Lexis as a Linguistic Level

M. A. K. HALLIDAY

At a time when few linguists, other than lexicographers themselves, devoted much attention to the study of lexis, and outlines of linguistics often contained little reference to dictionaries or to other methods in lexicology, J. R. Firth repeatedly stressed the importance of lexical studies in descriptive linguistics. He did not accept the equation of 'lexical' with 'semantic', and he showed that it was both possible and useful to make formal statements about lexical items and their relations. For this purpose Firth regarded the statement of collocation as the most fruitful level. General views on the levels of linguistic analysis, to the 'collocational patterns in language, and to suggest that it may be helpful to devise an approach, and he sometimes referred, within the framework of his general views on the levels of linguistic analysis, to the 'colloca-

tional level'.

The aim of this paper is to consider briefly the nature of lexical patterns in language, and to suggest that it may be helpful to devise methods appropriate to the description of these patterns in the light of a lexical theory that will be complementary to, but not part of, grammatical theory. In other words, the suggestion is that lexis may be usefully thought of (a) as within linguistic form, and thus standing in the same relation to (lexical) semantics as does grammar to (grammatical) semantics, and (b) as not within grammar, lexical patterns thus being treated as different in kind, and not merely in delicacy, from grammatical patterns. This view is perhaps implicit in Firth's recognition of a 'colloca
tional level'.

One of the major preoccupations of grammatical theory in present-day linguistics is the extension of grammatical description to a degree of delicacy greater than has hitherto been attained, and it is rightly claimed as a virtue of contemporary models that they permit more delicate statements to be made without excessive increase in complexity. A grammar is expected to explain, for example, the likeness and unlikeness between this brush won't polish and this floor won't polish, the three-way ambiguity of John made Mary a good friend, and the non-acceptability of beautiful hair was had by Mary beside the acceptability of the last word was had by

Mary. Such explanations require the recognition of distinctions which, as is well known, begin to cut across each other at a relatively early stage in delicacy, and the model has to accommodate cross-classification of this kind. The form of statement adopted (and the terminology) will of course depend on the model; but it is generally agreed that all such patterns need to be accounted for.

As part of the process of accounting for these distinctions the grammar attempts, both progressively and simultaneously, to reduce the very large classes of formal items, at the rank at which they can be most usefully abstracted (for the most part generally as words, but this is merely a definition truth from which we learn what 'word' means), into very small sub-classes. No grammar has, it is believed, achieved the degree of delicacy required for the reduction of all such items to one-member classes, although provided the model can effectively handle cross-classification it is by no means absurd to set this as the eventual aim: that is, a unique description for each item by its assignment to a 'microclass', which represents its value as the product of the intersection of a large number of classificatory dimensions.

If we take into account the amount of information which, although it is still far from having been provided for any language, contemporary grammatical models can reasonably claim to aim at providing, there would seem to be two possible evaluations of it. One is that, when the most delicate distinctions and restrictions in grammar have been explained, all formal linguistic patterns will have been accounted for; what is left can only be accounted for in semantic terms. The second is that there will still remain patterns which can be accounted for in formal linguistic terms but whose nature is such that they are best regarded as non-grammatical, in that they cut across the type of relation that is characteristic of grammatical patterning.

The particular model of grammar that is selected may suggest, but does not fully determine, which of these two views is adopted operationally. For example, a model which distinguishes sharply between the grammar of a language and the use of the grammar, regarding corpus-based statistical statements as proper only to the latter, and therefore as outside the range of validity of a descriptive statement, is less easily compatible with the second view than is a model which does not make this distinction and which allows statistical statements a place in linguistic description; nevertheless it is not wholly incompatible with it.

Lexical statements, or 'rules', need not be statistical, or even corpus-based, provided that their range of validity is defined in some other way,
as by the introduction of a category of ‘lexicalness’ to parallel that of grammaticalness.

One may validly ask whether there are general grounds, independent of any given model, for supplementing the grammar by formal statements of lexical relations, at least (given that the aim of linguistics is to account for as much of language as possible) until these are shown to be unnecessary. It may be a long time before it can be decided whether they are necessary or not, in the sense of finding out whether all that is explained lexically could also have been incorporated in the grammar; there still remains the question whether or not it could have been explained more simply in the grammar. The question is not whether formal lexical statements can be made; they are already made in dictionaries, although at a low level of generality, in the form of citations. The question of interest to linguists is how the patterns represented by such citations are to be stated with a sufficient degree of abstraction, and whether this can best be achieved within or outside the framework of the grammar.

Let us consider an example. The sentence he put forward a strong argument for it is acceptable in English; strong is a member of that set of items which can be juxtaposed with argument, a set which also includes powerful. Strong does not always stand in this same relation to powerful: he drives a strong car is, at least relatively, unacceptable, as is this tea's too powerful. To put it another way, a strong car and powerful tea will either be rejected as ungrammatical (or unlexical) or shown to be in some sort of marked contrast with a powerful car and strong tea; in either case the paradigmatic relation of strong to powerful is not a constant but depends on the syntagmatic relation into which each enters, here with argument, car or tea.

Grammatically, unless these are regarded as different structures, which seems unlikely, they will be accounted for in a way which, whatever the particular form of statement the model employs, will amount to saying that, first, strong and powerful are members of a class that enters into a certain structural relation with a class of which argument is a member; second, powerful (but not strong) is a member of a class entering into this relation with a class of which car is a member; and third, strong (but not powerful) is a member of a class entering into this relation with a class of which tea is a member. It would be hoped that such classes would reappear elsewhere in the grammar defined on other criteria. Argument, car and tea will, for example, already have been distinguished on other grounds on the lines of ‘abstract’, ‘concrete inanimate’ and 'mass'; but these groupings are not applicable here, since we can have a strong table and powerful whisky, while a strong device is at least questionable.

The same patterns do reappear: he argued strongly, I don't deny the strength of his argument, his argument was strengthened by other factors. Strongly and strength are paralleled by powerfully and power, strengthened by made more powerful. The same restrictions have to be stated, to account for the power (but not the strength) of his car and the strength (but not the power) of her tea. But these involve different structures; elsewhere in the grammar strong, strongly, strength and strengthened have been recognized as different items and assigned to different classes, so that the strong of his argument has been excluded on equal terms with the strong of his car. Strong and powerful, on the other hand, have been assigned to the same class, so that we should expect to find a powerful car paralleled by a strong car. The classes set up to account for the patterns under discussion either will cut across the primary dimension of grammatical classification or will need to be restated for each primary class.

But the added complexity involved in either of these solutions does not seem to be matched by a gain in descriptive power, since for the patterns in question the differences of (primary) class and of structure are irrelevant. Strong, strongly, strength and strengthened can all be regarded for this present purpose as the same item; and a strong argument, he argued strongly, the strength of his argument and his argument was strengthened all as instances of one and the same syntagmatic relation. What is abstracted is an item strong, having the scatter strong, strongly, strength, strengthened, which collocates with items argue (argument) and tea; and an item power (powerfully) which collocates with argue and car. It can be predicted that, if a high-powered car is acceptable, this will be matched by a high-powered argument but not by high-powered tea. It might also be predicted, though with less assurance, that a weak argument and weak tea are acceptable, but that a weak car is not.

As far as the collocational relation of strong and argue is concerned, it is not merely the particular grammatical relation into which these items enter that is irrelevant; it may also be irrelevant whether they enter into any grammatical relation with each other or not. They may be in different sentences, for example: I wasn’t altogether convinced by his argument. He had some strong points but they could all be met. Clearly there are limits of relevance to be set to a collocational span of this...
kind; but the question here is whether such limits can usefully be defined grammatically, and it is not easy to see how they can.

The items \textit{strong} and \textit{power} will enter into the same set as defined by their occurrence in collocation with \textit{argue}; but they will also enter into different sets as defined by other collocations. There is of course no procedural priority as between the identification of the items and the identification of the paradigmatic and syntagmatic relations into which they enter: ‘item’, ‘set’ and ‘collocation’ are mutually defining. But they are definable without reference to grammatical restrictions; or, if that is begging the question, without reference to restrictions stated elsewhere in the grammar. This is not to say that there is no interrelation between structural and collocational patterns, as indeed there certainly is; but if, as is suggested, their interdependence can be regarded as mutual rather than as one-way, it will be more clearly displayed by a form of statement which first shows grammatical and lexical restrictions separately and then brings them together.\(^7\) If therefore one speaks of a lexical level, there is no question of asserting the ‘independence’ of such a level, whatever this might mean; what is implied is the internal consistency of the statements and their referability to a stated model.

Possible methods of lexical analysis, and the form likely to be taken by statements at this level, are the subject of another paper, by J. McH. Sinclair, appearing in the present volume.\(^8\) Here I wish to consider merely some of the properties of this type of pattern in language, and some of the problems of accounting for it. Clearly lexical patterns are referable in the first place to the two basic axes, the syntagmatic and the paradigmatic. One way of handling grammatical relations on these two axes is by reference to the theoretical categories of ‘structure’ and ‘system’, with the ‘class’ definable as that which enters into the relations so defined. In lexis these concepts need to be modified, and distinct categories are needed for which therefore different terms are desirable.

First, in place of the highly abstract relation of structure, in which the value of an element depends on complex factors in no sense reducible to simple sequence, lexis seems to require the recognition merely of linear co-occurrence together with some measure of significant proximity, either a scale or at least a cut-off point. It is this syntagmatic relation which is referred to as ‘collocation’. The implication that degree of proximity is here the only variable does not of course imply how this is to be measured; moreover it clearly relates only to statements internal to the lexical level; in lexicogrammatical statements collocational restrictions intersect with structural ones. Similarly in place of the ‘system’ which, with its known and stated set of terms in choice relation, lends itself to a deterministic model, lexis requires the open-ended ‘set’ assignment to which is best regarded as probabilistic. Thus while a model which is only deterministic can explain so much of the grammar of a language that its added power makes it entirely appropriate for certain of the purposes of a descriptive grammar, it is doubtful whether such a model would give any real insight into lexis. Collocational and lexical set are mutually defining as are structure and system: the set is the grouping of members with like privilege of occurrence in collocation.

Second, in grammar a ‘bridge’ category is required between element of structure and term in system on the one hand and formal item on the other; this is the class. (This specific formulation refers to the ‘scale-and-category’ version of a system-structure model; but it is probably true that all models make use of a category analogous to what I am here calling the ‘class’.) In lexis no such intermediate category is required: the item is directly referable to the categories of collocation and set. This is simply another way of saying that in lexis we are concerned with a very simple set of relations into which enter a large number of items, which must therefore be differentiated qua items, whereas in grammar we are concerned with very complex and variable relations in which the primary differentiation is among the relations themselves: it is only secondarily that we differentiate among the items, and we begin by ‘abstracting out’ this difference. In other words there is a definable sense in which ‘more abstraction’ is involved in grammar than is possible in lexis.

Third, the lexical item is not necessarily coextensive on either axis with the item, or rather with any of the items, identified and accounted for in the grammar. For example, on the paradigmatic axis, in \textit{she made up her face} one can identify a lexical item \textit{make up\(_1\)} whose scatter and collocational range are also illustrated in \textit{your complexion needs a different makeup}. This contrasts with the lexical item \textit{make up\(_2\)} in \textit{she made up her team} and \textit{your committee needs a different makeup}. That the distinction is necessary is shown by the ambiguity of \textit{she made up the cast}, \textit{she was responsible for the makeup of the cast}. Grammatically, the primary distinction is that between \textit{made up} and \textit{makeup}; this distinction of course involves a great many factors, but it also relates to many other items which are distinguishable, by class membership, in the same way. If the grammar is at the same time to handle the distinction between \textit{make up\(_1\)} and \textit{make up\(_2\)} it must recognize a new and independent dimension.
LEXIS AS A LINGUISTIC LEVEL

some grammatical models, as has been noted, envisage that it is the grammar's task to distinguish strong from powerful as well as to distinguish a from the and 'past' from 'present'; while a lexicographical model in which a and the, as well as strong and powerful, are entered in the dictionary and described by means of citations could be regarded as in a similar way attempting to subsume grammar under lexis. Even where the model recognizes two distinct kinds of pattern, these still represent different properties of the total phenomenon of language, not properties of different parts of the phenomenon; all formal items enter into patterns of both kinds. They are grammatical items when described grammatically, as entering (via classes) into closed systems and ordered structures, and lexical items when described lexically, as entering into open sets and linear collocations. So in a strong cup of tea the grammar recognizes (leaving aside its higher rank status, for example as a single formal item expounding the unit 'group') five items of rank 'word' assignable to classes, which in turn expound elements in structures and terms in systems; and the lexis recognizes potentially five lexical items assignable to sets.

But, to take a further step, the formal items themselves vary in respect of which of the two kinds of pattern, the grammatical or the lexical, is more significant for the explanation of restrictions on their occurrence qua items. The items a and of are structurally restricted, and are uniquely specified by the grammar in a very few steps in delicacy; collocationally on the other hand they are largely unrestricted. For the item strong, however, the grammar can specify uniquely a class (sub-class of the 'adjective') of which it is a member, but not the item itself within this class; it has no structural restrictions to distinguish it from other members of the class (and if the members of its 'scatter' strong, strength, etc., turn out to operate collocationally as a single item then this conflated item is not even specifiable qua class member); collocationally, however, it is restricted, and it is this which allows its specification as a unique item. There might then appear to be a scale on which items could be ranged from 'most grammatical' to 'most lexical', the position of an item on the scale correlating with its overall frequency ranking. But these are three distinct variables, and there is no reason to assume a correlation of 'most grammatical' with either 'least lexical' or 'most frequent'. The 'most grammatical' item is one which is optimally specifiable grammatically: this can be thought of as 'reducible to a one-member class by the minimum number of steps in delicacy'. Such an item may or may not be 'least lexical' in the sense that there is no collocational environment...
in which its probability of occurrence deviates significantly from its unconditioned probability.

In a lexical analysis it is the lexical restriction which is under focus: the extent to which an item is specified by its collocational environment. This therefore takes into account the frequency of the item in a stated environment relative to its total frequency of occurrence. While *a* and *of* are unlikely to occur in any collocationally generalizable environment with a probability significantly different from their overall unconditioned probabilities, there will be environments such that *strong* occurs with a probability greater than chance. This can be regarded, in turn, as the ability of *strong* to 'predict' its own environment. As extreme cases, *fro* and *spick* may never occur except in environments including respectively *to* and *span* (the fact that *to* and *fro* accounts for only a tiny proportion of the occurrences of *to*, while *spick* and *span* may account for all occurrences of this item *span*, is immaterial to the specification of *fro* and *spick*); here it is likely that, for this very reason, *to* and *fro* and *spick* and *span* are to be regarded as single lexical items.

It is the similarity of their collocational restriction which enables us to consider grouping lexical items into lexical sets. The criterion for the definition of the lexical set is thus the syntactic (downward) criterion of potentiality of occurrence. Just as the grammatical system (of classes, including one-item classes) is defined by reference to structure, so the lexical set (of items) can be defined by reference to collocation. Since all items *can* be described lexically, the relation of collocation could be regarded as being, like that of structure, chain-exhausting, and a lexical analysis programme might well begin by treating it in this way; but this is not a necessary condition of collocation, and if closed-system items turn out, as may be predicted, to be collocationally neutral these items could at some stage be eliminated by a 'deletion-list' provided either by cross-reference to the grammar or, better, as a result of the lexical analysis itself. Once such 'fully grammatical' items are deleted, collocation is no longer a chain-exhausting relation.

Moreover while grammatical structures are hierarchically ordered, so that one can recognize a scale of 'rank' each of whose members is a chain-exhausting unit (text items being then fully accounted for in sentence structure and again in clause structure and so on), it does not seem useful to postulate such an ordered hierarchy for lexis. Lexical items may indeed enter into a sort of rank relation: it is likely, for example, that on collocational criteria we would want to regard *stone*, *grindstone* and *nose to the grindstone* each as a separate lexical item, and though triads of this kind may be rare it looks as though we need the categories of 'simple' and 'compound', and perhaps also 'phrasal', lexical item, in addition to 'collocational span', as units for a lexical description. Since the only 'structural' relation in lexis is one of simple occurrence, these represent a single serial relation: the item *stone* enters (say) into the collocation *grindstone*, which then does not itself collocate exactly like the sum of its parts but enters as an item into (say) the collocation *nose to the grindstone*, which likewise does not collocate like the sum of its parts but enters as an item into (say) the collocation *he's too lazy to keep his nose to the grindstone*. The first stage of such compounding yields a morphological (upward) grouping of items, the 'lexical series' which, like its analogue in grammar, may or may not coincide with the syntactic grouping recognized as a 'set': *oaktree* *ashtree* *planetree* *beechtree* presumably do operate in the same set, while *inkstand* *bandstand* *hallstand* *grandstand* almost certainly do not. The 'series' is formed of compound items having one constituent item in common; this item, here *tree* and *stand*, is the 'morphologically unmarked' member of the series and, likewise, if the series forms a set it may or may not be the 'syntactically unmarked' member of the set. Equivalence or non-equivalence between series and set is an interesting feature of lexical typology: one would predict that in Chinese, for example, practically all such series do form sets (with an unmarked member), whereas in Malay and English they very often do not.

The lexical item itself is of course the 'type' in a type-token (item-occurrence) relation, and this relation is again best regarded as specific to lexis. The type-token relation can be made dependent on class membership: just as in grammar two occurrences assigned to different primary classes, such as *ride* (verb) and *ride* (noun), can be regarded as different (grammatical) items, so in lexis two occurrences assigned to different primary sets can be regarded as different lexical items. This can then be used to define homonymy: if the two occurrences of *model* in the example above are shown to differ according to criteria which would assign them to different sets then they represent two homonymous items. It is not to be assumed, of course, that grammatically distinguished items such as *ride* (verb) and *ride* (noun) may not also operate as distinct lexical items, as indeed they may; merely that if they turn out to belong to the same set they will on that criterion be said to constitute a single lexical item, as also will *strong*, *strength*, *strongly* and *strengthen*, and perhaps also (if they can be suitably delimited) non-cognate 'scatters' such as *town* and *urban*. This would provide a basis for
deciding how many lexical items are represented by ‘expressions’ such as form, stand and term.

If we say that the criterion for the assignment of items to sets is collocational, this means to say that items showing a certain degree of likeness in their collocational patterning are assigned to the same set. This ‘likeness’ may be thought of in the following terms. If we consider \( n \) occurrences of a given (potential) item, calling this item the ‘node’, and examine its ‘collocates’ up to \( m \) places on either side, giving a ‘span’ of \( 2m \), the \( 2mn \) occurrences of collocates will show a certain frequency distribution. For example, if for 2,000 occurrences of \( \text{sun} \) we list the three preceding and three following lexical items, the 12,000 occurrences of its collocates might show a distribution beginning with \( \text{bright, hot, shine, light, lie, come out} \) and ending with a large number of items each occurring only once. The same number of occurrences of \( \text{moon} \) might show \( \text{bright, full, new, light, night, shine} \) as the most frequent collocates.

On the basis of their high probability of occurrence (relative to their overall frequency) in collocation with the single item \( \text{sun} \), the items \( \text{bright, hot, shine, light, lie, come out} \) constitute a weak provisional set; this resembles the weak provisional class recognizable in the grammar on the basis of a single ‘item-bound’ substitution frame—although in lexis it is relatively less weak because of the lower ceiling of generality: lexis is more item-bound than grammar. If we intersect these with the high frequency collocates of \( \text{moon} \) we get a set, whose members include \( \text{bright, shine and light} \), with slightly greater generality. That is to say, \( \text{bright, shine and light} \) are being grouped together because they display a similar potentiality of occurrence, this being now defined as potentiality of occurrence in the environment of \( \text{sun} \) and in that of \( \text{moon} \). The process can be repeated with each item in turn taken as the node; that is, as the environment for the occurrence of other items. The set will finally be delimited, on the basis of an appropriate measure of likeness, in such a way that its members are those items showing likeness in their total patterning in respect of all those environments in which they occur with significant frequency.

This is of course very much oversimplified; it is an outline of a suggested approach, not of a method of analysis. As Sinclair has shown, however, methods of analysis can be developed along these lines. Many other factors are involved, such as the length of the span, the significance of distance from the node and of relative position in sequence, the possibility of multiple nodes and the like. One point should be mentioned here: this is the importance of undertaking lexicogrammatical as well as lexical analysis. It is not known how far collocational patterns are dependent on the structural relations into which the items enter. For example, if a cozy discussion is unlikely, by comparison with a cozy chat and a friendly discussion, is it the simple co-occurrence of the two items that is unlikely, or their occurrence in this particular structure? All that has been said above has implied an approach in which grammatical relations are not taken into account, and reasons have been given for the suggestion that certain aspects of linguistic patterning will only emerge from a study of this kind. But it is essential also to examine collocational patterns in their grammatical environments, and to compare the descriptions given by the two methods, lexical and lexicogrammatical. This then avoids prejudging the answer to the question whether or not, and if so to what extent, the notion of ‘lexicalness’, as distinct from ‘lexicogrammaticalness’, is a meaningful one.9

An investigation on the lines suggested requires the study of very large samples of text. The occurrence of an item in a collocational environment can only be discussed in terms of probability; and, although cut-off points will need to be determined for the purpose of presenting the results, the interest lies in the degree of ‘lexicalness’ of different collocations (of items and of sets), all of which are clearly regarded as ‘lexical’. Moreover the native speaker’s knowledge of his language will not take the form of his accepting or rejecting a given collocation: he will react to something as more acceptable or less acceptable on a scale of acceptability. Likely collocations could be elicited by an inquiry in which the subject was asked to list the twenty lexical items which he would most expect to find in collocation with a given node;10 but the number of such studies that would be required to cover even the most frequent lexical items in the language is very large indeed. Textually, some twenty million running words, or 1,500–2,000 hours of conversation, would perhaps provide enough occurrences to yield interesting results. The difficulty is that, since lexical patterns are of low generality, they appear only as properties of very large samples; and small-scale studies, though useful for testing methods, give little indication of the nature of the final results.

It is hard to see, however, how the results could fail to be of interest and significance for linguistic studies. Their contribution to our knowledge of language in general, and of one language in particular, may perhaps be discussed in relation to the use of the term ‘semantics’. If lexis is equated with semantics, the implication is that lexical patterns can only be described either externally (that is, as relations between
LEXIS AS A LINGUISTIC LEVEL

in the paper already referred to in the present volume), can show in what ways and to what extent the introduction of formal criteria into the study of lexis, as implied by the recognition of a 'lexical level', are of value to any particular applications of linguistics. But there seem to be adequate reasons for expecting the results to be interesting; and if they are, this is yet another indication of the great insight into the nature of language that is so characteristic of J.R. Firth's contribution to linguistic studies.

Notes
2 'Modes of meaning', Essays and Studies (The English Association), 1951 (in Papers in linguistics 1934–1951; pp. 195–196): 'It must be pointed out that meaning by collocation is not at all the same thing as contextual meaning, which is the functional relation of the sentence to the processes of a context of situation in the context of culture.... Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words'. Compare also 'The Use and distribution of certain English sounds', English Studies 17.1 (1935) (in Papers in linguistics 1934–1951; p. 37), in reference to 'lexical substitution-counters' (lexical items): 'This (sc. the lexical) function should not be misnamed semantic'.
3 See 'A Synopsis of linguistic theory', Studies in linguistic analysis (Special Volume of the Philological Society), Oxford, Blackwell (1957), p. 12. In the present paper 'lexical level' has been used in preference to 'collocational level' in order to suggest greater generality and parallelism with the grammatical level.
4 It is also stated explicitly by Firth in 'A Synopsis of linguistic theory', p. 12: 'Collocations of a given word are statements of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order'. Note that here 'order' refers to the 'mutual expectancy' of syntagmatically related categories, such as elements of structure in grammar or phonology, and not to linear sequence: cf. ibid., pp. 5, 17 and my Categories of the theory of grammar', Word 17.3 (1961), pp. 254–255.
5 That is, distinctions are made which involve the recognition of more finely differentiated syntagmatic relations in the grammar, and that these in turn define further sub-classes on various dimensions within previously defined classes.
6 The place of collocational restrictions in a transformational grammar is considered by P. H. Matthews in 'Transformational grammar' (review article), ArchL 13.2 (1961).
7 For a discussion of the relation betweengrammatical and lexical patterns see Angus McIntosh, Patterns and ranges', Lgl 37.3 (1961).
8 See J. McH. Sinclair, 'Beginning the study of lexis', passim.
9 The implication is, in effect, that 'wellformedness' is best regarded as 'lexico-grammaticalness', and that a departure from wellformedness may be ungrammatical, unlexical or unlexico-grammatical. That the last two are distinct is suggested by

160

M. A. K. HALLIDAY

language and non-language, whether approached denotatively or contextually or lexicogrammatically (that is, in dependence on grammatical patterns). This restriction leaves two gaps in our understanding of language: the internal relations of lexis, and the external relations of grammar—that is, lexis (lexical form), and grammatical semantics. But linguistics is concerned with relations of both types, both internal (formal, within language) and external (contextual or 'semantic', between language and non-language); and all linguistic items and categories, whether operating in closed contrasts, like the and a, or 'past' and 'present', or in open ones, like strong and powerful, enter into both. Moreover, as Firth stressed, both these types of relation are 'meaningful': it is part of the meaning of 'past' that it contrasts with 'present', and it is part of the meaning of strong that it collocates with tea. The fact that the labels for grammatical categories are chosen on semantic grounds should not be taken to imply that they represent an adequate substitute for grammatical semantics; but equally the existence of traditional methods in lexical semantics does not mean that lexical items display no internal, formal patterns of their own.

A thesaurus of English based on formal criteria, giving collocationally defined lexical sets with citations to indicate the defining environments, would be a valuable complement to Roget's brilliant work of intuitive semantic classification in which lexical items are arranged 'according to the ideas which they express'. But even such a thing as a table of the most frequent collocates of specific items, with information about their probabilities, unconditioned and lexically and grammatically conditioned, would be of considerable value for those applications of linguistics in which the interest lies not only in what the native speaker knows about his language but also in what he does with it. These include studies of register and of literary style, of children's language, the language of aphasics and many others. In literary studies in particular such concepts as the ability of a lexical item to 'predict' its own environment, and the cohesive power of lexical relations, are of great potential interest. Lexical information is also relevant to foreign language teaching; many errors are best explained collocationally, and items can be first introduced in their habitual environments. A further possible field of application is information retrieval: one research group in this field is at present undertaking a collocational analysis of the language of scientific abstracts.

Only a detailed study of the facts, such as that now being undertaken by Sinclair (the principles and methods worked out by him are described in the paper already referred to in the present volume), can show in what ways and to what extent the introduction of formal criteria into the study of lexis, as implied by the recognition of a 'lexical level', are of value to any particular applications of linguistics. But there seem to be adequate reasons for expecting the results to be interesting; and if they are, this is yet another indication of the great insight into the nature of language that is so characteristic of J.R. Firth's contribution to linguistic studies.

Notes
2 'Modes of meaning', Essays and Studies (The English Association), 1951 (in Papers in linguistics 1934–1951; pp. 195–196): 'It must be pointed out that meaning by collocation is not at all the same thing as contextual meaning, which is the functional relation of the sentence to the processes of a context of situation in the context of culture.... Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words'. Compare also 'The Use and distribution of certain English sounds', English Studies 17.1 (1935) (in Papers in linguistics 1934–1951; p. 37), in reference to 'lexical substitution-counters' (lexical items): 'This (sc. the lexical) function should not be misnamed semantic'.
3 See 'A Synopsis of linguistic theory', Studies in linguistic analysis (Special Volume of the Philological Society), Oxford, Blackwell (1957), p. 12. In the present paper 'lexical level' has been used in preference to 'collocational level' in order to suggest greater generality and parallelism with the grammatical level.
4 It is also stated explicitly by Firth in 'A Synopsis of linguistic theory', p. 12: 'Collocations of a given word are statements of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order'. Note that here 'order' refers to the 'mutual expectancy' of syntagmatically related categories, such as elements of structure in grammar or phonology, and not to linear sequence: cf. ibid., pp. 5, 17 and my Categories of the theory of grammar', Word 17.3 (1961), pp. 254–255.
5 That is, distinctions are made which involve the recognition of more finely differentiated syntagmatic relations in the grammar, and that these in turn define further sub-classes on various dimensions within previously defined classes.
6 The place of collocational restrictions in a transformational grammar is considered by P. H. Matthews in 'Transformational grammar' (review article), ArchL 13.2 (1961).
7 For a discussion of the relation betweengrammatical and lexical patterns see Angus McIntosh, Patterns and ranges', Lgl 37.3 (1961).
8 See J. McH. Sinclair, 'Beginning the study of lexis', passim.
9 The implication is, in effect, that 'wellformedness' is best regarded as 'lexico-grammaticalness', and that a departure from wellformedness may be ungrammatical, unlexical or unlexico-grammatical. That the last two are distinct is suggested by
such examples as sandy hair, sandy gold and sandy desk: sandy desk is unlexical, in that this collocation is unlikely to occur in any grammatical environment, whereas sandy gold is merely unlexicogrammatical: there is nothing improbable about golden sand. An analogous distinction is observable in clichés: in shabby treatment the mutual expectancy is purely lexical, and is paralleled in they treated him shabbily, a shabby way to treat him and so on, whereas the collocation faint praise is restricted to this structure, in the sense that it will not occur with similar probability under other grammatical conditions.

Towards a Prosodic Statement of Vietnamese Syllable Structure

EUGÉNIE J. A. HENDERSON

The beginnings of my professional association with Professor J. R. Firth were coeval with those of my working acquaintance with the Vietnamese language, and the interaction between the two over the years was such as to suggest Vietnamese phonology as a fitting theme for a paper in the present volume.

When I made my first attempts at a phonetic description of Northern Vietnamese twenty years ago the prevailing tendency might fairly be said to be to use ‘one magic phoneme principle within a monosystemic hypothesis’. Such an approach produces ‘phonemic solutions’ that are in many ways unsatisfactory as statements of phonological function, whatever their merits from other points of view. There are aspects of Vietnamese pronunciation and of the distribution of Vietnamese sounds that are a challenge to phonetician and phonologist alike, as is witnessed by the variety in the phonetic descriptions and phonemic solutions that have been offered so far. Certain peculiarities of pronunciation and apparent disparities in distribution are so regular that it is clear that they should properly be viewed as systematic rather than accidental, as integral parts of a coherent whole rather than as irrelevant oddities. The ideal phonological statement would be one in which these ‘apparently eccentric features take a normal place’, one that would not so much solve the apparent problems as provide a framework within which they were found no longer to exist. In the quest for such a statement I have found the outlook and techniques advocated by Firth both provocative and illuminating. It is the aim of this paper, not to provide a definitive phonological statement of Vietnamese syllable structure, for which much further work would be required, but to demonstrate some of the lines of approach suggested by prosodic analysis to specific phonological problems, many of which have close parallels in others of the