Splitting Heads

Ronnie Cann

Dept. of Linguistics
University of Edinburgh
40 George Square
Edinburgh EH8 9LL
Email: ronnie@ling.ed.ac.uk

Earlier versions of this paper were delivered to the 1987 Autumn meeting of the LAGB in Bradford and the Linguistic Circles of the Universities of Edinburgh and Lancaster. I am grateful to all who gave comments on these presentations and to those who commented on earlier versions of this paper, particularly, Dick Hudson, Jim Miller, Mary Tait, and Nigel Vincent. None of those named are in any way responsible for any errors of fact, argumentation or judgement in the following discussion.

1. Heads

The use of some notion of head is almost universal in current grammatical theory. Indeed, such is its pre-eminence that, not only does it hold a central position in a number of frameworks[1], but it is also generally assumed to be a theory independent notion. This is asserted directly in, for example, Nichols (1986) where 'headedness' is introduced as a 'theory independent notion which in fact figures as a primitive in almost all theories - and which, although not directly given in linguistic data, is often directly reflected in such structural features as word order' (Nichols 1986:56)[2]. The apparent consensus among linguists concerning the existence of heads is also often assumed to be matched by a consensus about which expressions are the heads of particular constructions. Indeed, in the quoted article Nichols goes on to assert that 'linguists of divergent theoretical persuasions are in almost complete agreement as to what is the head and what is the non-head in given construction (Nichols 1986:57).

Nichols' confident assertions about the theory independence of the notion and about the consensus of opinion about what properties heads have and which expressions in a phrase are heads, is somewhat premature. Although, as will be shown below, it is possible to give some pre-theoretical content to the notion 'head', the relation between heads identified in this way and their use in theory is not straightforward by any means. Indeed, the connection between pre- theoretical ideas about headedness and theoretical applications of the notion are in certain cases so tenuous as to indicate that far from being theory independent, headedness has meaning only within particular theories. Merely because different theories use the same term does not mean that it has the same interpretation in each one. Furthermore, while it is true that for certain constructions, e.g. verb or preposition plus noun phrase or adjective plus noun, there is general agreement about which expression constitutes the head of the phrase, there are many other constructions in which the identification of the head is controversial, e.g. in determiner plus noun and complementizer plus sentence constructions. Such controversy is problematic for theories that assume the non-theoretical status of the concept of headedness, since the assumption that head is a pre-theoretical notion should require that there be some theory independent way of testing for heads in particular constructions.

The problem is that, as Nichols notes, heads are not directly given by the data and so must be identified indirectly, using independent criteria. Unfortunately, while many tests for heads have been proposed, which tests are appropriate, what their form should be and how they should be applied varies almost from writer to writer. For example, Nichols herself 'briefly' takes the head to be 'the word which governs, or is subcategorized for - or otherwise determines the possibility of occurrence of - the other word' and the expression that 'determines the category of its phrase' (Nichols (1986:57)). For Lyons, on the other hand, the head is 'syntactically equivalent to the whole phrase' (Lyons (1977:391)). In an article primarily concerned with the extension of the notion of head into morphology, Arnold Zwicky discusses some eight informal tests that may be used to identify heads in syntactic constructions (Zwicky 1985). These are summarized in (1).

1a The determinant of concord.
1b The governor.
1c The subcategorizand.
1d The morphosyntactic locus.
1e The distributional equivalent.
1f The obligatory constituent.
1g The 'ruler' in dependency relations.
1h The semantic argument.

However, from an application of these tests to the six English constructions shown in (2), Zwicky concludes that these tests yield conflicting results and that only the property of being the morphosyn-

1. Cf. X-bar theory (Jackendoff (1977)), Generalized Phrase Structure Grammar (Gazdar et al. (1985)), Head Driven Phrase Structure Grammar (Pollard and Sag (1988)) and Word Grammar (Hudson (1984)), inter alia.
2. Quoted also in Vincent (1988).

tactic locus is a reasonable diagnostic for being the head of a phrase. The results that lead him to this conclusion are shown in Table 1.[3]

2a V + NP: control those penguins.
2b P + NP: toward those penguins.
2c NP + VP: we control those penguins.
2d Det + N: those penguins.
2e Aux + VP: must control those penguins.
2f Comp + S: that we control those penguins.

Table 1: Zwicky's Results

Construction:	V+NP	P+NP	NP+VP	Det+N	Aux+VP	Comp+S
Concord	NP	*	NP	N	*	*
Governor	V	P	VP	*	Aux	*
Subcategorizand	V	P?	*	Det	?	Comp
Morphosyntactic Locus	V	P	VP	N	Aux	S
Distributional Equivalent	V	*	*	N	VP	S
Obligatory Constituent	V	P	VP	N	VP	S
Ruler	V	*	*	N	VP	Comp
Semantic Argument \NP	NP	NP	NP	N	VP	S

In a reply to Zwicky, Hudson (1987) contests this conclusion and sets out to show that, by changing some of the tests and some of Zwicky's assumptions, it is possible to arrive at an 'harmonious analysis' whereby the diagnostic tests converge on a single item in each of the chosen constructions, this being the head of that construction. In the first place, Hudson denies that being a determinant of concord is relevant to the status of an expression with respect to headedness and takes the semantic functor rather than the argument as a diagnostic. Having reduced Zwicky's eight tests to seven, Hudson re-analyzes the six canonical constructions in (2) and comes up with the pattern of results shown in Table 2.

Table 2: Hudson's Results

Construction:	V+NP	P+NP	NP+VP	Det+N	Aux+VP	Comp+S
Governor	V	P	VP	*	Aux	*
Subcategorizand	V	P	*	Det	*	Comp
Morphosyntactic Locus	V	P	VP	Det	Aux	Comp
Distributional Equivalent	V	*	*	Det	Aux	Comp
Obligatory Constituent	V	P	VP	Det	Aux	Comp
Ruler	V	*	*	Det	Aux	Comp
Semantic Functor	V	P	VP	Det	Aux	Comp

3. In Tables 1 and 2, * indicates that the test selects neither constituent and in Table 1? indicates that Zwicky does not discuss the test with respect to the particular construction.

Interestingly enough, Zwicky's decision to select the morphosyntactic locus as the only test for headedness and Hudson's analysis summarized in Table 2 result in their selecting the same constituents as heads in all but two of the constructions: Det+N and Comp+S. Zwicky takes the noun as the head of determiner plus noun constructions in English and the sentence as the head of complement clauses, while Hudson picks the other constituent in both cases.

The controversy between Zwicky and Hudson over which tests for heads are appropriate, the way they should be applied and the results they yield indicates that the unanimity among linguists on the issue is illusory. There are major problems for any attempt to identify heads (or indeed any other linguistic concept) non-theoretically. In the first place, of course, one must decide what tests are indeed relevant for headedness and then provide definitions of these that are robust enough to be applied by different researchers to a range of constructions. One of the reasons that Zwicky and Hudson can disagree so radically about which constituent is selected by different tests is that neither provide fully explicit definitions of the tests, relying solely on vague intuitive notions that can be interpreted in many different ways. The issue of headedness is also clouded by the assumption that heads are unique or that certain headlike properties can be associated with only a single constituent within a phrase. Such an assumption, too widely held and too rarely contradicted, is undermined by the argument between Zwicky and Hudson and, its abandonment has serious implications for theoretical uses of headedness.

The purpose of this paper is argue against the validity of certain of the tests in (1) as indicative of headedness and to provide more rigorous definitions of the ones that remain, so that their consequences can be properly examined. These definitions of the tests will be applied to the two constructions whose heads are in dispute, i.e. Det+N and Comp+S. The results obtained will show that at least three headlike properties are required in the grammar and that the tests to identify them may select both constituents in a phrase. This leads directly to the recognition that phrases may be doubly headed and this conclusion, together with the ancillary hypotheses that are needed to make the notion 'head' harmonious, will be shown to have consequences, both specific and methodological, for theoretical uses of the concept.

2. Testing for Heads

I shall begin the discussion of Zwicky's tests for heads by rejecting two of them immediately: the semantic argument or functor and the 'ruler' in dependency theory.

Both Zwicky and Hudson appeal to semantics as one criterion that may be used to identify heads, under the apparent assumption that syntactic and semantic heads coincide. Zwicky chooses the semantic argument, and Hudson the semantic functor, as corresponding to the syntactic head. In fact, Zwicky conflates two sorts of semantic property into his argument category. He begins by defining X as the semantic head in a phrase X+Y if X+Y describes 'a kind of the thing described by X' (Zwicky (1985:4)), pointing out that the phrase *that penguin* describes a kind of penguin and *will leave* describes a kind of leaving. Using the assumption that the noun in Det+N (and the VP in Aux+VP) is a semantic argument, Zwicky then extends the property of being a semantic head to all semantic arguments in his exemplifying constructions. Hudson objects to this move, pointing out that the 'kind of' property does not extend particularly well to the arguments of the other constructions not specifically mentioned by Zwicky in this context and that Zwicky's assumptions about the 'kind of' property with respect to determiner plus noun (and auxiliary plus verb phrase) can be challenged. Indeed, Hudson argues that it is the functor in each of Zwicky's six exemplifying constructions that best represents the 'kind of' property.

Unfortunately, neither functor nor argument are actually defined by the two authors and Zwicky himself points out in a footnote that caution is required in the selection of which of two co-constituents is functor and which argument because of the fact that a logic powerful enough to represent the meanings of natural language expressions can be tailored to pick out either expression as functor. He continues: 'The intuitively correct assignment of certain items as arguments and others as functors can be guaranteed only if substantive assumptions about the relationship of semantics and syntax are made' (Zwicky (1985:4)). This is something of an understatement. The whole notion of functor and argument depends almost entirely on the type of semantic theory selected (Montague Semantics, Situation Semantics, Discourse Representation Theory, and so on) and, despite Zwicky's (and Hudson's) assumption, there is, in fact, often no obviously intuitively identifiable functor and argument in non-verbal phrases. While intuition may favour the identification of predicates, relations and arguments in sentences and VPs, it does not reliably tell us anything about the functional status of the constituents in such constructions as Det+N, Comp+S, P+NP or even Aux+VP. Indeed, it is unlikely that any non-theoretical content can be given to notions such as semantic functor and argument which go beyond matters of predication and relation amongst individuals. Since the latter would have to be restricted almost exclusively to VP and S[4], its use as a test for headedness is too limited to be seriously considered as a diagnostic of headedness.

The 'kind of' test, however, seems more robust than whether an expression is a functor or an argument, but its apparently restricted applicability and imprecision make it an unreliable diagnostic. As mentioned in Hudson (1987), the 'kind of' test does not generalize well to other constructions not discussed by Zwicky. It is, for example, intuitively difficult to decide which constituent is picked out by the test in a prepositional phrase like *with the penguin*. It appears to be neither a kind of penguin or nor a kind of togetherness (or instrument, or whatever). The fact that it cannot be pre-theoretically defined with any precision allows Hudson and Zwicky to argue over which of the co-constituents in Det+N (or Aux+VP) is selected by the test. It seems better, therefore, to reject the 'kind of' test as a diagnostic for heads and concentrate on the more concrete syntactic tests.

There is another reason for ignoring semantics when investigating the property of headedness. One of the motivations for adopting the notion 'head' is that it may provide a means for 'syntax... to interface with other components of the grammar, in particular semantics, the lexicon and (inflectional) morphology' (Zwicky (1985:3)). However, apart from the fact that the notion is rarely, if ever, used in semantic theories, it is not clear that semantic heads do necessarily coincide with syntactic ones. Consider, for example, the 'case- marking' prepositions in English as in the expression *of the penguins* in the phrase the shooting *of the penguins*. As has often been recognized, the preposition of has no real semantic force in the phrase. It does not, for example, itself correlate with any particular participant role (cf. e.g. *the dancing of the penguins* (agent) or *the destruction of the city* (patient), etc.) or indicate any other specific property or relation between the noun *shooting* and the noun phrase *the penguins*. Because of this fact, we could say that *of the penguins* denotes some kind of penguin, i.e. those penguins under discussion. More persuasively, another possible test not discussed by Zwicky and Hudson but mentioned in Huddleston (1984) and Cruse (1986) strongly selects the NP and not the preposition.[5] This test selects as semantic head a constituent in a phrase whose appearance is semantically restricted by some other expression in a larger phrase in which it appears as in the location of selectional restrictions within verb phrase (e.g. *shoot the penguin* versus?shoot *the teardrop* indicates that N is restricted by the verb). Thus, while *Janet's shooting of the penguin* is semantically coherent and non-metaphorical, *Janet's shooting of the teardrop* is anomalous. Hence, it may be argued that *the penguins* is the semantic head in *of the penguins*. From the syntactic point of view, however, it is the preposition that is the head, being at least the subcategorizand, the governor and the morphosyntactic locus of the phrase. Hence we must conclude that syntactic and semantic heads (if the latter is really a coherent notion) do not necessarily coincide and this, together with the problems previously mentioned, must force us to abandon the test as a diagnostic for headedness.

Another head-like characteristic which I reject is one utilized by dependency grammarians. Zwicky calls this the property of being the 'ruler' in some construction, the term covering a number of different names that are given to the head in dependency constructions (cf. Zwicky, p. 14 and references there). In dependency theory, the ruler of a phrase is that expression in the phrase on which other expressions depend. Unfortunately, however, it is very difficult to pin down a more precise definition of this notion. In fact, it does not seem to be an independent criterion for headedness at all. Matthews (1981) seems to see rulership in terms of the government or the semantic connectedness of items with respect to predication, valency and (semantic) modification. For Hudson in his 1984 book, the ruler (Hudson's 'head') is defined by lexical selection (subcategorization and collocational restrictions), as

4. It could be extended to cover noun plus complement or noun plus possessive structures, but the application still remains very limited.
5. I am grateful to Nigel Vincent for bringing my attention to this further test.

well as government and word order. Zwicky refers to Hudson (1980) and claims that his definition in terms of a ruler being both governor and the controller of concord as well as the semantic argument and the subcategorizand is 'simply contradictory' (Zwicky (op.cit.:14)). He then suggests that the consensus view amongst dependency grammarians is that the ruler is, in endocentric constructions, 'the distributional equivalent/obligatory constituent; for exocentric constructions, it is the governor' (Zwicky: 15).

Whatever the merits and demerits of the particular definitions of the property given by the different authors, it seems to me they all rely on other properties or combinations of properties taken here and elsewhere to be characteristics of heads. Rulership, therefore, is a notion of the same sort as headedness, dependent on the same sort of criteria and so to define heads in terms of this property would be circular.

Abandoning rulership and semantic considerations as tests for headedness leaves us with six of Zwicky's eight tests: subcategorization, government, concord, distributional equivalence, obligatory constituency and the location of morphosyntactic properties.

2.1 Subcategorization

On the constituent picked out by the criterion of the subcategorizand, both Zwicky and Hudson agree: determiners subcategorize for nouns and complementizers for sentences, and not vice versa. So much seems self evident, but, as with many apparently self-evident things, this conclusion is not so easily demonstrable. The first problem that we face is the definition of subcategorization itself. Zwicky defines a constituent as being 'subcategorized with respect to its ability to occur with a particular set of co-constituents' (Zwicky (op.cit.:5)) and points to a close relation between subcategorization and semantic interpretation. Furthermore, he assumes, as do most other theorists, that subcategorization is a property exclusively associated with lexical expressions and so is not applicable, a priori, to any phrasal constituent.

Unfortunately, even though the latter condition automatically excludes S in Comp+S and, at least on the most popular analyses, N in Det+N (usually taken to be an exponent of a phrasal category Nom or N'), ascertaining whether Comp and Det are subcategorizands, given Zwicky's somewhat loose characterization, is no easy matter. There seem, in fact, to be at least three non-equivalent ways of interpreting the subcategorizand in any phrase, but before we go on to examine these let us put the discussion on a slightly more technical footing and give a more precise characterization of syntactic subcategorization. It is difficult to give precise, but completely theory independent, definitions of grammatical phenomena, but it is possible to make use of some non-linguistic tools which, although necessarily theory-bound in terms of mathematics, may be considered pre-theoretical from the linguistic point of view. Basic set theory forms the foundation of many grammatical theories and I shall use it in the following sections to make as precise as possible the properties that we are examining without recourse to particular syntactic theories.

Any syntactic category may be extensionally defined as a set of strings of word-forms (in the sense of Lyons (1977), etc.), i.e. just those strings that are of that particular category[6]. These expression-forms (which may, of course, consist of only a single word-form) may be grouped together in various ways, reflecting common morphosyntactic (or other) properties. One obvious way of partitioning them is by grouping together all those strings that are identical but for the inflectional form of one or more of its constituent expression-forms. Unfortunately, there is no generally accepted term for groups of inflectionally related strings, so, at the risk of yet further increasing the amount of linguistic jargon, I shall refer to them as 'phrasemes' on the model of the familiar lexeme, which can itself be viewed extensionally as a collection of the inflectional variants of a word (its paradigm) [7]. The set of phrasemes defined by some set of expression-forms, X, will be denoted as p(X) and given the definition in (3). Two strings are inflectional variants if they are identical but for the inflectional form of one or more word-forms in the string.

6. This, of course, begs the question of how these sets are defined, a subject that I have no intention of going into here. For the purposes of this paper, the sets of expressions that make up particular categories are assumed to be already defined.

7. Note that phrasemes are not defined by strings of lexemes but sets of strings of word-forms.

3 Given a set of expression-forms, X, p(X) $=_{def}$ {Y| Y is a subset of X and, for each x and y in Y, x is an inflectional variant of y}.

For example, in English the category VP will contain as a subset the set {*control that penguin, control them, control those penguins, controlling that penguin, shot the penguins, shoots the penguin, laughs, laughing, laugh*}. Calling this set VP_1, $p(VP_1)$, the set of phrasemes associated with VP_1, is given by the partition: {{*control that penguin, control those penguins, controlling that penguin*}, {*shoots the penguin, shot the penguins*}, {*control them*}, {*laughs, laughing, laugh*}}. Clearly, where the expression- forms consist of only a single word-form what we have in effect is a lexeme. In what follows, the distinction between a category taken as a set of forms and the set of phrasemes defined on it will be crucial. Where the distinction is not germane, however, the general term 'expression' will be used as the superordinate to both phraseme and expression-form.

We can now define subcategorization in terms of the possibilities of grammatical combinations of strings of different categories. Since in subcategorization we are not interested in the forms of the expressions being combined, we can define the property in terms of the combinations of phrasemes. A phraseme, p, of one category, X, is said to grammatically combine with a phraseme, q, of another, Y, if (and only if) there are expressions x in p and y in q such that x+y (the concatenation of x and y, considered here and below to be order-free) is a member of some category, A. Thus, X can be defined as subcategorizing for a category Y, if the ability to combine with phrasemes of Y partitions the phrasemes of X into two sets: one set containing all the phrasemes of X that can combine with some phraseme of Y, and the other containing all those that can combine with no phraseme of Y[8]. This gives us the definition in (4).

4 Subcategorization: X is the subcategorizand in X+Y if the ability to appear with Y partitions the set of phrasemes of X into two subcategories.[9]

We may test for subcategorizands by varying the phrasemes instantiating X and seeing whether the grammaticality of the phrase is altered. Furthermore, variation of the phraseme instantiating Y should not affect grammaticality, modulo selectional restrictions and form dependencies (for a discussion of the latter of which, see the next section). Thus, we can say that V is the subcategorizand in V+NP because *control those penguins* is grammatical but the replacement of control by another verb, e.g. *laugh*, yields an ungrammatical expression, **laugh those penguins*. On the other hand, varying the NP makes no syntactic difference (although it may give rise to semantic anomaly): e.g. *control every variable/his moods*, etc.

The definition in (4), however, is fairly weak and has two non-equivalent interpretations according to the sort of partition of X involved. In the strong case, the partitioning of X must be into two non-null sets: there must be some members of X that cannot appear with any expression of Y and some that can. The weaker version is the one apparently assumed by Zwicky in that it allows for all members of X to be able to appear in construction with expressions of Y (i.e. the partition of X may trivially consist only of X itself). This latter interpretation is satisfied simply if the expressions of a certain category can appear with a co-constituent of some other category, while the former actually divides categories into discrete subcategories[10].

---

8. In order to make this and following definitions as general as possible, we must allow a whole category to be associated with a property that just some of its members share. Thus, the properties of a syntactic category is just the union of all the properties of its individual members. Stronger conditions may be imposed by quantifying over the numbers exhibiting a particular property, but this is not attempted here.

9. Formally, this can be defined using first order predicate logic in the following way. Assume that $part_e(X)$ denotes the partition of a set X according to the ability of its elements to appear in construction with an element e of a set Y to give an expression of category A, i.e. $part_e(X) =_{def} \{\{x| x \in X \& e+x \in A\}, \{y \mid y \in X \& e+y \in A\}\}$, then X is a subcategorizand in X+Y iff. $\exists q \in p(Y)[part_q(p(X)) = \{p(X^+),p(X^-)\}]$.

10. Thus, for X, Y and A, categories in the language, the weak form reduces to the statement:

$\exists p \in p(X)\ \exists q \in p(Y)[p+q \in A]$,

while the strong form becomes:

$\exists p,r \in p(X)\ \exists q \in p(Y)[p+q \in A \ \& \ r+q \notin A]$

The non-equivalence of these two interpretations can be seen from an application to our two exemplifying constructions. Since all determiners may appear with a noun, modulo variation in number (see the next section), just as all complementizers may appear with a sentence, modulo finiteness, the strong version of (4) fails to mark either Det or Comp as subcategorizands. The weak version, however, gives the opposite result with both constituents being marked as weak subcategorizands by virtue of the fact that they can combine together to form noun phrases.

There is a third way of interpreting (4), one that often seems to be assumed in the literature. In this version, subcategorizands are defined over a range of environments yielding roughly the same type of grammatical phrase and not just the one being examined. It is the case that for some categories certain environments pick out a category as a weak subcategorizand and others as a strong one. So, for example, all prepositions in English may appear with a following NP, making P a weak subcategorizand in P+NP. However, P becomes a strong subcategorizand in P+PP (or P+S), since not all prepositions can appear with a following prepositional phrase (or sentence) (cf. Quirk et al. (1985)). This adds a third degree of strength to the interpretation of (4),coming between the strong and weak versions, allowing X to be defined as an intermediate strong subcategorizand in X+Y, if there is some environment, Z, replacing Y, that properly partitions X where X+Z and X+Y are phrases of similar grammatical type.

For our two exemplifying constructions, this gives the same result as the weak version. Det can be shown to be an intermediate strong subcategorizand in Det+N, because the replacement of N by A or PP properly partitions the category, as shown in (5).

5a Those/*the/many/*a on the shelf (are very good).
5b The/*a/some/*every/*all/?many poor (are always with us).

For Comp, intermediate status is granted by the fact that items such as *since* may take a following NP unlike *that* or *when,* while *because* may take a PP unlike *since,* as shown in (6).[11]

6a Since yesterday/*of yesterday.
6b Because *the penguins/of the penguins.
6c When *the penguins/*of the penguins.

The results for the three versions of subcategorization are summarized in table 3.

Table 3: Subcategorization

Subcategorizand	Det	N	Comp	S
Strong	-	*	-	*
Intermediate	+	*	%+	*
Weak	+	*	+	*

Notice that from these results we can see that the test prefers a single item in each construction: Det being an intermediate strong, and Comp a marginally intermediate strong, subcategorizand. This result obtains only because S and N are automatically excluded from consideration by the assumption that subcategorization is a property only associated with lexical items. There is nothing, however, in the definition of subcategorization in (4) that requires this restriction. Since there are at least two theoretical frameworks that allow non-lexical subcategorization (Head Phrase Structure Grammar and Categorial Grammar), it will be an instructive exercise to apply the test to N and S in the two constructions.

In Comp+S, S comes out as a weak subcategorizand only. It cannot be strong because any sentence can appear with any complementizer (apart from matters of finiteness which I take to be a matter of government and not subcategorization, see below), as illustrated in (7). Because there are no environments yielding a commensurate construction with Comp+S (V+S yielding a verbal phrase not a clause and traditional P+S being subsumed by Hudson under Comp+S), S cannot be an intermediate strong

11. This result does not hold if the traditional conjunctions like *because* and *since* are deemed not to be complementizers. Without them, Comp would only be weakly subcategorizing.

subcategorizand. On the other hand, it must weakly subcategorize for Comp simply by being able to appear with this category.

7a That/because/whenever/if/whether we control those penguins.

Interestingly enough, however, in Det+N, the noun comes out as strongly subcategorizing, because there are nouns (i.e. personal pronouns) that can never appear with a determiner, as shown in (8). Hence, Det properly partitions N into two non-null subcategories.

8a *The/*all/*some we (control those penguins).

The revised results for subcategorization, given in Table 4 show a cline in strength in subcategorization from the noun (strong) through the determiner (intermediate strong) to Comp (marginally intermediate) to S (weak)..

Table 4: Subcategorization

Subcategorizand	Det	N	Comp	S
Strong	-	+	-	-
Intermediate	+	+	%+	-
Weak	+	+	+	-

2.2 Government and Concord

The next two tests for headedness involve the dependence of one constituent on another: government and concord. To provide a pre- theoretical characterization of these properties, we must look to their traditional interpretations in terms of the formal dependence of one constituent on another. Form dependence of this sort may, therefore, be tested by holding the form of one of the expressions steady and varying that of the other. If ungrammaticality results for some variant, then the latter is form-dependent on the former. The difference between government and concord depends on whether the formally dependent item varies its form according to the form of the controller. If it does not vary, then we are dealing with government, and if it does, with concord.

Government can be more formally defined as a relation between the phrasemes of one category and expression-forms of another. One can think of government as holding when there is a binary partition of the expression-forms of Y into non-null subsets by at least one phraseme of X, one set containing all the forms of Y that can appear after some expression-form of the X phraseme (the positive subset, Y^+) and the other, all those that cannot (the negative subset, Y^-). Crucially, however, Y^+ should not be made up exclusively of complete phrasemes, i.e. $p(Y^+)$ is not a subset of p(Y), so that there is at least one phraseme of Y[12] whose forms are properly partitioned by a phraseme of X. Furthermore, the partition of the expressions of Y must remain constant for all expression-forms in the phraseme of X. In other words, for government to hold, it must be the case that not every form of a phraseme of Y can appear after phrasemes of X and there is no relation of c0-variation between the forms of the expressions instantiating X and those of the expressions instantiating Y. For the purposes of this paper, therefore, government is defined as in (9).

9 Government: X is a governor in X+Y if some phraseme of X induces the same binary partition on expression-forms of Y, irrespective of the expression-form instantiating X and the phrasemes defined on the positive subset of the partition, Y^+, are not a subset of those defined on Y.[13]

12. Again in order to make the definition as wide as possible, I allow government to be defined if only some subset of phrasemes of Y are properly partitioned when in construction with some phraseme of X. Hence, V may be said to govern NP in English merely on the basis of the personal pronouns. If no morphological variation occurs in a language, then there can be no pre-theoretical indication that form-government in this sense exists in that language.

In English, therefore, we can show that V governs NP because transitive verbs partition pronouns into two sets with only the accusative forms being able to appear in such positions (e.g. *control them* but not **control they*, etc.) Furthermore, this partition holds whatever form an expression of X takes and the same forms of pronouns appear after participles and third person singulars (e.g. *controls/controlling them* but **controls/*controlling they*). Hence, V governs NP according to (9).

We are now in a position to see whether government is involved in either of the two controversial constructions we are looking at. Both Zwicky and Hudson tacitly deny that the property is relevant to either of the constructions, as can be seen from tables 2 and 3. They do not, however, provide evidence for this assumption and, in fact, it can be shown from the definition in (9) that Comp governs S and Det governs N.

Complementizers, as mentioned in the previous section, select for the finiteness of their sister sentences, cf. (10). Since complementizers partition sentences into those that contain a finite verb and those that do not and no covariation occurs between forms of Comp and those of S (because complementizers in English are invariant), Comp must govern S.

10a Everyone knows that we control those penguins.
10b *Everyone knows that us to control those penguins.
10c *For we control those penguins is iniquitous.
10d For us to control those penguins is iniquitous.

The traditional wisdom that determiners do not govern their nouns can also be shown to be false on the interpretation of government given here. Consider the grammaticality patterns in (11).

11a the penguin/penguins.
11b some penguin/penguins.
11c a penguin/*penguins.
11d every penguin/*penguins.
11e all *penguin/penguins.
11f many *penguin/penguins.
11g this penguin/*penguins
11h these *penguin/penguins.

This shows that the determiners partition nominal expressions according to their grammatical number and the resulting partition is distributed across the phrasemes of NP. Furthermore, apart from the demonstratives (on which see below) there is no obvious formal sharing of properties (i.e. number) between the determiners and their sister nominal (a point made also in Huddleston (1984) inter alia). Neither *the* nor *some* change their form according to the number of the following noun and with the other non-demonstrative determiners there is no formal resemblance at all between those that take the plural and those that take the singular. It follows from the definition in (9), therefore, that, contrary to the traditional view, determiners must govern the grammatical number of their nominal sisters.

The demonstratives, on the other hand, do not govern their companion nouns because each member of the pairs that make up the two phrasemes {*this,these*} and {*that,those*} partition the set of nouns in different ways, the singular demonstratives selecting singular nouns and the plural ones, plural nouns. The covariation in number between Dem and N is the hallmark of grammatical concord, but to show that this is indeed the case a more precise definition of concord is required.

There are two ways of defining the concord relation using the techniques being developed here. Both cases are marked by the covariation of form of two expressions in a construction, but in the stronger case there is a definite controller of the concord relation along some parameter. As with government, both concord relations involve the binary partition by phrasemes of one category (the controller) of the expression-forms of another (the target, cf. Corbett (1983)). Unlike government, however, different expression-forms of phrasemes of X invoke different partitions of the expression- forms of

13. Assuming that the positive partition induced by a phraseme on some set, Y, is the union of all the positive partitions induced by each of its member expression-forms (and the negative subset is the complement of this set with respect to Y), then we can formally define this notion of form-government as:

$$\exists p \in p(X)[part_p(Y) = \{Y^+,Y^-\} \ \&\ p(Y^+) \not\subset p(Y) \ \&\ \forall e \in p[part_e(Y) = part_p(Y)]].$$

the target category, Y. (This is, of course, what is meant by co-variation.) Since we are interested in co-variation of form, concord must invoke partitions that distribute across the phrasemes of the target category, as with government. The difference between the two types of concord noted depends on which partitions are required to be distributed in this way. In the strong case, defined in (12), the partition invoked by a phraseme of X is required to distribute across the phrasemes of Y. The difference between the strong version of concord control and government, therefore, merely resides in the partitions induced by the expression-forms of some phraseme in X.

12 Concord (strong version): X is a concord controller in X+Y if some phraseme, p, of X induces a binary partition on expression-forms of Y, such that the phrasemes defined on the positive subset of the partition, Y^+, are not a subset of those defined on Y and there are at least two expressions in p which induce different partitions on the expression forms of Y.[14]

For an example of strong concord control, we can turn to Classical Greek. Here, the phraseme (lexeme) *ANE:R* 'man' partitions the set of Greek adjectives into the set of all masculine forms and the set of non-masculine forms, whatever number or case they may show. The masculine forms, however, are themselves partitioned in different ways by each expression-form associated with *ANE:R*, so that, for example, *andras*, the accusative plural, will select all the accusative plural masculine forms while *andri*, the dative singular will select all the dative singular forms, and so on. Since *ANE:R* partitions the phrasemes (lexemes, in the simplest case) of Adj in Classical Greek and its forms impose different refinements on the expression-forms of Adj, N is the strong controller of concord in N+Adj in Classical Greek.

With regard to the Det+N construction in English, however, Hudson asserts that 'the determinant of concord is the noun, and whether this is head or not is immaterial' (Hudson (1987:117)). Leaving aside his denial that concord has anything to do with headedness, a topic that will be taken up in section 3, his assertion that only nouns control concord is not guaranteed by the definition in (12) which makes no reference to specific categories and so for our purposes either the determiner or the noun is eligible to act as a strong concord controller. However, it is clear that neither of these expressions can be classed as strong concord controllers according to the definition in (12). The problem is that while the expression-forms of N (or Det) impose partitions on Det (or N) that distribute across the phrasemes of the latter (i.e. according to number), the phrasemes of the former induce partitions that are subsets of p(Det) (or p(N)). Indeed, for nouns the partition imposed by any lexeme yields Det as either the positive subset (for common nouns) or the negative subset (for personal pronouns).[15] By way of illustration, take the nominal phraseme *PENGUIN* (= {*penguin, penguins*}). Since any determiner can appear with at least one of the expression-forms in this phraseme, the partition induced on Det by the phraseme yields Det as the positive subset and the empty set as the negative subset. Hence, while the expression forms of *PENGUIN* do induce distinct partitions of Det that distribute across at least the demonstrative phrasemes {*this, these*} and {*that, those*}, the phraseme itself does not. Since this holds for all phrasemes in N, N does not strongly control concord in Det+N. Similar reasoning excludes Det from being a strong concord controller here. Since co-variation does occur, however, we may say that both expressions weakly control the concord in Det+N. Weak concord, instead of requiring the partition induced by the phraseme of X to distribute over the phrasemes of Y, merely requires one of the partitions induced by an expression-form in the phraseme of X to distribute over the phrasemes of Y. This is defined informally in (13).

13 Concord (weak version): X is a concord controller in X+Y if some phraseme, p, of X induces a binary partition on expression-forms of Y, such that there are at least two expressions in p which induce different partitions on the expression forms of Y

14. Formally, this comes out as: X is the strong concord controller in X+Y iff.

$\exists p \in p(X)[part_p(Y) = \{Y^+,Y^-\} \;\&\; p(Y^+) \not\subset p(Y) \;\&\; \exists e_1 \in p\; \exists e_2 \in p[part_{e1}(Y) \neq part_{e2}(Y)]]$.

15. In concord, unlike government, we do not require the partition of Y by the phraseme of X to yield two non-null subsets. This can be justified because of the requirement that the different expression-forms of X induce different partitions of Y, so that at least one of the partitions must be non-trivial.

and the phrasemes defined on the positive subset of one of these partitions, Y^+, are not a subset of those defined on Y.[16]

We must therefore conclude that neither Det nor N are strong concord controllers in Det+N, but both of them are weak concord controllers. Given that I am not assuming, contra Hudson, that a concord controller must be nominal, it is possible that either Comp or S (or both) control the concord in Comp+S. Although arguments in favour of agreement with non-nominal items tend to be theory specific, there is some, very slight evidence from English and stronger evidence from other languages that indicates that a (weak) concord relation does exist between complementizers and their sister sentences. The evidence from English depends on a number of assumptions which are open to discussion, but which I shall make for the sake of the argument. In the first place, we must assume that the WH-expressions *who*, *what* and *whom* form a phraseme in the category p(Comp).

Secondly, the reflexives *itself, herself, themselves,* etc., must also be grouped into a single phraseme, despite the fact that taking them to be inflectional variants of each other somewhat extends the notion of inflection. Be that as it may, the above assumptions allow us to define Comp and S as weak concord controllers of each other. The relevant data here are given by those clauses in which there is a dependency in animacy between a WH-Comp and a reflexive inside its sister sentence. Examples are given in (14).

14a Who/*what did Jo say shaved himself twice a day?
14b *Who/what did Jo say scratches itself all the time?
14c Here is the one who/*which Jo says shaves himself twice a day.
14d Here is the one *who/which Jo says scratches itself all the time.

The phraseme {*who, whom, what*} partitions S into those sentences that contain a gap (or lack a noun phrase) and those that do not, but the individual expression-forms will partition S further according to the animacy of any reflexive that is co-referential with the WH-expression. If the reflexives form a single phraseme then sentences which are identical but for the form of any reflexive it contains will also form a phraseme. Hence, by (13), Comp weakly controls the concord in Comp+S and, because sentences containing reflexives will properly partition the phraseme containing *who* and *what*, S also weakly controls the concord in the same construction.

In some languages, there is much stronger evidence for a concord relation between Comp and S where the form of a complementizer varies according to the mood of its companion sentence. For example, Romanian has different forms of the complementizer depending on whether the main verb of the subordinate sentence is subjunctive or indicative, cf. (15.a & b) while in Russian a distinction is maintained between realis and irrealis complements, cf. (15.b & c). (Examples taken from Noonan (1985)). Again weak concord is defined between the co-constituents in these languages, providing more support for the possibility of the relation in English.

15a El spune ca citeste o carte
he says COMP read(3SG INDIC) a book
'He says that he's reading a book'
15b El vrea sa citesca o carte
he wants COMP read(3SG SUBJ) a book
'He wants to read a book'
15c Ja verju cto Boris pridët/prisël
I believe COMP Boris will come/came(INDIC)
'I believe that Boris will come/came'
15d Ja ne verju ctoby Boris prisël
I NEG believe COMP Boris will come/came(SUBJ)
'I do not believe that Boris will come/came'

16. Formally, X is a weak concord controller in X+Y, iff.
$\exists p \in p(X)\ \exists e_1,e_2 \in p[part_{e1}(Y) = \{Y^+,Y^-\}\ \&\ part_{e2}(Y) \neq part_{e1}(Y)\ \&\ p(Y^+) \not\subset p(Y)]]$.

Table 5 sums up the government and concord properties of the two constructions we have been looking at.

Table 5: Government and Concord

	Det	N	Comp	S
Strong Concord	-	-	-	-
Weak Concord	+	+	%+	%+
Governor	+	-	+	-

2.3 The Operational Criteria

Zwicky gives two operational criteria for testing for heads: 'look for the constituent that is distributionally equivalent to the whole construct, and look for the obligatory constituent' (p. 11). Both Zwicky and Hudson disagree, however, over which of the two expressions in each of the two constructions under consideration should be considered as the distributional equivalents and obligatory constituents of their phrases. As shown in Table 1, Zwicky considers that both criteria pick out N and S, while Table 2 shows that Hudson takes the opposite view, that only Det and Comp are selected. Both protagonists provide some reasonable support for their positions, but their arguments are hampered by a failure to give proper definitions for the tests. I will first of all sketch the arguments used by the protagonists and then attempt to give a more rigorous statement of the tests and re-apply these to the two constructions under discussion. In support of his choice of the noun as the distributional equivalent of NP, Zwicky repeats the traditional claim that Det+N phrases have roughly the same distribution as proper names and plural common nouns. But he does point out in a footnote that the issue is not clear cut and that 'the criterion can be used to argue that N is the head, that neither constituent is, or that Det is, depending on what set of facts you look at' (fn. 8, p. 12). Hudson confirms this latter point and provides examples like those in (16) where it is the determiner and not the noun that appears to be distributionally equivalent to the whole phrase.

16a This likes that.
16b This likes that penguin.
16c This penguin likes that penguin.
16d This penguin likes that.
16e ?This penguin likes penguin.
16f *Penguin likes that penguin.

Hudson backs up his claim by alluding to the arguments for treating determiners as pronouns in Postal (1966), Sommerstein (1972) and Hudson (1984). Given such an assumption, the Det+N construction becomes N+N and so, if the distributional equivalent of a noun phrase is a noun, 'then this could be the determiner as easily as the common noun' (p. 123). Without more argument, Hudson then promptly adopts the determiner as the distributional equivalent of the NP. Very similar arguments are applied to the result of testing for the distributional equivalent within Comp+S structures. Zwicky points to grammaticality patterns like that in (17) to support his assertion that S is the distributional equivalent, while Hudson brings examples like (18) to the defence of his decision to take the complementizer as such.

17a I think that the penguins are ready to eat.
17b *I think that.
17c I think the penguins are ready to eat.

18a I haven't seen him since we had that argument.
18b I haven't seen him since.
18c *I haven't seen him we had that argument.

As with some of the other tests discussed above, distributional equivalence has a number of different interpretations with varying degrees of strength. The strongest form may be called recursive distributional equivalence and covers those (few) truly endocentric constructions where an expression of one category can be combined with an expression of another an indefinite number of times. The most obvious example of such a construction is attributive adjective plus noun, where the number of adjectives is not fixed: e.g. *black penguin, fat black penguin, happy fat black penguin* and so on. Although this test does not work for all members of N (it excludes the pronouns) it does work for a large subset and so I propose the definition of the recursive distributional equivalent (RDE for short) that appears in (19).

19 Recursivity: X is the RDE in X+Y if there is some subset, A, of X, such that there are expressions x in X and y in Y such that x is a member of A and x+y is a member of A.[17]

This definition fails to hold for either of the constituents in Det+N or Comp+S. In the former, it is not possible to iterate the nouns for any of the determiners: e.g. **many penguins linguists}*, **a penguin linguist happy singer*. The test fails because there are no noun phrases that form a subset of the determiners. Although it is possible to iterate some determiners (e.g. *the many books, all the three books*) this iteration is strictly bounded and the subparts of the expressions are not necessarily members of N (e.g. *the three books*), so that the noun cannot be the RDE in Det+N, either. For Comp+S, the results are the same. S cannot iterate, as witness the ungrammaticality of **(Jo believes) that we control those penguins those penguins are happy* so that Comp+S is not a member of Comp. It is also not a member of S given the non-iterability of Comp, e.g. **(Jo believes) that when if we control those penguins*. Hence, neither Comp nor S is an RDE for Comp+S.

Other versions of distributional equivalence, however, give different results. Another way of construing the property (and the one that Zwicky and Hudson seem to assume) is to look at the relevant constructions in combination with other expressions and seeing whether the omission of one of the expressions of the former yields a phrase of the same general sort. In other words, one takes expressions x of X and y of Y which combine to give a grammatical expression of category A and one or more expressions of some other category Z (say z) and sees whether z+x+y and z+x are both members of category A. If this is the case, then X can be called the Combinatory Distributional Equivalent (or CDE) in X+Y.

Different constraints may be placed on this sort of distributional equivalence to yield CDEs of different strengths. For example, one could require the property to hold for all categories with which X+Y may combine or, perhaps, for all members of X. Alternatively, we could merely require that there is just some Z that with which X+Y and X may combine or that just some members of X can appear with expressions of Z. These variants produce different results. For example, while some members of Det and N seem able to appear on their own in all environments in which noun phrases can occur (notably, of course, the demonstratives and plural or mass nouns) not all members of Det or N can do so. In Comp+S, on the other hand, sentences can appear without a complementizer in complement position (notably *that*), they cannot so appear in subject position (witness the ungrammaticality of **we control those penguins is obvious)*. In addition, although the traditional conjunctions (*since, because*, etc.) can appear on their own as post-VP modifiers (cf. (18.b)), they cannot do so as pre-sentential ones (cf. (20)).

20a Since the moon was last full, those penguins have been acting funny.
20b *Since, those penguins have been acting funny.

In view of the differences between the results obtained depending on which variables in the test one changes, I shall make a distinction between stronger and weaker CDEs. I shall, however, only make a two way distinction. In the previous tests, a positive result obtained for a subset of the expressions of the category concerned has been taken as positively selecting the whole of that category and this will be maintained for distributional equivalence, for the sake of consistency. The difference in

17. Formally, then X is the RDE in X+Y, iff.
$A \subseteq X$ and $\exists x \in A\ \exists y \in Y[x \in A\ \&\ x+y \in A]$.

strength for CDEs will, therefore, depend entirely on the number of constructions where Z+[X+Y] is equivalent to Z+X. One could give a scalar value for strength according to the number of categories that can instantiate Z for which the property holds. I shall not do so, however, but will only make a binary distinction according to whether the property holds for all categories that X+Y can combine with or just some of them. The definition of CDE with its two variants is given in (21).

21 Combinatory Distributional Equivalent: X is the CDE in X+Y if there are expressions, x of X, y of Y which combine to give an expression of category A, then (a) for every category, Z, there is an expression z in Z which combines with x + y to give an expression of category B such that z + x is also in B, or (b) there is some category Z such that there is an expression z in Z which combines with x + y to give an expression of category B such that z + x is also in B.[18]

From the rehearsal of Zwicky's and Hudson's arguments given above, it should be clear that both Det and N are strong CDEs in their construction, since demonstratives and proper names, plural nouns and pronouns can always stand in for a noun phrase. In Comp+S, on the other hand, both constituents are weak CDEs since there are constructions where neither of the constituents can be omitted. The notion of an expression being an obligatory constituent in a construction also has different interpretations that produce different results. Zwicky, for example, makes a distinction between elliptical constituents which 'must be interpreted from context (linguistic or otherwise)' and 'optionally present constituents [which] do not require such contextual interpretation' (p. 13), but allows only the latter to be relevant for choosing obligatory constituents. Hudson disagrees and adduces nounless NPs and sentenceless subordinate clauses as evidence for nouns and sentences being optional in their respective constructions, despite their apparently elliptical nature.

The question of whether elliptical expressions should or should not be included in the search for obligatory constituents is a tricky matter and one which I shall not go into here. I will, however, take a position closer to Hudson's than Zwicky's and allow those elliptical expressions where the morphosyntactic properties of the whole phrase can be reconstructed from those of one of its elements when in construction with some other expression. So, for example, I would accept phrases like *Mary may* as sentences because one can reconstruct the fact that this is an instance of NP+VP' from the presence of the auxiliary. On the other hand, gapped phrases like *Mary the book* are not obviously sentential and the fact that this might be NP+VP cannot be reconstructed from the second NP, *the book*, alone. I shall assume that the former sorts of expression are included in the appropriate categories (S in the cases under consideration), but the latter, which are totally dependent on context for their grammaticality, are not.

The definition of obligatory constituency parallels that of combinatory distributional equivalence above. Here again, we look at constructions in combination with other categories to see if one of the expressions can be omitted. If it can be omitted, then it is not obligatory and if it cannot, then it is. Furthermore, there are again strong and weak versions of the property, one where the constituent cannot be omitted in any construction and one where it can be omitted in some but not all. The definitions of strong and weak Combinatory Obligatory Constituency (COC) are given in (22).

22 Combinatory Obligatory Constituency: X is a COC in X+Y iff. (a) for all categories Z (strong), or (b) for some categories Z (weak), there are no expressions x+y (x in X and y in Y) of category A that can appear with some expression z in Z to give an expression, z+x+y of category B where z+y is also an expression in B.[19]

---

18. Formally, assuming Z, A and B are categories, then we have (i) for the strong version and (ii) for the weak version:

$\forall Z\ \exists z \in Z\ \exists x \in X\ \exists y \in Y\ [x+y \in A\ \&\ z+x+y \in B\ \&\ z+x \in B]$,

$\exists Z\ \exists z \in Z\ \exists x \in X\ \exists y \in Y\ [x+y \in A\ \&\ z+x+y \in B\ \&\ z+x \in B]$.

19. Formally, X is COC in X+Y iff. either

$\forall Z\ \exists z \in Z\ {\sim}(\exists x \in X\ \exists y \in Y[x+y \in A\ \&\ z+x+y \in B\ \&\ z+y \in B])$, or

$\exists Z\ \exists z \in Z\ {\sim}(\exists x \in X\ \exists y \in Y[x+y \in A\ \&\ z+x+y \in B\ \&\ z+y \in B])$

With respect to Det+N, neither of the co-constituents are strong or weak COCs, because there are no environments in which the demonstratives or plural nouns cannot appear on their own. (This result is, in fact, necessary given that both Det and N are defined as strong CDEs). Comp and S are not strong COCs either because, as we have seen, there are environments in which either Comp or S can be omitted. They are, however, weak COCs because there are environments where the constituent cannot be omitted (e.g. Comp cannot be omitted when complement clauses appear in subject position or where the clause forms an adjunct and S can never be omitted after the non-conjunctional complementizers, *that, for*, etc.).

Hence, as the results summarized in Table 6 indicate, the tests for the obligatory constituent gives almost the reverse results of those for the distributional equivalent. As with a number of the other tests, however, none of the versions of the operational criteria succeed in distinguishing the co-constituents in either construction. Comp is marked as marginal (%) in the table for combinatory distributional equivalence, since this result obtains only if the conjunctions are taken to be complementizers.

Table 6: Operational Criteria

	Det	N	Comp	S
Recursive Distributional Equivalent	-	-	-	-
Strong Combinatory Distributional Equivalent	+	+	-	-
Weak Combinatory Distributional Equivalent	+	+	%+	+
Strong Combinatory Obligatory Constituent	-	-	-	-
Weak Combinatory Obligatory Constituent	-	-	+	+

2.4 Morphosyntactic Locus

Our two authors again disagree about which constituent is the morphosyntactic locus in the two constructions under consideration. Zwicky takes the noun as the morphosyntactic locus of the NP (with respect to number) and S as that for Comp+S (with respect to tense). Hudson, on the other hand, goes for Det and Comp as the morphosyntactic loci of their respective constructions. Again, neither author provides a proper definition of what a morphosyntactic locus is, but it is possible to construct one using the set-theoretic approach used in earlier sections. An expression may be defined as a morphosyntactic locus of a phrase, if its form or the appearance of the expression itself is determined by an expression external to the phrase. Hence, one may test for it by taking a construction Z+[X+Y] and varying the expression instantiating the category being investigated, X. If there is some other expression in that category that cannot replace the original expression and maintain grammaticality, then the category in question is a morphosyntactic locus of the phrase. For example, taking P+NP as the construction under investigation we can see that P is a morphosyntactic locus for the phrase because certain verbs select the particular preposition that must instantiate P in V+[P+NP], as shown in (23).

23a credited with/*to/*from intelligence.
23b given *with/to/*from Jo.

The definition of Morphosyntactic Locus that I shall use in the following argument appears in (24). Using this definition, it is easy to show that, as with many of the other tests, neither of the co- constituents in the constructions are distinguished for this property.

24 Morphosyntactic Locus: X is a morphosyntactic locus in X+Y iff. there are expressions, x and x_1 in X and y in Y such that both x_1+y and x+y are both expressions of category A and for some expression z in Z z+x+y is of category B but z+x_1+y is not.[20]

Zwicky's choice of N as morphosyntactic locus in Det+N has data like that in (25) for [Det+N]+VP in its favour, showing that the number of the whole NP is taken from the noun and not the determiner.

25a The penguin/*penguins is happy.
25b The *penguin/penguins are happy.
25c All water/*penguins is wet.

On the other hand, the fact that there are nouns (e.g. *sheep, fish*) that do not signal number, and there are determiners that do (e.g. *this/these, both*, etc.) lends support for the alternative view that Det is a morphosyntactic locus, as shown in (26). Furthermore, if *we and you}* are to be analysed as determiners in expressions like *we linguists* and *you penguins*, as Hudson suggests, then determiners may signal person distinctions unlike common nouns, cf. (27).

26a This/*these sheep is stupid.
26b Both/*a/*every sheep are stupid.

27a We/*you linguists love ourselves.
27b *We/you penguins are enjoying yourselves.

With respect to noun phrases, therefore, it is abundantly clear that 'the few morphosyntactic markers that there are in the English NP... seem to be distributed fairly equally between Det and N' (Hudson p. 122). Hence, the most that one can fairly say about the English NP is that both Det and N are morphosyntactic loci in Det+N.

Unfortunately, Zwicky provides no evidence for taking the sentence as the morphosyntactic locus of the clause, but presumably one could point to the dependencies that hold between the tenses of the two clauses of a conditional in favour of this analysis, as illustrated in the examples from English in (28). The tense and mood dependencies between main and subordinate verbs that occur in other languages like Latin and Classical Greek might also be taken as further support for this position, cf. the sequence of tense examples in Latin given in (29).

28a If we try/*tried/*were to try to control those penguins, we are mad.
28b If we *try/tried/*were to try to control those penguins, we were mad.
28c If we *try/tried/were to try to control those penguins, we would be mad.

29a Dicebunt ut Cassandra furat/*fureret.
say(fut) that Cassandra be mad(subj.pres/impf)
'They will say that Cassandra is mad'
29b Dicebant ut Cassandra *furat/fureret.
say(impf)
'They used to say that Cassandra is mad'.

Hudson, on the other hand, argues persuasively for his position. In favour of the complementizer as morphosyntactic locus in the clause, he points out that verbs may select for complement clauses with a specific complementizer, as illustrated in (30).

30a Jo believes that/*whether/*if we control those penguins.

As with Det+N, therefore, both Hudson and Zwicky appear to be correct in their choice of morphosyntactic locus in Comp+S. What they are not correct in doing is assuming that only one of the co-con-

20. Formally, X is the ML in X+Y iff.
$\exists x,x_1 \in X\ \exists y \in Y\ \exists z \in Z[x+y \in A\ \&\ x_1+y \in A\ \&\ z+x+y \in B\ \&\ z+x_1+y \notin B]$.

stituents in the construction has this property. As table 7 shows, the test for the morphosyntactic locus of a phrase, like other tests selects both categories in the two constructions under consideration..

Table 7: Morphosyntactic Locus

	Det	N	Comp	S
Morphosyntactic Locus	+	+	+	+

2.5 Proper Government

Before discussing the results obtained so far, I should like to suggest a further test for headedness. This tests picks out those constituents in a phrase that license the unbounded dislocation of co- constituent(s). In a number of theories, e.g. GPSG (cf. Gazdar et al. (1985)) and GB (cf. Chomsky (1981), etc.), the gaps associated with displaced constituents in unbounded dependencies are constrained to be in the domain of certain lexical (subcategorizing) heads. Hence, this property should be able to be used to select heads in phrases and, following the terminology of GB we might call it Proper Government.

Defining this property is not quite as straightforward as for the other tests, as it requires reference to arbitrary strings over the vocabulary of a language. However, if we assume that all the grammatical strings containing unbounded dependencies are sorted into categories like other simple strings, then it is possible to give a fairly straightforward definition. Assume that there are expressions x and y in two categories X and Y, respectively, that combine to form a grammatical expression, x+y, which also forms a substring of some longer grammatical expression, s+x+y, of category A, where s is any non-null string of word-forms in the language that is itself not a member of any category. If it is also the case that y+s+x is a member of some category B and further that s+x is not a grammatical expression of the same category as s+x+y then X can be called a proper governor in X+Y.

31 Proper Government: X is a proper governor in X+Y iff. there is some string of word-forms, s, in the language and expressions x in X and y in Y such that x+y is grammatical, s+x+y is an expression of category A, y+s+x is an expression of category B and s+x itself is not an expression in A and s is not itself a member of any category.[21]

The sentences in adjunct and complement clauses and common nouns in NPs cannot be topicalized or clefted, as illustrated in (32) and (33), respectively. These constructions involve the unbounded dislocation of the sentence or noun and in all cases the resulting expressions are ungrammatical so that by (31) neither Comp nor Det are proper governors.

32a *He'll finish the race, Ronnie doesn't know whether.
32b *Marathon, the foolish lecturer will run a.

33a *(It was) he'll finish the race that Ronnie doesn't know whether.
33b *(It was) marathon that the foolish lecturer will run a.

It seems also that nouns and sentences are not proper governors either, witness the examples in (34).

34a *Whether, Ronnie doesn't know he'll finish the race.
34b *A, the foolish lecturer will run marathon.

Notice, however, that the dislocated demonstrative in (35b) from (35a) gives rise to a grammatical expression as do the subject- auxiliary inversion sentence in (36b) from (36a). However, the constraint in (31) against the constituent with the dislocated constituent and that without it being of the same category excludes (35b) as evidence for common nouns being proper governors, while (36b) is excluded

21. Formally, if E^+ is the set of all non-null strings defined over the vocabulary of the language in question and A, B and C are categories, then X is a proper governor in X+Y iff.
$\exists s \in E^+ \exists x \in X \exists y \in Y[x+y \in C \;\&\; s+x+y \in A \;\&\; y+s+x \in B \;\&\; s+x \notin A \;\&\; \sim(\exists Z \in CAT[s \in Z])]$.

by virtue of the fact that the modal may is a member of Aux and so VP is not defined as a proper governor here.

35a who saw those penguins (Relative Clause).
35b those who saw penguins (Noun Phrase).
35c who saw penguins (Relative Clause).

36a The penguins may swim.
36b May the penguins swim?

Hence, there appears to be no evidence that either of the co- constituents in the two constructions are proper governors, as shown in Table 8..

Table 8: Proper Governor

	Det	N	Comp	S
Proper Governor	-	-	-	-

3. Splitting Heads

Putting together the results obtained from the above tests for the two constructions, Det+N and Comp+S, we get the pattern shown in table 9..

Table 9: Summary

Test:	Det	N	Comp	S
Strong Subcategorizand:	-	+	-	-
Intermediate Subcategorizand:	+	+	%+	-
Weak Subcategorizand:	+	+	+	+
Strong Concord Controller:	-	-	-	-
Weak Concord Controller:	+	+	%+	%+
Governor:	+	-	+	-
Proper Governor:	-	-	-	-
Recursive Distributional Equivalent	-	-	-	-
Strong Combinatory Distributional Equivalent:	+	+	-	-
Weak Combinatory Distributional Equivalent:	+	+	%+	+
Strong Combinatory Obligatory Constituent:	-	-	-	-
Weak Combinatory Obligatory Constituent:	-	-	+	+
Morphosyntactic Locus:	+	+	+	+

It is clear from these results that, despite Hudson’s claims, the tests for headedness considered here are not harmonic and, indeed, apart from failing to uniquely select either of the constituents in the phrase in many cases, the distribution of positive and negative results seem almost haphazard. This would seem to support Zwicky’s somewhat negative conclusion that the morphosyntactic locus is the only key to headedness. Indeed, the fact that the latter test fails to distinguish Det from N or Comp from S seems to suggest an even less positive conclusion: that the tests suggested, far from being indicative of a central unifying grammatical notion, are merely independent properties arbitrarily distributed amongst syntactic constituents.

Such a pessimistic conclusion can, however, be ameliorated by re- arranging the results in Table 9 so that a more harmonious picture emerges, not as tidy as that offered by Hudson, but more so than Zwicky's. The tactic here is to separate out the into three independent groups within which heads may be identified: subcategorization, government and concord together and the rest. I shall argue for the independence of each group from the others and show how, within each group, the tests can be given an harmonic interpretation.

3.1 Subcategorization

In the discussion in section 2.2, I gave a definition of subcategorization with three non-equivalent interpretations. One, the weakest, under the assumption that subcategorization is not restricted to lexical items only, is extremely trivial in the sense that it is true of any two categories that contain phrasemes that may appear together in a construction. The test, therefore, can tell us nothing specific about either constituent as it is never selective and so is not useful as a definition of 'head'. Even if subcategorization is restricted to lexical items, however, weak subcategorization still tells us little about the syntactic status of the word, since all it implies is that the expression may appear in construction with its co- constituent and is lexical. Since the former is trivial and the latter an independent property, the weak version of the test should be excluded from consideration altogether.

This leaves the intermediate and strong interpretations of subcategorization. These are robust tests unlike weak subcategorization, but are independent of the other headlike properties. To see why this is so, consider simple adverbs in English. These subcategorize for (primarily non-auxiliary) verbs or verb phrases and must appear immediately before this constituent, as with *just* in (37).[22] This contrasts with other simple adverbials like *yet* and *soon* which may appear more freely in a string, cf. (38).

37a They may just do it. (Adv + VP)
37b *They just may do it. (Adv + VP')
37c *They may do it just. (S + Adv)[23]
37d *Just they may do it. (Adv + S)

38a They may soon/yet do it.
38b They?soon/*yet may do it.
38c They may do it soon/yet.
38d Soon/*yet they may do it.

These data show that simple adverbials are weak subcategorizands in Adv+VP (37a and 38a) but strong subcategorizands in S+Adv (37c and 38c or Adv+S (37d and 38d).[24] However, such adverbials are not morphosyntactic loci, distributional equivalents, proper governors or obligatory constituents in their constructions. Since this is true in situations in which the adverbials are strong subcategorizands, in order for headedness not to be a contradictory notion, subcategorization must be separated out from the rest of the headlike criteria.

Although this property is independent of the others, we can still pick out heads using it, but they will be heads with respect to this property only.[25] Thus, Comp may be identified as the head of Comp+S with respect to subcategorization because it is an intermediate strong subcategorizand whereas S is not a subcategorizand at all. Simply marking a category as a subcategorizand is not sufficient to identify it as the head of its construction, however, as shown by the result obtained for Det+N. Here N is a strong subcategorizand, and Det an intermediate strong one.

---

22. (36.b) is my own grammaticality judgement which I am aware may not be that of others.
23. Without heavy emphasis on *just*.
24. The categorial analysis of the co-constituents in these phrases is meant to be suggestive only. It is immaterial to the argument whether the (c) sentences are analysed as S+Adv or VP+Adv (order obviously significant) or what name is given to the constituent consisting of Aux+VP.
25. The notion of an expression's being a head relative to some property is one that is suggested for morphology, for example, in di Sciullo and Williams (1987).

Because strong subcategorization logically implies intermediate subcategorization, but not vice versa, we can order the former with respect to the latter and use this to choose between subcategorizing constituents. This enables N to be selected as the head with respect to subcategorization in Det+N because it is a stronger subcategorizand than Det. Thus, heads may be defined not only relative to particular sets of properties but also relative to the strength of the individual properties themselves. The idea of headedness being relative is pursued in later sections and its implications for theoretical approaches are taken up in section 4.

3.2 Form Control

As noted above, Hudson rejects concord as a basis for identifying heads, because he believes that only nominal expressions are involved in this relation. Although I do not share his belief, I consider that concord, together with government, are properties that are not directly linked to the other headlike criteria.

In the first place, weak concord will generally fail to distinguish either of the co-constituents in any construction, since in most cases if X weakly controls the concord in X+Y then so will Y. There is, in fact, only one, very obscure and, I believe, linguistically unlikely, situation in which only one of the co-constituents in a construction will be selected by this test. This would occur where there was some construction in a language where one of the categories governs the other except for the fact that some complete phrasemes of the latter may appear only with some of the forms of a phraseme of the former, but not with others. An example of such a language would be one where, for example, verbs determine the case of the noun phrases with which they combine but where there is (at least) one noun phrase in the language that is not morphologically marked for case but which can only appear with finite forms and not non-finite ones. Even if such a language should exist, which I strongly doubt, the fact that in the majority of cases weak concord control fails to select either of the co-constituents in a phrase is sufficient to exclude it from being a indicator of the headlike status of any expression.[26]

Even if we adopt only the strong version of concord, however, there remain compelling reasons for rejecting it, together with government, as necessarily selecting the expressions selected by the other headlike properties. Given the definitions of the headlike criteria in this paper, it is clear that concord and government (and subcategorization) are very different sorts of properties from the others. The other tests are defined only with respect to the category membership and combinatorial properties of expression-forms, whereas concord, government and subcategorization are defined in terms of partitions of one category by expressions of another. Although this fact does not preclude the latter tests from being directly related to the former, it does seem indicative of the independence of the partitional tests from the distributional ones.

This assumption is borne out by crosslinguistic evidence which strongly indicates that, synchronically speaking, concord and government do not directly reflect the distributional properties of expressions in particular constructions. Nichols (1986) suggests a distinction between what she calls head-marked and dependent-marked languages. In the former, the 'head', independently (but not explicitly) determined, carries information about the properties of its dependents (i.e. non-head co-expressions) as affixes. This is illustrated by the Swahili examples in (39) where the grammatical relations of the noun phrases are indicated by the prefixes on the verb stem and not by markings on the noun phrases themselves.

39a Juma a-li-wa-piga wa-toto
Juma he-past-them-hit pl-child
'Juma hit the children'

26. If the only situation in which weak concord control is selective is not in fact one that is possible in human language, then we can impose a restriction on both forms of concord control so that all phrasemes of Y that are not partitioned by expressions of X either appear in the positive subsets of the partitions induced by all expression-forms of the relevant phrasemes of X or in the negative subsets. This can be achieved by adding the following restriction to the two definitions of concord: where $part_{e1}(Y) = \{Y_1^+, Y_1^-\}$ and $part_{e2}(Y) = \{Y_2^+, Y_2^-\}$, $\forall q \in p(Y)[(q \in p(Y_1^+) \rightarrow q \in p(Y_2^+)) \,\&\, (q \in p(Y_1^-) \rightarrow q \in p(Y_1^-))]$.

39b wa-toto wa-li-m-piga Juma
pl-child they-past-him-hit Juma
'The children hit Juma'

In dependent marking languages, on the other hand, affixes indicating the relation between expressions go on the non-head expressions and are not marked on the head. An example of such a language is Diyari an Australian aboriginal languages where grammatical relations in a sentence are indicated exclusively by markings on the noun phrases, affixes on the verb indicating only mood and tense information. The examples in (40) are taken from Austin (1981) (texts 1.28 and 1.58, respectively).

40a kadi-yali wata dukara-yi
sister's husband-ERG snake-ABS take out-PRES
'The brother-in-law takes out some snake'.

40b paya-li nina kuna-li turipa-yi pita miri
bird-ERG him-ACC faeces-ERG pour-PRES tree above
'Some birds poured shit on him from the tree above.'

In strictly head-marking languages, it will be dependents that either govern or control the concord of the head, whereas in strictly dependent marking languages, either or both properties will indeed coincide with the head. Given that it is not necessary for distributional heads to be governors or to control the concord in their expressions across a range of languages, we must assume that the these properties are independent of each other.

Although the English evidence is scanty, it indicates that at least government does not necessarily follow the pattern of the other tests, as can be seen from table 9. Here, government selects Det and Comp, even though N and S have the same distributional properties and N is selected as a strong subcategorizand. This consideration, together with the facts of crosslinguistic variation mentioned above, yields sufficient grounds for treating the tests for government and concord as independent of the other tests for headedness.[27]

Although separate from the rest of the headlike properties, there are reasons for taking concord and government together as different aspects of the same property, since they both involve the dependency of the form of one constituent on its companion. Furthermore, the definitions given for the properties in this paper entail that they are mutually exclusive: no governor can be a concord controller and vice versa. Hence, these properties are not independent of each other and so we may class them together under the general term 'form control'. To be a head with respect to form control, therefore, an expression must be either a governor or a concord controller, making Det and Comp, as governors, the heads of their respective constructions for this property.

3.3 Distributional Heads

The separation of the partitional tests, subcategorization, government and concord control, leaves us with the tests for morphosyntactic locus, proper governor, distributional equivalent and obligatory constituent. Table 9 shows, however, that these tests do not all necessarily converge on particular constituents. No constituent is positively marked for all the headlike properties and their distribution is not identical in the two constructions. It might seem, therefore, that each property should be taken separately, thus reducing the pre-theoretical notion of head to vacuity. However, although Hudson's assertion that all the headlike tests are harmonious cannot be sustained, Zwicky's pessimism about the utility of these tests seems to be unjustified. This final group of properties do not in fact seem to be independent of each other, as they are independent of subcategorization and form control, but the relations between them are implicational, not conjunctive.

Although Hudson claims that the headlike criteria are independent properties, this is not strictly true for some of the definitions that I have proposed. For instance, the strong forms of Combinatory Distributional Equivalence and Combinatory Obligatory Constituency logically imply their weak

27. This is not to deny, however, that certain historical changes cannot be explained with reference to the interaction of the two sets of properties, merely that there is no necessary synchronic relation between them.

forms and Recursive Distributional Equivalence implies strong Combinatory Distributional Equivalence. Furthermore, it can be shown from the definitions given that no proper governor can be recursively distributionally equivalent to the whole phrase (since s+x must be of the same category as s+x+y if x and x+y are themselves of the same category) and if both constituents in a phrase are strong CDEs then neither of them can be weak COCs, and so not strong ones either, as readers can test for themselves. There are, therefore, certain logical connections between some of the tests and their variants. It seems also that there are, not logical, but linguistic, relations that hold between the remainder of the tests which may be partially accounted for by appealing to pre-theoretical ideas.

To begin with, proper government appears to be dependent on the obligatory status of a constituent in a construction. Intuitively, constituents have the property of being proper governors because the category of the gap associated with the dislocated constituent can be recovered from the presence of the proper governor alone. For this to be the case, the governing constituent must be obligatory (in the sense used above), otherwise it ought to be possible to extract a constituent without the proper governor being present. This would make the position associated with a dislocated constituent indeterminate and so greatly increase the processing load. Hence, we may hypothesize any proper governor must also be an obligatory constituent.

Other relations might well exist between (strong) obligatory constituency and (strong) distributional equivalence and the morphosyntactic locus. If a phrase consists of expressions of two categories where one must always be present whilst the other need not, then any formal dependencies between the phrase and a category it combines with must be realized on the obligatory one. A relation between distributional equivalence and morphosyntactic locus is also intuitively necessary given that a category is only distributionally equivalent to the whole phrase if it can appear in constructions without its co-constituent. Hence, any formal dependencies must be realizable on expressions of this category.

We can, therefore, postulate the hypotheses in (41) which indicate a strong interrelationship between the distributional tests for headedness.

41a If X is a proper governor, then X is a weak obligatory constituent.
41b If X is a strong obligatory constituent, then X is a morphosyntactic locus.
41c If X is a strong distributional equivalent, then X is a morphosyntactic locus.

These hypotheses are, as I have said, not logically necessary but empirically based and so falsifiable on that basis and we can provide an initial small test for these by reviewing the distribution of the properties for the four of Zwicky's six English constructions that have not yet been examined. Table 10 presents the results obtained from applying my definitions of the four criteria to the constructions: V+NP, P+NP, NP+VP and Aux+VP. Although I do not have space to argue for the results shown, the patterns must be correct if it is not presupposed that the test selects at least and at most one constituent in a construction. In most cases, the relevant data can be gleaned from the arguments in the cited articles by Zwicky and Hudson and the only peculiarities are with the properties ascribed to the noun phrases in V+NP and NP+VP. In the former, NP is marked as morphosyntactic locus because of possible reflexive dependencies exhibited between subjects and objects (cf. *The penguins admired themselves/*itself*, etc.). In the second construction, the subject NP is marked as a weak COC because there are situations in which a following VP cannot be omitted (e.g. after complementizers, cf. *I believe that penguins *(are wonderful)*).

Table 10: Further Results

Test:	V	NP	P	NP	Aux	VP	NP	VP
Strong Subcategorizand:	+	-	+	-	+	-	-	-
Proper Governor:	+	-	+	-	-	-	-	+
Strong Combinatory Distributional Equivalent:	+	-	-	-	+	+	-	-
Weak Combinatory Distributional Equivalent:	+	-	+	-	+	+	-	+
Strong Combinatory Obligatory Constituent:	+	-	-?	-	-?	-	-	+
Weak Combinatory Obligatory Constituent:	+	-	+	-	+	-	+	+
Morphosyntactic Locus:	+	+	+	-	+	+	-	+

The results shown in Table 10 clearly bear out the hypotheses in (41) so that we are justified in using these tests together to identify the distributional head of a phrase.

The results in Table 10 also show that in each phrase one of the constituents has more of the headlike properties than the other and the constituent so picked out is the one generally assumed to be the head, i.e. V in V+NP, P in P+NP, Aux in Aux+VP and VP in NP+VP. A head may, therefore, be defined as that constituent in a construction that does not have fewer of the headlike properties than its co- constituent(s). Distributional headedness, therefore, is defined relatively as in (42).

42 Distributional Heads: X is a distributional head in X+Y iff. X does not have fewer of the distributional head properties than Y.

Given this definition, it might seem that, although the combination of different headlike properties is constrained for a particular category in a phrase (i.e. by the hypotheses in (41)), the distribution of these properties across co-constituents within a phrase is arbitrary. Heads are defined merely with respect to the number of properties displayed by the co-constituents, but has nothing to say about the relation between the properties exhibited by heads and non-heads. If there were no constraints on this relation, then constructions could occur in which one constituent, for example, is marked as obligatory, distributionally equivalent and a morphosyntactic locus while the co-constituent is marked as a proper governor (and hence, a weak obligatory constituent). This and the other possible arbitrary combinations of properties would, if attested, seriously weaken any claims the 'head' has to being the central constituent of a phrase and having any predictive power at all. However, as Table 10 shows, the distribution of the headlike properties across constituents does not seem to be arbitrary. In fact, in every case, the non-head only displays properties already associated with the head. This leads me to propose the hypothesis in (43) which strongly constrains the properties that can be associated with a non-head, given those of the head.

43 Harmonic Head Hypothesis: If Y does not have more headlike properties than X in the construction X+Y, then X has all the properties associated with Y.

For the two constructions that have formed the central discussion point in this paper, the definition in (42) yields an interesting result: both co-constituents in Det+N and Comp+S are heads, as both exhibit the same number of distributional properties. Furthermore, the hypothesis in (43) is borne out because not only do each of the co- constituents in the two constructions show the same number of properties, but the properties themselves are the same for both, as predicted. This result, therefore, confirms the pre-theoretical existence of doubly headed constructions, the implications of which will be taken up in detail in the next section. This leaves us with the distribution of the properties in Table 11 and the picture of headedness for our two main constructions shown in Table 11. The fact that different constituents are picked out by different properties clearly illustrates the non-harmonic nature of the relation between the different primary headlike properties. Again this is a topic that will be taken up below..

Table 11: Summary of significant properties

Test:	Det	N	Comp	S
Strong Subcategorizand:	-	+	-	-
Intermediate Subcategorizand:	+	+	%+	-
Governor:	+	-	+	-
Strong Concord Controller:	-	-	-	-
Proper Governor:	-	-	-	-
Strong Combinatory Obligatory Constituent:	-	-	-	-
Weak Combinatory Obligatory Constituent:	-	-	+	+
Recursive Distributional Equivalent	-	-	-	-
Strong Combinatory Distributional Equivalent:	+	+	-	-
Weak Combinatory Distributional Equivalent:	+	+	%+	+
Morphosyntactic Locus:	+	+	+	+

Table 12: Heads

Test:	Det + N	Comp + S
Subcategorization:	N	Comp
Form Control:	Det	Comp
Distribution:	=	=

4. Theoretical Implications

In the above discussion, I have provided set-theoretic definitions of seven headlike properties with variants and applied these to two English constructions where the head of the phrase is in dispute, i.e. Det+N and Comp+S. The results obtained led to the postulation of three primary hypotheses concerning headedness: firstly, that the seven tests should be divided into three independent groups, subcategorization, form control and distribution. Secondly, it was suggested that heads can be defined within these groups according to the relative number of properties or their variants the co- constituents have. For subcategorization, this induces a hierarchical relation between strong, intermediate and non-subcategorizands and for form control, the head is either a governor or a concord controller. Distributional heads are defined relative to the number of properties exhibited by the constituents: if no other constituent has more of the headlike properties than the one under consideration then it is a head. Limits on which distributional properties a constituent may show are constrained by certain logical and non-logical relations that obtain between them.[28] In order to capture the idea that heads are central to their phrase, I have proposed that the relation of a co-constituent to a head be constrained by disallowing the former from showing different headlike properties from the latter. This implies that complements are indeed subordinate to heads as pre-theoretically defined since they are dependent on them with respect to the properties they may show.

This analysis, therefore, provides an account of headedness that is 'pre-theoretical' in the sense that it presupposes no particular syntactic theory and utilizes concepts from set theory and ideas about category membership as its basis, both tools that are foundational to most theories of syntax.[29]

By relativizing the notion of head to particular groups of properties, to the strength of the properties within these groups and to the distribution of these properties across co-constituents, I have provided a harmonious and, I believe, robust pre-theoretical definition of headedness. Although the discussion has concentrated almost exclusively on English, the tests provided are general enough to be applicable to other languages. No presuppositions are made that different languages have the same categories and constructions as English, since none of the tests makes reference to any named category of expressions. The tests merely require an initial sorting of expressions into sets according to their similarity in distribution, something that must be done before any syntactic analysis can be carried out. Thus, the tests are in principle applicable to any human language and so allow heads to be identified for crosslinguistic generalizations to be made. Nichols' confident claim that 'head' is a 'non-theoretical' concept therefore receives some support.

There are, however, at least two problems that the proposed analysis poses for theories of syntax that utilize the notion of 'head'. One is a general methodological problem, concerning the relationship between heads as pre-theoretically defined and as employed within a theory. The other is more specific and concerns the implications for particular theoretical approaches of the existence of constructions that contain two constituents which both have the same distributional head properties. In this last section, I shall examine these two problems, beginning with the latter. Because of space, the discussion will necessarily be curtailed and there will be aspects of the topics and theories that I do not touch upon. The arguments should, therefore, be viewed more as sketches than fully executed pictures, and I hope that the interested reader can fill in some of the blank spaces for herself.

4.1 Doubly Headed Constructions

Most syntactic theories, particularly those adopting some version of X-bar theory, assume that heads are unique in any construction. However, the results obtained above for the distributional tests applied to the English Det+N and Comp+S constructions show phrases may have two heads with identical properties (necessary by the Harmonic Head Hypothesis). The fact that Zwicky's and Hudson's results for the two phrases Det+N and Comp+S contradict each other can therefore be ascribed to their insistence that each construction contains only a single head.

The recognition that constructions may contain multiple heads is not new and has been suggested, for example, for coordination constructions (e.g. Sag et al. 1985) and, indeed, is asserted in Warner (1989) to be the case for just the structures I have been examining. Vincent (1988) makes a distinction between lexical heads which are of open class, major categories (e.g. N, V) and functional heads of closed class, minor categories (e.g. Det, Comp). The double headed constructions that we have been considering, therefore, can be accounted for at some level if one of the heads in a doubly headed structure is functional.

Although this generalization seems to be borne out across a range of constructions, and has some force from a historical perspective, the existence of doubly headed constructions has significant consequences for synchronic theories that use the notion of head to derive the syntactic properties of phrases.

28. Although these relations induce only a partial ordering on the distributional properties, they imply a hierarchy of strength with proper government being strongest, obligatory constituency next, then distributional equivalence and finally the morphosyntactic locus as the weakest.

29. It may be argued that my assumption of a constituency based syntax rather undermines my claim for theory-independence. However, given the weak equivalence that obtains between dependency and constituency analyses (cf. Matthews (1981)), the definitions I have given for the headlike properties should be convertible into the basic terminology of dependency theory without alteration to the results obtained.

There are three main areas where the notion of head is used in syntactic theory to determine grammatical properties: to generate structure, to account for feature distribution and to determine word order. All three areas rely to some extent on the uniqueness of heads within phrases and so must be revised to account of multiply headed structures. In each case considered, either the power of the head must be weakened to incorporate these structures or any direct relation between pre-theoretical and theoretical uses of the notion must be denied.

In feature based theories like GPSG and HPSG, 'head' is used as a primitive notion whose primary function is to enable features to be passed around a phrase structure tree in an economical and maximally general fashion.[30] The Head Feature Convention (HFC) of GPSG (or the Head Feature Constraint of HPSG) determine the inheritance of grammatical properties (encoded as syntactic features) by the head daughters from the mother category. In the earliest formulations of the principle, and the one adopted for HPSG in Sag (1987), the categorial specification of the head daughter is identical (modulo bar levels in GPSG and subcategorization information in HPSG) to that of its mother. Given constructions like Det+N or Comp+S that have two heads with different categorial definitions[31], such a strong version of the HFC must fail, unless the headlike status of the functional expression is ignored. To try to accommodate multiply headed constructions, later versions of the HFC within GPSG have suffered successive weakenings. In the version in Gazdar et al. (1985), a mother's category is constrained to be given as (a subset of) the intersection of the free feature specifications of all its head daughters. This revision, inspired by coordination facts, would not work for doubly headed structures such as Det+N and Comp+S, unless Det and Comp were also marked as N and S, respectively (cf. fn. 30). To get around this problem, Warner (1989) suggests a further alteration to the HFC to allow situations where the mother's category is essentially given as the unification of the head features of functional and major head daughters.[32] In order to ensure that the complementizer (he does not discuss noun phrase structure) does not inherit other properties from the mother like tense, Warner imposes a Feature Co-occurrence Restriction to the effect that a minor category cannot contain such features. While Warner's solution appears to work within the confines of the theory he adopts, the revision of the HFC he proposes and the necessity of imposing ad hoc constraints on feature co-occurrence to account for doubly headed constructions seriously weaken the case for using the notion of head to control feature percolation.

The second topic I wish to take up is that of word order. There have been many attempts within typological studies to reduce certain universal tendencies in word order to relations between heads and complements and there is a good deal about 'head-initial' and 'head- final' languages in the typological and theoretical literature. Unfortunately, attempts to bring these ideas within the bounds of a formal theory have been less than successful. The variability in the ordering of heads with respect to their dependents means that linearization rules based on heads are problematic and the problems are compounded by the recognition that some structures are doubly headed. If, for example, Det and N are both heads of their phrase and on the same level, their ordering must be defined by something other than the fact that they are heads. Interestingly enough, early suggestions within GPSG that ordering should make reference to heads (cf. Gazdar and Pullum (1981)) were dropped very quickly and in Gazdar et al. (1985) ordering is defined solely in terms of categorial properties.

---

30. In the former theory, it is also used to restrict the domain of application of metarules (cf. Gazdar et al. (1985)). Here, the recognition that Det is a head means that a metarule like that introducing a trace can apply to the rule expanding NP as Det+N, yielding ungrammatical sentences like those in (32.b and d), above. The restriction on metarule application needs, therefore, to be revised to apply to rules introducing only subcategorizing lexical heads, cf. Warner (1989).

31. Hudson, following suggestions by Lyons (1975), Postal (1966) and Sommerstein (1972), puts forward the view that determiners are 'in fact a subset of pronouns' (p.123) and, therefore, a subset of nouns. This would allow the two heads in NP to both be realized as N. But, whatever the merits of this hypothesis from a historical point of view, its synchronic status in contemporary English is far from certain. Even if we accept it, however, this leaves the problem of complementizers which it would be difficult to argue are a subcategory of sentences.

32. A very similar proposal was made in a series of papers delivered to various meetings beginning with Cann (1984).

Hudson (1984,1987) provides a partial theory of word order that determines the limits of how groups of words relate to each other linearly. He calls this the Adjacency Principle which he defines in terms of heads and dependents in the 1987 version given in (44).

44 The Adjacency Principle: If A is the head of B and some word C separates them, then it must be the case that (i) C is subordinate to A or (ii) that C is (subordinate to) the head of both A and B (Hudson 1987:127).

Together with independent ordering principles this can be used to account for why, given that *great* and *with* must precede *difficulty* the only possible order is *with great difficulty* and not **great with difficulty*. The problem, of course, is that if, for example, Det and N are both heads and independent rules ensure that *the* and *happy* precede *penguins* then the Adjacency Principle will still allow the generation of **happy the penguins*. Further problems arise with Comp+S where (if I understand the concepts correctly) taking the two expressions as co-heads will allow the derivation of **we that control penguins* where this a complement, rather than a relative, clause. To obviate this, it is necessary either to ignore the pre-theoretical evidence for the status of these constructions as doubly headed (as Hudson, in fact, does) or to have recourse to additional information, independent of headedness, to ensure the correct order of expressions.

Thirdly, X-bar syntax as used in current transformational theory (cf. Chomsky (1986), (1988), etc.), makes the syntactic head the central generator of the structure of the phrase that contains it. Each 'projection' of the head is associated with co-constituents which have particular properties: a lexical head is associated with its subcategorized (or θ-marked) complements[33] while its projection, X', is associated with a specifier or modifier. Thus, each head is minimally associated with the structure implied by the schemata in (45).

45a X" → Specifier, X'
45b X' → Modifier, X'
45c X' → Complements, X

There is an obvious problem with the generability of this structure where a construction contains more than one head, but a solution has emerged in recent years by extending the applicability of the schemata in (44) from the major categories (N, V, P and A) to the minor ones: Comp, Infl (the old Aux node) and Det (cf. e.g. Chomsky (1986), Fukui and Speas (1986)). This extension is not fully generalizable, however. The appearance of complements of major lexical items is generally considered to be determined by subcategorization properties which are independent of the X-bar schema like Case theory or Theta Theory (cf. Stowell (1981)). With the minor categories, on the other hand, the category of their complements needs to be specified directly, i.e. Comp must have a sentential (= IP) complement, Infl must be followed by VP and Det by NP. Hence, minor heads have different and more restricted collocational properties than major ones.

Furthermore, it is not certain that both major and minor categories induce the same structures. For example, it is commonly believed that VP does not have a specifier position (although for a different view cf. Manzini (1983)) whilst its associated Infl projection does. There may also be differences in terms of modifier adjunction, disallowed for minor categories but permitted for major ones (cf. Cann and Tait (1989)). If these observations are correct, then the X-bar schema must differentiate between functional and lexical heads, thus compromising the predictive power of the notion 'head' and the explanatory generality of the theory.

Even if the theory does not need to be weakened or if the revisions are considered trivial, the solution proposed within X-bar theory to handle doubly headed constructions does not really reflect the conclusion reached from an application of pre-theoretical tests for heads. The results obtained above indicate that Det and N and Comp and S have the same distributional properties and that both therefore constitute heads at the same level. The generalization of the schemata in (45) to the minor categories does not really capture this observation since NP (equivalent now to a common noun phrase) is subordinate the determiner, as S (IP) is subordinate to Comp. Thus, in fact, Det and Comp must function

33. Note that this assumes a theoretical connection between subcategorization and distributional properties, despite the fact that this relation is not pre-theoretically necessary.

uniquely as heads at the level of combination with their complements, so that the double headedness implied by the pre-theoretical tests is lost. The selection of heads within the theory, therefore, appears to be a theory-internal matter, only marginally related to the identification of heads using pre- theoretical ideas. Indeed, in a recent, as yet unpublished, paper, Chomsky (1988), following Pollock (1989), expands the number of minor categories to cover Agr(eement), Neg(ation), Tense and Asp(ect). Each of these categories projects according to the X-bar schemata and has an associated complement (whose category is again necessarily specified ad hoc), but no apparent specifier position. Since these positions are either empty in many languages (e.g. object agreement in English) or are filled by bound morphemes, it becomes clear that the notion of head being utilized here is not based on any pre-theoretical ideas about what properties heads have. There is no sense in which one could provide a pre-theoretical set of criteria for identifying heads in a language that are never phonetically realized or only realized by a bound morpheme. The identification of the heads in these instances must therefore be purely a matter of theoretical concern and the decision to take some element as a head is determined on purely theoretical grounds.

The existence of doubly headed constructions, therefore, poses problems for theoretical uses of the notion head that require considerable weakening, either of the grammatical power of headedness, or of the relation between the pre-theoretical and theoretical concepts of the notion. In either case, the theory independence of headedness is compromised and doubt is cast on any claim that the same concepts underlie every theoretical utilization of the term 'head'.

5. Heads and Theory

Although the notion of 'head' has undeniably had a good deal of success in typological studies and within particular theories in accounting for a wide range of data, the above discussion of the implications of the existence of doubly headed constructions highlights the problem of the relation between the pre-theoretical notion and the way this is developed theoretically. This is further compounded by some of the other implications for theory of the approach to heads developed in this paper. In the first place, although the tests for heads may be made harmonic within particular groups, as discussed above, Table 11 shows that the results obtained by the three groups may be disharmonic: in Det+N, for example, N is the head with respect to subcategorization while Det is the head with respect to form control and both are heads with respect to the distributional tests. This shows that, contrary to the assumptions of Zwicky and Hudson (amongst others), a single notion of head is not sufficient in syntactic theory. At least three sorts of head must be recognized and this poses problems for those theories that unify the properties into one concept as in HPSG and, to a great extent, X-bar theory.[34] In these theories, notions of theoretical headedness are again divorced from pre-theoretical ones, leaving the nature of the relation between them obscure.

Secondly, the fact that the distributional tests are not conjunctive and that heads may exhibit various combinations of the properties these identify makes the relationship between head as a theoretical primitive and the pre-theoretical criteria, problematic. The tests for heads as I have defined them are obviously syntactic, based on distribution and categorial membership and so a syntactic theory should be able to provide some way of characterizing the properties they reflect. Hence, if headedness is adopted as a theoretical primitive, and if heads are identified using the pre-theoretical criteria under discussion, then the theory should be able to explain the relationship between heads and the pre-theoretical tests used to identify them. In the simplest case, the theory should in fact predict the syntactic properties of a head, including whether it is a proper governor, an obligatory constituent, distributional equivalent and/or a morphosyntactic locus. If the number and type of these properties may vary from head to head, however, it is hard to see how a theory can make this prediction without further specifying other properties of the constituent in question to explain the variation. If the theory cannot predict which of the distributional properties a head will have, then again the relation between the theoretical notion and its pre-theoretical underpinnings becomes opaque.

34. GPSG does treat the three properties separately. Distributional heads fall within the domain of the HFC while subcategorization properties are introduced within rules themselves and not necessarily on heads and form control (or at least concord) is accounted for by the Control Agreement Principle.

A final problem in the relation between the headlike properties and the theoretical notion of 'head' lies in the fact that non-head constituents may show similar properties to their heads. If the point of adopting the notion 'head' as a theoretical primitive is to constrain certain syntactic processes like feature percolation, the distribution of morphosyntactic loci across a phrase seriously compromises the effectiveness of this decision. It will thus be necessary to allow feature percolation onto non-heads using some other means, thus weakening the motivation for the constraint in the first place. Indeed, the need for the notion of head becomes weaker the more properties heads and non-heads share, culminating in the doubly headed constructions where the notion has very little import, since it does not differentiate the behaviour of either of the co-constituents. Considerations like these have led linguists to assume that the relation between the headlike properties and the notion 'head' is only a 'very strong tendency' (Hudson (op.cit.:125)). But if this is true, then the claim that 'head' is a theory neutral notion loses most of its validity. Although the identification of tendencies is useful in typological and historical work and may be of some help in guiding theoretical research, unless they can be quantified and formally explained they cannot form the basis of a syntactic theory. Any correlation between syntactic properties and a theoretical syntactic primitive is arbitrary, unless the theory explains that relation. My point is that the methodological problem alluded to by Hudson, even if not peculiar to the notion 'head', is a significant one for attempts to justify theoretical concepts by appealing to pre-theoretical properties. If there is no way of linking heads with the pre- theoretical properties supposedly used to identify them, then one must conclude that there is no necessary relation between them and that their apparent interdependence is illusory.

There seem to be two ways out of the methodological dilemma posed for theoretical notions of 'head' by the variability in the properties associated with heads and the existence of doubly headed constructions. In the first place, the relationship between the pre- theoretical criteria and any theoretical notion of head can be denied. The identification of heads must therefore be a theory-internal matter only and no justification for the adoption of particular choices can be made using pre-theoretical notions. The apparent relation between theoretically identified heads and any pre-theoretical properties becomes an accidental one and claims for the theory-independence of the notion, like that made by Nichols quoted at the beginning of the paper, must be abandoned.

Alternatively, one could strengthen the position of the pre- theoretical criteria vis-a-vis the notion of 'head' by taking the latter as a derived notion and not a primitive one. Thus, the pre- theoretical criteria may themselves be given theoretical status and the notion 'head' derived from their interaction.[35] The power of such a derived notion of head will, however, necessarily be more restricted than the notion used as a primitive, since it obviously cannot be used to account for the properties used to define it without circularity.

Whichever solution is adopted (and there may well be others), confident assertions about the theory independent status of headedness and its apparent straightforwardness are clearly not tenable. Given that fairly robust and precise definitions can be given to pre- theoretical concepts of head, as I hope to have demonstrated, more serious attention should be given to the relation between these and their theoretical counterparts. It is not possible to assume without argument that theoretical and non-theoretical ideas of headedness have any direct relation. If the notion 'head' is to continue to be a popular and fruitful one within linguistic theory, it is essential that its status with regard to theoretical and pre-theoretical concepts be resolved. Without such a resolution, syntacticians may find themselves building strong hypotheses on very dubious foundations.

References:

Austin, P (1981) *A Grammar of Diyari, South Australia*. Cambridge, Cambridge University Press.

Cann, R (1984) Heads without bars. Paper delivered to the Autumn Meeting of the LAGB.

Cann, R (1988) Heads, modifiers and governors. Paper presented to the Spring Meeting of the LAGB, University of Durham.

35. Such a position is developed in Cann (1988) within a version of categorial grammar, where the implicational relations between the distributional properties are shown to follow from general combinatorial considerations.

Cann, R and M E Tait (1989) Free relatives revisited. Paper presented to the spring meeting of the LAGB, Queen's University, Belfast.

Chomsky, N (1981) *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Chomsky, N (1986) *Barriers*. Cambridge, Mass., MIT Press.

Chomsky, N (1988) Some notes on economy of derivation and representation. Unpublished ms., MIT.

Corbett, G (1983) *Hierarchies, Targets and Controllers.* London, Croom Helm.

Cruse, D (1986) *Lexical Semantics.* Cambridge, Cambridge University Press.

Fukui, N and M Speas (1986) Specifiers and Projection. In N Fukui et al. (eds.) *MIT Working Papers in Linguistics, 8*. MIT, Cambridge, Massachusetts.

Gazdar, G and G K Pullum (1981) Subcategorization, constituent order and the notion 'head'. In M Moortgat, H v.d. Hulst and T Hoekstra (eds.) *The Scope of Lexical Rules.* Dordrecht, Foris, 107-123.

Gazdar, G, E Klein, G K Pullum and I A Sag (1985) *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.

Huddleston, R (1984) *Introduction to the Grammar of English* Cambridge, Cambridge University Press.

Hudson, R (1980) Constituency and dependency. *Linguistics* 18, 179-198.

Hudson, R (1984) *Word Grammar*. Oxford: Basil Blackwell.

Hudson, R (1987) Zwicky on heads. *JL* 23. 109-132.

Jackendoff, R (1977) *X-bar Syntax*. Cambridge, Mass., MIT Press.

Lyons, J (1975) Deixis as a source of reference. In E Keenan (ed.) *Formal semantics of Natural Language*, Cambridge, Cambridge University Press, 61-83.

Lyons, J (1977) *Semantics*. Cambridge, Cambridge University Press.

Matthews, P (1981) *Syntax*. Cambridge, Cambridge University Press.

Nichols, J (1986) Head-marking and dependent-marking grammar. *Lg*. 62. 56-119.

Noonan, M (1985) Complementation. In T Shopen (ed.) *Language Typology and Syntactic Description II: Complex Constructions*. Cambridge, Cambridge University Press. 42-140.

Pollard, C and I A Sag (1988) HPSG. CSLI Publications, Stanford University.

Pollock, J-Y (1989) 'Verb movement, universal grammar and the structure of IP'. *Linguistic Inquiry* 20, 365-424.

Postal, P (1966) On so-called 'pronouns' in English. In Reibel, D and Schane, S (eds.) *Modern Studies in English*. Englewood Cliffs: Prentice Hall.

Quirk, R, S Greenbaum, G Leech and J Svartvik (1985) *A Comprehensive Grammar of English}*. Longman, London.

Sag, I A (1987) Grammatical hierarchy and linear precedence. In G J Huck and A E Ojeda (eds.) *Syntax and Semantics 20: Discontinuous Constituents.* Academic Press, New York. 304-342.

Sag, I A, G Gazdar, T Wasow and S Weisler (1985) Coordination and how to distinguish categories. *NLLT* 3.

di Sciullo, A-M and E Williams (1987) *On the Definition of Word*. MIT Press, Cambridge, Mass.

Sommerstein, A R (1972) On the so-called definite article in English. *LIn* 3. 197-209.

Stowell, T. (1981) Origins of Phrase Structure; unpublished doctoral dissertation, MIT.

Vincent, N (1988) Heads and history. Paper presented to the Autumn Meeting of the LAGB, University of Exeter.

Warner, A (1989) Multiple heads and minor categories in generalized phrase structure grammar. *Linguistics 27.*

Zwicky, A (1985) 'Heads'. *JL* 21. 1-30.