

Verb Concatenation in Hmong Njua

A syntactic description and its treatment in natural language processing

Dr. Bettina Harrichausen-Mühlbauer
Scientific Center
IBM Germany
Institute for Knowledge Based Systems
Wilckensstr. 1a
6900 Heidelberg
Germany
e-mail: HARRIEHA at DIIDIBM1.BITNET
phone: +49 (0)6221-404216

Abstract

Hmong Njua, a language mainly spoken in the highlands of Laos, is a language without inflection, case and agreement system, tense, number and gender markings. It uses neither derivational nor inflectional morphology. Its strategy of forming sentences is, by and large, similar to that of many other East Asian languages of the isolating type: sentences are formed by concatenating nouns and verbs, and grammatical words seldom occur.

A linguist studying a language such as Hmong Njua is, thus, faced with the challenge of describing its multi-verb-constructions: What are the semantic relations between verbs? What are the constituent structures, or the lack of it, in those constructions? Do those concatenated verbs constitute one proposition or a complex or compound structure?

In this paper, the phenomenon of verb-concatenation in Hmong Njua is described in detail.

Furthermore, the PLNLP natural language processing system is described, demonstrating how verb-concatenations can be processed computationally, including the possibility of detecting syntactic errors, such as the incorrect verb-ordering in a concatenated verbal structure.

Verb-Concatenation

Verb-concatenation, a well-known and well-distributed feature among Southeast-Asian languages, describes also a characteristic phenomenon in Hmong Njua.

When speaking of verb-concatenation, several similar terms are used in the literature, which might lead to confusion: *multi-verb construction*, *verb-concatenation*, and *serial verbs*. Although all these structures share many features, a clear distinction is necessary:

- *Multi-verb constructions* contain more than one verb in a sentential structure. Two types of multi-verb constructions can be distinguished: those with one predication, i.e. sentences of the form SVO or SOV, and those with more than one predication, i.e. coordinated structures.
- By *verb-concatenation*, we understand the phenomenon of combining verbs in a sequence without interruption of other word-classes:

$$V + V + V + V + V^1$$

- *Verb-series*, or serial verbs, describe the linear combination of verbs with possible interruptions of the sequence:

$$VI + OBJ1 + V2 + OBJ2$$

According to these definitions, this paper will describe verb-concatenation of multi-verb constructions with one predicate.

While analyzing the phenomenon of verb-concatenation, certain questions will arise and will need to be answered:

1. Which elements or morphemes are modified or modify other elements?
2. Which elements are words of a word-class and which are particles?
3. Which constructions are lexical compounds, i.e. have to be analyzed morphologically, and which are syntactic constructions?

In order to explain this phenomenon, a differentiation into the following 3 main categories or types of verbs is necessary:

1. Verbs of motion (MOTION)
2. Verbs of direction (DIR)
3. Verbs of deictic nature (DEICT)

¹ According to Matisoff (1973) this is a structure mainly found in SOV-languages.

Verbclasses

Verbs of motion describe a certain motion by focussing on the kind of motion.

dhla	to run/jump
yaa	to fly
dlaum	to climb
nkaag	to crawl
moog	to go

Fig.1: Examples for verbs of motion

Verbs of direction refer to the direction of a motion, like 'up', 'down', 'into', or 'out of', without any reference to the kind of motion. The meaning of these verbs change according to their use as full verbs, i.e. when they occur alone, or as verbs of direction, i.e. when used as complements to the main verb in a concatenated structure.

	full-verb	concatenation
	-----	-----
nce	to climb, ascend	up
ncig	to circle	around
hlaa	to cross	across
nqeg	to descend	down
nkaag	to enter	into

Fig.2: Examples for verbs of direction

The third class, that of *deictic verbs*, can be regarded as a subclass of either verbs of motion or verbs of direction, as their meaning implies **motion** as well as **direction of a motion** with reference to the speaker.

But as this third class, which only consists of the 2 verbs,

lug	to come / here
moog	to go / away

Fig.3: The closed class of deictic verbs

carries certain very special features, it gets treated as a separate class. Again, as with verbs of direction, these verbs can be used alone (then they carry their full-verb meaning) or in a concatenated structure, where they carry a meaning which can be compared with the English adverbs **away** and **here**.

Deictic verbs in Hmong Njua and separable prefixes in German

While analyzing and describing syntactic structures and features of a language, we often compare these structures with similar phenomena in other, usually linguistically related, languages.

During the analysis of Hmong Njua, many features were found that Hmong Njua shares with other Southeast-Asian languages, such as Thai, Burmese, Lahu, or Chinese. Surprisingly, an interesting parallel could be drawn to "deictic particles" in German, with one major difference, that the German deictic particles have already lost the possibility of occurring alone, i.e. they can only be used in combination with a main verb as separable prefixes. Just as in Hmong Njua, these prefixes point to the direction of a motion with reference to the speaker:

herkommen (to come here)	: Er kommt her. (He comes here.)
	MOTION DEICT
herüberkommen (to come over here)	: Er kommt her- über. (He comes over here.)
	MOTION DEICT DIR
hingehen (to go there/away)	: Er geht hin. (He goes away.)
	MOTION DEICT
hinübergehen (to go away/across)	: Er geht hin- über. (He goes away/across).
	MOTION DEICT DIR

Fig.4: German separable prefixes with deictic implication

Of course this does not imply that Hmong Njua and German are linguistically or typologically related, but such comparisons and parallels can be used as helpful "devices" in natural language processing, as parallel structures can be treated identically by the rules.

Multi-verb constructions with a single predicate

Multi-verb constructions with a single predicate allow 3 different interpretations:

1. They form a compound, i.e. 2 or more verbs form a semantic and syntactic unit, which cannot be separated by other words or particles.
2. According to Li/Thompson (1981), they can function as "serials":
"A construction that refers to a sentence that contains two or more verb phrases or clauses juxtaposed without any marker indicating what the relationship is between