading, implying the that reconstruction of the ellipsis copies the relationship between the *ana*s and their antecedent, *Pov*. The copied forms would then be dependent on the A′-constituent *txhua tug tub hluas*.

3.2 Some problematic sentences

Unfortunately, my VIP does not solve all the problems in the interpretation of A′-bound variables in Hmong. There are a number of sentence types, mentioned briefly here, that present problems for this interpretive principle.

3.2.1 Donkey sentences

It is possible for a full clause to appear as the topic of a Hmong sentence. It is also possible for this clause to contain an indefinite pronoun. In these cases, the patterns of dependence are different than that predicted by the VIP:

(35) a. [Tug twg]$_i$ pum *pro*$_j$ los [**tug twg**]$_i$ yeej nyam *pro*$_j$.
CLF which see TOP CLF which always like

'Whoever sees him will surely like him.'

b. [Tug twg]$_i$ pum *pro*$_j$ los **nwg**$_{i/*j}$ yeej nyam *pro*$_{j/*i}$
CLF which see LOS 3SG always like

'Whoever sees him surely like him.'

c. [Tug twg]$_i$ pum *pro*$_j$ los **pro**$_i$ yeej nyam *pro*$_j$.
CLF which see TOP always like

'Whoever sees him will surely like him.'

According to the VIP, in (35b) *pro* rather than *nwg* should receive the variable interpretation. Furthermore, in (35c), the VIP would predict that either of the instances of *pro* could receive the variable reading, but this does not appear to be the case.

Interestingly enough, (35a) looks rather like certain famous examples of donkey sentences in Chinese such as:

(36) Shei$_i$ xian lai **shei**$_i$ (jiu) xian chi.
who first come who then first eat

'Whoever comes first eats first.' (Cheng and Huang 1996)

However, it is not possible to produce Mandarin equivalents of the Hmong sentences in (35):

(37) a. * Shei kan jian *pro*, **shei** (jiu) xihuan *pro*.
who look see who then like

Intended: 'Whoever sees (him), likes (him).'

b. Shei$_i$ kan jian ta$_j$, **shei**$_i$/**pro**$_i$ (jiu) xihuan ta$_j$.
who look see 3SG who/*pro* then like 3SG

'Whoever sees him, likes him.' (Patrick Chew p.c.)

Nevertheless, it appears that the principle that determines dependency relations in both languages makes reference to the relative structural position of the anaphoric elements in the protasis clause and the apodasis clause.

3.2.2 Some QP-bound variables

So far we have not examined QP-bound variables except vary briefly. This is partly because they do not present as straightforward a means of comparing the dependence of two anaphoric elements as the topic construction that has dominated our discussion of A′-bound variables. This is also because QP-bound variables provide some interesting cases that appear to contradict the VIP.

Example (39) presents a relatively tame example of a QP-bound anaphor in Hmong:

(38) [Txhua tug tub hluas]$_i$ xaav has tas [nwg [tug kheej]]$_i$ zoo tshaaj *pro*$_j$.
every CLF boy young think say that 3SG CLF self good surpass

'Every boy thinks that he is better (than another/the others).'

Nothing in this sentence seems to contradict the VIP, since *tug kheej* forms are assumed to be more dependent than *pro*. However, if *nwg tug kheej* is replaced by *nwg*, it still gets the dependent reading:

(39) [Txhua tug tub hluas]$_i$ xaav has tas nwg$_i$ zoo tshaaj *pro*$_j$.
every CLF boy young think say that 3SG good surpass

'Every boy thinks that he is better (than another/the others).'

The behavior observed in this example may well be due to the selectional properties of *tshaaj* 'to surpass'—it may demand a plural argument, thus preventing the *pro* that is its argument from acting as a distributive variable.

4 Conclusion

None of the examples given here are necessarily fatal for the FTIP, VALC, or VIP, but they do point out the need for further investigating what facts govern anaphoric dependence in Hmong, and the relationship between the syntactic phenomena of this under-studied language and those of better studied languages. For example, it is still uncertain whether the relationship described by the VALC applies to *ana* in all languages that have this lexical item, or whether a reformulation of this principle can actually describe the distribution of a class of long-distance anaphors other than *ana*. Furthermore, it should ultimately be determined whether the patterns of dependence for A′-bound variables can be extended to other languages, or whether the observation made here about the relation between dependence and variable binding is an epiphenomenon of some other universal principle.

What should be clear from the previous discussion, however, is that what appear to be bound R-expressions and pronouns in Hmong are actually instances of an anaphor *ana*. This form is variable in the sense that its phonological form is determined by its referent, but is not a variable in the sense that it is bound by a WH-expression or quantifier expression. In fact, it is always A-bound. And like other A-bound elements in Hmong, Safir's notion of dependence and FTIP provide a useful means of explaining its interpretive possibilities. It is distinct, in this respect, from A′-bound variables in Hmong, for which the FTIP is inadequate, but for which the notion of dependence is still useful.

References

AROONMANAKUN, WIROTE 1999. "Referent resolution for zero pronouns in Thai". Ms, Chulalalongkorn University.

BISANG, WALTER 1993. "Classifiers, quantifiers, and class nouns in Hmong". *Studies in Language* 17(1):1–51.

BLACK, CHERYL A. 2000. *Quiegolani Zapotec Syntax*. Dallas, Texas: SIL and the University of Texas at Arlington.

CHENG, LISA LAI-SHEN AND JAMES HUANG 1996. "Two types of donkey sentences". *Natural Language Semantics* 7:121–163.

HEIMBACH, ERNEST E. 1969. *White Meo–English Dictionary*. Ithaca: Cornell Southeast Asia Program.

HOONCHAMLONG, YUPHAPHANN 1991. *Some Issues in Thai Anaphora: A Government and Binding Approach*. Ph.D. thesis, University of Wisconsin-Madison.

LASNIK, HOWARD 1989. *Essays on Anaphora*. Boston: Kluwer Academic.

LEE, FELICIA 2002. "Anaphoric R-expressions as bound variables". In "Proceedings of the Berkeley Lingustics Society", 28. Berkeley, CA: Berkeley Lingustics Society.

— 2003. "Anaphoric R-Expressions as bound variables". To appear in *Syntax*.

LI, YEN-HUI AUDREY 1992. "Indefinite *wh* in Chinese". *Journal of East Asian Linguistics* 1(2):125–155.

NARAHARA, TOMIKO 1995. "Alternatives to reflexives in Thai and Vietnamese: Binding theory and language variations". In Mark Alves (ed.), "Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society", Tempe, Arizona: Arizona State University, Program for Southeast Asian Studies.

RATLIFF, MARTHA 1992. *Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes, and Expressive Phrases in White Hmong*. Number 27 in Monograph Series on Southeast Asia. DeKalb Illinois: Northern Illinois University Center for Southeast Asian Studies.

— 1997. "Hmong-Mien demonstratives and pattern persistence". *Mon-Khmer Studies* 27:317–328.

SAFIR, KENNETH J. 2002a. "The syntax of anaphora". Ms Rugers University.

— 2002b. "The syntax of (in)depedence". Ms Rugers University.

THOMPSON, LAWRENCE C. 1965. *A Vietnamese Grammar*. Seattle: University of Washington Press.