R. A. HUDSON

LEXICAL INSERTION IN A TRANSFORMATIONAL GRAMMAR

At some point in the derivation of any sentence it becomes necessary to give idiosyncratic information about particular morphemes (or larger lexical units, such as two-morpheme words like *understand* – but for simplicity we can ignore these in this article). For instance, in generating 'Sheep sleep' a grammar will generate some kind of underlying structure containing just general categories (S, NP, N, etc.) and features ([+plural]), [−past] etc), but this structure would presumably be indistinguishable from that for 'Cows moo', which is obviously a different sentence, needing a different structural description. The differences between them are lexical differences, in that they reside in the particular properties of *sheep*, *cow*, *sleep* and *moo*: in their phonological, morphological, syntactic and semantic properties, to be a little more precise.

This much would presumably be acceptable to all linguists, including all types of transformational linguist. The disagreement arises over the questions of how lexical information should be *inserted* into a structure consisting of general categories and features, and how such information should be *stored* in the grammar (using 'grammar' in this case to include the lexicon or dictionary). Among current transformational proposals, the main difference seems to be that between Chomsky's approach and McCawley's: roughly speaking, Chomsky says that all the bits of idiosyncratic information about a morpheme should be both stored and inserted together, while McCawley says that even if all the information is stored together (and on this he's much less clear than Chomsky) it shouldn't all be inserted together. For example, Chomsky has a single lexical entry for both *refuse* and *refusal* and all the information relevant to either or both words is stored together and also inserted into any sentence-structure together (1970: 190); whereas McCawley would have two separate lexical entries for *cook*, one for the verb and the other for the noun, and would therefore insert only one of them into any given sentence structure, though he might store them together (1968: 581).

I shall argue below that the two approaches complement each other: Chomsky is wrong to insert all the information about a morpheme en bloc into sentence structures, and McCawley is right not to do so; and Chomsky is right to store all the information together, and McCawley would be right

Foundations of Language 14 (1976) 89–107.

to do so too if he could, but I can see no way in which he in fact could, given other aspects of his theory. What we need is the best of both worlds, and I shall offer one suggestion as to how this might be achieved.

I. THE CHOMSKY APPROACH

In 'Aspects of the Theory of Syntax' Chomsky proposed that the grammar should contain a lexicon, which would consist of an unordered list of lexical entries; and each lexical entry would include all the information (phonological, morphological, syntactic and semantic) about the lexical item in question (1965: 84, 87). It is interesting to recall that Katz, Postal and Fodor had previously considered the possibility of having two separate lexicons, one for syntactic and semantic information, the other for morphological and phonological information (Katz and Fodor, 1963: 399; Katz and Postal, 1964: 161; cf. the discussion in Weinreich, 1966: 400). In addition to the lexicon, there should be a lexical rule (*ibid.*: 84) which would select a lexical entry and add its contents to one of the terminal nodes in a deep-structure tree.

As a digression, the way in which Chomsky formulates his lexical rule makes nonsense as it stands, although it is reminiscent of the earlier approach, where the main job of the lexicon was to give phonological shapes for morphemes. His formulation is as follows:

(1) "If Q is a complex symbol of a preterminal string and (D, C) is a lexical entry, where C is not distinct from Q, then Q can be replaced by D." (*ibid.*: 84).

Taken literally, this tells us to *replace* all the syntactic features of the node concerned by the phonological information contained in the lexical entry. This is absurd: it means that every lexical node in the deep structure of a sentence will carry nothing but phonological information, although it obviously *ought* to carry semantic, syntactic and morphological information as well, since these three types of information have to be available to the semantic, transformational and phonological rules. What the lexical rule needs to do is to add the whole of (D, C) to Q, and just delete one out of each pair of identical features. Let us assume that this is what Chomsky really intended. (Since I shall later *criticise* him for suggesting that the whole lexical entry should be added to Q, rather than what he appears to have said, some readers may feel that I'm fighting a straw man, but I think it is hard to see how his own formulation of the lexical rule could possibly mesh with his other proposals in 'Aspects'.)

II. THE PROBLEM OF DISJUNCTIONS

There are a number of problems with the view that I am attributing to Chomsky. The first is that lexical entries can include *disjunctions*, so if one were to literally insert the whole lexical entry into the sentence structure one would have to insert all the alternatives listed in the entry, whether they were relevant or not. For instance, in 'Remarks on Nominalisation' he suggests that wherever a verb has a nominalised form (such as *refuse – refusal*) the two should be covered by a single lexical entry, which would contain at least one disjunction of features: [noun] or [verb] (1970: 190). According to which of these features is chosen the phonological form will be either *refuse* or *refusal*, so clearly there must be a way of eliminating either [noun] or [verb] from the lexical entry, and thereby also eliminating one of the two phonological forms (and also any other syntactic, semantic or morphological peculiarities of these forms), but so far as I can see there is no provision for doing this at present. Moreover, so far as I know there is no rule that distinguishes between words that can be *either* nouns *or* verbs, and those that can be only one or the other (and similarly for every other disjunction). This would follow automatically from a decision never to enter more than one of any set of alternative features.

The obvious place to look for a mechanism to sort out the relevant alternatives from the others is in the condition that the syntactic features of the lexical entry should be 'non-distinct' from those of the node to which it is added. In 'Aspects of the Theory of Syntax', however, there is no discussion of how this principle applies to disjunctive sets. There are presumably just two ways in which we could interpret the definition of distinctness given there (*ibid.*: 81): "two segments are distinct just in case one is positively specified with respect to a feature with respect to which the other is negatively specified, and, more generally, that two matrices with the same number of columns are distinct if the *i*th segment of one is distinct from the *i*th segment of the other." Either we take it literally, and say that the joint lexical item for *refuse/refusal* can't be inserted at a node bearing the feature [+noun], because the feature [−noun] is included among its features (paired with [+verb]); and similarly it can't be inserted as a verb, because of the feature [−verb], so that in effect it can *never* be inserted. Or we take it liberally, and allow entries like *refuse/refusal* to be entered at any node provided that any distinct features are just alternatives to feature-sets that are non-distinct; but in this case, of course, the *whole* of the lexical entry for *refuse/refusal* is bound to be entered whether it is being entered as a noun or as a verb, so there is still no way of telling whether it should be pronounced 'refuse' or 'refusal'.

Having said that there is no provision for selecting among alternatives in Chomsky's theory as it stands, I must admit that there is no difficulty in principle to making such provision: on the contrary, all we need is some simple rewording of the lexical rule such as the following (which also allows for the insertion of more than just the phonological matrix, unlike Chomsky's rule quoted in (1) above).

(2) If Q is a complex symbol of a preterminal string and $(f_1 \ldots f_n)$ is a lexical entry, consisting of a set of phonological, morphological, syntactic and semantic features, $(f_1 \ldots f_n)$ can be added to Q, with the exception of (a) any features in $(f_1 \ldots f_n)$ that are distinct from features in Q, (b) any features in a conjoint relation to features excluded under (a), (c) any features in $(f_1 \ldots f_n)$ that are the same as features in Q.

This rule treats each lexical entry as a simple set of features, where the relations between the features referring to different levels (phonology, morphology, etc.) is simple conjunction; so if there is a syntactic feature in the lexical entry that clashes with one in Q, and this feature is conjunctively related to every other feature in the entry, then nothing can get inserted at Q – in other words, the lexical item concerned can't be inserted at all in that environment. And if there are disjunctive sets of features in the entry, as in the case of *refuse/refusal*, only those features that are compatible with Q will be entered. In this way the lexical rule sorts out the alternative features, as we want, and the problem is solved.

Incidentally, it will be seen that the discussion so far has dealt only with Chomsky's first proposal, according to which contextual features (selectional and strict subcategorisation features) are included among the features attached to the lexical node. It may be, however, that the reader prefers Chomsky's second proposal, (1965: 120), as one surely must (cf. for instance McCawley, 1968: 253; Matthews, 1967: 127), in which the lexical rule is context-sensitive and matches contextual features not against the features on the lexical node but against the context of the lexical node. Whether one adopts this proposal or the earlier one makes little difference: the second proposal would simply need a lexical rule with different wording, making it somewhat more complex. In particular, the features specified under (a) in (2) would have to be respecified as follows:

(3) ... (a) any features in $(f_1 \ldots f_n)$ that are distinct from features in Q, or any contextual features in $(f_1 \ldots f_n)$ that conflict with the context of Q, (b)....

However, contextual features raise rather special problems, and we shall return to them below.

With these changes in the lexical rule we seem to have solved the problem of disjunctions in lexical entries. However, in doing so we have raised another problem: the simplicity of Chomsky's original approach has been spoiled in that it is no longer the case that lexical entries are inserted en bloc, in exactly the form in which they are stored. Originally, the notion lexical entry could be defined with reference *both* to the lexicon *and* to the lexical rule, and there was no conflict between the two definitions: lexical entries were the objects that were entered in an unordered list in the lexicon (corresponding to the lemmas of the dictionary), and they were also the objects that were inserted by the lexical rule. Now, with the revisions I have just proposed, the lexical rule inserts just *parts* of the individual entries in the lexicon, so we may ask whether the notion of the lexical entry is in fact as simple as it seemed to be. The following sections will suggest that it isn't.

III. CONTEXTUAL FEATURES

'Contextual features' is the term Chomsky uses (1965: 93) for features of the form [*X*–*Y*], where *X* and *Y* are strings (perhaps null) of symbols, and the '–' stands for the place in which the lexical entry itself can occur. The clearest examples of contextual features are strict subcategorisation features, such as [___ NP], which is assigned to items that can occur as sisters of an immediately following NP; the other main type is selectional features, such as [[+abstract] Aux –], but since it's not clear what the status of such features is, I'll restrict the discussion to strict subcategorisation features.

The role of contextual features is primarily to define the contexts into which a lexical item can be inserted, so one might well wonder whether it is right that contextual features themselves should be inserted, along with all the other, non-contextual features. After all, once a feature like [–NP] has been referred to by the lexical rule, to check it against the environment of the lexical node, there is no need for it to be referred to again, or so one might think. Moreover, once it has been inserted, there is no provision for removing it from the structure, so it will inevitably still be there in *surface* structure, telling us what the *deep* structure environment of the verb concerned was – which looks like a case of deep structure information 'leaking' through into the surface structure, where it has no business to be, any more than, say, semantic information would. It is therefore tempting to argue that contextual features, unlike all the other features in lexical entries, should not be inserted into the sentences structure, but should simply be available to the lexical rule.

Before arguing the pros and cons of this approach, I should point out that there is no problem in principle in adapting Chomsky's theory to fit it, if it should prove desirable. In his discussion of the second view of lexical insertion, where contextual features are present in the lexical entry but not in the preterminal string, Chomsky points out (1965: 122) that the lexical rule would be a substitution transformation, with the syntactic features of each lexical entry as the structure index for the replacement of *Δ* by the phonological and other features of the same lexical entry. There is no reason why the lexical rule should include the contextual features among the features by which it replaces *Δ*, although Chomsky himself seems to assume the contrary. The passage in question is sufficiently vague for it to make very little difference either way.

The question, then, is whether the features to be inserted should include contextual features and, in particular, strict subcategorisation features. The main advantages of *not* inserting them are firstly that this seems the natural thing to do, for the reasons given above, and secondly that they could be treated as genuine structural indices of the substitution transformation, in the normal way, whereas according to Chomsky's proposal they would have to act *both* as indices for the environment of *Δ and* as part of the complex symbol by which *Δ* is replaced, which makes them odd and problematic. The advantages, on the other hand, of the other approach, in which contextual features are inserted, are also two: firstly, this allows the other features to be divided into disjoint sets according to which contextual features they occur with; and secondly, there do seem to be rules that need the contextual features to be present in surface structure. I shall now explain these two claims.

The first claim rests on the assumption that at least some lexical items have several different contextual features as alternatives, and that minor differences in the other features can be associated with the differences in context. For instance, let us assume, following Chomsky (1972: 174), that (4a) and (b) are not synonymous.

(4a) Bees are swarming in the garden.
(4b) The garden is swarming with bees.

(Bees could be swarming in just one corner of the garden, but in that case (a) would be true but (b) wouldn't.) Let us also assume, as seems reasonable, that both sentences contain the same lexical item *swarm*, and not two otherwise unrelated homophones. In that case, the entry for *swarm* will have to contain at least three sets of features: (i) the phonological, morphological, syntactic and semantic features common to *both* uses of *swarm*, (ii) the syntactic and semantic features of the (a) use, (iii) the syntactic and semantic

features of the (b) use. Now, if the difference in deep structure environments between (4a) and (4b) is as one would expect from their surface structures, it can easily be shown by contextual features on *swarm*: [___] for (4a), [___PP] for (4b); and the relations among the three sets of features listed above can be shown very simply: (set (i), *and* ((set (ii), including [___]) *or* (set (iii), including [___PP])), or, more simply still, (i *and* (ii *or* iii)). However, this is true only if contextual features have the same status within the lexical entry as other types of feature, so that they can be treated as co-members of sets, as in the standard theory.

Let us now consider the alternative, in which contextual 'features' are separate from the other features, in that they define the contexts in which the other features can be inserted but aren't inserted themselves. It's not too clear quite how this proposal would be formalised, but it would probably be easiest not to treat the contextual 'features' as features at all, but rather to treat them as the structural descriptions for the transformations that introduce the features in question. But if that were the case, they would represent properties of the context, in contrast with the properties of the lexical item itself represented by the non-contextual features, and it would make very little sense to think of the two kinds of property as co-members of a single set. Instead, we need ordered pairs of a context with a set of features permitted in that context – but that means that we should have to treat each context as defining a different lexical item, so that the *swarms* in (4) a would be a completely different lexical item from the *swarm* in (4)b, any similarity between them being purely fortuitous. If this is in fact what follows if we separate contextual and non-contextual features, it looks as though we should think twice before doing so.

The second advantage of not separating contextual features from the rest, but of allowing them to be inserted in the normal way, is that at least one transformation, conjunction-reduction, makes use of them. Consider the following:

(5a) John gave Bill a book and a record to Fred.
(5b) *John presented Bill with a book and a record to Fred.

There is some doubt as to the acceptability of (5a), but no doubt as to its relation to (5b): it is much better. How can we explain this difference? In both sentences conjunction-reduction has to delete the subject (*John*) and the main verb (*gave* or *presented*), which are in the same positions in both sentences to which I assume conjunction-reduction has to apply:

(6a) John gave Bill a book and *John gave* a record to Fred.
(6b) John presented Bill with a book and *John presented* a record to Fred.

If the difference between (5a) and (b) was due to differences between the internal structures of the conjoined VP's, then if anything (5a) should be worse than (5b), since the differences are greater: NP–NP versus NP–PP, whereas both conjuncts are NP–PP in (5b).

Thus the difference can't be in the broad outlines of the underlying structures in (6). Instead, I suggest, it is in the features on the verbs *give* and *present*, reflecting the *deep* structures of the conjuncts: the deep structures of the VPs in (5a) and (6a) are both the *same* (V–NP–PP), and the differences between the conjuncts are the result of a transformation (Dative movement) applying in one conjunct but not in the other; whereas in (5b) and (6b) the deep structures are *different*, and the differences can't be attributed to transformations. (It isn't clear how to represent these deep structure differences, so let us represent them as V–NP-with NP versus V–NP-to NP). If contextual features *are* inserted, then they will be among the verb's features when the conjunction-reduction transformation applies, and must therefore be taken into account when deciding whether or not two verbs are identical and can be reduced to a single verb by conjunction reduction. In (6a), *give* was inserted in the context ___NP PP (or NP to NP) in both conjuncts, and therefore has the same contextual feature in both cases, and can be deleted from the second conjunct; but in (6b) the first *present* has the feature [___NP with NP], and the second [___NP to NP], making their features different and therefore not allowing deletion by conjunction-reduction.

The proposal that contextual features are present throughout the derivation, from deep structure to surface structure, leads to a prediction: that words which can have contextual features (mainly verbs and nouns) will be considered identical for transformational purposes only if they occur in the same immediate deep-structure context. All the examples I have considered confirm this hypothesis, though I have to confess it is hard to test where deep structures are in doubt, as they usually are.

Assuming that deep-structure contextual features are indeed needed for transformations, the proposal about disjunctive sets of features in the previous section is confirmed: contextual features are relevant in surface structure only where they represent the *actual* deep structure context of the verb concerned, and not just a deep structure context that it *could* have had. Therefore all contextual features except the relevant one have to be eliminated, as would be done automatically according to the arguments of the previous section.

To summarise, I have tried to show that there is at least an issue to be discussed in connection with the insertion of contextual features, although I have tended to favour the conservative view that contextual features *should* be inserted along with the others. The arguments I have given are anything

but conclusive, and it is to be hoped that better arguments can be found to decide the question one way or the other. If it should turn out eventually that my tentative conclusion is wrong, and that contextual features should not in fact be inserted, this will support the general theme of this article, that the insertion of lexical features requires an apparatus that can *select among* the features in the lexicon; on the other hand of course, if my conclusions here are correct, the general theme will not be weakened.

IV. THE SECOND LEXICAL LOOK-UP

What we have argued so far is that when a lexical item is inserted into deep structure, the rule that inserts it must select among its features in order to weed out all disjunctions, so that if 'refuse/refusal' is inserted as a verb for instance, it gets inserted *only* as a verb, and not also as a noun, or even as a *potential* noun. We have also suggested that it *might* be necessary for this rule also to distinguish between contextual features and non-contextual features, and only insert the latter; but the balance of evidence at present seems to be against this possibility. We now consider evidence for further fragmentation of the lexical entry, this time distinguishing between features on different levels.

In an approach like Chomsky's, where all lexical insertion is meant to be done early on in the derivation of a sentence, it is inevitable that there will be a residue of morphemes in surface structure that need to be given a morphological and phonological shape because they couldn't be looked up in the main lexical look-up for one reason or another (Stockwell *et al.*, 1973: 793). In general, this is the case with most 'grammatical' root morphemes (as opposed to lexical roots), and there are two main reasons why these couldn't be looked up in the first look-up: either because they weren't there when the first look-up happened, as with any morpheme introduced by a transformation (eg. the *do* introduced by do-support); or because they can't reasonably be given any phonological form until after all transformations have applied, which means until surface structure (eg. *I* versus *me* versus *my*).

The arguments for looking up these words so late in the derivation are impeccable; but what now is the relation between the two look-ups? The first look-up inserts sets of *four* types of features (phonological, morphological, syntactic, semantic); and the second inserts sets of *two* types (phonological and morphological), but since both look-ups insert phonological and morphological features, why can't their jobs be more neatly demarcated, so that the first look-up inserts *just* syntactic and semantic features, and the second inserts *all* morphological and phonological features? There can be

no objection in *principle* to this reallocation of features, since the second lexical look-up, as conceived by Stockwell and his colleagues, already inserts phonological (and, presumably, morphological) features at nodes carrying defined complexes of syntactic features, so why not extend the principle to add phonological and morphological information to complexes of syntactic and semantic features? In other words, why not insert the semantic and syntactic features of morphemes like *sheep* in deep structure, and the phonological and morphological features both of morphemes like *sheep*, and of morphemes like *do* and */I/we/*etc., in surface structure? After all, if the syntactic and semantic features of *sheep* are sufficiently precise to distinguish *sheep* from *cow* in deep structure, they could also do the same in surface structure.

V. A REVISED VIEW OF THE LEXICON AND LEXICAL INSERTION

Before considering the pros and cons of the two possible approaches, there is a distinction to be made between having two separate *lexicons* and having two separate *look-up systems*. Taking what I shall call the 'standard' view (that the main look-up is done in the base, and only a tidying-up operation is needed at the level of surface structure) we could interpret it in either of two ways. Either we could take it to mean that there are two separate lexicons, which are consulted at different points in the derivation, or we could take it to mean that there is a single lexicon, containing *all* lexical features, which is consulted at two different stages in the derivation, by a different look-up mechanism each time. The latter interpretation seems to be the one intended by Stockwell and his colleagues, and it is the one I favour (for reasons I shall discuss immediately below), but it has an interesting consequence: that the lexicon, *unlike every other component of the grammar*, has to be brought in at two separate stages in the derivation. By way of contrast the PS rules, for instance, all apply before any other type of rule applies, and then never apply at any later stage. This property of the lexicon raises the question of whether the lexicon really *is* a single component at all, but this is a question that can't be explored here.

The purpose of this digression on the question of one versus two lexicons was to prepare for a discussion of the relative merits of the standard view and what I shall call the 'revised' view (that the first look-up only looks up syntactic and semantic properties, while the second looks up all morphological and phonological properties). It was necessary first to show that either view was compatible with the idea of a single lexicon, since this is the only way to make the radical view seem at all plausible: the alternative would have been to say that every set of syntactic and semantic features would have to

be listed twice, once in the 'upper' lexicon (whence they were extracted for insertion in deep structure) and once in the 'lower' lexicon, where they would each define the conditions for inserting some combination of morphological and phonological features.

Fortunately this duplication is unnecessary: what we need is a single lexicon, containing all types of lexical features in sets, but also divided into levels so that each insertion rule can insert just features from a defined level or levels. Thus one rule would insert syntactic and semantic features that are shown in the lexicon to be compatible with each other and with the deep structure context into which they are to be inserted, and the other rule would insert phonological and morphological features that are shown in the same lexicon to be compatible with the syntactic and semantic features attached to the same node in surface structure.

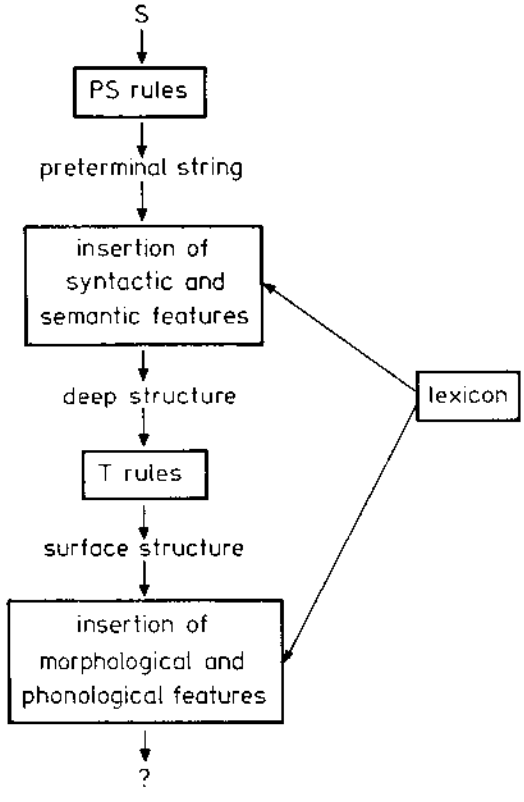
Let me make it as clear as I can what the revised view on lexical insertion is. First of all, we can assume for the present that there is a thing called a lexicon, which specifies a very large number of sets of features. In the simplest cases, such as *sheep*, there will be a single conjunctive set containing features of four types: phonological, morphological, syntactic and semantic, and what the lexicon tells us, in effect, is that all the features in the set are compatible with one another, and can therefore be associated with the same morpheme. Each feature will be classified in some way according to the level (phonological, morphological, etc.) to which it belongs, but there is no special ordering among the features. This is important for more complex cases like *refuse/refusal*, where *disjunctions* appear among the sets: each of the alternative sets can contain features from more than one level, as can the set of features common to all occurrences of *refuse/refusal*, so it has to be possible for set-boundaries to cut across the levels of phonology and so on. Finally, in cases like the grammatical morpheme *do*, we must allow for a set which contains syntactic, morphological and phonological features but no semantic features. The three cases are summarised below, in a very simplified form:

(7) *sheep*: ([ʃi:p] *and* [plural=∅] *and* [noun] *and* ['sheep'])
refuse/refusal: ([rəfju:z] *and* ['refuse'] *and* (([verb] *and* [regular]) *or* ([noun] *and* [+ −əl] *and* [regular])))
do: ([du:] *and* [∼d ∧ ∼did] *and* [verb] *and* [auxiliary] and [neutral])

Against the background of this view of the lexicon, lexical insertion becomes a matter simply of building up complete conjunctive sets of features including the features that are already present in the structure of the sentence. It can be thought of as taking two steps: first, at the level of deep structure, sets

containing just syntactic and semantic features are built up on the basis of the purely syntactic features already in the deep structure, using a rule similar to the rule in (2), with the possible revision of (3) and a revision to prevent phonological and morphological features from being added. The second step takes place in surface structure, where the terminal nodes bear sets of syntactic and semantic features due either to the first step or to subsequent transformations; the second lexical insertion rule then repeats the process of building up sets of features, just as in the first step except that this time phonological and morphological features *can* be included. The steps are summarised below:

(8)



VI. THE MERITS OF THE REVISED VIEW

We now have to justify the revised view, as just described, as against the standard view, in which all types of features are inserted en bloc into deep structure and the second lexical look-up just has a tidying-up function. The arguments are as follows.

(i) *General Aesthetic Considerations*
The revised view has no rules which just 'tidy up' the output of other rules. The concept of rules for tidying-up goes against the general trend in structural linguistics, including TG versions, in that such rules are essentially ad hoc. In the standard view, moreover, it is a fairly arbitrary matter whether one inserts the phonological shape of a word in the first or the second look-up: if pronoun-forms like *I* versus *me* are included because they are irregular, why not also include all irregular noun and verb roots which undergo internal changes in inflected forms? In the revised view, this decision need never be made: *all* morphological and phonological features are added in the same way.

(ii) *Morphological Irregularities*
In cases where words inflect irregularly, the revised view allows the whole of the irregularity to be easily incorporated into the lexicon, where it belongs, although it will add complexity to the grammar, as irregularities should. For example, we could develop the entry for *do* given in (7) as follows:

(9) (([du:] *and* [present] *and* [neutral]) *or* ([d ∧ z] *and* [present] *and* [singular]) *or* ([did] *and* [past]) *or* ([d ∧ n] *and* [past participle])) *and* [verb] *and* [auxiliary] *and* [neutral])

As in (7), this entry includes no semantic features, as it is the entry for the 'dummy' and therefore meaningless, *do*, but we shall relate it to the meaningful *do* below. It will be seen, however, that (9), unlike (7), includes no morphological features as such: this is because there is nothing left to be said about its morphology. (Morphological features in general are meant to show *widespread* idiosyncratic facts about morphemes, such as declension classes.)

(iii) *Generalisations Covering Grammatical and Lexical Morphemes*
It is hard to say whether the standard view could treat lexical and grammatical items as instances of the same lexical entry. For instance, can the lexical *do* in 'I do all the painting' be covered by the same entry in the lexicon as the grammatical *do* in 'Do you work here?' In the revised view, there is no difficulty in doing this as one surely should, given their morphological and phonological identity: if we use ['do'] to stand for whatever semantic features the lexical *do* has, we simply expand the entry in (9), replacing the last three features of (9) by the following:

(10) (... *and* [verb] *and* (([auxiliary] *and* [neutral]) *or* ([main] *and* ['do']))

The difference between the two views becomes particularly clear in cases like this: in the standard view, exactly the same set of phonological and morphological features are introduced at two completely separate stages in the derivation, although it is precisely these features that are *neutral* as to when they are introduced; whereas in the revised view they are all introduced at the same stage, namely the stage where the difference between lexical and grammatical items is no longer relevant.

(iv) *Suppletion*

When two phonologically unrelated roots alternate in a paradigm, (eg. go~went) this is a problem for the standard view, which sees a single underlying phonological shape as the basis for the lexical item (we shall return to this point below): for instance, in the case of *go* and *went* are we to say that there is a single entry, with *go* as the underlying form and a unique phonological rule converting *go* to *wen-*; or are there two different entries, one for *go* and the other for *wen-*? The objections to both analyses are obvious. Again, there is no problem at all for the revised view, since *go* and *went* can both be included in a single lexical entry, as follows:

(11) (([gou] *and* [present]) *or* ([went] *and* [past]) *or* ([gon] *and* [past participle])) *and* [verb] *and* ['go'])

As well as inflexional suppletion, like *go~went*, there is also derivational suppletion, to which we shall return below.

(v) *Surface Features in Deep Structure*

One of the objections that has been levelled at TG theory by its critics (cf., for example Lamb, 1966 and the rather irrelevant reply in Postal, 1968: 205) is that phonological and morphological features are introduced too early in the derivation: in the standard view, these features are introduced, along with the syntactic and semantic features, into deep structure, although they become relevant only in surface structure. Apart from general aesthetic objections, this raises the problem that the surface features are available for transformational rules to refer to, but transformational rules never *do*, so the metatheory needs to explain why not. In the revised view, on the other hand, this problem never arises: phonological and morphological features aren't inserted until after all T-rules have applied, so the latter couldn't possibly refer to them.

The above arguments suggest quite clearly that the standard view of lexical insertion has some serious drawbacks which are avoided by the revised view. Moreover it should be remembered that these advantages in-

volve virtually no change in the grammar: all that is needed is a change in the formulation of the two rules for lexical insertion.

How do these conclusions affect the concept of the lexical entry? Paradoxically, by asserting the unity of the lexicon we have weakened yet further the unity of the lexical entry: the unity of the lexicon allows us to insert surface features separately from deep features, without having to repeat all the latter in introducing the former; but in order to insert deep and surface features separately we have to divide each lexical entry into at least two parts, one part to be inserted by each insertion rule. We thus find ourselves even further from the original Chomskyan position where lexical entries, as listed in the lexicon, are inserted in toto into sentence structures.

VII. THE MCCAWLEY APPROACH

McCawley's approach to the lexicon and lexical insertion is much less clear than Chomsky's, but it seems to have similarities to the approach advocated in this paper. One of the main differences between McCawley's approach and either Chomsky's or the present approach is that McCawley's includes 'prelexical syntax' – the notion that lexical insertion has to follow certain transformations. As one would expect, it is hard to separate this particular aspect of McCawley's views from other aspects which are more directly relevant to this paper, but I shall try to do so.

The first relevant point in McCawley's theory is that he rejects all disjunctions in lexical entries: "there is no a priori reason for grouping items together in a dictionary at all: one could perfectly well take the notion 'lexical item' to mean the combination of a single semantic reading with a single underlying phonological shape, a single syntactic category, and a single set of specifications of exceptional behaviour with respect to rules." (1968b: 125). This passage occurs in the context of a discussion of the four meanings of *bachelor*: Katz and Fodor had suggested that all four should be inserted together, as alternatives, into sentence-structures, and McCawley was objecting to this idea. This is in line with the view of the present paper, which also advocates the elimination of disjunctions during lexical insertion, rather than after it.

It's less clear, however, whether McCawley means to allow disjunctions in the lexicon, which we certainly do. In the continuation of the passage just quoted, he says: "Under this conception of 'lexical item', first proposed by Weinreich (1966), there would simply be four lexical items pronounced *bachelor* rather than a single four-ways ambiguous lexical item." (*ibid.*: 126). From this it looks as though he proposes to list the phonological, morphological and syntactic properties of the four *bachelor*'s separately in the lexicon,

in the same way as one might list the two *banks* (money-bank and river-bank). The trouble with this approach, of course, is that it vastly increases the size of the lexicon, and denies the distinction between accidental and non-accidental similarities in the lexicon. On the other hand, in his review of Weinreich, 1966, he envisages treating *cook* as "two separate lexical units (noun and verb) in a single composite dictionary entry" (1968:581), which is obviously a more satisfactory approach, assuming that it is possible to give some content to the notion 'composite dictionary entry'.

Unfortunately, I don't see how one could do so, given another aspect of his theory: that "each 'dictionary entry' could be regarded as a transformation which replaces a portion of a tree that terminates in semantic material by a complex of syntactic and phonological material." (1968:72). Let us assume that the structure to be replaced by the verb *cook* is an abstract representation corresponding to 'prepare food by heating it', and (following McCawley's own analysis) that of the noun *cook* is a more complex structure including this structure ('one who prepares food by cooking it'): how could a *single* transformation substitute *cook* for *either* structure? The claim that lexical substitution is just a particular type of transformational operation presumably means that it is subject to the normal constraints on transformations, but there can be few linguists who would be happy to allow transformations such as the one McCawley seems to envisage for *cook*:

(12) $\underset{1}{x} - \underbrace{([\text{ONE}\ [\text{WHO} - \text{PRESENT} -}_{2}$

$\underbrace{\text{PREPARE} - \text{DET} - \text{FOOD} - \text{BY} - \text{ING} - \text{HEAT} - \text{IT}]]}_{3} - \underset{4}{Y}$

$1 \quad \underbrace{(2)\,3}_{[\text{kuk}]} \quad 4$

On the other hand, if a different transformation is responsible for inserting the noun *cook* and the verb *cook*, there is no sense in which the two can be said to be subsumed under a single 'composite dictionary entry'.

What I have tried to show in this discussion of McCawley's approach to lexical insertion, which is the second main view on lexical insertion in a TG framework, is that he avoids the introduction of disjunctions of features, as advocated in this paper, but does so simply by removing disjunctions from the lexicon – a case of throwing out the baby with the bath-water, it seems. With reference to the other main theme of this paper, that surface features need to be inserted later than deep features, one might say that McCawley scores slightly better than Chomsky, inasmuchas his phonological features get inserted a little nearer to the surface than Chomsky's, but this is simply

because *all* his features get inserted later (1968d: 78), rather than because he recognises the need to separate surface from deep features.

VIII. THE CONCEPT OF THE LEXICAL ENTRY

In conclusion, how does the lexical entry itself fit into the view proposed in this paper? Briefly, it doesn't fit in at all well. Let me explain why not.

The lexical entry has a clear theoretical status for both Chomsky and McCawley, as the element that is (a) inserted en bloc into sentence structures and (b) listed in the lexicon independently of all other lexical entries. I have shown that there is a conflict between these two properties: that what is inserted need not be more than a *sub*-set of a total entry, in sense (b), since disjunctions exist in the lexicon but not in what is inserted into sentence structures. Moreover, it isn't even the case that whatever features get inserted are all inserted en bloc: rather, the high-level features (semantic and syntactic) get inserted into deep structure, and the low-level ones into surface structure. Thus the lexical insertion rule has the job not only of comparing the content of the lexical node with the place in structure where it is to be inserted, but also of choosing among the features in the entry, so as to insert some but not others.

Of course, at this point we find ourselves playing with terminology. We have two concepts (what is inserted and what is stored) which both used to be referred to by the same name, 'lexical entry', but we have found that they often define different objects so the common name is at best confusing. To revise the terminology, we could use 'lexical entry' for *either* of the original concepts and introduce a new term for the other, or introduce two new terms, in order to avoid confusion. This course seems the safer of the two, so I hereby propose 'lexical insert' as the name for what is inserted, and 'lemma' for what is stored in the lexicon.

But what *is* the 'the unit that is stored' (the lemma) now that it has been separated from the lexical insert? We might define it as the largest unit of structure in the lexicon, so that the only relation between two lemmas is a disjunctive one, and the lexicon is seen as just a disjunctive (and, of course, unordered) set of lemmas. Any *conjunctive* relation would therefore by definition have to be within a lemma, though, as we have seen, there would have to be some disjunctive ordering within the sets that are conjunctively related, and therefore within the lemma. For instance, a hypothetical lexicon containing only the words *sheep*, *refuse* and *refusal* would look something like this (cf. the individual entries in (7) above):

(13) (([ši:p] *and* [Plural=∅] *and* [noun] *and* ['sheep']) *or* ([rəfju:z]

and ['refuse'] *and* (([verb] *and* [regular]) *or* ([noun] *and* [+ −əl] *and* [regular])))

Here there would be just two lemmas, separated by the first '*or*', with a mixture of *and*'s and *or*'s within the second lemma.

Given this view of the lexicon, the lexical insertion rule given above in (2) (with the possible modification in (3)) is exactly right: it resolves all disjunctions in favour of one of the alternatives, and makes sure that all the features that are assigned to the same node are allowed to be in a conjunctive relation. However, it is worth pointing out that it makes no reference to the notion 'lemma': it resolves the choice among lemmas in just the same way as it resolves the choice among alternatives *within* lemmas. So at least from the point of view of lexical insertion it makes no difference at all whether we distinguish between lemmas and parts of lemmas – which casts doubt on the significance of this distinction, and therefore of the concept 'lemma' itself.

How seriously this doubt should be taken I find it quite hard to say. Certainly there would be several advantages for the linguist in doing away with the concept of the lemma. For example, it is only because of this concept that we have to distinguish between cases of homonymy (two separate lemmas which happen to contain the same phonological shape) and polysemy (one lemma containing two meanings), a distinction which is notoriously difficult to make. Similarly, it is hard to see why *radio* and *wireless* shouldn't be covered by the same lemma, for just the same reasons (mutatis mutandis) as the two meanings of *cook* would. Again, without the lemma the problem wouldn't arise. The trouble is, that it is difficult to see how one *could* do without the lemma, since relations between features on different levels involve often complex sets of features on either level. For instance, one needs to relate *all* the semantic features of *sheep* to *all* its phonological features, which means in effect that one has to deal with *sheep* all on its own.

IX. CONCLUSIONS

The main conclusions of this paper are the following:

firstly, that the lexical unit which is inserted into sentence structures (the 'lexical insert') is different from the unit that is stored in the lexicon (the 'lemma'), and therefore the concept of the 'lexical entry' needs either revising or, better still, abandoning;

secondly, that lexical insertion takes place in two steps: first, just semantic and syntactic features are inserted into deep structure, and then *all* phonological and morphological features are inserted into surface structure.

University College London

BIBLIOGRAPHY

Chomsky, N.: 1965, *Aspects of the Theory of Syntax*, Cambridge, Mass.: MIT Press.
Chomsky, N.: 1970, 'Remarks on Nominalization', in R. A. Jacobs, and P. S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, London, etc.: Ginn.
Chomsky, N.: 1972, 'Some Empirical Issues in the Theory of Transformational Grammar', in Peters, S. (ed.), *Goals of Linguistic Theory*, Englewood Cliffs: Prentice-Hall; also reprinted in N. Chomsky, *Studies on Semantics in Generative Grammar*, The Hague: Mouton.
Gunter, R.: 1972, 'English Derivation', *Journal of Linguistics* **8**, 1–20.
Katz, J. J. and Fodor, J. A.: 1963, 'The Structure of a Semantic Theory', *Language* **39**, 170–210.
Katz, J. J. and Postal, P. M.: 1964, *An Integrated Theory of Linguistic Descriptions*, Cambridge, Mass.: M.I.T. Press.
McCawley, J. D.: 1968a, 'Concerning the Base Component of a Transformational Grammar', *Foundations of Language* **4**, 243–69.
McCawley, J. D.: 1968b, 'The Role of Semantics in a Grammar', in E. Bach and R. T. Harms (eds.), *Universals in Linguistic Theory*, Holt, Rinehart and Winston.
McCawley, J. D.: 1968c, Review of T. A. Sebeok (ed.), 'Current Trends in Linguistics', Vol. 3: Theoretical Foundations, *Language* **44**, 556–592.
McCawley, J. D.: 1968d, 'Lexical Insertion in a Transformational Grammar without Deep Structure', in W. J. Darden *et al.* (eds.), *Papers from the Fourth Regional Meeting of the Chicago Linguistic Society*; Chicago: Dept. of Linguistics, Univ. of Chicago.
Matthews, P. H.: 1967, Review of N. Chomsky, *Aspects of the Theory of Syntax*, *Journal of Linguistics* **3**, 119–152.
Stockwell, R. P., Schachter, P., and Partee, B. H.: 1973, *The Major Syntactic Structures of English*, New York, etc. Holt, Rinehart and Winston.
Weinreich, U.: 1966, 'Explorations in Semantic Theory', in T. Sebeok (ed.), *Current Trends in Linguistics, Vol. III*, 395–474; The Hague: Mouton.