

## THE WORD

Not long ago, when general linguistics tended to be identified with the philosophy of language, a linguist entitling a short article "The Word" would have been rightfully accused of unbearable presumptuousness. Any discussion of the word involved, in fact dealing with the relationship between thought and language, that is, penetrating into an area which the linguist neither dared nor desired to exclude from his researches but in which he felt ill-equipped to accomplish anything worthwhile alone. In the second place, it involved all the questions raised by the nature of the sign, in other words, semiology in its entirety. Finally it led to reconsidering the relationships between the "word" and the sentence on the one hand, the word and the "lower" elements of the sequence, syllables and "phonemes" on the other. The question, however, that the linguist could not ask was whether criteria existed that permitted, for all languages and all cases, the identification and delimitation of a segment of the sequence as a determined word. For this it would have been necessary for a linguist to feel that it was his duty to define precisely the terms he used. He would also have needed enough courage to foresee

Translated by Victor A. Velen.

that the term "word" itself might have to be put aside if his researches showed that this term could not be given a universally applicable definition.

Today's linguist, encouraged by thirty years of progress in his field, feels that he has the right and obligation to set aside certain problems because they are not in his province and because the way in which he envisages his research does not depend upon solutions which he might be able to find for these problems. The effort of linguistics during the last decades has tended toward the assurance of its own autonomy and this effort has been largely crowned with success. It will certainly be necessary some day to confront language and thought, but this effort cannot bear fruit until the study of thought (should we say psychology?) is capable of freeing itself completely from certain idealistic hypotheses and of furnishing data, on the basis of observation, comparable to those provided by contemporary linguistics.

The effort that has been made, in linguistics, to define exactly the technical vocabulary naturally entails that we may not contrast "word" with "thought" but rather "language" with "thought", and that semiological considerations concern a unit, the "sign," which is defined as being of variable extension (a sign may be composed of a number of signs), but which there is no reason to identify with what, in current non-scientific usage, is termed a word. Semiology, as revealed by recent studies, has no need of the word. And we should not believe that semiologists are, in fact, thinking of "word" when they write "sign." There are those who might actually think "sentence" or "statement" without, however, forgetting that the *-r-* in French *payera* is also a sign.

In contemporary linguistics the only way to pose the problem of the word seems to be the following: there exists in the current usage of the languages of contemporary cultures a term *word* (French *mot*, German *Wort*, Italian *parola*, Spanish *palabra*, Russian *slovo*, etc.) which indicates a segment in a spoken or written sequence that can be isolated from its context by either pronouncing it alone in speech or separating it by a space from other elements in a text and attributing to it a specific function or meaning. Broad agreement in European languages has led us to believe that the notion itself was universal i.e. that all languages have a term to designate such a segment. But it is enough to note

## The Word

in this matter that the divergence between the Greek which distinguishes *lexis* from *logos* and Latin which contents itself with one word *verbum*, for us to conclude that this agreement is the result of a community of culture beginning with Latin and perhaps facilitated by the original identity of *verbum*, *word* and *Wort*, divergent forms of one and the same more ancient vocable. On reflection it seems necessary to reach a high cultural level before one feels the need to analyse the flow of speech. The use of dictation, for example, which requires the speaker to divide his text into segments that are not so long as to overburden the memory of his scribe, may have brought about an awareness that in order to divide a statement into segments there is a more adequate method than simply making a break after just any syllable.

It is clear, however, that what interests the linguist is not whether all linguistic communities provide a term to designate an isolatable segment of a statement. It is not even a question of finding out if all languages have such isolatable segments. Probable as this may be, some languages, such as Eskimo, permit us to imagine what would be a language in which utterances could be divided only into successive sentences. The real problem is to find out whether the isolatable segments that are designated as words correspond to a definite linguistic reality and if there is some means of analysing utterances that would better illuminate the way language functions. It may be necessary to consider giving up physically isolating all the elements with which one works if this were to permit a clearer view of true analogies or profound differences.

Grammarians and linguists have long known that the analysis of utterances can be pursued beyond the word without going into phonetics, that is, ending with segments of speech, such as syllables or phonemes, which have nothing to do with meaning. One doesn't have to be very audacious to discover and affirm that, in the French word *donnerons* "we will give" three successive units can be identified: a root *donn-*, an element *-(e)r-* that denotes the future, and an ending *-ons*, which denotes the first person plural. Any one of these elements may be replaced by something else without affecting the remaining elements in either form or meaning. In place of *donn-(e)r-ons*, we may have *chant-er-ons*, *donn-i-ons*, *donn-er-ez*; in *chant-er-ons*, the form and the sense of *-er-*

and of *-ons* are not changed; in *donn-i-ons*, the identity of *donn-* and of *-ons* is intact; in *donn-er-ez*, *donn-er-* is always the “future of *donner*.” This analysis, inasmuch as it is based on and limited to the frame of the word could not be accomplished without respecting an obvious hierarchy: *donn(e)* is a central element which may appear without *-(e)r-* and without *-ons*, for example in the imperative *donne!*; *-(e)r-* and *-ons* do not exist without *donn(e)* or some other form like *chant(e)* or *mang(e)*, which can be substituted in any proper grammatical context. The way the facts are presented is normal for modern linguistics. Fifty or sixty years ago one would have sought primarily semantic reasoning to justify the hierarchy: *donn(e)* was recognized as the sole element conveying meaning, the only one which corresponded to a “representation,” and it was designated as a *semanteme*; *-(e)r-* and *-ons*, which hardly evoked any image, seemed to impose themselves more through their form than through their meaning, as suggested by the term *morpheme* which was used to describe them.

What required more daring was the extension of this analysis beyond the framework of the word. In a complex of three words like *dans le château*, *château* is that element which may be treated as *donn(e)* was treated above, while *dans* and *le* are treated like *-(e)r-* and *-ons*. Three successive units are identified: *dans* which marks the withinness, *le* which is the “definite article,” and the element *château*. Each of these three may be replaced by something else without affecting the form or sense of the remaining two elements. In place of *dans le château* we might have *pour le château*, *dans un château* or *dans le chemin* where neither the form nor the meaning of *dans*, *le* or *château* have been modified. Here, again, we note the hierarchy: *château* may be used without *dans* and without *le* as a vocative or in an enumeration, while *dans* and *le* cannot be employed without *château* or some other form such as *chemin* or *jardin*, which can be substituted in all proper grammatical contexts. The same semantic argument which results in the designation of *donn(e)* as a *semanteme*, *-(e)r-* and *-ons* as morphemes, could be used here so that *château* would be classified as among the former and *dans* and *le* among the latter. So far, however, *dans* and *le* remain words since they are separated from their neighbors in the text by two spaces while *-(e)r-* and *-ons* are not words. Can *dans* and *le*, at the same time,

## The Word

be both words and morphemes, while many morphemes are parts of words? Which analysis should have preference? That which produces words or that which produces semantemes and morphemes? If the reasons we have for using a single term for the designation of both *dans* and *le* on the one hand and *-(e)r-* and *-ons* on the other appear to us to be decisive, we may be led to ask ourselves how valid the reasons are which led us to see a word in *dans* and in *le*, parts of words in *-(e)r-* and in *-ons*.

Linguists who worked with semantemes and morphemes certainly sensed the problem raised by this double possibility of analysis but they don't seem to have attempted to solve it or even state it clearly. They tended to define morphology as the study of morphemes and no longer, as their predecessors and many of their contemporaries, as an examination of the variations in the form of words. But they did not dare to draw all the inferences of the point of view which they adopted and they did not propose, for example, to include in morphology French prepositions as well as Latin declensions.

Other linguists came along later and, with good sense, put aside the semantic considerations that had led to the adoption of the terms semanteme and morpheme. Instead they put the emphasis on what is held in common by the "semantemes" *donn(e)* and *château* and the "morphemes" *-(e)r-*, *-ons*, *dans* and *le*, that is to say their value as minimum signs, elements endowed with a form and a meaning not susceptible of being conceived as resulting from the combination of two or more units endowed with meaning. They unfortunately designated these minimal signs as "morphemes," and this has created within general linguistic usage a permanent source of confusion between morpheme as opposed to semanteme and the morpheme which encloses it. They have not, however, given up making use of the hierarchical order between elements like *donn(e)* and *château*, which can exist by themselves and which are designated as "free forms" and those of the type of *-(e)r-*, *-ons*, *dans* and *le* which are designated as "bound forms." These designations are obviously much preferable to those used by the "mentalists" of the preceding generation. But if those who make use of them think they have thus salvaged the traditional notion of the word they are mistaken; it remains up to them to explain why the bound forms *dans* and *le* are words

while *-(e)r-* and *-ons* are not. Moreover, when the principle of the hierarchy of forms (founded on the possibility for some to remain when their neighbors are eliminated) is applied systematically, it leads to a classification of many more levels than the rather oversimplified one, which opposes free forms to bound forms. The conditions under which, in French, one may hear all by itself a supposed free form like *mémoire* are just about as exceptional as those where the bound forms *pour* or *-être* might constitute by themselves a complete statement. There are degrees in the hierarchy thus obtained: the subject is properly the complement of the predicate without which the latter cannot exist. In French a substantive never exists outside of fixed forms without an article but it appears frequently without being preceded by a preposition; *dans* and *le* are not, therefore, to be placed on the same plane.

We remain, in fact, faced with our problem: why is *dans le château* three words while *donnerons* one word only? Could it be that in the former phrase the nucleus, the free form, the semanteme of our predecessors, comes *after* its satellites while in the latter it precedes them? This, of course, is not an explanation but a simple statement of what is quite normal in French. When one extends the observation to the other languages one notes that the satellites of similar function are, in writing, treated as independent words when they precede the nucleus, and are tagged to that nucleus when they follow it. Among the European languages which have a definite article, some (the majority) put it before the substantive. In this case, it is not attached to the substantive in writing, and, invariable as the English *the* or inflected as the German *der, die, das*, it represents what is called a word. In those languages, which place the article after the substantive, the article is attached to the substantive: Danish *bord*, "table," *bordet* "the table," Rumanian *nas*, "nose," *nasul*, "the nose." It is interesting to compare Rumanian and French because the article has the same origin in the two languages (Latin *illum* in both *nasul* and *le nez*) and, nevertheless, it appears to maintain its independence in *le nez* while it has lost it in *nasul*.

There is no doubt that the writing habits which we have pointed out are symptomatic of some actual differences in speech. In those languages which place the article before the substantive,

## *The Word*

it is possible to place an adjective between the article and the substantive: *the long nose*; but this is not possible when the article is placed after the substantive. In Danish, if a descriptive adjective is added to the substantive accompanied by the definite article, the latter is no longer placed after the substantive but assumes a different and independent form and precedes the adjective and the substantive: *bordet* "the table," but *det store bord* "the large table." In Swedish and in Norwegian, the independent article is employed before and the agglutinated article after: Swedish *bordet*, but *det stora bordet*. The spaces left when writing the article which is placed before the substantive bear witness to the real independence of the article, which is assured by the permanent possibility of inserting one or more determinants between the article and the substantive.

It is not difficult to understand the reasons which make possible the insertion of an element between the article and the substantive but prevent it between the substantive and the article. When the two elements of the spoken sequence are constantly in contact, the end of the first and the beginning of the second tend inevitably to assimilate each other, so much so that the phonic identity of the two elements may be affected. This infringement on the integrity of the form is less dangerous, in general, in a satellite than in a nucleus. The reason for this is that in a given context the choice between different satellites is rather limited (for example, between the definite article and the indefinite article or between the singular and plural) so that linguistic communication will not be endangered as long as the least difference exists between one form and the other. On the contrary the nucleus, substantive or verbal root, is a form chosen from among hundreds and its phonic identity must be well marked. On the other hand the identity of any element of the sequence is better assured by its beginning than by its end. The reason for this is that the speaker naturally enough perceives the beginning at the outset and the languages are so made that the listener very often identifies the word before it is entirely uttered. Since the speaker also has experience as a listener and since the relations of his partner in the dialogue frequently indicate that the message has been understood before it has been completed, the distinctions tend to be less clear towards the end

and may actually disappear completely. Thus a German who distinguishes perfectly between *Torf* and *Dorf* will make no difference in the pronunciation of *Rat* and *Rad*. Once this is accepted it will be understood that while it is natural enough to bring two elements like the substantive and its article together, speakers won't be tempted to separate them if the changes which are apt to result from their contact affect the end of the substantive, less important for the identification of the latter than its beginning. On the contrary, if the article precedes, its permanent presence in contact with the beginning of the substantive might dangerously affect this beginning and as a consequence the identification of the term. The occasional insertion of an epithet between the article and the substantive will be welcome and will be favored as contributing to a sound separation of the article from its substantive.

From the examples cited above, it is clear that the separability of successive elements of speech is, fundamentally, what normally brings about the writing of separate words. The three words *dans le château* affirm their independence in an expression such as *dans tout le grand château* where the insertion of *tout* and of *grand* justifies the traditional transcription and interpretation. In *donnerons*, on the other hand, only a single insertion may be envisaged, that is *-i-* after the *-r-*, so that we have *donnerions*; but we have here a satellite, the same one we opposed to the element *-(e)r-* above and which here turns out to be combinable with it. There is only one unit in the series of satellites which may be combined with a verb root like *donn(e)*, each occupying the place which is reserved for it by tradition.

\*

As long as the tendency to assimilate the elements in permanent contact does not lead to the elimination of the borders between them there is no unresolvable conflict between the tradition which analyses the sequence into words and a more refined method which operates with minimal signs. The inseparability which characterizes a complex like the Danish *bordet* is a trait which deserves recognition whatever the method adopted. Declaring that *bordet* is a word is therefore a perfectly acceptable process.



## The Word

Writing *bord-et*, with a space on each side of the word and a hyphen between the two minimal signs, should be perfectly satisfactory to everyone.

Problems arise when it is no longer possible to analyse the word physically into its successive elements. This may either be because there is no decisive reason for drawing a dividing line here rather than there, or because the phonic feature which carries the meaning does not appear where the meaning would demand it.

Let us take the Latin accusative *rosam*. If it is compared to the nominative *rosa* which seems to be the same form, one is tempted to cut *rosa-m* where *rosa-* means "rose" and *-m* indicates the complex of functions that are designated as accusative. If, however, we consider the dative plural of the same word, *rosis*, we will not find *rosa-* any more. The only common element is *ros-* and so *-is* should be interpreted as the mark of the dative plural. But then must we not analyse *rosam* as *ros-am* and *rosa* as *ros-a* with the same root as in *ros-is*? In this case it would be necessary to analyse *clava* "club" as *clav-a*. But, following the same principles, it would be necessary to analyse *clavus* "nail" as *clav-us*, the *-us* of *clavus* and the *-a* of *clav-a* being different indications of the nominative; "club" and "nail" would be homonyms of the form *clav-*, since what follows *clav-* in these two cases would correspond to the different inflectional satellites. Of course this point of view is not acceptable because "club," that is, in Latin, *clav-* is followed by certain endings and "nail" *clav-* is followed by other endings. The nucleus and the satellites overlap formally. There would be analogous difficulties if one were to try to separate a particular segment for each of the satellites, if for example, one regarded the *-is* of *rosis* as a succession *-i-s* where *-i* would mark the dative and *-s* the plural.

In French as it is spoken and heard, the difference between singular and plural subjects in *il peut* and *ils peuvent* is marked only by the *-v-* of *peuvent* with, accessorially, a difference in the sound of the vowel written *-eu-*. To get past the problem, it is said that the verb is in the plural. But the plurality is not that of the action but that of the subject, as subject which is either a distinct nucleus (a noun), or a satellite (a pronoun) which, being set ahead of the verb, is treated in writing as a word.

In English, the difference between *he cuts* and *he cut* is in the *-s* in the first form. Formally, the past is recognized by the absence of this *-s*. But, of course, this absence can be taken as the mark of the past only for the third person singular and it functions this way only for those few verbs which the accidents of phonetic evolution make formally identical in the present and past tenses.

Faced with the problems posed by *rosam* and *rosis*, the first reaction will be conservative. Any attempt to analyse an inflected form is vain. The only valid analysis is that which ends in words and it is not necessary to go beyond that. Behind a form like *rosis*, one finds, of course, besides the notion of "rose," those so-called "categories" of dative (or ablative) and the plural. Since the analysis into successive segments comes up against difficulties which cannot be surmounted except by arbitrary decisions, it is better here to disregard completely all questions of sequence and to consider the word as a complex of meaningful features in the same way that, on the level of distinctive units, the phoneme is recognized as a complex of relevant phonic features. It is understood that the phoneme *p* is not physically homogeneous from the beginning to the end of its enunciation: a *p* implies an initial closing, a holding the speech organs and a brusque release, and all this does not prevent it from being linguistically indivisible. The *ch* of Spanish *mucho* begins with *t* and ends with a hush. The *a* of English *pane* begins with the *e* of *pen* and ends with something bordering upon *y*, neither one is less than a single unit. One could say as much for *rosis* by postulating the concomitance of its semantic components.

Appealing as this methodological point of view may be, it will not however, solve our problems and this simply because the difficulties we encounter when we try to analyse statements into segments corresponding to minimal signs, exist not only in the body of the word but, as we have seen, in complexes where, rightly or wrongly, tradition sees a succession of distinct words: *il peut, ils peuvent* written in two words. The argument will be advanced, to justify this tradition, that the pronouns are inflected (singular *il*, plural *ils*, with a different pronunciation in *il aime* and *ils aiment*); they are therefore the nuclei of satellites, and one would balk at treating these nuclei as the satellites of other nuclei.

## *The Word*

In fact, what is largely operative here is the reluctance, inherited from the origins of the language, to inflect the words at the beginning. Where French writes the three words *je le porte*, Basque does not hesitate to write *dakart* for the spoken *da-kar-t*, in which *da-* corresponds to *le*, *kar-* to *porte* and *-t* to *je*. It was pointed out long ago that if French had no traditional spelling some explorer-linguist from another continent would be tempted to write as a single word, *jaluidonne*, that which is actually spelled out *je la lui donne*.

We are touching here on what renders the notion of the word so suspect to all true linguists. They cannot accept traditional writing without verifying first whether it reproduces faithfully the true structure of the language which it is supposed to record. If the term word is to be kept to designate the segments of speech which appear separately in writing, then this can be only insofar as the spaces in the text correspond to a well defined type of limit and encompass all groups with a certain degree of homogeneity. In fact, we know very well today that this is not the case. We should hence try to define the word without aiming at a formulation that would permit tracing down all the "words" of all existing writing systems. In other words, it should be decided in which case one should posit one word and in which case several words, and this not only for French or Basque but for French, Basque and all languages.

But if it is finally necessary to discard tradition, can we be sure that it will be worthwhile positing a meaningful unit, called word by definition, intermediate between the minimal sign and the clause such that any utterance can be analysed into an uninterrupted succession of such units? Is anything to be gained by employing the same term to designate satellites such as *dans* or *là*, isolated nuclei like *donne*, *château* and complexes like *donnerions*, *timbre-poste* and *anticostitutionnellement*?

There is an entire category of facts which we have refrained from including in the above considerations namely the phonic features, whose function is no longer to establish identities of the significant units of the sequence by opposing them to those that could appear in their place, but to mark the individuality of those units in relation to their neighboring units in a speech. Among these characteristics, often called demarcatives, it is neces-

sary above all to note the accent. But it should not be forgotten that certain phonemes or groups of phonemes that only appear at the limits of significant units function equally as demarcatives: in spoken German, *b* is likely to mark the beginning of what we have designated as a nucleus. In many languages accent is clearly demarcative. In Czech, for example, it is regularly on the initial syllable of what corresponds to the written word. Elsewhere, in English, in Italian, in Russian, its place is not fixed in the relationship to the limits of the word, but the presence of an accent at a certain point in the sequence indicates the presence of a significant unit which is often what one designates as a word.

All this explains why it is often considered that the word has a physical individuality which would help pointing out the articulation of speech even before the meaning intervenes. Yet, if it is true that accent, and, to a lesser degree, the other demarcative characteristics, are precious adjuncts for the listener in his unconscious analysis of the utterances he hears, it cannot be claimed that it characterizes each of the elements that spelling spaces. Prepositions, regarded as distinct words, can only rarely be accented. When they are, it is generally because they receive the accent of the following substantive or pronoun (Russian *pód nosom*, "under the nose," Czech *dó školy*, "to school," Danish *fór hende*, "for her,") in such a way that the preposition and what follows it form a single "accentual word" which does not coincide with the written words.

In quite another connection, in order to delimit the successive units of speech, a calculation can be made of the probability of each minimal segment (phoneme or letter). Let us take a statement beginning with *il apparât...* It is clear that an initial *i-* has some chance of being followed by *-l-* because the pronoun *il* is frequent initially; thus let us say that after *i-*, *-l-* is more likely than any other letter. After the *-l-* of *il-*, one can expect some verb and in consequence *-a-* is not more likely than any other letter. If then we were to represent the probability by a curve, this curve which rose in passing from *i-* to *-l-*, will fall between *-l-* and *-a-*. From *-a-* to *-p-*, the curve rises a little since a consonant is here a little more likely than a vowel. From the first *-p-* to the second, the ascent is sharpened, because most

## The Word

French words beginning with *ap-* take a second *p*. From the second *-p-* to *-a-*, the curve drops a little because *-r-*, *-l-* or any other vowel can be expected there, but far less than when passing from *il-* to *-app-*. Gradually as one advances within the word the curve rises, and when *-i-* is reached the probability that the following letter may be *-t-* becomes a certainty. After *-apparâit-*, the probability of an inflectional segment such as *-ra* or *-rait* is quite high but that of any other letter but *-r-* is very low since one can expect almost anything at this point. If *-d-* follows *-apparâit-*, one notes a sharp fall in the curve of probability between *-t-* and *-d-*. We see that falls coincide with the passage of the word *il* to the word *apparâit* and from *apparâit* to what follows, unless this is *-ra* or *-rait*, elements which belong to the same word (*apparâitra*, *apparâitrait*). There could therefore be a way to determine scientifically, on the basis of an exhaustive preliminary study of the probability of occurrence of letters or phonemes, the limits between the segments of an utterance which would coincide perhaps well enough with written words. It will be noted with interest that endings or suffixes like *-ra* and *-rait* above have, in the same sequence, a fairly high probability, whereas the choice of a proposition like *de* or *à* does not permit us to predict the substantive or adjective that is to follow. The sharp descent of the curve of probability that follows here coincides well with the word status of prepositions. On the other hand, the probability that the definite article will follow such prepositions is considerable; this would account for the fact that each time the substantive is masculine or plural, *de* and *à* form with the following article the unique words *du*, *des*, *au*, *aux*.

It is clear that the considerations based on the probability of successive elements recall those which have been developed above regarding the different role of the beginning and the end of the word in its identification by the listener. If the curve of probability rises abruptly toward the end of the word it is evidently because everything that has preceded it almost certainly allows identification of the word. However, probability makes it easier to comprehend that there is a natural break between a grammatical element placed in front of a word and the following nucleus and that such a break does not exist between a nucleus and a satellite that follows it.

It is clear, however, that the spaces in a text do not necessarily coincide with falls in the probability curve, one reason for it being that orthography in general and French orthography in particular abound in inconsistencies: the probability of *de terre* occurring after *pomme* is in modern French more likely than *-poste* occurring after *timbre*, but *timbre-poste* is written as one single word while *pomme de terre* is written in three words. If *du* and *au* are written without lifting the pen in contrast to the spacing in *de la*, *à la*, it is evidently because phonetic evolution has combined intimately in one place what it leaves separate in another; the probability of the feminine after the preposition is indeed not less than that of the masculine or the plural. Using the curve of probability to delimit the units in a sequence cannot be considered, because it would be impossible to calculate the actual probability at each point, the context upon which it depends being too vast to be taken into consideration. In addition, a criterion is desirable if its application permits obtaining, if not always perfectly clear-cut classes, at least groups of units forming generally distinct fields of dispersion; this would not be the case where all degrees of probability would be represented: would the probability of *de Sèvres* after *vase* make a word of *vase de Sèvres*? And what might happen in the case of *vase de Soissons*? Could today's news abruptly promote the complex *bijoux de la Bégum* to the level of a word only to let it founder and break into four pieces a few days later?

\*

All efforts to give to the term "word" a really scientific status ran up against the fact that aside from clear-cut cases, there are others where none of the usable criteria permit a yes or no reply. The notion of word, completed by that of enclitic, has permitted us to account for the sentence structure of inflected languages such as Greek or Latin. In other languages it permits useful groupings of certain facts, but its extension to all utterance of all languages complicates grammatical description more often than it simplifies it.

It seems that the solution of the problem might be found replacing, in linguistic practice, "word" by the much more supple concept of "syntagm." This term designates any group of several

## The Word

minimal signs. But it is understood that the minimal signs, conceived as parts of a syntagm, maintain amongst themselves more intimate relations than those that bind them to the rest of the sentence: the sentence *un énorme rocher surplombait la voie ferrée*, naturally breaks into three syntagms: *un énorme rocher*, *surplombait* and *la voie ferrée*.

To be able to speak of groups of minimal signs it is of course necessary to know precisely what a minimum sign is. For those who call the minimum sign a "morpheme," this has to be a particular segment of the utterance: *donnerons* is broken up easily into three morphemes, each with its separate pronunciation (*donn-*, *-(e)r-*, *-ons*), each with its own contribution to the meaning of the message. But it often happens that a segment cannot be divided up so easily, although it contains two distinct meanings: the *-arum* of *rosarum* "of roses" expresses the genitive relationship on one hand and, on the other, the plural. In a case of this kind one may either give up the analysis and call the unit a genitive plural morpheme or risk an arbitrary analysis (*-a-rum*, *-ar-um* or *-aru-m*) which will satisfy no one. Elsewhere, it may be easy to see where to make a cut but it will not be more acceptable because it will not comply with what the meaning requires. Faced with the singular *animal* and the plural *animaux*, the analyst is tempted to see in *-al* the mark of the singular and in *-aux* that of the plural. But the segment which corresponds to the meaning "animal" is not *anim-* but *animal* and *animaux*, and it would therefore seem preferable to look for the expression of the plural in the article *les* or *des* which normally precedes *animaux*. However, since there are contexts in which *animaux* is used without an article, one cannot help considering *animaux* an amalgam of two minimal signs.

When one experiences some difficulty in isolating, localizing and identifying a segment which corresponds to a minimum sign one often designates the latter as a category. This is a slackness consecrated by usage. Is the French definite article a "category," given the difficulty one has trying to locate it precisely within *au*, *aux*? If a declension case is a "category," is the Basque comitative also one, despite its being uniformly expressed by means of an isolable element, *-ekin*? It is doubtless necessary, wherever the presence of a new meaningful element manifests

itself by means of a formal change of any sort, to diagnose a minimum sign even if it is not possible, without being arbitrary, to attribute to this sign a specific segment of the sentence: in *rosarum* there is the sign of the genitive and the sign of the plural even if one is not able to localize exactly one or the other. We can say in this case that there is an amalgam of what designates the genitive and what designates the plural. On the other hand, since, as we have seen, it is not easy to isolate the root in this case, one shall treat the complex *rosarum* as an amalgam, and the inflected word of classical languages will often be defined as an amalgamated syntagm.

To anticipate the ambiguities that plague the term "morpheme" and to make quite clear that analysis should, on occasion, go beyond the physical segmentation, it is useful to designate the minimum sign as a moneme.

All utterances break down into monemes; most often they are successive but occasionally they are amalgamated. There is a moneme every time that the speaker is forced to give to his utterance a particular turn in order to convey exactly the message that he has in mind and not some other message that the language might permit him to transmit. There are as many monemes as there are choices. The French phrase *Au fur et à mesure* which seems so easy to analyse, represents without question five traditional words. But it is nothing more than a moneme because once the speaker has chosen to use *fur*, he cannot refrain from uttering the remainder of the phrase.

The analysis of statements into monemes is not a simple operation and it will not necessarily meet with success. The criterion of choice does not apply when the problem involves compounds and derivatives chosen all at once and not by successive detached elements. When one says *telephone* or *thermometer* one chooses *telephone* or *thermometer* without worrying about a possible meaning of *tele-* and *-phone*, *thermo-* and *-meter*. The particular value of these elements can only be revealed when *telephone* is compared with *television* and *grammophone* and *thermometer* with *thermoscope* and *barometer*. But no linguistic description can avoid the task of an analysis permits the analysis to go farthest.

down to the minimum sign, and the moneme is the unit which



## *The Word*

We are now able to formulate more precisely what we said above about the syntagm: two or more than two monemes which have more intimate relationships with each other than with those that bind them to the rest of the statement form a syntagm. Not everything, in this sense, is a syntagm in a statement. *You leave me* is not a syntagm but three monemes which form an utterance. On the subject of the syntagm, we can here reply negatively to the question asked above concerning the word: there is no point whatsoever in postulating, as intermediate between the moneme and the complete minimum utterance, which is the sentence, a troublesome unit, of which each segment of the statement necessarily is a part. The linguist is free to operate with syntagms wherever it will contribute to clarity. He is also free to work with words wherever the structure of the language under study appears to demand that the emphasis be put on the formal and semantic coherence of certain syntagms. But what must be avoided is confounding, under the same term, elements as functionally disparate as prepositions, bare lexical radicals and complexes in which the lexical element is surrounded by grammatical modifiers and accompanied by indications of its relationships with the context.



What a contemporary linguist can say about the word well illustrates to what extent the functional and structural research of the last thirty-five years has had to revise traditional concepts in order to lay the scientific foundations for the observation and description of language. Certain applications of linguistics, such as the research applying to mechanical translation, by the emphasis which they place on the written form of the language might seem to lend importance to spaces in the written text and lead us to forget that it is from speech that one should always start in order to understand the real nature of human language. Thus, more than ever, it is indispensable to insist on the necessity of pushing the examination beyond the immediate appearances and beyond the structures most familiar to the researcher. The fundamental traits of human language are frequently to be found behind the screen of the word.