

Hmong Elaborate Expressions are Coordinate Compounds

David Mortensen
dmort@socrates.berkeley.edu
UC Berkeley

December 11, 2003

Tyger! Tyger! burning bright
In the forests of the night,
What immortal hand or eye
Dare frame thy fearful symmetry?

—William Blake (1757-1827), *Songs of Experience*.

1 Introduction

On exposure to Hmong discourse¹ discourse, or that of many other structurally-similar Southeast Asian languages—the speaker of English can hardly help but be impressed by the pervasiveness of parallelism in this discourse tradition. Almost everything seems to be repeated, often in balanced pairs. Of course, parallelism in language is by no means confined to Southeast Asia—as a stylistic device it is ubiquitous. However, in Hmong at least, it is hard to escape the conclusion that parallelism plays a greater role in the grammatical structure of the language than it does in English. In addition to the usual types of stylistic parallelism, Hmong features an impressive array of symmetrical grammatical structures. Of particular interest in this regard are two morphological constructions: headless coordinate compounds consisting of pairs of synonyms, near-synonyms, antonyms, and converses like *caij-nyoog* ‘time-time’ (‘time’), *num-tswv* ‘official-lord’ (‘leader’), and *muag-nug* ‘sister-brother’ (‘brother and sister’); and so-called elaborate expressions—parallel expressions superficially like the English idioms *easy-come easy-go* and *like father like son*. In this latter category are compounds like

¹Hmong is a Southeast Asian language, belonging to the Western Hmongic group within the Hmong-Mien family. The data in this paper are from the Mong Leng dialect, which is widely spoken in China, Vietnam, Laos, and Thailand (often alongside the better-known Hmong Daw or White Hmong dialect). The data are recorded in the Hmong Romanized Popular Alphabet, the most widely used Hmong orthography. This orthography is exceptionally practical, but has a number of oddities that take some time to get used to. Among these is the use of consonant symbols at the end of each syllable to represent tone, the use of doubled vowel symbols to represent nasalization, the use of the symbol <x> to represent a voiceless dental fricative, the use of the symbol <s> to represent a voiceless alveopalatal fricative, and the use of <r> to represent a retroflex stop.

ua-tsuu-ua-roq ‘do-tiger-do-war’ (‘engage in warfare’) and *yaug-ncauj-yaug-lu* ‘rinse-mouth-rinse-mouth’ (‘receive as compensation for services performed’). While these two phenomena have been traditionally been described separately, the differences between them are ultimately shallow. This paper will argue, taking an extended version of Construction Morphology as a framework, that Hmong elaborate expressions are instances of the same construction as coordinate compounds.

Extending this claim to cover all elaborate expressions, however, has a number of significant formal consequences for Construction Morphology. First, it requires that coordinate compounds—words—be allowed to have internal syntactic structure. Second, some mechanism is required for capturing an apparent bracketing paradox in which simple coordinate compounds occur as discontinuous constituents within an elaborate expression. I argue that the problem of representing two overlapping structures (on “syntactic” and one lexical/morphological) can be largely resolved by introducing into Construction Morphology a mechanism of indexation and slash features similar to that found in syntactic theories such as GPSG and HPSG.

2 Coordinate Compounds

Coordinate compounds are defined in this paper as headless compounds having a parallel structure. Similar compounds occur in many languages, and have been given almost as many labels (though, admittedly, some of these do not describe exactly the same phenomena): dvandva compounds, copulative compounds, appositional compounds, parallel compounds, coordinating compounds, co-compounds, symmetrical compounds, redundant compounds, fixed binomials, binoms, and probably others. In some languages, like Sanskrit and English, there are no strong symmetry requirements imposed upon these constructions. When speaking English, it is possible to freely coin noun compounds like ‘doctor-poet’, ‘beast-man’, or ‘singer-songwriter’, in which cases the referent of the composed forms is more-or-less the intersection of the referents of the individual nouns (but see Olsen 2001 for a more nuanced view). In these cases, the only semantic requirement seems to be that there be some possible intersection between the two referents.

In Hmong and many other languages of East and Southeast Asia, the restrictions that hold upon coordinate compounds are far more stringent, and the semantics of the composed expressions are proportionally more interesting. While the semantic features of an English coordinate compound is the union of the semantic features of its parts, entailing that its referent is a subset of the referents of each of its parts, the semantic features of Hmong coordinate compounds are an intersection of those of its parts, meaning that their referents are either equivalent to those of their parts (as in synonym compounds), the union of the referents of the parts (as in converse constructions), or a superset of the referents of the parts (as in near-synonym or antonym compounds).

It is also noteworthy that the the Hmong compounds of this type are both obligatorily binary and are fixed in their order. While English, to return to a familiar example, has both fixed binary expressions (‘law and order’, ‘up and down’, ‘cat and mouse’) and has many binary coordinate compounds, English coordinate compounds are not typically fixed in order and can contain any number of members. One can

as easily say ‘activist-singer-songwriter’, ‘singer-songwriter-activist’, or ‘songwriter-activist-singer’ as ‘singer-songwriter’. In contrast, in Hmong, the equivalent structures would not be grammatical. *Qab-npua* ‘chicken-pig’ (‘small livestock’) cannot be reversed as **npua-qab* ‘pig-chicken’, nor can a third member be added to the compound as in **qab-npua-nyuj* ‘chicken-pig-cow’.

2.1 Semantic Types

It is possible to delineate a number of different relationships that may exist among the members of coordinate compounds in Hmong. However, it is possible to reduce these several types to just two types: symmetrical and converse.

2.1.1 Converse pairs

The members of some Hmong coordinate compounds are in a strictly converse relationship to each other. If the two members are treated as relations \mathfrak{A} and \mathfrak{B} , then $(x\mathfrak{A}y \rightarrow y\mathfrak{B}x) \wedge (y\mathfrak{B}x \rightarrow x\mathfrak{A}y)$. Not surprisingly, most such compounds consist of kinship terms:

- (1) a. *kwv* + *tij* → *kwv-tij*
 ‘younger brother’ ‘older brother’ ‘brothers; clansmen’
 b. *muag* + *nug* → *muag-nug*
 ‘sister’ ‘brother’ ‘brother and sister’

In other cases, the converseness of the relationship between the two members is less perfect, but it is still clearly a reciprocal relationship:

- (2) a. *nam* + *tub* → *nam-tub*
 ‘mother’ ‘son’ ‘mother and son’
 b. *txiv* + *tub* → *txiv-tub*
 ‘father’ ‘son’ ‘father and son’

These are a kind of “second-order” converse, where the members refer to subsets of relations that are converses of one another, such as PARENT/CHILD. But despite the differences in the logical structure of this “second-order” group and the strict converses, they are parallel in their (very simple) semantics: the referent of the compound is simply a collective containing instances of one or more of each of its members.

2.1.2 Synonym pairs

It is also extremely common for the members of coordinate compound to be in a relationship of synonymy, where the referent of the two members overlap entirely. Synonym compounds consisting of both nouns and verbs are attested:

- (3) a. *caij* + *nyoog* → *caij-nyoog*
 ‘time’ ‘time’ ‘time’

- b. ncauj + lu → ncauj-lu
 ‘mouth’ ‘mouth’ ‘mouth’
- c. cuab + yim → cuab-yig
 ‘household’ ‘household’ ‘household’
- (4) a. mov + tshua → mov-tshua
 ‘love’ ‘love’ ‘love’
- b. taj + tug → taj-tug
 ‘level’ ‘level’ ‘level’

Cross-linguistically, this type of coordinate compound is surprisingly common. Similar types of compounds are found in Hindi (Singh 1982), Khmer (Ourn and Haiman 2000), Vietnamese (Thompson 1965:130-131), Pacoh (Watson 2001), and Mandarin (Li and Thompson 1981). The semantic content of these compounds is easy to predict—it is essentially the same as that of the individual members. Some of these compounds are confined to a poetic or oratorical register. For example, *ncauj* is the normal word for ‘mouth’ and *ncaujlu* is only used in flowery or ritual speech. However, *cuabyig* ‘household’ can be heard in most registers of speech.

2.1.3 Antonym pairs

Another crosslinguistically attested type of symmetrical compounds found in Hmong are those consisting of antonym pairs Li and Thompson (1981). Like synonym pairs, these occur with both nouns and verbs, as well as other lexical categories:

- (5) a. toj + taug → toj-taug
 ‘hill’ ‘dale’ ‘hill and dale; terrain’
- b. nkauj + nraug → nkauj-nraug
 ‘bachelor’ ‘maiden’ ‘unmarried people; amorous matters’
- c. nub + mo → nub-mo
 ‘day’ ‘night’ ‘day and night; all the time’
- (6) a. caj + tuag → caj-tuag
 ‘live’ ‘die’ ‘whether living or dead; no matter what’
- b. noj + haus → noj-haus
 ‘eat’ ‘drink’ ‘to subsist’
- c. moog + lug → moog-lug
 ‘go’ ‘come’ ‘go back and forth; come and go’

The semantics of these expressions are understandably more complicated than those of the synonym compounds. The meanings of some of the expressions are quite idiomatic. For *nkaujnraug* to mean ‘amorous matters’ is a clear example of this. On the other hand, some of these compounds, like *tojtaug* ‘hill and dale; terrain’ are quite

transparent. In general, the referents of these expressions are the union of the referent of their parts, or some superset of this union. Put more formally, let \mathcal{A} and \mathcal{B} be relations corresponding to the referents of the daughters and let \mathcal{C} be a relation corresponding to the referent of the compound. \mathcal{C} must be some predicate such that $\forall x((\mathcal{A}(x) \rightarrow \mathcal{C}(x)) \wedge (\mathcal{B}(x) \rightarrow \mathcal{C}(x)))$. That is, the referent of the compound is some hypernym of the referents of its constituents. The non-compositional examples seem to be the result of metonymic extensions of this type of original, composition meaning.

2.1.4 Near synonym pairs and representative members of a class

Although there are many clear cases of synonym and antonym pairs, there is a significant body of cases that seem to occupy a problematic netherworld between these two poles. In fact, it is sometimes difficult to distinguish synonym pairs and antonym pairs, because few words are truly synonymous, and the differences between synonyms (when they are not distinguished by factors such as register) often create very local semantic oppositions that are reminiscent of antonymy.

This type of compound is found in many of the same languages that allow The majority of Hmong coordinate compounds actually fall into this category:

- (7) a. kwj + haav → kwj-haa
 ‘canyon’ ‘valley’ ‘valley (e.g. a river valley)’
 b. yaaj + tshis → yaaj-tshis
 ‘sheep’ ‘goats’ ‘ovines’
 c. kaab + kev → kaab-ke
 path way ‘ceremony; ritual’
 d. quaj + nyav → quaj-nyav
 ‘cry’ ‘mourn’ ‘lament’
 e. qhuab + qha → qhuab-qha
 ‘exhort’ ‘teach’ ‘instruct, especially in matters of behavior’

Some of these “near synonyms” are of a special type where two members of the compound refer to representative members or subparts of the category to which the whole compound refers:

- (8) a. lab + cuam → lab-cuam
 monkey gibbon ‘simians’
 b. xyoob + ntoo → xyoob-ntoos
 bamboo tree ‘woody (as opposed to herbaceous) vegetation’
 c. vaaj + tsev → vaaj-tse
 garden house ‘home’
 d. pluj + tuag → pluj-tuag
 ‘be lost’ ‘die’ ‘perish’

- e. dlawb + huv → dlawb-huv
 white clean 'immaculate'

The same two sorts of compounds in Khmer are distinguished by Ourn and Haiman (2000). Thompson (1965:128-132) draws a similar distinction for Vietnamese. However, in the Hmong cases it is often not possible. to do this, especially in the case of verb compounds where the meanings of the two constituent verbs are related as parallel aspects of the same type of event, but are not clearly synonymous. An interesting additional example are numeral compounds:

- (9) a. ob + peb → ob-peb
 'two' 'three' 'a few'
- b. cuaj + kaum → cuaj-kaum
 'nine' 'ten' 'many; however many'

The relationship between the semantics of the daughters and the semantics of the whole compound is subject to the same hypernymy generalization made for the antonym compounds, namely that the compound, as a whole, is a hypernym of its daughters.

2.1.5 Conflating categories

In fact, assuming that this generalization is formalized as in the section on antonym pairs (above), it would apply to the antonym pairs, the class-representative pairs, the near-synonym pairs and the synonym pairs—that is to say, all the types of coordinate compounds except the converse pairs. This appears, at first blush, to be an odd statement, since we do not usually consider a word to be its own hypernym or hyponym. However, if this relationship is formalized as above, this problem disappears. The fact that the semantics of all coordinate compounds except the coordinate type seem to be reducible to a singular formalization points to a fundamental division of these compounds into two major types.

2.2 Symmetry Requirements

Both of these major types are characterized by a type of semantic symmetry: a similar relationship must hold between the meaning of the compound as a whole and each of the parts. There are other symmetry requirements that must be met in order for a Hmong coordinate compound to be well-formed.

2.2.1 Syntactic symmetry

Each construction of this type must have two and only two daughters. Neither of these daughters may depend syntactically upon the other and both daughters must always of the same syntactic type. The compound as a whole is always of this syntactic type as well. It is likely that this set of facts could simply be derived from semantic considerations, because if the semantic features of the two members of the compound meet the converseness or co-hyponymy requirements described above, they almost certainly are of the same syntactic type as well, and the inheritance of this syntactic type by the compound as a whole can be attributed to these same principles.

2.2.2 Phonological symmetry

It is somewhat easier to demonstrate that this construction imposes a requirement of phonological (prosodic) symmetry. Specifically, it requires that both daughters contain the same number of syllables. If one constituent of the compound has two syllables, the other must as well.

- (10) a. laajmej + pejxeem → laajmej-pejxeem
 ‘citizenry’ ‘citizenry’ ‘citizenry’
 b. pejxeem + fuabfwm → pejxeem-fuabfwm
 ‘citizenry’ ‘citizenry’ ‘citizenry’
 c. faajtim + fuabtais → faajtim-fuabtais
 ‘emperor’ ‘emperor’ ‘emperor’

Novel compounds where one member has two syllables and the other has one syllable are unacceptable. But novel compounds with plausible semantics and the proper balance of syllables are, at worst, merely odd.

- (11) a. fuabtais + vaaj → *fuabtais-vaaj
 ‘emperor’ ‘king’
 b. fuabtais + vaajntxwv → ?fuabtais-vaajntxwv
 ‘emperor’ ‘king’ ‘monarch’

The second example, in fact, would probably be acceptable if it was not blocked by this very similar compound *faajtim-huabtais* ‘emperor’, given above.

The coordinate compound construction in Hmong enforces symmetry in at least semantics and phonology. The constructions are characterized by syntactic symmetry as well, but this fact may simply be a side-effect of the other symmetry requirements. Nevertheless, the syntactic symmetry of these compounds is as important, descriptively, as their semantic and phonological symmetry.

2.3 Ordering Principles

So far, our treatment of coordinate compounds has emphasized the ways in which they are symmetrical. However, there is one sense in which the two halves of such a compound are trivially non-coequal, and this is in their linear sequence. All lexicalized coordinate compounds have a fixed order. Furthermore, there are significant restrictions upon the order of novel compounds. From the standpoint of theories of the morphology-phonology interface, it is significant that the primary locus of these restrictions is prosodic.

2.3.1 Phonological factors

The order of constituents, in most coordinate compounds, can be predicted based upon the historical tones of the rightmost syllable of each constituent². The seven modern

²Johns and Strecker (1987:107-108) suggest that tonal patterns may influence the ordering of coordinate compounds and elaborate expressions (though they do not use this terminology). However, they did not have

tones of Mong Leng reflect an older four-tone system (*A, *B, *C, and *D)³ that underwent a conditioned tonal split, yielding eight tonal categories. Two tonal categories subsequently merged, yielding a seven tone system and occasionally rendering phonological phenomena that depend upon the historical four-tone system opaque. These four tones may be arranged in a hierarchy that predicts the ordering of constituents whose target tones belong to different historical categories.

$$(12) *A \succ *D \succ *B \succ *C$$

The Mong Leng reflexes of these proto-tones are somewhat messy, and do not suggest a synchronic rationale for this hierarchy:

$$(13) \text{ a. } -b, -j \succ -s, -m, \succ -v, (-g) \succ -\emptyset, -g$$

$$\text{ b. } /55/, /52/ \succ /22/, /21?/ \succ /35/, /21fi/ \succ /33/, /21fi/$$

However, the pattern is very robust:

- (14) a. tub + ntXHais → tub-ntXHais
 ‘son’ ‘daughter’ ‘children’
- b. Maab + Suav → Maab-suav
 ‘lowlander’ ‘Han’ ‘foreigner’
- c. koob + npe → koob-npe
 ‘reputation’ ‘name’ ‘fame’
- d. noj + haus → noj-haus
 ‘eat’ ‘drink’ ‘subsist’
- e. roj + ntshaav → roj-ntshaav
 ‘oil’ ‘blood’ ‘blood’
- f. tuj + taug → tuj-taug
 ‘poison’ ‘poison’ ‘poison’
- (15) a. nam + txiv → nam-txiv
 ‘mother’ ‘father’ ‘parents’
- b. num + tswv → num-tswv
 ‘official’ ‘lord’ ‘ruler’
- c. tis + taw → tis-taw
 ‘wing’ ‘leg’ ‘limbs (of a bird)’
- (16) a. rig + tsho → rig-tsho
 ‘pants’ ‘shirt’ ‘clothing’

access to a large enough database in order to make a predictive generalization.

³Possibly having the phonetic values *high*, *rising*, *falling*, and *low*, respectively.

- b. teg + taw → teg-taw
 ‘arm/hand’ ‘leg/foot’ ‘limbs’
- c. thov + caw → thov-caw
 ‘beg’ ‘invite’ ‘entreat’
- d. ntshuav + ntshw → ntshuav-ntshw
 ‘camel’ ‘elephant’ ‘exotic beasts of burden’

Curiously, this same pattern of tonal ordering in coordinate compounds was operative in Old Chinese. Ting (1975) shows that 80% of the coordinate compounds and coordinate phrases in the Book of Odes (*shijing*), the *Analects* and the works of Mengzi follow a generalization regarding ordering that is remarkably similar to that made here for Hmong. The Proto-Hmong tones *A, *B, *C, and *D can be shown, through loanword evidence, to correspond to the tonal categories of (genetically unrelated) Middle Chinese (Chang 1953). The hierarchy set up by Ting (1975) is stated in terms of these Middle Chinese tonal categories, and is exactly equivalent to the hierarchy given in (12)⁴.

But this pattern is not exceptionless, but fortunately, the exceptions are not sporadic—this pattern applies almost perfect regularity in coordinate compounds of the co-hyponym type, but does not apply in coordinate compounds of the converse type:

- (17) a. nam + tub → namtub (*tubnam)
 ‘mother’ ‘son’ ‘mother and son’
- b. txiv + tub → txivtub (*tubtxiv)
 ‘father’ ‘son’ ‘father and son’
- c. kwv + tij → kwvtij (*tijkwv)
 ‘younger brother’ ‘older brother’ ‘brothers’

It does, however, apply in coordinate compounds of kinship terms that are not of the reciprocal type:

- (18) a. puj + yawm → pujyawm
 ‘grandmother’ ‘grandfather’ ‘grandparents’
- b. ntshais + vauv → ntshaisvauv
 ‘daughter’ ‘son-in-law’ ‘daughters and sons-in-law’

Beside this difference in the ordering properties of converse-type compounds and the semantic differences already discussed, converse compounds display syntactic behavior that is different from compounds of the co-hyponym type. Specifically, converse-type compounds can be quantified by a numeral without a noun classifier (unlike other nominal elements in the language). This numeral does specify the number of sets, but the number of members in the set:

⁴Ting’s hierarchy is *pingsheng* > *rusheng* > *shangsheng* > *qusheng*.

- (19) a. ob kwv - tij
 two younger brother older brother
 ‘the two brothers’
- b. peb txiv - tub
 three father son
 ‘the three, father and sons (one father and two sons)’
- (20) a. *ob puj - yawm
 two grandmother grandfather
 Intended: ‘the two grandparents’
- b. ob nkawm puj - yawm
 two pair grandmother grandfather
 ‘two pairs of grandparents’
- c. *ob ntxhais - vauv
 two daughter son-in-law
 Intended: ‘the two, daughter and son-in-law’

These facts, taken together, provide additional evidence for the division of Hmong coordinate compounds into two categories: Co-hyponym compounds cannot be quantified without a classifier, may refer to a superset of the referents of their constituents, and must be ordered according to the tonal hierarchy described here. Converse compounds can be quantified without a classifier, may not refer to a superset of the referents of their constituents, and are not ordered according to their tones.

2.3.2 Semantic factors

In converse compounds, the linear order of constituents appears to be determined primarily by semantic factors, although more data is needed before these factors can be stated as an exhaustive and predictive generalization. For most of the known cases of kinship converses, the following principles make the correct predictions:

1. Parent before child
2. Female before male
3. Younger sibling before older sibling

There is at least one problematic case that defies generalization (3), the compound *vivn-caus* ‘sisters’. This compound is no longer transparent in Mong Leng (or the closely related Hmong Daw dialect), since neither of its constituents occur as free morphemes (or, indeed, occur at all, except in this compound). However, the cognate to the first syllable in Xuyong Hmong is *ve*³ ‘older sister (of a female)’, suggesting that the second syllable is a word for ‘younger sister (of a female)’. Based on this fact, and the principles given above, we would predict the opposite linear order for this compound. In

Xuyong, the corresponding form is *ku³ve³*, ‘younger sibling-older sister’, which *does* follow the generalizations made above.

A good deal more research is needed before we can say confidently what the determinants of linear sequence for Hmong converse coordinate compounds are, but the semantic criteria listed here do appear to be operative. The exceptions may be forms that were lexicalized before the current set of semantic conditions came into effect.

2.4 Coordinate Compounding is Morphological

Lexicalization has left its mark on several facets of coordinate compounds in Hmong. Some have non-compositional semantics, and others preserve a tonal sequence the motivation for which has been rendered opaque by a tonal merger. It can hardly be questioned that coordinate compounds are lexical objects. However, it is well understood—even by those who argue most persistently for compounding as a morphological rather than simply syntactic phenomenon—that the lexicalization of constructions does not prove that they are words in the morphological sense (Sadock 1998:165-166). In fact, many undeniably syntactic constructions are non-compositional (and therefore must be stored in the lexicon). It is doubtful, for example, that we would wish to treat English binomials of the kind discussed by Malkiel (1959)—expressions like ‘salt and pepper’ or ‘law and order’—as morphological entities, even though they are obviously lexicalized syntactic entities⁵.

Recently, Olsen (2001) has argued that some “copulative” compounds—like those of English—are morphological while others—like those of Sanskrit—are merely syntactic. This argument was supported, in part, by the fact that English coordinate compounds have a single referent, while Sanskrit coordinate compounds refer to collectives. This argument would seem to favor the syntactic analysis for Hmong, were it not that bare nouns in Hmong have as their referents not entities, but properties Bisang (1993). As such, this argument is not helpful in the Hmong case since coordinate compounds are often neither more or less capable of referring to collections of discreet entities than are ordinary bare nouns.

There are several reasons to believe that the coordinate compound construction in Hmong is, in fact, morphological. By themselves, none of these arguments would seal the case regarding the status of these constructions. However, taken together they provide ample reason to view coordinate compounds in Hmong as words rather than phrases.

2.4.1 Sandhi Domains

The usual domain of tone sandhi interactions in Hmong is the phonological word. Tone sandhi in coordinate compounds is not as predictable as in numeral-classifier constructions or in subordinating compounds, but it nevertheless occurs in a great number of

⁵If fact, lexicalized—or partially lexicalized—syntactic expressions are so common that if all idiosyncratic, and therefore lexicalized, expressions were treated as morphological constructions, there would be little work left for syntax, which would then take on the appearance of a very regular and productive fringe at the edge of morphology.

cases (Ratliff 1992:44). Tone sandhi in Hmong compounds seems to function in a manner analogous to stress alternations in English compounds, marking the construction as a single word rather than a phrase of two words.

2.4.2 Bound Morphemes

It is not uncommon for both of the constituents in a coordinate compound to no longer occur except in that word. The word *viv-ncaus* ‘sisters’ is an example of this type of fossilized compound, as both its members are preserved only in combination. A great many other compounds contain a component about which the language provides no information other than the meaning associated with the compound as a whole. Johns and Strecker (1987:109) examine the case of *qoob-loo* ‘crops’. The first element occurs frequently in isolation, and means ‘crops’ or ‘grain’; the second element never occurs except in this compound, though Johns and Strecker show that this word must be cognate to words in other Hmong-Mien languages meaning ‘earth’ or ‘field’ such as Qo Xiong *lu*⁵³ ‘dry-land field’.

But it should be noted that similar types of isolated forms occur in lexicalized syntactic constructions as well. For example, the *kith* in English *kith and kin* seems now to occur only in this binomial, but this fact would not establish *kith and kin* as a word in any morphological sense.

2.4.3 Blocking

An additional feature of these compounds that makes them seem unlike syntactic constructions is their frequent participation in blocking phenomena. For example, when both members of a co-hyponym compound belong to the same tonal category, both sequences of the two constituents should be well-formed. However, for lexicalized compounds, only one of the orders is acceptable; the second order is blocked by the presence of the compound having the first order in the lexicon:

- (21) a. *tub* + *nyaab* → *tub-nyaab* (**nyaabtub*)
 ‘son’ ‘daughter-in-law’ ‘son and daughter-in-law’
 b. *nyaj* + *kub* → *nyaj-kub* (**kubnyaj*)
 ‘silver’ ‘gold’ ‘wealth’

In all fairness, though, it should be noticed that similar blocking effects can be found in English fixed binomials with internal syntactic structure like *aid and abet*, which seems to block the syntactically well-formed ?*abet and aid*. While this effect does not render such expressions unacceptable in the same degree as the blocked expressions in Hmong like **nyaab-tub* and **kub-nyaj*, it does show that blocking effects can extend to lexical items with internal syntactic structure.

2.4.4 Relation to Reduplication

Recently, Inkelas and Zoll (2003) have argued persuasively that this type of coordinate compound construction is similar to reduplication in significant ways. Under Morphological Doubling Theory, reduplication can be described in terms of morphological

constructions that repeat the same morpheme twice (or select for two morphemes having the same, or almost the same, featural content). This theory predicts the existence of reduplication-like constructions that do not simply consist of two instances of the same morph, or two allomorphs of the same morpheme, but instead contain two morphemes sharing most or all of their semantic features. Inkelas and Zoll, in fact, provide a rather complete descriptive typology of types of morphological doubling that are not—in the conventional sense, at least—reduplication, and the compounds described here seem to fall neatly into this typology. If Inkelas and Zoll are correct in drawing this parallel between synonym compounds and reduplication, we would expect coordinate compounding constructions and reduplication constructions to be of the same general type and to employ the same module of grammar. It would seem to follow that if reduplication is morphological, coordinate compounding of this type must be morphological as well.

In general, the upshot of these arguments is that Hmong coordinate compounds are words, not simply collocations. Phonologically, they act like words. Semantically, the relationship between the members is constrained in a manner that is rather unlike normal syntactic restrictions, and the semantics of the compound as a whole are of a more complex character than would typically be assumed for simple syntactic coordination. Furthermore, these compounds frequently contain fossilized elements that only exist in the coordinate compound construction. Furthermore, lexicalized coordinate compounds block semantically identical but otherwise well-formed novel forms. Perhaps most persuasively, Hmong coordinate compounds have distinct parallels with reduplication (in line with the earlier insights of Singh 1982 regarding Hindi coordinate compounds, Ourn and Haiman 2000 regarding similar pairs in Khmer, and Inkelas and Zoll (2003) regarding Hindi, Khmer, Madurese, and other languages). These parallels imply that the type of composition under discussion here should take place in the same part of the grammar as reduplication.

2.5 A Constructional Account of Coordinate Compounds

Construction Morphology (specifically Sign-Based Morphology, or SBM; Orgun 1996, 1999; Inkelas and Zoll 2003)—a declarative theory of phonology that treats constructions as theoretical primitives—provides a useful framework for formalizing those generalizations that can be made regarding the types of doubling constructions attested in Hmong. The relationship between these constructions may be described in terms of the inheritance hierarchy symbolized informally in Figure 1. We model the shared properties of the various constructions requiring two symmetrical daughters as shared inheritance from a common abstract base construction called *Double*. Likewise, the two major types of coordinate compounds share common characteristics (most importantly, a requirement that the daughters of the construction should be prosodically symmetrical) by virtue of inheriting them from a single abstract ancestor construction *Coordinate Compound*.

In any individual coordinate compound, the syntactic properties, meaning, and phonological content of the compound are a function of those of the daughters. We may capture these relationships notationally through the use of a schema corresponding to each construction type. The schema for the *CONCoCOMP* is given in in Figure

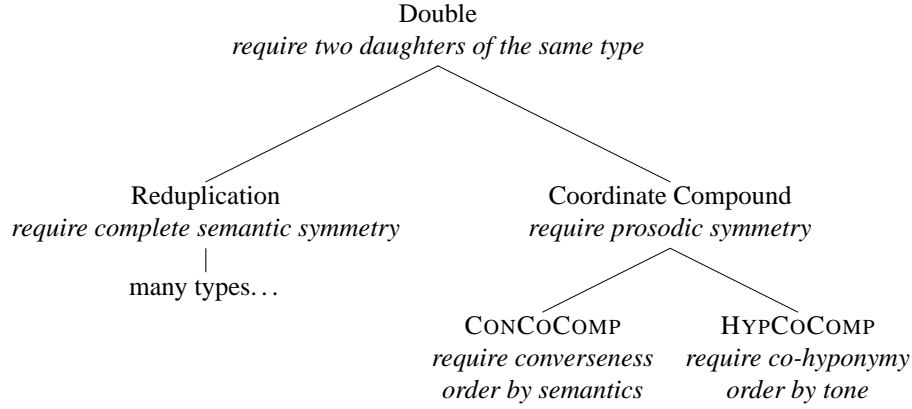


Figure 1: Inheritance hierarchy of morphological doubling constructions in Hmong.

2. Each node is associated with three functions which I will call ι , ς , and ϕ , representing syntactic, semantic, and phonological subgrammars, respectively. The inherited attributes of constructions are encoded in these subgrammars. Thus, the prosodic sym-

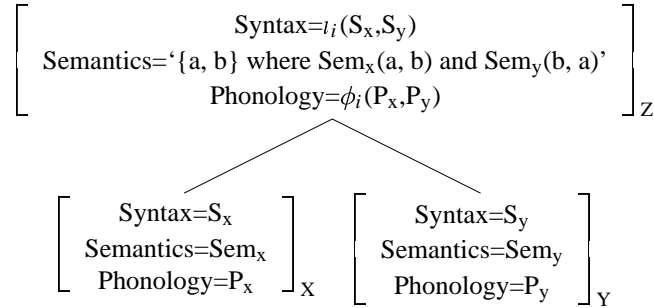
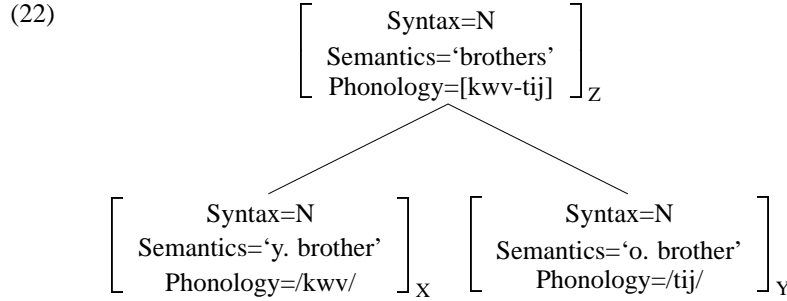
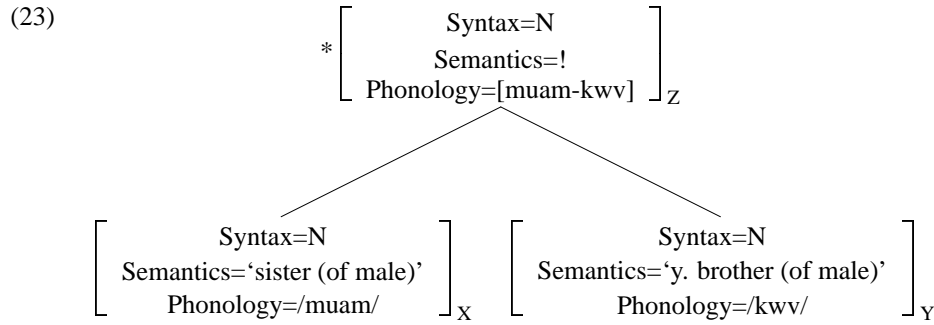


Figure 2: CONCoCOMP construction schema.

metry requirement that CONCoCOMP inherits from Coordinate Compound is encoded in ϕ_i , the phonological subgrammar (or cophonology) associated with CONCoCOMP. This construction is different from its ancestor, however, in that it is associated with ς_i , and thus imposes a semantic requirement of converseness (stated somewhat informally in Figure 2). These attributes have the effect of ruling out potential converse coordinate compounds that do not have these characteristics, and are thus ungrammatical as instances of this construction. A CONCoCOMP instance like *kwv-tij* ‘brothers’ would have the following notational structure:



On the other hand, the string *muam* ‘sister (of a male)’ + *kwv* ‘younger brother (of a male)’ → *muamkwv* is ruled out on semantic grounds (since its members are not converses), even though it is well formed phonologically and semantically:



This same notational scheme makes possible the formation of a somewhat more insightful formalization of the second, and more important, class of coordinate compound constructions in Hmong, those here assigned to the construction HYPCoCOMP (or hypernym coordinate compound). A tentative schema for this construction is given in Figure 4. This construction is associated with a different cophonology, ϕ_j , which en-

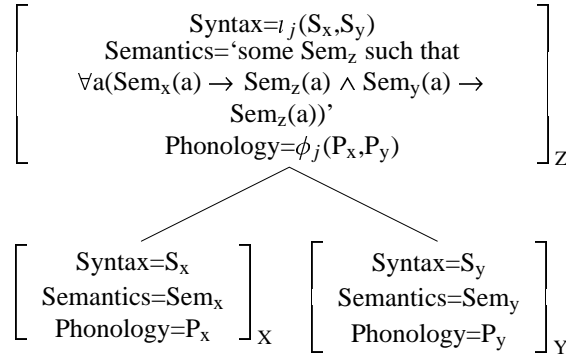


Figure 3: HYPCoCOMP construction schema: first approximation.

forces the order of constituents according to the tones of their last syllable, as discussed above. The syntax function ι_j ensures that the mother node and the two daughters are all of the same syntactic type.

The ς_j function ensures that the referent of the construction as a whole encompasses the referents of its parts. However, this definition should be refined. Under the current analysis, an expression like *nyuj* ‘cattle’ + *neeg* ‘horse’ \rightarrow *nyujneeg* ‘large domestic ungulates’ could instead refer to ‘ungulates’, ‘hairy animals’, ‘living things’, or ‘physical entities’. There must be some generalization that constrains the reference of these constructions.

If it is assumed that the semantics of each lexical word can be described as a set of semantic features that are predicates restricting the reference of the lexical word and that the ‘meaning’ of the word is the conjunction of this set of predicates, then the semantic content of the HYPCoCOMP construction can be restated as the intersection of the semantic features of its two daughters. In other words, $\varsigma(X, Y) = X \cap Y$, where X , Y , and $\varsigma_j(X, Y)$ are sets of semantic features rather than predicates formed by the conjunction of these features. This formulation of the semantics of this construction—in terms of its features rather than its referent—has the effect of predicting both the generalization and the restriction the generalization associated with this construction. This analysis also unites the semantics of the several ‘types’ of HYPCoCOMP elegantly. The intersection of the set of semantic features of a synonym pair should be identical to the set of semantic features of each of the synonyms. Likely, the intersection of the semantic features of two co-hyponyms should correspond to their most ‘local’ hyponym. This analysis still does not address a certain problem of over-generation that

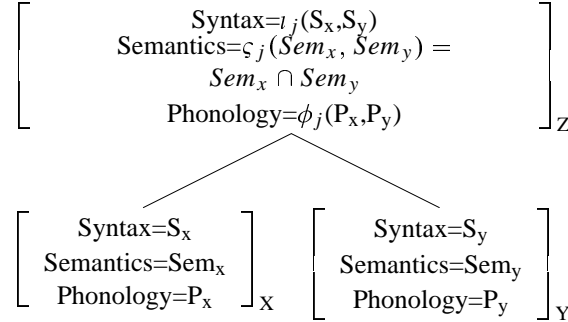


Figure 4: HYPCoCOMP construction schema: final version.

was also characteristic of the first analysis, namely that this construction would allow as daughters words having no recognizable semantic relationship to each other. In the odd chance that the two constituents did not share any semantic features in common, $\varsigma_j = \{\}$, and these cases would thus be excluded as uninterpretable, but words sharing semantic features but not belonging to a recognizable class should still be able to form an interpretable coordinate compound. This fact could be ignored in the formalism, and attributed to pragmatic and semantic factors, but that would weaken the parallelism between this type of construction and reduplication—a bad move in light of the strong empirical case that they should be analyzed together. However, this is ultimately not a real cause for alarm since HYPCoCOMP and its daughter-constructions (if the subtypes of this construction ultimately prove irreducible) could implement $\varsigma_j(X, Y)$ in such a way that it requires the cardinality of $X - (X \cap Y)$ and $Y - (X \cap Y)$ to be minimal

by some criteria. The first clause would give the construction the appropriate semantics, based upon the semantics of the daughters, and the second clause would rule out unattested pairings of semantically unrelated daughters.

It is significant that the schemas I have given here do not subcategorize for the syntactic type of the compound constituents. Instead, they make the syntactic category of the mother node a function of that of the daughter nodes without placing any such restrictions. This has the effect of allowing coordinate compounds to be of any syntactic type for which a semantically plausible pair could be found, and allows us to capture generalizations among compounds of nouns, verbs, numerals, quantifiers, and noun classifiers (all of which compounds display a similar cluster of properties). By assigning this structure, we are assuming that the constituents of a coordinate compounding construction may themselves be morphologically complex⁶.

3 Elaborate Expressions

Essentially, the above analysis of coordinate compounds predicts the existence of so-called elaborate expressions: parallel constructions usually characterized by the repetition of some element and typically have four parts. These constructions are found in many languages of Southeast Asia including Lahu (Matisoff 1973:81-86, 297-301), Burmese (Wheatley 1990:850), Pacoh (Watson 2001), and Thai (Hudak 1990:768). Hmong elaborate expressions are most common of the form ABAC, where each letter represents a monosyllabic element. The form ABCB is also quite well attested, and the form ABCD is common as well (although Matisoff (1973) does not consider these to be true elaborate expressions).

All of the (four part) elaborate expressions have the following structure:



The two constituents of the expression, 1-2 and 3-4 are typically headed compounds (although coordinate compounds and other types of expressions are also to be found in elaborate expressions). They are always of the same type and are always semantically parallel. They are coordinated to form the elaborate expression 1-2-3-4.

3.1 Types

In Hmong, a wide array of different kinds of constructions can be coordinated to form elaborate expressions. These range from N-N and N-V compounds whose properties are exactly those of ordinary nouns to a pronoun-N possessive construction that does not appear to exist except in elaborate expression.

⁶Of course, even a strict subcategorization restriction would probably still allow morphologically complex constituents, unless a means was devised to specifically exclude them.

3.1.1 Noun-Noun Compounds

A great many elaborate expressions consist of the coordination of two headed N-N compounds. If an element is repeated, it is usually the head (the right element). Syntactically, these elaborate expressions behave like bare nouns.

- (25) a. ntXHais + num + ntXHais + tswv → ntXHais-num-ntXHais-tswv
 ‘girl’ ‘official’ ‘girl’ ‘lord’ ‘princess’
- b. txiv + maab + txiv + ntoo → txiv-maab-txiv-ntoo
 ‘fruit’ ‘vine’ ‘fruit’ ‘tree’ ‘fruit’
- c. rig + tsuj + tsho + tsuj → rig-tsuj-tsho-tsuj
 ‘pants’ ‘silk’ + ‘shirt’ + ‘silk’ ‘silk clothing’
- d. hluas + nkauj + hluas + nraug →
 ‘youth’ ‘maiden’ ‘youth’ ‘bachelor’
 hluas-nkauj-hluas-nraug
 ‘young men and women; lovers’

3.1.2 Noun-Stativ Verb Compounds

Compounds consisting of a noun and stative verb are also common in elaborate expressions. The individual compounds behave as ordinary nouns, as do elaborate expressions composed of them.

- (26) a. kev + mob + kev + nkeeg → kev-mob-kev-nkeeg
 ‘way’ ‘sick’ ‘way’ ‘tired’ ‘sickness’
- b. sab + dlawb + sab + zoo → sab-dlawb-sab-zoo
 ‘liver’ ‘white’ ‘liver’ ‘good’ ‘noble-hearted’
- c. teb + taj + teb + tug → teb-taj-teb-tug
 ‘land’ ‘level’ ‘land’ ‘level’ ‘level land’

3.1.3 Transitive Verb-Noun Compounds

The Hmong lexicon contains many compounds⁷ headed by transitive verbs with nominal “direct objects” as their complements. With few exceptions, these compounds act as intransitive verbs. These types of compounds are extremely common in elaborate expressions. Elaborate expressions composed of this type of compound display the same syntactic behavior as individual compounds of this type.

- (27) a. kawm + txuj + kawm + ci → kawm-txuj-kawm-ci
 ‘learn’ ‘art’ ‘learn’ ‘art’ ‘to become educated’
- b. ndlav + dlaag + ndlav + zug → ndlavdlaagndlavzug
 ‘hire’ ‘strength’ ‘hire’ ‘strength’ ‘to hire labor’

⁷These are, of course, not prototypical compounds, and could be interpreted as lexicalized verb phrases, a possibility that will be discussed further below.

- c. ua + plaub + ua + ntug → ua-plaub-ua-ntug
 ‘do’ ‘trial’ ‘do’ ‘trial’ ‘to engage in legal proceedings’

3.1.4 Intransitive/Stative Verb-Noun Compounds

Somewhat more peculiar is a class of Hmong compounds consisting of an intransitive verb head and a nominal complement that specifies the place, manner, or undergoer of the event to which the verb head refers. Again, the syntactic properties of elaborate expressions composed of this type of are the same as those of individual compounds.

- (28) a. tawg + vaaj + tawg + tsev →
 ‘shatter’ ‘garden’ ‘shatter’ ‘house’
 tawg-vaaj-tawg-tsev
 ‘for one’s household to split; to divorce’
- b. tsw + qej + tsw + dlob →
 ‘smell’ ‘garlic’ ‘smell’ ‘onion’
 tsw-qej-tsw-dlob
 ‘to smell of garlic and onions; pungent or savory’
- c. mob + teg + mob + taws → mob-teg-mob-taws
 ‘hurt’ ‘arm/hand’ ‘hurt’ ‘leg/foot’ ‘to hurt in the limbs’

3.1.5 Verb-Stative Verb Resultative Compounds

The members of another class of Hmong compounds consist of two verbs. The verb on the left is the head, is transitive, and specifies the event to which the compound refers. The verb on the right is a modifier that specifies the result state of the verb. Both compounds of this type and elaborate expressions composed of two such expressions act as verbs.

- (29) a. noj + qaab + nyob + zoo → noj-qaab-nyob-zoo
 ‘eat’ ‘tasty’ ‘live’ ‘good’ ‘to be healthy’
- b. noj + taag + haus + taag → noj-taag-haus-taag
 ‘eat’ ‘finish’ ‘drink’ ‘finish’ ‘to finish eating and drinking’

3.1.6 Stative Verb-Stative Verb Compounds

There is a very small and exceptional class of elaborate expressions that consist of coordinated compounds of stative verbs with the head on the right and the modifier on the left:

- (30) a. nrawm + caj + nrawm + tuag → nrawm-caj-nrawm-tuag
 ‘fast’ ‘alive’ ‘fast’ ‘dead’ ‘very quickly’
- b. coob + caj + coob + tuag → coob-caj-coob-tuag
 ‘numerous’ ‘alive’ ‘numerous’ ‘dead’ ‘very quickly’

The modifier specifies the manner or intensity of the event described by the compound as a whole.

These compounds are significant in that they only occur within elaborate expressions, and are not licensed to occur as free words (unlike the other compounding constructions discussed thus far).

3.1.7 Verb-Expressive Collocations

It is also possible, in some cases, to create elaborate expressions by coordinating two verbs modified by an expressive or ideophone:

- (31) a. ua + dlog + ua + dlig → ua-dlog-ua-dlig
 ‘do’ ‘sloppily’ ‘do’ ‘sloppily’ ‘to do sloppily’

The relationship between expressives and the verbs they modify is generally seen as being syntactic in nature (Ratliff 1992:138). If this is true, and if elaborate expressions are words, then they must be words that accommodate syntactic structure internally.

3.1.8 Classifier-Noun Collocations

Classifier-noun collocations are another example of an apparently syntactic construction that occurs within elaborate expressions:

- (32) a. lub + koob + lub + npe → lub-koob-lub-npe
 ‘CLF’ ‘reputation’ ‘CLF’ ‘name’ ‘the glory’
 b. cov + nam + cov + txiv → cov-nam-cov-txiv
 ‘CLF’ ‘mother’ ‘CLF’ ‘father’ ‘the parents’
 c. raab + teb + rooj + ntuj → raab-teb-roob-ntug
 ‘CLF’ ‘land’ ‘CLF’ ‘heaven’ ‘world; country’

In Hmong and other East Asian languages, noun classifiers are never treated as affixes, but always as independent words. In Hmong, this analysis is supported by the fact that a closed class of adjectives can intervene between nouns and their classifiers, and the fact the classifiers can occur in a noun phrases that contains no noun root, provided there is a quantifier or demonstrative in the phrase. Furthermore, there is no tone-sandhi domain that includes both the classifier and the noun root in normal syntactic noun phrases. However, in elaborate expressions it is possible for noun classifiers to trigger tone sandhi in following nouns, as shown in example (32c) where the classifier *rooj* seems to be triggering a tone change in to following noun, *ntuj* ‘sky’.

3.1.9 Numeral/Quantifier-Classifier Collocations

From a phonological standpoint, quantifier-classifier and numeral-classifier constructions seem to be words, but from a morphological standpoint, their status is less certain. Whatever their actual nature may be, they also occur in elaborate expressions:

- (33) a. cuaj + lo + kaum + lo →
 ‘nine’ ‘CLF (word)’ ‘ten’ ‘CLF (word)’
 cuajloskaumlo
 ‘many words; however many words’
- b. cuaj + txub + kaum + txub → cuajtxubkaumtxub
 ‘nine’ ‘CLF (topic)’ ‘ten’ ‘CLF (topic)’ ‘all things; all topics’
- c. txhua + yaam + txhua + tsaav → txhuayaamtxhuatsaav
 ‘every’ ‘CLF (kind)’ ‘every’ ‘CLF (sort)’ ‘every kind of thing’

These collocations behave differently in elaborate expressions than when they occur independently. Independent quantifier-classifier and numeral-classifier constructions typically appear as part of a noun phrase containing a noun head (over which the quantifying expression has scope). In contrast, elaborate expressions consisting of numerals or quantifiers in collocation with classifiers act as complete noun phrases and cannot have scope over any nominal element outside of the expression.

3.1.10 Pronoun-Noun Possessive Constructions

A final construction that is peculiar to elaborate expressions are pronoun-noun possessive constructions:

- (34) a. luag + tuj + luag + taug → luag-tuj-luag-taug
 ‘other’s’ ‘poison’ ‘other’s’ ‘poison’ ‘other’s poison’
- b. peb + vaab + peb + tshaus → peb-vaab-peb-tshaus
 ‘1PL’ ‘tray’ ‘1PL’ ‘sifter’ ‘our sifting tray’
- c. koj + teb + koj + chaw → koj-teb-koj-chaw
 ‘2SG’ ‘land’ ‘2SG’ ‘place’ ‘your country’

Normally, Hmong nouns can only be alienably possessed if they are marked for specificity by the presence of a noun classifier.

- (35) a. peb lub vaabtshaus
 1PL CLF sifting tray
 ‘our sifting tray’
- b. *peb vaabtshaus
 1PL sifting tray
 Intended: ‘our sifting tray’

Without the classifier, such expressions are ungrammatical. There are relatively few nouns that can be possessed inalienably (*nam* ‘mother’, *txiv* ‘father’, *tsev* ‘home’, etc.) If these nouns are possessed inalienably, the possessor is simply juxtaposed with the noun phrase, with no additional marking:

- (36) a. Pov nam
 Pao mother
 ‘Pao’s mother’
- b. kuv tsev
 1SG house
 ‘my house’

The possessive construction that occurs in elaborate expressions is formally the same as the inalienable possession construction, but is not bound by the same semantic restrictions.

3.2 Semantic/Syntactic Properties

Elaborate expressions are very similar to coordinate compounds of the co-hyponym type in their semantic structure. The two constituents in these constructions seem to be constrained by the need to share most (or even all) of their semantic features. This fact has been noted by earlier investigators, especially Ratliff (1992:44), who notes that simple coordinate compounds are like a “miniature” version of the more complex elaborate expression. Both elaborate expressions and their simpler kin form “either a repetitive compound involving words with only slight meaning differences or a compound pair involving objects or people that belong together.”

Syntactically, the situation is more complex. Some of the constructions that occur as members of elaborate expressions do not occur outside of this larger construction, or appear to display different properties when coordinated as an elaborate expression than when isolated. In some cases, apparent differences in syntactic properties can be attributed to facts of constituency structure. For example, noun classifiers and numerals or quantifiers probably do not form a constituent, and any construction containing only these elements must still behave as if it has a noun head from which it is projected (even if there is no overt noun in the phrase). If this was the case, it would be entirely expected that two such phrases, when coordinated, would not be able to make specific or quantify another noun phrase. But other cases cannot be explained away so simply. Constituency structure provides little insight into the fact that pronominal possessors can be associated with bare nouns (nouns lacking a classifier) in elaborate expressions but not in normal syntactic contexts.

This set of phenomena suggest that the syntactic features of elaborate expressions are not simply those of their daughters. Rather, there is some function $\iota_k(x, y)$ that relates the syntactic features of the daughters (x and y) to those of the mother node. For most values of x and y , $\iota_k(x, y) = x = y$, but this is not always the case. If it were, then elaborate expressions composed from instances of the pronoun-noun possessive construction would be as unacceptable as free syntactic objects as their daughters. This type of stipulation was not necessary in my discussion of coordinate compounds, but it is in no way incompatible with the analysis I have given of them. In other words, my account of coordinate compounds can be extended to elaborate expressions without compromising the insights that motivated it.

3.3 Phonological Properties

From a phonological point of view, there are other reasons for treating elaborate expressions as a special case of coordinate compounds. Chief among these is the fact that elaborate expressions follow both the prosodic symmetry and tonal ordering generalizations that characterize simpler coordinate compounds. The halves of an elaborate expression are always metrically balanced. Likewise, the order of the halves of an elaborate expression are predictable based upon the same tonal hierarchy that was established to describe co-hyponym compounds. It is not necessary to modify this generalization in any way in order to predict the order of elaborate expressions: the locus of evaluation is the rightmost syllable in each constituent of the construction, regardless of the internal structure of the constituents.

The phonological evidence, like the semantic evidence, unites elaborate expressions not simply with coordinate compounds as a class, but with the specific co-hyponym type of coordinate compound.

4 Uniting the Analyses of Coordinate Compounds and Elaborate Expressions

The confluence of characteristics between coordinate compounds is so striking that any analysis of one of these phenomena that did not account for the principal details of the other would be missing an important and obvious generalization. But while it is clearly desirable to give a unified account of these phenomena, there are a few of significant complications that present themselves in the face of any such account. Resolving these complications appears to require the extension of the morphological framework in which I have presented my account.

4.1 Challenges to a Unified Analysis

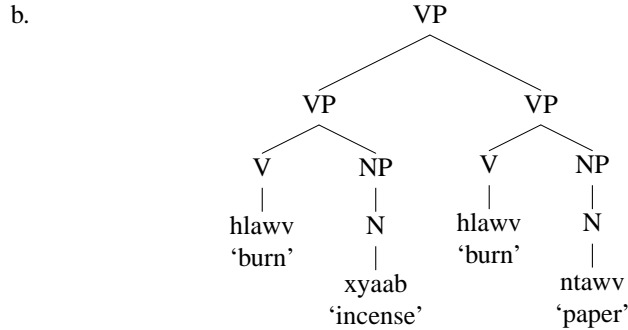
Thus far, in my description of elaborate expressions, I have brought little attention to the fact that their internal structure appears, in some cases, to be syntactic. This observation seems to be at odds with the assertion that coordinate compounds are morphological objects. On the other hand, I have treated the parallelism that exists between the halves of an elaborate expression as the simple result of semantic requirements, without drawing attention to the fact that the two halves of an elaborate expression can often be shown to contain each of the two (discontinuous) parts of simpler coordinate compound.

4.1.1 Syntactic structure within morphological objects

The members of some elaborate expressions are clearly morphological compounds—noun compounds and compounds of intransitive or stative verbs and nouns. These present no real challenge to our account. More problematic are elaborate expressions that coordinate what I have called transitive verb-noun compounds. These appear to

have the same structure and syntactic properties as ordinary verb phrases. One could easily impute to an example like (37a) a structure like that in (37b):

- (37) a. hlawv - xyaab - hlawv - ntawv
 burn incense burn paper
 ‘burn incense and paper’

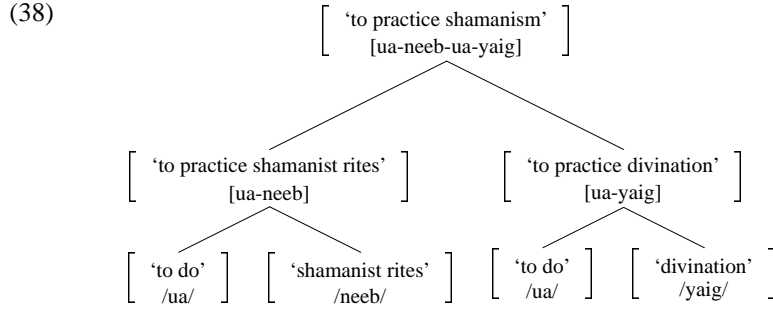


It would seem odd for a process akin to reduplication to target a constituent like a verb phrase, and seems especially odd to call the output of such a process a word. However, it is also the case that a structural configuration can superficially resemble a syntactic construction without actually being syntactic (as is the case with compounds in many languages). In cases like that of (37a), there is unfortunately little evidence for or against the proposition that the structure seen here is a verbal compound rather than a syntactic verb phrase. However, for other types of subconstituents, there is evidence that distinguishes the type of construction that occurs within an elaborate expression from superficially similar structures occurring in the syntax. For example, the noun-classifier collocations that occur in elaborate expressions are clearly based upon the syntactic structure of the noun phrase, where the classifier appears as a specifier for the noun phrase. We can tell that the construction that occurs in elaborate expressions is different, however, because within this construction, the classifier and noun are in the same tone sandhi domain—a situation that never seems to occur outside of elaborate expressions.

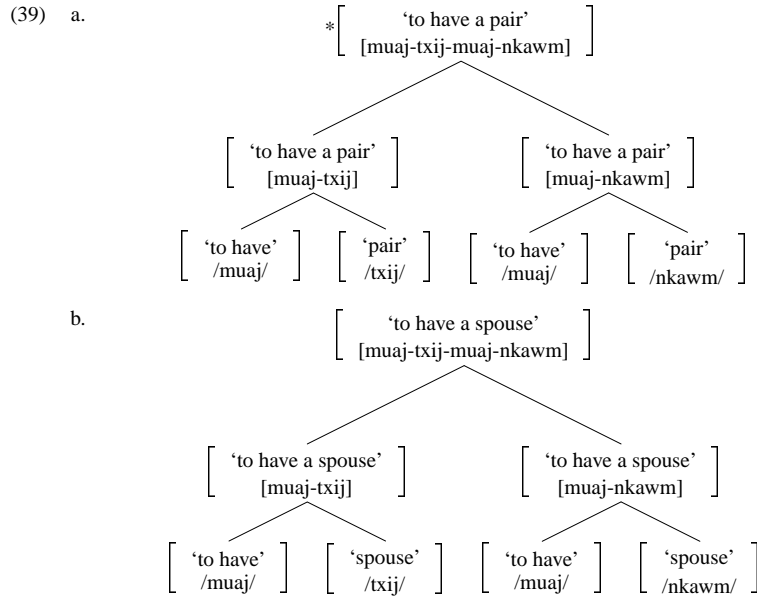
4.1.2 The “ionization” problem

A second and more interesting problem in the analysis elaborate expressions is the phenomenon that Matisoff (1973:82) calls “ionization”: many elaborate expressions seem to be derived from coordinate compounds by the intercalation of a reduplicated element. This means that the original coordinate compound is discontinuous in the output. For example, the elaborate expression *ua-nee-ua-yaig* ‘do-shamanist rites-do-divination’ (‘to practice shamanism’) is derived from *nee-yaig* ‘shamanist rites-divination’ (‘shamanist practices’). When the semantics of an elaborate expression are composition, this presents no special problem. The ordering of the elements, in either the simple or the complex expression, is predictable based upon the same tonal criterion. The principle by which the semantic features of *ua-nee* ‘practice shamanist

rites’ and *ua-yaig* ‘practice divination’ are intersected to yield ‘practice shamanism’ is the same as that which yields *neeb-yaig* ‘shamanist practices’ from *neeb* ‘shamanism’ and *yaig* ‘divination’:



However, this kind of lexical relationship is by no means restricted to coordinate compounds with compositional semantics. In fact, any coordinate compounds of the HYP-COCOMP type can be “ionized” to produce an elaborate expression. When the relationship between the meaning of a coordinate compound and the meaning of its parts is not compositional, then this type of an analysis no longer works. For example, the expression *txijnkawm* ‘spouse’ is derived from the parts *txij* ‘pair(?)’ and *nkawm* ‘pair’. Assuming a purely compositional analysis, the expression *muaj-txij-muaj-nkawm* ‘have-*txij*-have-*nkawm*’ should mean ‘to have a pair’ as in (39a), but instead it means ‘to have a spouse’ as in (39b).



This is only possible if, in this case, the semantics of *txij* and *nkawm* are constrained to be synonymous with or hyponymous to *txij-nkawm* ‘spouse’, rather than bearing the semantic features they would include as individual lexical items. In other words, these “words” are behaving as if they are discontinuous parts of the same word.

The data supporting this conclusion is quite strong. There are three principle sources for it:

1. The occurrence of bound forms in elaborate expressions, but only apposition to their counterparts from lexical coordinate compounds. The second element in the coordinate compound *txuj-ci* ‘arts’ does not occur independently, but does occur in such expressions as *kawm-txuj-kawm-ci* ‘learn-arts-learn-arts(?)’ (‘become educated’).
2. The appearance of monomorphemic forms as discontinuous elements in elaborate expressions (apparently as a result of their reanalysis as coordinate compounds). Thus, *phoojywg* ‘friend’, from Mandarin *pengyou* ‘friend’, which was borrowed as an unanalyzable unit can be “ionized” in an elaborate expression, yielding compounds like *ua-phooj-ua-ywg* ‘to be friends’.
3. The existence of what Martha Ratliff has called Sandhi Form Promotion (Ratliff 1992:77) in elaborate expressions (Johns and Strecker 1987). Forms in elaborate expressions often display the sandhi tone-allomorph even when they are separated from the trigger by an intervening element. For example, *tub* ‘son’ + *kiv* ‘daughter’ → *tubki* ‘children’. The sandhi form of *kiv* ‘daughter’ is still present in an elaborate expression like *tsegtubtsegki* ‘bereft-son-bereft-daughter’ even though the sandhi trigger is no longer contiguous to the sandhi target.

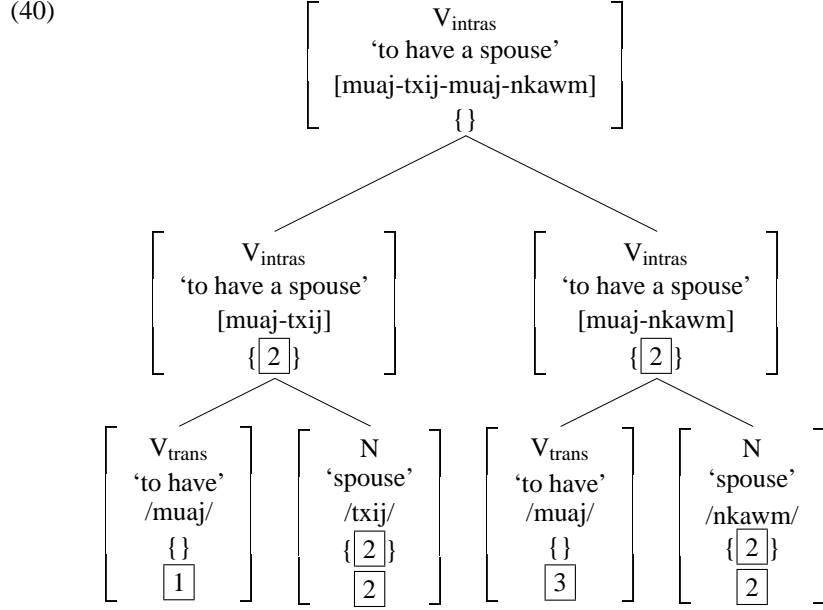
It does not appear that the Construction Morphology formalism employed in the analysis thus far has the ability to capture the relationship between the two parts of this discontinuous constituent and their immediate sisters simultaneously.

4.2 An Analysis Employing Coindexation

Capturing these relationships seems to require an additional formal device, preferably one that is independently motivated. Non-derivational syntactic theories such as GPSG and HPSG employ the formal device of indexes to express relationships between elements that are not directly derivable from constituent structure. It is possible to account for the three-way relationship between the lexical representation of the simpler coordinate compounds from which elaborate expressions may be derived, the discontinuous parts of the simpler coordinate compounds within the elaborate expression, and the hierarchical constituency structure of the elaborate expression as a whole. The mechanism employed here is not fully formalized. It is adapted from, but not formally identical to, the system of slash features introduced in GPSG (Gazdar 1981) and employed in modified form in HPSG (Pollard and Sag 1994).

Under this analysis, each terminal node is assigned an index, represented in (40) with “tags” (digits in boxes). Terminal nodes sharing the same index share the same semantic content and syntactic features (either through structure sharing or a constraint enforcing identity). Terminal nodes that do not represent a given lexical entry exhaustively have a feature corresponding to a particular index in their slash set, which is inherited by a parent node. The slash-set of each node is the complement of the union of the slash-sets of its daughters and the intersection of the slash-sets of its daughters. Such a morphological dependency is satisfied when the corresponding index is a

member of the intersection of the slash-sets of a coordinative construction of the appropriate type. An informal diagram of *muaj-txij-muaj-nkawm* ‘to have a spouse’ is given in (40):⁸



Here, the verb *muaj* [3] is able to take as its complement the most local part of [2], which from the perspective of *muaj* [3] is a normal noun meaning ‘spouse’. This semantic sleight-of-hand, for a piece of phonology that should mean ‘pair’, is enforced through its coindexation with its other half, setting up a constrained relationship between this sign and another sign in the lexicon, *txij-nkawm* ‘spouse’.

In order to keep this model from badly over-generating, it is necessary to stipulate a number of constraints, subject to revision upon further investigation of their implications. A provisional statement of these constraints is as follows:

(41) A lexical item α that is a subtype of Coordinate Compound may occur discontinuously as β_1 and β_2 in a morphological word just in case that:

1. The resulting dependencies are satisfied within the morphological word at a constructional node that is of the class Coordinate Compound, or one of its subclasses.
2. The phonological content of α is the concatenation of the phonology of β_1 and β_2 .
3. The semantic and syntactic features of β_1 and β_2 are identical to those of α .

⁸This should actually probably be modeled asymmetrically, as is the SLASH and TO-BIND features of HPSG, but at present I have no good arguments for the direction of the asymmetry. The current model is intended as a precursor to a full formalism, rather than a working grammar fragment.

While the adoption of this formal architecture would be undesirable if it can be shown that the current theoretical apparatus is adequate for dealing with these expressions, it seems that there is currently no way of accounting for this set of phenomena insightfully without a device of equivalent formal power.

5 Conclusion

Whatever the complications in analysis, an account of Hmong morphological parallelism that treats coordinate compounds and elaborate expressions as manifestations of the same fundamental phenomenon is more insightful than one that does not. It is clear that these two sets of compounds are more similar than different. The both share an interesting set of conditions on semantic, syntactic, and prosodic symmetry. They also share an interesting tonal condition on the ordering of constituents (the history of which needs to be investigated further). However, it seems that the cost of capturing these similarities is the extension of the theoretical framework of Construction Morphology that is currently growing in both conceptual coherence and theoretical influence. It is hoped that further research will reveal that other types of discontinuities and non-hierarchical relationships in morphology can be explained using the same set of principles.

References

- BISANG, WALTER 1993. "Classifiers, quantifiers, and class nouns in Hmong". *Studies in Language* 17(1):1–51.
- CHANG, KUN 1953. "On the tone system of the Miao–Yao languages". *Language* 29:374–378.
- GAZDAR, GERALD 1981. "Unbounded dependencies and coordinate structure". *Linguistic Inquiry* 12:155–184.
- HUDAK, THOMAS JOHN 1990. "Thai". In Bernard Comrie (ed.), "The World's Major Languages", New York: Oxford University Press, 757–776.
- INKELAS, SHARON AND CHERYL ZOLL 2003. "Morphological doubling theory". Ms. UC Berkely/MIT.
- JOHNS, BRENDA AND DAVID STRECKER 1987. "Lexical and phonological sources of Hmong elaborate expressions". *Linguistics of the Tibeto-Burman Area* 10(2):106–112.
- LI, CHARLES N. AND SANDRA A. THOMPSON 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.
- MALKIEL, YAKOV 1959. "Studies in irreversible binomials". *Lingua* 8(2):113–160.

- MATISOFF, JAMES A. 1973. *The Grammar of Lahu*. Berkeley: University of California Press.
- OLSEN, SUSAN 2001. "Copulative compounds: a closer look at the interface between syntax and morphology". In Geert Booij and Jaap van Marle (eds.), "Yearbook of Morphology 2001", Dordrecht: Kluwer Academic Publishers, 279–320.
- ORGUN, CEMIL ORHAN 1999. "Sign-based morphology: a declarative theory of phonology-morphology interleaving". In Ben Hermans and Marc van Oostendorp (eds.), "The derivational residue in phonological Optimality Theory", Amsterdam: John Benjamins, 247–267.
- ORGUN, ORHAN 1996. *Sign-based morphology and phonology: with special attention to Optimality Theory*. Ph.D. thesis, University of California, Berkeley.
- OURN, NOEURNG AND JOHN HAIMAN 2000. "Symmetrical compounds in Khmer". *Studies in Language* 24(3):483–514.
- POLLARD, CARL AND IVAN SAG 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CLSI/University of Chicago Press.
- 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.
- SADOCK, JERROLD M. 1998. "On the autonomy of compounding morphology". In Steve LaPointe, Diane K. Brentari, and Patrick M. Farrell (eds.), "Morphology and Its Relation to Phonology and Syntax", Stanford: CLSI Publications, 161–187.
- SINGH, RAJENDRA 1982. "On some 'redundant compounds' in Modern Hindi". *Lingua* 56:345–351.
- THOMPSON, LAURENCE C. 1965. *A Vietnamese Grammar*. Seattle: University of Washington Press.
- TING, PANG-HSIN 1975. "Tonal relationship between the two constituent of the coordinate construction in the Analects, the Meng-tze, and the Book of Odes". *Bulletin of the Institute of History and Philology, Academia Sinica* 47(1):17–52.
- WATSON, RICHARD L. 2001. "A comparison of some Southeast Asian ideophones with some African ideophones". In F. K. Erhard Voeltz and Christa Kilian-Hatz (eds.), "Ideophones", Number 44 in Typological Studies in Language. Amsterdam: John Benjamins, 386–405.
- WHEATLEY, JULIAN 1990. "Burmese". In Bernard Comrie (ed.), "The World's Major Languages", New York: Oxford University Press, 834–854.