

RICHARD A. HUDSON
London University

COMPLEX SYMBOLS DOMINATING BRANCHING STRUCTURES IN BEJA

In *Aspects of the Theory of Syntax* (p. 170—184) Chomsky suggested that the traditional 'word-and-paradigm' approach is a better basis for a generative grammar for inflectional patterns than the American structuralist approach, with its exclusive emphasis on segmenting words into morphemes. For instance, instead of showing the distributional and semantic similarities between *took* in relation to *take*, and *baked*, in relation to *bake*, by segmenting both *took* and *baked* into stem + tense suffix, it is better in every way to show these similarities by means of syntactic features in a complex symbol attached to the node representing the *word* as a whole, and leave the question of whether and how to segment into morphemes to be decided on its own merits. In this Chomsky is making the same suggestion as Robins (1959) and Matthews (1965, 1967), and I myself am in entire agreement with him.

In this paper I shall illustrate the advantages of this approach in the analysis of a small area of the complex morphology of Beja, a Cushitic language spoken in the Sudan, between the Red Sea and the Nile (see Andrzejewski, 1968 and Hudson, 1964). I shall also make a number of suggestions as to how the approach might be extended in a generative grammar, and point out the implications of these extensions for the transformational-generative theory of language.

1. The lack of correspondence between syntactic and morphological groupings

All the Beja examples that I shall quote will be regular verbs: both the noun and the verb are inflected, but the inflectional patterns of the verb are far richer and more interesting than those of the noun. However, I shall not try to give an overall view of the verbal patterns in this paper, since there are too many complications. Instead I shall concentrate on a few particularly interesting phenomena which can be illustrated with reference to small parts of a few representative verb paradigms.

One of the first things that strikes one about the forms of verbs in Beja is the lack of a neat correspondence between their (external) syntactic or semantic properties and their internal structures: verbs which are the same as far as their mood, tense or person-number are concerned can be very different in their shapes, and vice versa. To take an extreme instance of this, we can consider the following forms (the transcription is a „morpho-phonemic“ one; if a „phonemic“ one had been used the discrepancies would have been even more dramatic).

A. 1 adbíł	‘I collected’	B. 1 tamán	‘I ate’
A. 2 tidbíla	‘you (m. s.) collected’	B. 2 tamtă:	‘you (ms) ate’
A. 3 idbíł	‘he collected’	B. 3 tamyă	‘he ate’
A. 4 tidbíł	‘she collected’	B. 4 tamtă	‘she ate’
A. 5 adanbí:l	‘I collect’	B. 5 tamány	‘I eat’
A. 6 danbí:la	‘you (m. s.) collect’	B. 6 tamtini:a	‘you (ms) eat’
A. 7 danbí:l	‘he collects’	B. 7 tamí:ny	‘he eats’
A. 8 danbí:l	‘she collects’	B. 8 tamtíny	‘she eats’

The only difference between A. 1 and B. 1 is that they involve different lexical items: ‘collect’ and ‘eat’; in all other syntactic and semantic respects they are the same (1st person singular preterite). And yet in A. 1 there is a prefix *a* followed by a ‘stem’ *dbíl*, whilst in B. 1 there is a stem *tam* followed by a suffix *án*. In other words the very order of elements within the word is

determined by purely internal, syntactically irrelevant, factors (the class of the stem), just as in other languages syntactically identical words can belong to different ‘declensions’ or ‘conjugations’. It will be seen that the same is true of all the A and B forms: where A forms have prefixes B forms have suffixes.

The converse of this rule does not hold however, since some A forms have suffixes: A. 2 and A. 6 both have a suffix *a*. Moreover, in three cases (A. 6—8) there is no prefix at all, so we cannot even make the generalisation that A forms contain prefixes and B forms contain suffixes, but that the prefixes in A forms and the suffixes in B forms serve the same function: in some cases the same function is served by both a prefix (*ti*) and a suffix (*a*) in an A form and by a suffix in a B form, and in others (A. 6, 7 and A. 8) it is the lack of any prefix at all in an A form that serves the same function as a suffix in the B form.

Given these facts it seems clear that there is little point in trying to „deepen“ the morphological representations in such a way that all verbs that are 1st person, say, should contain the same morpheme, alongside a morpheme showing that they are singular and present (or whatever other syntactic properties we may need to show). A much more promising approach is to cross-classify verbs by means of syntactic features, which might be introduced either in the base (in the case of tense features) or by transformations (in the case of person/number/gender features). These features would then be mapped onto the appropriate phonological patterns by interpretive rules which could also take account of the kind of verb-root that was involved (like *d-b-l*, ‘collect’, or like *tam*, ‘eat’, for instance).

2. The need to segment words into morphemes

The suggestions made so far raise no particular theoretical problems for current versions of the transformational model: all we need is some kind of rule (possibly among the phonological readjustment rules of Chomsky and Halle, 1968?) which can expand whatever phonological representation the verb-root has in

the lexicon, taking account of the features associated with it. So for instance the root *d-b-l* would be expanded by the prefixing of *ti* if it was accompanied by the feature [2nd person] (or [2-person], to follow Chomsky, 1965), by the suffixing of *a* if it was with [1-gender] (that is 'masculine'), and by the infixing of *i* between the second and third consonant if the attached tense feature was [2-tense] (meaning 'preterite'). The rules would also have to get the accent in the right place, on the last vowel in the root.

However, when we look at a wider range of verb-forms it becomes clear that it would be helpful to segment words into morphemes that is, to analyse them further in the syntactic component, before embarking on the task of relating syntactic structures to phonological structures.

There are two main arguments in favour of analysing Beja verbs into morphemes (or, more accurately, into morphs). Firstly, it is easier to state the sequence of certain elements if we treat them as morphs than if we leave them simply as parts of the word's phonological structure without any particular status in the surface structure. For instance, in the word *kátamtā:na*, 'you (pl.) don't eat', it is both possible and useful to identify the following morphs: *ká* (negative), *tam* (eat), *t* (2nd person), *a* (preterite — sic!), *ĩn* (plural), *a* (2nd person again). It is technically easy to get a string of morphemes in the right order in the surface structure, and let the order of their respective morphs be determined in a one: one way by this, and this is what we need to be able to do, since the elements involved have to be identified on a relatively abstract level. For instance, the word *kátamtā:na* just discussed contains two morphs both of which are *a*, but one of them represents the morpheme {preterite} whilst the other represents {2nd person}. If we want to generalise about the positions of the *t* and the *ĩn* morphs we can easily do so in terms of morphemes: the morphemes realised by these two morphs respectively precede and follow the {preterite} morpheme, realised by the first *a* morph. But to state the same generalisation without reference to morphemes, but just in terms of the morphs concerned, would be impossibly complicated, since it would be very hard to distinguish the two *a*'s from each other.

The second argument in favour of identifying morphemes in the surface structure of Beja verbs is that the same range of morphs can occur, under the same range of conditions, in otherwise unrelated classes of words. If we simply say that in both classes we find the same morpheme we can then cover the variant-forms in a single statement about the one morpheme, but if we have no morpheme we have to cover these patterns of variation twice.

For instance there are two classes of verbs which there is no other reason for identifying (future tenses, which occur in independent clauses, and infinitives, which do not); in both cases we find the same range of forms (*e:t* or *t*) used according to the same rules (*e:t* after a root like *tam*, and *t* after a root like *d-b-l*). In one form (the infinitive) this suffix is also followed by an *i* suffix, and is in fact optional (hence either *diblĩ* or *dibiltĩ*, *tamĩ* or *tame:tĩ*) whereas in the other form this *ĩ* suffix does not occur, and the forms mentioned are found only before a range of optional pronoun suffixes (*dibiltó:k*, 'I'll collect you', *tame:tó:k*, 'I'll eat you'), being replaced under all other circumstances by *át* (after both kinds of root): *diblát*, 'I'll collect' and *tamát*, 'I'll eat'.

What these examples suggest is that we should abstract a morpheme (with very little identifiable meaning such as could suggest a 'name' for the morpheme) which (a) is obligatory in 'future' words but optional in 'infinitives', in which the obligatory suffix is the one marked by *ĩ*, and (b) is signalled (morpho)phonologically by *át* (if not followed by any other suffix) or by *e:t* or *t*, according to the preceding root. (Precisely the same argument would lead us to identify a morpheme which occurs both in the infinitives and in the future tenses of Italian or written French, realised by *er*, *ir*, etc.).

Having argued in favour of generating a layer of structure below the word in surface structure, I should like to restore the balance by repeating what I said in (1) above: that there are also cases where it is no help at all, but rather a positive hindrance, to postulate such a level. The general principle is, of course, that it is helpful to postulate morphemes when the relation between them will be 'agglutinative', i.e. when we can put them in a left-to-right sequence which is matched by the sequence of their

morphs — but not when several morphemes would regularly share a single morph. If we apply this principle, then (practically) all Beja verbs will consist of at least two morphemes, but the number of morphemes in a word's structure will never be as great as the number of features we assign to the word in order to define its syntactic properties. So we shall be able to analyse *tidbīla*, 'you (m. s.) collected', into three morphs, *ti*, *dbīl*, and *a*, corresponding to three morphemes, but these three morphemes between them will reflect a much larger number of features; and a word like *danbi:l*, 'he/she collects', will consist of just one morpheme, which reflects a number of properties similar to the number reflected by the three morphemes in *tidbīla*.

If these conclusions are correct, then we must be able to generate surface structures in which complex symbols dominate branching structures. (The alternative, of attaching all the features to the root-node and adding the other morphemes to this as appropriate, would be intolerably complex, as I shall show in (3) below). This probability was foreseen by Chomsky in „Aspects“, but he did not develop its theoretical implications. The most important consequence of what I have argued in this section is that the syntactic features in the complex symbol and the phonological and morphological specification of the root between them *fully determine* whatever structure the word may be given in terms of morphemes: given the former, it is always possible to predict the latter.

This being so, it is probably best in a TG grammar to introduce all the morphemes except the root morpheme by means of transformations which are sensitive to features in complex symbols; in other kinds of grammars (e.g. systemic grammars — Hudson, 1971) realisation rules would be more appropriate. However, I shall argue in the following section that there are more far-reaching implications if we adopt the transformational approach, since the transformations will not only have to introduce future but will also have to be sensitive to features attached to nodes dominating these features. I am not aware of such rules having been proposed before, and it seems to me that such transformational rules are not very different from realisation rules, as defined by Lamb, 1964.

3. The need for separate feature complexes to be given to morphemes and to words

By analysing words in terms of morphemes we shall identify not only a number of separate morphemes but whole classes of morphemes which can all occur in the same place in the structure of the verb; when we come to state the rules for ordering morphemes we can make them much more general by referring to these classes. For instance, for paradigms like the one below we can define the structures of the verb-forms concerned in the formula: person + stem (+ number) (+ gender).

person	stem	number	gender	
a	dbīl			I collected
ti	dbil		([~]) a	you (m. s.) collected
ti	dbil		([~]) i	you (f. s.) collected
i	dbīl			he collected
ti	dbīl			she collected
ni	dbīl			we collected
ti	dbil	[~] n	a	you (pl.) collected
i	dbil	[~] na		they (pl.) collected

In the paradigm of the finite tenses of a Beja verb there are typically eight distinct forms, as there are in the paradigm above; these eight forms will be defined by eight different combinations of person/number/gender features attached to the verb, as such. The point I want to make in this section is that this classification is different from the natural classification of any one of the verb's constituents, so that we cannot legitimately suppress either of the systems of classification as redundant.

Let us see how many morphemes we need to distinguish in the paradigm above. Starting on the right, we need two „gender“ morphemes: a neutral or unmarked one, realised as *a*, and a feminine one, *i*. In the plural there is no gender-distinction for any of the persons, but the unmarked gender morpheme occurs in 2nd-person forms. In the column before, we need one number morpheme, which is either [~]*n* or [~]*na* according to whether or not it is followed by the gender suffix. Before this there is one stem morpheme, *dbīl* or *dbil*, according to whether another accent is

'imposed' on it by a following morpheme. (Note the change of terminology from 'root' to 'stem'; this will be explained shortly). And right at the beginning of the structure there are four person morphemes, *a*, *ti*, *i* and *ni*, used respectively in 1st-person singular verbs (*a*), in all 2nd-person and 3rd-person singular feminine verbs (*ti*), in 3rd-person singular masculine and 3rd-person plural verbs (*i*).

This gives a total of exactly seven morphemes, apart from the root morpheme which doesn't mark any contrasts in this paradigm; and there are eight ways in which these seven morphemes can combine, each combination characterising a different set of features of the word. It should be clear that there is no one place in the structure of the verb at which all the differences between the eight person/gender/number categories are located: it is the overall pattern of the verb's structure that reflects its person/number/gender, and not just a single morpheme. Moreover, there is no place in the verb's structure at which there as many as eight morphemes in contrast. Therefore if we were to attach the eight person/number/gender features to the stem let us say, rather than to the verb as a whole, this would be a choice imposed by limitations in our theory, and not by the patterns inherent in the morphological structure of Beja.

One class of morphemes which one might expect in a verb's structure would be 'tense', especially since some of the examples already given have reflected tense distinctions. Much the same arguments apply to the various tense-features as applied to the person/number/gender features: there are three distinct tenses (present, 'preterite' and 'past'), apart from the future tense, which is an oddity, and various kinds of 'imperative' forms, but these three tenses represent categories of the verb as a whole and not of any of its constituent morphemes, as I shall now explain.

If anything the situation regarding tense-forms is even more complex than the person/number/gender contrasts were, since in the case of tense it is the form of the stem itself which reflects the contrast, and not that of any affix (at least in verbs like *d-b-l*, 'collect'); with verbs like *tam* the situation is rather more straightforward since there is a tense-morpheme, so we can ignore such verbs here; to take account of them too would simply be to

strengthen my argument. In Beja (but in none of the other Cushitic languages, so Dr. Andrzejewski tells me) there is a very large class of roots which consist simply of two or three consonants (like *d-b-l*) among which a variety of vowels and consonants can be 'interdigitated' (as in Semitic paradigms). It is these vowels and consonants which mark the tense of the verb, as can be seen from the following paradigms.

tense	person	stem	person	stem
of verb				
preterite:	a	dbíl	ni	dbíl
				'we collected'
past:	—	i:dbíl	n	i:dbíl
				'we used to collect'
present:	a	danbí:l	n	e:dbíl
				'we collect'

The arguments which I used to justify the isolation of other morphemes seem not to apply nearly as clearly in the case of the stem forms like *dbíl*, *danbí:l*, etc., and on balance it seems best *not* to analyse these forms further. The rules for interspersing vowels and consonants fit better with the rules specifying the phonological shapes of morphemes. Indeed, it seems to me that a good case can be made for attaching features directly to the „root“ node to determine its full phonological shape — the shape of the „stem“, as I have been referring to it.

However this may be (and there are several other complicating factors to be taken into account not least of which is the possibility of a variety of „derived“ patterns, such as 'causative' and 'reciprocal', within the 'stem'), what is relevant is that there are about fifteen syntactically and semantically different forms of each verb apart from distinctions that involve simply person, gender and number, and those involving the categories like causative which I have just referred to; but there are never more than six distinct forms of the stem for any given root. There are three relevant kinds of root: with two consonants (CC), with three consonants, the third being *y* (CCy), and with three consonants (CCC). There are five distinct forms for the CC roots, and

six for each of the other two. However, the distribution of the two sets of six do not coincide exactly, so that if we take account of the distributions of all the classes, we arrive at seven different distributional classes of stems, shown below:

class of stem	form of stem according to root		
	CC root	CCy root	CCC root
I	CiC	C(i)Ci	C(i)CiC
II			CiC(i)C
III	i: CiC	Ci: C	i: CCiC
IV		Ca: C	
V	e: CiC	Ce: C	e: CCiC
VI	Ci: C	CaCi	CaCi: C
VII	nCi: C	CanCi	CanCi: C

As we might expect, the distribution of these seven stem-classes among the fifteen verb-classes is fairly complex: at one extreme, stem-classes II and VI are found in four verb-classes each, whereas at the other extreme classes V and VII share one verb-class between them, as can be seen from a set of paradigms quoted three paragraphs back: singular present tense forms contain stems of class VII (e.g. *adanbi:l*, 'I collect'; *anri:b*, 'refuse'; *adangĩ*, 'I return'), but plural present tense forms contain class V stems (cf. *ne:dbil*, *ne:rib*, *nidẽ:g*, 'we collect/refuse/return'). The different verb-forms that share the same stem-class are further distinguished by means of affixed morphemes: for instance the preterite and imperative forms both have class I stems, but preterites have person prefixes whilst imperatives don't (*tirĩba*, 'you (m. s.) refused', but *ribã*, 'refuse (m. s.)!').

The conclusion to which I believe these facts lead us is that we should not attach to the root-node in the surface structure the features showing the syntactic and semantic properties of the verb, such as tense and person/number/gender, which should be attached to the node representing the verb as a whole; but that we should attach to the root-node features which take account both of the verb-features with which it can be associated, and also of the phonological shape which will be assigned to it by the morphophonological rules. Moreover, precisely the same should be done to all the other morphemes, since they are all related in

some way to the syntactic and semantic features of the verb as a whole, and they all have to be assigned a phonological form. In other words, the features assigned to morphemes should be taken from a completely different inventory from those assigned to verbs as a whole. If it was simply a question of 'copying' features from verbs to verb-constituents, or vice versa, the whole basis of my argument would collapse, and there would be no particular benefit from assigning features to the higher nodes (in this respect my suggestions are completely different from those of Weinreich, 1966).

Yet another bit of evidence in favour of my position is that there are some features of stems which are totally irrelevant to the syntactic and semantic features of the verb, and should therefore be kept strictly separate from these. I am referring in particular to the four kinds of root that I have mentioned: those like *tam*, 'eat', which have a constant form (comprising an integral number of syllables, incidentally), and those like *d-b-l*, 'collect', *d-g-y*, 'return' and *r-b*, 'refuse' where the lexical item determines only two or three consonants within the stem, the rest of the stem being determined by the stem-class (I—VII) and the number of consonants determined (two, three, or two-plus-*y*). Verbs containing these classes of root differ markedly in their internal structures — even to the extent of having suffixes in one case but prefixes in another, as I showed in (1) above — but in every other respect are the same. The distinction is thus entirely inappropriate at the level of the word. (The same arguments would separate the „declension“ features from the rest in Chomsky's 1965 example of „der Brüder“).

4. Is syntax really different from morphology?

All that I have said so far is compatible with at least the spirit of Chomsky's discussion in „Aspects“, and in particular with his conclusion that it may be necessary to draw a theoretical line between morphology and syntax (in the narrower sense of the rules for combining words (phrases, etc.), rather than morphemes), allowing complex symbols to dominate branching struc-

tures in morphology but not in syntax. However, since then he has (rightly, in my opinion) suggested that it might be necessary to allow this even in syntax, and that categorial labels could be replaced by complexes of features (Chomsky, 1970; cf. also Dougherty, 1970). I believe that this is a suggestion that deserves to be taken much more seriously than it seems (to an 'outsider' at least) to have been, and I should like to conclude this paper by outlining one particular direction in which it might be developed.

The main question arising out of the suggestions I have made in this paper, is whether there is a fundamental difference between morphology and syntax such as would lead us to use a different range of theoretical apparatus for them. The same question arises when we compare Chomsky's *Aspects* position with his suggestions in *Remarks on Nominalisations*: in the former he allowed for branching structures dominated by complex symbols within the word, but not above the word, so that in effect he was suggesting a clear theoretical difference between morphology and syntax; whilst in the later suggestions he considered the other possibility, that this configuration might also occur higher in the tree.

It seems to me that there are four possible answers that one could give:

a) Morphology and syntax should be treated separately, since each needs a different kind of theoretical apparatus (Chomsky's 1965 position; the traditional word-and-paradigm approach as defended and developed by Robins and Matthews).

b) Morphology and syntax cannot be separated and treated in different sections of the grammar, but neither can they be handled, whether together or separately, in terms of any of the models we have available now. This position can be discounted since it is unknowable.

c) Morphology and syntax can be handled in terms of the same theoretical apparatus, namely that developed for syntax with its emphasis on segmentation and structures consisting of strings of items. (Roughly speaking this was the typical position of American structuralist linguists, and those influenced by them). This position is untenable, for the reasons that I gave in (1).

d) Morphology and syntax can be handled in terms of the

same theoretical apparatus, namely that developed for morphology, with its emphasis on cross-classification of words and possibly complex relations between classificatory features and internal shape. This latter is the position towards which I believe Chomsky has moved, and which I feel may be the most fruitful.

The question is thus reduced to a choice between the first and last positions: either we separate morphology from syntax, in the same way as we keep phonology and (perhaps) semantics separate from syntax; or we allow the morphological apparatus to be used in syntax as well, possibly even to the exclusion of the apparatus that is used there now. Of course it would be silly to even start to offer arguments for or against either position in an article of this size (but cf. Hudson, 1973, for my reasons for preferring the last position). The issues at stake are far too big for that, and it is not even obvious in what terms the discussion could take place. For instance, if it is agreed that in the morphology we do need transformational rules that are like realisational rules (as I argued in (3) above), building up a word's internal structure according to its own features, then can this be used as an argument for allowing the same apparatus in syntax wherever it might be helpful, or would this count as making the model too strong so that it ceases to tell us anything about human language?

However, perhaps I can merely make a small contribution to the discussion in the form of a statement about Beja: that for Beja, at least, it would be hard to separate syntax from morphology since syntactic structures (structures of clauses and phrases) can be embedded in morphological structures. For instance, relative constructions are formed by using a clause as „stem“ in a word which could also contain a variety of other elements including number/gender/case concord markers and comparative markers, and which is treated as an ordinary adjectival word.

Compare the following examples, and notice the distributional similarity between adjective morphemes and whole clauses:

independent (dawri-, 'beautiful') *bū:n gwá?tni*, 'she drinks coffee'

as stem, not	<i>dawri:t kittě</i>	<i>bũ:n gwa?tni:t kittě</i> , 'she can't
comparative	'she isn't beautiful'	drink coffee', 'she's not the
		kind of person who drinks
		coffee'
as stem,	<i>dawri:kađ:t kittě</i>	<i>bũ:n gwa?tni:kađt kittě</i> , 'she
comparative	'she isn't more	doesn't drink coffee more'
	beautiful'	

This means that the only way in which the entire syntax could be generated before the morphological rules were applied would be if we could readjust the output of the syntactic component (the syntactic surface structure) so that even if morphemes like the comparative *kaa* are treated in this as sisters of whole clauses, they will be treated as sisters of verb stems instead in the structures on which the morphological rules have to operate. So if we do add morphology as another component separate from syntax, we shall also have to introduce a set of 'adjustment' rules between the two components; which means that the arguments in favour of separating morphology from syntax have to be that much stronger in order to make it clearly worthwhile.

REFERENCES

- Andrzejewski, B. W. (1968), *The Study of the Bedauye Language: The Present Position and Prospects*, Khartoum: Sudan Research Unit, University of Khartoum, African Studies Seminar Paper no 4.
- Chomsky, N. (1970), *Remarks on Nominalization*, in: Jacobs, R. A. and Rosenbaum, P. S., eds., *Readings in English Transformational Grammar*, London.
- Chomsky, N., Halle, M. (1968), *The Sound Pattern of English*, New York.
- Dougherty, R. D. (1970), *A Grammar of Coordinate Conjoined Structures*: I, „Language“ 46, pp. 850—898.
- Hudson, R. A. (1964), *A Grammatical Study of Beja*. Unpublished London University Dissertation.
- Hudson, R. A. (1971), *English Complex Sentences: An Introduction to Systemic Grammar*, Amsterdam.
- Hudson, R. A. (1973), *A 'Word-and-Paradigm' Approach to Beja Syntax and Morphology*, „Foundations of Language“.

- Matthews, P. H. (1965), *The Inflectional Component of a Word-and-Paradigm Grammar*, „Journal of Linguistics“, 1, pp. 139—171.
- Matthews, P. H. (1967), *The Main Features of Modern Greek Verb Inflection*, „Foundations of Language“, 3, pp. 261—283.
- Robins, R. H. (1959), *In Defence of WP*, „Transactions of the Philological Society“, pp. 116—144.
- Lamb, S. (1964), *On Alternation, Transformation, Realisation and Stratification*, in: C. I. Stuart, ed., *Monograph Series on Languages and Linguistics*, No 17, Washington.
- Weinreich, U. (1966), *Explorations in Semantic Theory*, in: T. A. Sebeok, ed., *Current Trends in Linguistics*, 3, The Hague.