MANS HULDEN
SHANNON T. BISCHOFF
University of Arizona
Tucson, Arizona, USA
mhulden@email.arizona.edu
bischoff@email.arizona.edu

Concatenative Frames: toward a compact description of verbal morphology in polysynthetic languages

Abstract

A significant challenge in polysynthetic verbal morphology is to provide a complete and concise morphotactic description so that the combinatorics of various morphemes in a language is properly constrained to attested surface forms only. This paper briefly presents a descriptive methodology focusing on core Navajo morphology. Building on—and departing from—the template-based view of Young and Morgan (1987) and Faltz (1998), we propose that the Navajo verb can instead be concisely captured by a mechanism of simple concatenation together with a sparse set of feature-unification operations.

1 Background

A pure description of well-formedness on the morphological level in Navajo will have to account for a wide range of phenomena: metathesis of morphemes, long-distance agreement of morphemes, cross-serial agreement patterns of morphemes, a wide array of morphologically conditioned phonological processes, etc. Because of this apparent complexity, some scholars have gone so far as to propose that the Navajo verb complex be described in terms of syntactic phenomena (see e.g. Hale (2000)) as morphological processes commonly attributed to other languages do not seem to find direct counterparts in Navajo.

In the course of designing a computational parser for Navajo morphology (Hulden and Bischoff, in prep.), we have been prompted to perform a detailed analysis of available Navajo grammars. The current grammars for the language do—in the general case—successfully describe the legal patterns of word-formation. However, they do so by sacrificing simplicity in formalism and description. Even putting aside semantic considerations, the task of separating the "possible" Navajo verbs from the "impossible" ones seems to produce large volumes of complicated descriptive observations.

2 The Navajo verb

The morphotactics of the Navajo verb are overwhelmingly described in the literature in terms of a slot-and-filler template. The exact function and position of the different templatic elements vary, but in general most scholars decompose the verb into roughly a dozen components (see Figure 1 for a slightly abbreviated template).[1] The most important element in the verb is the last element, the stem. Meaning, however, is not determined by the stem alone, but by a combination of the stem and lexical and inflectional prefixes occupying certain slots in the template. The stem is nearly without exception monosyllabic, and there exist only an estimated 500 or so unique stems in the language (Willie, p.c.).

outer prefixes	plural	object	inner prefix	subject	classifier	stem
disjunct prefixes		*conjunct prefixes*				

Figure 1: *A somewhat simplified Navajo verb template, after Faltz (1998).*

(1) hajooltáál	"One dashed out of something"			
ha	J	oo	l	táál
lexical prefix	subject marker	subject marker	classifier	"run"

Figure 2: *A simple example of a Navajo verb. Noteworthy is the presence of two subject markers, **j** and **oo**, which together signal a 4th person subject. The j is actually positioned in the object "slot", whereas **oo** is found in the subject "slot."*

3 Templates as concatenation

Many of the complexities of template-based and other descriptions of Navajo verbal patterns seem to arise from the fact that, in the various paradigms, the notion of morpheme order is not isolated from the notion of co-occurrence constraints, yielding complex interaction between the two. To avoid the problems of simultaneously

[1] The finiteness of this template is part of our reason to contend that the Navajo verb complex is a matter of morphology, and not syntax.

describing order and co-occurrence, we propose—instead of a template—that a Navajo verb is simply a concatenation of elements taken from n sublexicons.

To be more precise, we say that the concatenative process is an ordered collection of sets $\omega = <S_1, S_2, ..., S_n>$ which must be consulted consecutively in such a way that a completely constructed morphological object consists of morphemes $<M_1, M_2,..., M_n>$ where M_1 is in S_1, M_2 is in S_2, etc. Thus, ω could informally be considered an ordered collection of sublexicons. We call this a "concatenative frame."

In reference to the template in figure 1, one can say that under such a description, S_1 consists of whatever morphemes are seen to be positionable in the slot numbered 1. So far, this concatenative description does not depart from the templatic one. However, we do not label these sets in any way, and defer the rest of the word-formation process to the unification of features connected to the various elements (or morphemes) in the sets.

4 Features

The sets $<S_1 ... S_n>$ in the concatenative frame all consist of a phonological component PHON plus a bundle of features in the format OP [FEAT VALUE], where the OP is an operator defining the mode of unification, FEAT a label for the feature in question, and VALUE the value of the feature. VALUE may also be omitted, which is really a shorthand for the idea that we do not care what the value of the feature is. The usefulness of this will become clear as the operations on feature bundles are defined.

To control for co-occurrence, we find it necessary to expand on the standard processes of "feature unification" found in the syntactic literature (Kay (1976), Shieber (1986)). Our basic operation is indeed the U operator that unifies OP [FEAT VALUE] combinations. In addition to this, we define two new operators: + for coercion, and - for exclusion. The operator + simply requires that there exist some feature [FEAT], or feature-value [FEAT VALUE] combination in the current word. Call M the morphological object produced by the concatenation process, then a morpheme carrying the feature +OP [FEAT VALUE] is permitted if and only if there exists some other feature with the same [FEAT VALUE] combination in M. We define - symmetrically to reject all such features in the word at hand. It should be noted that - is really a shorthand, the effect of which could be duplicated by the operator U and by inserting "dummy" features. However, this is not the case for +.[2]

Also, feature-value pairs can take advantage of the logical connective v to signal optionality among a group of features. This is simply a method to avoid redundant elements in the morpheme sublexicon, as can be seen in the example in figure 3. An entry in S_n may also contain a null PHON element (ø) with features associated to it.

Thus, what we have is 1) a concatenation of elements chosen from an ordered collection of sets S where 2) the individual elements consist of a phonological component PHON, together with a feature-matrix which operates with a sparse set of meta-features that specify the manner of unification. The result of this process encodes the morphotactics of the language, and can be assumed to be subsequently transported to a phonological layer that dictates the various sound changes that occur in the language. This phonological layer would operate independently of the morphotactic description.

The subject morpheme, **oo**, in figure 2 would, for example, have an entry in some S_N in ω that consists of:

	[PHON oo]
U	[SUBJ 3] v +[SUBJ 4]
U	[MODE pY]
U	[CL l v d]
+	[CONJUNCT]
-	[DISJUNCT]

where each of the features acts to restrict co-occurrence of the morphemes in the various ways described above.

The complete derivation of the example word **hajooltáál** in figure 2 can now be diagrammed as:

outer prefix(S_1)	**object (S_3)**	**subject(S_5)**	**classifier(S_6)**	**stem(S_7)**
ha	**j**	**oo**	**l**	**táál**
	U[SUBJ 4]	U[SUBJ 3] v +[SUBJ 4]		
+[MODE pY]		U[MODE pY]		U[MODE pY]
		U[CL l v d]	U[CL l]	
	U[CONJUNCT]	+[CONJUNCT]		
U[DISJUNCT]				

Figure 3: *An example of the interplay of features and the extended unification operations proposed. Sublexicons*

[2] We also do not take advantage of so-called "feature structures" sometimes employed in syntactic theories.

that produce ø phonological elements (i.e. slots that are not filled) are here omitted.

Here, the features U[SUBJ 4] associated with j, and U[SUBJ 3] v +[SUBJ 4] associated with **oo** license the combination **j+oo**. Noteworthy is that, for instance, **j** (signaling a 4th person subject) cannot surface by itself—it must be accompanied with what is traditionally analyzed as a 3rd person subject in a later "slot" in the template. We capture this by the feature +[SUBJ 4] associated with **oo** to the same effect, i.e. **oo** can only surface if a) there is no morpheme where SUBJ does no have the value 3 or b), there exists some morpheme in the derivation carrying [SUBJ 4]. Likewise, **oo** requires the presence of a feature [CONJUNCT], satisfied by **j**, and likewise disallows any [DISJUNCT] feature. Further, **oo** cannot surface if the stem, for instance, carries a conflicting [MODE] feature.

The above example in figure 3 contains only a fragment of the information associated with each prefix. However, it can be seen that this simple calculus is powerful enough to specify many long-distance dependencies in word formation, a common circumstance in Navajo. For instance, the very first lexical prefix, **ha,** in effect dictates that other mode-specific prefixes, if any, match it. In fact, **ha**'s features do more than this, and a closer analysis reveals that the dependency does not actually run left-to-right as might be expected. It is the prefix **oo** and the stem **táál** that each, individually, licence **ha**. The prefix **ha** cannot surface without one of the other two prefixes specified for [MODE], whereas **oo** and **táál** could surface without **ha**.

This long-distance relationship between a prefix and a stem is strong morphotactically, and the result is not surprising given that the lexical prefixes (which tend to appear first in the verb) together with the stem (which appears last) define the meaning.

Other effects, such as metathesis, can also be naturally captured with the formalism: let us suppose that a pair of morphemes M1 and M2 undergo metathesis in certain circumstances, triggered by the presence of some morpheme M3. If M3 is absent, metathesis does not occur. Let us further assume that, under normal circumstances, M1 < M2 in the ordered concatenation, where both would carry a feature disallowing the presence of M3, call it -[MT]. The following example derivation would, for instance, fail.

S_x	...	S_y	...	S_z
M3	...	M1	...	M2
U[MT]	...	-[MT]	...	-[MT]

Here, M1 cannot be drawn from Sy, nor M2 from Sz because they both carry –[MT] in their feature matrix, whereas M3, the metathesis-triggering morpheme carries U[MT], and is present. What would unify, however, is the following:

S_x	...	S_y	...	S_z
M3	...	M2	...	M1
U[MT]	...	+[MT]	...	+[MT]

That is, all we have to specify is that M1 and M2 are both present in the sublexicons S_y and S_z respectively, but with a different feature-matrix in each to achieve the effect. Specifying that this occurs at a long distance is no more cumbersome than any local metathesis, as can be seen from the formalism.

5 Discussion

The advantages in such a methodology are considerable. Significantly, the proposed notational formalism allows for a succinct description of Navajo verbal morphology in an exhaustive and accessible manner.[3]

Thus, it provides a simplified account for discussing and analyzing Navajo morphology. Further, this approach can be directly extended to morphological phenomena in other polysynthetic languages where morphotactic co-occurrence constraints and allomorphy are difficult to capture concisely with existing formalisms. As the goals of this design are to simply capture the given morphological facts of a particular language, in this case Navajo, it does not run into the problems that similar (in general terms) theoretical frameworks such as LFG and HPSG Kaplan and Bresnan (1982); Sag et al. (1999) encounter: this model is not constrained by the need to postulate any abstract accounts of language universals.

However, the model does serve to illustrate some rather interesting cross-linguistic phenomena. Notably, morphotactic elements are strictly regulated by the grammar, and morphological processes are isolated from the phonology in a natural way. It seems likely that such a model could be extended beyond the class of languages

[3] It must be noted that the proposal here could be interpreted in the light of ideas present in earlier works such as Kay (1976), Shieber (1986). The co-occurrence constraints have been formulated slightly differently in Functional Unification Grammars using an atom none to signal what we call -. The ideas of Lexical-Functional Grammars (Kaplan and Bresnan, 1982) contain a counterpart in what is called coherence conditions. Our operator + however, cannot be expressed in terms of U.

we have considered. Tackling issues such as morphological strata dealt with in lexical phonology and morphology (LPM) (Mohanan, 1986), for instance, reduces in the current model to the use of one feature that controls which “stratum” morphemes can be drawn from.

Also, the rigor of the proposed formalism has the additional advantage that it readily translates to possible computational formalization using standard methods of computational phonology and morphology along the lines of e.g. Koskenniemi (1983); Beesley and Karttunen (2003).

References

Beesley, K. and Karttunen, L. (2003). *Finite-State Morphology*. CSLI, Stanford.

Faltz, L. M. (1998). *The Navajo Verb*. University of New Mexico Press.

Hale, K. (2000). Remarks on the syntax of the Navajo verb. *Diné Bizaad Naalkaah: Navajo Language Investigations: Working Papers on Endangered and less familiar languages*, 3:55—96.

Kaplan, R. M. and Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In Bresnan, J., editor, *The Mental Representation of Grammatical Relations*, pages 173-281. MIT Press, Cambridge, MA.

Kay, M. (1976). Functional grammar. *Proceedings of the 5th meeting of the Berkeley Linguistics Society*.

Koskenniemi, K. (1983). *Two-level morphology: A general computational model for word-form recognition and production*. Publication 11, University of Helsinki, Department of General Linguistics, Helsinki.

Mohanan, K. P. 1986.*The theory of lexical phonology.* Dordrecht, Holland: Reidel.

Sag, I., Wasow, T., and Bender, E. (1999). *Syntactic Theory. A Formal Introduction*. CSLI.

Shieber, S. M. (1986). *An introduction to unification-based approaches to grammar*. CSLI.

Young, R. and Morgan, W. (1987). *The Navajo Language: a Grammar and Colloquial Dictionary*. University of New Mexico Press, Albuquerque.