Zwicky on heads¹

RICHARD A. HUDSON

Department of Phonetics and Linguistics, University College London

(Received 22 July 1986)

I. INTRODUCTION

An interesting development in the last decade or so has been the increasing use that theoretical linguists have made of the notion ‘head’ – or rather, in order not to beg the question, of notions to which they have given the name ‘head’. The term has been around for a long time in linguistics, of course – for example Bloomfield uses it in relation to endocentric constructions (1933: 195), where the head is the daughter constituent which has the same distribution as the mother. Before that, Sweet had used ‘head-word’ to refer to any word to which another is subordinate (1891: 16, quoted in Matthews, 1981: 165). However, theoretical linguists made very little use of the term, or of the constellation of associated concepts, until quite recently. Its present status is due largely to work on X-bar syntax dating from Chomsky (1970), and especially to its recent manifestation in Generalised Phrase Structure Grammar (Gazdar & Pullum, 1981; Gazdar et al., 1985) – and even more so in the ‘head-driven’ variant of this (Pollard, 1985). But the improved status of ‘head’ is also due to some extent to the renewed interest in dependency grammar (Anderson, 1971, 1977; Matthews, 1981; Atkinson, et al, 1982; Hudson, 1984; Nichols, 1986). All these treatments agree not only in using the term ‘head’, but also in using it to refer to the element in some construction to which all the other parts of that construction are (in some sense) subordinate.

It is reasonable to be suspicious of notions with as chequered a history as ‘head’ on the grounds that it developed in the days before formal theories of syntactic structure were available, as a rather metaphorical and vague way of referring to notions which we may now be able to define more economically and insightfully in terms of other, more primitive notions. It is sometimes tempting in such cases to reintroduce the traditional term just because it is traditional. There is then a danger of multiplying entities unnecessarily – as if one were to try to find a place for notions like ‘guttural’ or ‘lilting’ in modern phonetics. Worse still, a traditional term may be used to refer to a multiplicity of notions which are more or less closely related to the (vague) traditional concept, but which are in fact independent of each other. Arnold

¹ Arnold Zwicky gave me most valuable comments on an earlier version of this paper, for which I should like to thank him. I am also grateful to Nigel Vincent for his help with the present version.
Zwicky has recently argued (1985a) that this is in fact what has happened with the term ‘head’. He starts with the observation that after its short (recent) life in syntax, the term has now started to be applied in morphology as well, in order to allow the very general principle of ‘percolation’ to apply at both levels. This principle is generally formulated in terms of a mother constituent and its ‘head’, and requires the category membership of the latter to ‘percolate up’ to the former. Thus, if we know that the head is a noun, we can deduce that the mother is also a noun (perhaps with a different number of bars); and this is true whether the head is a single word or a suffix (e.g. -ness). However, after surveying a variety of head-like concepts in syntax, Zwicky concludes that percolation applies to only one of them – what he calls the ‘morphosyntactic locus’ (the constituent on which inflexions relevant to the mother are located, or would be if there were any such inflexions). He considers seven other head-like notions, but they all seem to be independent of the morphosyntactic locus, so the notions ‘syntactic head’ and ‘morphosyntactic locus’ appear to be one and the same. Moreover, when possible uses in morphology are considered, he shows that ‘morphosyntactic locus’ is not the concept that is needed for category percolation – which of course means that the apparent generalizations and insights to be gained by extending the category ‘head’ into morphology are spurious.

Apart from the consequences for morphology, the general conclusion is that ‘head’ is much less important in syntax than many of us have supposed; indeed, if we have some way to characterize morphosyntactic loci, ‘head’ can be dropped entirely from our theoretical vocabulary. The logic of Zwicky’s argument is impeccable, and there is no doubt that this paper is a valuable contribution to the development of grammatical theory – as so many of his other papers have been, such as those on clitics (1985b) and on constituents (1978). However, there are serious reasons to doubt his conclusions about the role of ‘head’ in syntax – and presumably also for questioning his pessimism regarding its role in morphology, though I shall have nothing to say about this. The doubts arise not from his logic, but from his assumptions about how various types of structure should be analysed.

2. ZWICKY’S ANALYSIS

He considers the following syntactic notions as possible candidates for ‘head’ – i.e. for the category to which percolation applies.

A. The SEMANTIC ARGUMENT – the constituent whose meaning has the status of ‘argument’ in relation to some ‘functor’. This category will be explained and discussed in the next section, but Zwicky assumes, for example, that a determiner’s meaning is a functor whose argument is the meaning of the following common noun.

B. The DETERMINANT OF CONCORD – the constituent with which some other fellow-constituent must agree.
C. The morphosyntactic locus — as explained earlier, this is the constituent on which any inflexions which are relevant to the mother are located (or on which they would have been located had the language included any such inflexions).

D. The subcategorizand — the constituent which is subcategorized with respect to its sisters, in the familiar sense.

E. The governor — the constituent which determines the morphosyntactic form of some sister.

F. The distributionally equivalent constituent — the constituent whose distribution is similar to that of the mother.

G. The obligatory constituent — the one which has to be present if the mother is to be categorized as it is.

H. The ruler of dependency theory — in a dependency-based analysis, the 'ruler' (Zwicky's term for what is often called 'head' in actual dependency analyses) is the word on which other words depend.

Zwicky discusses these categories in relation to six English constructions:

(i) V + NP e.g. control those penguins
(ii) P + NP toward those penguins
(iii) NP + VP we control those penguins
(iv) Det + N those penguins
(v) Aux + VP must control those penguins
(vi) Comp + S that we control those penguins

Given the analyses that Zwicky assumes for these constructions, his eight categories apply to them as shown in Table 1. The rows show the head-like notions listed above (A to H), and the columns show the six English constructions. Each entry in the table shows which of the constituents in these constructions is taken, in Zwicky's analysis, to be the head-like constituent, but these entries are given in terms of the corresponding entries for an extra row which I have supplied at the top of the table, for 'semantic functor'. This is of course just the converse of the row for 'semantic argument', as can be seen from the row of stars in the latter, but it will be helpful for the later discussion to take the semantic functor entries as a standard against which we can measure other entries.

If all the entries in Table 1 had been either '=' or blank then we could have claimed that all of the eight head-like categories were in fact the same category, presented with respect to different properties. A natural name for this super-category would have been 'head', and the value — indeed, indispensability — of 'head' would have been established. The big attraction of the notion 'head', for those of us who believe in it, is precisely that it integrates a wide range of different phenomena such as the eight 'head-like' concepts. (It goes without saying that it is immaterial what we actually call this super-concept; if one linguist calls it 'head' and another calls it something else the dispute is about nothing but terminology.) But of course the main
thrust of Zwicky’s paper is that this is **NOT** how things are – the various concepts in fact pick out different ranges of items across the various constructions, as shown by the distribution of = and * in Table 1.

<table>
<thead>
<tr>
<th>Semantic functor . . .</th>
<th>V + NP (V)</th>
<th>P + NP (P)</th>
<th>NP + VP (VP)</th>
<th>Det + N (Det)</th>
<th>Aux + VP (Aux)</th>
<th>Comp + S (Comp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Semantic argument</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(B) Determinant of concord</td>
<td><em>(</em>)</td>
<td>.</td>
<td>*</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>(C) Morphosyntactic locus</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
<td>*</td>
</tr>
<tr>
<td>(D) Subcategorizand</td>
<td>=</td>
<td>.</td>
<td>=</td>
<td>.</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>(E) Governor</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>.</td>
<td>=</td>
<td>*</td>
</tr>
<tr>
<td>(F) Distributional equivalent</td>
<td>=</td>
<td>.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(G) Obligatory</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(H) Ruler</td>
<td>=</td>
<td>.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>=</td>
</tr>
</tbody>
</table>

Key: = same as entry for ‘Semantic functor’
* different from entry for ‘Semantic functor’

Table 1
Zwicky’s analysis of six constructions in terms of eight head-like categories

One advantage of presenting the information as I have done in this table is that it allows us to pinpoint the sources of discrepancy by isolating the rows and columns in which * appears. These are as follows:

(a) the row for ‘(A) Semantic argument’,
(b) the row for ‘(B) Determinant of concord’,
(c) the column for ‘Det + N’,
(d) the column for ‘Aux + VP’,
(e) the column for ‘Comp + S’.

In the following discussion I shall try to show that the analyses which Zwicky assumes, and on which these entries are based, are either irrelevant or open to improvement. The outcome of the discussion will be a revised version of the table in which the discrepancies are removed, and which can be taken as evidence for the position which Zwicky rejects, namely that there is a general category which subsumes many – though not all – of the ‘head-like’ concepts. Naturally I shall suggest that this supercategory is what has traditionally been called ‘head’.

Before embarking on this exercise it is perhaps worth repeating that the scope of the discussion is extremely limited – just six constructions, and all in relation to English. All that is at issue for present purposes is whether the various head-like categories are in conflict as far as this particular set of data is concerned. At the end of the discussion it will still be an important matter
of debate to what extent the conclusions reached here can be generalized to other constructions, or to other languages.

3. ARGUMENTS AND BEING ‘A KIND OF’

I start with Zwicky’s introduction to his discussion of semantic arguments:

   We could take the head/modifier distinction to be at root semantic: in a combination X + Y, X is the ‘semantic head’ if, speaking very crudely, X + Y describes a kind of the thing described by X. (p. 4)

However crude it may be, I find this description very insightful, and I think it corresponds closely to the intuition that underlies many of the dependency analyses with which I am familiar. For example, in control those penguins (and other V + NP patterns), the whole refers to a kind of controlling, and not to a kind of penguins, so control is the head; and similarly in We control those penguins, in which both we and those penguins are subordinate to control. This interpretation of ‘head’ applies even more obviously to compounds like jam sandwich, whose head is sandwich because the whole defines a kind of sandwich, and not a kind of jam. Thus the head’s sisters have the function of modifying the meaning of the head, in the everyday, non-technical meaning of ‘modify’.

Having given this excellent introduction, however, Zwicky immediately suggests ‘a sharpening (and extension)’ of it in terms of the semantic distinction between functors and arguments. The train of thought is as follows: let us assume that in Det + N the semantic head is N because ‘those penguins describes a kind of penguin [sic]’. Now in standard functor/argument analyses it is the N which is taken as the argument, so let us assume that the notion ‘argument’ is just a sharpened-up equivalent of the element of which the whole defines ‘a kind’. I shall suggest in Section 6 that the standard analysis of Det + N on which this conclusion rests is faulty, but it is surely easy to see that the analysis does not in fact generalize to the other constructions.

Take the V + NP construction, to which we have already referred. According to the standard analyses the semantic argument is clearly the NP, but we have seen that control those penguins refers to a kind of controlling, and not to a kind of penguins. Similar remarks apply to NP + VP examples: as I suggested above, We control those penguins refers to a kind of controlling, namely one in which we and the penguins are involved. Of course, there is a long tradition of taking NP + VP as a semantically exocentric construction, in which neither constituent has a uniquely privileged status. This seems, however, to confuse two questions. One is about the relation of the NP to the VP – does the whole refer to an instance of whatever is referred to by the VP? Consider the well-established logical tradition of representing the meaning of John loves Mary as ‘Loves (John, Mary); here the respective
relations of subject and object to the predicate (provided by the verb) are clearly very similar, and it would be correct to say that ‘Loves (John, Mary)’ is an instance of ‘Loves (x, y)’. So, in V + NP and in NP + VP, the NPs both have the same relation to the other constituent. A similar assumption about the equal statuses of subjects and objects is made by all the familiar semantic analyses which use semantic relations.

Both types of semantic analysis reflect the intuition that a sentence and its verb both refer to some kind of predicate, with the difference that for the verb various argument ‘slots’ are filled by variables (or indefinite default values), whereas for the sentence they have values defined by the subject and object. Take the verb control for instance; in the lexicon it might be represented as ‘Control (x, y)’ (I am not advocating such structures, just assuming them for simplicity), and this structure would be used as the basis for the semantic structure of a sentence like We control those penguins, with the variables replaced by ‘we’ and ‘those penguins’ (or their equivalents). All this is presumably very familiar and rather uncontroversial; but it all indicates that NP + VP refers to a kind of the thing referred to by VP.

This question about the semantic relation of NP to VP is separate from the other one, which is about the relations between the whole structure NP + VP (i.e. sentences) and the structure V + NP. Sentences seem to have something which none of their parts have, namely the semantic basis for an illocutionary force; so in that respect the meaning of a sentence is indeed more than a modified version of the meaning of its VP. However, this addition of illocutionary force is not due to the addition of the subject, because it also happens in the case of imperatives, where no subject is added (compare the infinitive (to) control those penguins with the imperative Control those penguins!). Moreover no illocutionary force is added when NP + VP is embedded in sentences like I know we control those penguins, so the illocutionary force must be due to the linguistic context – i.e. lack of dominating constituents – rather than to the NP + VP structure itself. I conclude, then, that we are justified in taking NP + VP as referring to a kind of what is referred to by VP.

How about P + NP, in which once again it is the NP that is the argument? Surely it is nonsense to say that toward those penguins refers to a kind of penguins – as it should do if the whole referred to a ‘kind of’ the semantic argument? On the contrary, toward those penguins refers to a kind of place or direction, which is defined in relation to the penguins. Similarly, behind the table refers to a place, and not to any kind of table; and because of the penguins refers to a kind of reason, not to the penguins. In case it is not obvious that this is so, let me just point out that if toward those penguins did refer to a kind of penguins it should correspond to pronouns like what, which and them or it, but it does not (compare for example *What he saw was toward those penguins). Instead it corresponds to where, there and so on (compare Where he pointed his camera was toward the penguins).
The less straightforward constructions are similar in their interpretation: the whole refers to something which is a particular case of (‘kind of’) the thing referred to not by the argument, but by the functor. Take the Comp+S construction, as in because we control those penguins. (In case it is objected that because is better analysed as a preposition, I should explain that I shall be introducing an analysis in which prepositions and subordinating conjunctions/complementizers are members of the same class.) According to Zwicky the S is the argument, so the whole construction should refer to a kind of controlling (by us) of those penguins, but it could be taken at least equally convincingly as referring to a kind of reason – namely the kind of reason in which the cause is our controlling of those penguins (notice that it could be paraphrased by for a certain reason, namely that we control those penguins). Similarly, when we control those penguins refers to a kind of time, and so that we can control those penguins to a kind of purpose. Admittedly the semantic analyses which I am invoking are very rough and ready, but at least they are nearer to the mark than analyses in which Comp+we control those penguins refers just to some instance of us controlling those penguins (as if Comp were a modifier like yesterday or firmly).

Another construction to be tested is Aux+VP, as in may control those penguins. Here Zwicky takes VP as the argument, so it ought to refer to a kind of controlling of those penguins. This might seem fairly reasonable, especially in view of sentences like Maybe they control those penguins, which means much the same as They may control those penguins, but where the effect is achieved by means of an adverb. If an adverb + verb combination generally defines a kind of the event to which the verb refers (as in e.g. often control), then perhaps the same should be true of cases like maybe control and therefore also of may control. However, it could also be argued that maybe control is a kind of possibility rather than a kind of controlling and that in this case the syntax and semantics are out of step; a more transparent rendering of the same meaning would be It is possible that they control those penguins, in which it is much easier to see that the whole refers to a kind of possibility, rather than to a kind of controlling of penguins.

In conclusion, then, the notion of ‘semantic argument’ seems particularly badly suited as a basis for formalizing the ‘kind of’ notion – I have shown that in five out of the six constructions it is the functor, and not the argument, that defines the thing of which the whole structure defines ‘a kind’; and I have promised a discussion of the sixth case (Det+N) which will lead to the same conclusion. Indeed, it would have been very odd had it turned out otherwise, because I take it that in a functor–argument structure it is the arguments, not the functor, which would be represented in the lexicon by means of variables. Thus the functor stays constant as the arguments vary, and each new combination of arguments defines a different ‘kind of’ the thing defined by the functor.

It is because of the specially close relation between ‘kind of’ and functors
that I included the notion ‘semantic functor’ as a ‘head-like’ notion in addition to those which Zwicky gives. In Table 1 I even go so far as to take this extra notion as the standard against which other notions can be measured, because of the very great value of Zwicky’s original insight about the whole defining a kind of the thing denoted by the ‘semantic head’ – i.e. by the functor. This means in effect that we can now ignore the whole of the ‘semantic argument’ line of entries in Table 1, on the grounds that stars there are irrelevant.

4. Concord

The line for ‘(B) controller of concord’ contains only three entries because there is no concord at all in most of the constructions – concord applies only to Det + N and to NP + VP. There is also a bracketed entry for V + NP on the basis of Zwicky’s observation that some languages – Hungarian for example – have concord between verb and object. However, the three entries which are supplied agree with those for the semantic argument, which we are now ignoring as irrelevant to headship. Zwicky’s entries for concord are of course entirely in conformity with the widely held view, which he quotes, that concord is determined by the semantic argument, but we must now look for a way to reconcile the entries for concord with those for semantic functors.

An alternative view of concord is as follows. Let us assume two elements X and Y and a rule which requires them to agree with each other with respect to some feature or set of features F. The question is what else needs to be specified about X and Y in order for this rule to operate correctly, and the simplest assumption is clearly that NOTHING else need be specified. We must presumably accept this assumption unless some good reason is found for rejecting it. Now there are two obvious essentials for any concord pattern between two elements, which we may call the ‘concordants’. One requirement is that one of the concordants should be fixed, with respect to F, by some rule other than the concord rule. For example, if one of the concordants is a noun and the other an adjective, the number of the noun is fixed by its meaning (or by the lexicon, if its number is arbitrary). The other requirement is that the remaining concordant should NOT be fixed in this way, because otherwise its value for F will not be free to be fixed by the concord rule.

Of course it is logically possible for a concord rule to apply to concordants which both have their values for F fixed independently. In that case the concord rule would act as well-formedness constraint, ruling out cases where the values for X and Y differ. However, so far as I know this is not in fact how concord rules generally, or even ever, work. A candidate in English for such a rule would be the one which is often said to apply between a predicative noun and its subject, giving rise to pairs like He is my friend versus They are my friends. Since either noun could in principle have its number value fixed lexically, irreconcilable conflicts could arise between them; but when such
situations arise, the result may be either impeccable (as in *They are a nuisance*) or slightly problematic (as in *My main problem is these scissors*) — but not ungrammatical, as one would expect if there were a concord rule that had been infringed. The conclusion to which I come on this matter, then, is that there is no such rule (in English at any rate), and that the predicative noun selects its value for the number feature independently of that of its subject, taking account only of the sentence’s meaning.

<table>
<thead>
<tr>
<th>Semantic functor...</th>
<th>V + NP (V)</th>
<th>P + NP (P)</th>
<th>NP + VP (VP)</th>
<th>Det + N (Det)</th>
<th>Aux + VP (Aux)</th>
<th>Comp + S (Comp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C) Morphosyntactic locus</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
<td>*</td>
</tr>
<tr>
<td>(D) Subcategorizand</td>
<td>=</td>
<td>.</td>
<td>.</td>
<td>=</td>
<td>.</td>
<td>=</td>
</tr>
<tr>
<td>(E) Governor</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(F) Distributional equivalent</td>
<td>=</td>
<td>.</td>
<td>.</td>
<td>=</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(G) Obligatory</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(H) Ruler</td>
<td>=</td>
<td>.</td>
<td>.</td>
<td>*</td>
<td>*</td>
<td>=</td>
</tr>
</tbody>
</table>

*Table 2*

Zwicky’s analysis of six constructions in terms of the six genuinely head-like categories

It seems reasonable to suppose that all concord rules will take the form described earlier, with only one concordant fixed externally. If that is so, there is no reason for specifying which of the concordants is the ‘determinant of concord’, because this will be shown independently by whatever rules apply to one of them. Typically one of them is a noun and the other is not, so the rules which fix features on the noun will naturally apply to this, and the values of the corresponding features on the other concordant will be left to the concord rule. If this is so, it can be seen that the direction of concord determination has nothing at all to do with the notion ‘head’, *pace* Zwicky. Putting it somewhat crudely, the determinant of concord is the noun, and whether this is head or not is immaterial. We can now remove the line for ‘(B) Determinant of concord’ from Table 1, thereby further reducing the number of deviations from the standard. Table 2 shows the effects of this reduction.

5. Auxiliaries and Complementizers

According to Zwicky’s analysis the Aux + VP and Comp + S constructions each contribute three discrepancies relative to the standard. We shall now consider whether these judgments are correct. We start with the analysis of auxiliaries. Here is what Zwicky says about them in connexion with the category of ‘distributional equivalents’:
VP is the distributional equivalent of Aux + VP, since the distribution of Aux + VP is roughly the same as the distribution of VPs like *control those penguins* and *go to Fresno.* (p. 12)

This claim is hard to assess because of the word ‘roughly’. For example, VP clearly cannot substitute for Aux + VP in sentences like *He will control those penguins* (compare *He control those penguins*) or *He is controlling those penguins* (compare *He controlling those penguins*). Indeed, in order to take VP as interchangeable with Aux + VP we have to ignore the morphosyntax of the verb in VP entirely. Moreover, Aux + VP cannot be replaced by VP in examples like *Why are you worrying?*, even if we ignore questions of morphosyntax and pretend that the difference between *worrying* and *worry* does not matter (compare *Why you worrying/worry?*). It would be at least as persuasive to say that the distributional equivalent of Aux + VP is Aux. So far as I know there are no contexts in which a sequence of Aux + VP cannot be replaced by the Aux on its own, because of the effects of VP ellipsis. For example, *He will control those penguins* is matched by *He will; He is controlling those penguins by He is;* and *Why are you worrying?* by *Why are you?* Admittedly these cases are all elliptical, but this fact is irrelevant as far as the principles which Zwicky gives for identifying distributional equivalence are concerned.

Similar conclusions apply to the related question of which constituent is obligatory in Aux + VP. According to Zwicky, VP is the obligatory element, in spite of the well-known facts mentioned above, because he excludes elliptical constructions by fiat as irrelevant to the head-like notion of ‘obligatory constituent’. He mentions two kinds of ellipsis in this context. One is gapping (e.g. *I ate sushi, and Kiyoko a hamburger*), where I sympathize with his decision – so much can be elided in gapping that if we took account of gapping we should surely have to say that everything is optional. Moreover, there is no special connexion between the rule for gapping and the Aux + VP construction – all that is needed is for the gap to contain a verb of some kind, whether auxiliary or not. But the situation is very different with the ellipsis of VP (e.g. *I can swallow goldfish, but you can’t*). Here the ellipsis is made possible by the specific subcategorization properties of auxiliaries, according to which the complement of any auxiliary (i.e. the VP) is optional, so there is a special connexion between the ellipsis and the Aux + VP construction (namely, that the VP is always subject to ellipsis). Moreover in this respect auxiliaries are just like some kinds of non-auxiliary verb such as *watch, contribute* (Fillmore, 1986) and *know*, whose complements are also optional and, if absent, are to be recovered anaphorically (compare *Fred was watching, Fred contributed a pound* and I *know*). Zwicky accepts the optionality of the NP in V + NP as relevant, so it is hard to see why we should exclude that of VP in Aux + VP. If this is so, it is clearly Aux, and not VP, which is the obligatory element. The conclusion of the discussion, then, is that
ZWICKY ON HEADS

Aux is both the distributional equivalent of Aux + VP and also its obligatory constituent, contrary to Zwicky's analysis.

We now turn to Comp + S, where of course we have to take account of examples like since he left and what he did as well as of those like that we control those penguins which Zwicky cites. Once again the questions about which part is the distributional equivalent of the whole and about which part is obligatory are closely interconnected. It is true that if we concentrate on clauses introduced by that, it is easy to believe that the obligatory part, and therefore the part which is distributionally equivalent to Comp + S, is S. For example, in He thinks that we control those penguins, it is possible to omit that but not to omit the following clause, we control those penguins. However, that is a very untypical complementizer in this respect. Take a complementizer like since, which is much more typical. Here it is possible to omit the clause, by anaphoric ellipsis similar to the process responsible for VP ellipsis (compare I haven't seen him since we had that argument and I haven't seen him since); but it is not possible to omit the complementizer (compare *I haven't seen him we had that argument). Similarly for interrogative pronouns like what (e.g. I don't know what he did, I don't know what, and *I don't know he did), and for a number of other complementizers.

Not all complementizers are like since and what in this respect, of course – with many of them the clause is obligatory (e.g. if, as in *I won't come if). But this is precisely like the relation in V + NP, in which Zwicky accepts that the distributional equivalent and the obligatory element is V, in spite of the fact that some verbs have to be followed by a NP, others may, and others again must not be. What I am suggesting, then, is that a complementizer like since may be treated as a kind of adverb which occurs in three different structures: as sister of S (e.g. I've not seen him since he left), as sister of NP (e.g. I've not seen him since the party), or without a sister (e.g. I've not seen him since). Matthews (1981: 150) seems to favour a similar analysis and traces it back to Jespersen and Strang; and of course it is similar to the view that prepositions may be either transitive or intransitive – Jackendoff (1983: 49), referring to earlier work by Klima and Emonds.

This analysis has a welcome consequence, namely that it fits precisely with Zwicky's view that Comp is the subcategorizand in Comp + S. He bases this view on the observation that different complementizers allow different kinds of S – finite, non-finite, with or without gaps, and so on – so our observations about optionality can be taken as further evidence for his analysis. Moreover, we can even use the above discussion as a basis for filling in some of the entries for 'subcategorizand' which Zwicky left blank (for perfectly good reasons – because they presupposed too much discussion of theory-particular analyses). Even without entering into theory-particular issues, it is clear that there is at least a close connexion among adverbs, prepositions and non-pronominal complementizers, to the extent that there are single words which may be used in all three ways (e.g. since, discussed above; other examples are before, after
and maybe except). It seems at least worth exploring the possibility, in any theory, that prepositions are adverbs which have an NP as their complement, that complementizers are adverbs which have S as their complement, and that traditional adverbs are simply adverbs (now taken more broadly) which have no complement. But if this is so we can clearly identify P as subcategorizand in P+NP, thereby filling one of the boxes which Zwicky left empty.

<table>
<thead>
<tr>
<th>Semantic functor</th>
<th>V+NP (V)</th>
<th>P+NP (P)</th>
<th>NP+VP (VP)</th>
<th>Det+N (Det)</th>
<th>Aux+VP Comp+S (Comp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C) Morphosyntactic locus</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(D) Subcategorizand</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
</tr>
<tr>
<td>(E) Governor</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
</tr>
<tr>
<td>(F) Distributional equivalent</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
</tr>
<tr>
<td>(G) Obligatory</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
</tr>
<tr>
<td>(H) Ruler</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 3

Zwicky’s analysis of six constructions in terms of the six genuinely head-like categories, showing first revisions

The conclusion so far reached in this section, then, is that Aux and Comp are the distributional equivalents and the obligatory elements in their respective constructions, and that P is the subcategorizand in P+NP. Table 3 is a revised version of Table 2, in which these changes are included.

Two other questions arise in connexion with Aux + VP and Comp + S. The first is about the morphosyntactic locus in Comp + S. According to Zwicky it is the S, and the reason for this choice is that S is where the morphological markers of tense are. I take it that he accords this criterial status to tense because he believes that tense should percolate up from S on to the node dominating Comp + S, so presumably he has reasons for believing that this node should carry a tense feature – perhaps something to do with subcategorization of matrix verbs. However, since he does not explain his reasons it is hard to comment on the choice. What is clear, however, is that at least as good a case can be made for taking Comp as morphosyntactic locus, on precisely the same grounds as he offers for taking P as morphosyntactic locus in P + NP: that the selection of particular prepositions is analogous to case features (a very theory-particular decision associated with Generalised Phrase Structure Grammar (Gazdar et al., 1985: 23) and Lexical-Functional Grammar (Kaplan & Bresnan, 1982: 197)). Since there are verbs in English which select particular prepositions (e.g. inform of, tell to), we must show this preposition as a property of the verb’s sister, although it is actually manifested on the verb’s ‘niece’. Whatever the merits and demerits of this assumption, much the same is true, though on a smaller scale, of complementizers.
For example, the verb *doubt* allows either *whether* or *if* as complementizer in its complement (e.g. *I doubt whether/if it's true*). This fact cannot be stated in semantic terms, by allowing an interrogative structure as complement, because most kinds of interrogative structure are not in fact allowed after *doubt*—compare *I doubt why he did it*—and in any case, *whether* and *if* are here interchangeable with *that* (e.g. *I doubt that it's true*) without change of meaning. Similarly verbs seem to differ as to whether they accept either *if* or *whether*—so *discuss*, for example, seems to prefer *whether* strongly to *if* (e.g. *We were discussing whether/if we had enough money*). In other words, the subcategorization for verbs must be able to distinguish between clausal complements which are introduced by different complementizers just as it must be able to be sensitive to the prepositions in prepositional complements. In short, Comp has at least as good a claim to be taken as the morphosyntactic locus in Comp+S as S does.

The other question relates to Zwicky's view about the dependency analysis of Aux+VP, which leads him to take VP as what he calls the 'ruler'—alias the 'régissant' or 'Regens' of some dependency scholars, and the 'head' of others. Admittedly Zwicky is presenting what he takes to be the consensus of opinion among dependency grammarians, and he may well be right in suggesting that most would take the VP as 'ruler'. This is certainly a traditional view, which is firmly entrenched in the term 'auxiliary' verb, suggesting subordinating in relation to the 'main' verb inside the VP. However, many linguists now question this traditional view, and it is at least reasonable to suggest that all the arguments which are used to show that Aux is a sister of VP, in a constituency-based analysis, carry over to a dependency-based one. (This is certainly the view that I have taken in my own dependency-based work—cf. Hudson, 1984: 91.) I think, then, that if the entries for the 'ruler' of dependency grammar are to be taken seriously they must be updated for Aux+VP; so I shall assume that in this construction it is Aux, not VP, that is the ruler. This revision, together with the analysis of Comp as morphosyntactic locus of Comp+S, will be incorporated into the next version of our table.

6. Determiners

All the remaining stars in Table 3 are in the column for Det+N. However, let us start with the one entry which agrees with the standard. According to Zwicky, Det is the subcategorizand, on the grounds that 'determiners are lexically subcategorized according to whether they can combine with singular count nouns (*each penguin/*/penguins/*sand*), plural count nouns (many *penguin/penguins/*sand*) or mass nouns (*much *penguin/*/penguins/sand*). This judgment is again supported by other distributional facts about the common noun, namely that the possibility of ellipsis varies from determiner to determiner. For example, it is possible after *each*, but not after *every* (compare *The winners lined up and each/*every was given a standing ovation*);
after this but not after the; after some but not after a(n); and so on. All these facts can be dealt with very easily if the determiner is analysed as the subcategorizand, with lexical differences among determiners.

Zwicky’s other claims about Det+N are open to serious doubt. Consider first his view that N is the morphosyntactic locus. He justifies this by pointing to the distinction between singular and plural which is marked on the common noun, and which is relevant to distinctions in the VP. There are a number of other facts, however, which he does not mention although they are highly pertinent.

(a) Several determiners also mark the distinction between singular and plural – this/these, that/those, much, many, few, little, both, either, neither a(n) and all the numbers. And some common nouns do not show the number contrast morphologically – sheep, fish, etc. Notice too that in some varieties of English measure nouns are singular when they follow a number – five mile, three pound, etc. So it is at least debatable whether number is indicated by the noun rather than by the determiner.

(b) Two determiners are of particular interest: you and we (e.g. you Americans, we linguists). It seems clear that these words are determiners when combined with a common noun (Postal, 1966; Sommerstein, 1972), as witness for example the fact that they cannot combine with other determiners (in the absence of an intonation break to mark non-defining apposition) – e.g. *we the linguists, *you those students. One relevant fact about these determiners is that they have sole responsibility for marking person (relevant, for instance, to choice of reflexive pronouns: You children must behave yourselves/*themselves). Likewise they alone mark ‘case’ – the difference between subject and object forms. For example, we students contrasts with us students (at least in some varieties): We/*us students work hard, vs. The government is against us/*we students. In contrast, common nouns indicate nothing about either person or case.

The few morphosyntactic markers that there are in the English NP, then, seem to be distributed fairly equally between Det and N, so I shall reverse Zwicky’s decision by taking Det as the morphosyntactic locus. It should be clear that we could extend the discussion interestingly to other languages, and in at least some languages we should find even clearer evidence that Det is the morphosyntactic locus, rather than N – e.g. case is more clearly indicated on Det in German, and number is more clearly marked on Det in spoken French. However, for the present I shall restrict my discussion to English, following Zwicky’s lead.

What about the related concepts ‘distributional equivalent’ and ‘obligatory element’? Regarding distributional equivalence, Zwicky writes: ‘N is the distributional equivalent of Det+N, since the distribution of Det+N is roughly the same as the distribution of Ns like penguins and Kim’ (12). However, as Zwicky himself notes, it is by no means easy to decide whether Det+N is endocentric, so it is also hard to decide whether it has any
distributional equivalent at all. If you take examples like the boys, it is clear that boys is distributionally (more or less) equivalent, and that the is not (e.g. I know the boys/boys/*the); on the other hand, this boy points in precisely the opposite direction because boy on its own cannot be used as an NP, whereas this can (e.g. I know this boy/this/*boy). To quote Zwicky: 'the criterion [of distributional equivalence] can be used to argue that N is the head, that neither constituent is, or that Det is, depending on which set of facts you look at' (12 n. 8).

Part of the problem with thinking through this area of grammar is that we are constrained by a rather unsatisfactory set of terminology. In this respect the situation is rather similar to the one we considered earlier in relation to the terms 'preposition', 'adverb' and 'complementizer'. The terminology in the present case consists of 'noun', 'pronoun' and 'determiner'. The traditional assumption is that 'determiner' is a basic word class which cannot be subsumed under other classes, but there are good reasons for thinking that determiners are in fact a subset of pronouns, which are generally agreed to be a subset of nouns (or more precisely, a subset of nouns which are distributionally equivalent to NPs). Some of these arguments are already well known (e.g. Postal, 1966; Sommerstein, 1972), but others have been pointed out more recently (Hudson, 1984: 90 f.). The most striking connexion between 'determiner' and 'pronoun' is the large overlap of membership, if we follow the tradition in saying that any determiner which occurs without a following noun is a pronoun. In fact all but a handful of the determiners are also pronouns in this sense. But there are other striking similarities between determiners and pronouns, such as the fact that the same range of grammatical features are relevant in both cases (e.g. possessive, relative, interrogative, demonstrative, definite, indefinite, negative). In view of similarities as great as this it seems at least reasonable to assume that determiners are pronouns, and therefore nouns.

If we now return to the question of the distributional equivalent in Det + N, the situation looks rather different if we update the notation by writing 'N' instead of Det: N + N. In other words, we are now assuming that the words in this boy are both nouns. Thus if the distributional equivalent of NP must be an N, then this could be the determiner as easily as the common noun. And if we apply the test of distributional equivalence in some other way, which results in the determiner being selected as the distributional equivalent of the NP, then this result should now be more acceptable to those who believe, with X-bar theory, that the head of a phrase should have the same category features as the phrase.

I recognize that many readers will want much more detailed argumentation before they accept that determiners are just 'transitive nouns', as it were (parallel with prepositions as 'transitive adverbs'). However, I submit that the case for choosing the common noun as the distributional equivalent of the NP is at least less clear than Zwicky assumes. Similar remarks apply to
the choice of the obligatory element, but they follow in a fairly obvious way from what I have just said, so there is no need to make them in detail. The main point is that it is at least as easy to omit a common noun as it is to omit a determiner (e.g. *I didn't read either book/either/*book). Admittedly the former omission is an example of ellipsis, which Zwicky excludes a priori; but as I have already explained, we need not follow Zwicky in this respect. I assume, then, that it is Det, rather than N, which is the obligatory element in Det + N.

Finally we come to what Zwicky takes as the consensus of opinion among dependency grammarians regarding the status of Det and N. The situation here is much the same as it was in the case of Aux + VP: he reports a traditional view against which a number of arguments can be mustered even within the dependency framework. I shall present one such argument in Section 8. For the present I shall simply note that many of the things I have said above about Det + N are relevant to a dependency analysis, and indicate that Det, not N, should be taken as 'ruler'.

7. A HARMONIOUS ANALYSIS

Table 4 presents the results of the discussion. Zwicky's main theme is that the various head-like notions which he identifies in fact represent different, and competing, analyses of sentence structure. My contention is that exactly the contrary is the case: there is a quite remarkable degree of agreement among those of them that are genuinely relevant to headship, to the extent that there is no conflict whatsoever among them. This is not of course to say that no conflict will ever arise, if we apply the same categories to a different set of constructions from the six that Zwicky chose. All it shows is that a harmonious analysis of these particular constructions is possible. It is natural to draw the conclusion that the notion 'head' does indeed contribute a great deal to linguistic theory, because it - and it alone - is suited to bringing together the six distinct notions, which we can now take as (more or less) independent properties of the head of a construction: it is the semantic functor, the morphosyntactic locus, the subcategorizand, the governor, the distributional equivalent, and the obligatory element.

It is worth reminding ourselves that this list of head-like notions includes the semantic notion 'semantic functor', as well as the five strictly syntactic ones. This shows that 'head' is an important point of contact and congruence between the syntactic and semantic structures of a sentence, the point at which the semantic category 'functor' is mapped on to syntactic categories. If the congruence does indeed generalize beyond the six structures considered here, then clearly it limits very strictly the extent to which syntactic structures are arbitrary vis-à-vis semantic ones, and allows us to identify all sorts of mappings between syntactic and semantic structures which should not be found in natural languages. Further research will have to develop some of
these consequences for typology, child language acquisition and other fields. An excellent start has already been made in the typological field by Nichols (1986), in which it is shown that the morphological marker of a dependency need not be located on the dependent (as in the prepositions, case-markers and so on of familiar Western European languages) but may also be located on the head.

<table>
<thead>
<tr>
<th>Semantic functor . . .</th>
<th>V + NP (V)</th>
<th>P + NP (P)</th>
<th>NP + VP (VP)</th>
<th>Det + N (Det)</th>
<th>Aux + VP Comp + S (Aux) (Comp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C) Morphosyntactic locus</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>(D) Subcategorizand</td>
<td>=</td>
<td>=</td>
<td>*</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>(E) Governor</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>*</td>
</tr>
<tr>
<td>(F) Distributional equivalent</td>
<td>=</td>
<td>*</td>
<td>*</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>(G) Obligatory</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>(H) Ruler</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
</tr>
</tbody>
</table>

Table 4
Final revised version of Zwicky's analysis of six constructions in terms of the six genuinely head-like categories

A methodological problem arises in using generalizations based on categories which, like 'head', unify a range of different properties. The problem is that what unifies these properties is only a very strong tendency, so deviant cases can arise. We have already seen one such case in our discussion of English, when we noted that the semantic structure for *maybe control* is not as expected, with the head (*control*) providing the functor. If *They maybe control the penguins* has the same semantic structure as *It is possible that they control the penguins*, it is clear that the syntactic relations must be out of step with the semantic ones in at least one of the sentences because in one *maybe* is subordinate to *control*, and in the other the syntactic relation is reversed. Such cases are easy to accommodate in Word Grammar – indeed the theory leads us to expect them – but methodologically they make it somewhat harder to be sure in particular cases which element is the head, because different properties may point in different directions. The problem is not peculiar to the concept 'head', however – other important categories like 'subject' raise precisely analogous problems – so I see no reason to take it as evidence against the general claims I have just made.

I should also like to point out that these connexions between semantic and syntactic structure do not depend on acceptance of the distinction between 'argument' and 'functor' – a distinction which Zwicky himself notes is itself rather uncertain. Instead of referring to functors we could refer to Zwicky's original idea about the 'kind of' relation between the referents of the head.
and of the whole construction (i.e. the construction refers to a kind of X, where X is what its head refers to). This is a much more general notion which can be formulated in terms of a variety of theories of semantic and cognitive structure. The general picture of how syntax and semantics relate to one another is then a very simple one, in which the head of a construction provides a general semantic notion which is narrowed down by the various other elements in the construction.

A further consequence of our discussion is that the notions of ‘head’ as this term has been applied in a variety of theories are not, in fact, as disparate as Zwicky suggests. In particular, he is wrong in suggesting that ‘the notion of head that Anderson and Hudson [two advocates of dependency theory] are using is... probably not the same notion as the head in syntactic percolation (11 n. 7). Different linguists may use the notion ‘head’ for different purposes – one for percolation, another for government, and so on – but this is to be expected in view of the multiplicity of properties that we have found for heads. In the same way, one might refer to the notion ‘subject’ in rules for word order, for agreement, and so on, but it is the same concept that is being referred to in each case. One could thus see the increased use of the term ‘head’ as a genuine case of convergence among theories, a particularly encouraging development considering how different in other respects some of the theories concerned are.

I shall finish the paper by discussing two further uses of the term ‘head’ – in other words, two more ‘head-like concepts’ – which are not on Zwicky’s list.

8. Adjacency

Zwicky refers rather briefly (and somewhat sceptically) to the possibility of formulating word-order rules in relation to ‘head’. The rules he has in mind are those for linear ordering of sisters in a construction, such as the rule for Japanese: heads follow non-heads. This possibility has been recognized for a long time in the dependency literature, and is already quite well developed in Tesnière (1959). Unfortunately the generalizations that can be made on this basis for English are not particularly impressive, because some heads follow and others precede. For some other languages the benefits of being able to refer to heads are much more obvious, but I shall not pursue this point here. Instead I shall take up another aspect of word-ordering to which Zwicky does not refer, namely the question of adjacency. Recent versions of dependency theory have used a principle called the Adjacency Principle (and first formulated) by Robinson (1970). This principle has a crucial part to play in dependency theories which eschew constituent structure, because it is solely responsible for keeping phrases continuous. (A phrase can be defined, derivatively, as a word plus all the words which are subordinate to it, in the sense of subordinate to be defined below.) Roughly speaking it says that a word must be as close as possible to its head – i.e. to the head of the
construction containing it. (From now on I shall assume that in ‘head of X’ X is a word depending on the head, rather than a phrase containing it; the two terminologies are largely interchangeable, but only the former is appropriate to dependency theory.)

In my own theory, Word Grammar, the Adjacency Principle can be expressed as follows (similar formulations have been given by others, e.g. Robinson, 1970; Anderson, 1976).

**ADJACENCY PRINCIPLE**

If A is the head of B, and some word C separates them, then it must be the case either (i) that C is subordinate to A, or (ii) that C is (subordinate to) the head of both A and B.

The term ‘subordinate to’ is used in the sense of Anderson & Durand (1986) to refer to an extension of the dependency relation. If X depends on Y and Y on Z, then X is subordinate to Z, although it does not ‘depend’ on it in the strict sense of having it as its head. (This distinction between ‘dependent’ and ‘subordinate’ is intended to replace the distinction I made in previous work between ‘direct’ and ‘indirect’ dependency.) Formally speaking:

- A depends (or is dependent) on B if B is the head of A;
- A is subordinate to B if A is dependent on B, or if A depends on C and C is subordinate to B.

To make the discussion more concrete, imagine three words, *with, difficulty* and *great*, and assume that *difficulty* depends on *with*, and *great* on *difficulty*. We can indicate these dependencies in the traditional way by means of the vertical dimension:

```
with
difficulty
great
```

Let us also assume two rules for the ordering of heads relative to their dependents, whose effects guarantee the following orderings:

```
with difficulty
great difficulty
```

The question is, what determines the ordering of *with* in relation to *great*? The rules given so far permit either of *with great difficulty* and *great with difficulty*, so how do we exclude the second of these? This is the function of the Adjacency Principle. (Since this is a universal principle, special dispensations will be needed for cases in other languages which seem to infringe it, such as Latin *summa cum laude*, ‘highest with praise’ – i.e. ‘with the highest praise’.)

To see how the principle works let us look at the dependency structures, using the Word Grammar notation in which arrows point from the head towards its dependents:
Structure (a) is permitted by the Adjacency Principle because difficulty is separated from its head, with, by a word – great – which is subordinate to with, thereby satisfying clause (i). It will be seen that great is subordinate to with because the former depends on difficulty, which in turn depends on with. In contrast, structure (b) does not satisfy clause (i) because great is separated from its head, difficulty, by with, which is not subordinate to difficulty. Nor does it satisfy clause (ii), because with is not the head of both great and of difficulty. Consequently (b) is ill formed.

I should explain why clause (ii) is needed. It is meant to cover a variety of constructions in which an element is displaced – so-called subject-raising, subject-control, tough-movement, extraposition, long extraction (alias unbounded dependencies – e.g. wh-movement), and a number of other possibilities. In each of these cases a special rule assigns some word a second head, so one of its heads is allowed (by clause (ii)) to intervene between it and its other head. Explaining and justifying these analyses would take us much too far afield for present purposes, so I shall just illustrate the operation of clause (ii) in relation to subject-raising verbs, as in Fred keeps talking. I assume that Fred is subject not only of keeps, but also – and in the same structure – of talking. (Similar structures are also assumed in Lexical-Functional Grammar – see Bresnan, 1982a.) If X is subject of Y, then Y is the head of X, so Fred has two heads; the two verbs, one of which – keeps – is also head of the other. The structure is as follows:

Fred keeps talking

It can be seen that Fred is separated from one of its heads, talking, by a word which is not subordinate to the latter, so it is not permitted by clause (i). However, this word is also the head of both of them, so clause (ii) covers the structure and it is well formed.

The point of this discussion is to show the important part played by the Adjacency Principle in dependency theory, which in turn shows how important the notion of ‘head’ is in specifying adjacency, because the principle refers crucially to the asymmetrical relation between ‘head’ and ‘dependent’. Let us take just one particularly simple instantiation of clause (i), expressed as follows. If A is head of B, and some word C separates them, then C may depend on B. This covers cases like our earlier example with great difficulty, where great depends on difficulty; and it excludes cases like *great with difficulty because the intervening word with is head rather than dependent of difficulty. If the direction of dependency – i.e. the difference between ‘head’ and ‘dependent’ – had been irrelevant, there would have been no difference in status between these two examples, because in both cases the dependency
relations are in other respects the same (each of great and with has a dependency relation to difficulty).

As I promised in the discussion of determiners, the Adjacency Principle throws extra light on the relation between Det and N in Det+N. Imagine a sequence of determiner + adjective + common noun, such as the big book. Now if the common noun were the head of the determiner, as in the traditional analysis, the dependency relation between them would be precisely the same as that between big and book, where book is clearly the head. But if that were the case, there would be no explanation for the fact that the determiner always precedes the adjective; why should it not be allowed to follow, possibly subject to some pragmatic constraints such as those on the relative order of adjectives? E.g. why should *big the book be so much worse than a phrase like red big books (compared with big red books)? If on the other hand we adopt the analysis which I assumed in the earlier discussion, according to which it is the determiner which is head of the common noun, then the word-order restrictions follows absolutely automatically from the Adjacency Principle. That is, *big the book is bad for exactly the same reason as *great with difficulty.

The discussion in this section has shown the crucial importance of the notion ‘head’ in dealing with the continuity of phrases in a dependency-based analysis. A good deal of research is needed before we can be sure that the Adjacency Principle is compatible with all we know about English, not to mention other languages, and of course even more research is needed to evaluate the general principles of dependency grammar, but so long as dependency theories invoke the Adjacency Principle, they must also invoke the notion ‘head’.

9. Grammatical relations

One further point is worth making in connexion with the notion of ‘head’. Let us consider again Zwicky’s list of constructions, together with the constituent in each which we have now decided is its head:

(i) V + NP     head: V
(ii) P + NP     P
(iii) NP + VP   VP
(iv) Det + N   Det
(v) Aux + VP   Aux
(vi) Comp + S  Comp

This list suggests a generalization about grammatical relations, namely that any constituent to which an established grammatical relation category (GRC) applies is not the head. Thus in V + NP, the only relevant GRC is ‘object’ – i.e. the NP, not the head V; and similarly for P + NP. In the case of NP + VP, the head is VP, and the relevant GRC is ‘subject’, which applies
RICHARD A. HUDSON

to the NP. So far so good. When we come to the last three constructions, it is less obvious how to apply this generalization, because GRCs are not usually applied at all to these constructions, partly because of the analyses that have traditionally been assumed. We have considered the analysis of these three constructions, and have decided that in each case the head – Det, Aux or Comp – is the subcategorizand, which means that we can now apply at least one established GRC, namely ‘complement’. This is after all the term that is now widely used to refer to any element with respect to which some word or class of words is subcategorized. But of course it is again the non-head in each case which is classified as complement, as predicted by our generalization.

These remarks are meant to be only suggestive, but they can be developed in terms of the current Word Grammar treatment of grammatical relations (see Hudson, 1985a, b, 1986). In this theory GRCs are assumed to be distinct from non-relational categories such as ‘noun’ or ‘auxiliary verb’, but similar to them in being organized in an ‘isa’ hierarchy. In this hierarchy the first contrast is that between ‘head’ and ‘dependent’, and then ‘dependent’ is subdivided into GRCs, some familiar and others unfamiliar. The hierarchy for English may include the one shown in Figure 1.

If this hierarchical view of GRCs is in fact correct, at least in outline, then once again we find that ‘head’ is a crucial category in grammatical theory, because it is at the very top of the hierarchy. Moreover, every other GRC is defined in relation to ‘head’, because wherever X is the R of Y (where R stands for a GRC other than ‘head’), Y is the head of X. (It should of course be remembered that in dependency theory a GRC, like the more general dependency relation, is defined in relation to single words, and does not define the relation between a construction and its constituent parts as in constituency-based theories.) For example, if X is the object of Y, then Y is the head of X; if X is the complement of Y, Y is the head of X; if X is the subject of Y, Y is the head of X; and so on.

Figure 1
In conclusion, then, ‘head’ is the name of a grammatical relation category, on a par with categories like ‘subject’ and ‘object’, but on a higher level of generality than these. Like them it serves the essential function of allowing generalizations across rules which could not otherwise be made – for example, a category which is referred to in a word-order rule can be identified, by name, with a category referred to in a rule concerned with agreement, or subcategorization, or whatever. It is possible to justify the category in terms of the kind of analysis that Zwicky himself assumes – one based on constituent structure – and I tried in the first six sections of this paper to present my arguments in such terms. However, if we adopt instead a dependency-based theory we can go further, and integrate ‘head’ into a comprehensive system of grammatical relation categories – something which I doubt that one could do in a constituent-structure theory.

REFERENCES


I31 5-2