Linguistic typology and Sinospheric languages

Elizabeth Riddle and Herbert Stahlke
Ball State University

1. Introduction

The well-known fact that Chinese and many Southeast Asian, or Sinospheric languages (Matisoff 1990:113), share a number of characteristics (Matisoff 1976; Matisoff 1986) is commonly attributed to contact influence. However, less well known to Southeast Asianists is the fact that some of the same properties cluster in a number of Kwa and Benue-Congo (Niger-Congo) languages spoken in West Africa as well. Foley and Olson (1985) note that these languages share the following properties with many languages of Southeast Asia: phonemic tone, a tendency toward monosyllabicity, isolating morphology, verb medial word order, and verb serialization. We have observed other shared properties as well. Since there is no question of contact influence between these two regions, some other explanation for this clustering must be sought unless it is to be considered merely coincidental.

In this paper we sketch a typology, involving what we call the Specification Scale, which offers a rationale for the clustering of similar morphological and syntactic properties of some representative languages from the two geographic areas, namely Mandarin Chinese, White Hmong, Thai, and Vietnamese from East and Southeast Asia, and Yoruba, Yatye, and Ewe from West Africa. The last three are Niger-Congo languages belonging to different subgroups of the Kwa (Ewe) and Benue-Congo (Yoruba and Yatye) subgroups within Niger-Congo (Bendor-Samuel 1989), with a time depth of differentiation probably similar to that of Indo-European.

2. Shared Properties

Matisoff (1986) provides a convenient listing of many Southeast Asian areal properties. Among these are tone,
relatively isolating morphology, verb concatenation, reduplication as an adverbializing process, and adjectives as a subclass of verbs. Tone, isolating morphology, and verb concatenation or serialization also occur in Foley and Olson's list of properties shared between Sinospheric and West African languages. They list SVO word order as a shared characteristic as well, but this does not show up as a clustering property when our typology is applied to other languages in either area. Lahu and Ijo, for example, are SOV and serializing, and Ewe shows some SOV traits. We will therefore ignore SVO word order although the seven languages we deal with here all have this as the basic main clause order. In addition to Matisoff's and Foley and Olson's lists, we have found that they also share the following properties: reduplication in general, compounding as a major word formation strategy, "underspecification" of some word classes (for example, words with adjectival and prepositional meanings frequently cannot be differentiated clearly from verbs by means of morphosyntactic criteria), and relatively restrictive phonotactic and morpheme structure constraints.

3. Rationale for clustering

Foley and Olson (1985:51) offer the following rationale for the clustering which they observed:

This cluster of properties is not accidental; they are all interrelated. Phonological attrition causes syncope of segments or syllables, with the result that phonemic tone or complex vowel systems develop to compensate for phonemic distinctions being lost. On the grammatical side, phonological attrition causes gradual loss of bound morphemes, which, being prefixes or suffixes, are most at risk to loss. With respect to noun phrases, this means loss of case marking, with the result that word order becomes rigid, in order to distinguish the semantic roles of noun phrases. Verb-medial order seems most favored in this function because the verb separates the actor and the undergoer and no ambiguity is possible. With the verb, phonological attrition gradually reduces verbal morphology. Now, valence increasing and decreasing are functions of verbal morphology (Comrie 1985). As this verbal morphology is lost, a
new device for valence adjustment must be found. Verb serialization begins to be used in this function, provided no new pattern is being created. Rather, the function of an already existing pattern is simply extended. In sum, then, the use of serial constructions to indicate valence increases is a function of the isolating structure of these languages.

We modify and extend this rationale to propose a general typology of language in which all the clustering properties find a natural explanation. Specifically, we propose that a useful dimension along which languages can be compared is the way and degree to which they overtly specify dependency relationships in surface structure. We suggest that there are two polar prototypical target structures to which languages adhere to varying degrees, overspecification and underspecification of relationships in surface structure. Overspecification in syntax involves a strong tendency to signal dependence by means of redundant grammaticized markers, such as, for example, the combination of prepositions and case in Classical Latin or gender and number concord in Bantu. In word formation, attachment of a dependent, bound form to a more dominant stem is very common, for example, noun morphology in Finnish. Languages characterized by underspecification tend to signal relations by word order and by the use of independent lexical items in both syntax and morphology, for example, the lexical expression of tense and aspect or the widespread use of body part nouns for temporal and spatial relations in Yoruba and many other West African languages.

4. Target Structure

The concept of target structure has appeared in various forms in linguistic theory but has been most clearly applied to syntax by Green (1974; 1980) and Haiman (1974), based on the notion "phonological conspiracy" discussed in Kisseberth (1970), among others. A target structure arises when several distinct rules "conspire" to produce the same result in surface structure. For example, Green (1980) claims that in English there are two target inversion structures derived from a large variety of syntactic constructions. Haiman argues that the verb second position in German main
clauses is a target structure. Riddle (1990) proposes parataxis as a target structure for White Hmong.

This notion has been applied primarily to individual phenomena within languages. We propose that overspecification and underspecification are two different extreme targets for languages as a whole, cutting across varied phenomena in syntax and morphology, in addition to being targets for individual aspects of grammar. These are prototypical targets to which languages and constructions conform in varying degrees. Thus some languages strongly overspecify relations in surface structure, some strongly underspecify them, and still others fall in between, depending in part on the degree to which semantic and pragmatic information is redundantly grammaticized.

5. Overspecification vs. Underspecification

Prototypical underspecification and overspecification provide the endpoints of a scale along which a number of seemingly disparate grammatical phenomena can be ranged and thus interrelated.

Prototypically underspecifying languages have juxtaposition and the inclusion of independent lexical items as their primary signals of semantic dependencies. There are fewer grammaticized indicators of particular semantic dependencies than in overspecifying languages. Examples are the parataxis of clauses, verb serialization, and the compounding of free morphemes. Linguistic elements are often not anchored to each other by overt tagging such as in case marking. Interclausal relations and argument structure tend to manifest themselves through linear order rather than through grammaticization.

Prototypically overspecifying languages emphasize the grammatical dependence of one element upon another, forming pairs (or groups) of dominant and dominated elements such as main vs. subordinate clauses and free vs. bound morphemes. Grammaticized markers overtly signal the semantic dependencies of surface elements to each other and explicitly tag certain elements as “belonging together” grammatically, as in case marking.
6. West African and Sinospheric languages

In applying the Specification Scale to the West African and Sinospheric languages under consideration here, we argue that these languages are characterized by relative underspecification in general, in comparison to languages such as English and Polish, which are relatively overspecifying. Our discussion will focus on White Hmong (henceforth Hmong) and Yoruba in order to give a general picture of how the properties cluster in single languages.

6.1 Clauses

Verb serialization or concatenation, i.e., the stringing together of verbs with or without accompanying arguments with no formal surface marking of subordination or coordination, is a hallmark of clause structure in both groups of languages. Just as Matisoff (1969) notes for Lahu, in these two groups of languages, concatenations of two or more verbs forming complex phrases by juxtaposition often serve in place of the formal subordination devices found in a language like English.

For example, in Hmong, serialization is extensive both in the functions it serves and in its frequency, and other forms of parataxis in which items such as separate clauses are juxtaposed with no overt signalling of subordination or coordination are very common (Riddle 1990). Hmong generally does not mark overt subordination. It does have two complementizers translatable as that, i.e., tias and haiś tias, but the latter is a compound including the former (Jaisser 1984), and they appear to be basically synonymous. They are derived historically from verbs of saying. Moreover, haiś is still used as an independent verb of saying and tias as a quotative (Clark 1989). (Li (1988) also describes tias in Green Hmong, a closely related dialect, as a quotative.) Thus, the only subordinate clause type functioning as a complement that is overtly marked as such appears very similar to a serial construction on the surface, the level relevant to the notion of target structure. Since Hmong has no inflection of any sort, there is also no finite-nonfinite distinction. The only other apparent type of clausal subordination is in relative clauses (Riddle 1989), which follow the head NP and may be indicated by a contextually optional invariant relative marker uas, as shown in (1):
1. Cov neeg (uas) haus cawv khee v
   GRP person that drink alcohol possible
   khee v muaj mob taub hau . . .
   possible have sick head . . .

   People who drink often have headaches.
   (adapted from (Community Health Care Center n.d.))

Adverbial clauses are often introduced not by subordinating conjunctions but rather by NPs which simply have a paratactic relationship to the rest of the sentence, as in (2):

2. Koj yuav tsum tsis txhob tsav lum fais
   2sg must not drive car
   thaum uas tseem noj cov tshuav no.
   time that still eat GRP medicine this

   You should not drive while taking this medicine.
   (Xiong 1980:20)

or take the form of existential clauses juxtaposed to the beginning of the sentence, as in (3):

3. Muaj ib tag kis nws tuaj.
   have one morning s/he come

   One morning s/he came.

Similar tendencies toward surface underspecification of subordination are found in the other Sinospheric languages under consideration here.

The West African languages we have examined are also similar. For example, Yoruba uses serial constructions for a large variety of functions and they are far more frequent than subordinate clauses. (See Stahlke (1970; 1974b) for a fuller discussion.) Although it does have an infinitive-like construction for desideratives, subordinate conditional clauses, a subordinate declarative clause structure, and quite common relativization, it has about one-third the number of subordinating conjunctions that English has, and as in Hmong, the subordinate declarative clause uses as its complementizer a verb meaning ‘say.’ Ewe and Yatye are
similar in their treatment of complementizers, including their use of a 'say' verb for subordinate declaratives. With regard to clause structure, then, these languages are somewhat less underspecifying than the Sinospheric languages but still predominantly underspecifying in comparison to languages such as English and Polish.

Compare these to Polish, a strongly overspecifying language. Polish has numerous types and markers of clausal subordination, including że 'that' for subject and object complements, several conditional complementizers, infinitives, participles, and inflected relative pronouns. There is no serialization, and every verb and every clause is clearly marked as to whether it is main or subordinate. In English we find a few constructions which appear to mimic serialization, such as "I want to go see Mary," where there is no marker of subordination or coordination for the verb see and no special dependent verb form, but this is not a characteristic pattern of the language as a whole, and in any case, an infinitive marker can be used before see as well, as in "I hope to go to see Mary," and is required in the past "I hoped to have gone to see Mary by now." Treating the English construction as an unmarked infinitive rather than as a serial construction conforms better to the structure of English.

An example of a somewhat different type is the polysynthetic Algonquian language Menomini (Teeter 1976), where the verb is central to the sentence, and every sentence has just one predicative phrase, with particular inflectional patterns showing other phrases to be non-predicative. Although this is different from the subordination found in many Indo-European languages, the fact that virtually all major components of a sentence are overtly signalled as dependent on one verb shows Menomini clause structure to tend towards overspecifying.

6.2 Lexicon, Morpheme Structure, and Phonotactics

Both the West African and the Sinospheric languages have a high percentage of monosyllabic words or compounds formed from monosyllabic words. For example, Hmong has no derivational affixation and Thai has only limited derivational prefixation in comparison to the affixation found in overspecifying languages. Thai also has many polysyllabic Indic loanwords, but monosyllabicity is still a general characteristic of native Thai and other Sinospheric
lexical items. Vietnamese and Chinese are also predominantly monosyllabic. All three West African languages have predominantly monosyllabic verb and noun roots, with strict constraints on vowel sequences and medial consonants in disyllabic roots. This contrasts with overspecifying languages such as English and Polish, which have a large percentage of polysyllabic and polymorphemic words in addition to monosyllabic words.

This correlation has also been observed by Foley (1986), who notes that many Papuan languages have a comparative shortage of verb stems and that they make up for this lack by the use of adjunct constructions and/or verb compounding or serialization. Stahlke (1974a) reports a similar shortage of verb roots in Yoruba, where phonotactic constraints and chance lexical gaps reduce the number of actual verb stems to about 330. This contrasts with the list of some 3300 distinct verb stems listed for English in The Longman Dictionary of Contemporary English. Such small inventories of verbs may be due to small phonemic inventories and/or highly restrictive phonotactic and morpheme structure constraints, as is generally the case in the West African examples.

Hawaiian, with its eight consonants and ten vowels (including both long and short) (Elbert and Pukui 1979) and very simple syllable structure, or Japanese with its CV structure, have no such impoverished supply of possible word forms in comparison to other languages because they also allow long words and polysyllabic morphemes. On the other hand, White Hmong, which tends not to allow polysyllabic morphemes, has a large number of consonant phonemes and tones and allows initial clusters, thus permitting about 3500 possible syllable forms. Because of the limits on word length, however, this syllable limit is very near to the the total number of morphemes the language allows without compounding. This number pales in comparison to English, which exhibits nearly that many forms for existing verbs alone. Hmong, Yoruba, Chinese, and Foley’s Papuan languages overcome their relative lack of verbs in part by means serialization and verb compounding. Languages with longer morphemes and richer phonotactics will tend to have more verbs and therefore be less likely to exhibit serialization and other prototypical traits of underspecification.

We consider derivational affixation to be related to overspecification in that affixes are attached to stems on
which they are dependent in an asymmetric relationship. In contrast, compounding is a form of underspecification since it typically strings together two free morphemes in a morphologically symmetrical relationship. Reduplication of words (as opposed to syllables or bound morphemes) is similar, consisting as it does of the linear repetition of an independent element, although again, the degree to which this is the case depends on the particular type of reduplication.

Strongly overspecifying languages as a group and individually may exhibit a range of word derivation types, but usually have affixation as a primary means of word formation. For example, Polish uses affixation almost exclusively. There is very little compounding, and some compounds have one part marked as bound or dependent on the other by means of a combining element. For example, in the Polish form wielobarwny 'multi-colored' the o marks wiel 'many' as a bound form. Reduplication does not occur at all in Polish. English, which is somewhat less overspecifying than Polish, uses both affixation and compounding productively, with reduplication restricted largely to a hodge-podge of affective and onomatopoetic words. Algonquian languages have affixation, compounding, and reduplication (Teeter 1976). Strongly underspecifying languages such as Hmong, on the other hand, tend to use little affixation, favoring compounding and reduplication. Yoruba has some limited prefixation as well as widespread compounding and reduplication. This makes Yoruba somewhat less underspecifying than Hmong but still on the underspecifying end of the scale.

6.3 Inflection

Overspecifying languages exhibit a range of mechanisms for indicating grammatical and semantic relationships, but inflectional morphology plays a major role. Markers such as noun and verb inflections explicitly symbolize the grammatical dependence of one item upon another. Strongly overspecifying languages exhibit a high degree of inflection. For example, Polish nouns show case, number and gender marking, and verbs are inflected for tense, person and number. Aspect is also indicated partly through inflection (and partly through verb stem and suppletion). The past participle is also inflected for gender. Algonquian languages are polysynthetic, showing a great degree of verbal inflection, with a weak development of case.
Verbs are marked for gender (animate/inanimate), number, obviation, and presence, and there are five verbal orders. Verbs also take suffixes to indicate local and instrumental meanings among others (Teeter 1976, citing Sapir).

English shows a smaller degree of specification in this regard, although it clearly leans towards overspecification. For example, it has verbal inflection, including special dependent verb forms, as well as plural marking on nouns. On the other hand, the case system has been lost except for some pronouns and possessives, and even in those forms there is evidence that the function of the morphology is shifting from marking grammatical case to indicating pragmatic functions (Stahlke 1984). There is also a periphrastic possessive (of). Major grammatical relations are indicated by word order and/or prepositions.

In contrast, Hmong is inflectionally a strongly underspecifying language since it has no inflection of any sort, i.e., no case, number, or gender marking on nouns and no verbal inflection such as agreement, tense or aspect. Rather, aspect and future tense are indicated by separate lexical items, as are other time references. If it is contextually necessary to indicate number, a quantifier plus a noun classifier must be used. Word order alone indicates major grammatical relations. Hmong has a number of prepositions, but several of these are homophones of verbs with related meanings and thus mimic serial structure. Chinese, Vietnamese, and Thai have similar properties.

The West African languages are relatively less underspecifying than the Sinospheric languages but still more so than English and Polish. Although Yoruba, Yatye, and Ewe have some pronominal affixation on verbs, it is fairly limited, and they have little or no number marking on nouns and far less verbal inflection than typical overspecifying languages. Also, they use serial constructions to express many of the semantic relations indicated by case marking and adpositions in other languages. An example from Yoruba is given in (4),

4. Olú bá mi mú ìwé wá sí
   Olu help me pick up book come to

   ilé fún àwọn ọmọ
   home give they child
Olu brought me home a book for the children.

where a verb meaning ‘give’ is used in Yoruba instead of a preposition as in English.

6.4 Underspecification of word classes

As noted earlier, the inclusion of adjectives as a subclass of verbs is an areal feature of Sinospheric languages. This is true of the West African languages as well. For all three West African languages considered it is difficult to define a separate category of adjective. Many other Kwa and Benue-Congo languages exhibit at best highly limited, closed classes of adjectives. We claim that this is related to another characteristic shared by the Sinospheric and West African languages, namely that in many cases it is difficult to draw a clear-cut line between verbs and prepositions. Clark (1989), for example, states that both Southeast Asian and West African languages have locus prepositions synchronically derived from verbs (see also Schiller this volume). We cite Yoruba as an extreme case, with not a single unequivocal example of a preposition. Words with some verbal properties or clear-cut verbs are used instead, as fún ‘give’ in (4). Another example is the word sí, which in some contexts such as in (4), seems to have a general spatial prepositional sense, and in others functions as an existential verb in negative sentences, as in (5):

5. Kò sí wàhálà.
   not be trouble

There is no trouble.

sí thus has a use as a first verb and has positional limitations on its use in non-first verb positions. The only other word in Yoruba that might be analyzed as a preposition is ní, which is semantically empty in non-first verb position in verb series, its role being determined by the meaning of its argument. ní is also homophonous with a stative verb meaning ‘have.’ The other West African languages are also very sparse in their inventory of adpositions. Similarly, the Sinospheric languages under consideration here display a fuzzy boundary between verbs and prepositions, viz. the ‘coverb’ category in Chinese.
In some languages, there is also a somewhat fuzzy boundary between complementizers and verbs. As noted earlier, the Hmong complementizers are transparently related to verbs of saying. In Yoruba as well, one of the two complementizers, i.e. kpe, which is used like *that* in English, can be used alone as a verb of saying or as a complementizer after a verb of saying. The desiderative complementizer ki, which normally must follow a verb, may follow kpe when the latter is preceded by a verb of saying.

This underspecification of such word classes as adjective and adposition is another characteristic of underspecifying as opposed to overspecifying languages in general.

6.5 Underspecification of NPs

Finally, underspecifying languages seem to have less strict constraints on the unique anchoring of referential NPs to individual verbs than do overspecifying languages. This follows from the fact that formal surface marking of grammatical relationships is not the norm. For example, underspecifying languages such as Hmong and Yoruba often allow a single NP to appear as the surface syntactic argument of two verbs simultaneously, as in (6) and (7):

6. Nws nyeem ntwv rau kuv niam
   s/he read book to my mother
   nloog.
   listen

   S/he's reading to my mother.
   (Strecker and Vang 1986:14)

7. Olú ti qomo náà şubu
   Olu pushed child the fall

   Olu pushed the child and the child fell.
   (Bamgboşe 1974:23)

   In (6), kuv niam ‘my mother’ functions both as the object of rau ‘to’ and as the subject of nloog ‘listen.’ (7) is a serial construction where the surface object of the first verb is also the surface subject of the second verb. It can also be translated as “Olu pushed the child down.”
In contrast, overspecifying languages tend to show explicit evidence of clause merging or Raising (or its theoretical equivalent). That is, they do not tend to string items together without signalling a clause affiliation of some sort.

7. Summary

We have proposed that languages can be distinguished according to their manner and degree of overt specification of relations in surface structure. Both the West African and the Sinospheric languages considered here tend toward relative underspecification. This can be seen as a target to which many different aspects of structure relate.

This discussion recalls the distinction made by Tsao (1978) and Huang (1984) between discourse-oriented and sentence-oriented languages. They offer Chinese as an example of the former and English of the latter. Similarly, Li, Harriehausen and Litton (1986) claim that Green Hmong and isolating languages in general are more iconic than languages with many grammatical devices such as Finnish or Navajo. Thus they consider isolating languages to emphasize pragmatics over grammar. It is important to note, as Matisoff (1969) states for Lahu, that this does not mean that isolating languages are *supra grammaticam* as once claimed by H. A. Giles for Chinese. Matisoff shows that there are many complex syntactic and semantic constraints on the concatenation of verbs in Lahu. The target structure analysis proposed here mediates these apparently opposing points of view by focusing on the absence of formal grammatical marking of particular types of relationships in surface structure alone.

8. Conclusion

Given that many areal features of Sinospheric languages also cluster in some West African languages, with which there is no question of contact, we suggest that typological pressure as sketched here contributed to the spread of a number of Sinospheric areal features. In other words, this clustering of properties is not purely an areal phenomenon resulting from language contact. Rather, there is a principled basis on which many of these characteristics cluster.
References


by Joseph Drogo, Veena Mishra, and David Testen, 358--64. Chicago: Chicago Linguistic Society.


