Deseret Language and Linguistic Society Symposium

Volume 13 | Issue 1 Article 12

3-27-1987

Morpho-LFG

Dan W. Higinbotham

Morpho-LFG

Dan W. Higinbotham, Executive Communication Systems

Lexical-Functional Grammar (LFG), a syntactic theory recently developed by Joan Bresnan and Ronald Kaplan, is an attempt to model the syntactic processes at work in the minds of human beings as they code and decode sentences. The theory as described in The Mental Representation of Grammatical Relations (Bresnan 1982), assumes that lexical processes have occurred at the inclusion of each word in the lexicon to enter it in all of its forms with the appropriate functional features.

A lexicon including all word forms is plausible enough for English, since it is morphologically relatively simple. But for a language like Finnish, in which a verb may have tens of thousands of forms, it is hardly conceivable that all forms of every word could be stored separately.

This paper suggests treating morphemes functionally in the same way that words are treated in LFG, but with an additional unificational mechanism at morpheme boundaries to handle strictly local constraints. This simple addition to the theory allows morphological processes from a variety of languages to be described in a general way.

Word Formation Rules

Functionally, word formation involves the combination of elements chosen from specific morpheme classes in a definite order. This is similar to phrase formation in syntax, where elements chosen from specific grammatical classes combine in some order. In fact, simplified rules for Japanese verb formation could include

```
(1) V => VSTEM VAF        (AUX)                     (AUX)
                  (↑ ASPECT)==DESIRE  (↑ NEGATIVE)==PLUS

    AUX => VSUFFIX VAF
```

The '=>' notation, rather than the usual '->', implies that the rule is for word internal syntax. For example, the word 'isogitakunai' means "does not want to hurry":

```
iso   gi  ta   ku  na  i
hurry VAF want VAF not VAF
```

VAF is a class of suffixes whose forms depend both on what immediately precede them and on what immediately follow them. The usual functional features of LFG alone are not enough to constrain the affixes which are allowed to occur. The desirative morpheme requires the preceding VAF to be of the 'renyokei' form. The negative morpheme requires the preceding VAF to be of the 'mizenkei' form. If each morpheme adds a feature to the functional structure to reflect these requirements, (such as BASE-FORM RENYOKEI and BASE-FORM MIZENKEI, respectively), the differing values of the features will prevent unification, causing the rule to fail. In this particular case, these clashing features are not really functional in the usual LFG sense; they really just apply locally at morpheme boundaries. If a certain value of the feature is found in the AUX, and a matching feature is found in the preceding VAF, the sequence is allowed.

I therefore propose that two new metavariables, ◄ and ►, be added to the usual ↑ and ↓ metavariables of LFG. The ◄ metavariable can refer to a set of morphological features which must match those of the preceding morpheme. The ► metavariable can refer to a set of morphological features which must match those of the following morpheme. By 'matching', we mean that at any morpheme boundary, the ► features of the first morpheme must be able to unify with the ◄ features of the second morpheme. For the Japanese example above, we might use a set of morphological features as follows:

Feature	Values	
STEM-TYPE	GODAN	consonant-stem verbs
	ICHIDAN	vowel-stem verbs
	IRREGULAR	irregular verbs
	TRUE-ADJ	
	QUASI-ADJ	
CONSONANT	K,S,T,N,M,R,K,G	sub-consonant-types
BASE-FORM	MIZENKEI	a- form
	RENYOKEI	i- form
	SHUSHIKEI	u-form
	KATEIKEI	e-form
	TERENYOKEI	te-form
	KAKO	ta-form

Consider the following lexical entries:

```
'iso':: VSTEM, (↑ PRED) = 'HURRY<(↑ SUBJ)>'
               (► STEM-TYPE) = GODAN
               (► CONSONANT) = G

'gi':: VAF, (◄ STEM-TYPE) = GODAN
            (◄ CONSONANT) = G
            (► BASE-FORM) = RENYOKEI
```

```
'ga':: VAF, (◄ STEM-TYPE) = GODAN
            (◄ CONSONANT) = G
            (► BASE-FORM) = MIZENKEI

'gu':: VAF, (↑ TENSE) = PRESENT
            (◄ STEM-TYPE) = GODAN
            (◄ CONSONANT) = G
            (► BASE-FORM) = SHUSHIKEI

'ta':: VSUFFIX, (↑ ASPECT) = DESIRE
                (◄ BASE-FORM) = RENYOKEI
                (► STEM-TYPE) = TRUE-ADJ

'ku':: VAF, (◄ STEM-TYPE) = TRUE-ADJ
            (► BASE-FORM) = MIZENKEI ¦ RENYOKEI

'na':: VSUFFIX, (↑ NEGATIVE) = PLUS
                (◄ BASE-FORM) = MIZENKEI
                (► STEM-TYPE) = TRUE-ADJ

'i':: VAF, (↑ TENSE) = PRESENT
           (◄ STEM-TYPE) = TRUE-ADJ
           (► BASE-FORM) = SHUSHIKEI
```

In 'isogitakunai', the ► of 'iso' is the structure

```
[ STEM-TYPE  GODAN
  CONSONANT  G ]
```

and the ◄ of 'gi' is exactly the same structure, so the two structures can unify. The ► of 'gi' is the structure [BASE-FORM RENYOKEI] and the ◄ of 'ta' is the structure [BASE-FORM RENYOKEI] so the two unify. The entry 'ga', however, has the ► structure [BASE-FORM MIZENKEI] , which would not unify with the ◄ of 'ta', so it is not allowed between the stem 'iso' and the desirative 'ta'. It is, however, allowed in the word 'iso<u>ga</u>nai', meaning "does not hurry", because the ◄ of the negative 'na' requires BASE-FORM MIZENKEI.

The above entries, together with rule (1) above, also predict the forms 'isogu' for "hurries" and 'isogitai' for "wants to hurry" (present tense conclusive forms always end with the BASE-FORM SHUSHIKEI).

In this way, the morphological work is done by markings in the lexicon, just as in the original formulation of LFG. The addition of the ◄ and ► metavariables has allowed the morphological features of morphemes to insure the proper combinations of allomorphs, without interfering with the constituent structure given by the rules in (1), or the associated functional structure.

Finnish Morphology

Jäppinen and Ylilammi (1986) describe some of the problems of Finnish morphology. Their system divides the problem into two parts, morphotactics and stem alternation.

Their approach to morphotactics was to define a set of numbered morphological rules. Each rule corresponded to a string of characters, possibly null, and a set of features. There was also a binary relation defined on these rules, such that if a pair <MR-n,MR-m> was a member of the relation, the rule MR-n was allowed to immediately precede the rule MR-m. An example they give includes the following:

```
MR-α   = ''       -> []
MR-β   = ''       -> []
MR-3   = ''       -> []
MR-4   = ''       -> []
MR-10  = ''       -> [sg,nom]
MR-54  = 'mme'    -> [1pp]  (poss. 1st-person-plural)
MR-61  = ''       -> [pl,nom]
MR-62  = ''       -> [sg,gen]
MR-63  = ''       -> [act,ind,pres]
```

```
{ <MR-β,MR-63>,<MR-β,MR-62>,<MR-β,MR-61>,<MR-β,MR-10>,
  <MR-63,MR-54>,<MR-62,MR-54>,<MR-61,MR-54>,
  <MR-10,MR-54>,<MR-54,MR-4>,<MR-4,MR-3>,<MR-3,MR-α> }
```

These predict that the four possible interpretations of 'kalamme' will be

```
kala+[sg,nom,1pp]
     MR-β < MR-10 < MR-54 < MR-4 < MR-3 < MR-α
kala+[sg,gen,1pp]
     MR-β < MR-62 < MR-54 < MR-4 < MR-3 < MR-α
kala+[pl,nom,1pp]
     MR-β < MR-61 < MR-54 < MR-4 < MR-3 < MR-α
kala+[act,ind,pres,1pp]
     MR-β < MR-63 < MR-54 < MR-4 < MR-3 < MR-α
```

All of these include the 'mme' morpheme, which is a possessive morpheme with value first person plural; it can appear in nouns or nominalized verbs. The first three can attach to the noun stem 'kala' ('fish'), but there is no verb stem 'kala', so the last one fails.

Viewed from the perspective of this paper, each of the MR rules is one allomorph, and the associated features would be assigned to its functional structure. An equivalent form of the binary relation can be constructed by assigning the morphological rule number as the value of a feature MR in ►, and listing the rule numbers of those that may precede it in ◄, as follows:

```
NOUN -> NSTEM STEM-ALT AFFIX* WORD-FINAL
VERB -> VSTEM STEM-ALT AFFIX* WORD-FINAL
  (* is for Kleene-Star, meaning zero or more AFFIXes)

'':: WORD-FINAL, (► MR) = α
                 (◄ MR) = 3

'':: STEM-ALT, (► MR) = β        (see below)

'':: AFFIX, (► MR) = 3
            (◄ MR) = 4

'':: AFFIX, (► MR) = 4
            (◄ MR) = 54

'':: AFFIX, (► MR) = 10
            (◄ MR) = β
            (↑ NUMBER) = SINGULAR
            (↑ CASE) = NOMINATIVE

'mme':: AFFIX, (► MR) = 54
               (◄ MR) = 63 ¦ 62 ¦ 61 ¦ 10
               (↑ POSSESSIVE) = FIRST-PERSON-PLURAL

'':: AFFIX, (► MR) = 61
            (◄ MR) = β
            (↑ NUMBER) = PLURAL
            (↑ CASE) = NOMINATIVE

'':: AFFIX, (► MR) = 62
            (◄ MR) = β
            (↑ NUMBER) = SINGULAR
            (↑ CASE) = GENITIVE

'':: AFFIX, (► MR) = 63
            (◄ MR) = β
            (↑ VOICE) = ACTIVE
            (↑ MOOD) = INDICATIVE
            (↑ TENSE) = PRESENT

'kala':: NSTEM, (↑ PRED) = 'FISH'   (see below)
```

The MR feature now does the work of the binary relation. The MR feature of ► only unifies with the MR feature of ► if the disjunctive lists (with ¦) have a non-empty intersection.

These rules and lexical entries will create three functional structures for 'kalamme' corresponding to the three interpretations found by the Jäppinen and Ylilammi system. It is possible to match their morphotactic rules as given, one by one, and handle all the same data.

The facts of stem alternation in Finnish consist basically of two sets of paradigms, one for noun stems and one for verb stems. A paradigm for 'kala' ('fish') is given below.

FEATURES	singular					plural		
	nom	gen	part	ess	ill	gen	part	ill
STEM-ALTERNATION	a	a	a	a	a	o	o	o
AFFIX		n	a	na	an	jen	ja	ihin

The genitive plural can also have STEM-ALTERNATION 'a' and AFFIX 'o'. All other affixes on the noun use the nominative singular stem.

Rather than discussing Jäppinen and Ylilammi's solution, I will simply give the solution in the current framework. The words belonging to each paradigm are assigned a unique STEM-TYPE in ►. A sufficient number of entries of the morpheme class STEM-ALT are also added, with a STEM-TYPE feature in ◄ and a number of LFG constraint equations. The constraint equations require the functional structure to have certain functional features, which will have had to come from combined AFFIXes following. Sample entries for the above paradigm could be as follows:

'kala':: NSTEM, (↑ PRED) = 'FISH'
(► STEM-TYPE) = 10

'a':: STEM-ALT, (► MR) = β
(◄ STEM-TYPE) = 10
(↑ NUMBER) == SINGULAR

'o':: STEM-ALT, (► MR) = β
(◄ STEM-TYPE) = 10
(↑ NUMBER) == PLURAL
(► SPECIAL-TYPE) = 1

'a':: STEM-ALT, (► MR) = β
(◄ STEM-TYPE) = 10
(↑ NUMBER) == PLURAL
(↑ CASE) == GENITIVE
(► SPECIAL-TYPE) = 2

'jen':: AFFIX, (► MR) = x
(◄ MR) = β
(◄ SPECIAL-TYPE) = 1
(↑ NUMBER) = PLURAL
(↑ CASE) = GENITIVE

'in':: AFFIX, (► MR) = y
(◄ MR) = β
(◄ SPECIAL-TYPE) = 2
(↑ NUMBER) = PLURAL
(↑ CASE) = GENITIVE

'ja':: AFFIX, (► MR) = z
(◄ MR) = β
(↑ NUMBER) = PLURAL
(↑ CASE) = PARTITIVE

Similar entries could be created for all of the paradigms.

Finnish has two other problems for morphology. The first is that each stem potentially has two grades, a weak and a strong. In the framework proposed here, both forms of the stem should be entered in the lexicon, with values WEAK or STRONG for the feature GRADE. Similar features on STEM-ALTs ensure the proper constraints.

The second problem is vowel harmony. Finnish has three groups of vowels, (a,o,u), (ä,ö,y), and (i,e). Those of group1 never occur with those of group2, except in compounds. Those of group3 occur with either; stems of group3 take group2 affixes (Lehtinen 1963, pp.102-103). In our system, stems, STEM-ALTs, and AFFIXes of group1 would be marked (► HARMONY) = GROUP1. Stems, STEM-ALTs, and AFFIXes of group2, and stems of group3 would be marked (► HARMONY) = GROUP2. All AFFIXes and STEM-ALTs will have an equation enforcing harmony, namely (◄ HARMONY) = (► HARMONY). In this way, morphemes with a definite vowel harmony value introduce it, and every morpheme is required to have the same harmony as the morpheme before it, so the harmony applies to the whole word. Since all stems introduce a definite harmony value, and stems are not required to have the same harmony as anything before them, compounds (which have two stems) can have different harmony values in each part.

English

English basically has only one layer of affixes, so each stem can be marked with a STEM-TYPE feature in ►, and each affix with a STEM-TYPE feature in ◄. There will be separate STEM-TYPE features for each possible plural ending, such as 's', 'es', 'im', 'a', etc. Stems can be entered in the lexicon in usual form, but with a dot between the stem and dictionary form ending, if any, so that only the invariant part is necessary to find it in the lexicon. So 'bite' would be 'bit·e', with affixes 'e', 'es', '', '', and 'ing'.

Some of the STEM-TYPEs are phonologically predictable. For example, if nouns have no irregular STEM-TYPE marking, words ending in sibilants will have the STEM-TYPE for 'es', and others will have the STEM-TYPE for 's'. In Japanese, the CONSONANT feature of godan verbs is phonologically predictable. All of the

STEM-TYPEs of Finnish are phonologically predictable, except two noun paradigms (Jäppinen and Ylilammi 1986, p.265). Our system assumes that phonologically predictable features are added by lexical rules, and all idiosyncratic features are also specified, before stems and allomorphs are entered into the lexicon.

Conclusion

This paper has suggested the addition of two metavariables, ◄ and ►, to an LFG system to handle morphology. The unification the ► morphological features of allomorphs on the left of morpheme boundaries with the ◄ morphological features of allomorphs on the right of the boundaries, appears to be a sufficient mechanism to handle many of the morphological phenomena in such divergent languages as English, Finnish, and Japanese.

REFERENCES

Bresnan, Joan, ed. 1982. The Mental Representation of Grammatical Relations. MIT Press, Cambridge, Massachusetts.

Jäappinen, Harri, and Matti Ylilammi. 1986. "Associative Model of Morphological Analysis: An Empirical Inquiry", Computational Linguistics 12(4): 257-272.

Lehtinen, Meri. 1963. Basic Course in Finnish. Uralic and Altaic Series, Vol. 27. Indiana University, Bloomington.