PROCEEDINGS
OF THE
ELEVENTH INTERNATIONAL
CONGRESS
OF LINGUISTS

BOLOGNA-FLORENCE, AUG. 28-SEPT. 2, 1972

Edited by

LUIGI HEILMANN

I

MORPHOLOGY IN A GENERATIVE GRAMMAR*

Morris Halle

In the pages that follow I have attempted to study the component of a generative grammar that account for the form and composition of words in a language. I shall therefore be concerned here with trying to account for such facts as that English speakers know that

(1a)   dog think write love transformation antidisestablishmentarianism

are words in their language, whereas

(1b)   švan pensāre katav mīle Donaudampfschiffahrtsgesellschaft

are not.
I shall also be concerned with the further fact that speakers know that some of the words in (1a) are simple entities whereas others are composed of several pieces or morphemes. For instance, English speakers know that

trans-form-at-ion-al

is composed of the five morphemes above and that these five morphemes cannot be concatenated in a different order such as, for example,

ion-trans-al-at-form

al-from-at-trans-ion

I am, of course, not the first to be concerned with questions of this kind and one can find in the literature a number of reasonable suggestions as to how these questions are to be dealt with. The proposal that seems to me most *à propos* is that speakers know the list of morphemes of their language and can, thus, readily determine which of the items in (1a) and (1b) are on their list and hence part of their language. The items in the list are not represented just as sequences of phonetic segments; i.e., as distinctive feature matrices; they are provided also with some grammatical information. E.g., the entry for the English morpheme *write* is provided with the information that it is a verbal root, that it is a member of the so-called "nonlatinate" portion of the list (it is by virtue of this fact that it is allowed to combine with certain affixes, but not with others), that it is

among the small class of verb stems of the so-called "strong" conjugation, etc. Moreover, the list need not be closed; items may from time to time be added to or removed from the list, for speakers learn new morphemes (and words) as well as forget old ones.

To account for the type of fact illustrated above with the word *transformational* it seems to me plausible to adopt the suggestion that speakers have knowledge not only of the list of morphemes but also the rules that govern the manner in which morphemes are concatenated. These *rules of word formation* are of considerable interest in themselves and will be examined to an extent in a later part of this paper. At this point I am primarily concerned not so much with the character of these rules as with the limits of their applicability.

The morpheme list and word formation rules together, however, are not by themselves sufficient to characterize a further property of words. Individual words often exhibit idiosyncratic behavior which they share with few or no other words in the language, and to capture formally these idiosyncrasies, rules and the list of morphemes are not adequate; the grammar needs to contain a special component dealing with these idiosyncrasies. The idiosyncrasies I have in mind can be conveniently grouped into three types.

An examination of the deverbal nouns in (2)

(2) a. approval   recital   proposal   transmittal   reversal

    b. arrival   refusal   rehearsal   acquittal   renewal   carousal
        betrayal   withdrawal   denial   survival   betrothal

shows that while the majority have the meaning "the act of V-ing" and, if derived from a transitive verb, also "the fact of being V-ed", this formula does not hold in all cases. For example, *recital* commonly refers to a concert by a soloist, or *transmittal* is used to refer to the transfer of official documents or information, but not of other things that are quite normally said to be transmitted. The information about the common meaning of these nouns — i.e., "the act of V-ing" and/or "the fact of being V-ed" — can readily be included as part of the rule for forming deverbal nouns in *-al*. It seems rather forced to include in the rule also information about the special meanings of particular words. Nor does it seem appropriate to include this information in the morpheme list, especially in view of the fact that the special meanings are commonly limited only to the given noun and are lacking in other formations with the same base; e.g., the verb *recite* never refers to the act or fact of giving a recital.

A second type of idiosyncrasy that is not readily dealt with by means of rules or entries in the morpheme list is of a phonetic or phonological character. As an example consider the English nouns formed from adjectives by adding the suffix *-ity*:

(3) serene [sɔriyn] - serenity [sɔrɛnitiy]; obscene - obscenity;
     sincere - sincerity;   severe - severity;   profane - profanity;
     divine - divinity

All the nouns above are regularly subject to the Tri-syllabic shortening Rule (cf. Chomsky and Halle [1968], p. 181). The nouns

(4) entirety   probity   nicety   obesity

differ from other words in *-ity* in that they are exceptions to Trisyllabic Shorten-

ing. Again, it seems unnatural to include this information either in the morpheme list or in the rule forming nouns in -ity.

A third type of idiosyncrasy that is not readily incorporated into either the morpheme list or the word formation rules is the so-called "accidental gaps in the dictionary". Compare to this end the deverbal nominals in -al with those in -at-ion or -ion. While, in general, the restrictions on nominals in -al are quite different from those in (-at)-ion, there is a number of verbal stems from which both types of nominal can be derived; e.g., the nominals listed in (5a) are formed from the same stems as those in (2a).

(5a)   approbation   recitation   proposition   transmission   reversion

On the other hand, the nominals in (5b) do not have counterparts in -al:

(5b)   derivation   description   conversion   confusion   permission
        observation   obligation   omission   accusation

Moreover, the nouns in (2b) lack counterparts in (-at)-ion. In other words, the language lacks words such as those in (6).

(6) a.   *derival   *describal   *conversal   *confusal   *permittal   *observal
         *accusal   (but cf. Jespersen, MEG 6.22.22)

    b.   *arrivation   refusation   *refusion   *rehearsion   *acquitation

Once again, it appears somewhat forced to incorporate this information in the morpheme list or the word formation rules. But if it is not to be incorporated there, then one must immediately ask how this information is to be reflected in a grammar.

I propose that idiosyncrasies of the type just illustrated be listed in a special filter through which the words have to pass after they have been generated by the word formation rules. The special information given in the filter under each entry is then added to the representation of the word. In the case of semantic idiosyncrasies such as those exemplified by the special meaning of nouns like *recital* and *transmittal* the filter would supply the appropriate indications about their meaning. In the case of phonological idiosyncrasies like those exhibited by nouns like *obesity*, the filter would supply the information that the noun in question is not subject to the Trisyllabic Shortening Rule, or, more formally, would supply the noun with the feature [-Trisyllabic Shortening Rule]. Finally, "gaps" in the dictionary like those illustrated in (6) and (7) would be accounted for by providing the "missing" words with the rule feature [-Lexical Insertion]. In other words, the fact that English lacks the nouns *derival and *arrivation would be reflected in the grammar by marking these nouns, which would be generated by the word formation rules, as not being subject to lexical insertion and, therefore, incapable of appearing in any actual sentence of the language, in spite of the fact they are neither semantically nor syntactically or phonologically anomolous [1].

---

[1] The proposal just sketched might be modified somewhat as regards the treatment of words formed by rules that traditionally have been called "nonproductive". One might propose that all words formed by nonproductive rules are marked by these rules as [— Lexical Insertion]. The smaller subset of actually occurring words formed by such rules would then be listed in the filter with the feature [+ Lexical Insertion]. E.g., the nouns formed with the suffix -al would all be generated with the feature [— Lexical Insertion]; the relatively small number of actually occurring nouns of this type, like those listed in (3a), will appear in the filter marked [+ Lexical Insertion]. In other words, it is assumed that words generated by a

In other words, I am proposing that the list of morphemes together with the rules of word formation define the set of *potential* words of the language. It is the filter and the information that is contained therein which turn this larger set into the smaller subset of actual words. This set of actually occurring words will be termed the *dictionary of the language*. Observe especially that the dictionary as here conceived is nothing but a product of what is contained in the three components identified above; i.e., the dictionary is completely determined by the contents of the morpheme list, the word formation rules, and the filter listing the exceptions. It might appear, therefore, at this point that there is no need to postulate the dictionary as a separate entity of the grammar, for its contents can readily be derived from that of the other components. There is some evidence suggesting that the dictionary plays a role above and beyond that of the logical product of the three other components. I shall mention this evidence at a later point.

The examples discussed above have been chosen from the domain that traditionally has been called *derivational morphology*. As far as I can tell, facts that traditionally have been treated under the separate heading of *inflectional morphology* must be handled in completely parallel fashion to those discussed above. I know of no reason why the list of morphemes should not include also the *inflectional affixes* or *desinences*, or why the rules of word formation should not also include rules for positioning the inflectional affixes appropriately or for handling such other inflectional phenomena as reduplication, stem Ablaut, etc.

It is important in this connection to realize that the three types of exceptional behavior that above have been handled by means of the filter are not restricted to word derivation but are found also in the inflection. Thus, one finds that particular case forms of particular words idiosyncratically possess meanings that are in general not those of either the base noun or the case. For instance, in Russian the instrumental case of certain nouns designating times of the year and of the day has special adverbial force that is not possessed by other nouns in the instrumental case. In particular, *letom* may mean "in summer", *noč'ju* "at night", *zimoj* "in winter". However, *avgustom* may *not* mean "in the month of August", or *obedom* may not mean "at dinner (or noon) time".

The second type of idiosyncratic behavior which I proposed to handle with the help of the exception filter was phonological irregularity of the kind illustrated in (4). Parallels to this type of exceptional behavior are found also within paradigms; i. e., some forms in a paradigm are subject to a given phonological rule, others are not. An interesting illustration of this is provided by the accentuation of Russian nouns, which I have recently had occasion to study in some detail (see Halle [1973]). As discussed there, a considerable number of Russian nouns must be entered in the dictionary without stress marked on any of their vowels. Such words will then be subject either to the OXYTONE rule, which places stress on the last syllable, or to the CIRCUMFLEX rule, which stresses the initial syllable. It can be shown that the OXYTONE rule must be ordered towards the beginning of the phonological rules, whereas the CIRCUMFLEX rule must be ordered towards the end of the phonology. Moreover, in paradigms in which one or more forms are subject to the CIRCUMFLEX rule, the remaining form will be subject to the OXYTONE rule, but not vice versa. Finally no form is subject to both stress rules; words in this class are either OXYTONE or CIRCUMFLEX. These somewhat elaborate facts can be handled straightforwardly by marking words subject to the CIRCUMFLEX rule with the rule feature [-OXYTONE].

productive process are all actually occurring and that only exceptionally may a word of this type be ruled out of the language. On the other hand, words generated by a nonproductive rule are assumed not to be occurring except under special circumstances. In this fashion we capture the difference between productive and nonproductive formations.

To see this more clearly consider the feminine nouns of Russian that are stressless. As shown in (7) we find in this class at least three distinct types of accent pattern. The first, illustrated in (7a), has stress on the desinence in all case forms; the second, illustrated in (7b), has initial stress (CIRCUMFLEX) in the nom. pl. and desinential stress elsewhere; the third group, illustrated in (7c), has initial stress (CIRCUMFLEX) in the nom. pl. and acc. sg., and desinential stress elsewhere.

	Nom. sg.	*Inst. sg.*	*Acc. sg.*	*Nom. pl.*	*Dat. pl.*
(7) a.	kočerga	kočergoj	kočergu	kočergi	kočergam
	gospoža	gospožoj	gospožu	gospoži	gospožam
b.	skovoroda	skovorodoj	skovorodu	skovorody	skovorodam
	arba	arboj	arbu	arby	arbam
c.	boroda	borodoj	borodu	borody	borodam
	golova	golovoj	golovu	golovy	golovam

The facts illustrated in (7) will be accounted for in the following fashion. The three types of noun forms under discussion will be entered in the morpheme list without stress and none of the word formation rules will supply stress to them. This will be sufficient to obtain the correct output for the nouns of type (7a). In the case of the other two types of noun it will be necessary to provide the information that some of their case forms are [-OXYTONE], so that the CIRCUMFLEX rule can apply to them. In the light of the discussion above, this can be done quite straightforwardly by listing in the filter the appropriate words, e.g., the acc. sg. *golovu* and the nom. pl. *skovorody* as [-OXYTONE]. In sum, the rules of word formation will generate the inflected forms in the fashion to be expected; in most cases these will pass through the filter without further effects. In the nom. pl. and acc. sg. forms under discussion, however, the filter will supply a special marker by the phonology. But this implies that each of the case forms will appear as a special entry in the dictionary [2].

Finally, paralleling the "accidental gaps" in derivation illustrated in (6) above one finds various kinds of defective paradigms in the inflection. For instance, in Russian there are about 100 verbs (all, incidentally, belonging to the so-called "second conjugation") which lack first person singular forms of the nonpast tense. Russian grammar books frequently note that such forms as (8)

(8) *lažu "I climb" *pobežu (or *pobeždu) "I conquer"

 *deržu "I talk rudely" *muču "I stir up"

 *erunžu "I behave foolishly"

"do not exist", or "are not used", or "are avoided". The most recent scholarly grammar of Russian (Švedova [1970] § 988) remarks that no ready explanation

---

[2] A consequence of the treatment proposed here is that the absence of the fourth type of accentual pattern (initial stress in the acc. sg. only) will be regarded as a pure accident and not as a special subregularity of the language which must be captured in the grammar. There are about a dozen nouns that have the stress pattern of (7c). I know of no other interesting features that these nouns share in common.

for these gaps has been offered. Thus, it has been suggested that the first three forms cited above are avoided because they are homophonous with 1. sg. forms of other verbs. There are, however, other verbs in the language where the identical homophony has not resulted in any gaps (cf. *vožu* "I lead" or "I cart")[3]. Equally unconvincing is the suggestion sometimes made with regard to the last two verbs cited in (8). It is said that these 1. sg. forms are "difficult to pronounce" or "unusual". But as is noted in the Švedova grammar, since the language has 1. sg. forms exhibiting precisely the same behavior (e.g., *vonžu* "I thrust (a knife)" or *šuču* "I joke") this hardly is a plausible explanation. It would appear, therefore, that we are faced here with an "accidental gap" in the dictionary. In view of what has been said above, the natural way to handle these facts is to mark such forms as those in (8) as being [-Lexical Insertion]. In other words, just like the forms in (6), those in (8) are incapable of appearing in any well-formed sentence of the language in spite of the fact that they exhibit no semantic, syntactic, morphological or phonological abnormality.

At this point it might be useful to review briefly the proposal that has been made above. To aid in this review I have prepared a block diagram of the proposal in Figure 1. I have suggested that morphology consists of three distinct components:
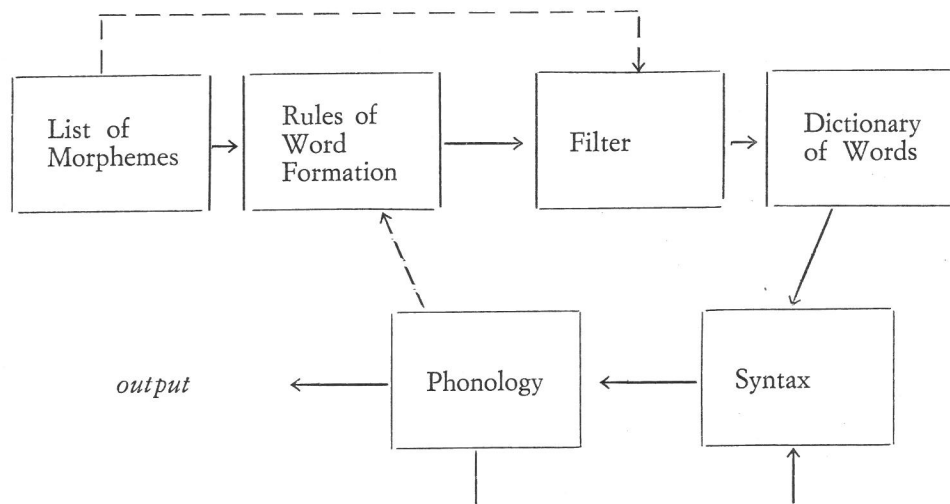

FIGURE 1.

a list of morphemes, rules of word formation, and a filter containing the idiosyncratic properties of words. The list of morphemes and the rules of word formation together define the *potential* words of the language. The set of *actual* words is obtained from that of the potential words by applying to the latter the modifications indicated in the filter. One can think, then, of the morphology as producing a long list of words: it is this list that is designated by the term *dictionary*. I should like to propose further that the lexical insertion transformations be thought of as selecting items from the dictionary and as entering these in appropriate slots in structures representing the underlying constituent structure of particular sentences. It is to these underlying representations that the syntactic transformations apply in the by now familiar fashion and generate what has been called the *surface structure*. I shall assume that the large body of phonological rules — rules like the overwhelming majority of those discussed in *The Sound Pattern of*

---

[3] This tendency might also be one of the reasons for the "gaps" in the paradigms cited in (8) above.

: first three
ig. forms of
he identical
')³. Equally
t two verbs
e" or "unu-
has 1. sg.
knife)" or
:, therefore,
In view of
mark such
ust like the
sentence of
, morpholo-

at has been
he proposal
omponents:

ctionary
Words

the idiosyn-
d formation
al words is
ie modifica-
s producing
lictionary. I
be thought
ppriate slots
ular senten-
isformations
called the
les — rules
Pattern of

iradigms cited

English or other works of generative phonology — apply to the surface structure, and I shall disregard here the refinement that is necessitated by the fact that at least some phonological rules apply as part of the transformational cycle of syntax; cf. J. Bresnan (1971) and (1972).

We must briefly review a problem which arises with regard to the lexical insertion transformations. I proposed above that the lexical insertion transformations have access only to words in the dictionary and, moreover, that the dictionary contains only (and all) fully inflected forms of the language. This proposal might appear to run into a difficulty, for the case which a given noun takes in a sentence is normally determined by its position in surface structure, whereas lexical insertion takes place at a much earlier stage in the derivation. This difficulty, however, is not insurmountable. One might propose that instead of inserting a single item, the lexical insertion transformations insert partial or entire paradigms, i.e., certain or all inflected forms of a given "word". A perfectly general convention can then eliminate all but the one inflected form that fits syntactically into the configuration in which the word is found in surface structure. There are, of course, other equally plausible ways of surmounting this difficulty. Since I am concerned, however, only to show that the difficulty is not one of principle, I shall not explore here any of the other alternatives.

I do not find it surprising that the lexical insertion transformations characteristically affect paradigms rather than single dictionary entries. It is well known that paradigm pressure plays a potent role in the evolution of languages. For example it is because of paradigm pressure that Russian has lost the consonantal alternations k ~ č ~ c in the nominal inflections. In fact, paradigm pressure provides a very plausible explanation for the "accidental gaps" in the Russian conjugations illustrated in (8) above. If paradigms can influence the evolution of language then there is every reason to expect that paradigms must appear as entities in their own right somewhere in a grammar. If my proposal is correct one such place would be the dictionary from where the lexical insertion transformations draw items for insertion into a sentence. Note, incidentally, that if this is correct then the dictionary must be organized into paradigms in some way and it would then no longer be equivalent to the logical product of the morpheme list, the word formation rules and the exception filter.

I now turn to an examination of the character of the word formation rules. It is hardly to be doubted that languages have words which consist of linear strings of morphemes without internal structure, e.g. (10a).

(10a) serendip+i+ty vac+ant tot+al bro+ther hand+some
       be+lieve

I shall assume that words of this type have the structure shown below by virtue of the existence of word formation rules that express this fact perhaps in the manner of templates such as those in (10b):

(10b) $[\text{STEM}+i+ty]_N$ $[\text{STEM}+ant]_A$ $[\text{STEM}+al]_A$ $[\text{STEM}+er]_N$
       $[\text{STEM}+some]_A$ $[be+\text{STEM}]_V$

Presumably the stems in the list of morphemes will be appropriately marked so that a given stem will be substitutable only in certain frames and not in others.

A somewhat more complex type of word is constituted by those words that are derived from other words. Thus, in English we have nouns derived from verbs as in *arrival, refusal,* or *condensation*; nouns derived from adjectives as in

profanity, obesity; verbs derived from adjectives as in *darken, stiffen, blacken*; or adjectives from verbs as in *explanatory, anticipatory, obligatory*, etc To capture such facts as there we would presumably have to have word formation rules of a form such as

(11)   [VERB+al]ₙ
        [ADJ+en]ᵥ

Word formation rules define, in part, the content of the dictionary. However, in order for rules of the type illustrated in (11) to operate properly, they must have access to the dictionary, for it is only there that such crucial information as that *arrive* is a verb of English will be found. In other words, word formation rules must be presumed to have available not only the information contained within the string on which they are to operate, but they must also have access to the content of another component of the grammar. This is an important formal difference between word formation rules and the more familiar phonological rules.

That the content of the dictionary affects the formation of words is hardly to be doubted. Thus the existence of *arrival* and *confusion* is one reason why English lacks *arrivation and *confusal. However, it must immediately be noted that this cannot be the whole explanation for doublets such as *recital* and *recitation, transmittal* and *transmission* do exist.

I must also explain at this point that formulas such as those in (11) do not imply that phonological rules will invariably apply to words derived from words in the familiar cyclical fashion. While some derived words have nested constituent structure at the stage where phonological rules apply, others do not. For example it has been argued in *The Sound Pattern of English* and elsewhere that words such as

(12a)   exaltation relaxation elasticity obligatory declarative
        assimilatory generative

must be presumed to have internal constituent structure, whereas words such as

(12b)   consultation   information

must be presumed to be formed by a linear concatenation of morphemes. The reason for this differential treatment of what appear to be very similar words is obscure at present. I know, however, of no other plausible way of accounting for the differences in stress and reduction in the pretonic vowel than by postulating a difference in constituent structure.

We have just seen that word formation rules must have access to the dictionary in order that they may function properly. This fact would require that word formation rules be formal devices of considerable power. It would, however, seem that even this increase in power is not sufficient and that additional power is required by these rules. To see this, consider the formation of the inchoative verbs in English with the suffix *-en*. These verbs have recently been studied by Siegel (1972) who has noted that these verbs are formed from adjectives and are, moreover, subject to the phonetic condition that the base must be monosyllabic and end with an obstruent which optionally may be preceded by a sonorant. It is for this reason that we have the verbs in (13a) but not those in (13b).

(13)   a. blacken whiten toughen dampen harden
        b. *dryen *dimmen *greenen *laxen

It should be noted that the "words" in (131b) are phonetically well formed in English as shown by the existence of the words in (14).

(14)   lion   women   Keenan   flaxen

The above phonetic condition, therefore, is limited to the particular verbs under discussion. The condition, however, appears to be a condition not on the string formed by the rule but rather on the string after the phonological rules have applied to it.

To see this, consider the verbs

(15)   soften   fasten   moisten

These verbs are clearly derived from the adjectives and would have in their underlying representation the form

(16)   soft+n   fast+n   mɔys+n

These strings, however, violate the constraint against obstruent sequences appearing before the inchoative suffix. It is, of course, immediately obvious that because of the phonological rule which deletes [t] in the environment $s - \left\{ \begin{matrix} 1 \\ \overline{n} \end{matrix} \right\}$ these forms will appear in the output with a single obstruent, thus conforming to the constraint. But if this is indeed the correct reason for the well-formedness of *moisten* as contrasted with the ill-formedness of *laxen*, then it must be supposed that the rules of word formation must have access not only to the dictionary but also the output of the phonology. In other words, we are saying, in effect, that a word formation rule will produce acceptable words if the words formed by this rule conform to certain conditions when they have been acted upon by the rules of the phonological component. This is clearly quite unlike the more familiar rules of phonology which will or will not apply to a string depending only on the characteristics of the string in question.

When examples of the sort just presented have been discussed the assumption has usually been made that rules of word formation are part of the phonology and the conclusion was, therefore, drawn that phonological rules must be much more powerful devices than had heretofore been supposed. In particular, in past studies of generative phonology it has been assumed that the decision of whether a given rule should or should not be applied to a string depends solely on the composition of the string at the point in the derivation where the rule in question applies; it does not depend in any way on the shape of the string at a later or at an earlier point in the derivation. But as we have just seen, there are aspects of language that require the power of rules which have access to several stages in the derivation at once. It is obvious that, in general, one would not wish to replace less powerful by more powerful devices especially when it is known that the less powerful devices are capable of handling a very large part of the task at hand. Under such circumstances, it would be essential to attempt to limit as much as possible the domain in which the more powerful devices may be invoked. There seems to be a fairly natural way of achieving this, given the framework of grammar sketched above (cf. Figure 1) where the rules of word formation are distinct and separate from the rules of phonology. I would like to propose that the added power of having access to different stages in a derivation be made available only to word formation rules, whereas the rules of phonology be restricted, as in previous work, to information overtly present in the string at

the point in the derivation at which the phonological rule applies. This restriction appears to be tenable: to the extent to which I have been able to investigate proposed instances of "derivational constraints" these could always be handled as part of the word formation component.

The proposal just made amounts to saying that word formation is a fundamentally different process than phonology. In fact, it may well be useful to speak not of "rules of word formation" but rather, to use a term suggested by Lakoff and others, and speak of "derivational constraints that hold in word formation". In the case of word formation, we are dealing with conditions that no string of morphemes can ever violate if it is to be admitted to the dictionary as a legitimate word of the language. While my own investigations of word formation in different languages are not extensive enough to allow me to place too much confidence in general impressions that I have gathered, nonetheless it seems to me significant that I have yet to come across any clear instances where word formation rules have to be ordered in that tightly constrained fashion that is constantly encountered in true phonological rules. I should like to propose therefore that the word formation component differs from the phonology by having completely different principles of interaction among rules. Whereas in the phonology this interaction is captured by means of the convention of linear order of rule application, the interaction among word formation constraints may require a different principle altogether; e.g. simultaneous application. Needless to say at this point this must remain in the realm of pure speculation.

M.I.T.
Cambridge Mass.

BIBLIOGRAPHY

Bresnan, J. (1971), « Sentence Stress and Syntactic Transformations », *Language 47*, pp. 257-281.

— — (1972), « Stress and Syntax », *Language 48*, pp. 326-342.

Chomsky, N. and Halle M. (1968), *The Sound Pattern of English* (Harper and Row, New York).

Halle M. (1973), « The Accentuation of Russian Words » *Language 49*, pp. 312-348.

Jespersen O. (1946), *Modern English Grammar Part VI: Morphology* (Allen and Unwin Ltd., London).

Lightner T. (1965), *Segmental Phonology of Modern Standard Russian*, unpublished Ph.D. thesis, M.I.T.

Siegel D. (1971), « Some Lexical Transderivational Constraints in English » (Unpublished paper, M.I.T.).

Švedova, Ju., ed. (1970), *Grammatika sovremennogo russkogo literaturnogo jazyka* (Nauka, Moscow).